Correct Identification Decisions:

How to Obtain Them and How to Judge Them

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Philosophie des Fachbereichs 06

der Justus-Liebig-Universität Gießen

vorgelegt von

Kristina S. Kaminski

aus Linsengericht

2016

Dekan: Prof. Dr. Dr. Jürgen Hennig

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

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

Tag der Disputation: 14. Juli 2016

DANKSAGUNG

Mein besonderer Dank gilt meinem Betreuer Prof. Siegfried L. Sporer, Ph.D., der es mir ermöglicht hat, im spannenden Bereich der experimentellen Aussagepsychologie tätig zu werden. Seine fachliche Unterstützung und seine stets hilfreichen und kritischen Rückmeldungen haben deutlich zum Gelingen der vorliegenden Arbeit beigetragen. Bei Prof. Dr. Joachim Stiensmeier-Pelster bedanke ich mich für sein allseits offenes Ohr und die universitäre Unterstützung. Valerie Hauch, Emma Halfmann, Dr. Kerstin Kocab und Dr. Sarah Tandler danke ich für die vielen hilfreichen Diskussionen, den fachlichen Austausch und die gute Zusammenarbeit. Für die großartige Hilfe bei der Erstellung der Stimulusmaterialien und den umfangreichen Datenerhebungen möchte ich mich ganz besonders bei Maike Davids, Dr. Dawn McQuiston, Sarah Lichtenthäler, Roman Briker, Monja Lehnigk, Raffaela Witting, Natalia Gadek, Valerie Erkens, Katja Böhme, Emilie Eckstein, Fabian Becker, Peter Mross, Sebastian Trentepohl, Andreas Reis, Svenja Haußner, Zita Pielka, Julia Kettenring, Christine Steinmetz, Benjamin Sklarek und Monique Winkler bedanken. Bei Reinhard Bentrup und Prof. Dr. Colin Tredoux bedanke ich mich für ihre kompetente Beratung bei komplexen statistischen Fragestellungen.

Zuletzt gilt mein besonderer Dank meinen Eltern Birgit und Fred Kaminski, meinem Bruder Tim Kaminski und meinem Freund Stephen Braun, die mir in jeder Phase der Erstellung der vorliegenden Arbeit zur Seite standen und mich stets emotional unterstützt haben. Stephen Braun danke ich besonders für seinen unermüdlichen Optimismus und seine Lebensfreude, die er immer wieder auf mich übertragen konnte.

Linsengericht, im April 2016 *Kristina Kaminski*

ABSTRACT

False eyewitness identifications as well as erroneous judgments of identification decisions in court have been shown to contribute to wrongful convictions in many criminal cases (e.g., Garrett, 2008, 2011, 2012). The present dissertation reports three experiments, which were aimed to investigate a new system variable (Wells, 1978) to increase identification accuracy (Experiment 1) as well as to analyze fact finders' judgmental processes when evaluating the accuracy of an identification decision (Experiments 2 and 3).

In Experiment 1, potentially beneficial effects of re-reading one's own person description prior to the identification task were examined (cf. Cutler, Penrod, O'Rourke, & Martens, 1986; Sporer, 2007). Moreover, description effects on identification accuracy were investigated under more realistic retention intervals, expecting a verbal facilitation effect instead of verbal overshadowing (Schooler & Engstler-Schooler, 1990). Participants watched a video film (1) without describing the perpetrator (no description control group), (2) with describing the perpetrator or (3) with describing and re-reading their own description prior to the identification task. Two days or five weeks later all participants were asked to identify the perpetrator in a target-absent or target-present lineup. As predicted from a context reinstatement framework, results revealed higher identification accuracy after re-reading compared to the control group. This result was replicated using a second target and corroborated by several small meta-analyses of similar studies ($k = 4$). Moreover, there was a tendency for a verbal facilitation effect with the likelihood of a correct identification decision being almost three times higher when the perpetrator was described compared to the control condition.

In Experiment 2, the Brunswikian lens model (Brunswik, 1956, 1965) was applied to the evaluation of eyewitness identifications (for the first time to our knowledge). To explain observers' judgment accuracies when evaluating the accuracy of an identification decision it was examined (1) which cues observers use to evaluate an identification decision (”subjective utilities”), (2) how they interpret and weight these cues, and (3) if these cues as perceived by observers are indeed related to identification accuracy ("ecological validities"). Study 1 presented participant-observers with literal transcripts of 48 choosers' identification decisions, whereas Study 2 used the original videotapes. A “think-aloud” method was applied to make discriminating cues more salient to observers, which was compared to retrospective reasoning protocols. Both studies demonstrated that observers tended to overestimate cue discriminability and used these cues as indicators of identification accuracy independently of type of decision protocol. However, when videotaped think-aloud protocols were evaluated discriminating cues were visible to observers resulting in a high correspondence between subjective utilities and ecological validities. Advantages of think-aloud methods and videotapes to increase fact finders' judgment accuracy when evaluating identification decisions are discussed.

In Experiment 3, persuasive effects of more peripheral, indirect measures (i.e., perceived witness speech style and attributed witness traits) and ratings of different person and event description qualities on observers' judgments were investigated. Although none of these cues was a valid indicator of objective identification accuracy, observer judgments were related to several description qualities and perceived witness confidence. Persuasive effects of speech style characteristics depended on the presentation of additional descriptions, that is, only when identification statements were presented alone did ratings affect observer judgments.

To conclude, this dissertation suggests re-reading one's own person descriptions as a promising approach to increase identification accuracy. As it is common police practice to ask eyewitnesses for a description of the perpetrator, re-reading one's description is an easily applicable system variable that does not require any additional procedures, training, or resources. In contrast to former research (Alogna et al., 2014; Schooler & Engstler-Schooler, 1990), describing a perpetrator does not seem to impair identification performance under more realistic conditions. Thus, differences in description effects due to experimental procedures used are discussed.

Considering fact finders' judgmental processes when evaluating the accuracy of an identification decision, the present studies demonstrated that original videotaped or transcribed identification statements indeed contain some valid indicators of identification accuracy that are perceivable and usable by observers. Especially the application of videotaped think-aloud protocols seems to be fruitful to make valid cues more salient to observers. However, future researchers are encouraged to test and optimize these instructions for the evaluation of identification decisions. Moreover, the Brunswikian lens model framework offers an appropriate method to contrast relationships between empirically valid and intuitively used cues. Thus, to increase fact finders' judgment accuracy, the model further allows recommendations for an appropriate weighting of information contained in identification protocols.

TABLE OF CONTENTS

INTRODUCTION

> As the pace of DNA exonerations has grown across the country in recent years, wrongful convictions have revealed disturbing fissures and trends in our criminal justice system. Together, these cases show us how the criminal justice system is broken--and how urgently it needs to be fixed. (Innocence Project, retrieved from *http://www.innocenceproject.org/causes-wrongful-conviction*, February 13, 2016).

According to the American *Innocence Project* (*www.innocenceproject.org*), an organization aiming to exonerate wrongfully convicted individuals through DNA testing, false eyewitness identifications are one of the major causes that contributed to wrongful convictions in more than 70% of cases (beside other factors; see also Garrett, 2008, 2011, 2012).

A popular case describes the wrongful identification of Ronald Cotton, a 22-year old, black man, who was accused of having raped a female student, Jennifer Thompson, at night in her apartment in July 1984. Three days after the rape Jennifer Thompson identified Ronald Cotton in a photo lineup. It is reported that she carefully studied each lineup picture and it took her up to five minutes to make her choice. Some days later, a live lineup was conducted, and again, Jennifer Thompson identified Ronald Cotton. She reported of having been 100 percent sure that she had identified the right man. Moreover, the police officers reinforced her by saying that she identified the same person as in the photo lineup, which made her inappropriately confident of having identified her true rapist. Consequently, in

January 1985, she identified Ronald Cotton in court a third time. The jury believed her testimony and judged her identification decision to be accurate. Thus, Ronald Cotton was found guilty and sentenced for life and 50 years. However, after having served two years in prison, Ronald Cotton’s new inmate, Bobby Poole, admitted the crime Cotton was convicted for. Consequently, Cotton’s advocates requested a new trial, but even at this moment Jennifer Thompson did not recognize her true rapist. She was still convinced that Ronald Cotton raped her and thus Ronald Cotton was convicted a second time for life and another 54 years.

Seven years later, in 1995, Ronald Cotton was exonerated through DNA testing and Bobby Poole was convicted instead. He was cleared of all charges, however, he had served 10.5 years in prison while being innocent (for a detailed case description see, Thompson-Cannino, Cotton, & Torneo, 2009; *www.pickingcottonbook.com*).

A careful examination of this and similar other cases demonstrates the powerful consequences that may follow a wrongful identification. To date 337 individuals were exonerated successfully by the Innocence Project after having served an average number of 14 years in prison (*www.innocenceproject.org*). However, not only a misidentification per se, but also the incorrect evaluation of an identification decision in court may cause a wrongful conviction (cf. Garrett, 2011, 2012). When it comes to a trial, fact finders (i.e., police officers, jurors, lay judges or judges) have to evaluate the accuracy of an eyewitness’s identification decision to arrive at their verdicts. This is problematic especially when the identification decision is the only or one of the major sources of evidence, and thus has a high relevance for the following verdict.

In the present dissertation, factors are addressed that are supposed to increase the probability of witnesses making a correct identification decision (Experiment 1). Second, to contribute to the prevention of erroneous judgments in court, fact finders’ evaluation and decision processes when judging the accuracy of an identification decision are analyzed in detail (Experiments 2 and 3).

How To Obtain Correct Identifications?

When a crime is observed, it is common police practice to first ask the eyewitness for a detailed *description* of the incident, the crime scene and the perpetrator’s physical appearance. When a suspect is made out, eyewitnesses usually are asked to *identify* the person in a lineup. As the police commonly do not know the perpetrator who committed the crime, an identification procedure is conducted to examine if the suspect indeed is the perpetrator. The eyewitness is confronted either with a live lineup or a photo array, containing the suspect next to several innocent foils. Objectively, the suspect may (*target-presence: TP*) or may not be the perpetrator (*target-absence: TA*). In contrast to real life situations, in experimental investigations the perpetrator is known and thus target-presence can be controlled. To mitigate a witness’s expectation that the perpetrator indeed is in the lineup and only has to be selected (e.g., this is what Jennifer Thompson thought), the eyewitness should be made aware of the possibility that the perpetrator might not be present in the lineup.

To construct a fair lineup, that is, a lineup that does not increase the suspect’s probability of being misidentified, it is important that the foils generally match the suspect’s physical appearance (e.g., they should have the same height, stature, skin and eye color). Based on the "*match-to-description*" method described by Luus and Wells (1991), the suspect as well as the selected foils should also match the witness's description of the perpetrator. Thus, ideally all lineup members should show all of the described features to prevent witnesses of making a positive identification simply based on the presence of one outstanding feature. As an example, a witness might remember the perpetrator's dark brown eyes, but only two persons in the lineup match to this description. Consequently, the witness probably will choose between these two without considering the other lineup members. Thus, the lineup is biased.

Finally, eyewitnesses have the opportunity to make a positive identification (*choosers*), that is, to pick one lineup member that is identified as the perpetrator. On the other hand, they may reject the lineup, that is, they do not choose any of the lineup members (*nonchoosers*). The possible results of an identification procedure are displayed in Table 1 (cf. Sporer & Sauerland, 2008; Wells, 1993).

Table 1

Possible Outcomes of an Identification Procedure

Target-presence	Witness Response: No Identification "Nonchoosers"	Witness Response: Positive Identification "Choosers"	
Target-absent (TA)	Correct rejection	False identification (false alarm)	Foil identification [a]
Target-present (TP)	False rejection	Correct identification (hit)	Foil identification [a]

Note. [a] As there is only one suspect per lineup and all foils are known to be innocent a foil identification always reflects an incorrect identification decision. In many studies foil identifications are also treated as false identifications (cf. Sauerland & Sporer, 2008).

Since many years researchers have investigated factors that affect identification accuracy (for a review see National Academy of Sciences, 2014; Wells & Olson, 2003) and concluded that eyewitness testimonies and especially identifications are highly error-prone. Wells (1978) proposed a distinction between two areas of applied eyewitness research focusing on *system* and *estimator* variables. System variables are modifiable in the criminal justice system and thus can be controlled and used to increase identification accuracy (e.g., interrogation techniques, foil selection and lineup presentation mode). In contrast, estimator variables cannot be controlled by the criminal justice system and have to be estimated retrospectively. These estimations refer to the *situational witnessing conditions* (e.g., lighting, distance, duration of the crime, weapon presence) as well as to *characteristics of the witness* (e.g., age, emotional state, attention focus and expertise) and the *target* (e.g., distinctiveness, ethnic group).

According to Sporer (2007a, 2008) system- and estimator variables affect eyewitness testimony in different phases of information processing reaching from perception and retention to a final recall or recognition phase (see also Sporer & Sauerland, 2008). First, eyewitnesses perceive external crime relevant information in a specific situation (e.g., after a party on the street at night) and in a specific internal context (e.g., intoxicated and highly emotional). Then, they have to keep the information in mind for an indeterminate time, in which they may be influenced by post-event information (e.g., suggestions, media influence) until they need to retrieve it in another specific internal state and situational context (e.g., at the police station).

In the individual case, the knowledge about system and estimator variables serves to assess the probability of a correct identification retrospectively. Moreover, certain assessment variables can be used to postdict the accuracy of an identification decision (Sporer, 1993). However, to directly increase the number of correct identifications, and more importantly, to reduce the number of false identifications, modifications of system variables should be considered in criminal proceedings in line with significant research findings (e.g., National Academy of Sciences, 2014;

Wells & Olson, 2003; Wells, Small, Penrod, Malpass, Fulero & Brimacombe, 1998).

Experiment 1: Re-reading person descriptions as system variable to increase identification performance?

The first experiment examines the potentially facilitating role of *re-reading one's own person description* prior to the identification task as a system variable to increase identification accuracy (cf. Cutler, Penrod, O'Rourke, & Martens, 1986; Sporer, 2007b). It is assumed that re-reading one's own person description might serve as a self-generated retrieval cue (cf. Mäntylä, 1986), which mentally reinstates the perceptual context, in which the original stimulus face was encoded. Based on an associative network perspective (Anderson, 1983) a stimulus is never stored in memory alone. If one information unit in memory is activated, activation is assumed to spread along the associative network, and thus related information units are activated as well. Consequently, the presentation of contextual cues is supposed to facilitate the recall of the target face.

These assumptions contrast former work on *the verbal overshadowing effect* (VOE: Alogna et al., 2014; Schooler & Engstler-Schooler, 1990) demonstrating an overall negative effect of giving person descriptions on identification accuracy. However, the VOE seems to be restricted to certain experimental conditions and thus has not been unconditionally replicable (e.g., Meissner, Brigham, & Kelley, 2001; Sauerland, Holub, & Sporer, 2008). With more realistic experimental conditions the effect disappeared or even tended to be reversed, for example, when a longer delay (> 20 minutes) between the description and the identification task was embedded (Finger & Pezdek, 1999; Meissner & Brigham, 2001). Based on Paivio's (1971) dual process theory and Craik and Lockhart's (1972) levels-of-processing account, in the present experiment facilitating description effects on identification accuracy were expected using more realistic retention intervals.

How To Judge Identification Accuracy?

Sporer (2007a, 2008) proposed an integrative model of eyewitness testimony (Figure 1) to explain how fact finders evaluate an identification decision. In this model three levels of eyewitness research are distinguished. At the bottom there is the *information processing level* addressing factors that affect eyewitness testimony during perception, retention and retrieval (i.e., estimator and system variables described above).

At the *meta-memory level* witnesses' evaluations of their own memory processes are focused. For example, these evaluations include the witness's perceived ease of retrieval resulting in a subjective confidence rating that the recalled information is correct. Meta-memory judgments may be expressed either verbally (e.g., "I am absolutely sure! He is the perpetrator!") and/or through nonverbal behaviors (e.g., in terms of signs of nervousness or gestures) and paraverbal cues (e.g., in terms of response latencies or certain speech characteristics like hedges and hesitations). Moreover, witnesses may be asked to evaluate the decision strategies (cf. Dunning & Stern, 1994; Sauerland & Sporer, 2007; Wells, 1984) they have applied to make an identification decision.

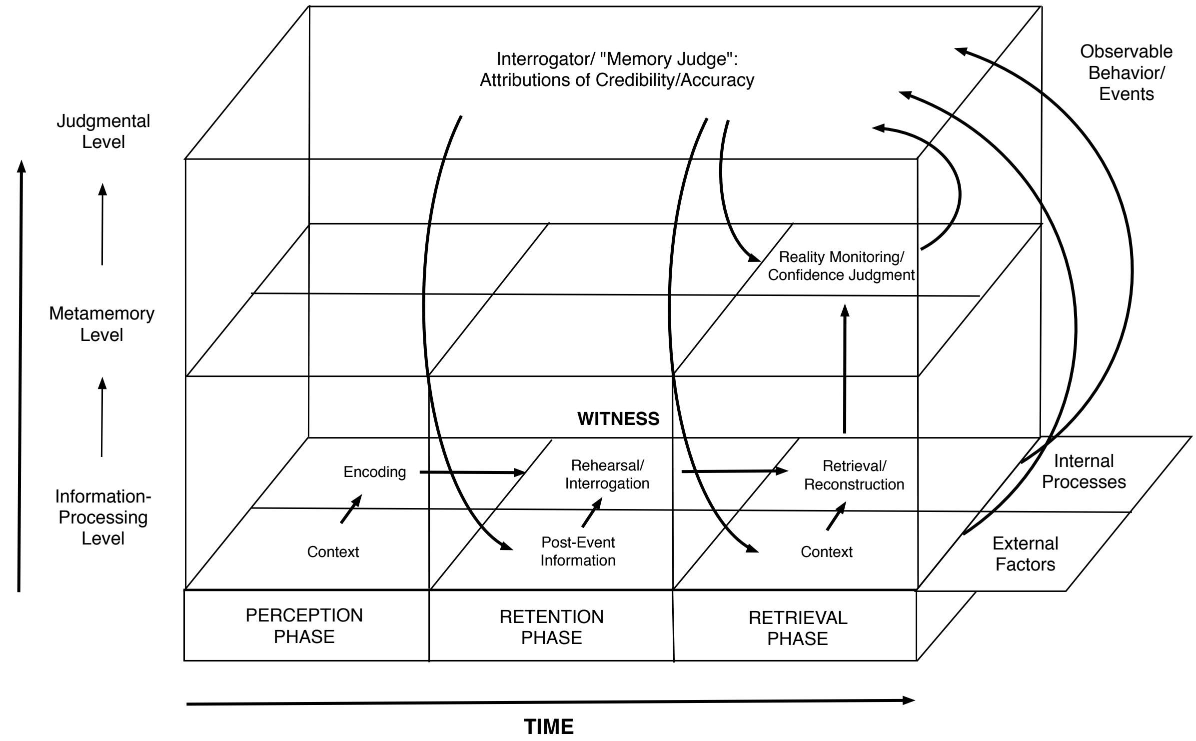

Figure 1. Integrative model of eyewitness testimony and its evaluation (Sporer, 2008).

Many researchers have extensively investigated the “postdictive” value of different meta-memory judgments so far, demonstrating reliable relationships between identification accuracy and the witness’s confidence, response latency at the identification and self-reported decision processes. At least for choosers, reliable associations were observed (e.g., *decision processes*: Dunning & Stern, 1994; Ross, Benton, McDonnell, Metzger, & Silver, 2007; Smith, Lindsay, & Pryke, 2000; Wells, 1984; *response latency*: Brewer, Caon, Todd, & Weber, 2006; Dunning & Perretta, 2002; Sporer, 1992, 1993, 1994; Weber, Brewer, Wells, Semmler, & Keast, 2004; *confidence*: Brewer & Wells, 2006; Sauerland & Sporer, 2007, 2009; Sporer, Penrod, Read, & Cutler, 1995; Wixted, Mickes, Clark, Gronlund, & Roediger, 2015).

At the *judgmental level*, interpersonal memory judgments are made, and thus, fact finders' judgmental processes when evaluating the accuracy of an eyewitness's memory are investigated. To make these judgments, fact finders rely on observable witness behaviors and the witness's report of the event itself, but also take witnessing conditions at the perception, retention and retrieval stage into account. Fact finders also use witnesses' meta-memory statements as assessment variables (cf. Sporer, 1993; Sporer & Sauerland, 2008) to evaluate the witness's memory (Sporer, 2007a). For example, a notable body of research has demonstrated that observers heavily rely on the witness's confidence when they were asked to judge the accuracy of an identification decision (e.g., Cutler, Penrod, & Stuve, 1988; Lindsay, Wells, & Rumpel, 1981; Wells, Lindsay, & Ferguson, 1979).

However, judgments are also influenced by fact finders' common sense beliefs about eyewitness testimony and their knowledge about estimator and system variables that were supposed to be influential in a specific case (for a review see Boyce, Beaudry, & Lindsay, 2007; Leippe, 1994). This is problematic to the extent that several studies have shown that fact finders only have a poor knowledge about factors affecting identification accuracy (e.g., Benton, Ross, Bradshaw, Thomas, & Bradshaw, 2006; Desmarais & Read, 2011; Wise & Safer, 2004). For example, it seems to be common sense to intuitively assume that police officers are "better eyewitnesses" than laypersons (Yarmey, 1986). Consequently, police officers receive higher credibility ratings in court and their testimonies are rated more favorably (Linz et al., 1982, as cited in Penrod, 1983; Yarmey, 1986; Yarmey & Jones, 1983). However, and this is important, research results are largely heterogeneous and do not support such a clear police officers' memory advantage (for an own experimental study on this topic see Kaminski & Sporer, 2016; for meta-analytic results see Sporer, Zimmerman, & Kaminski, in preparation).

With regard to these inappropriate beliefs it is not surprising that observers often showed only poor judgment accuracies, which were comparable to a chance level of 50%, when they were asked to evaluate the accuracy of an identification decision (e.g., Beaudry, Lindsay, Leach, Mansour, Bertrand, & Kalmet, 2015; Brigham & Bothwell, 1983; Lindsay et al., 1981; Wells et al., 1979).

As an explanation for these poor judgment accuracies, it is assumed that observers (1) might not be sensitive to cues that indeed are associated with identification accuracy (i.e., they did not use them to make their judgments at all), (2) give these cues an inappropriate weight (i.e., they under- or overestimate cue discriminability), and/or (3) they use wrong cues to make their judgments (i.e., cues that are not valid indicators of identification accuracy; cf. *validity-intuition model*: Leippe, 1994; see also Semmler, Brewer, Bradfield Douglass, 2012).

Thus, to increase fact finders' judgment accuracies, it is necessary to investigate observers' judgmental processes in more detail. Two experimental studies were conducted to investigate (1) which cues observers use to evaluate identification accuracy, (2) how they weight them and (3) if these cues indeed are valid indicators of identification accuracy. Therefore, the Brunswikian lens model (Brunswik, 1956, 1965) was adapted to judgments of identification decisions for the first time. In this model, *ecological validities*, that is, relationships between certain eyewitness characteristics and objective identification accuracy, and *subjective utilities*, that is, relationships between these characteristics and observer judgments, are contrasted. High correspondences between ecological validities and subjective utilities reflect an appropriate use of the investigated characteristics, which should result in increased judgment accuracy.

Experiment 2: Discriminating between correct and incorrect eyewitness identifications: The use of appropriate cues.

Many studies demonstrated reliable relationships between identification accuracy and different meta-memory variables (e.g., Dunning & Stern, 1994; Sporer et al., 1995; Sporer, 1992, 1993, 1994). As correct and incorrect identifications can be discriminated by taking the witness's confidence, response latencies and self-reported decision processes into account, in the present experiment it was investigated if observers were sensitive to these relationships, too. Observers were presented with a positive eyewitness identification statement and had to evaluate its accuracy. Instead of presenting observers with witnesses' explicit self-reports, observers had to rate the witnesses' perceptual basis, confidence, response latency, and decision processes based on their subjective impression of the witness and his/her testimony. To make witnesses' meta-memory thoughts and decision processes more apparent for observers, think-aloud protocols (cf. Ericsson & Simon, 1993) were used and compared to the use of retrospective reasoning statements.

Two studies were conducted presenting the identification decision as literal transcripts (Study 1) or as videotapes (Study 2). Videotapes of the original identification decision were expected to facilitate observers' evaluations by presenting more peripheral witness characteristics that are not observable or less salient in written transcripts (e.g., nonverbal and paraverbal behaviors, response latencies; cf. *Elaboration Likelihood Model*: Petty & Cacioppo, 1986; Leippe, 1994). It was hypothesized that think-aloud protocols as well as the use of videotapes would increase cue discriminability as well as observers' ability to discriminate between correct and incorrect identifications. As a result higher correspondences between ecological validities and subjective utilities were expected under these conditions.

Experiment 3: Evaluating eyewitness identification decisions by indirect measures.

In a third experiment, persuasive effects of different indirect measures on observer judgments as well as their objective relationships with identification accuracy were investigated. Based on dual-process models of persuasion (e.g., *Elaboration Likelihood Model*: Petty & Cacioppo, 1986) it was assumed that not only content-related aspects of a memory message influence fact finders' judgments, but also more peripheral witness characteristics. Thus, this study focused on the witness's speech style (i.e., *powerless vs. powerful speech*: Erickson, Lind, Johnson, & O'Barr, 1978; O'Barr, 1982) and several attributed witness traits (cf. *Witness Credibility Scale:* Brodsky, Griffin, & Cramer, 2010), which are assumed to convey an impression of the overall witness's confidence and memory accuracy, and thus are assumed to affect observers' judgments.

Based on the US Supreme Courts recommendation (Neil vs. Biggers, 1972) to use person description quality to assess the accuracy of an identification decision (see also the criteria used by the German Supreme Court: Sporer & Cutler, 2003) and based on heuristic decision rules like "the more arguments the better" (cf. *Elaboration Likelihood Model*: Petty & Cacioppo, 1986) persuasive effects of different, easily accessible person and event description qualities on observer judgments were investigated in this study as well.

References

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

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior, 22,* 261-295. doi: 10.1016/S0022537183902013

Beaudry, J. L., Lindsay, R. C. L., Leach, A.-M., Mansour, J. K., Bertrand, M. I., & Kalmet, N. (2015). The effect of evidence type, identification accuracy, line-up presentation, and line-up administration on observers' perceptions of eyewitnesses. *Legal and Criminological Psychology, 20,* 343-364. doi: 10.1111/lcrp.12030

Benton, T. R., Ross, D. F., Bradshaw, E., Thomas, W. N., & Bradshaw, G. S. (2006). Eyewitness memory is still not common sense: Comparing jurors, judges and law enforcement to eyewitness experts. *Applied Cognitive Psychology, 20,* 115-129. doi: 10.1002/acp.1172

Boyce, M. A., Beaudry, J. L., & Lindsay, R. C. L. (2007). Belief of eyewitness identification evidence. In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Handbook of eyewitness psychology: Memory for people (pp. 25-45)*. Mahwah, NJ: Lawrence Erlbaum and Associates.

Brewer, N., Caon, A., Todd, C., & Weber, N. (2006). Eyewitness identification accuracy and response latency. *Law and Human Behavior, 30,* 31-50. doi: 10.1007/s10979-006-9002-7

Brewer, N., & Wells, G. L. (2006). The confidence-accuracy relationship in eyewitness identification: Effects of lineup instructions, foil similarity,

and target-absent base rates. *Journal of Experimental Psychology: Applied, 12*, 11-30. doi: 10.1037/1076-898x.12.1.11

Brigham, J. C., & Bothwell, R. K. (1983). The ability of prospective jurors to estimate the accuracy of eyewitness identifications. *Law and Human Behavior, 7,* 19-29. doi: 10.1007/BF01045284

Brodsky, S. L., Griffin, M. P., & Cramer, R. J. (2010). The witness credibility scale: An outcome measure for expert witness research. *Behavioral Sciences & the Law, 28,* 892-907. doi: 10.1002/bsl.917

Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. Berkeley: University of California Press.

Brunswik, E. (1965). *The conceptual framework of psychology*. Chicago, London: The University of Chicago Press.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11,* 671-684. doi:10.1016/S0022-5371(72)80001-X

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*(2), 113-134.

Cutler, B. L., Penrod, S. D., & Stuve, T. E. (1988). Juror decision-making in eyewitness identification cases. *Law and Human Behavior, 12,* 41-55. doi: 10.1007/BF01064273

Desmarais, S. L., & Read, J. D. (2011). After 30 years, what do we know about what jurors know? A meta-analytic review of lay knowledge regarding eyewitness factors. *Law and Human Behavior, 35,* 200-210. doi: 10.1007/s10979-010-9232-6

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

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

Erickson, B., Lind, E. A., Johnson, B. C., & O'Barr, W. M. (1978). Speech style and impression-formation in a court setting: Effects of powerful and powerless speech. *Journal of Experimental Social Psychology, 14,* 266-279. doi: 10.1016/0022-1031(78)90015-X

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. Massachusetts: A Bradford Book.

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

Garrett, B. L. (2008). Judging innocence. *Columbia Law Review, 108,* 55-142.

Garrett, B. L. (2011). *Convicting the innocent: Where criminal prosecutions go wrong*. Cambridge, Massachusetts: Harvard University Press.

Garrett, B. L. (2012). Eyewitnesses and exclusion. *Vanderbilt Law Review, 65,* 451-506.

Kaminski, K. S. & Sporer, S. L. (2016). Sind Polizisten/-innen bessere Augenzeugen/-innen? Ein Vergleich zwischen Polizisten/-innen und Zivilpersonen hinsichtlich Beschreibungen und

Identifizierungsaussagen [Are police officers the better eyewitnesses? A comparison of police officers and laypersons regarding the accuracy of descriptions and identifications]. *Recht & Psychiatrie, 34*(1), 18-26.

Leippe, M. R. (1994). The appraisal of eyewitness testimony. In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 385-418). Cambridge, England: Cambridge University Press.

Lindsay, R. C. L., Wells, G. L., & Rumpel, C. M. (1981). Can people detect eyewitness-identification accuracy within and across situations. *Journal of Applied Psychology, 66,* 79-89. doi: 10.1037/0021-9010.66.1.79

Luus, C. A. E., & Wells, G. L. (1991). Eyewitness identification and the selection of distracters for lineups. *Law and Human Behavior, 15,* 43-57. doi: 10.1007/BF01044829

Mäntylä, T. (1986). Optimizing cue effectiveness: Recall of 500 and 600 incidentally learned words. *Journal of Experimental Psychology: Learning Memory and Cognition, 12,* 66-71. doi: 10.1037/0278-7393.12.1.66

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

Meissner, C. A., Brigham, J. C., & Kelley, C. M. (2001). The influence of retrieval processes in verbal overshadowing. *Memory & Cognition, 29,* 176-186. doi: 10.3758/bf03195751

National Academy of Sciences (2014). *Identifying the culprit: Assessing eyewitness identification*. Washington, D. C.: The National Academies Press.

Neil v. Biggers, 409 U.S. 188 (1972).

O'Barr, W. M. (1982). *Linguistic evidence: Language, power, and strategy in the courtroom*. San Diego: Academic Press.

Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt.

Penrod, S. (1983). *Social psychology.* Englewood Cliffs, N. J.: Prentice Hall.

Petty, R. E., & Cacioppo, J. T. (1986). *Communication and persuasion: Central and peripheral routes to attitude change*. New York: Springer.

Ross, D. F., Benton, T. R., McDonnell, S., Metzger, R., & Silver, C. (2007). When accurate and inaccurate eyewitnesses look the same: A limitation of the 'pop-out' effect and the 10- to 12-second rule. *Applied Cognitive Psychology, 21,* 677-690. doi: 10.1002/acp.1308

Sauerland, M., Holub, F. E., & Sporer, S. L. (2008). Person descriptions and person identifications: Verbal overshadowing or recognition criterion shift? *European Journal of Cognitive Psychology, 20,* 497-560. doi: 10.1080/09541440701728417

Sauerland, M., & Sporer, S. L. (2007). Post-decision confidence, decision time, and self-reported decision processes as postdictors of identification accuracy. *Psychology, Crime & Law, 13,* 611-625. doi: 10.1080/10683160701264561

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

Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology, 22,* 36-71. doi: 10.1016/0010-0285(90)90003-m

Semmler, C., Brewer, N., & Bradfield Douglass, A. (2012). Jurors believe eyewitnesses. In B. L. Cutler (Ed.), *Conviction of the innocent: Lessons from psychological research* (pp. 185-209). Washington, DC: American psychological Association.

Smith, S. M., Lindsay, R. C. L., & Pryke, S. (2000). Postdictors of eyewitness errors: Can false identifications be diagnosed? *Journal of Applied Psychology, 85,* 542-550. doi: 10.1037//0021-9010.85.4.542

Sporer, S. L. (1992). Post-dicting eyewitness accuracy: Confidence, decision-times and person descriptions of choosers and non-choosers. *European Journal of Social Psychology, 22,* 157-180. doi: 10.1002/ejsp.2420220205

Sporer, S. L. (1993). Eyewitness identification accuracy, confidence, and decision times in simultaneous and sequential lineups. *Journal of Applied Psychology, 78,* 22-33. doi: 10.1037/0021-9010.78.1.22

Sporer, S. L. (1994). Decision times and eyewitness identification accuracy in simultaneous and sequential lineups. In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 300-327). Cambridge: Cambridge University Press.

Sporer, S. L. (2007a). Evaluating witness evidence: The fallacies of intuition. In C. Engel & F. Strack (Eds.), *The impact of court procedure on the psychology of judicial decision making* (pp. 111-150). Baden-Baden, Germany: Nomos Verlag.

Sporer, S. L. (2007b). Person descriptions as retrieval cues: Do they really help? *Psychology Crime & Law, 13,* 591-609. doi: 10.1080/10683160701253986

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., & Cutler, B. L. (2003). Identification evidence in Germany and the United States. In P. J. vanKoppen & S. D. Penrod (Eds.), *Adversarial versus inquisitorial justice: Psychological perspectives on criminal justice systems* (pp. 191-208). New York: Springer.

Sporer, S. L., Penrod, S., Read, D., & Cutler, B. (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

Sporer, S. L., & Sauerland, M. (2008). Personenidentifizierung [Person identification]. *Forensische Psychiatrie, Psychologie und Kriminologie, 1,* 1-9. doi: 10.1007/s11757-008-0057-9

Sporer, S. L., Zimmerman, L., & Kaminski, K. S. (in preparation). Police vs. laypersons as eyewitnesses: A systematic review and meta-analysis. Manuscript to be submitted for publication.

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

Weber, N., Brewer, N., Wells, G. L., Semmler, C., & Keast, A. (2004). Eyewitness identification accuracy and response latency: The unruly 10-12-second rule. *Journal of Experimental Psychology: Applied, 10,* 139-147. doi: 10.1037/1076-898x.10.3.139

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

Wells, G. L. (1984). The psychology of lineup identifications. *Journal of Applied Social Psychology, 14,* 89-103. doi: 10.1111/j.1559-1816.1984.tb02223.x

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

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

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

Wells, G. L., Small, M., Penrod, S., Malpass, R. S., Fulero, S. M., & Brimacombe, C. A. E. (1998). Eyewitness identification procedures: Recommendations for lineups and photospreads. *Law and Human Behavior, 22,* 603-647. doi: 10.1023/a:1025750605807

Wise, R. A., & Safer, M. A. (2004). What US judges know and believe about eyewitness testimony. *Applied Cognitive Psychology, 18,* 427-443. doi: 10.1002/acp.993

Wixted, J. T., Mickes, L., Clark, S. E., Gronlund, S. D., & Roediger, H. L. (2015). Initial eyewitness confidence reliably predicts eyewitness identification accuracy. *American Psychologist, 70,* 515-526. doi: 10.1037/a0039510

Yarmey, A. D. (1986). Perceived expertness and credibility of police officers as eyewitnesses. *Canadian Police College Journal, 10*(1), 31-52.

Yarmey, A. D., & Jones, H. P. T. (1983). Is the psychology of eyewitness identification a matter of common sense? In S. M. A. Lloyd-Bostock &

B. R. Clifford (Eds.), *Evaluating witness evidence: Recent psychological research and new perspectives* (pp. 13-40). Chichester: John Wiley & Sons.

EXPERIMENT 1

The Verbal Facilitation Effect: Re-Reading Person Descriptions as a System Variable to Improve Identification Performance

> Appropriate arrangements must be made to make sure, before witnesses attend the identification parade, they are not able to (...) see, or be reminded of, any photograph or description of the suspect or be given any other indication as to the suspect's identity. (Police and Criminal Evidence Act, 1984, Code D, p. 53)

When a crime is reported to the police, a common practice is to ask the victim or witnesses for a description of the perpetrator. Most descriptions contain references to some aspect of the physical appearance of the perpetrator, that usually are not very distinctive, as well as vague estimates of age, height and weight. Among these attributes, the face is the most useful part of the body when trying to identify a person. However, our vocabulary for describing physical attributes of a face is rather limited (cf. Sporer, 1989). Moreover, research has shown that faces are best encoded configurally or holistically (Ellis, 1984; Tanaka & Farah, 2003), whereas giving a description requires accessing the memory of the face by its individual features (e.g., Schooler & Engstler-Schooler, 1990; see Meissner, Sporer, & Schooler, 2007). Thus, on the one hand, obtaining a useful description from a witness can be quite difficult (Fahsing, Ask, & Granhag, 2004; Lindsay, Martin, & Webber, 1994; Shepherd, Davies, & Ellis, 1981), while on the other hand being indispensable for furthering the criminal investigation (e.g., Loftus, 1979; Wells & Olson, 2003). In most cases person descriptions precede a later person identification task. Moreover, the US Supreme Court recommended considering the accuracy of person descriptions when evaluating the reliability of an identification decision (*Neil vs. Biggers*, 1972).

Due to the practical importance of person identifications, research has focused on the relationships between quantitative and qualitative description measures and identification performance by investigating the postdictive value of person descriptions to discriminate between correct and incorrect identification decisions (see the meta-analysis by Meissner, Sporer, & Susa, 2008). Other researchers have focused on the impairing effects that giving a description can have on a future identification decision (Schooler & Engstler-Schooler, 1990; see also Meissner & Brigham, 2001).

In contrast, in the present study special emphasis was placed on the potentially beneficial effect of giving a person description on lineup identification accuracy (*verbal facilitation*), while ensuring the study's ecological validity. Furthermore, we employ person descriptions as a system variable to improve identification performance (Wells, 1978). Specifically, descriptions were treated and used as self-generated retrieval cues by allowing witnesses to re-read their own descriptions prior to the identification task, thus reinstating the previous retrieval context.

Description Effects: Verbal Overshadowing and Verbal Facilitation

A description given by a witness can be helpful in finding a suspect. But what if the very act of describing the perpetrator impaired a witness's ability to later identify him or her? Schooler and Engstler-Schooler (1990) found that exactly this might be the case. In their Experiment 1, participants who had been encouraged to give a detailed description of the perpetrator's face for five minutes performed significantly worse (decrease of 25%) in an immediately following target-present identification task than participants in the control condition who did not give a description, an effect the authors termed *verbal overshadowing* (VOE). Recently, Alogna et al. (2014) conducted a registered replication report of Schooler and Engstler-Schooler's (1990) original experiment , including 22 studies.[1] The results supported a robust and consistent VOE with an average difference of 16% between a verbal description condition and a control condition.

However, note that encouraging participants to describe a face for *five minutes* may place undue emphasis on verbalization, thus provoking potentially misleading descriptors. In a recent study from our laboratory (*N* = 197) the description of *faces, body and clothing* lasted on average only *Mdn* = 34 seconds (inter-quartile range = 26 seconds). In an archival analysis of person descriptions in criminal cases, only 2.88 face descriptors were mentioned (Sporer, 1996). Hence, we wonder why participant-witnesses are encouraged to describe *a face* for *five minutes*?

Since Schooler and Engstler-Schooler's (1990) original work, quite a few studies have been conducted on the verbal overshadowing effect, some replicating (e.g., Fallshore & Schooler, 1995; MacLin, 2002; Smith & Flowe, 2014; Sporer, 1989), others failing to replicate or even demonstrating a reversal of the effect (e.g., Brown & Lloyd-Jones, 2005, 2006; Chance &

[1] Sample sizes of Alogna et al.'s (2014) two replication studies varied due to an error in the initial experimental protocol. The replication of the traditional verbal overshadowing condition (cf. Experiment 1, Schooler & Engstler-Schooler, 1990) was conducted as a follow-up experiment and consequently only 22 of the initially participating 31 laboratories completed this experiment.

Goldstein, 1976; Itoh, 2005; Kitagami, Sato, & Yoshikawa, 2002; McKelvie, 1976; Meissner, Brigham, & Kelley, 2001; Read, 1979; Sauerland, Holub, & Sporer, 2008; Yu & Geiselman, 1993). A major restriction of past studies showing a VOE is that in most of the experiments no or only very short delays of a few minutes were inserted between the description and the identification task. In contrast, in real world cases, a description-identification interval of several days, or even weeks or months is more likely to occur (cf. the median delay of 13 to 14 days in a recent field study by Wells, Steblay, and Dysart, 2015). If, however, a delay was inserted between the description and the recognition task, the negative effects of verbalization disappeared in most studies (e.g., Yu & Geiselman, 1993; see the meta-analysis by Meissner & Brigham, 2001; an exception is Schooler & Engstler-Schooler, 1990, Experiment 5). In a direct test of this argument by Finger and Pezdek (1999), the VOE disappeared when an interval of 24 minutes or one hour, respectively, was inserted between description and recognition task. Similarly, in Alogna et al.'s (2014) replication report, the VOE was much smaller (4%) when an interval of 20 minutes was used.

These findings were supported in the meta-analysis by Meissner and Brigham (2001). The VOE occurred in studies in which the identification task immediately (Z_r = -0.16), or with a short delay (Z_r = -0.13) followed the verbalization task. Differences were marginally significant in the opposite direction in studies employing a delay of more than 30 minutes (Z_r = 0.07), suggesting a *verbal facilitation effect* (VFE).

Facilitative effects of verbalization on recognition performance can be explained by Paivio's (1971) *dual process* theory of encoding and Craik and Lockhart's (1972) *levels-of-processing* account. Both accounts share the assumption that multimodal (verbal and visual) and thus deeper encoding (e.g., by adding self-generated semantic associations while describing the face) should result in a retrieval benefit for the encoded stimulus. As memory strength decreases with time (Ebbinghaus, 1913; Rubin & Wenzel, 1996), we assume that an early first recall, that is, describing the face soon after the witnessed event, leads to a deeper level of processing and elaboration, and thus, consolidates the recalled information into memory and reduces the amount of forgetting (e.g., Ebbesen & Rienick, 1998; see also Hope, Gabbert, & Fisher, 2011). Moreover, in terms of an associative network perspective, an early recall increases the activation level of the recalled items and strengthens the associations between them (Anderson, 1983). Thus, new retrieval routes are produced and related concepts are activated that later can serve as additional retrieval cues. Consequently, later recall attempts are facilitated.

These assumptions are in line with learning studies investigating the beneficial effect of an early memory test prior to the final memory task, called the *testing effect*.[2] A retrieval-induced facilitation of material related to the tested material is observed when early testing activities are included compared to no-testing control groups (for a review see McDermott, Arnold, & Nelson, 2014; Karpicke & Grimaldi, 2012; Roediger & Karpicke, 2006), especially with longer delays (e.g., Butler & Roediger, 2007; Chan, 2009). Extrapolating from this literature, one would also expect a (stronger) VFE after longer description-identification delays.

---

[2] We are indebted to an anonymous reviewer to direct us to pointing out the testing effect as a possible explanation for our results.

Indeed, several studies have shown a beneficial effect of verbalization on later face recognition. For example, Sporer (1988) exposed participants to pairs of faces for 10 s and told them afterwards, which of the two they were to commit to memory. In one group, participants were telephoned and read their own descriptions to visually rehearse these faces. At a final recognition test, the verbally described and rehearsed faces showed a 12.5% advantage compared to the nonrehearsed faces. It appears that the description fostered the original memory trace. However, most of these studies used old-new recognition paradigms with multiple faces to be described and remembered (e.g., Brown & Lloyd-Jones, 2005, 2006; Brown, Gehrke, & Lloyd-Jones, 2010; Wickham & Lander, 2008). But most of them did not include realistic post-description delays or used target-present lineups only (e.g., LaPaglia & Chan, 2012). Thus, the main goals of the present study were to investigate description effects using an eyewitness identification paradigm using (1) realistic description instructions (2) longer description-identification delays and (3) both target-present and target-absent lineups.

Context Reinstatement

From an associative network perspective (Anderson, 1983), a to-be-remembered stimulus is never encoded into memory alone. Accordingly, a variety of environmental, emotional, and other contextual information of the episode in which the stimulus was encountered is encoded in an associative network into which the to-be-remembered stimulus is embedded (Bower, 1981). If a later memory search fails to activate the direct path to the stimulus node, alternative routes can be primed by using contextual cues, making it more likely for the stimulus node to be activated and the required information to be recalled. This effect was termed *context reinstatement effect*.

In an eyewitness identification study by Cutler, Penrod, and Martens (1987) a context reinstatement interview was used, consisting of Geiselman, Fisher, MacKinnon, and Holland's (1985) "*mnemonic instructions*", pictures of the location and the victim of the incident, and a review of the original description. They found that when the perpetrator was disguised, the context reinstatement interview significantly improved identification performance (51% vs. 29%, d = 0.49, *OR* = 2.43)[3], whereas it had no significant effect if the perpetrator was non-disguised (47% vs. 57%, d = -0.22, *OR* = 0.67). Beneficial effects of context reinstatement procedures on identification accuracy were observed in a field experiment by Krafka and Penrod (1985) and a staged event study by Malpass and Devine (1981) when target-present lineups were used (both with *OR*s greater than 2.00).

In an early meta-analysis of both facial recognition and lineup identification studies, Shapiro and Penrod (1986) reported a large beneficial effect of context reinstatement on hits (d = 1.91, k = 23), but also a smaller increase in false alarms (d = -0.44, k = 18). Recently, Wong and Read (2011) similarly reported a significant positive effect of context reinstatement on the hit rate in target-present lineups (*OR* = 3.12), but a nonsignificant effect on the false alarm rate when the target was absent (*OR* = 1.84).

---

[3] Odds ratios (*OR*) > 1 illustrate higher observed frequencies for the context reinstatement condition compared to a particular control group. Odds Ratios for the results of Cutler et al.'s (1987) and Cutler, Penrod, O'Rourke, and Martens' (1986) studies were converted from the reported d values (for the exact formulae, see Borenstein, 2009).

Consequently, there is a need to develop a method of context reinstatement that will increase hit rates without increasing false identifications.

Context Reinstatement by Re-Reading one's own Descriptions

Based on the expected facilitating effect of person descriptions on identification accuracy, the question arises whether one could even further increase this positive effect. Hence, we propose that person descriptions be used as a form of context reinstatement as a simple system variable to further improve identification accuracy.

Cutler, Penrod, O'Rourke, and Martens (1986) attempted to unconfound different context reinstatement procedures and observed that re-reading one's own description of the target and the event was the only context variable yielding significant effects on identification accuracy. However, beneficial effects of re-reading were present only under certain circumstances, viz., under less optimal retrieval conditions. When the target was absent in the lineup, re-reading had a positive effect (control: 60% vs. re-reading: 74% correct rejections, d = 0.30, OR = 1.72), whereas with target-present lineups it had a negative effect (control: 68% vs. re-reading: 50% hits, d = -0.39, OR = 0.49). Moreover, re-reading had positive effects when the perpetrator was disguised *and* absent from the lineup (control: 50% vs. re-reading: 70% correct rejections, d = 0.45, OR = 2.26).

Sporer (2007) also explored possible effects of re-reading descriptions using a relatively shorter exposure time of the target and a retention interval of one week. There was a tendency for participants who re-read their descriptions (52% correct identification decisions) to perform better at the identification task than participants in the no-reread condition

(36%), *OR* = 1.90. However, this effect did not reach significance, due to the small sample size (*N* = 54).

In sum, re-reading one's own description does not only not impair identification performance, but may actually activate an associative memory network for the target face, resulting in an increase in identification accuracy, especially after long delays and with target-absent lineups. Re-reading descriptions may function as self-generated retrieval cues, which have been shown to induce even higher memory performance compared to other-generated cues (analogous to the studies with word lists: e.g., Mäntylä, 1986). Thus, the present study aimed for a replication of the re-reading effect with new stimulus material to further test its effectiveness, while taking extensive care to ensure ecological validity.

Do Quantity and Quality of Person Descriptions Matter?

The benefit of re-reading is likely to depend on the quantity and quality of a witness's description. Re-reading should be helpful to the extent that the description includes many details, which act as retrieval cues to activate the original memory for the target face, thus enabling better identification. Hence, the question is: Does context reinstatement by means of re-reading depend on a "good" person description containing many correct details?

Sporer (1996) identified five aspects that can be related to identification accuracy: the length of the description (i.e., the number of words), the number of details reported, the accuracy, the internal consistency between different descriptions by the same witness, and the general quality of the statement. To judge the "goodness" of a description, both the total number of details reported and the proportion of accurate and inaccurate details have to be considered. Relationships between different aspects of a description and identification are generally weak but stronger if person descriptions are measured with methodological rigor (cf. the meta-analysis by Meissner et al., 2008). Sometimes description properties are not related to identification accuracy but to choosing rates with participants who were allowed to re-read their descriptions prior to the identification task (Sauerland et al., 2008).

This supports the idea that, within re-readers, increased lineup rejections might be due to the perceived inaccuracy of their descriptions making participants more skeptical of their own memory and thus, more reluctant to chose someone from the lineup (cf. the *criterion shift* account of the VOE, Clare & Lewandowsky, 2004; Sauerland et al., 2008).

In the present study, relationships between identification accuracy and description accuracy as well as the number of details were examined. Because in actual criminal cases, there is no way of assessing the actual accuracy of a description--the true identity of the perpetrator is unknown--different aspects of the perceived description quality (cf. Valentine, Pickering, & Darling, 2003) were additionally measured, including ratings of a description's precision, specificity and informativeness. Especially for re-readers we expected positive relationships between these ratings and identification accuracy.

The Present Study

The main goal of the present study was to investigate the effects of verbalization and re-reading one’s own description on subsequent identification accuracy. We were also interested in examining possible associations with description quality.

Using three groups, a no description control group, a description only group, and a description plus re-reading group, orthogonally crossed with both target-present and target-absent lineups, allowed us to test rival predictions from the verbal overshadowing theories and the context reinstatement literature. We inserted two ecologically valid delays of two days and five weeks between the exposure to the target and the identification task, which we predicted would result in a positive effect of verbalization on recognition performance. Based on accounts of retrieval-based learning (cf. Karpicke & Grimaldi, 2012) we expected greater identification accuracy for participants who gave a target description compared to those in a control group who did not (verbal facilitation hypothesis).

Furthermore, we expected that re-reading one's description prior to the identification task would serve as a self-generated retrieval cue and, based on an associative memory network model, a mental reinstatement of the encoding context. Re-reading should increase the probability of a correct identification decision compared to a description only and a no description control group (context reinstatement hypothesis).

To further substantiate this assumption, we investigated whether mock witnesses who had not seen the stimulus film but were only given a person description from a yoked witness-participant would be equally able to make a correct identification decision as participant-witnesses who had re-read their own descriptions. Here we sought to rule out the alternative explanation that not the activation of an associative memory network by self-generated retrieval cues was responsible for the expected improvement in identification accuracy but the simple use of *anyone's* person descriptions (i.e., other-generated retrieval cues).

Method

Design

To assess the effect of person descriptions and context-reinstatement on the accuracy of a subsequent lineup identification decision two experiments were conducted (see Table 1). In Experiment 1, a 3 x 2 factorial between-participants design was used. Participants were randomly assigned to one of three groups: A control group in which participants gave no description, a description only group in which participants provided a description of the perpetrator, and a description re-reading group in which participants provided a description of the perpetrator and were allowed to re-read prior to the identification task that took place two days later. The presence versus absence of the target face in the lineup was orthogonally varied. Experiment 2 was identical, however, there was no description only group and the post-description delay was five weeks. Preliminary analyses revealed that the effects reported did not interact with Experiment (1 vs. 2), so we combined the data from both experiments to increase statistical power.

Participants

Across both experiments 208 students participated as a course requirement. Ninety-five students were tested at the Arizona State University (32 males and 59 females, age 17-52, *Mdn* = 20.0), and 117 at the Justus-Liebig University Giessen in Germany (32 males and 85 females, age 19-53, *Mdn* = 22.0). The participants were mainly Caucasian. Participants of other ethnic groups (Mexican-American and Native American, Asian, African) originated from the American sample and were equally distributed across the different conditions. Participants were tested in groups of one to five persons and individually seated in front of personal computers. Preliminary analyses revealed that there were no differences in results across countries. Therefore, location of study will not be considered as a factor.

Materials

Stimulus film.

The stimulus was a high-quality color and sound video film showing a theft of a wallet from a young man's backpack. It was filmed at a sidewalk in a quiet residential area and had two actors. A young Caucasian male ("victim"), 22 years old, was searching through his backpack when he was approached by another male ("thief"), 25 years old, who asked for directions. After the victim had finished giving his directions, his cell phone rang and he walked a few steps away to take the call, turning his back to the thief and leaving his backpack lying on the floor. The thief then quickly bent down and took the victim's wallet out of the backpack. He thanked the victim for the directions and walked off. The victim was still speaking on the phone, saying: "Yes, I've got her number, it's in my wallet, hold on, I'll get it for you." He then went back to his backpack and searched for his wallet. When he could not find it he took several things (water bottle, book, sweater, sheets of paper, keys) out of his backpack. When his wallet still did not show up, he angrily got up and told the person on the phone that the guy who had just asked for directions must have stolen it. The film lasted 1 minute and 34 seconds. Altogether, the thief could be seen for 42 seconds (24 s in close-up).

Photo lineups.

Lineups consisted of six frontal portrait color photographs, presented simultaneously in two rows of three pictures each, and an additional picture next to them on the right depicting a white silhouette of a face on a black background labeled "Not Present". In a pilot study foils were selected (out of 17 possible) based on their similarity to the target. In order to avoid recognition on the basis of clothing cues, all individuals wore a dark blue cloth covering their upper torso.

Four different lineups were composed, one target-present [TP] one target-absent [TA], each with two different target positions (3 = top right and 4 = bottom left) in the lineup. Target-presence and target position were completely counter-balanced across participants and description conditions. Lineups were constructed for both the perpetrator and the victim.

Procedure

After arriving at the laboratory, participants were seated in front of a 35-inch TV screen. Participants were instructed to watch the film attentively because they were going to be asked several questions about it afterwards. Following the film, control group participants (*n* = 80) were allowed to leave.

Description task.

Shortly after seeing the film, participants of the two experimental groups (*n* = 128) were seated in front of the 15-inch screen of a Macintosh computer. The program SuperLab 1.75 (*www.cedrus.com*) was used for all instructions, lineup tests and data collection. At first, participants were instructed to imagine having to give a statement about the incident seen in the film to the police. They were asked to describe the incident in as much detail as possible on an answer sheet. Following this, they were asked some non-leading specific questions about the event (approximately 5 to 10 min). After that, they first gave a free description of the perpetrator, followed by 12 specific, non-leading questions about his appearance (see Appendix A). Participants were instructed to describe the perpetrator as precisely as possible on an answer sheet, so that another person could find him in a crowd. Importantly, they had the opportunity to give "don't know" answers, if they were not able to remember any of the specific features asked for. In Experiment 1, participants were also asked to describe the victim with equivalent instructions and questions.

Re-reading manipulation and identification tasks.

Two days (*n* = 144) or five weeks (*n* = 64) later all participants, including the control group, returned to the laboratory. All participants were presented with the lineup task on a 15-inch computer screen. Prior to the lineup task, only participants in the re-reading group were allowed to read again their own free description of the thief they had given earlier. All participants were presented with unbiased lineup instructions, which stressed the possibility that the thief may or may not be present in the lineup. Following this, they either saw a TP or TA lineup (*n* = 104 in each condition). Afterwards all participants had to give a confidence rating about their lineup decision on an 11-point scale (*0%* to *100%*). Next, the same procedure was repeated for the identification of the victim (in Experiment 1 only). At the end, participants were thanked, asked not to talk about the experiment with future participants, and released. All procedures were in accordance with departmental (based on APA) ethical guidelines.

Rating and Coding of Person Descriptions

The free reports as well as the specific questions regarding the perpetrator's physical appearance were analyzed to obtain separate scores for description quantity (total number of details), the number of correct details, the number of false details and description accuracy (= number of correct details/[number of correct details plus number of false details]). A comprehensive coding scheme was prepared in a pilot study, which was used by two independent coders to code the descriptions.

Preparation of a coding scheme.

First, three persons were instructed to watch the stimulus film carefully and to create as many feature categories considered necessary to capture all possible aspects of the perpetrator's physical appearance that could have been described by any participant of the main study. This resulted in a total of 130 features of the perpetrator's face, hair, body, clothes and accessories.

Then, in a pilot study N = 20 participants individually watched the stimulus film and were asked to describe the perpetrator based on those preset feature categories, as precisely as possible (e.g., "Which eye-color does the perpetrator have?" or "Describe the perpetrator's skin texture."). They were allowed to watch the film as often as necessary and/or to stop it anytime to find all the information needed to answer the questions. For every of the 130 features the most frequently stated answer (the *mode*) was defined as correct. In most cases there were clear modal answers (e.g., 18 out of 20 persons described the perpetrator's eyes as "brown", thus brown eyes were adopted as correct answer in the coding scheme). If there was no clear majority in the participants' answers or if there were any ambiguities in the formulations, two additional coders had to agree on the correct answer. Age was coded correct if it matched the perpetrator's true age (25 years +/- one year). For perpetrator's height and weight answers within a specific range were defined that should be coded as correct (184.5 to 190.5 cm and 77.5 to 82.0 kg).

The final coding scheme resulted in a total of 132 variables (one variable each for age, height and weight, as well as 51 variables describing the perpetrator's face, 8 variables describing his hair, 18 his body, 26 his clothes and 26 describing further accessories).

Ratings of description quality.

First, all descriptions in the free report were rated on three different dimensions regarding their content quality by two independent coders. Ratings referred to a description's precision, specificity and informativeness using 7-point Likert scales (i.e., *1 = not at all precise/specific/informative*; *7 = very precise/specific/informative*).[4] *Precision* was defined as a measure for the elaborateness and clearness of the description. If a description was rated as highly precise, lots of different features were explicitly stated (e.g., facial form, nose, mouth, skin, eyes, ears, hair structure, hair color, clothes). In contrast, descriptions containing just a few vaguely described features were to be rated low in precision (e.g., hair color and length). *Specificity* explicitly referred to the degree of differentiation the features were described

[4] The verbatim definitions used are available from the authors.

with (e.g., "Hair was short, 5-7 cm, brown and curly" vs. "brown hair"). *Informativeness* referred to the description's ability to differentiate the perpetrator from other persons (i.e., to find him in a crowd). Thus in a highly informative description the perpetrator's unique or distinctive features were emphasized (e.g., "He had a small tattoo on his right arm.").

Coding of descriptions for number and accuracy of details.

Descriptions given in free report and in specific questions were coded separately. The coding scheme as well as the descriptions of *N* = 101 participants[5] from the main study were imported into Maxqda2 (*www.maxqda.de*), a software program for qualitative data and text analysis. Two independent coders who were familiar with the complex coding procedure, coded every descriptor as "correct", "incorrect", "confabulated", or "subjective". If a feature was mentioned that the perpetrator did not have (e.g., he wore glasses, although he did not wear any) the item was treated as a confabulation. Items coded as subjective were descriptions idiosyncratic for a participant (e.g., "He looked like my brother"; "He was handsome"). For details not included in the coding scheme a "rest" category was used.

After the coding process, incorrect and confabulated items were merged into a category of "false" descriptors due to low frequencies. Categories of subjective details and remaining features were excluded from further analyses for the same reason (in the free reports, less than 50% and

[5] Unfortunately, 27 free report and specific questions descriptions were lost in the process of moving offices. Thus, 101 descriptions could be used for these analyses.

less than 25% of participants mentioned one or more feature that was coded in these categories).

Subsequent data collection

To eliminate the mere use of descriptions as an alternative explanation of the proposed re-reading effect an additional $N = 128$ participants (76 male, 52 female, age 16-79, $Mdn = 26.0$), who did not see the stimulus film, were instructed to identify the perpetrator solely based on the 128 free person descriptions given by the participants in the two main experiments. First, every participant had to read one yoked participant's free target description, and was then immediately shown the same lineup the corresponding participant of the main experiments had seen. Participants received the same unbiased lineup instructions used in the main study, that is, they were informed about the possibility to reject the lineup.

Results

Preliminary Analyses: Description Use as Alternative Explanation for the Expected Re-Reading Effect?

To rule out the possibility that simply reading anyone's description might be responsible for correct identification decisions, we compared identification accuracy of the two re-reading groups from both experiments and mock witnesses who had not seen the film but were asked to select the target based on a yoked description of an experimental participant. Identification accuracies were clearly higher for re-readers (85.0%) than mock witnesses (37.5%), LR $chi^2(1, N = 160) = 40.15$, $p < .001$, $OR = 9.44$. Likewise, participants in the description only group of Experiment 1 (79.2%) made significantly more correct identification decisions than mock witnesses given these descriptions (33.3%), LR $chi^2(1, N = 96) = 21.35$, $p < .001$, OR = 7.60.

Overview of Further Analyses

Preliminary analyses revealed that the results of Experiment 1 and 2 were completely parallel, with the exception of the influence of delay. Therefore, we first present the joint results of Experiment 1 and 2 regarding the potential benefit of re-reading on the accuracy of identification decisions. Subsequently, we analyze the effect of prior describing a perpetrator on subsequent identifications (i.e., no description control group [CG] vs. the two description groups combined) to differentiate between verbal overshadowing vs. verbal facilitation effects. [6]

In light of the ongoing controversy regarding the most appropriate analysis of identification data, we first report *diagnosticity ratios* (DR) as well as the *signal-detection theory* based analyses of receiver operating characteristic (ROC) curves (cf. Gronlund, Mickes, Wixted, & Clark, 2015; Gronlund, Wixted, & Mickes, 2014; Mickes, Flowe, & Wixted, 2012; Mickes, Moreland, Clark, & Wixted, 2014; Wells, Smalarz, & Smith, 2015; Wells, Yang, & Smalarz, 2015; Wixted & Mickes, 2012, 2014). We are aware that conducting repeated comparisons potentially inflate the type-1 error rate. However, the main purpose here was to demonstrate that similar conclusions can be drawn no matter which of these analyses are conducted.

---

[6] Effects of verbalization and re-reading on choosing rates and confidence are available on request. We do encourage future researchers to investigate whether the confidence-accuracy relationship is affected by re-reading, or more generally, by verbalization of the target's face.

Further, data were analyzed via four hierarchical loglinear frequency analyses (SPSS Hiloglinear, hierarchical backward elimination method), which are considered appropriate when both independent and dependent variables are categorical in nature (Howell, 2013; Tabachnik & Fidell, 2007; cf. also Meissner et al., 2001).

To describe the size of effects for categorical variables we report *odds ratios* (*OR*), that is, the ratio of the odds for a given outcome (e.g., the odds for a correct identification decision) in one group divided by the odds for the same outcome in another group (cf. Fleiss, 1994; Lipsey & Wilson, 2001). For example, in the re-reading group, the odds of a correct lineup decision are the proportion of correct decisions (e.g., 0.85) divided by the proportion of incorrect decisions (e.g., 0.15), which equals 5.7. Thus, in this condition the odds for correct decisions are almost six times higher than those for incorrect decisions. In contrast, in the no description CG the odds for a correct decision were 0.63/0.37 = 1.7. The *OR* is the ratio between these two odds, which is 5.7/1.7 = 3.4 in this case. Hence, the odds for a correct lineup decision in the re-reading group are 3.4 times higher than in the no description CG. An *OR* of 1 means the odds of correct lineup decisions to be the same for both groups.

At the end, point-biserial correlations between the different description measures and identification accuracy as a function of delay and experimental condition (description only vs. re-reading group) are reported.

Diagnosticity Ratios vs. Receiver Operating Characteristic Analyses

Diagnosticity ratios (= correct identifications in target-present lineups/[false identifications in target-absent lineups/lineup size]), signal detection theory performance *d*' and response criterion *C* (cf. Macmillan & Creelman, 2005) for the different description groups are displayed in Table 2. The indices *d*' and *C* were calculated based on hit and false alarm rates, with foil identifications in TP lineups excluded from the analysis (cf. Meissner, Tredoux, Parker, & MacLin, 2005). False alarm rates were calculated in the same way as for diagnosticity ratios, that is, false identifications in TA lineups were divided by nominal lineup size (see "Estimated FAs" in Table 2).

Moreover, receiver operating characteristic (ROC) curves for each description condition were constructed based on the current approach illustrated by Gronlund et al. (2014) using participants' identification response data and confidence ratings. Therefore, the cumulative hit and false alarm rates (HR and FAR) were plotted at each confidence level, ranging from conservative responding at the one end (i.e., identifications with high confidence levels only) to a more liberal response criterion at the other end. Again, false alarm rates were corrected by the nominal size of the lineup, and foil identifications in TP lineups were excluded from the analysis. Then, partial area under the curve (pAUC) analyses were conducted to compare identification discriminability across the different description conditions. The procedure yielding the higher pAUC is considered diagnostically superior, that is, yielding a higher ability to distinguish between innocent and guilty suspects in a lineup. Partial AUC analyses are appropriate here, because the data do not include the full range of HR and FAR from 0 to 1. For each comparison we selected the maximum FAR as cutoff point. pAUCs were computed and compared using the data analysis package pROC for R (Robin et al., 2011), applying the bootstrapping method (with the number of bootstraps set to 10000). For comparing the two pAUCs the following formula was used: *D* = (AUC1-AUC2)/*s*, where *s* is the standard deviation of the bootstrap differences and AUC1 and AUC2 are the areas under the curve of the two ROC curves (Robin et al., 2011).

Parallel to our further analyses described below, we first compared the identification performance of the no description control group and the re-reading groups in both experiments. The pAUC of the no description control group (pAUC = 0.012, $CI_{95\%}$ = 0.006 - 0.022) was significantly smaller than the pAUC of the re-reading groups (pAUC = 0.031, $CI_{95\%}$ = 0.021 - 0.041), *D* = -2.75, *p* = .006. The ROC curves are displayed in Figure 1, whereby FARs were divided by the total number of persons in the lineup (i.e., six).

Next, possible effects of verbalization on identification discriminability (VOE or VFE) were tested. The difference between the pAUCs of both description conditions joined (pAUC = 0.024, $CI_{95\%}$ = 0.015 - 0.037) and the no description control group (pAUC =0.012, $CI_{95\%}$ = 0.006 - 0.022) was only marginally significant, *D* = -1.85, *p* = .064.

In Experiment 1, comparing the pAUCs of the description only group (pAUC = 0.012, $CI_{95\%}$ = 0.007 - 0.024) with the no description control group (pAUC = 0.010, $CI_{95\%}$ = 0.005 - 0.023), showed no significant difference, *D* = -0.26, *p* = .798.

Effects of Re-Reading on Identification Accuracy

Comparing the re-reading with the no description control group.

First, we focus on the conditions present in both experiments (*N* = 160), thus excluding the description only condition in Experiment 1. Table 2 and Figure 2 give an overview of the accuracy of identification decisions and other outcomes in the different conditions in both experiments.

Data were analyzed via a four-way hierarchical loglinear frequency analysis including re-reading condition (no description CG vs. re-reading group), delay (2 days vs. 5 weeks), target-presence (TA vs. TP) and accuracy of identification decisions (coded as a binary variable: *0 = incorrect* [false rejection or filler identification in TP lineups; any positive identification in TA lineups]; *1 = correct* [hit in TP lineups or correct rejection in TA lineups]).

Stepwise selection by simple deletion of effects produced a model that included only first-order effects and two-way associations. Tests of k-way interactions revealed that none of the three-way or four-way interactions were significant. The final model had a likelihood ratio $chi^2(8, N = 160) = 5.93$, $p = .655$, indicating a good fit between observed and expected frequencies generated by the model. There were highly significant partial associations between identification accuracy and re-reading, partial $chi^2(1, N = 160) = 12.37$, $p < .001$, identification accuracy and delay, partial $chi^2(1, N = 160) = 15.67$, $p < .001$, and identification accuracy and target-presence, partial $chi^2(1, N = 160) = 7.62$, $p = .006$. Re-reading one's prior description improved identification accuracy from 62.5% in the no description CG to 85.0% in the re-reading group (*OR* = 3.40). Thus, the odds of correct identification decisions were 3.4 times higher in the re-reading group than in the no description CG. Performance deteriorated from 84.4% after two days to 57.8% after five weeks (*OR* = 3.94). Perhaps surprisingly, participants were better with TA lineups than when tested with TP lineups (82.5% vs. 65.0%, *OR* = 0.39).

Although there was no evidence for the existence of three-way interactions, it is worth noting that the benefit of re-reading was at least as strong, if not stronger, in TA lineups (95.0% vs. 70.0% correct rejections; *OR* = 8.14) as in TP lineups (75.0% vs. 55.0% hits; *OR* = 2.45). Notably, the effect was comparable in magnitude for the two-day (91.7% vs. 77.1% correct decisions; *OR* = 3.27) and five-week delay (75.0% vs. 40.6% correct decisions; *OR* = 4.38).

Comparing the re-reading with the description only group.

Although re-readers apparently yielded slightly better identification outcomes than participants in the description only group (see Table 2), none of the differences, neither for TP nor for TA lineups, were statistically significant, all *p*s > .148.

Effects of Verbalization on Identification Accuracy: Verbal Facilitation or Verbal Overshadowing?

The data presented above demonstrate that re-reading one's description facilitates performance rather than impairing it. In Experiment 1, a description only control group was used to test whether this beneficial effect is restricted to re-reading or if giving a description itself is sufficient to increase identification accuracy.

Comparing both description groups of Experiment 1 and 2 with the control group.

Data were analyzed via a four-way hierarchical loglinear frequency analysis including description condition (no description [CG] vs. description [description only plus re-reading]), delay (2 days vs. 5 weeks), target-presence (TA vs. TP) and identification accuracy (*0 = incorrect*; *1 = correct*).

The likelihood ratio for the final model was $chi^2(2, N = 208) = 0.45$, $p = .798$, revealing a good model fit. Tests of k-way effects revealed that all three-way or four-way interactions could be excluded from the model. Besides the already known highly significant associations between identification accuracy and delay as well as target presence (all *p*s < .002), there was a significant two-way effect of Description Condition x Identification Accuracy, partial $chi^2(1, N = 208) = 8.06$, $p = .005$. Combining the description only group of Experiment 1 (79.2%) and the two re-reading groups (85.0%; *M* of both conditions = 82.8%) and comparing them with the no description CG (62.5%) yielded a highly significant VFE (*OR* = 2.89). Thus, when person descriptions were given, the odds of correct identification decisions were almost three times higher than in the no description CG.

Although the three-way interaction was not significant, separate analyses within TA and TP lineups yielded similar conclusions. In the TA condition, performance was significantly higher in the two description conditions combined (92.2%) than in the control condition (70.0%), LR $chi^2(1, N = 104) = 8.67$, $p = .003$, *OR* = 5.06. In the TP condition the effect failed to reach significance, with a tendency for higher identification accuracies in the description conditions (73.4%) than in the no description CG (55.0%), LR $chi^2(1, N = 104) = 3.71$, $p = .054$, *OR* = 2.26.

Comparing the description only with the control group.

Only in Experiment 1, it was possible to test the traditional VOE by comparing the description only group (79.2%) with the no description CG (77.1%), which was far from significant, LR $chi^2(1, N = 96) = 0.06$, $p = .805$, *OR* = 1.13. Separate analyses for TA and TP lineups also revealed neither a

VOE nor a VFE effect, both *p*s > .40 (correct rejections: *OR* = 1.84; hits: *OR* = 0.81; see the means in Table 2).

Replication of the Effects With Another Target: Victim Identification Accuracy

In order to replicate the observed effects of re-reading and verbalization on identification accuracy, parallel analyses were conducted for the identification of the victim in Experiment 1. There was a significant effect of re-reading, LR $chi^2(1, N = 96) = 6.25$, $p = .012$, $OR = 2.87$, with 70.8% correct identification decisions for re-readers compared to 45.8% for the no description CG. Moreover, participants who gave a victim description (re-reading plus description only group combined) performed better at the identification task (67.7%) than the no description CG (45.8%), LR $chi^2(1, N = 144) = 6.35$, $p = .012$, $OR = 2.48$. There was a marginally significant VFE without re-reading, with 64.6% of the description only group making a correct lineup decision compared to 45.8% of the control group, LR $chi^2(1, N = 96) = 3.43$, $p = .064$, $OR = 2.16$.

Quantity and Quality of Person Descriptions as "Postdictors" of Identification Accuracy

Results are presented separately for free reports and specific questions. As only free descriptions were re-read prior to the identification task, we focused on the relationship between free report description measures and identification accuracy. Features mentioned twice (i.e., in both free report and in specific questions) were not coded separately, and thus, no overall description score was calculated.

Details coded for different feature categories (face, hair, body, clothes and accessories) were summed up for the total number of details reported regarding the whole person. Accuracy scores were only calculated for the whole person.

Inter-coder reliabilities.

Different measures of inter-coder reliability were computed separately for free reports and for specific questions (Pearson correlation coefficient [*r*] and intra-class correlation coefficients [*ICC*]). All values were highly satisfactory.

Pearson correlations between the two coders for the total number of descriptors in free reports were $r(99) = .90$, and $r(99) = .93$ for specific questions (both *ICC*s = .95). Codings for the number of correct details correlated by $r(99) = .89$ in free reports, and by r(99) = .92 in specific questions (both *ICC*s =.94). Correlations for the number of false details were $r(99) = .71$ (*ICC* =.83), and r(99) = .82 (*ICC* =.90), respectively. Accuracy scores correlated by $r(99) = .67$ in free reports, and by $r(99) = .87$ in specific questions (*ICC*s = .81 and .91, respectively).

For the ratings, reliabilities were satisfactory as well (precision: $r(99) = .66$, *ICC* = .78; specificity: $r(99) = .62$, *ICC* = .75; informativeness: $r(99) = .61$, *ICC* = .71). For further analyses the means of the two coders' ratings were used.

Descriptive statistics of description quantity and quality.

In free reports $M = 10.68$ ($SD = 4.18$) details were mentioned concerning the perpetrator's physical appearance. Thereof, $M = 7.51$ ($SD = 3.76$) descriptors were coded as correct, and $M = 3.17$ ($SD = 1.87$)

descriptors as false. This resulted in an overall accuracy rate of 69.14% (*SD* = 17.74). In specific questions participants reported *M* = 14.05 (*SD* = 3.57) details in total, whereof *M* = 8.79 (*SD* = 3.42) were coded as correct and *M* = 5.26 (*SD* = 1.98) were coded as false. Accuracy rate was *M* = 61.76% (*SD* = 14.01). The distribution of details across the different feature categories (face, hair, body, clothes and accessories), separately for free report and specific questions is displayed in Table 3. Specific questions yielded more correct and more false details but lower accuracy than free reports, all ps < .001.

Relationship between identification performance and description quantity and accuracy.

All correlations between identification accuracy and description measures are shown in Table 4. For the total sample, in free reports there were significant positive correlations between identification accuracy and the number of descriptors, $r_{pb}(99) = .23$, and the number of correct descriptors, $r_{pb}(99) = .22$. These correlations were also significant in the re-reading group, $r_{pb}(66) = .29$ and .28, whereas all correlations in the description only group remained nonsignificant. For specific questions the same pattern of results emerged. Correlations after five weeks tended to be larger than for identifications after two days but the samples are too small to test for significance of differences in correlations.

Relationship between identification performance and ratings of description quality.

Finally, we examined whether identification performance covaried with the rated degree of description precision, specificity and informativeness. Additionally, based on high inter-correlations between these variables, we computed a mean composite rating across all four ratings (Cronbach's *alpha* = .95). Point-biserial correlations are displayed in Table 5. In the re-reading group each rating variable was significantly positively correlated with identification accuracy. In contrast, for the description only group none of these correlations were reliable.

Discussion

The present study aimed to investigate the potentially beneficial effect of describing a target's physical appearance on a subsequent identification. In particular, we explored the usefulness of re-reading one's own free target description prior to the identification task as a new system variable. As a possible explanation for the expected beneficial effect of re-reading, the relationships between identification performance and description quantity, accuracy as well as quality ratings were considered. A primary concern was to ensure ecological validity. Hence, we inserted a retention interval of either two days or five weeks between observation of the crime and the identification task. Person descriptions were collected according to common police practice, that is, a free report was followed by open-ended questions. By allowing a “Don't Know” option we discouraged the self-generation of false details. Both unbiased TA and TP lineups were used to assess the effects on identification accuracy.

Verbal Facilitation vs. Verbal Overshadowing

Contrary to assumptions in the VOE literature (e.g., Meissner & Brigham, 2001; Meissner et al., 2008; Schooler & Engstler-Schooler, 1990) identification performance across all description groups, with or without re-

reading (83%) compared to a no description control group (63%) was superior. The odds of correct identification decisions were almost three times larger when prior descriptions were given compared to the no description group (*OR* = 2.89).

The effect was especially strong for TA lineups (correct rejections: 93% vs. 70%, *OR* = 5.06), while there was only a nonsignificant tendency of facilitation when the target was present (hits: 73% vs. 55%, *OR* = 2.26). Even when re-readers were not considered in the analysis, thus comparing the description only group with the no description CG in Experiment 1, results showed that verbalization per se did not impair identification performance (*OR* = 1.13). Note that Meissner and Brigham's (2001) meta-analysis also showed a VFE after the insertion of a post-description delay of more than thirty minutes. Hence, our results support the assumption that realistic delays of more than two days or even several weeks, in combination with conservative description instructions, seem to annihilate or even reverse the VOE.

The present results are in line with the assumed memory advantages that are accompanied by early first retrieval attempts (cf. the *testing effect*; McDermott et al., 2014; Roediger & Karpicke, 2006), especially after longer delays (cf. Butler & Roediger, 2007; Chan, 2009).

In sum, these findings suggest that under ecologically valid conditions, if identification performance is affected by verbalization at all, it seems to be in a positive, and not, as assumed by the VOE hypothesis, in a negative way. Longer retention intervals of several days or weeks are more representative of real crime situations (see Sporer, 1996; Wells, Steblay, et al., 2015)

Re-Reading as an Effective Retrieval Cue after a Long Delay

As predicted, we found a positive effect of context reinstatement by re-reading one's description prior to the identification task. Identification accuracy was higher for participants that were allowed to re-read their description prior to identification (85%) than for participants who were not allowed to do so (63%), with odds for re-readers to make an accurate identification decision more than three times as high than the odds for non-re-readers (*OR* = 3.40). The effect was internally replicated for the identification of the victim (*OR* = 2.87).

Moreover, *diagnosticity ratios* were much higher in the re-reading group (90.0) compared to the no description control group (11.0). Thus, after re-reading witnesses were 90 times more likely to identify the target than to choose an innocent suspect. Due to the current criticisms concerning *diagnosticity ratios* (e.g., Wixted & Mickes, 2012) signal-detection based ROC analyses were computed, which led exactly to the same conclusions.

The observed beneficial effect of re-reading tended to be stronger for TA (*OR* = 8.14) than for TP lineups (*OR* = 2.45). Thus, there was a tendency for re-reading to be more effective to reduce false identification in TA lineups than to increase hits in TP lineups. The positive re-reading effects did not differ as a function of the different delays, a finding future studies should explore. As re-readers and participants of the description only group did not differ in their identification performance, a joint effect of giving a person description and re-reading is assumed. However, based on Experiment 1, describing the perpetrator's face alone is not sufficient to explain the observed improvement in identification accuracy on its own.

Furthermore, the observed re-reading benefit seems not just to be due to the use of an appropriate person description. As the mock witness data reveal, merely reading someone else's person description does not suffice to select the target in a lineup (nor to correctly reject it). Mock witnesses who did not watch the stimulus film but only read someone else's person description yielded much lower identification accuracy. Thus, our results are compatible with the view that re-reading one's own description of the target does induce processes akin to spreading activation in a memory network in which the original target face is embedded (cf. Anderson, 1983). In line with Paivio's (1971) dual coding hypothesis, we assumed that describing the target should lead to coding features about the target not only visually but also verbally. Presumably, re-reading the description prior to identification would lead to a reinstatement of this verbal code as a self-generated retrieval cue as well as to a re-activation of the associated encoding context resulting in better recognition performance.

To further understand the underlying mechanism that affects person memory and identification, future research should examine whether the observed benefits due to re-reading are restricted to the use of one's own description as a self-generated retrieval cue compared to being given another person's (e.g., a co-witness's) description (other-generated cues; cf. Mäntylä, 1986; Sporer, 1991).

Robustness of the Re-Reading Effect: A Meta-Analysis

Several studies have found some form of re-reading effect (Cutler et al., 1986; Cutler et al., 1987; Sauerland et al., 2008; Sporer, 2007). To demonstrate the robustness of our findings including conceptual replications with different stimulus materials (films, targets, and lineups), we conducted several small meta-analyses, which include the four studies mentioned above as well as the present results. Following the procedures of Lipsey and Wilson (2001), mean weighted effect sizes (*OR*s) for different lineup outcomes (overall correct identification decisions [TA and TP], hits [TP] and correct rejections [TA]) were calculated, comparing re-reading vs. no description control groups (Table 6) and re-reading vs. description only groups (Table 7). All effects for the comparison between a re-reading and a no description CG were reliable, with *OR*s between 2 and 3. Regarding the comparison between a re-reading and a description only group, there was only a beneficial effect for correct rejections (*OR* = 1.79), which seems to be most important regarding preventing mistaken identifications. Moreover, the results support the idea that verbalization and re-reading make people aware of their poor memory of the target, which in turn may make them more cautious at the identification task (cf. Sauerland et al., 2008; Sporer, 2007).

Hence, re-reading one's own description may be worth considering as a simple and easy to implement system variable to improve identification accuracy.

Quantity and Quality of Person Descriptions as Possible Reasons for the Re-Reading Effect

To test the assumption that the beneficial effect of re-reading might be associated with the quantity and quality of the re-read person descriptions, relationships between description measures and identification accuracy were investigated.

Description quantity and accuracy.

Different from the findings in Meissner et al.'s (2008) meta-analysis, in free reports and in specific questions, we found moderate positive associations between identification accuracy with the number of details reported (cf. Sporer, 1992) as well as with the number of correct details. However, these relationships were significant among re-readers only, being nearly equal in size for free reports and specific questions. It seems that recognition performance was facilitated when the descriptions that participants were allowed to re-read contained a large number of (correct) details. In contrast, in the description only group, none of the relationships were significant. However, relationships with specific questions were comparable in magnitude to those in the re-reading group, although not significant due to the smaller sample sizes.

But how could identification accuracy be related to the quantity of the re-read description for the perpetrator, while it was independent of description accuracy? As a possible explanation we assume that the mere number of (correct) details was decisive, irrespective of the presence of false details or the overall description accuracy. Hence, re-reading may have been beneficial not via directly activating the memory for the true target's physical appearance, but by activating the associative network (Anderson, 1983) in which the face of the perpetrator was embedded, and thus providing participants a sound basis for the identification task.

Although these analyses are correlational, and thus, do not allow causal conclusions, natural variations in description quantity and accuracy, which often occur in real cases (Sporer, 1996) seem to be associated with the benefits of re-reading.

Correlations between identification accuracy and number of details as well as number of correct details were moderate in size ($r_{pb} > .30$) after the longer delay but tended to be smaller after a delay of two days (but sample sizes are too small to test for differences in correlations). The results are in line with the assumption that a large number of initially recalled descriptors might be more beneficial for identifications after longer delays (cf. Butler & Roediger, 2007). We encourage future researchers to use even longer retention intervals to test this.

Quality of description ratings.

As the true accuracy of a given description is not known in actual criminal cases, we additionally rated descriptions quality given in free reports in terms of precision, specificity and informativeness. For re-readers all of the quality ratings as well as the composite mean rating were positively associated with identification accuracy. Thus, the current results confirm our expectation that the beneficial effect of re-reading on identification performance is associated with the quality of the re-read descriptions.

The present results are consistent with findings from police records by Valentine et al. (2003), who observed positive associations between identification of the suspect (which may or may not have been the perpetrator) and description completeness, which is vaguely comparable to our "precision" variable.

Conclusions and Practical Implications

This research provides a new look at the role of person descriptions and identifications. We replicated across two ecologically valid delays that the practice of having witnesses re-read their own prior descriptions showed better identification performance in both TP and TA lineups. Re-reading one's description is an easily applicable system variable (Wells, 1978) that does not require any additional procedures, training, or resources. Our results also contradict the VOE that has not only "declined" over decades (Schooler, 2011) but may be reversed if examined in situations representative of real crimes. Consequently, the initially cited proscription made by the Police and Criminal Evidence Act (1984) not to show the witness any descriptions of the suspect prior to the identification task appears inappropriate in light of our data.

References

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

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior, 22,* 261-295. doi: 10.1016/S0022537183902013

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. 221-235). New York: Russel Sage Foundation.

Bower, G. H. (1981). Mood and memory. *American Psychologist, 36,* 129-148. doi: 10.1037//0003-066X.36.2.129

Brown, C., Gehrke, J., & Lloyd-Jones, T. J. (2010). A visual and semantic locus to beneficial effects of verbalization on face memory. *American Journal of Psychology, 123,* 51-69. doi: 10.5406/amerjpsyc.123.1.0051

Brown, C., & Lloyd-Jones, T. J. (2005). Verbal facilitation of face recognition. *Memory & Cognition, 33,* 1442-1456. doi: 10.3758/bf03193377

Brown, C., & Lloyd-Jones, T. J. (2006). Beneficial effects of verbalization and visual distinctiveness on remembering and knowing faces. *Memory & Cognition, 34,* 277-286. doi: 10.3758/BF03193406

Butler, A. C., & Roediger, H. L. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology, 19,* 514-527. doi: 10.1080/09541440701326097

Chan, J. C. K. (2009). When does retrieval induce forgetting and when does it induce facilitation? Implications for retrieval inhibition, testing effect, and text processing. *Journal of Memory and Language, 61,* 153-170. doi: 10.1016/j.jml.2009.04.004

Chance, J., & Goldstein, A. G. (1976). Recognition of faces and verbal labels. *Bulletin of the Psychonomic Society, 7,* 384-386. doi: 10.3758/bf03337223

Clare, J., & Lewandowsky, S. (2004). Verbalizing facial memory: Criterion effects in verbal overshadowing. *Journal of Experimental Psychology-Learning Memory and Cognition, 30,* 739-755. doi: 10.1037/0278-7393.30.4.739

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11,* 671-684. doi: 10.1016/S0022-5371(72)80001-X

Cutler, B. L., Penrod, S. D., & Martens, T. K. (1987). 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

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*(2), 113-134.

Ebbesen, E. B., & Rienick, C. B. (1998). Retention interval and eyewitness memory for events and personal identifying attributes. *Journal of Applied Psychology, 83,* 745-762. doi: 10.1037/0021-9010.83.5.745

Ebbinghaus, H. (1913). *Memory: A contribution to experimental psychology.* New York: Teachers College, Columbia University.

Ellis, H. D. (1984). Practical aspects of face memory. In G. L. Wells & E. F. Loftus (Eds.), *Eyewitness testimony* (pp. 12-37). Cambridge: Cambridge University Press.

Fahsing, I. A., Ask, K., & Granhag, P. A. (2004). The man behind the mask: Accuracy and predictors of eyewitness offender descriptions. *Journal of Applied Psychology, 89,* 722-729. doi: 10.1037/0021-9010.89.4.722

Fallshore, M., & Schooler, J. W. (1995). Verbal vulnerability of perceptual expertise. *Journal of Experimental Psychology-Learning Memory and Cognition, 21,* 1608-1623. doi: 10.1037/0278-7393.21.6.1608

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

Fleiss, J. L. (1994). Measures of effect sizes for categorical data. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 245-260). New York: Russell Sage Foundation.

Geiselman, R. E., Fisher, R. P., MacKinnon, D. P., & Holland, H. L. (1985). Eyewitness memory enhancement in the police interview: Cognitive retrieval mnemonics versus hypnosis. *Journal of Applied Psychology, 70,* 401-412. doi: 10.1037//0021-9010.70.2.401

Gronlund, S. D., Mickes, L., Wixted, J. T., & Clark, S. E. (2015). Conducting an eyewitness lineup: How the research got it wrong. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 63, pp. 1-43). Amsterdam: Elsevier.

Gronlund, S. D., Wixted, J. T., & Mickes, L. (2014). Evaluating eyewitness identification procedures using receiver operating characteristic analysis. *Current Directions in Psychological Science, 23,* 3-10. doi: 10.1177/0963721413498891

Hope, L., Gabbert, F., & Fisher, R. P. (2011). From laboratory to the street: Capturing witness memory using the self-administered interview. *Legal* *and Criminological Psychology, 16,* 211-226. doi: 10.1111/j.2044-8333.2011.02015.x

Howell, D. C. (2013). *Statistical methods for psychology.* Belmont, CA: Wadsworth Cengage Learning.

Itoh, Y. (2005). The facilitating effect of verbalization on the recognition memory of incidentally learned faces. *Applied Cognitive Psychology, 19,* 421-433. doi: 10.1002/acp.1069

Karpicke, J. D., & Grimaldi, P. J. (2012). Retrieval-based learning: A perspective for enhancing meaningful learning. *Educational Psychology Review, 24,* 401-418. doi: 10.1007/s10648-012-9202-2

Kitagami, S., Sato, W., & Yoshikawa, S. (2002). The influence of test-set similarity in verbal overshadowing. *Applied Cognitive Psychology, 16,* 963-972. doi: 10.1002/acp.917

Krafka, C., & Penrod, S. (1985). Reinstatement of context in a field experiment on eyewitness identification. *Journal of Personality and Social Psychology, 49,* 58-69. doi: 10.1037/0022-3514.49.1.58

LaPaglia, J. A., & Chan, J. C. K. (2012). Retrieval does not always enhance suggestibility: Testing can improve witness identification performance. *Law and Human Behavior, 36,* 478-487. doi: 10.1037/h0093931

Lindsay, R. C. L., Martin, R., & Webber, L. (1994). Default values in eyewitness descriptions: A problem for the match-to-description lineup foil selection strategy. *Law and Human Behavior, 18,* 527-541. doi: 10.1007/bf01499172

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

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

MacLin, M. K. (2002). The effects of exemplar and prototype descriptors on verbal overshadowing. *Applied Cognitive Psychology, 16,* 929-936. doi: 10.1002/acp.923

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide.* New York: Lawrence Erlbaum Associates.

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

Mäntylä, T. (1986). Optimizing cue effectiveness: Recall of 500 and 600 incidentally learned words. *Journal of Experimental Psychology: Learning Memory and Cognition, 12,* 66-71. doi: 10.1037/0278-7393.12.1.66

McDermott, K. B., Arnold, K. M., & Nelson, S. M. (2014). The testing effect. In T. J. Perfect & D. S. Lindsay (Eds.), *The SAGE handbook of applied memory* (pp. 183-200). Los Angeles, CA: SAGE.

McKelvie, S. J. (1976). The effects of the verbal labeling on recognition memory for schematic faces. *Quarterly Journal of Experimental Psychology, 28,* 459-474. doi: 10.1080/14640747608400573

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

Meissner, C. A., Brigham, J. C., & Kelley, C. M. (2001). The influence of retrieval processes in verbal overshadowing. *Memory & Cognition, 29,* 176-186. doi: 10.3758/bf03195751

Meissner, C. A., Sporer, S. L., & Schooler, J. W. (2007). Person descriptions as eyewitness evidence. In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Handbook of eyewitness psychology: Memory for people* (Vol. 2, pp. 3-34). Mahwah, NJ: Lawrence Erlbaum and Associates.

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

Meissner, C. A., Tredoux, C. G., Parker, J. F., & MacLin, O. H. (2005). Eyewitness decisions in simultaneous and sequential lineups: A dual-process signal detection theory analysis. *Memory & Cognition, 33,* 783-792. doi: 10.3758/bf03193074

Mickes, L., Flowe, H. D., & Wixted, J. T. (2012). Receiver operating characteristic analysis of eyewitness memory: Comparing the diagnostic accuracy of simultaneous versus sequential lineups. *Journal of Experimental Psychology: Applied, 18,* 361-376. doi: 10.1037/a0030609

Mickes, L., Moreland, M. B., Clark, S. E., & Wixted, J. T. (2014). Missing the information needed to perform ROC analysis? Then compute d', not the diagnosticity ratio. *Journal of Applied Research in Memory and Cognition, 3,* 58-62. doi: 10.1016/j.jarmac.2014.04.007

Neil v. Biggers, 409 U.S. 188 (1972).

Paivio, A. (1971). *Imagery and verbal processes.* New York: Holt.

Police and Criminal Evidence Act (1984). Code D: Code of practice for the identification of persons by police officers. Retrieved from:

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/253831/pace-code-d-2011.pdf

Read, J. D. (1979). Rehearsal and recognition of human faces. *American Journal of Psychology, 92,* 71-85. doi: 10.2307/1421480

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J. C., & Muller, M. (2011). pROC: An open-source package for R and S plus to analyze and compare ROC curves. *BMC Bioinformatics, 12,* 77. doi: 10.1186/1471-2105-12-77

Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1,* 181-210. doi: 10.1111/j.1745-6916.2006.00012.x

Rubin, D. C., & Wenzel, A. E. (1996). One hundred years of forgetting: A quantitative description of retention. *Psychological Review, 103,* 734-760. doi: 10.1037/0033-295X.103.4.734

Sauerland, M., Holub, F. E., & Sporer, S. L. (2008). Person descriptions and person identifications: Verbal overshadowing or recognition criterion shift? *European Journal of Cognitive Psychology, 20,* 497-560. doi: 10.1080/09541440701728417

Schooler, J. W. (2011). Unpublished results hide the decline effect. *Nature, 470,* 437. doi: 10.1038/470437a

Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology, 22,* 36-71. doi: 10.1016/0010-0285(90)90003-m

Shapiro, P. N., & Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological Bulletin, 100,* 139-156. doi: 10.1037/0033-2909.100.2.139

Shepherd, J., Davies, G., & Ellis, H. (1981). Studies of cue saliency. In G. Davies, H. Ellis, & J. Shepherd (Eds.), *Perceiving and remembering faces* (pp. 105-132). London: Academic Press.

Smith, H. M. J., & Flowe, H. D. (2014). ROC analysis of the verbal overshadowing effect: Testing the effect of verbalization on memory sensitivity. *Applied Cognitive Psychology*. Advance online publication. doi: 10.1002/acp.3096

Sporer, S. L. (1988). Long-term improvement of facial recognition through visual rehearsal. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 182-188). Chichester: Wiley.

Sporer, S. L. (1989). Verbal and visual processes in person identification. In H. Wegener, F. Loesel, & J. Haisch (Eds.), *Criminal behavior and the criminal justice system* (pp. 303-324). New York: Springer.

Sporer, S. L. (1991). Deep-deeper-deepest? Encoding strategies and the recognition of human faces. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 323-333. doi: 10.1037/0278-7393.17.2.323

Sporer, S. L. (1992). Post-dicting eyewitness accuracy: Confidence, decision-times and person descriptions of choosers and non-choosers. *European Journal of Social Psychology, 22,* 157-180. doi: 10.1002/ejsp.2420220205

Sporer, S. L. (1996). Psychological aspects of person descriptions. In S. L. Sporer, R. S. Malpass, & G. Koehnken (Eds.), *Psychological issues in person identification* (pp. 53-86). Mahwah, NJ: Lawrence Erlbaum.

Sporer, S. L. (2007). Person descriptions as retrieval cues: Do they really help? *Psychology Crime & Law, 13,* 591-609. doi: 10.1080/10683160701253986

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics.* Boston: Pearson.

Tanaka, J. W., & Farah, M. J. (2003). The holistic representation of faces. In M. A. Peterson & G. Rhodes (Eds.), *Perception of faces, objects and scenes: Analytic and holistic processes* (pp. 53-74). Oxford: Oxford University Press.

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

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

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

Wells, G. L., Smalarz, L., & Smith, A. M. (2015). ROC Analysis of lineups does not measure underlying discriminability and has limited value. *Journal of Applied Research in Memory and Cognition, 4,* 313-317. doi: 10.1016/j.jarmac.2015.08.008

Wells, G. L., Steblay, N. K., & Dysart, J. E. (2015). Double-blind photo lineups using actual eyewitnesses: An experimental test of a sequential versus simultaneous lineup procedure. *Law and Human Behavior, 39,* 1-14. doi: http://dx.doi.org/10.1037/lhb0000096

Wells, G. L., Yang, Y. R., & Smalarz, L. (2015). Eyewitness identification: Bayesian information gain, base-rate effect-equivalency curves, and reasonable suspicion. *Law and Human Behavior, 39,* 99-122. doi: 10.1037/lhb0000125

Wickham, L. H. V., & Lander, K. (2008). The effect of verbal description and processing type on face identification. *European Journal of Cognitive Psychology, 20,* 577-586. doi: 10.1080/09541440701728433

Wixted, J. T., & Mickes, L. (2012). The field of eyewitness memory should abandon probative value and embrace receiver operating characteristic analysis. *Perspectives on Psychological Science, 7,* 275-278. doi: *10.1177/1745691612442906*

Wixted, J. T., & Mickes, L. (2014). A signal-detection-based diagnostic-feature-detection model of eyewitness identification. *Psychological Review, 121,* 262-276. doi: *10.1037/a0035940*

Wong, C. K., & Read, J. D. (2011). Positive and negative effects of physical context reinstatement on eyewitness recall and identification. *Applied Cognitive Psychology, 25,* 2-11. doi: 10.1002/acp.1605

Yu, C. J., & Geiselman, R. E. (1993). Effects of constructing identi-kit composites on photospread identification performance. *Criminal Justice and Behavior, 20,* 280-292. doi: 10.1177/0093854893020003005

Tables

Table 1

Summary of Experimental Design with Corresponding Number of Participants per Condition

		Experiment 1		Experiment 2	
		Two days delay		Five weeks delay	
Condition	*n*	TA	TP	TA	TP
No description CG	80	24	24	16	16
Description only	48	24	24	-	-
Re-reading	80	24	24	16	16
Total	208	72	72	32	32

Note. CG = control group; TA = target-absent; TP = target-present. Due to the study's primary focus on the effects of re-reading, there was no description only group in Experiment 2.

Table 2

Identification Outcomes (%; with Frequencies in Parentheses), Diagnosticity Ratios, Signal Detection Performance d' and Response Bias C for Description Conditions in Experiment 1 (Two Days Delay) and Experiment 2 (Five Weeks Delay)

Condition	*n*	Correct decisions	Hits	FAs	Estimated FAs	DR	*d'*	*C*
Experiment 1								
No description CG	48	77.1 (37)	75.0 (18)	20.8 (5)	3.5	21.6	2.49	0.57
Description only	48	79.2 (38)	70.8 (17)	12.5 (3)	2.1	34.0	2.58	0.74
Re-reading	48	91.7 (44)	87.5 (21)	4.2 (1)	0.7	125.0	3.61	0.65
Experiment 2								
No description CG	32	40.6 (13)	25.0 (4)	43.8 (7)	7.3	3.4	0.78	1.06
Re-reading	32	75.0 (24)	56.2 (9)	6.2 (1)	1.0	54.4	2.48	1.09
Experiment 1 and 2								
No description CG	80	62.5 (50)	55.0 (22)	30.0 (12)	5.0	11.0	1.77	0.76
Description only	48	79.2 (38)	70.8 (17)	12.5 (3)	2.1	34.0	2.58	0.74
Re-reading	80	85.0 (68)	75.0 (30)	5.0 (2)	0.8	90.0	3.08	0.87

Note. Hits = correct identifications in TP lineups; FAs = any positive identification in TA lineups; Estimated FAs = false identifications in TA lineups/lineup size; DR = diagnosticity ratio.

Table 3

Means and Standard Deviations of Description Quantity, Number of Correct and False Descriptors and Description Accuracy in Free Reports and in Specific Questions (N = 101)

	Free report		Specific questions	
Category	*M*	*SD*	*M*	*SD*
		Total number of details		
Face	2.96	1.90	3.26	1.11
Hair	2.13	0.73	2.40	0.83
Body	0.96	0.85	1.86	0.85
Clothes	4.63	2.79	6.53	2.41
Sum	10.68	4.18	14.05	3.57
		Number of correct details		
Face	2.18	1.65	1.85	1.07
Hair	1.88	0.75	2.07	0.83
Body	0.28	0.49	0.69	0.66
Clothes	3.18	2.66	4.18	2.48
Sum	7.51	3.76	8.79	3.42
		Number of false details		
Face	0.78	0.87	1.41	0.96
Hair	0.25	0.48	0.33	0.62
Body	0.68	0.77	1.17	0.76
Clothes	1.46	1.24	2.36	1.22
Sum	3.17	1.87	5.26	1.98
		Accuracy of details (%)		
Face	67.09	36.63	56.02	27.90
Hair	88.05	23.68	88.37	21.12
Body	19.64	36.01	35.23	34.84
Clothes	60.89	32.21	60.72	20.46
Sum	69.14	17.74	61.76	14.01

Table 4

Point-biserial Correlations Between Identification Accuracy and Description Quantity, Number of Correct and False Descriptors and Description Accuracy for Free Reports and Specific Questions

		Details			
Sample	*n*	Number	Correct	False	Accuracy
		Free report			
Total sample	101	.23*	.22*	.06	.06
Delay					
2 days	69	.13	.15	-.03	.10
5 weeks	32	.34°	.30°	.15	-.01
Condition					
Description only	33	.13	.13	.04	.01
Re-reading	68	.29*	.28*	.09	.07
		Specific questions			
Total sample	101	.27*	.24*	.09	.09
Delay					
2 days	69	.21	.19	.04	.10
5 weeks	32	.39*	.31°	.18	.03
Condition					
Description only	33	.32	.26	.13	.09
Re-reading	68	.26*	.24*	.06	.10

Note. Number = number of details; Correct = number of correct details; False = number of false details.

° p < .10; * p < .05.

Table 5

Point-biserial Correlations Between Quality Ratings of Perpetrator Descriptions in Free Reports with Identification Accuracy

Sample	*n*	Spec	Info	Prec	Comp
Total sample	101	.14	.16	.17^{o}	.16
Delay					
Two days	69	.07	.04	.08	.07
Five weeks	32	.17	.31^{o}	.24	.25
Condition					
Description	33	.00	.02	.08	.03
Re-reading	68	.25*	.28*	.24*	.27*

Note. Spec = specificity; Info = informativeness; Prec = precision; Comp = mean composite rating.

^{o} $p < .10$; * $p < .05$.

Table 6

Mean Weighted Effect Sizes for the Re-Reading Effect (Re-Reading vs. No Description Control Group) for Overall Correct Identification Decisions, Correct Identifications in Target-Present Lineups and Correct Rejections in Target-Absent Lineups

					95% CI						
DV	*k*	*N*	*LOR*	*OR*	*LL*	*UL*	*Z*	p_Z	*Q*	p_Q	I^2
Overall Correct Identification Decisions	4	352	0.764	2.15	1.34	3.44	3.179	.001	6.924	.074	56.671
Correct Identifications in TP Lineups	4	176	0.668	1.95	0.99	3.84	1.932	.053	3.937	.268	23.796
Correct Rejections in TA Lineups	4	176	1.008	2.74	1.27	5.92	2.564	.010	3.837	.280	21.805

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 = indicator of heterogeneity.

Table 7

Mean Weighted Effect Sizes for the Re-Reading Effect (Re-Reading vs. Description Only Group) for Overall Correct Identification Decisions, Correct Identifications in Target-Present Lineups and Correct Rejections in Target-Absent Lineups

					95% CI						
DV	*k*	*N*	*LOR*	*OR*	*LL*	*UL*	*Z*	p_Z	*Q*	p_Q	I^2
Overall correct identification decisions	6	919	0.144	1.15	0.88	1.52	1.031	.302	5.167	.396	3.225
Correct identifications in TP lineups	5	315	-0.262	0.77	0.48	1.24	1.072	.284	7.315	.120	45.320
Correct rejections in TA lineups	5	315	0.585	1.79	1.08	2.98	2.254	.024	0.432	.980	-825.575

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 = indicator of heterogeneity.

Figures

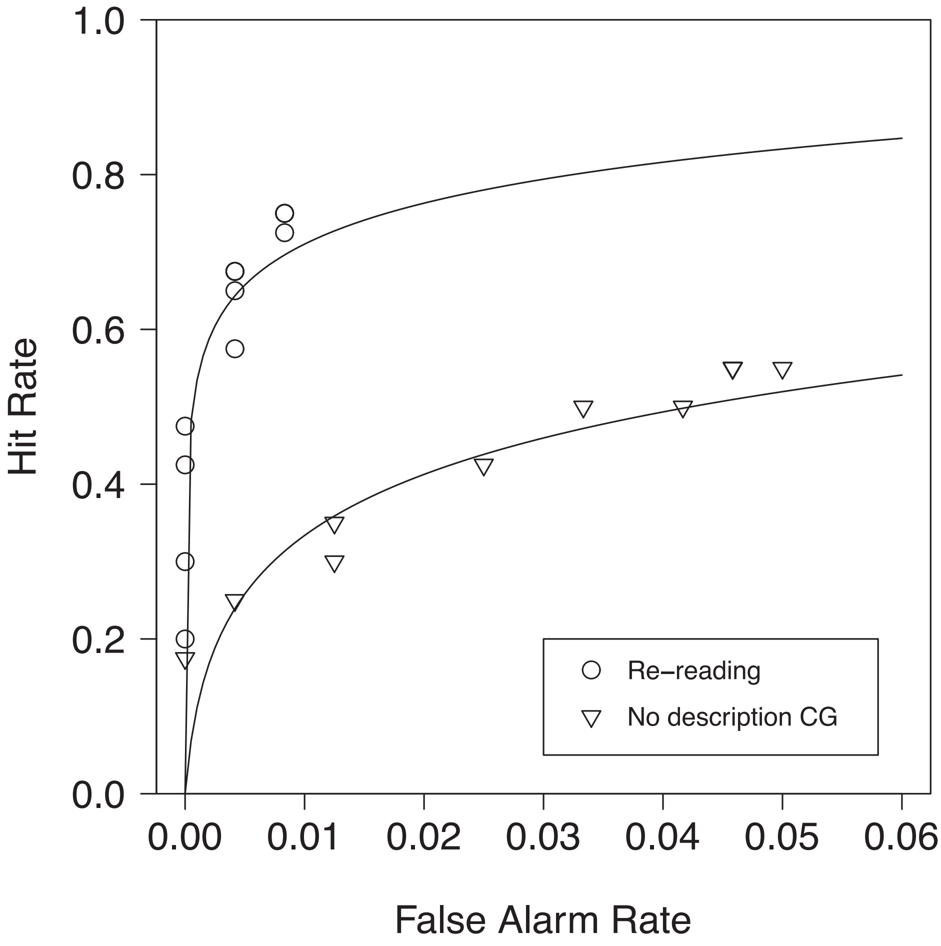

Figure 1. Confidence-based receiver operating characteristic (ROC) curves for the re-reading and no description control group. Data points reflect cumulative hit and false alarm rates (i.e., false alarms in TA lineups/lineup size) for each confidence level. The rightmost point of the ROC represents the hit and false alarm rate across all confidence levels (i.e., most liberal decision criterion). The remaining points were computed using ever lower cutoff values on the confidence scale (i.e., an increasingly conservative decision criterion).

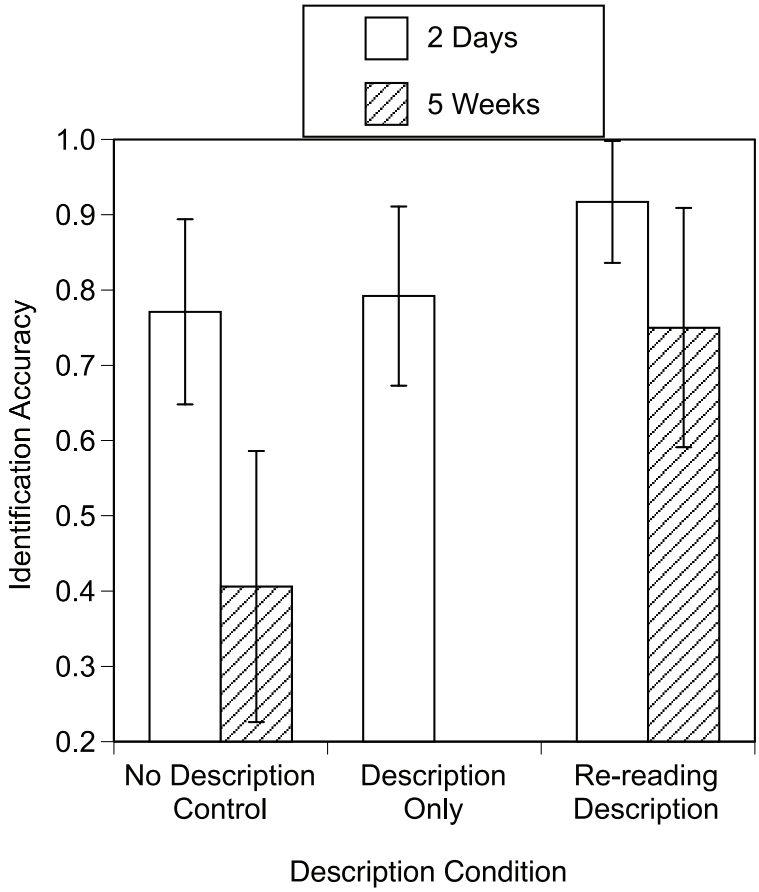

Figure 2. Means (and 95% CIs) for identification accuracy (hits and correct rejections) in the no description control, the description only, and the re-reading condition after a 2 day vs. 5 weeks delay. Note that there was no description only group after five weeks.

Appendix

Appendix A

Specific Questions About the Perpetrator's Physical Appearance

Now we will ask you some specific questions regarding the outer appearance of the perpetrator. Please answer these questions on answer sheet B2. If you have already answered one of the questions in your previous report, please answer it again. If you don't know the answer to one of the questions, write "don't know" in the corresponding line.

1. What age do you estimate the perpetrator was?
2. What size do you estimate the perpetrator was?
3. Please describe the figure of the perpetrator in detail!
4. Please describe the clothes of the perpetrator in detail!
5. Please describe the color of the perpetrator's hair in detail!
6. Please describe the perpetrator's hairdo in detail!
7. Please describe the shape of the perpetrator's face!
8. Did you notice any special features on the perpetrator? If so, which?
9. Did the perpetrator wear a headdress? If so, which?
10. Did the perpetrator wear glasses? If so, what did they look like?
11. Did the perpetrator have a beard? If so, what did it look like?
12. Did the perpetrator speak in a certain dialect or did he have an accent? If so, which?

EXPERIMENT 2

Discriminating Between Correct and Incorrect Eyewitness Identifications: The Use of Appropriate Cues

Police investigations often involve asking eyewitnesses to make an identification decision. If a positive identification is made and it comes to a trial, judges and/or jurors have to evaluate the accuracy of these identification decisions to arrive at their verdicts. Indeed, eyewitness testimonies including identification decisions are often the only, or the major, source of evidence available and consequently serve as an essential basis for later convictions.

Many studies focused on estimator and system variables influencing identification decisions (Wells, 1978; Wells & Olson, 2003). However, not a misidentification per se leads to judicial errors, but rather the wrong evaluation of these identification decisions (in light of other evidence). At this judgment level (cf. Sporer, 2007, 2008) fact finders do not only evaluate witnesses' reports, but also try to consider witnessing conditions at the perceptual, the retention and the recall/recognition stage. They also take witnesses' meta-memory statements (e.g., confidence, self-reports about their decision processes) as well as behavioral aspects (e.g., response latency to make an identification decision) into account (Garrett, 2011; Sporer, 2007).

Moreover, fact finders' beliefs and knowledge about indicators of correct eyewitness testimonies affect the weight given to specific factors in the evaluation process and thus influence fact finders' judgments in a given case (Leippe, 1994; Semmler, Brewer, & Bradfield Douglass, 2012).

However, observers seem to have only limited knowledge about factors affecting eyewitness memory (e.g., Benton, Ross, Bradshaw, Thomas, & Bradshaw, 2006; Desmarais & Read, 2011). This holds true even for judges, attorneys and law enforcement personnel, although they are more knowledgeable than jurors (Benton et al., 2006; Brigham & Wolfskeil, 1983; Wise & Safer, 2004).

But are people able to discriminate between correct and incorrect eyewitnesses nonetheless? Using mock-juror paradigms that asked observers to evaluate the accuracy of presented identification decisions, observers showed only poor judgment accuracies comparable with chance level (e.g., Beaudry, Lindsay, Leach, Mansour, Bertrand, & Kalmet, 2015; Brigham & Bothwell, 1983; Wells & Lindsay, 1983). As a possible explanation, observers often use commonsense evaluations and rely on cues erroneously believed to be postdictive of accurate identification decisions (Boyce, Beaudry, & Lindsay, 2007; Semmler et al., 2012). Thus, they either use invalid cues and/or weight these cues inappropriately to make their judgments (cf. Lindsay, 1994; Semmler et al., 2012).

To account for fact finders' decisions, Leippe (1994) proposed the *validity-intuition model* that locates accuracy cues in a memory message somewhere along two dimensions. The validity dimension refers to the degree to which cues are diagnostic of identification accuracy, whereas the intuition dimension describes the degree to which people have preconceived beliefs that the cues are associated with identification accuracy. Consequently, judgment accuracy should increase, if highly valid cues are appropriately perceived and intuitively used by observers and if invalid cues are neglected appropriately. In contrast, judgment accuracy should decrease if observers intuitively rely on cues that are not (or only weakly) related to identification accuracy or if they fail to use cues that are actually good indicators of identification accuracy.

Hence, to improve judgment accuracy, it is necessary to get deeper insight into observers' evaluation and decision processes. Thus, we focus on the discriminant validity of cues that observers use to make their judgments and on observers’ interpretation of these cues when judging identification accuracy.

Valid Indicators of Identification Accuracy vs. Intuitively Used Cues

There are several suggestions by the courts in the United States (cf. *Biggers criteria*, Neil vs. Biggers, 1972), in the United Kingdom (cf. Police and Criminal Evidence Act, 1984), and in Germany (Sporer & Cutler, 2003) about which criteria should be used as reliable indicators of an accurate identification decision (for a review see Semmler et al., 2012). However, the validity of these criteria is not always decisively given (Garrett, 2011, 2012; National Academy of Sciences, 2014; Wells, Small, Penrod, Malpass, Fulero, & Brimacombe, 1998). Thus, both legal scholars and psychologists tried to identify empirically valid factors to be used to distinguish between witnesses who made an accurate and those who made an inaccurate identification decision, especially concentrating on characteristics of the testimony and the identification procedure itself.

Confidence and response latency.

Positive identifications made by highly confident eyewitnesses (who reported their confidence directly after the identification decision) are more likely to be correct than identifications made by less confident witnesses

(Brewer & Wells, 2006; Sporer, Penrod, Read, & Cutler, 1995; Wixted, Mickes, Clark, Gronlund, & Roediger, 2015). Moreover, correct positive identifications are associated with shorter response latencies than false identifications (e.g., Dunning & Perretta, 2002; Sporer, 1992, 1993, 1994).

While fact finders seem to be sensitive to the existing confidence-accuracy relationship (at least for choosers) but tend to overestimate it when judging identification accuracy (e.g., Cutler, Penrod, & Stuve, 1988; Wells, Lindsay. & Ferguson, 1979; for a review, see Boyce et al., 2007) results regarding observers' usage of decision time as a marker of identification accuracy are inconsistent. On the one hand observers intuitively associate response latency with identification accuracy (Benton et al., 2006) and hence judge witnesses who make fast identifications as more credible than those deciding slower (Neal, Christiansen, Bornstein, & Robicheaux, 2012). On the other hand, the exoneration cases analyzed by Garrett (2011) demonstrate that fact finders often ignore this relationship. Garrett (2011) reported that many eyewitnesses who later were judged as credible in court had been uncertain when they first saw the lineup and took up to several minutes to make an identification decision.

Decision processes.

Indicators of identification accuracy also arise from self-reports of decision strategies applied by an eyewitness during the identification procedure. Wells (1984) distinguished between *absolute* and *relative* decision processes, with correct lineup choices being more often accompanied by absolute than by relative decision processes than incorrect choices (e.g., Lindsay & Bellinger, 1999; Smith, Lindsay, & Pryke, 2000). According to Wells (1984), relative decisions go along with a comparison of the different lineup members, resulting in the choice of the person that is the closest relative match to the eyewitness' memory of the culprit. In contrast, eyewitnesses making an absolute decision compare each lineup member with their memory for the perpetrator (absolute match).

Based on the reality/source monitoring framework (Johnson, Hashtroudi, Lindsay, 1993; Johnson & Raye, 1981), Dunning and Stern (1994) suggested that eyewitnesses differ especially in their employed *cognitive effort* when making a positive identification. Accordingly, correct identifications are made *more quickly* and *automatic* without explicit cognitive effort (Behrman & Richards, 2005; Dunning & Stern, 1994). In contrast, poor memories may force people to make more logical inferences and to actively reconstruct memory (cf. Schooler, Gerhard, & Loftus, 1986; Wells & Lindsay, 1983). Thus, eyewitnesses making a false identification use more *deliberative* and *reflective* explicit cognitive strategies to arrive at their decision (Dunning & Stern, 1994).

The reality monitoring framework has also been fruitfully employed to judge *other people's* memory termed interpersonal source monitoring (Mitchell & Johnson, 2000; Sporer, 2004). If people use certain qualities to evaluate the source of their own memory, these cues might also be useful for the evaluation of someone else's memory report (Leippe, 1994; Sporer, 1997, 2004, 2008). For example, Leippe (1994) expected that people might be sensitive to an eyewitness’s "noisy thinking out loud" (p. 395) reflecting more cognitive operations in the witness’s report and in turn indicating a poor memory.

In line with this, Dunning and Stern (1994) reported that mock-jurors’ associated an automatic recognition process with higher identification accuracy and also used it to make their judgments. However, participants failed to use elimination processes as a marker of an inaccurate identification. Moreover, witnesses who reported a negligible influence of nonchosen photos and a higher impact of their own memory (compared to the other photos) were judged as more likely to be accurate.

Encoding conditions.

There are recommendations of the US Supreme Court to use the witness's *quality of view* as well as the *attention a witness paid to the culprit* to evaluate an identification decision (Neil vs. Biggers, 1972). Indeed, studies manipulating witness attention and viewing conditions (i.e., exposure time, distance) provide some empirical evidence supporting the assumptions of these two Biggers criteria (e.g., Lindsay, Semmler, Weber, Brewer, & Lindsay, 2008; Palmer, Brewer, & Weber, 2010). Similarly, observers seem to use information about encoding conditions to make their judgments to some extent (e.g., Lindsay, 1994).

Amount of details and report consistency.

Although research on their validity is inconsistent, fact finders sometimes intuitively rely on content cues, like the amount of details described by an eyewitness (Bell & Loftus, 1989; Wells & Leippe, 1981), as well as on intra-subjective report consistency or contradictions included in the testimony (Berman & Cutler, 1996; Berman, Narby, & Cutler, 1995; Semmler & Brewer, 2002). Thus, in the present study we investigate if these cues are indeed valid to discriminate between correct and incorrect identifications and if observers use them.

The Present Study

Numerous studies have investigated whether there are valid cues to discriminate between correct and incorrect identifications using witnesses' self-reports or objectively measurable decision times. However, only few studies examined if observers are also able to infer these cues based on original identification protocols and how they interpret them to evaluate the accuracy of an identification decision. In criminal cases witnesses are commonly asked to give reasons for their identification decision, to state their confidence, and to estimate duration and the degree of attention they have paid to the perpetrator. The witnesses' answers are then used to postdict their identification accuracy. However, witnesses are often not able to describe reasons for their decisions adequately (Nisbett & Wilson, 1977). Moreover, these questions are often asked at trial and not directly after the identification decision and thus might be biased (e.g., due to feedback effects: Wells & Bradfield, 1998). To prevent fact finders' from relying on biased eyewitness statements when evaluating the accuracy of their identification decisions, it would be of high practical relevance if there are valid indicators of identification accuracy in a witness's original identification protocol that are perceivable for observers *without asking* the witness to evaluate his/her decision before.

Consequently, in our study, *observers' perceptions* (instead of witnesses' self-report measures) of witnesses' confidence, response latency, decision processes, perceptual basis and quality of reasons given for the identification decision are analyzed and tested for their discriminative value. To explain observers' judgment accuracy (and to extrapolate recommendations how to increase it) the discriminative value of observer ratings is contrasted with *their interpretation and use* of these cues.

Application of the Brunswikian lens model.

The Brunswikian lens model (Brunswik, 1956, 1965) provides a theoretical framework to describe the relationship between the validity of certain perceived cues to predict an objective outcome on the one hand and perceivers' usage of these cues to predict their judgments of this outcome on the other hand. The model has been applied to medical decision-making (e.g., Hammond, Hursch, & Todd, 1964), social perception (e.g., Vicaria, Bernieri, & Isaacowitz, 2015) and to many fields of psychology, for example to the detection of deception (e.g., Fiedler, 1989; Hartwig & Bond, 2011; Sporer, 1997; Sporer & Küpper, 1995; Sporer, Masip, & Cramer, 2014).

The model can be divided into two parts. First, relationships between the presence of certain cues (e.g., a patient's symptoms) and an objective outcome (e.g., an appendicitis) are investigated. These correlations describe if the observed cues are valid indicators of the outcome (*ecological validities*).

Second, relationships between the observed cues and perceivers' judgments (e.g., a doctor's decision to operate and extract the appendix based on the observed symptoms) are analyzed (*subjective utilities*). High correlations indicate a strong usage of these cues.

To ensure high judgment accuracy (e.g., a correct decision to operate based on the observed symptoms, because there was an appendicitis) the perceived cues should be strongly related to the objective outcome of interest and perceivers should use and interpret these cues appropriately to arrive at their judgments. Thus, a *correspondence* measure is computed, to estimate the agreement between cue validities and perceivers usage of these cues.

A perfect agreement between ecologically valid and subjectively used cues is demonstrated when the correlations between the cues and the objective outcome as well as between the cues and perceivers' judgments are in the same direction and equal in size. In this case observers are able to distinguish between valid and invalid cues. First, they are sensitive to discriminating cues and thus use and weight them appropriately. Second, they realize that other cues do not have discriminative value and thus do not weight them strongly. To the extent, that the two sets of correlations differ, correspondence between ecological validities and subjective utilities decreases.

We applied the Brunswikian lens model to judgments of identification decisions (to our knowledge) for the first time to (a) examine the validity of several cues perceived by observers based on identification protocols to discriminate between correct and incorrect identification decisions (*ecological validities*), and (b) to investigate if these cues are indeed used by observers to make their judgments (*subjective utilities*) and how much weight observers give to each of these cues. To the extent that there is a correspondence between ecological validities and subjective utilities observers' judgment accuracy of identification decisions should be improved.

It can be assumed that jurors are not aware of factors that influence their decisions and thus, asking them which cues they have used and how they have weighted them is not expected to result in appropriate statements (cf. Nisbett & Wilson, 1977). Consequently, the application of the

Brunswikian lens model seems to be an adequate solution to contrast ecological validities and subjective utilities.

Retrospective reasoning vs. thinking-aloud protocols.

To make the witnesses' decision processes and thoughts more apparent for observers, in the present study half of the witnesses were instructed to verbalize their thoughts ("think-aloud") while arriving at their identification decision (cf. Ericsson & Simon, 1993). In contrast, the other half of participants had to list reasons for their decision afterwards.

In the think-aloud condition participants had to verbalize every thought or idea that came to their mind while trying to identify the target person.[1] By this procedure, we hoped to gain deeper insight into the reported cognitive processes compared to witnesses who were instructed to just give reasons for their identification decision afterwards. Nisbett and Wilson (1977) claimed that people only have limited access to cognitive processes that prompt their actions and decisions and thus are often unable to explain the causes of their behavior appropriately. Thus, when asked to describe reasons for their behavior afterwards, participants were found to only retrospectively justify their behavior.

Consequently, we hypothesized that think-aloud protocols will include more (salient) cues about the accuracy of the identification decision than retrospective reasoning protocols. This is expected to result in an increased observer ability to perceive valid indicators of identification accuracy and

---

[1] It is noteworthy, that thinking aloud instructions do not impair task performance, as reported in a meta-analysis by Fox, Ericsson and Best (2011).

hence an increased ability to discriminate between correct and incorrect identifications when think-aloud protocols are used compared to retrospective reasoning reports.

Written transcripts vs. videotapes.

To evaluate the reliability of an eyewitness's testimony Garrett (2011) suggested to rely on the original (videotaped) identification decision instead of considering testimonies about the identification made at court. Several researchers have recommended the videotaping of eyewitnesses during the identification procedure since many years (e.g., Kassin, 1998; Sporer, 1992, 1993; Wells et al., 1998). Videotaping results in a conservation of behavioral cues and information about the original identification procedure to be used by jurors as potentially diagnostic information to determine identification accuracy in court. Based on dual-process theories (e.g., the Elaboration Likelihood Model: Petty & Cacioppo, 1986) as well as on Leippe's (1994) model of eyewitness persuasion, not only the message content but also peripheral and heuristic cues like the message delivery style and witness attributes affect fact finders' judgments. Peripheral cues are assumed to be more salient in videotapes, showing, for example, the witness's appearance, the powerfulness of speech (cf. Erickson, Lind, Johnson, & O'Barr, 1978), response latencies as well as nonverbal and paraverbal behaviors. In turn, these behaviors allow observers to make inferences about the witness's confidence and credibility (Erickson et al., 1978; Schooler et al., 1986).

Former judgment studies often used videotaped mock trial procedures as stimulus materials in which only the type of presented material was varied (e.g., the presentation of the witness examination at trial vs. presentation of the witness examination and the original identification) to assess its effects on observer judgments (e.g., Beaudry et al., 2015; Kassin, Rigby, & Castillo, 1991; Reardon & Fisher, 2011). For example, Reardon and Fisher (2011) compared observers' ability to discriminate between accurate and inaccurate witnesses showing them videotapes of witness examinations at trial with or without an additional videotape of the original description of the perpetrator and the identification decision. In contrast to the examination-only condition, observers' judgment accuracy increased when the videotaped identification procedure was presented. This supports the assumption that observers benefit from videos of the original identification decision that seem to contain useful additional information to evaluate identification accuracy (for similar conclusions see also Bradfield Douglass & Jones, 2013).

To our knowledge there are no studies that investigated different presentation media to test its effects on mock jurors' discrimination between correct and incorrect identifications. Thus, we conducted two studies, in which observers were presented with the original identification decisions only (i.e., without further eyewitness testimonies at trial) either presented as written transcripts (Study 1) or as videotapes (Study 2). We hypothesized that presenting observers with videotaped identification decisions should result in an increased discriminability of the perceived cues as well as in higher judgment accuracies, both due to an increased cue salience.

Method

First, we describe the study conducted to construct eyewitness statements (Phase 1), followed by a detailed description of the two judgment studies (Phase 2) using written transcripts (Study 1) and videotapes (Study 2).

Phase 1: Witnessing and Identification

Design and participants.

A 2 (stimulus film: short vs. long) x 2 (type of decision protocol: retrospective reasoning vs. think-aloud) x 2 (target presence in the lineup: target-absent [TA] vs. target-present [TP]) factorial between-participants design was used with 192 witness-participants (48 male, 144 female, aged 19 to 61, *Mdn* = 22.0) who were randomly assigned to the experimental conditions. They participated voluntarily in the study and were mainly Caucasian students of a German university.

Stimulus film and description task.

Witness-participants saw a short video film showing a theft of a bicycle in a pedestrian area. There were two versions of the film lasting 1:44 minutes and 2:24 minutes with the perpetrator, a young woman, seen in close-up for 3 vs. 5 seconds. After having completed an unrelated filler task (for approximately 40 minutes), participants were instructed to imagine having to give a statement about the incident seen in the film to the police. First, they were asked to give a free report of the perpetrator's physical appearance followed by several non-leading questions. Participants were asked to describe the perpetrator as precisely as possible but had the opportunity to give "don't know" answers, if they were not able to remember any of the specific features asked for. Afterwards, participants described the observed incident and the crime scene in as much detail as possible, again followed by several non-leading questions.

Training to think-aloud vs. retrospective reasoning.

Half of the witness-participants received a short training to get familiar with the "think-aloud" procedure. First, a short film showing a man thinking-aloud while playing "Four in a Row" was presented. Then, participants were instructed to "think-aloud" concurrently while solving a cognitive puzzle (Tower of Hanoi). They were asked to say out loud everything (i.e., every idea, thought or recollection) that comes to their mind while they were solving the task, even if it seemed to be irrelevant (cf. Ericsson & Simon, 1993). The other half of participants (in the retrospective reasoning condition) also had to solve the puzzle, but had to give reasons for their solution process afterwards instead.

Photo lineups and identification procedure.

Lineups consisted of six frontal portrait photographs, presented simultaneously in two rows of three pictures each. In a pilot study, foils were selected based on the "match-to-description-of-culprit" strategy (Luus & Wells, 1991). For TA lineups, the photo that matched the description most was taken as replacement for the target. Lineup fairness (Tredoux, 1998) was satisfactory, with Tredoux's E = 5.25 for TA lineups (N = 50) and E = 5.79 for TP lineups (N = 50).

Half of the participants were either instructed to think-aloud while making their identification decision or to give reasons for their decision afterwards. Participants had to state their identification decision orally. Afterwards, they were asked to rate their confidence regarding the accuracy of their identification decision (0 to 100%) as well as their willingness to testify in court (0 to 100%).

Testing modalities.

All experimental instructions were given via the speakers of the computer the witness-participants were seated in front of. They were instructed to give their answers orally while focusing on a video camera, which was placed diagonally in front of them. Witness-participants were videotaped during the description and the identification task. Afterwards, all statements were literally transcribed.

Selection of the stimulus identification statements.

In sum, 126 (65.6%) participants made a positive identification. Of these choosers' statements, 48 were randomly selected to be judged in Phase 2, counterbalancing identification accuracy (incorrect choices in TA lineups vs. correct choices in TP lineups) and type of decision protocol (retrospective reasoning vs. think-aloud).[2] Thus, there were 12 identification statements (of 6 male and 6 female witnesses) per condition to be judged by observer-participants in the subsequent study. Hence, our experiment was 12 times internally replicated, satisfying stimulus sampling requirements (Wells & Windschitl, 1999).

---

2 Preliminary analyses showed no differences in choosers' identification accuracy for the short (36.7%) and the long (28.8%) stimulus film, $chi^2(1, N = 126) = .889$, $p = .346$, $OR = 0.70$, 95% CI [0.33, 1.48]. Consequently, length of the stimulus film was not considered in the statement selection. The final selection resulted in $n = 26$ witnesses who had watched the short and $n = 22$ witnesses who had watched the long film version.

Phase 2: Judgment of identification statements

Overview and Design.

Two parallel experiments with *N* = 288 observer-participants of an Internet survey (Study 1) evaluating written transcripts and *N* = 96 observer-participants evaluating videos of the identification decisions in the laboratory (Study 2) were conducted. In both studies each observer-participant had to judge only one of the 48 stimulus identification statements created in Phase 1.

For both studies a 2 (objective identification accuracy: incorrect choice [TA] vs. correct choice [TP]) x 2 (type of decision protocol: retrospective reasoning vs. think-aloud) x 2 (person and event descriptions: not presented vs. presented) factorial between-participants design was used. Dependent variables were a number of ratings described below.

Participants.

In Study 1, a total of *N* = 288 observer-participants (96 male, 192 female, age 18 to 70, *Mdn* = 25.0, 67% students) were recruited via circular e-mails and postings in social networks (e.g., Facebook). They received an Internet link guiding them to the experiment and completed it on their personal computers.

In Study 2, a total of *N* = 96 observer-participants (32 male, 64 female, age 18 to 34, *Mdn* = 23.0, 96% students) were individually tested in the laboratory. Participation in both studies was completely voluntary and was rewarded with the chance to win one of several vouchers or to gain course credit.

Materials.

Eyewitness statements.

The same 48 eyewitness statements were used in both studies and differed only due to presentation medium. Observer-participants were either presented with a written transcript (Study 1) or with a video film (Study 2) of a witness's given person and event descriptions as well as the to be judged identification decision.

The description statements included the entire instructions to give a free report of the target's physical appearance, the incident and the crime scene as well as the witness's given descriptions and answers to the specific questions. The identification statement contained the lineup instructions and the photo lineup used in Phase 1, as well as the recorded identification decision, with either thinking aloud or retrospective reasoning protocols for the lineup choice given by the eyewitness. The length of the identification statements varied between 22 and 453 words (*Mdn* = 134 words) and lasted from 11 to 270 seconds (*Mdn* = 73 sec) with retrospective reasoning protocols (*M* = 85.0 words) being much shorter than think-aloud protocols (*M* = 190.2 words). To avoid undue reliance of observer-judges on confidence the witnesses' explicit numerical confidence statements and their willingness to testify in court, which was asked for after the identification task, were *excluded* from statements/videos.

Every statement was accompanied with details of the witness's age, gender and occupation. The mean age of the selected witnesses was 24.65 years (*SD* = 7.08) and all of them were students. In the video films used in Study 2 the upper part of the witness's body was visible, while he/she was sitting in front of a computer. Most of the time the witness's face was seen in three-quarter profile. In contrast, participants of Study 1 did not see a photo of the witness's physical appearance.

Photo lineups.

Observer-participants were presented with the corresponding photo-lineup, the to be judged witness had seen before (see Phase 1). In Study 1, the lineups were included in the presented transcript, shown in a separate browser window. In Study 2, the lineups were printed on high quality photo paper (A4 format) and handed to the observer-participants. In both studies the lineups remained visible for the participants during the whole experiment.

Procedure.

Both studies followed exactly the same sequence and used the same instructions. All observer-participants were instructed to imagine being a judge or juror in court who had to evaluate an eyewitness identification decision. Additionally, participants were briefly informed about the incident (i.e., the witnessed theft) and the identification that took place approximately one hour after the film. Next, the participants were presented with the transcript (Study 1) or video (Study 2) of the eyewitness statement and were instructed to read it, or watch the video, carefully. All participants had the opportunity to read or watch the whole statement again during the following judgment process.

Observer ratings.

Participants were asked to evaluate the identification statements regarding different qualities described below. In doing so they had to infer these qualities based on their overall impression of the witness gained from the witness testimony and the used lineup.

Based on the measures of witnessing experience used in Wells and Bradfield's (1998) study, first, all observer-participants completed five ratings concerning their subjective impression of the witnessing conditions, followed by another five ratings reflecting the quality of the witness's identification decision (Table 1).

Afterwards, the spontaneity of the witness's identification decision (*1 = spontaneous* vs. *7 = hesitant*; *1 = without any deliberations* vs. *7 = with lots of deliberations*) was evaluated, followed by evaluations of the eyewitness's reasons for his/her identification decision (*1 = not at all precise/detailed/consistent/convincing* vs. *7 = very precise/detailed/consistent/convincing*).

Then, participants were asked to give a dichotomous judgment whether they believed that the witness had made a correct identification (*0 = incorrect choice* vs. *1 = correct choice*). This measure reflected the main dependent variable in the present study. Afterwards, observer-participants indicated their confidence in their judgment (11-point scale ranging from *0 to 100%*).

Finally, all observer-participants rated 15 variables concerning the witness's decision processes (Table 2). Most of these items are based on measures used by Dunning and Stern (1994) and Sauerland and Sporer (2007), reflecting automatic vs. deliberative and absolute vs. relative identification decisions.

Results

Overview

First we present the data of the Internet survey using written transcripts (Study 1: N = 288) followed by the results of the laboratory study using videotapes (Study 2: N = 96) with observer judgments as unit of analysis.

For each study we first present logistic regressions predicting observers' *judgments* of identification accuracy as well as observers' *judgment accuracies* as a function of the manipulated independent variables. Subsequently, we focus on the *observer ratings*. In separate analyses of variance we compared the observer ratings for objectively and for perceived correct and incorrect identifications. We were particularly interested whether think-aloud protocols would increase differences in observer ratings to increase their discriminability between correct and incorrect identifications and between observer judgments, respectively.

Finally, to test the Brunswikian lens model, relationships between observer ratings and objective identification accuracy (ecological validities) as well as between observer ratings and perceived identification accuracy (subjective utilities) were examined. Two multiple regression analyses were computed to compare the predictions of the objective and the perceived identification accuracy. To test the agreement between ecological validities and subjective utilities the intra-class correlation coefficient (*ICC*: two-way mixed effect model, average measure, estimated absolute agreement; cf. McGraw & Wong, 1996; Orwin & Vevea, 2009; Shrout & Fleiss, 1979) between the two sets of correlations was computed (using Fisher's Z_r transformations).

The corrected standardized mean difference Hedges g_u and point-biserial correlations are reported as effect sizes (cf. Borenstein, 2009; Lipsey & Wilson, 2001). According to Cohen (1988) a small effect is represented by a point-biserial correlation of .10 that equals a standardized mean difference of .20, whereas moderate and large effects are reflected by r_{pb}s of .24 and .37 equaling *d*s of .50 and .80, respectively. For dichotomous variables we report odds ratios (*OR*) as effect size (cf. Lipsey & Wilson, 2001).

Study 1: Written Transcripts

Observer judgments and judgment accuracy.

Separate hierarchical (sequential) logistic regression analyses were conducted to predict (1) observer judgments (identification perceived as incorrect vs. perceived as correct) and (2) judgment accuracy. Predictor variables were objective identification accuracy (incorrect ID vs. correct ID), type of decision protocol (retrospective reasoning vs. think-aloud) and description presence (with vs. without descriptions), and their two-way and three-way interactions. Proportions for the different experimental conditions are displayed in Table 3.

With observer judgments (identifications perceived as incorrect vs. perceived as correct) to be classified we started by considering a fully specified logistic regression model including the three-way interaction between the three predictors as its most complex term. In a backward hierarchical sequence we removed nonsignificant higher-order interactions. We ended up with a model containing the three main effects, $chi^2(3, 288) = 7.57$, $p = .056$, *Nagelkerke* $R^2 = .04$, with none of them being a significant predictor of observer judgments, all *Wald* statistics < 3.64, all *p*s > .056.

Parallel analyses were conducted with judgment accuracy (incorrect vs. correct judgment) to be predicted. After removing nonsignificant higher-order interactions we ended up with a model containing the three main effects $chi^2(3, 288) = 78.23$, $p < .001$, *Nagelkerke* $R^2 = .32$. Overall predictive accuracy of the model was 75.0%. In this model objective identification accuracy was a significant predictor of judgment accuracy with incorrect identifications (79.9%) being judged correctly more often than correct identifications (29.9%), $B = -2.25$, $SE_B = 0.28$, *Wald* statistic = 65.28, $p < .001$, $OR = 0.11$. Effects of type of decision protocol and description presence were not significant, both *Wald* statistics < 1.87, both *p*s > .172.

Factor structure of observer ratings.

Perceived witnessing conditions and perceived identification decision.

A factor analysis was conducted on the 16 items concerning the observer ratings of the perceived witnessing conditions, the perceived identification decision and the perceived quality of reasons. Four factors were extracted (using the maximum likelihood method), explaining 64.8% of the variance. To permit inter-correlations between factors oblique rotation (direct oblimin) was used (cf. Fabrigar, Wegener, MacCallum, & Strahan, 1999).[3] Factor loadings after rotation are displayed in Appendix A. Items loading on the respective factors were averaged, yielding four scales with high internal consistencies and satisfactory corrected item-total correlations

---

[3] We followed the advice of an anonymous reviewer who recommended this method instead of principal components analysis.

(all *CITC*s > .30): Perceived perceptual basis (*alpha* = .82), perceived confidence (*alpha* = .86), perceived decision time and difficulty (*alpha* = .71), perceived persuasiveness of reasons (*alpha* = .79).

Perceived decision processes.

Another factor analysis was conducted on 14 items concerning the observer ratings of the perceived decision processes in Table 2 (Item 10 was removed from the analysis due to low inter-correlations). Four factors were extracted (using the maximum likelihood method), explaining 61.8% of the variance. Factor loadings after oblique rotation (direct oblimin) are displayed in Appendix B (cf. Fabrigar et al., 1999). Based on our theoretical assumptions items loading on the same factors were averaged, yielding four scales with satisfactory internal consistencies and corrected item-total correlations (all *CITC*s > .32): Perceived automatic decision (*alpha* = .70), perceived deliberative and effortful decision (*alpha* = .81), perceived absolute decision and a good memory for the perpetrator (*alpha* = .61) and perceived eliminative and relative decision (*alpha* = .68).

Observer ratings as a function of objective identification accuracy (ecological validities).

For every scale separate 2 x 2 analyses of variance (ANOVAs) were conducted with objective identification accuracy (incorrect ID vs. correct ID) and type of decision protocol (retrospective reasoning vs. think-aloud) as classification variables. Means and standard deviations of each scale as well as the effect sizes g_u for objectively correct and incorrect identification statements in the retrospective reasoning and the think-aloud condition are displayed in Table 4.[4]

For five of the eight observer ratings we found a significant main effect of objective identification accuracy, all Fs(1, 284) > 3.98, all ps < .047 (see Table 4). For ratings of perceived deliberative and effortful decision processes the effect was qualified by a significant interaction between objective identification accuracy and type of decision protocol, $F(1, 284) = 6.28$, $p = .013$.[5]

---

[4] In preliminary analyses of Studies 1 and 2 we had included description presence as a third independent variable. However, for all eight rating scales neither the interaction between description presence and objective identification accuracy, description presence and perceived identification accuracy, nor the three-way interactions reached significance. Consequently, the factor was excluded from further analyses.

[5] There were significant main effects of type of decision protocol for four scales. Ratings of perceived decision time and difficulty, $g_u = 0.28$, 95% CI [0.05, 0.51], perceived deliberative and effortful decision processes, $g_u = 0.36$, 95% CI [0.13, 0.59], and perceived eliminative and relative decision processes, $g_u = 0.58$, 95% CI [0.34, 0.81], were higher in the think-aloud than in the retrospective reasoning condition, all Fs(1, 284) > 5.73, all ps < .017. In contrast, ratings of perceived automatic decision processes were higher when retrospective reasoning protocols were used compared to think-aloud protocols, $F(1, 284) = 9.23$, $p = .003$, $g_u = -0.35$, 95% CI [-0.58, -0.12].

Observer ratings as a function of perceived identification accuracy (subjective utilities).

To focus on the perceived cues observers used to make their judgments, we conducted separate 2 x 2 ANOVAs for each scale with perceived identification accuracy (perceived incorrect ID vs. perceived correct ID) and type of decision protocol (retrospective reasoning vs. think-aloud) as classification variables.

Seven rating scales differed due to observer judgments of perceived identification accuracy, all Fs(1, 284) > 9.51, all ps < .002. For none of the eight scales were there any interactions between observer judgments and type of decision protocol, all Fs(1, 284) < 2.02, all ps > .157. Means and standard deviations of each scale as well as the effect sizes are displayed in Table 5.

Brunswikian lens model analyses: Do observers use appropriate cues?

To test the Brunswikian lens model two multiple regression analyses including objective and perceived identification accuracy as dependent variables and observer ratings as predictors were conducted. The analyses were conducted separately for participants in the think-aloud and in the retrospective reasoning condition (n = 144 judgments each).

In the think-aloud condition, ratings explained 27% of the variance of perceived identification accuracy (p < .001), while 12% of the variance of objective identification accuracy were explained (p = .015). For every rating scale zero-order correlations with the two dependent variables were computed. Seven of the eight zero-order correlations demonstrated significant relationships between the ratings and the perceived identification accuracy. In contrast, only three ratings were related with objective identification accuracy (see Figure 1). The two sets of correlation coefficients (Fisher's Z_r transformations were used) were related by *ICC* = .68 demonstrating high correspondence between ecological validities and subjective utilities, but also some minor discrepancies.

In the retrospective reasoning condition ratings explained 25% of the variance of perceived identification accuracy ($p < .001$) and 12% of the variance of objective identification accuracy ($p = .023$). Seven scales were significantly correlated with observer judgments of identification accuracy. In contrast, only one of the eight zero-order correlations demonstrated a significant relationship with objective identification accuracy (see Figure 2). The two sets of correlation coefficients were weakly related by *ICC* = .22 reflecting essential non-correspondence in the validity of the perceived cues (ecological validities) on the one hand and the use of these cues on the other hand (subjective utilities).

Study 2: Videos

Observer judgments and judgment accuracy.

Parallel to Study 1, hierarchical logistic regressions were conducted to predict (1) observer judgments (identification perceived as incorrect vs. perceived as correct) and (2) judgment accuracy as a function of objective identification accuracy (incorrect ID vs. correct ID), type of decision protocol (retrospective reasoning vs. think-aloud) and description presence (with vs. without descriptions). Proportions of observer judgments and judgment accuracy for the different experimental conditions are displayed in Table 3.

With observer judgments again we started by considering a fully specified logistic regression model including the three-way interaction between the three predictors as most complex term. In a backward hierarchical sequence we removed nonsignificant higher-order interactions. We ended up with a model containing the three main effects, $chi^2(3, 96) = 12.32$, $p = .006$, *Nagelkerke* $R^2 = .16$. Overall predictive accuracy of the model was 69.8%. In this model only objective identification accuracy was a significant predictor of observer judgments with correct identifications (52.1%) being judged as correct more often than incorrect identifications (25.0%), $B = 1.25$, *Wald* statistic = 7.47, $SE_B = 0.47$, $p = .006$, $OR = 3.26$. Effects of type of decision protocol and description presence were not significant, both *Wald* statistics < 2.36, both *p*s > .124.

Parallel analyses were conducted with judgment accuracy (incorrect vs. correct judgment) to be classified. After removing nonsignificant higher-order interactions we ended up with a model containing the three main effects $chi^2(3, 96) = 7.91$, $p = .048$, *Nagelkerke* $R^2 = .11$. Overall predictive accuracy of the model was 69.8%. In this model objective identification accuracy was a significant predictor of judgment accuracy with incorrect identifications (75.0%) being judged correctly more often than correct identifications (52.1%), $B = -1.04$, $SE_B = 0.45$, *Wald* statistic = 5.41, $p = .020$, $OR = 0.36$. Effects of type of decision protocol and description presence were not significant, both *Wald* statistics < 2.31, both *p*s > .129.

Scale construction.

Considering the relatively small sample size ($N = 96$) in Study 2 the factor structures obtained in Study 1 were applied on observer ratings of Study 2 to make both studies comparable. The eight rating scales were constructed parallel to Study 1 (factor loadings of observer ratings in Study 2 after oblique rotation are displayed in Appendices C and D). Internal consistencies of the eight scales and corrected item-total correlations (all *CITC*s > .22) were satisfactory: Perceived perceptual basis (*alpha* = .76), perceived confidence (*alpha* = .91), perceived decision time and difficulty (*alpha* = .92), perceived persuasiveness of reasons (*alpha* = .81), perceived automatic decision (*alpha* = .65), perceived deliberative and effortful decision (*alpha* = .91), perceived absolute decision and a good memory for the perpetrator (*alpha* = .44) and perceived eliminative and relative decision (*alpha* = .76).

Observer ratings as a function of objective identification accuracy (ecological validities).

Parallel to Study 1 separate 2 x 2 ANOVAs were conducted for each scale with objective identification accuracy (incorrect ID vs. correct ID) and type of decision protocol (retrospective reasoning vs. think-aloud) as classification variables. Means and standard deviations of each scale as well as the effect sizes are displayed in Table 6.

Similar to Study 1, five observer rating scales differed as a function of objective identification accuracy, all Fs(1, 96) > 6.63, all ps < .012. Moreover, there were significant interactions between objective identification accuracy and type of decision protocol for ratings of perceived perceptual basis, perceived decision time and difficulty and for perceived absolute decision processes accompanied by a good memory for the perpetrator, all Fs(1, 92) > 4.79, all ps < .031. For ratings of perceived eyewitness confidence and perceived deliberative and effortful decision processes the interactions were only marginally significant, all *Fs*(1, 92) > 3.39, all *p*s < .069.[6]

Observer ratings as a function of perceived identification accuracy (subjective utilities).

To test for differences in observer ratings due to judgments of perceived identification accuracy (identifications perceived as incorrect vs. correct) and type of decision protocol (retrospective reasoning vs. think-aloud) again separate 2 x 2 ANOVAs were conducted for each rating scale.

Six rating scales differed due to observer judgments of perceived identifications accuracy, all *F*s(1, 92) > 9.37, all *p*s < .003. For none of the eight scales were there any interactions between observer judgments and type of decision protocol, all *F*s(1, 92) < 1.73, all *p*s > .192. Means and standard deviations of each scale as well as the effect sizes are displayed in Table 7.

Brunswikian lens model analyses: Do observers use appropriate cues?

Parallel to Study 1, the Brunswikian lens model was constructed separately for the think-aloud and the retrospective reasoning conditions (*n* = 48 judgments each).

---

[6] Significant main effects of type of decision protocol were found for ratings of perceived decision time and difficulty, $F(1, 92) = 4.64$, $p = .034$, $g_u = 0.42$, 95% CI [0.01, 0.82], and perceived eliminative and relative decision processes, $F(1, 92) = 12.03$, $p = .001$, $g_u = 0.68$, 95% CI [0.27, 1.09], with higher ratings in the think-aloud than in the retrospective reasoning condition.

In the think-aloud condition, ratings explained 44% of the variance of objective identification accuracy (p = .002) and 37% of the variance of judgments of perceived identification accuracy (p = .015). Six of the eight zero-order correlations demonstrated significant relationships between the ratings and the objective identification accuracy. In contrast, five ratings were related with perceived identification accuracy (see Figure 3). The two sets of correlation coefficients (Fisher's Z_r transformations were used) were related by *ICC* = .93 demonstrating high correspondence between ecological validities and subjective utilities.

In the retrospective reasoning condition ratings explained only 13% of the variance of objective identification accuracy (p = .660), while 49% of the variance of judgments of perceived identification accuracy were explained (p < .001). Seven scales were significantly correlated with observer judgments of identification accuracy. In contrast, only one of the eight zero-order correlations demonstrated a significant relationship with objective identification accuracy (see Figure 4). The two sets of correlation coefficients were not related, *ICC* = -.12, reflecting essential non-correspondence between ecological validities and subjective utilities.

Discussion

The purpose of these studies was to examine observers' judgmental processes when evaluating the accuracy of an identification decision, focusing on observers' subjective perceptions and interpretations of different identification qualities. Three major research questions were addressed: (1) Do cues as perceived by observers discriminate between correct and incorrect identification decisions (*ecological validities*)? (2) How do observers use these cues to make their judgments (*subjective utilities*) and how do they weight them? (3) Are these relationships better visible in think-aloud protocols and video presentations due to an increased cue saliency than in retrospective reasoning protocols and transcripts?

Ecological Validities: Observer Ratings as Indicators of Identification Accuracy?

In both studies, there were five cues, as perceived by observers, discriminating between objectively correct and incorrect identifications. Observer ratings of perceived confidence (Study 1: g_u = 0.28; Study 2: g_u = 0.54) and perceived automatic decision processes (Study 1: g_u = 0.48; Study 2: g_u = 0.76) were higher for objectively correct than for incorrect identifications. In contrast, perceived decision time and difficulty (Study 1: g_u = -0.25; Study 2: g_u = -0.50), perceived deliberative and effortful decision processes (Study 1: g_u = -0.32; Study 2: g_u = -0.64) as well as perceived relative and eliminative decision processes (Study 1: g_u = -0.23; Study 2: g_u = -0.55) were rated lower for correct than for incorrect identifications.

These results are in line with previous studies on variables to postdict identification accuracy (e.g., Dunning & Stern, 1994; Sporer, 1992, 1993; Sporer et al., 1995; Wells, 1984), supporting the existence of ecologically valid indicators to discriminate between correct and incorrect identifications. However, in the present study we did not examine eyewitnesses' self-report measures but observers' subjective ratings of these variables. We found that witness self-report measures do not seem to be necessary to postdict identification accuracy. Instead, observers were also able to correctly gauge the underlying decision processes based on witnesses’ identification decision protocols. From an applied perspective, this is of high practical relevance when witness self-reports at later stages of the investigation or at a trial are assumed to be biased (e.g., through post-identification feedback: Steblay, Wells, & Bradfield Douglass, 2014).

In the present studies the witnesses' numerical estimates of confidence per se were not presented to observers to avoid their strong influences on observers' judgments that were observed in former studies (e.g., Beaudry et al., 2015; Bradfield & Wells, 2000; Cutler et al., 1988; Wells et al., 1979). Nevertheless, our results suggest that witness confidence will be inferred by observers in a way that it discriminates between correct and incorrect identifications. In line with Martire and Kemp (2009), observers seem to use "more than just these numerical statements to evaluate confidence, possibly incorporating verbal and nonverbal cues into their estimates" (p. 233). As these estimates show similarly high correlations with objective identification accuracy as witness self-reports, observer ratings of confidence may be used as fruitful alternative indicators of identification accuracy.

However, not all scales as perceived by observers were valid indicators of identification accuracy. Neither the attributed eyewitness's perceptual basis, the perception of markers of an absolute decision strategy nor the rated quality of reasons differed as a function of identification accuracy. However, in both studies the observed relationships between observer ratings and objective identification accuracy tended to depend on the type of decision protocol the observers were presented with.

Effects of type of decision protocol and presentation medium on ecological validities.

Although not all of the two-way interactions reached significance, several observer ratings were valid indicators of identification accuracy only when think-aloud protocols were presented. When videotapes were used (Study 2) observer ratings of six of the eight scales strongly discriminated between correct and incorrect identifications only in think-aloud protocols (i.e., perceived perceptual basis, g_u = 0.79; perceived confidence, g_u = 0.97; perceived decision time and difficulty, g_u = -0.94; perceived automatic decisions, g_u = 1.01; perceived deliberative and effortful decision processes, g_u = -1.10; perceived absolute decisions, g_u = 0.73). When retrospective reasoning protocols were presented, ratings for these scales did not differ significantly (g_us ranged from -0.25 to 0.47).

In contrast, when written transcripts were presented (Study 1) the effect of type of decision protocol was far from being as strong as with the use of videotapes. Differences in observer ratings were only shown for perceived confidence (g_u = 0.43) and perceived deliberative and effortful decisions (g_u = -0.63), which were valid indicators of identification accuracy in the think-aloud condition only, but not with retrospective reasoning protocols (g_us ranged from -0.03 to 0.14).

Increased cue saliency due to think-aloud procedures?

In line with our hypotheses, discriminating cues seemed to be more apparent for observers when witnesses were instructed to think-aloud during the identification task than when they had to give reasons for their decisions afterwards. Think-aloud procedures allow observers to trace decision steps the witness engages in and display reconstructive memory activities and qualifiers that are associated with poor memories (Schooler et al., 1986; Wells & Lindsay, 1983). This procedure probably facilitates an adequate rating of the witness's cognitive effort to make a decision, the degree of deliberation as well as the witness's decision speed. Moreover, observers may form an impression of the witness's perceptual basis more easily based on verbalized thoughts about specific physical features of the perpetrator or references to the witness's memory.

The results support Nisbett and Wilson's (1977) claim that people are aware about their results of thinking (i.e., their identification decision) but are unable to access past thought processes correctly that describe *how* they made their final decisions. Thus, Nisbett and Wilson (1977) assumed that people cannot explain the causes for their behavior adequately afterwards. Moreover, our results are in line with Ericsson and Simon's (1993) theory that retrospective reports (i.e., answers to questionnaires) are often incomplete due to forgetting or selective reporting of thoughts and mental processes (cf. also Bainbridge, 1999). Consequently, witnesses in the reasoning conditions did not mention enough cognitive details retrospectively, either due to limited access to their actual thought processes during the identification task or due to a biased appraisal of their decision processes going along with an individual report threshold (cf. strategic regulation of memory accuracy: Koriat & Goldsmith, 1996; Pansky, Koriat, & Goldsmith, 2005).

Support for the assumption of less detailed reasoning protocols was obtained by analyzing protocol length showing that retrospective reasoning protocols were much shorter than think-aloud protocols. Nonetheless, in the reasoning condition there were also significant differences of perceived automatic (Study 1: $g_u = 0.56$) and perceived relative and eliminative decision processes (Study 2: $g_u = -0.63$) as a function of identification accuracy. Thus, although reasoning reports were clearly shorter than think-aloud protocols, witnesses did report these decision strategies at least to some extent when they justified their decisions. We encourage future research to explicitly focus on objectively measurable content differences in think-aloud and retrospective reasoning reports. Think-aloud and reasoning protocols could be coded for the presence of different decision processes (and not just rated based on observers' subjective impressions), to further test our explanation of the observed results.

Increased cue saliency due to the use of videotapes?

Like a magnifying glass, think-aloud protocols seemed to be much more effective to allow observers to assess discriminating cues when they were presented as videotapes compared to written transcripts. Videotaped peripheral witness characteristics that were visible *during* the identification task (e.g., decision time, hesitations, hedges or facial expressions) facilitated the assessment of valid indicators of identification accuracy (cf. Erickson et al., 1978; Leippe, 1994). Thus, our results support the common recommendation to videotape the original identification procedure (Garrett, 2011; Sporer, 1992, 1993; Wells et al., 1998) and encourage the application of think-aloud methods.

In contrast, in both retrospective reasoning conditions observer ratings only had low predictive value to explain objective identification accuracy. Thus, videotapes of witnesses who have to give reasons for their identification decision afterwards do not seem to be more effective than written transcripts in general. As there were almost no valid indicators of identification accuracy that were perceivable for observers, the use of reasoning protocols does not seem to be sufficient to allow fact finders' evaluation of identification accuracy at all.

In sum, observers seem to be able to perceive several ecologically valid cues to discriminate between correct and incorrect identifications, especially when think-aloud protocols are used and the identification decision is videotaped. Thus, videotaped think-aloud protocols seem to be a fruitful approach to increase the salience of valid indicators of identification accuracy, and to make them perceivable and usable by observers. As a result an increase in judgment accuracy can be expected. However, in Study 2, judgment accuracy in the videotaped think aloud condition (70.8%) was not significantly higher compared to the videotaped retrospective reasoning condition (56.3%), probably due to the small sample size (n = 48 in each condition). Further improving think-aloud instructions and pilot testing them might increase this difference.

Subjective Utilities: Which Cues Did Observers Use to Make Their Judgments?

Independently of type of decision protocol, observers heavily used seven of the eight scales to make their judgments in the theoretically expected way (for additional support see Semmler et al., 2012), with results having been almost parallel in both studies.

Based on the interpersonal source monitoring framework (Mitchell & Johnson, 2000), an increased number of cognitive operations are perceived as an indicator of an erroneous memory (Leippe, 1994; Sporer, 2004, 2008). In line with this assumption, in both studies observers evaluated identification decisions as incorrect more often when they were perceived as highly deliberative and effortful (Study 1: g_u = -0.44; Study 2: g_u = -0.57), that is, when the decisions contained a higher number of cognitive operations. In contrast, identification decisions that were perceived as highly automatic were more likely judged as correct than incorrect in Study 1 (g_u = 0.50), but the effect failed to reach significance in Study 2 (g_u = 0.28).

Similarly, identifications were judged as incorrect more frequently when the identification decision was perceived as slow and difficult (Study 1: g_u = -0.63; Study 2: g_u = -0.62), which was also found by Neal et al. (2012). However, from his archival analysis Garrett (2011) concluded that fact finders often seemed to ignore witnesses' initial hesitations and judged the identification as correct. Perhaps training of judges, expert testimony or jury instructions might help decision makers.

Regarding observers' use of absolute and relative decision processes (Wells, 1984), results were mixed. Observers heavily used an attributed good memory for the perpetrator accompanied by an absolute decision as an indicator of a correct identification (Study 1: g_u = 1.00; Study 2: g_u = 1.06). Dunning and Stern (1994) also found that observers judged identifications as correct more often, when witnesses reported a higher impact of their own memory than of the other photos in the lineup, which reflects an absolute decision strategy to some extent. However, contradictory to our assumptions, in both studies observers did not use perceived relative and eliminative decision processes as an indicator of an incorrect identification. In Study 2, they even tended to interpret them in the opposite direction (g_u = 0.41), however the effect failed to reach significance due to the small sample size. In Dunning and Stern's (1994) study, observers also failed to associate eliminative decision processes with identification accuracy. Perhaps, observers interpret their perception of eliminative strategies as an indicator for a conscientious and reliable witness.

As suggested by different courts (e.g., Neil vs. Biggers, 1972), identifications were judged as correct more often when the witness's attributed perceptual basis (Study 1: g_u = 1.08; Study 2: 0.99) and perceived confidence (Study 1: g_u = 0.87; Study 2: g_u = 0.99) were high. Several studies have shown observers' high reliance on confidence for a long time (e.g., Wells et al., 1979). As expected observers also seemed to infer the witnesses' perceptual basis from their identification protocols and used this cue heavily to make their judgments.

Finally, identifications were judged as correct when the reported reasons were perceived as very persuasive (Study 1: g_u = 0.94; Study 2: 1.11). As persuasiveness was defined in terms of detailedness and consistency of the reported reasons, our results replicate previous findings that observers use these factors intuitively (cf. Bell & Loftus, 1989; Berman & Cutler, 1996).

To sum up, observers seem to heavily use the investigated cues to make their judgments. However, high judgment accuracies can only be observed if the used cues validly discriminate between correct and incorrect identifications as well and are given an appropriate weight.

Ecological Validities vs. Subjective Utilities: Did Observers Use the Cues Appropriately?

In Study 1, observer judgment accuracy was 54.9%, which is comparable to chance level, and thus mirrors previous research results (for reviews see Boyce, Beaudry, & Lindsay, 2007; Wells & Lindsay, 1983). In contrast, when videotapes were used (Study 2) judgment accuracy (63.5%) exceeded chance level.

We applied the Brunswikian lens model to explain the observed judgment accuracies by focusing on observers' weighting of cues. Therefore, objectively valid and subjectively used cues were contrasted to investigate if observers used the cues appropriately (cf. Hartwig & Bond, 2011; Sporer & Küpper, 1995).

Brunswikian analyses in the retrospective reasoning conditions.

In the retrospective reasoning conditions of both studies observer ratings showed greater predictive value for the judgments of perceived identification accuracy (Study 1: $R^2 = .25$; Study 2: $R^2 = .49$) than for objective identification accuracy (Study 1: $R^2 = .12$; Study 2: $R^2 = .13$). Although there was only one valid indicator of identification accuracy in each study, observers heavily relied on the majority of the investigated cues and overestimated their discriminative value (Figures 1 and 3). In Study 2, observers also failed to use eliminative and relative decision processes as a valid indicator of an incorrect identification decision and even interpreted it in the opposite direction.

A comparison of the correlations between observer ratings and objective identification accuracy on the one hand and between ratings and perceived identification accuracy on the other hand reveals that observers gave far too much weight to non-diagnostic cues. Thus, correspondences between ecological validities and subjective utilities were low (Study 1: *ICC* = .22; Study 2: *ICC* = -.12).

To conclude, the low judgment accuracies in the retrospective reasoning conditions (i.e., 51.4% for written transcripts and 56.3% for videotapes) can be explained by observers' inappropriate use of non-diagnostic cues. However, it is noteworthy that ratings explained only 25%

and 49% of the variance of observer judgments. Thus, there must be also other influence factors that are not investigated here (e.g., perceived description qualities, perceived witness characteristics: see Kaminski & Sporer, 2016).

Brunswikian analyses in the think-aloud conditions.

Although correspondences between ecological validities and subjective utilities were much higher in both think-aloud conditions compared to the retrospective reasoning conditions, the analyses differed due to presentation medium.

Study 1: Written transcripts of think-aloud protocols.

When transcribed think-aloud protocols were used observer ratings showed greater predictive value for the judgments of perceived identification accuracy (R^2 = .27) than for objective identification accuracy (R^2 = .12), which is comparable to both reasoning conditions. However, contrary to the reasoning conditions (Study 1: *ICC* = .22; Study 2: *ICC* = -.12), correspondence between ecological validities and subjective utilities was moderately high when think-aloud protocols were used (*ICC* = .68), reflecting an appropriate weighting of the investigated cues, but also some discrepancies.

First, observers were highly sensitive to the three discriminating cues (i.e., perceived confidence, perceived automatic decision, perceived deliberative and effortful decision) and weighted them appropriately. For these scales the correlations with the objective and perceived identification accuracy were in the same direction and comparable in size (see Figure 2). Second, observers tended to appropriately weight perceived decision time and difficulty, although this cue did not significantly discriminate between correct and incorrect identifications. Third, observers gave low weight to eliminative and relative decision processes, which indeed were not valid indicators of identification accuracy.

However, there were also discrepancies. Observers highly overestimated the discriminative value of the witnesses' attributed perceptual basis, the perceived quality of reasons and the perceived witness's memory quality for the perpetrator resulting in an absolute decision. Although these cues were not related to objective identification accuracy, observers gave them the highest weights. These results support the assumption that observers tend to overestimate some situational factors, like the attributed witness's attention and his/her quality of view of the perpetrator, when making their judgments (Boyce et al., 2007). Lindsay (1994) had also reported that participants rated these factors as most important cues to identification accuracy. Although we do not deny the importance of situational factors (cf. Shapiro & Penrod, 1986) their impact may not be gauged easily from witness statements.

Study 2: Videotapes of think-aloud protocols.

Contrary to the other conditions, observer ratings based on think-aloud protocols showed greater predictive value for objective identification accuracy (R^2 = .44) than for the judgments of perceived identification accuracy (R^2 = .37). Six of eight cues showed moderate to high relationships with objective identification accuracy with observers being sensitive to these relationships and weighting these cues highly appropriately. Thus, correspondence between ecological validities and subjective utilities was almost perfect. Minor discrepancies are reflected in observers' underestimation of the discriminative value of automatic and deliberative and effortful decision processes (Figure 4). Moreover, they failed to give sufficient weight to eliminative and relative decision processes although this cue was moderately related to objective identification accuracy (similar results were found by Dunning & Stern, 1994).

According to Leippe's (1994) validity-intuition model, judgment accuracy increases if observers intuitively use highly valid cues and ignore invalid cues. This was the case when videotaped think-aloud protocols were used resulting in a judgment accuracy of 70.8% that dropped to 58.3% with literally transcribed think-aloud protocols. In videotaped think-aloud protocols discriminative cues were much more visible for observers explaining 44% of the variance of objective identification accuracy (vs. only 12% in transcripts of think-aloud protocols). Thus, observers' intuitive use and correct weighting of these cues produced higher judgment accuracy.

Conclusions and Practical Implications

To sum up, observers were able to detect several valid cues to discriminate between correct and incorrect identifications that were present in eyewitnesses' identification decision protocols. However, this finding was mainly restricted to the use of videotaped think-aloud procedures. Videos showing witnesses who think-aloud during the identification task probably make discriminating cues more salient for observers compared to the use of written transcripts and compared to retrospective reasoning protocols. The latter just seem to elicit a post hoc rationalization to explain the identification decision, with presentation medium being not helpful to detect discriminating cues. In contrast, videotaped think-aloud protocols seem to allow observers to detect witnesses who just beat around the bush and to distinguish them from those who rely on their memory and make absolute and automatic decisions. However, due to the lack of random assignment no direct statistical comparisons of the two studies are legitimate. Hence, effects of presentation medium have to be interpreted with caution.

Independent of type of decision protocol observers heavily relied on the investigated cues to make their judgments and correctly interpreted their relationships with objective identification accuracy in both studies. However, the investigated cues did not explain more than 44% of the variance in observers’ judgments. Thus, other cues have to be considered to gain a deeper understanding of observers’ evaluation processes.

To investigate how much weight is given to each cue the Brunswikian lens model offers a useful method to understand the relationships between empirically valid and intuitively used cues. In the videotaped think-aloud condition there were good correspondences in the use of valid cues, whereas in the reasoning condition cue discriminability was highly overestimated.

To increase fact finders’ judgment accuracy when evaluating an identification decision we recommend (1) to use think-aloud protocols to record witnesses’ decision processes which make discriminating cues more salient, (2) to videotape the identification decision and (3) to inform fact finders about valid factors that are associated with identification accuracy. We encourage to pilot test and refine think-aloud instructions to optimize them for the evaluation of identification decisions. Ultimately, guidelines about the appropriate weighting of valid indicators of identification decisions can be devised on the basis of Brunswikian analyses.

References

Bainbridge, L. (1999). Verbal reports as evidence of the process operator's knowledge. *International Journal of Human-Computer Studies, 51,* 213-238. doi: 10.1006/ijhc.1979.0307

Beaudry, J. L., Lindsay, R. C. L., Leach, A.-M., Mansour, J. K., Bertrand, M. I., & Kalmet, N. (2015). The effect of evidence type, identification accuracy, line-up presentation, and line-up administration on observers' perceptions of eyewitnesses. *Legal and Criminological Psychology, 20,* 343-364. doi: 10.1111/lcrp.12030

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

Bell, B. E., & Loftus, E. F. (1989). Trivial persuasion in the courtroom: The power of (a few) minor details. *Journal of Personality and Social Psychology, 56,* 669-679. doi: 10.1037/0022-3514.56.5.669

Benton, T. R., Ross, D. F., Bradshaw, E., Thomas, W. N., & Bradshaw, G. S. (2006). Eyewitness memory is still not common sense: Comparing jurors, judges and law enforcement to eyewitness experts. *Applied Cognitive Psychology, 20,* 115-129. doi: 10.1002/acp.1172

Berman, G. L., & Cutler, B. L. (1996). Effects of inconsistencies in eyewitness testimony on mock-juror decision making. *Journal of Applied Psychology, 81,* 170-177. doi: 10.1037/0021-9010.81.2.170

Berman, G. L., Narby, D. J., & Cutler, B. L. (1995). Effects of inconsistent eyewitness statements on mock-jurors evaluations of the eyewitness, perceptions of defendant culpability and verdicts. *Law and Human Behavior, 19,* 79-88. doi: 10.1007/BF01499074

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. 221-235). New York: Russell Sage Foundation.

Boyce, M. A., Beaudry, J. L., & Lindsay, R. C. L. (2007). Belief of eyewitness identification evidence. In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Handbook of eyewitness psychology: Memory for people* (Vol. 2, pp. 25-45). Mahwah, NJ: Lawrence Erlbaum and Associates.

Bradfield, A. L., & Wells, G. L. (2000). The perceived validity of eyewitness identification testimony: A test of the five Biggers criteria*. Law and Human Behavior, 24,* 581-594. doi: 10.1023/a:1005523129437

Bradfield Douglass, A., & Jones, E. E. (2013). Confidence inflation in eyewitnesses: Seeing is not believing. *Legal and Criminological Psychology, 18,* 152-167. doi: 10.1111/j.2044-8333.2011.02031.x

Brewer, N., & Wells, G. L. (2006). The confidence-accuracy relationship in eyewitness identification: Effects of lineup instructions, foil similarity, and target-absent base rates. *Journal of Experimental Psychology: Applied, 12,* 11-30. doi: 10.1037/1076-898x.12.1.11

Brigham, J. C., & Bothwell, R. K. (1983). The ability of prospective jurors to estimate the accuracy of eyewitness identifications. *Law and Human Behavior, 7,* 19-30. doi: 10.1007/BF01045284

Brigham, J. C., & Wolfskeil, M. P. (1983). Opinions of attorneys and law-enforcement personnel on the accuracy of eyewitness identifications. *Law and Human Behavior, 7,* 337-349. doi: 10.1007/bf01044736

Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. Berkeley: University of California Press.

Brunswik, E. (1965). *The conceptual framework of psychology*. Chicago, London: The University of Chicago Press.

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

Cutler, B. L., Penrod, S. D., & Stuve, T. E. (1988). Juror decision-making in eyewitness identification cases. *Law and Human Behavior, 12,* 41-55. doi: 10.1007/BF01064273

Desmarais, S. L., & Read, J. D. (2011). After 30 years, what do we know about what jurors know? A meta-analytic review of lay knowledge regarding eyewitness factors. *Law and Human Behavior, 35,* 200-210. doi: 10.1007/s10979-010-9232-6

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

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

Erickson, B., Lind, E. A., Johnson, B. C., & O'Barr, W. M. (1978). Speech style and impression-formation in a court setting: Effects of powerful and powerless speech. *Journal of Experimental Social Psychology, 14,* 266-279. doi: 10.1016/0022-1031(78)90015-X

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. Massachusetts: A Bradford Book.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4,* 272-299. doi: 10.1037/1082-989X.4.3.272

Fiedler, K. (1989). Lügendetektion aus alltagspsychologischer Sicht [Lie detection based on commonsense]. *Psychologische Rundschau, 40*(3), 127-140.

Fox, M. C., Ericsson, K. A., & Best, R. (2011). Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods. *Psychological Bulletin, 137,* 316-344. doi: 10.1037/a0021663

Garrett, B. L. (2011). *Convicting the innocent: Where criminal prosecutions go wrong.* Cambridge, Massachusetts: Harvard University Press.

Garrett, B. L. (2012). Eyewitnesses and exclusion. *Vanderbilt Law Review, 65*(2), 451-506.

Hammond, K. R., Hursch, C. J., & Todd, F. J. (1964). Analyzing the components of clinical inference. *Psychological Review, 71,* 438-456. doi: 10.1037/h0040736

Hartwig, M., & Bond, C. F. (2011). Why do lie-catchers fail? A lens model meta-analysis of human lie judgments. *Psychological Bulletin, 137,* 643-659. doi: 10.1037/a0023589

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114,* 3-28. doi: 10.1037/0033-2909.114.1.3

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88,* 67-85. doi: 10.1037//0033-295X.88.1.67

Kaminski, K. S. & Sporer, S. L. (2016). *The evaluation of eyewitness identification decisions by indirect measures.* Manuscript submitted for publication.

Kassin, S. M. (1998). Eyewitness identification procedures: The fifth rule. *Law and Human Behavior, 22,* 649-653. doi: 10.1023/A:1025702722645

Kassin, S. M., Rigby, S., & Castillo, S. R. (1991). The accuracy-confidence correlation in eyewitness testimony: Limits and extensions of the retrospective self-awareness effect. *Journal of Personality and Social Psychology, 61,* 698-707. doi: 10.1037//0022-3514.61.5.698

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103,* 490-517. doi: 10.1037/0033-295x.103.3.490

Leippe, M. R. (1994). The appraisal of eyewitness testimony. In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 385-418). Cambridge, England: Cambridge University Press.

Lindsay, R. C. L. (1994). Expectations of eyewitness performance: Juror's verdicts do not follow from their beliefs. In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 362-384). Cambridge, England: Cambridge University Press.

Lindsay, R. C. L., & Bellinger, K. (1999). Alternatives to the sequential lineup: The importance of controlling the pictures. *Journal of Applied Psychology, 84,* 315-321. doi: 10.1037//0021-9010.84.3.315

Lindsay, R. C. L., Semmler, C., Weber, N., Brewer, N., & Lindsay, M. R. (2008). How variations in distance affect eyewitness reports and identification accuracy. *Law and Human Behavior, 32,* 526-535. doi: 10.1007/s10979-008-9128-x

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

Luus, C. A. E., & Wells, G. L. (1991). Eyewitness identification and the selection of distracters for lineups. *Law and Human Behavior, 15,* 43-57. doi: 10.1007/BF01044829

Martire, K. A., & Kemp, R. I. (2009). The impact of eyewitness expert evidence and judicial instruction on juror ability to evaluate eyewitness testimony. *Law and Human Behavior, 33,* 225-236. doi: 10.1007/s10979-008-9134-z

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*, 30-46. doi: 10.1037/1082-989X.1.4.390

Mitchell, K. J., & Johnson, M. K. (2000). Source monitoring: Attributing mental experience. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 179-195). Oxford, UK: Oxford University Press.

National Academy of Sciences (2014). *Identifying the culprit: Assessing eyewitness identification.* Washington, DC: The National Academies Press.

Neal, T. M. S., Christiansen, A., Bornstein, B. H., & Robicheaux, T. R. (2012). The effects of mock jurors' beliefs about eyewitness performance on trial judgments. *Psychology, Crime & Law, 18,* 49-64. doi: 10.1080/1068316X.2011.587815

Neil vs. Biggers, 409 U.S. 188 (1972).

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84,* 231-259. doi: 10.1037//0033-295X.84.3.231

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

Palmer, M. A., Brewer, N., & Weber, N. (2010). Post-identification feedback affects subsequent eyewitness identification performance. *Journal of Experimental Psychology: Applied, 16,* 387-398. doi: 10.1037/a0021034

Pansky, A., Koriat, A., & Goldsmith, M. (2005). Eyewitness recall and testimony. In N. Brewer & K. D. Williams (Eds.), *Psychology and law: An empirical perspective* (pp. 93-150). New York: Guilford Press.

Petty, R. E., & Cacioppo, J. T. (1986). *Communication and persuasion: Central and peripheral routes to attitude change.* New York: Springer.

Police and Criminal Evidence Act (1984). *Code D: Code of practice for the identification of persons by police officers.* Retrieved from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/253831/pace-code-d-2011.pdf

Reardon, M. C., & Fisher, R. P. (2011). Effect of viewing the interview and identification process on juror perceptions of eyewitness accuracy. *Applied Cognitive Psychology, 25,* 68-77. doi: 10.1002/acp.1643

Sauerland, M., & Sporer, S. L. (2007). Post-decision confidence, decision time, and self-reported decision processes as postdictors of identification accuracy. *Psychology, Crime & Law, 13,* 611-625. doi: 10.1080/10683160701264561

Schooler, J. W., Gerhard, D., & Loftus, E. F. (1986). Qualities of the unreal. *Journal of Experimental Psychology: Learning Memory and Cognition, 12,* 171-181. doi: 10.1037//0278-7393.12.2.171

Semmler, C., & Brewer, N. (2002). Research report: Effects of mood and emotion on juror processing and judgments. *Behavioral Sciences & the Law, 20,* 423-436. doi: 10.1002/bsl.502

Semmler, C., Brewer, N., & Bradfield Douglass, A. (2012). Jurors believe eyewitnesses. In B. L. Cutler (Ed.), *Conviction of the innocent: Lessons from psychological research* (pp. 185-209). Washington, DC: American psychological Association.

Shapiro, P. N., & Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological Bulletin, 100,* 139-156. doi: 10.1037/0033-2909.100.2.139

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86,* 420-428. doi: 10.1037//0033-2909.86.2.420

Smith, S. M., Lindsay, R. C. L., & Pryke, S. (2000). Postdictors of eyewitness errors: Can false identifications be diagnosed? *Journal of Applied Psychology, 85,* 542-550. doi: 10.1037//0021-9010.85.4.542

Sporer, S. L. (1992). Post-dicting eyewitness accuracy: Confidence, decision-times and person descriptions of choosers and non-choosers.

European Journal of Social Psychology, 22, 157-180. doi: 10.1002/ejsp.2420220205

Sporer, S. L. (1993). Eyewitness identification accuracy, confidence, and decision times in simultaneous and sequential lineups. *Journal of Applied Psychology, 78,* 22-33. doi: 10.1037/0021-9010.78.1.22

Sporer, S. L. (1994). Decision times and eyewitness identification accuracy in simultaneous and sequential lineups. In D. F. Ross, J. D. Read, & M. P. Togila (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 300-327). Cambridge: Cambridge University Press.

Sporer, S. L. (1997). The less travelled road to truth: Verbal cues in deception detection in accounts of fabricated and self-experienced events. *Applied Cognitive Psychology, 11,* 373-397. doi: 10.1002/(SICI)1099-0720(199710)11:5<373::AID-ACP461>3.0.CO;2-0

Sporer, S. L. (2004). Reality monitoring and the detection of deception. In P. A. Granhag & L. A. Stroemwall (Eds.), *Deception detection in forensic contexts* (pp. 64-102). Cambridge: Cambridge University Press.

Sporer, S. L. (2007). Evaluating witness evidence: The fallacies of intuition. In C. Engel & F. Strack (Eds.), *The impact of court procedure on the psychology of judicial decision making* (pp. 111-150). Baden-Baden, Germany: Nomos Verlag.

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., & Cutler, B. L. (2003). Identification evidence in Germany and the United States. In P. J. van Koppen & S. D. Penrod (Eds.), *Adversarial versus inquisitorial justice: Psychological perspectives on criminal justice systems* (pp. 191-208). New York: Springer.

Sporer, S. L., & Küpper, B. (1995). Realitätsüberwachung und die Beurteilung des Wahrheitsgehaltes von Erzählungen: Eine experimentelle Studie [Reality monitoring and the judgment of the truthfulness of accounts: An experimental study]. *Zeitschrift für Sozialpsychologie, 26*(3), 173-193.

Sporer, S. L., Masip, J., & Cramer, M. (2014). Guidance to detect deception with the Aberdeen Report Judgment Scales: Are verbal content cues useful to detect false accusations? *American Journal of Psychology, 127*(1), 43-61.

Sporer, S. L., Penrod, S., Read, D., & Cutler, B. (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., Wells, G. L., & Bradfield Douglass, A. (2014). The eyewitness post identification feedback effect 15 years later: Theoretical and policy implications. *Psychology, Public Policy, and Law, 20,* 1-18. doi: 10.1037/law0000001

Tredoux, C. G. (1998). Statistical inference on measures of lineup fairness. *Law and Human Behavior, 22,* 217-237. doi: 10.1023/a:1025746220886

Vicaria, I. M., Bernieri, F. J., & Isaacowitz, D. M. (2015). Perceptions of rapport across the life span: Gaze patterns and judgment accuracy. *Psychology and Aging, 30,* 396-406. doi: 10.1037/pag0000019

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

Wells, G. L. (1984). The psychology of lineup identifications. *Journal of Applied Social Psychology, 14,* 89-103. doi: 10.1111/j.1559-1816.1984.tb02223.x

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

Wells, G. L., & Leippe, M. R. (1981). How do triers of fact infer the accuracy of eyewitness identifications: Using memory for peripheral detail can be misleading. *Journal of Applied Psychology, 66,* 682-687. doi: 10.1037/0021-9010.66.6.682

Wells, G. L., & Lindsay, R. C. L. (1983). How do people infer the accuracy of eyewitness memory? Studies of performance and a metamemory analysis. In S. Lloyd-Bostock & B. R. Clifford (Eds.), *Evaluating witness evidence: Recent psychological research and new perspectives* (pp. 41-55). Chichester: John Wiley & Sons.

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

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

Wells, G. L., Small, M., Penrod, S., Malpass, R. S., Fulero, S. M., & Brimacombe, C. A. E. (1998). Eyewitness identification procedures: Recommendations for lineups and photospreads. *Law and Human Behavior, 22,* 603-647. doi: 10.1023/a:1025750605807

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

Wise, R. A., & Safer, M. A. (2004). What US judges know and believe about eyewitness testimony. *Applied Cognitive Psychology, 18,* 427-443. doi: 10.1002/acp.993

Wixted, J. T., Mickes, L., Clark, S. E., Gronlund, S. D., & Roediger, H. L. (2015). Initial eyewitness confidence reliably predicts eyewitness identification accuracy. *American Psychologist, 70,* 515-526. doi: 10.1037/a0039510

Tables

Table 1

Observer Ratings (Based on Wells & Bradfield, 1998) of the Witnesses' Identification Statements

	Rating Variables	Scale Anchors
	Witnessing conditions	
(1)	How good was the witness's quality of view of the target?	*1 = very poor; 7 = very good*
(2)	How long was the target face in view for the witness?	*1 = very short; 7 = very long*
(3)	How well was the witness able to make out specific features of the target's face?	*1 = not at all; 7 = very well*
(4)	How much attention did the witness pay to the target' s face?	*1 = no attention; 7 = total attention*
(5)	Did the witness have a good basis (enough information) to make an identification?	*1 = no basis at all; 7 = very good basis*
	Identification decision	
(6)	How easy or difficult was it for the witness to make an identification?	*1 = very difficult; 7 = very easy*
(7)	How long did it take the witness to make an identification?	*1 = very short; 7 = very long*
(8)	How confident was witness that the identified person is the target?	*1 = not confident at all; 7 = totally confident*
(9)	How willing would the witness be to testify about his/her identification in court?	*1 = not willing at all; 7 = totally willing*
(10)	How willing would the witness be to swear an oath about his/her identification in court?	*1 = not willing at all; 7 = totally willing*

Table 2

Observer Ratings of the Witnesses' Decision Processes Using 7-point Likert Scales (1 = not at all; 7 = absolutely so)

	Rating Variables
(1)	The witness matched the image in his/her head to the pictures in front of him/her.
(2)	The witness first looked at all the photos before making a decision.
(3)	The witness compared each face with the others to make a decision.
(4)	The witness used a process of elimination.
(5)	The witness relied on specific facial features (e.g. nose, hair, eyes) when making an identification.
(6)	The person chosen by the witness seemed to be an exact match to his/her memory.
(7)	The witness just recognized the target and could not explain why.
(8)	The target's face seemed to just "pop out" at the witness.
(9)	The witness first eliminated the ones definitely not the target, then chose among the rest.
(10)*	The target face seemed to be the person closest to what the witness remembered but not exact.
(11)	The faces seemed to be all so similar that they made the decision more difficult.
(12)	The faces confused the witness, which made the task more difficult.
(13)	The witness had to think carefully to make a decision.
(14)	Much effort was necessary to make the decision.
(15)	The witness still seemed to have a clear picture of the target in mind.

Note. *Item 10 was dropped from later analyses due to low inter-correlations with other items.

Table 3

Observer Judgments (% of Identifications Perceived as Correct) and Judgment Accuracy (% Correct Judgments) for each Experimental Condition of Study 1 (N = 288) and Study 2 (N = 96)

	Study 1		Study 2	
	Observer Judgments	Judgment Accuracy	Observer Judgments	Judgment Accuracy
Total	25.0	54.9[a]	38.5	63.5[b]
Identification Accuracy				
Incorrect	20.1	79.9	25.0	75.0
Correct	29.9	29.9	52.1	52.1
Type of decision protocol				
Reasoning	27.8	51.4	31.3	56.3
Think-aloud	22.2	58.3	45.8	70.8
Description presence				
Without descriptions	20.8	55.6	31.3	64.6
With descriptions	29.2	54.2	45.8	62.5

Note. [a] Judgment accuracy did not differ from chance level of 50%, $t(287) = 1.66$, $p = .099$; [b] Judgment accuracy significantly differed from chance level of 50%, $t(95) = 2.74$, $p = .007$.

Table 4

Means and Standard Deviations of Observer Ratings of Study 1 (Using Written Transcripts) for Objectively Correct and Incorrect Identifications

	Total (*N* = 288)					Retrospective Reasoning (*n* = 144)					Think-aloud (*n* = 144)				
	Incorrect ID		Correct ID			Incorrect ID		Correct ID			Incorrect ID		Correct ID		
Observer ratings	*M*	*SD*	*M*	*SD*	g_u 95% CI]	*M*	*SD*	*M*	*SD*	g_u [95% CI]	*M*	*SD*	*M*	*SD*	g_u [95% CI]
Perceptual basis	3.59	1.05	3.69	1.14	0.09 [-0.14, 0.32]	3.73	1.06	3.75	1.12	0.02 [-0.30, 0.35]	3.46	1.02	3.63	1.16	0.16 [-0.17, 0.48]
Confidence	**3.14**	**1.31**	**3.53**	**1.47**	**0.28 [0.05, 0.51]**	3.26	1.43	3.46	1.41	0.14 [-0.19, 0.46]	**3.01**	**1.17**	**3.60**	**1.53**	**0.43 [0.10, 0.76]**
Decision time and difficulty	**4.78**	**0.96**	**4.53**	**1.01**	**-0.25 [-0.48, -0.02]**	4.61	0.95	4.43	1.00	-0.18 [-0.51, 0.15]	4.95	0.95	4.64	1.02	-0.32 [-0.65, 0.00]
Persuasiveness of reasons	3.35	1.06	3.20	1.11	-0.15 [-0.38, 0.09]	3.40	1.03	3.13	0.97	-0.27 [-0.60, 0.06]	3.31	1.09	3.27	1.24	-0.04 [-0.36, 0.29]
Automatic decision	**2.41**	**1.34**	**3.16**	**1.74**	**0.48 [0.25, 0.72]**	**2.65**	**1.31**	**3.47**	**1.59**	**0.56 [0.23, 0.89]**	**2.17**	**1.33**	**2.85**	**1.83**	**0.42 [0.09, 0.75]**
Deliberative/ effortful decision	**4.45**	**1.27**	**4.03**	**1.46**	**-0.32 [-0.55, -0.09]**	4.02	1.41	3.98	1.32	-0.03 [-0.36, 0.29]	**4.91**	**0.93**	**4.08**	**1.60**	**-0.63 [-0.97, -0.30]**
Absolute decision/ good memory for perpetrator	4.23	1.15	4.34	1.23	0.09 [-0.14, 0.32]	4.34	1.08	4.33	1.17	-0.01 [-0.34, 0.31]	4.11	1.21	4.35	1.29	0.19 [-0.14, 0.51]
Eliminative/ relative decision	**4.38**	**1.31**	**4.07**	**1.42**	**-0.23 [-0.46, 0.01]**	3.96	1.36	3.73	1.38	-0.17 [-0.49, 0.16]	4.80	1.12	4.42	1.39	-0.30 [-0.63, 0.02]

Note. g_u = corrected standardized mean difference [and 95% confidence interval]; Bold values represent significant mean differences ($p < .05$) and effect sizes.

Table 5

Means and Standard Deviations of Observer Ratings of Study 1 (Using Written Transcripts) for Perceived Correct and Incorrect Identifications

	Total (*N* = 288)					Retrospective Reasoning (*n* = 144)					Think-aloud (*n* = 144)				
	Perceived Incorrect ID		Perceived Correct ID			Perceived Incorrect ID		Perceived Correct ID			Perceived Incorrect ID		Perceived Correct ID		
Observer ratings	*M*	*SD*	*M*	*SD*	g_u 95% CI]	*M*	*SD*	*M*	*SD*	g_u [95% CI]	*M*	*SD*	*M*	*SD*	g_u [95% CI]
Perceptual basis	**3.38**	**0.99**	**4.45**	**1.00**	**1.08 [0.80, 1.36]**	**3.46**	**0.99**	**4.46**	**1.00**	**1.00 [0.62, 1.38]**	**3.29**	**0.98**	**4.43**	**1.01**	**1.15 [0.73, 1.56]**
Confidence	**3.05**	**1.31**	**4.19**	**1.33**	**0.87 [0.60, 1.15]**	**3.03**	**1.29**	**4.23**	**1.40**	**0.91 [0.53, 1.28]**	**3.07**	**1.33**	**4.16**	**1.25**	**0.83 [0.42, 1.23]**
Decision time and difficulty	**4.81**	**0.99**	**4.20**	**0.86**	**-0.63 [-0.90, -0.36]**	**4.68**	**0.99**	**4.10**	**0.82**	**-0.61 [-0.98, -0.24]**	**4.93**	**0.98**	**4.33**	**0.90**	**-0.62 [-1.02, -0.22]**
Persuasiveness of reasons	**3.04**	**0.93**	**3.99**	**1.22**	**0.94 [0.67, 1.22]**	**3.04**	**0.90**	**3.84**	**1.05**	**0.84 [0.47, 1.22]**	**3.04**	**0.95**	**4.18**	**1.40**	**1.07 [0.66, 1.48]**
Automatic decision	**2.59**	**1.58**	**3.37**	**1.49**	**0.50 [0.23, 0.77]**	**2.87**	**1.52**	**3.55**	**1.39**	**0.45 [0.09, 0.82]**	**2.33**	**1.60**	**3.14**	**1.61**	**0.50 [0.11, 0.90]**
Deliberative/ effortful decision	**4.40**	**1.38**	**3.80**	**1.30**	**-0.44 [-0.71, -0.17]**	**4.16**	**1.40**	**3.59**	**1.17**	**-0.42 [-0.78, -0.05]**	**4.62**	**1.33**	**4.05**	**1.42**	**-0.42 [-0.82, -0.03]**
Absolute decision/ good memory for perpetrator	**4.01**	**1.16**	**5.10**	**0.86**	**1.00 [0.73, 1.28]**	**4.08**	**1.14**	**4.98**	**0.77**	**0.85 [0.47, 1.23]**	**3.94**	**1.17**	**5.26**	**0.94**	**1.17 [0.75, 1.58]**
Eliminative/ relative decision	4.17	1.39	4.40	1.30	0.17 [-0.10, 0.43]	3.75	1.45	4.08	1.11	0.24 [-0.13, 0.60]	4.56	1.22	4.80	1.43	0.19 [-0.20, 0.58]

Note. g_u = corrected standardized mean difference [and 95% confidence interval]; Bold values represent significant mean differences ($p < .05$) and effect sizes.

Table 6

Means and Standard Deviations of Observer Ratings of Study 2 (Using Videos) for Objectively Correct and Incorrect Identifications

	Total (*N* = 96)					Retrospective Reasoning (*n* = 48)					Think-aloud (*n* = 48)				
	Incorrect ID		Correct ID			Incorrect ID		Correct ID			Incorrect ID		Correct ID		
Observer ratings	*M*	*SD*	*M*	*SD*	g_u [95% CI]	*M*	*SD*	*M*	*SD*	g_u [95% CI]	*M*	*SD*	*M*	*SD*	g_u [95% CI]
Perceptual basis	3.65	1.05	3.93	1.00	0.27 [-0.13, 0.67]	3.87	1.25	3.66	0.81	-0.19 [-0.75, 0.36]	**3.43**	**0.77**	**4.19**	**1.12**	**0.79 [0.21, 1.36]**
Confidence	**3.41**	**1.56**	**4.24**	**1.50**	**0.54 [0.14, 0.94]**	3.83	1.54	4.10	1.66	0.16 [-0.40, 0.72]	**2.99**	**1.50**	**4.39**	**1.34**	**0.97 [0.38, 1.56]**
Decision time and difficulty	**4.57**	**1.48**	**3.81**	**1.55**	**-0.50 [-0.91, -0.10]**	3.93	1.37	3.81	1.47	-0.08 [-0.64, 0.48]	**5.22**	**1.32**	**3.80**	**1.65**	**-0.94 [-1.52, -0.35]**
Persuasiveness of reasons	3.58	1.22	3.70	1.20	0.09 [-0.30, 0.49]	3.58	1.32	3.31	0.99	-0.23 [-0.79, 0.33]	3.58	1.14	4.08	1.28	0.41 [-0.16, 0.97]
Automatic decision	**2.49**	**1.54**	**3.77**	**1.79**	**0.76 [0.35, 1.17]**	2.83	1.36	3.58	1.72	0.47 [-0.09, 1.04]	**2.15**	**1.65**	**3.96**	**1.87**	**1.01 [0.42, 1.60]**
Deliberative/ effortful decision	**4.29**	**1.69**	**3.23**	**1.59**	**-0.64 [-1.05, -0.23]**	3.69	1.68	3.27	1.65	-0.25 [-0.80, 0.31]	**4.90**	**1.50**	**3.19**	**1.56**	**-1.10 [-1.70, -0.50]**
Absolute decision/ good memory for perpetrator	4.51	1.01	4.79	1.20	0.26 [-0.14, 0.65]	4.74	1.14	4.53	1.13	-0.18 [-0.74, 0.38]	**4.27**	**0.82**	**5.05**	**1.23**	**0.73 [0.16, 1.31]**
Eliminative/ relative decision	**4.42**	**1.42**	**3.59**	**1.58**	**-0.55 [-0.95, -0.15]**	**3.97**	**1.39**	**3.03**	**1.55**	**-0.63 [-1.20, -0.05]**	4.88	1.33	4.15	1.42	-0.52 [-1.09, 0.05]

Note. g_u = corrected standardized mean difference [and 95% confidence interval]; Bold values represent significant mean differences ($p < .05$) and effect sizes.

Table 7

Means and Standard Deviations of Observer Ratings of Study 2 (Using Videos) for Perceived Correct and Incorrect Identifications

	Total (*N* = 96)					Retrospective Reasoning (*n* = 48)					Think-aloud (*n* = 48)				
	Perceived Incorrect ID		Perceived Correct ID			Perceived Incorrect ID		Perceived Correct ID			Perceived Incorrect ID		Perceived Correct ID		
Observer ratings	*M*	*SD*	*M*	*SD*	g_u [95% CI]	*M*	*SD*	*M*	*SD*	g_u [95% CI]	*M*	*SD*	*M*	*SD*	g_u [95% CI]
Perceptual basis	**3.43**	**0.84**	**4.36**	**1.06**	**0.99 [0.56, 1.42]**	**3.44**	**0.87**	**4.48**	**1.08**	**1.09 [0.45, 1.73]**	**3.42**	**0.83**	**4.27**	**1.06**	**0.90 [0.31, 1.48]**
Confidence	**3.28**	**1.48**	**4.70**	**1.34**	**0.99 [0.56, 1.42]**	**3.46**	**1.47**	**5.07**	**1.29**	**1.11 [0.47, 1.75]**	**3.04**	**1.48**	**4.45**	**1.35**	**0.98 [0.39, 1.57]**
Decision time and difficulty	**4.55**	**1.43**	**3.61**	**1.59**	**-0.62 [-1.04, -0.20]**	**4.23**	**1.29**	**3.07**	**1.35**	**-0.88 [-1.50, -0.25]**	**4.95**	**1.51**	**3.99**	**1.66**	**-0.60 [-1.17, -0.03]**
Persuasiveness of reasons	**3.18**	**1.08**	**4.37**	**1.02**	**1.11 [0.68, 1.55]**	**3.01**	**0.92**	**4.42**	**1.06**	**1.43 [0.77, 2.10]**	**3.40**	**1.24**	**4.34**	**1.02**	**0.81 [0.22, 1.39]**
Automatic decision	2.94	1.74	3.43	1.82	0.28 [-0.13, 0.68]	3.15	1.54	3.33	1.73	0.11 [-0.49, 0.71]	2.67	1.97	3.50	1.92	0.42 [-0.15, 0.98]
Deliberative/ effortful decision	**4.13**	**1.67**	**3.18**	**1.65**	**-0.57 [-0.98, -0.15]**	**3.84**	**1.64**	**2.68**	**1.47**	**-0.72 [-1.33, -0.10]**	**4.49**	**1.67**	**3.51**	**1.72**	**-0.57 [-1.14, 0.00]**
Absolute decision/ good memory for perpetrator	**4.24**	**0.95**	**5.30**	**1.05**	**1.06 [0.62, 1.49]**	**4.33**	**1.03**	**5.30**	**1.08**	**0.91 [0.28, 1.54]**	**4.13**	**0.85**	**5.30**	**1.05**	**1.21 [0.61, 1.82]**
Eliminative/ relative decision	3.76	1.52	4.39	1.54	0.41 [0.00, 0.82]	**3.21**	**1.44**	**4.13**	**1.59**	**0.61 [0.00, 1.22]**	4.46	1.35	4.57	1.51	0.07 [-0.49, 0.63]

Note. g_u = corrected standardized mean difference [and 95% confidence interval]; Bold values represent significant (p < .05) mean differences and effect sizes.

Figures

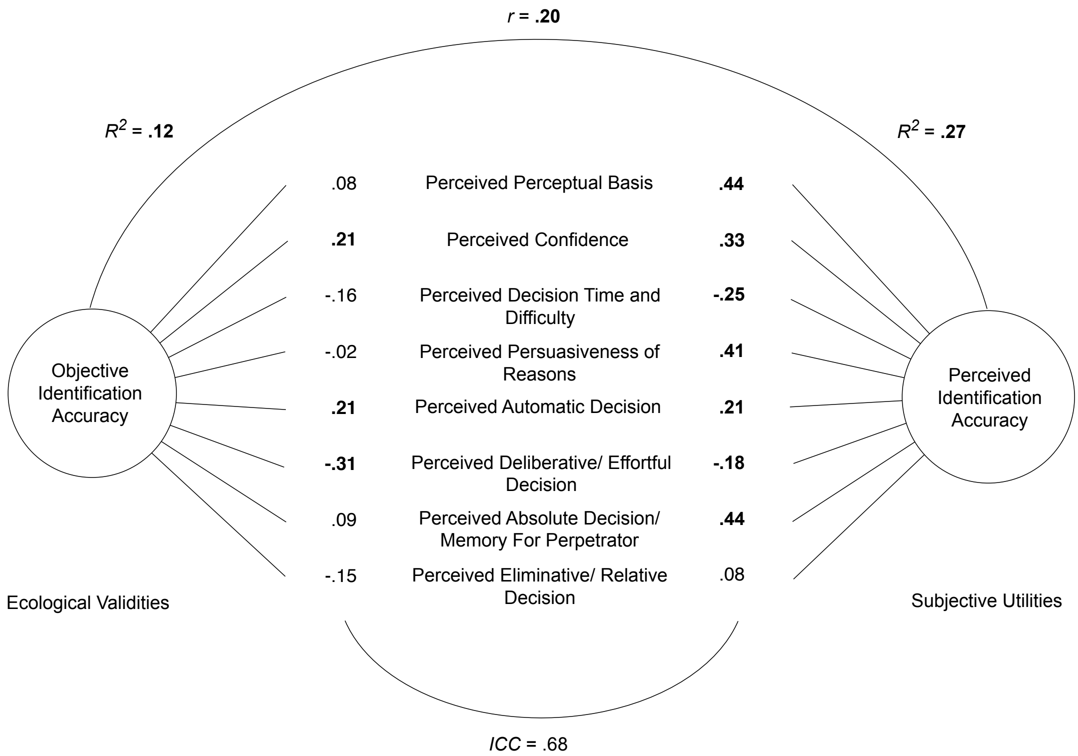

Figure 1. Brunswikian lens model of relationships between ratings of identification decision protocols and objective and perceived identification accuracy (Pearson correlations) in the think-aloud condition of Study 1 (n = 144 judgments). Bold values are significant at $p < .05$.

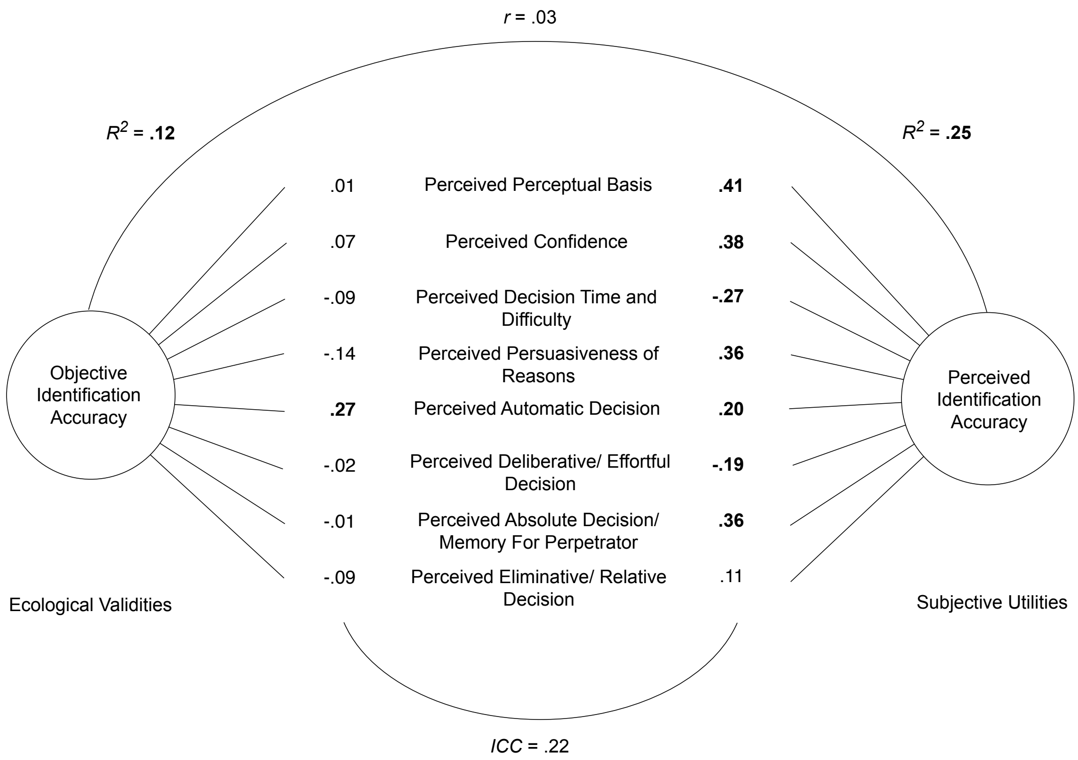

Figure 2. Brunswikian lens model of relationships between ratings of identification decision protocols and objective and perceived identification accuracy (Pearson correlations) in the retrospective reasoning condition of Study 1 (n = 144 judgments). Bold values are significant at $p < .05$.

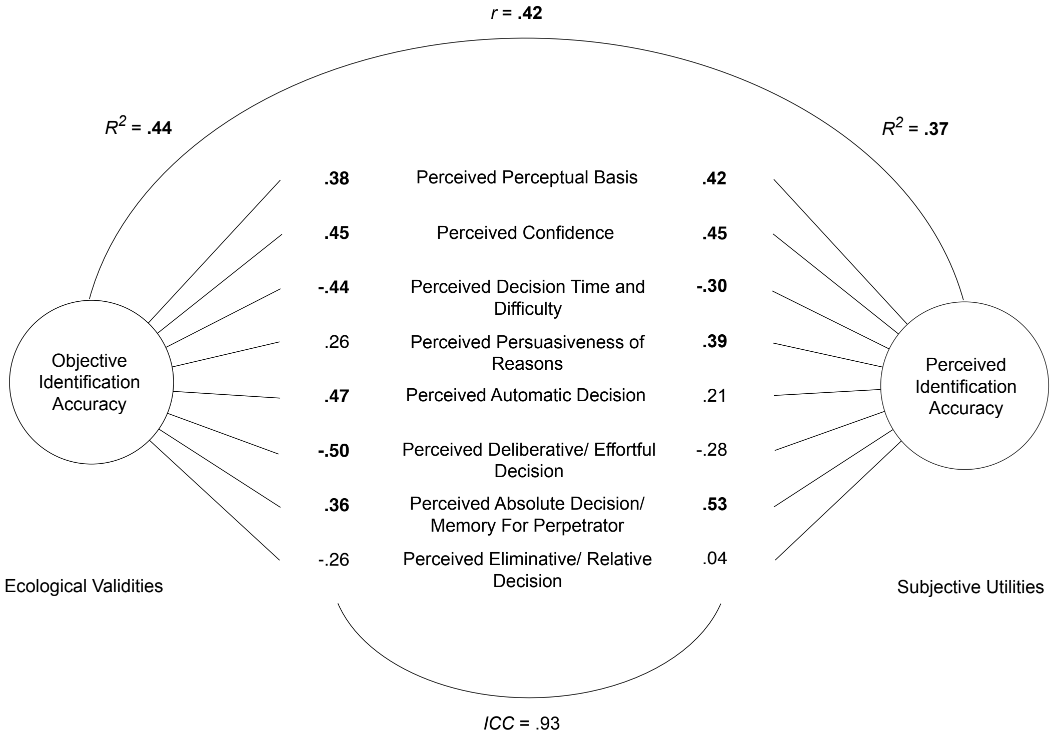

Figure 3. Brunswikian lens model of relationships between ratings of identification decision protocols and objective and perceived identification accuracy (Pearson correlations) in the think-aloud condition of Study 2 (n = 48 judgments). Bold values are significant at $p < .05$.

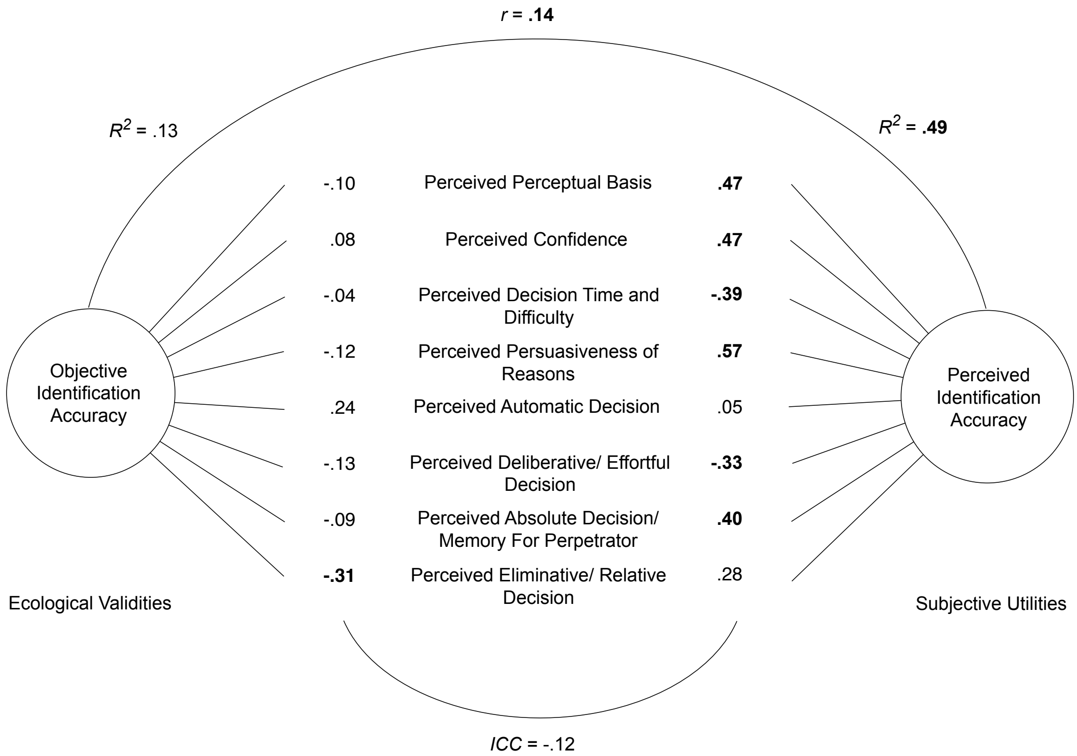

Figure 4. Brunswikian lens model of relationships between ratings of identification decision protocols and objective and perceived identification accuracy (Pearson correlations) in the retrospective reasoning condition of Study 2 (n = 48 judgments). Bold values are significant at $p < .05$.

Appendix

Appendix A

Factor Loadings of the 16 Original Items Measuring Perception and Identification Qualities in Study 1 (N = 288) and Inter-Correlations Between Factors

Items	Factor 1	Factor 2	Factor 3	Factor 4
Good view of perpetrator (1)	**.43**[a]	-.03	.03	-.12
Time perpetrator's face was in	**.59**[a]	-.02	.02	.01
Special facial features visible (3)	**.74**[a]	-.03	-.12	-.15
Attention paid to face (4)	**.73**[a]	.07	.04	.04
Good basis for identification (5)	**.81**[a]	.09	.07	-.05
Difficulty of identification decision	.39	-.37[b]	.23	.13
Response latency Identification	-.04	**.86**[b]	.02	-.04
Post-identification confidence (8)	.18	-.16	**.60**[c]	.07
Willingness to testify in court (9)	-.07	-.01	**.86**[c]	-.09
Willingness to swear oath in	-.04	.08	**.96**[c]	-.01
Decision made with hesitations	-.08	**.26**[b]	-.09	.19
Decision made with deliberations	.13	**.81**[b]	-.01	.00
Accuracy of reasons	-.04	-.08	-.01	**-.96**[d]
Detailedness of reasons	.06	.20	.07	**-.66**[d]
Consistency of reasons	.15	-.08	.03	**-.31**[d]
Convincingness of reasons	.28	-.02	.13	**-.50**[d]
Factor 1	1.00	-.25	.53	-.58
Factor 2		1.00	-.39	.02
Factor 3			1.00	-.38
Factor 4				1.00

Note. The number in brackets refers to the original item of Table 1; Bold values refer to primary loadings for a factor. [a] Perceived perceptual basis (*alpha* = .82); [b] Perceived decision time and difficulty (*alpha* = .71; item 6 was recoded to build the scale); [c] Perceived confidence (*alpha* = .86); [d] Perceived persuasiveness of reasons (*alpha* = .79).

Appendix B

Factor Loadings of 14 Original Items Measuring Decision Processes in Study 1 (N = 288) and Inter-Correlations Between Factors

Items	Factor 1	Factor 2	Factor 3	Factor 4
Matched image in head to	-.16	.05	-.18	**.51**[d]
First looked at all photos (2)	.14	.16[b]	-.08	.29
Compared each face with others	.17	**.31**[b]	.08	.17
Used process of elimination (4)	-.04	**.82**[b]	.03	-.04
Relied on special facial features	.14	.02	-.05	**.53**[d]
Exact match with memory (6)	-.17	.02	.36	**.42**[d]
Just recognized the target, cannot explain why (7)	-.01	.01	**.63**[c]	-.13
Face just popped out (8)	-.09	-.07	**.81**[c]	-.02
First eliminated faces, then chose among rest (9)	-.06	**.85**[b]	-.06	-.06
Faces were all similar (11)	**.61**[a]	.00	-.11	-.13
Photos were confusing (12)	**.52**[a]	.02	.01	-.13
Had to think carefully (13)	**.82**[a]	.01	.03	.29
Much effort necessary (14)	**.80**[a]	.04	-.09	.01
Clear picture in mind (15)	-.15	-.05	.39	**.60**[d]
Factor 1	1.00	.37	-.39	-.05
Factor 2		1.00	-.18	.32
Factor 3			1.00	.06
Factor 4				1.00

Note. The number in brackets refers to the original item of Table 2; Bold values refer to primary loadings for a factor. [a] Perceived deliberative and effortful decision (*alpha* = .81); [b] Perceived eliminative and relative decision (*alpha* = .68); [c] Perceived automatic decision (*alpha* = .70); [d] Perceived absolute decision and a good memory for the perpetrator (*alpha* = .61).

Appendix C

Factor Loadings of the 16 Original Items Measuring Perception and Identification Qualities in Study 2 (N = 96) and Inter-Correlations Between Factors

Items	Factor 1	Factor 2	Factor 3	Factor 4
Good view of perpetrator (1)	.10	.31	-.10	.10[d]
Time perpetrator's face was in	.01	.03	-.05	**.48**[d]
Special facial features visible (3)	-.05	.04	-.02	**.81**[d]
Attention paid to face (4)	.03	-.02	.19	**.77**[d]
Good basis for identification (5)	.00	.05	-.15	**.63**[d]
Difficulty of identification decision	.18	.02	**-.69**[c]	.18
Response latency Identification	.06	-.05	**.90**[c]	.02
Post-identification confidence (8)	**.58**[a]	-.03	-.30	.10
Willingness to testify in court (9)	**.97**[a]	.05	.01	-.06
Willingness to swear oath in	**.86**[a]	-.03	-.06	.06
Decision made with hesitations	-.23	.07	**.69**[c]	.02
Decision made with deliberations	.00	.03	**.93**[c]	.03
Accuracy of reasons	-.05	**1.01**[b]	-.06	-.06
Detailedness of reasons	-.11	**.79**[b]	.15	.05
Consistency of reasons	.22	**.32**[b]	.06	.05
Convincingness of reasons	.08	**.55**[b]	-.11	.16
Factor 1	1.00	.28	-.59	.41
Factor 2		1.00	-.02	.67
Factor 3			1.00	-.21
Factor 4				1.00

Note. The number in brackets refers to the original item of Table 1; Bold values refer to primary loadings for a factor. a Perceived confidence (*alpha* = .91); b Perceived persuasiveness of reasons (*alpha* = .81); c Perceived decision time and difficulty (*alpha* = .92; item 6 was recoded to build the scale); d Perceived perceptual basis (*alpha* = .76).

Appendix D

Factor Loadings of 14 Original Items Measuring Decision Processes in Study 2 (N = 96) and Inter-Correlations Between Factors

Items	Factor 1	Factor 2	Factor 3	Factor 4
Matched image in head to	.17	.12	**.37**[c]	-.01
First looked at all photos (2)	.15	**.47**[b]	.00	.04
Compared each face with others	-.03	**.82**[b]	.06	.04
Used process of elimination (4)	.06	**.61**[b]	.19	.12
Relied on special facial features	-.01	.03	.04[c]	.28
Exact match with memory (6)	-.45	.17	**.48**[c]	.08
Just recognized the target, cannot explain why (7)	-.11	-.12	.24	**-.79**[d]
Face just popped out (8)	-.34	-.28	.50	-.26[d]
First eliminated faces, then chose among rest (9)	-.10	**.70**[b]	-.06	.01
Faces were all similar (11)	**.76**[a]	.12	-.07	-.05
Photos were confusing (12)	**.79**[a]	.00	-.02	-.11
Had to think carefully (13)	**.84**[a]	.09	.11	.09
Much effort necessary (14)	**.98**[a]	-.08	.10	.15
Clear picture in mind (15)	-.22	-.26	**.42**[c]	.36
Factor 1	1.00	.42	-.21	.04
Factor 2		1.00	.05	.36
Factor 3			1.00	.17
Factor 4				1.00

Note. The number in brackets refers to the original item of Table 2; Bold values refer to primary loadings for a factor. [a] Perceived deliberative and effortful decision (*alpha* = .91); [b] Perceived eliminative and relative decision (*alpha* = .76); [c] Perceived absolute decision and good memory for the perpetrator (*alpha* = .44); [d] Perceived automatic decision (*alpha* = .65).

EXPERIMENT 3:

The evaluation of eyewitness identification decisions by indirect measures

Eyewitness testimony including identification decisions is often the only, or the major, source of evidence available in court and consequently serves as an essential basis for later convictions (Garrett, 2011; Wells & Olson, 2003). At trial, judges and/or jurors have to evaluate the accuracy of identification decisions to arrive at their verdicts. Importantly, it is not only a witness's misidentification per se that leads to judicial errors, but rather the erroneous evaluation of these identification decisions that leads to wrongful convictions.

Many studies have shown that observers' ability to discriminate between correct and incorrect identification decision is limited, that is, their judgment accuracy is often no better than chance level (e.g., Beaudry, Lindsay, Leach, Mansour, Bertrand, & Kalmet, 2015; Brigham & Bothwell, 1983; Wells & Lindsay, 1983). Thus, to explain fact finders' judgment accuracy it is necessary to get a deeper insight into observers' judgmental processes.

In these evaluation processes an identification decision serves as a persuasive message that affects fact finders' judgments. Similarly, fact finders' judgmental processes can be explained in line with typical two-process models of persuasion (cf. Leippe, 1994). In these models, for example the Elaboration Likelihood Model (ELM: Petty & Cacioppo, 1986) or the Heuristic Systematic Model (HSM: Chaiken, 1980), there are two ways of evaluating a message. When a message is processed at a systematic or central route, fact finders carefully consider the message's content (e.g., the strength and logic of the arguments) and form opinions based on the detailed elaboration of the presented arguments. A message can also be processed through a peripheral route, which is characterized by a limited elaboration of the message's arguments and the use of heuristic decision rules. Here, fact finders focus on simple and easily accessible cues, for example, the length of the testimony (e.g., number of details), the perceived credibility of the witness (e.g., expert witness or police officer), the witness's appearance, nonverbal behaviors (e.g., maintenance of eye contact, attractiveness) or the message delivery style (e.g., speech style or voice loudness). Based on the ELM, the message content is elaborated centrally when fact finders' motivation (e.g., when the message has a personal relevance) and cognitive ability (e.g., appropriate message complexity, time, attention) to process the information are high. Otherwise persuasion develops trough peripheral processing, which takes place when personal involvement is low or when the message content is too complex and difficult to be carefully processed (cf. Carpenter, 2015; Petty & Cacioppo, 1984, 1986; Petty, Wegener, & Fabrigar, 1997).

Many studies have examined the persuasive impact of content related aspects of an identification decision (i.e., aspects that have been shown to be related with identification accuracy) on observer judgments (i.e., witnessing conditions, confidence, response latency and witnesses' decision processes; e.g., Dunning & Stern, 1994; Kaminski & Sporer, 2016; Neal, Christiansen, Bornstein, & Robicheaux, 2012; for a review see Semmler, Brewer, & Bradfield Douglass, 2012). In contrast, the present study focused on the persuasive impact of *indirect measures* (i.e., measures that are not directly linked to eyewitnesses' memory performance) that are assumed to be processed peripherally. In particular, it is analyzed if observer judgments are related to (1) an overall impression of certain witness traits, (2) perceived witness speech style characteristics and (3) different easily accessible quantitative and qualitative description characteristics. Observers' use of these cues when evaluating the accuracy of an identification decision is contrasted with the validity of these cues to discriminate between correct and incorrect identifications.

Perceived Witness Traits

Based on social perception theories, a first global impression of a person is built within a few milliseconds (for a recent review see Todorov, Olivola, Dotsch, & Mende-Siedlecki, 2015). Inferences about emotions, intentions and personality traits are made automatically and quickly from facial appearance, vocal and behavioral aspects.

Based on implicit personality theories trait inferences are highly correlated and guide social attributions and social judgments (e.g., credibility judgments, sentencing decisions, guilt verdicts). For example, based on a "what is beautiful is good" stereotype, attractive people are also perceived as more trustworthy and honest (cf. Spellman & Tenney, 2010; Zebrowitz, Voinescu, & Collins, 1996), and are rated as more intelligent and competent (Eagly, Ashmore, Makhijani, & Longo, 1991; Zebrowitz, Hall, Murphy, & Rhodes, 2002).

In a legal context it has been shown that witnesses who were perceived as very likeable were also rated as more credible (Garcia & Griffitt, 1978), whereas nervous witnesses were perceived as being less believable in court than witnesses showing only few signs of nervousness

(Bothwell & Jalil, 1992; Pryor & Buchanan, 1984). Porter, ten Brinke, and Gustaw (2010) even showed that participants requested less evidence to convict a defendant who was previously rated as untrustworthy compared to a defendant rated as trustworthy.

In sum, many such peripheral cues seem to affect the evaluation of witness credibility. In the present study we extended this idea by investigating the persuasive impact of certain witness traits on observer judgments of perceived identification accuracy. Therefore, we adapted the Witness Credibility Scale (WCS), which was developed by Brodsky, Griffin, and Cramer (2010). The WCS is designed to measure *expert* witnesses' credibility in court, which is assumed to be associated with the perceived persuasiveness of the expert's report. The scale consists of four highly inter-correlated subscales: confidence, likeability, trustworthiness and knowledge. Higher ratings in these scales were associated with higher credibility ratings (e.g., Neal, Guadagno, Eno, & Brodsky, 2012), as well as with mock-juror judgments of perpetrator blame and sentencing recommendations (Cramer, Titcomb Parrott, Gradner, Stroud, Boccaccini & Griffin, 2014).

In the present study, we adapted the *WCS* as a measure for the perceived accuracy of *non-expert witnesses'* identification decisions, but also included additional items to assess further perceived witness characteristics. Although it is an empirical question, if these trait measures are associated with objective identification accuracy, it is expected that observers will rely on the four dimensions of the *WCS* to evaluate the accuracy of an identification decision. In particular, perceived witness confidence, which has been shown to heavily affect observers' judgments of witness accuracy (for a review, see Boyce, Beaudry, & Lindsay, 2007), is assumed to be used by observers as an indicator of identification accuracy. In line with this assumption Leippe (1994) describes a witness's holistic "confident look" (p. 396) integrating many verbal and nonverbal witness aspects that affect fact finders' judgments of eyewitness accuracy.

Perceived Speech Style

Speech markers in addition to witness traits contribute to a global impression of the witness and also affect the evaluation of an eyewitness's memory report (e.g., Erickson, Lind, Johnson, & O'Barr, 1978). Based on empirical investigations of witnesses' natural speech variations in the courtroom, O'Barr and colleagues (Conley, O'Barr, & Lind, 1978; Erickson et al., 1978; O'Barr, 1982) differentiated between a *powerless* and a *powerful* speech style, which affects the perceived power of a witness's testimony. A powerful style of speaking was more likely to be used by witnesses with high social status and was characterized by the *infrequent* use of "*intensifiers* ("so", "very", "surely" as in "I surely did"), *hedges* ("kinda", "I think", "I guess" etc.), *especially formal grammar* (the use of bookish grammatical forms), *hesitation forms* ("uh", "well", "you know", etc.), *gestures* (e.g., the use of hands and expressions such as "over there" while speaking), *questioning forms* (e.g., the use of rising question intonation in declarative contexts), and *polite forms* ("please", "thank you", etc.)" (Erickson et al., 1978, p. 267).

A powerful speech style has been shown to affect observers' evaluations of the speaker in a favorable way (e.g., they were perceived as more truthful, credible, convincing, intelligent and attractive) as well as to heighten the message's persuasive influence (e.g., Areni & Sparks, 2005; Blankenship & Holtgraves, 2005; Bradac, Hemphill, & Trady, 1981; Clancy & Bull, 2014; Erickson et al., 1978; Holtgraves & Lasky, 1999; O'Barr, 1982;

Ruva & Bryant, 2004). Laboratory studies also provide more direct evidence that hedges and hesitations negatively affect evaluations of witness credibility and guilt (Hosman & Wright, 1987; Schooler, Gerhard, & Loftus, 1986).

It is assumed that a powerful speech style also serves as an indicator of witness confidence (Erickson et al., 1978; Leippe, 1994). Conversely, a powerless speech style, including hedges and hesitations, reflects the speakers' lack of confidence. Focusing on other linguistic features, a speaker's confidence is also expressed in higher *voice loudness* and faster *speech rate* (Scherer, London, & Wolf, 1973). Moreover, a higher speech rate is associated with higher ratings of perceived credibility (Miller, Maruyama, Beaber, & Valone, 1976), whereas people speaking louder than others are perceived as more friendly and logical (Robinson & Zebrowitz McArthur, 1982).

Leippe (1994) concluded that all these linguistic characteristics convey a general impression of expertise, which can be applied to a witness's ability to give an accurate memory report. In line with this assumption, Jules and McQuiston (2013) demonstrated that observers rated witnesses' recollections of details as well as witnesses' identification decisions as more accurate when witnesses used a powerful compared to a powerless speech style.

Another potentially persuasive characteristic of a witness's speech style derives from the interpersonal source monitoring approach (Mitchell & Johnson, 2000; Schooler et al., 1986). In the classical reality monitoring framework (Johnson & Raye, 1981) it is suggested that reports of imagined or internally generated events differ from memory reports of perceived events by an increased inclusion of the speaker's cognitive processes ("it must have been", "lets see", "If I think about it"). These cognitive processes could be reflected in *hedges* and *hesitations* as well as in *long, indirect and evasive answers* (cf. Schooler et al., 1986) that imply the impression of a witness "beating around the bush". Observers are assumed to use the same phenomenal characteristics when they evaluate other people's memories--a judgment process referred to as interpersonal reality monitoring (Mitchell & Johnson, 2000; Sporer, 2004).

In the present study we expect observers' judgments of identification accuracy to vary with differences in the perceived linguistic features. A powerful speech style, high speech rate, a loud voice as well as the infrequent use of long, indirect and evasive answers are assumed to be associated with observers' tendency to evaluate an identification decision as correct. It is an empirical question, whether or not speech style is associated with objective identification accuracy as well.

Perceived Description Qualities

There are many characteristics of eyewitness descriptions that may intuitively affect fact finders' evaluations of an identification decision (e.g., the perceived number of details, the frequent use of "don't know" answers, clear contradictions, or the fit ["congruence"], between the description and the identified person). These characteristics are easily accessible for observers and do not need a careful elaboration or knowledge about the culprit and the crime.

Indeed, it is recommended to use eyewitness descriptions as an indicator of identification accuracy (Neil vs. Biggers, 1972; Sporer & Cutler, 2003). A notable body of research demonstrated small but reliable relationships between person description quality and quantity and identification accuracy (e.g., Meissner, Sporer, & Susa, 2008; Sauerland, Holub, & Sporer, 2008; Sporer, Kaminski, Davids, & McQuiston, 2015). In line with these findings, witness credibility is enhanced when witnesses display a “good memory” (Bell & Loftus, 1988, 1989; cf. Spellman & Tenney, 2010). Bell and Loftus (1988, 1989) showed that ratings of witness credibility increased with the degree of details in an eyewitness’s testimony. A detailed testimony was also associated with a better memory for the culprit’s face and a higher degree of attention paid to the culprit. Similarly, Wells and Leippe (1981) found mock jurors to judge identifications to be more likely to be accurate when a witness’s memory for peripheral details was good; A finding that contradicts the negative correlation found between objective identification accuracy and memory for peripheral details.

According to Petty and Cacioppo (1984), simply increasing the number of arguments should increase the persuasive impact of a message by applying the simple decision rule “the more arguments the better”. Bell and Loftus’s (1988, 1989) as well as Wells and Leippe’s (1981) findings are also in line with the interpersonal source monitoring approach (Mitchell & Johnson, 2000; Schooler et al., 1986). It is assumed that a highly detailed memory report including more sensory details and more contextual information is associated with an externally perceived and well-remembered event, resulting in a higher perceived credibility of the witness (cf. Sporer, 2004).

Besides the degree of details in a testimony, report consistency (Berman & Cutler, 1996; Berman, Narby, & Cutler, 1995; Leippe, Manion and Romanczyk, 1992) and a high congruence between the description of the perpetrator and the identified person (Bradfield & Wells, 2000) have been shown to favorably affect jurors' judgments as well.

Unfortunately, only little is known about the persuasive impact of person and event descriptions. Therefore, we (1) manipulated the content of the to be judged eyewitness statements by presenting identification decisions with or without a description of the perpetrator, the event and the crime scene and (2) created a series of ratings that assessed various aspects of description quantity and quality (see Appendices A and B). We investigated the persuasive influence of these description characteristics on observer judgments of identification accuracy and expected observers to heavily rely on these cues.

Application of the Brunswikian Lens Model

For a detailed analysis of observers' judgmental processes and the persuasive influence of different indirect measures, we applied the Brunswikian lens model (Brunswik, 1956, 1965). The model has been used to explain judgments of medical decision-making (Hammond, Hursch, & Todd, 1964), social perception (Vicaria, Bernieri, & Isaacowitz, 2015) as well as credibility judgments in the context of the detection of deception (Fiedler, 1989; Hartwig & Bond, 2011; Sporer, 1997; Sporer & Küpper, 1995; Sporer, Masip, & Cramer, 2014). Recently, we also applied it to judgments of identification decisions (Kaminski & Sporer, 2016).

The Brunswikian lens model provides a theoretical framework to describe the relationship between the validity of certain perceived cues to predict an objective outcome on the one hand (*ecological validities*) and perceivers' usage of these cues to predict their judgments of this outcome on the other hand (*subjective utilities*). To ensure high judgment accuracy the perceived cues should be strongly related to the objective outcome of interest and perceivers should use, interpret and weight these cues appropriately to arrive at their judgments. To estimate the agreement between cue validities and perceivers usage of these cues a *correspondence* measure is computed. This measure shows whether observers are sensitive to indicators of identification accuracy, that is, if they use valid cues to discriminate between correct and incorrect identification decisions and if they weight them appropriately. Consequently, high correspondences should result in increased judgment accuracy.

In the present study, we contrasted ecological validities of observers' perceptions of the investigated indirect measures (i.e., relationships between these characteristics and objective identification accuracy) with observers' usage of these measures (i.e., relationships between these characteristics and observer judgments) when evaluating the accuracy of an identification decision.

Method

First, we describe Study 1, which was conducted to obtain judgments of identification accuracy and a series of ratings of perceived person and event description qualities. In Study 2, we collected ratings of different witness traits and speech style characteristics. Finally, data of both studies were combined to examine the relationships between observer ratings and observer judgments.

Study 1: Ratings of Description Qualities and Judgment of Identification Accuracy

Study 1 has already been reported by Kaminski and Sporer (2016). However, in the present study we focus on the manipulation of description presence and on observer ratings of perceived description qualities, which have not been analyzed before.

Observer-participants and design.

Ninety-six observer-participants (32 male, 64 female; 75% students, 25% working) between the ages of 18 and 34 years (*Mdn* = 23.0) voluntarily participated in this study. Each observers-participant judged one of 48 videotaped identification statements with or without the presentation of an additional person and event description. A 2 (objective identification accuracy: incorrect choice in a target absent lineup [TA] vs. correct choice in a target present lineup [TP]) x 2 (type of presented decision protocol: retrospective reasoning vs. think-aloud)[1] x 2 (presence of person and event descriptions: not presented vs. presented) factorial between-participants design was used. Thus, there were 12 identification statements (of 6 male and 6 female witnesses) per condition, which internally replicates our experiment 12 times to satisfy stimulus sampling requirements (Wells & Windschitl, 1999).

---

[1] Effects of type of decision protocol were investigated in detail by Kaminski and Sporer (2016). Thus, we do not focus on this manipulation in the present study.

Stimulus eyewitness statements.

To construct the videos a pilot-study was conducted (for a detailed description of this study see Kaminski & Sporer, 2016). In this study witness-participants first watched a short stimulus film and then gave a free description of the perpetrator, the event and the crime scene, which was followed by several non-leading questions. Witness-participants were asked to give the description as precisely as possible but also had the opportunity to give "don't know" answers. Afterwards, they were asked to identify the perpetrator from a target-absent or target-present lineup. The witness-participants were either instructed to think-aloud while making their identification decision (i.e., to say out loud everything, that is, every idea, thought or recollection, that comes to their mind even if it seemed to be irrelevant; cf. Ericsson & Simon, 1993) or to give reasons for their decision afterwards.

The identification statements contained the lineup instructions given to the witness-participants and the witnesses' identification decision. The witnesses' explicit numerical confidence statements, which we had also collected, were cut from the videos and not presented to observers to avoid a halo effect of confidence to influence all other judgments.

Procedure.

All observer-participants were tested individually and were instructed to imagine being a lay judge or juror in court who had to evaluate a videotaped eyewitness identification statement. Additionally, participants were briefly informed about the incident the witness had observed (i.e., a short film of a bicycle theft) and the identification that had taken place approximately one hour after the film. Next, the observer-participants were presented with the videotaped testimony and were instructed to watch it carefully. All participants had the opportunity to watch the video repeatedly during the whole judgment process. Every video was accompanied with details of the witness's age, sex and occupation. The mean age of the videotaped witnesses was 24.65 years (*SD* = 7.08). Moreover, the six-person photo lineup the witnesses saw in the video was printed on high quality photo paper (A4 format) and was handed to the observer-participants.

Observer ratings.

When person and event descriptions were presented (depending on condition), different description qualities were evaluated first (see Appendices A and B). Then, observer-participants were asked to evaluate the identification statements regarding different qualities, which are not considered here (see Kaminski & Sporer, 2016). Finally, observer-participants were asked to give a dichotomous judgment whether they believed that the witness had made a correct identification (*0 = incorrect choice* vs. *1 = correct choice*), which reflects the main dependent variable in the present study.

Study 2: Ratings of Indirect Measures

This study was designed to evaluate the same 48 eyewitness statements exclusively regarding the witness’s speech style and observers’ subjective impression of different witness traits.

Observer-participants and design.

Thirty-two students (8 male, 24 female), aged from 19 to 34 years (*M* = 23.71, *SD* = 3.41) participated in this study to gain course credit. We used a 2 (identification accuracy: incorrect choice [TA] vs. correct choice [TP]) x 2 (type of decision protocol: retrospective reasoning vs. think-aloud) x 2 (person and event descriptions: not presented vs. presented) mixed design with identification accuracy as repeated measures factor. Thus, per condition each participant rated six correct and six incorrect identifications, which were presented in random order. In sum, each of the 48 identification statements was evaluated eight times, with four raters each evaluating the identification statement with and four raters without the presentation of person and event descriptions.

Procedure and materials.

Raters were informed that they would have to watch 12 videos showing different witness-participants of a former study who had testified on a witnessed crime, described the perpetrator (depending on condition) and made an identification decision. They were instructed to evaluate each witness regarding his/her speech style as well as regarding several personality adjectives. Rating order was counterbalanced, that is, half of the participants evaluated the speech style first whereas the other half started with the evaluation of the witness traits. It was emphasized that there are no correct or incorrect answers for these ratings, which should only reflect their subjective impression of the witness statements. The videos were shown and evaluated one after the other in a randomized order. The lineups used were not presented.

Rating variables.

Ten items were constructed to describe the witness's speech style (Appendix C), of which seven items described the powerless-powerful speech style dimension (Erickson et al., 1978; O'Barr, 1982).

To assess the participants' subjective impression of the witness, 37 items were used (Appendix D), including 20 items adapted from the *Witness Credibility Scale* (Brodsky et al., 2010) as well as items adapted from the *Observed Witness Efficacy Scale* (Cramer, DeCoster, Neal, & Brodsky, 2013). All further items were added by us.

For all ratings observers had to evaluate the degree to which each speech characteristic, adjective or behavior was applicable to the witness (*1 = not at all*; *7 = very much*). Items within each category were presented in a newly randomized order for each part to avoid carry-over effects.

Results

Overview

First, we present the results of observer judgments of identification accuracy as a function of the manipulated independent variables (with *N* = 96 judgments as unit of analysis; cf. Study 2 in Kaminski & Sporer, 2016). Subsequently, we focus on observer ratings of witness traits, speech style and description qualities with *N* = 48 stimulus identification decisions as unit of analysis.

To test the Brunswikian lens model (Brunswik, 1956, 1965) relationships between observer ratings and objective identification accuracy (ecological validities) as well as between observer ratings and perceived identification accuracy (subjective utilities) were examined. Two multiple regression analyses were computed to compare the predictions of the objective and the perceived identification accuracy. Further, we computed zero order correlations between ratings and (1) objectively and (2) perceived identification accuracy. To test for agreement between ecological validities and subjective utilities the intra-class correlation coefficient (*ICC*: two-way mixed effect model, average measure, estimated absolute agreement; cf. McGraw & Wong, 1996; Orwin & Vevea, 2009; Shrout & Fleiss, 1979) between the two sets of correlations was computed (using Fisher's Z_r transformations).

The corrected standardized mean difference Hedges g_u, point-biserial correlations (r_{pb}), and odds ratios (*OR*) are reported as effect sizes (cf. Borenstein, 2009; Lipsey & Wilson, 2001). According to Cohen (1988), a small effect is represented by a point-biserial correlation of .10 that equals a standardized mean difference of .20, whereas moderate and large effects are reflected by r_{pb}s of .24 and .37 equaling *d*s of .50 and .80, respectively.

Judgments of Identification Accuracy

Observer judgments.

In sum, 38.5% of identifications were perceived as correct. Objectively correct identifications were judged as correct (52.1%) more often than objectively incorrect identifications (25.0%), $chi^2(1, N = 96) = 7.43$, $p = .006$, *OR* = 3.26, 95% CI [1.37, 7.74]. Judgments did not differ due to the presence of person and event descriptions (identification statement only: 31.3% perceived as correct vs. identification and descriptions: 45.8%, *OR* = 1.86), nor were there differences in judgments due to type of decision protocol (retrospective reasoning: 31.3% perceived as correct vs. think-aloud: 45.8%, *OR* = 1.86).

Judgment accuracy.

Overall, 63.5% of the identification decisions were judged correctly above the chance level of 50%, $t(95) = 2.74$, $p = .007$. Incorrect identifications were judged correctly (75.0%) more often than correct identifications (52.1%), $chi^2(1, N = 96) = 5.44$, $p = .020$, *OR* = 0.36, 95% CI [0.15, 0.86]. Judgment accuracy did not differ due to the other manipulated variables (identification statement only: 64.6% vs. identification and descriptions: 62.5%, *OR* = 0.91; retrospective reasoning: 56.3% vs. think-aloud: 70.8%, *OR* = 1.89).

Observer Ratings of Witness Speech Style and Witness Traits

Inter-rater reliabilities.

Eight different raters evaluated each identification video in Study 2. As measures of inter-rater reliability we computed the average correlation between the eight raters using the Spearman-Brown correction (r_{SB}; cf. Rosenthal, 1995) as well as the intra-class correlation coefficient (*ICC*: one-way random effects model, average measure, estimated consistency; cf. McGraw & Wong, 1996; Orwin & Vevea, 2009; Shrout & Fleiss, 1979).

Across all raters and ratings, there was an average r_{SB} of .65 (mean *ICC* = .63) ranging from .38 (for the rated “use of intensifiers” in the witness’s speech) to .80 (for the rated witness’s “attractiveness”). Results for the subgroups were nearly comparable: for the 10 ratings of the witness speech style (r_{SB} = .68; mean *ICC* = .67) and for the 37 trait ratings (r_{SB} = .64; mean *ICC* = .62). However, due to relatively low inter-reliabilities (r_{SB} < .45) the trait items 31 (“control behavior”), 34 (“admit memory failures”) and 35 (“act natural”) were excluded from further analyses.

As four raters judged the speech style and witness traits solely based on the identification statement, while another four raters made their ratings based on the identification statement presented along with additional person and event descriptions, ratings were averaged within these two conditions, respectively.

Scale construction.

Based on the factor structure underlying the *Witness Credibility Scale* (Brodsky et al., 2010), ratings of witness traits that were highly inter-correlated were combined separately in both description conditions, yielding four scales each measuring perceived witness likeability, trustworthiness, knowledge and confidence (Appendix E; cf. WCS: Brodsky et al., 2010). For each scale high internal consistencies (all Cronbach’s *alpha*s > .89) and satisfactory corrected item-total correlations were observed (Appendix E). Items 26 (“reserved”), 27 (“spontaneous”) and 36 (“thinking hard”) were excluded due to low inter-correlations with the other items.[2]

[2] The seven items describing characteristics of a powerless speech style were not averaged to build a scale, which was due to low internal consistencies (ratings based on identification statements only: Cronbach’s *alpha* = .35; ratings based on identification statements and descriptions: Cronbach’s *alpha* = .61) and low corrected item-total correlations (all *CITC*s < .55).

Perceived witness traits as a function of objective and perceived identification accuracy.

In separate 2 x 2 x 2 mixed model ANOVAs we compared observer ratings with (1) objectively and (2) perceived correct and incorrect identifications as classifying variables. Type of decision protocol (retrospective reasoning vs. think-aloud) was included as between-participants factor. Description presence was treated as a repeated measures factor.

None of the four rating scales differed due to objective identification accuracy (means, standard deviations and effect sizes are displayed in Table 1). However, for ratings of witness confidence there was an interaction between type of decision protocol and objective identification accuracy, $F(1, 44) = 4.75$, $p = .035$, partial $eta^2 = .098$. In the think-aloud condition confidence ratings were higher for correct ($M = 4.69$, $SD = 0.51$) than for incorrect identifications ($M = 4.22$, $SD = 0.45$), $F(1, 44) = 4.23$, $p = .046$, $g_u = 0.96$, 95% CI [0.14, 1.78]. With retrospective reasoning protocols no differences were observed (correct ID: $M = 4.30$, $SD = 0.76$ vs. incorrect ID: $M = 4.54$, $SD = 0.52$), $F(1, 44) = 1.06$, $p = .310$, $g_u = -0.36$, 95% CI [-1.14, 0.42].

Regarding observer judgments, ratings of perceived confidence differed as a function of perceived identification, $F(1, 44) = 9.88$, $p = .003$. The main effect was also marginally significant for ratings of perceived witness knowledge, $F(1, 44) = 3.73$, $p = .060$ (means, standard deviations and effect sizes are displayed in Table 1).

Ratings did not differ due to type of decision protocol and description presence, nor were there any other significant interactions including objective or perceived identification accuracy.

Brunswikian lens model analysis for perceived witness traits.

Two multiple regression analyses with objective and perceived identification accuracy as dependent variables and observer ratings of witness traits as predictors were conducted. Ratings of witness traits showed greater predictive value for the perceived (R^2 = .21, p = .039) than for the objective identification accuracy (R^2 = .02, p = .926). There were no significant relationships between the rating scales and objective identification accuracy, whereas the perceived confidence was significantly correlated with observer judgments (r_{pb} = .43; Figure 1). The two sets of zero-order correlations between ratings and observer judgments and between ratings and objective identification were related by *ICC* = .37, demonstrating differences between ecological validities and subjective utilities.

Ratings of perceived speech style as a function of objective and perceived identification accuracy.

Parallel mixed model ANOVAs were conducted to compare observers' speech style ratings for (1) objectively and (2) perceived correct and incorrect identifications.[3] Means and standard deviations of each

---

[3] For none of the ten speech style ratings were there significant main effects of type of decision protocol, all *F*s < 2.98, all *p*s > .092. Nor were there significant interactions with objective or with perceived identification accuracy (except for the perceived speech rate), *F*s < 3.25, all *p*s > .078. Thus, this factor is not further discussed here.

speech rating for objectively and perceived correct and incorrect identification statements (as well as effect sizes) are displayed in Tables 2 and 3.

Effects of objective identification accuracy.

Ratings of the perceived use of intensifiers and perceived speech rate differed due to objective identification accuracy, both *F*s(1, 44) > 5.03, both *p*s = .030. For three speech ratings (i.e., hedges, hesitations, and long, indirect and evasive answers) there were interactions between description presence and objective identification accuracy, all *F*s > 4.44, *p*s < .041, all partial eta^2 > .092 (see Table 2).

Moreover, for ratings of perceived speech rate there was a significant interaction between objective identification accuracy and type of decision protocol, *F*(1, 44) = 4.15, *p* = .048, partial eta^2 = .086. Only when think-aloud protocols were presented were ratings higher for correct (*M* = 3.75, *SD* = 0.59) than for incorrect identifications (*M* = 3.11, *SD* = 0.67), *F*(1, 44) = 9.18, *p* = .004, g_u = 0.97, 95% CI [0.15, 1.79]. With retrospective reasoning protocols, perceived speech rate was comparable (correct ID: *M* = 3.66, *SD* = 0.38 vs. incorrect ID: *M* = 3.63, *SD* = 0.34), *F*(1, 44) = 0.02, *p* = .882, g_u = 0.08, 95% CI [-0.69, 0.86]. All further main effects and interactions failed to reach significance.

Effects of perceived identification accuracy.

Ratings of the perceived use of hedges and questioning forms differed as a function of observers' judgments of perceived identification accuracy, both *F*s(1, 44) > 4.45, both *p*s < .041 (see Table 3). For perceived speech rate there also was a significant interaction between perceived identification accuracy and type of decision protocol, $F(1, 44) = 4.05$, $p = .050$, partial $eta^2 = .084$. When think-aloud protocols were used ratings were higher for perceived correct ($M = 3.59$, $SD = 0.71$) than for perceived incorrect identifications ($M = 3.05$, $SD = 0.54$), $F(1, 44) = 4.97$, $p = .031$, $g_u = 0.78$, 95% CI [-0.10, 1.65]. With retrospective reasoning protocols perceived speech rate did not differ (ID judged as correct: $M = 3.59$, $SD = 0.31$ vs. ID judged as incorrect: $M = 3.70$, $SD = 0.40$), $F(1, 44) = 0.29$, $p = .592$, $g_u = -0.32$, 95% CI [-1.10, 0.46]. There were no significant interactions between perceived identification accuracy and description presence.

Main effects of description presence.

Six speech ratings (i.e., intensifiers, questioning forms, gestures, long, indirect and evasive answers, speech rate and voice loudness) differed due to description presence, all Fs > 6.69, ps < .019, g_us ranging from -0.41 to -0.75, showing higher ratings when identifications were presented without descriptions compared to ratings when descriptions were added.

Brunswikian lens model analyses for speech style ratings.

Due to the reported interactions between objective identification accuracy and description presence, regression analyses were conducted separately for evaluations based solely on the identification statements and for evaluations that were based on both, the identification statement and the descriptions.

In the *identification-only* condition, ratings showed somewhat greater predictive value for the objective ($R^2 = .38$, $p = .038$) than for the perceived identification accuracy ($R^2 = .26$, $p = .285$). Two of ten zero-order correlations demonstrated significant relationships with objective identification accuracy (i.e., long, indirect and evasive answers and speech rate), whereas observers primarily used two speech style characteristics (i.e., hedges and questioning forms) to make their judgments (Figure 2). In sum, the two sets of zero-order correlations between ratings and objective and perceived identification accuracy were moderately related by *ICC* = .64, demonstrating correspondence between ecological validities and subjective utilities, but also discrepancies.

When identification statements were *presented along with descriptions* predictive value of observer ratings was R^2 = .32 (*p* = .113) for the objective identification accuracy compared to R^2 = .18 (*p* = .604) for the perceived identification accuracy. Two of ten zero-order correlations demonstrated significant relationships with objective identification accuracy (i.e., intensifiers and gestures; Figure 3). However, none of the perceived speech characteristics was significantly correlated with observers' judgments of identification accuracy. The two sets of correlation coefficients were not related (*ICC* = -.11). Thus, essential non-correspondence in the validity of the perceived cues (ecological validities) and the use of these cues (subjective utilities) is reflected.

Observer Ratings of Description Qualities

First, description ratings were averaged to yield general rating scales of quality of person (Cronbachs's *alpha* = .85) and of event and scene descriptions (Cronbachs's *alpha* = .91). Due to low (and even negative) corrected item-total correlations (*CITC*s < .40) person description items 12 and 13 (see Appendix A) as well as event and crime scene description items 3, 8, 17 and 18 (see Appendix B) were excluded. To build the scales ratings of description difficulty and response latency were reversed.

Two 2 x 2 analyses of variance were conducted to compare description ratings for (1) objectively and (2) perceived correct and incorrect identifications with type of decision protocol (retrospective reasoning vs. think-aloud) as between-participant factor. There were no significant effects of objective identification accuracy, type of identification decision protocol, nor any significant interactions, all *F*s < 2.08, all *p*s > .157. In contrast, for both rating scales, there was a significant main effect of perceived identification accuracy. When identifications were judged as correct person description quality was rated higher ($M = 4.64$, $SD = 0.81$) compared to identifications judged as incorrect ($M = 3.86$, $SD = 0.78$), $F(1, 44) = 10.91$, $p = .002$, $g_u = 0.98$, 95% CI [0.39, 1.57], $r_{pb} = 0.45$. Similarly, event and scene description quality was rated higher when identifications were perceived as correct ($M = 5.42$, $SD = 0.77$) than incorrect ($M = 4.59$, $SD = 0.95$), $F(1, 44) = 10.58$, $p = .002$, $g_u = 0.93$, 95% CI [0.35, 1.52], $r_{pb} = 0.44$.

Brunswikian lens model analyses for perceived description characteristics.

To test which of the different description characteristics observers used to make their judgments, Brunswikian lens model analyses were conducted with the single description ratings.

Ratings of *person description qualities* showed greater predictive value for perceived ($R^2 = .42$, $p = .074$) than for objective identification accuracy ($R^2 = .17$, $p = .898$). There were significant positive correlations between perceived identification accuracy and perceived description accuracy, description congruence, witness confidence and the witness's picture in mind (Figure 4). However, none of the ratings was a valid indicator of objective identification accuracy. The two sets of zero-order correlations were moderately related by *ICC* = .64, reflecting moderate correspondence between ecological validities and subjective utilities.

Ratings of *event and scene description qualities* also showed greater predictive value for perceived (R^2 = .55, p = .055) than for objective identification accuracy (R^2 = .26, p = .889). For seven ratings there were significant correlations with observer judgments; however none of the variables was related to objective identification accuracy (Figure 5). Correspondence between ecological validities and subjective utilities was low (*ICC* = .37).

Discussion

The present study aimed at investigating the persuasive impact of different indirect measures on observers' judgments of perceived identification accuracy. Applying the Brunswikian lens model, it was examined (1) if these measures were related to objective identification accuracy (*ecological validities*) and (2) if observers used these measures to make their judgments (*subjective utilities*) and how they weighted them.

Perceived Witness Traits: Subjective Utilities vs. Ecological Validities

Observer ratings mainly supported the factor structure of the *Witness Credibility Scale*, which was developed to measure expert witness credibility (Brodsky et al., 2010). As expected, observers heavily relied on the *perceived witness confidence* to judge the accuracy of an identification decision (r_{pb} = .43). Note that witnesses' explicit confidence ratings had been edited out from the videotapes so they would not influence observers' judgments. Moreover, observers also tended to use *perceived* *trustworthiness* (r_{pb} = .21) and *perceived witness knowledge* (r_{pb} = .24) to make their judgments. Although the effect sizes were moderate in size, these effects failed to reach significance.

In contrast, none of the perceived traits discriminated between objectively correct and incorrect identifications (all r_{pb} < .11; Figure 1). However, when think-aloud protocols were presented, confidence ratings were higher for correct than for incorrect identifications (r_{pb} = .46). In contrast, with retrospective reasoning statements confidence was no indicator of identification accuracy.

A comparison of subjective utilities and ecological validities revealed that observers tended to give an inappropriate high weight to cues that were not indicative of objective identification accuracy at all (except for confidence ratings in the think-aloud condition).

Former studies using the *WCS* with expert witnesses found positive relationships between the four subscales and attributed expert witness credibility (e.g., Cramer et al., 2014; Neal, Guadagno, et al., 2012). However, applied to the evaluation of non-expert witnesses' identification decisions only the confidence scale was related to observers' judgments. Several other studies also demonstrated that observers heavily relied on witness confidence (e.g., Beaudry et al., 2015; Kaminski & Sporer, 2016; Wells, Lindsay, & Ferguson, 1979). However, as the witnesses' explicit numerical confidence statement was edited out from the videotapes in the present study, observers had to form a general impression of the witness's confidence based on other peripheral cues like nonverbal and paraverbal witness behaviors. Apparently, this impression formation is facilitated when witnesses were instructed to think-aloud while identifying the target, which is supposed to make the witness's cognitive processes (e.g., doubts or a faded memory) more visible for observers (cf. Kaminski & Sporer, 2016). Thus, only when think-aloud protocols were used this subjective impression of confidence discriminated between correct and incorrect identifications. Similarly, Kaminski and Sporer (2016) demonstrated that asking observers to explicitly judge the witness's confidence regarding his/her identification decision resulted in appropriate confidence ratings that were moderately related to objective identification accuracy only when think-aloud protocols were used, but not with retrospective reasoning protocols.

However, as the use of think-aloud protocols is no common police practice, it is recommended to ask witnesses to explicitly state their post-identification confidence to protect fact finders from relying on invalid confidence inferences. Many studies have shown that self-reported witness confidence at the time of the original identification is strongly positively related to choosers' identification accuracy (mean weighted $r = .37$: Sporer, Penrod, Read, & Cutler, 1995). However, this is the case only when confidence is asked for directly after the identification decision which should be videotaped for further investigations (National Academy of Sciences, 2014; Sporer, 1992, 1993, 1994; Wells, Small, Penrod, Malpass, Fulero, & Brimacombe, 1998; Wixted, Mickes, Clark, Gronlund, & Roediger, 2015).

In sum, persuasive influences of witness traits on observer judgments of identification accuracy were only partially supported by the present results.

Perceived Speech Style Characteristics

Subjective utilities.

O’Barr and colleagues distinguished between a powerless and powerful speech style dimension that has been shown to be associated with observers’ evaluations of the speaker and his or her communicated message (e.g., Erickson et al., 1978; Hosman et al., 2002; Hosman & Wright, 1987; Jules & McQuiston, 2013). In the present study, relationships between rated power of speech and objective as well as perceived identification accuracy depended on the type of material presented (identification statement only vs. identification plus descriptions). Only in the identification-only condition were *hedges* (r_{pb} = -.30) and *questioning forms* (r_{pb} = -.41) associated with observer judgments. In contrast, when additional descriptions were presented none of the ten speech style ratings differed due to perceived identification accuracy. Moreover, perceived *speech rate* was positively related with observer judgments when think-aloud protocols were used (r_{pb} = .36), but not with retrospective reasoning statements (cf. findings by Miller et al., 1976).

Consequently, the persuasive impact of special speech style characteristics was not independent of the presented message content and the number of presented arguments. Ratings served as persuasive cues only when testimony was short, and thus, only a few arguments were available to be processed. Similarly, Hosman, Huebner and Siltanen (2002) demonstrated that a powerless speech style produced more negative thoughts about a message than a powerful speech style only when argument strength was weak. When arguments were strong, speech power was not influential. Moreover, Sparks and Areni (2008) demonstrated that observers relied on language markers especially when their ability to carefully process the message was low. They concluded that when observers are unable to process message content, they tend to elaborate speaker qualities instead. Additionally, we assume that the descriptions themselves may have served as a strong persuasive cue that affected observer judgments in a way that speech style effects probably were masked. In sum, a persuasive effect of certain linguistic features may emerge only when identification protocols are presented in isolation.

Ecological validities.

Considering ecological validities, there were different speech characteristics in each description condition that discriminated between correct and incorrect identifications. In the identification-only condition, correct identifications were moderately associated with fewer *long, indirect and evasive answers* (r_{pb} = -.30), a higher *speech rate* (r_{pb} = .34) and with less perceived *hesitations* (r_{pb} = -.28; however, the effect was only marginally significant) than incorrect ones.

These results are in line with the reality monitoring approach (Johnson & Raye, 1981; Mitchell & Johnson, 2000; Sporer, 2004) which suggests that correct memories are accompanied by less cognitive processes compared to incorrect or generated memories, which might be also reflected in a witness's speech style (cf. Schooler et al., 1986). Several studies have shown that correct identifications are often made faster (e.g., Sauerland & Sporer, 2007, 2009; Sporer, 1992, 1993, 1994) and more automatically, that is, with fewer deliberations and with less eliminative thought processes than incorrect ones (e.g., Dunning & Stern, 1994; Wells, 1984).

In contrast, when identifications were presented with descriptions, *intensifiers* (r_{pb} = .32) and *gestures* (r_{pb} = .34) were significantly related with objective identification accuracy. Interestingly, intensifiers and gestures originally were supposed to reflect a powerless speech style used by uncertain and incompetent, low status witnesses (Erickson et al., 1978). However, our results suggest an opposite interpretation. Here, accurate witnesses used intensifiers and gestures, presumably to emphasize the strength of their memory.

Did observers use speech style characteristics appropriately?

Although results differed for the two description conditions, in both conditions speech style ratings explained more variance of objective than of perceived identification accuracy. Thus, observers did not use speech style cues as extensively to make their judgments as suggested by the observed ecological validities (Figure 2 and 3). In both conditions observers were not sensitive to objectively valid indicators of identification accuracy underestimating their discriminative value.

In the *identification-only condition*, there was an overall moderate correlation (*ICC* = .64) between ecological validities and subjective utilities. Although observers underestimated cue discriminability of valid indicators (i.e., speech rate and long, indirect and evasive answers) most of the correlations between ratings and objective and perceived identification accuracy were in same direction. In particular, observers realized that some cues were not postdictive of identification accuracy (i.e., intensifiers, gestures, polite forms and voice loudness), and thus they did not weight them, respectively.

In the *identification-plus-description condition* there was an overall non-correspondence between ecological validities and subjective utilities (*ICC* = -.11). Observers were not sensitive to discriminating speech characteristics (i.e., intensifiers and gestures), and thus did not use them to make their judgments. Although, they tended to use some other speech characteristics in line with our theoretical assumptions (e.g., hedges, hesitations, questioning forms, long, indirect evasive answers and voice loudness), the observed relationships between these ratings and objective identification accuracy were in the opposite direction or not present at all.

In sum, observers' utilization of speech style characteristics was more appropriate when identification statements were presented without additional descriptions. When descriptions, and hence more persuasive content arguments were present, speech style was not used to judge identification accuracy at all. In line with these findings, Hosman et al. (2002) concluded that the effects of speech style are assumed to be generally small, particularly in comparison to the persuasive effects of argument strength. These findings are in line with Reinhard and colleagues' dual process theory accounts of detecting deception (e.g., Reinhard, Sporer, Scharmach, & Marksteiner, 2011).

Perceived Description Qualities

As only less is known about ratings of person, event and crime scene descriptions and their persuasive impact on observers' judgments of identification accuracy, different ratings of description quality were collected in the present study. It was expected that observers would heavily rely on easily accessible qualities like the perceived description detailedness, the perceived cognitive effort to make a description (cf. interpersonal reality monitoring approach: Mitchell & Johnson, 2000) as well as on different aspects reflecting witness confidence.

In line with these expectations, description qualities were rated much more favorable when identifications were perceived as correct compared to identifications perceived as incorrect (person descriptions: r_{pb} = 0.45; event and crime scene descriptions: r_{pb} = 0.44).

However, there was no effect of description presence on observers' judgment accuracy suggesting that perceived description qualities did not seem to help observers to correctly discriminate between correct and incorrect identifications.

Ratings of person descriptions.

Although ratings of person description qualities explained 42% of the variance in observers' judgments, only 17% of the variance in objective identification accuracy was explained (Figure 4). Observers heavily relied on the perceived *accuracy* (r_{pb} = .49), the perceived *congruence* (r_{pb} = .46; cf. results by Bradfield & Wells, 2000), the perceived witness *confidence* (r_{pb} = .34) as well as the perceived quality of the *witness's picture in mind* (r_{pb} = .33). Although not significant, all further description characteristics (except the ratings of the witness's admission of memory gaps) were moderately weighted by observers.

In contrast, none of the ratings were valid indicators of objective identification accuracy. Nevertheless, there were nonsignificant tendencies for perceived congruence (r_{pb} = .23), perceived description difficulty (r_{pb} = -.19) as well as perceived witness's response latency (r_{pb} = -.20) to discriminate between correct and incorrect identifications. High correspondences between ecological validities and subjective utilities were observed for the perceived description difficulty and the witness's response latency only. Consequently, correspondence between ecological validates and subjective utilities was still moderate (*ICC* = .64).

In sum, observers overestimated the discriminative value of the investigated person description qualities. An attributed good memory for the perpetrator combined with a high witness confidence was intuitively associated with good identification performance (cf. Bell & Loftus, 1988, 1989; Wells & Leippe, 1981). However, objective relationships between these ratings and identification accuracy could not be shown in the present study. Thus, the US Supreme court's recommendation (Neil vs. Biggers, 1972) to use the perceived person description's accuracy to evaluate an identification decision is only partially supported by the present results.

Ratings of event and crime scene descriptions.

Ratings explained 55% of the variance in observers' judgments, whereas only 26% of the variance in objective identification accuracy was explained (Figure 5). Correspondence between ecological validities and subjective validities (*ICC* = .37) was low. Although none of the individual ratings significantly discriminated between correct and incorrect identifications (all r_{pb}s between -.16 and .19), observers overestimated the discriminative value of perceived event and crime scene description qualities, and used several description qualities to make their judgments.

Observers used cues as indicators of a good memory, that is, perceived *usefulness* (r_{pb} = .44) and perceived *accuracy* of the event (r_{pb} = .33) and crime scene descriptions (r_{pb} = .42), to infer a good identification performance. Especially description *consistency* (r_{pb} = .41) and the *reconstructability* of the event (r_{pb} = .43), that is, the absence of contradictions and a logical structure, seemed to be persuasive (cf. Berman & Cutler, 1996; Berman et al., 1995).

As expected, the perceived number of details was higher for identifications judged as correct than for those judged as incorrect (r_{pb} = .46). However, this held true only for *relevant event details*. Scene details and irrelevant details were not persuasive, which contradicts the results by Wells and Leippe (1981) and Bell and Loftus (1988, 1989). Thus, the simple decision rule "the more arguments the better" which was derived from the ELM (Petty & Cacioppo, 1986), as well as suggestions from the interpersonal reality monitoring approach (Mitchell & Johnson, 2000; Sporer, 2004) found only partial support. Instead, more content related description aspects (e.g., consistency) seem to affect observers' judgments more heavily

Moreover, short *response latencies* were associated with higher perceived identification accuracy (r_{pb} = -.38), which is in line with the interpersonal reality monitoring approach (Mitchell & Johnson, 2000; Sporer, in press). Witnesses responding fast convey the impression that they do not need to search their memory for the correct answer or have to engage in extensive deliberations or other reflective processes. Thus, they seem confident and credible regarding their own memory. Similarly, Kaminski and Sporer (2016) showed that observers used ratings of identification response time and identification difficulty as well as perceived presence of deliberations as indicators of incorrect identifications.

Conclusions and Practical Implications

To sum up, observer judgments of perceived accuracy of an identification decision indeed seem to be affected by certain peripheral cues. However, observers tend to overestimate the discriminative value of these cues (except for speech style ratings). Observers' subjective impression of a witness demonstrating a good memory for the event and the target as well as high confidence seem to be most powerful to affect observers' judgments. Thus, observers do not only consider central, content related aspects of an eyewitness's testimony, but also use some less valid heuristic decision rules to make their judgments. But only few of these peripheral measures were valid indicators of objective identification accuracy which resulted in low correspondences between ecological validities and subjective utilities. Taken together, this pattern explains fact finders' low judgment accuracies observed in former studies (e.g., Wells & Lindsay, 1983).

Based on the present and other studies, we recommend fact finders to concentrate on those aspects of a testimony that have been shown to be reliably related to identification accuracy (e.g., post-identification confidence, response latency, automatic and effortless decision processes: e.g., Brewer & Wells, 2006; Kaminski & Sporer, 2016; Sauerland & Sporer, 2007, 2009; ; Weber, Brewer, Wells, Semmler, & Keast, 2004; Wells, 1984), and to ignore invalid aspects.

Based on the present study, one fruitful peripheral indicator of identification accuracy is the subjective impression of the witness's confidence that is formed by observers when think-aloud protocols were used. Moreover, certain witness's speech style characteristics discriminated between correct and incorrect identifications. However, the effects of speech style depended on the presence of person and event descriptions. The perceived congruence between the target description and the person identified should be considered as well, provided that future research corroborates this relationship.

References

Areni, C. S., & Sparks, J. R. (2005). Language power and persuasion. *Psychology & Marketing, 22,* 507-525. doi: 10.1002/mar.20071

Beaudry, J. L., Lindsay, R. C. L., Leach, A.-M., Mansour, J. K., Bertrand, M. I., & Kalmet, N. (2015). The effect of evidence type, identification accuracy, line-up presentation, and line-up administration on observers' perceptions of eyewitnesses. *Legal and Criminological Psychology, 20,* 343-364. doi: 10.1111/lcrp.12030

Bell, B. E., & Loftus, E. F. (1988). Degree of detail of eyewitness testimony and mock juror judgments. *Journal of Applied Social Psychology, 18,* 1171-1192. doi: 10.1111/j.1559-1816.1988.tb01200.x

Bell, B. E., & Loftus, E. F. (1989). Trivial persuasion in the courtroom: The power of (a few) minor details. *Journal of Personality and Social Psychology, 56,* 669-679. doi: 10.1037/0022-3514.56.5.669

Berman, G. L., & Cutler, B. L. (1996). Effects of inconsistencies in eyewitness testimony on mock-juror decision making. *Journal of Applied Psychology, 81,* 170-177. doi: 10.1037/0021-9010.81.2.170

Berman, G. L., Narby, D. J., & Cutler, B. L. (1995). Effects of inconsistent eyewitness statements on mock-jurors evaluations of the eyewitness, perceptions of defendant culpability and verdicts. *Law and Human Behavior, 19,* 79-88. doi: 10.1007/BF01499074

Blankenship, K. L., & Holtgraves, T. (2005). The role of different markers of linguistic powerlessness in persuasion. *Journal of Language and Social Psychology, 24,* 3-24. doi: 10.1177/0261927X04273034

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. 221-235). New York: Russell Sage Foundation.

Bothwell, R. K., & Jalil, M. (1992). The credibility of nervous witnesses. *Journal of Social Behavior and Personality, 7*(4), 581-586.

Boyce, M. A., Beaudry, J. L., & Lindsay, R. C. L. (2007). Belief of eyewitness identification evidence. In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Handbook of eyewitness psychology: Memory for people* (Vol. 2, pp. 25-45). Mahwah, NJ: Lawrence Erlbaum and Associates.

Bradfield, A. L., & Wells, G. L. (2000). The perceived validity of eyewitness identification testimony: A test of the five Biggers criteria. *Law and Human Behavior, 24,* 581-594. doi: 10.1023/a:1005523129437

Brewer, N., & Wells, G. L. (2006). The confidence-accuracy relationship in eyewitness identification: Effects of lineup instructions, foil similarity, and target-absent base rates. *Journal of Experimental Psychology: Applied, 12,* 11-30. doi: 10.1037/1076-898x.12.1.11

Brigham, J. C., & Bothwell, R. K. (1983). The ability of prospective jurors to estimate the accuracy of eyewitness identifications. *Law and Human Behavior, 7,* 19-29. doi: 10.1007/BF01045284

Brodsky, S. L., Griffin, M. P., & Cramer, R. J. (2010). The witness credibility scale: An outcome measure for expert witness research. *Behavioral Sciences & the Law, 28,* 892-907. doi: 10.1002/bsl.917

Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. Berkeley: University of California Press.

Brunswik, E. (1965). *The conceptual framework of psychology*. Chicago, London: The University of Chicago Press.

Carpenter, C. J. (2015). A meta-analysis of the ELM's argument quality x processing type predictions. *Human Communication Research, 41,* 501-534. doi: 10.1111/hcre.12054

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology, 39,* 752-766. doi: 10.1037/0022-3514.39.5.752

Clancy, D., & Bull, R. (2015). The effect on mock-juror decision-making of power-of-speech within eyewitness testimony and types of scientific evidence. *Psychiatry, Psychology and Law, 22,* 425-435. doi: 10.1080/13218719.2014.960029

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

Conley, J. M., O' Barr, W. M., & Lind, E. A. (1978). Power of language: Presentational style in the courtroom. *Duke Law Journal*, 1375-1399. Retrieved from http://scholarship.law.duke.edu/cgi/viewcontent.cgi?article=2686&context=dlj

Cramer, R. J., DeCoster, J., Neal, T. M. S., & Brodsky, S. L. (2013). The observed witness efficacy scale: A measure of effective testimony skills. *Journal of Applied Social Psychology, 43,* 1696-1703. doi: 10.1111/jasp.12124

Cramer, R. J., Titcomb Parrott, C., Gardner, B. B., Stroud, C. H., Boccaccini, M. T., & Griffin, M. P. (2014). An exploratory study of meta-factors of expert witness persuasion. *Journal of Individual Differences, 35,* 1-11. doi: 10.1027/1614-0001/a000123

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

Eagly, A. H., Ashmore, R. D., Makhijani, M. G., & Longo, L. C. (1991). What is beautiful is good, but...: A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin, 110,* 109-128. doi: 10.1037/0033-2909.110.1.109

Erickson, B., Lind, E. A., Johnson, B. C., & O' Barr, W. M. (1978). Speech style and impression-formation in a court setting: Effects of powerful and powerless speech. *Journal of Experimental Social Psychology, 14,* 266-279. doi: 10.1016/0022-1031(78)90015-X

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. Massachusetts: A Bradford Book.

Fiedler, K. (1989). Lügendetektion aus alltagspsychologischer Sicht [Lie detection based on commonsense]. *Psychologische Rundschau, 40*(3), 127-140.

Garcia, L. T., & Griffitt, W. (1978). Impact of testimonial evidence as a function of witness characteristics. *Bulletin of the Psychonomic Society, 11,* 37-40. doi: 10.3758/BF03336759

Garrett, B. L. (2011). *Convicting the innocent: Where criminal prosecutions go wrong.* Cambridge, Massachusetts: Harvard University Press.

Hammond, K. R., Hursch, C. J., & Todd, F. J. (1964). Analyzing the components of clinical inference. *Psychological Review, 71,* 438-456. doi: 10.1037/h0040736

Hartwig, M., & Bond, C. F. (2011). Why do lie-catchers fail? A lens model meta-analysis of human lie judgments. *Psychological Bulletin, 137,* 643-659. doi: 10.1037/a0023589

Holtgraves, T., & Lasky, B. (1999). Linguistic power and persuasion. *Journal of Language and Social Psychology, 18,* 196-205. doi: 10.1177/0261927X99018002004

Hosman, L. A., Huebner, T. M., & Siltanen, S. A. (2002). The impact of power-of-speech style, argument strength, and need for cognition on impression formation, cognitive responses, and persuasion. *Journal of Language and Social Psychology, 21,* 361-379. doi: 10.1177/026192702237954

Hosman, L. A., & Wright, J. W. (1987). The effects of hedges and hesitations on impression formation in a simulated courtroom context. *Western Journal of Speech Communication, 51,* 173-188. doi: 10.1080/10570318709374263

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88,* 67-85. doi: 10.1037//0033-295X.88.1.67

Jules, S. J., & McQuiston, D. E. (2013). Speech style and occupational status affect assessments of eyewitness testimony. *Journal of Applied Social Psychology, 43,* 741-748. doi: 10.1111/jasp.12002

Kaminski, K. S., & Sporer, S. L. (2016). *Discriminating between correct and incorrect eyewitness identifications: The use of appropriate cues.* Manuscript submitted for publication.

Leippe, M. R. (1994). The appraisal of eyewitness testimony. In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 385-418). Cambridge, England: Cambridge University Press.

Leippe, M. R., Manion, A. P., & Romanczyk, A. (1992). Eyewitness persuasion: How and how well do fact-finders judge the accuracy of adults and children's memory reports. *Journal of Personality and Social Psychology, 63,* 181-197. doi: 10.1037//0022-3514.63.2.181

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

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1,* 30-46. doi: 10.1037/1082-989X.1.4.390

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

Miller, N., Maruyama, G., Beaber, R. J., & Valone, K. (1976). Speed of speech and persuasion. *Journal of Personality and Social Psychology, 34,* 615-624. doi: 10.1037/0022-3514.34.4.615

Mitchell, K. J., & Johnson, M. K. (2000). Source monitoring: Attributing mental experience. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 179-195). Oxford, UK: Oxford University Press.

National Academy of Sciences (2014). *Identifying the culprit: Assessing eyewitness identification*. Washington, D. C.: The National Academies Press.

Neal, T. M. S., Christiansen, A., Bornstein, B. H., & Robicheaux, T. R. (2012). The effects of mock jurors' beliefs about eyewitness performance on trial judgments. *Psychology, Crime & Law, 18,* 49-64. doi: 10.1080/1068316X.2011.587815

Neal, T. M. S., Guadagno, R. E., Eno, C. A., & Brodsky, S. L. (2012). Warmth and competence on the witness stand: Implications for the credibility of male and female expert witnesses. *Journal of the American Academy of Psychiatry and the Law, 40*(4), 488-497.

Neil vs. Biggers, 409 U.S. 188 (1972).

O'Barr, W. M. (1982). *Linguistic evidence: Language, power, and strategy in the courtroom*. San Diego: Academic Press.

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

Petty, R. E., & Cacioppo, J. T. (1984). The effect of involvement on responses to argument quantity and quality: Central and peripheral routes to persuasion. *Journal of Personality and Social Psychology, 46,* 69-81. doi: 10.1037//0022-3514.46.1.69

Petty, R. E., & Cacioppo, J. T. (1986). *Communication and persuasion: Central and peripheral routes to attitude change*. New York: Springer.

Petty, R. E., Wegener, D. T., & Fabrigar, L. R. (1997). Attitudes and attitude change. *Annual Review of Psychology, 48,* 609-647. doi: 10.1146/annurev.psych.48.1.609

Porter, S., ten Brinke, L., & Gustaw, C. (2010). Dangerous decisions: The impact of first impressions of trustworthiness on the evaluation of legal evidence and defendant culpability. *Psychology Crime & Law, 16,* 477-491. doi: 10.1080/10683160902926141

Pryor, B., & Buchanan, R. W. (1984). The effects of a defendant's demeanor on juror perceptions of credibility and guilt. *Journal of Communication, 34,* 92-99. doi: 10.1111/j.1460-2466.1984.tb02176.x

Reinhard, M.-A., Sporer, S. L., Scharmach, M., & Marksteiner, T. (2011). Listening, not watching: Situational familiarity and the ability to detect deception. *Journal of Personality and Social Psychology, 101,* 467–484. doi:10.1037/a0023726

Robinson, J., & Zebrowitz McArthur, L. (1982). Impact of salient vocal qualities on causal attribution for a speaker's behavior. *Journal of Personality and Social Psychology, 43,* 236-247. doi: 10.1037/0022-3514.43.2.236

Rosenthal, R. (1995). Methodology. In A. Tesser (Ed.), *Advanced social psychology* (pp. 17-49). New York: McGraw-Hill.

Ruva, C. L., & Bryant, J. B. (2004). The impact of age, speech style, and question form on perceptions of witness credibility and trial outcome. *Journal of Applied Social Psychology, 34,* 1919-1944. doi: 10.1111/j.1559-1816.2004.tb02593.x

Sauerland, M., Holub, F. E., & Sporer, S. L. (2008). Person descriptions and person identifications: Verbal overshadowing or recognition criterion shift? *European Journal of Cognitive Psychology, 20,* 497-560. doi: 10.1080/09541440701728417

Sauerland, M., & Sporer, S. L. (2007). Post-decision confidence, decision time, and self-reported decision processes as postdictors of identification accuracy. *Psychology, Crime & Law, 13,* 611-625. doi: 10.1080/10683160701264561

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

Scherer, K. R., London, H., & Wolf, J. J. (1973). The voice of confidence: Paralinguistic cues and audience evaluation. *Journal of Research in Personality, 7,* 31-44. doi: 10.1016/0092-6566(73)90030-5

Schooler, J. W., Gerhard, D., & Loftus, E. F. (1986). Qualities of the unreal. *Journal of Experimental Psychology: Learning Memory and Cognition, 12,* 171-181. doi: 10.1037//0278-7393.12.2.171

Semmler, C., Brewer, N., & Bradfield Douglass, A. (2012). Jurors believe eyewitnesses. In B. L. Cutler (Ed.), *Conviction of the innocent: Lessons from psychological research* (pp. 185-209). Washington, DC: American psychological Association.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86,* 420-428. doi: 10.1037//0033-2909.86.2.420

Sparks, J. R., & Areni, C. S. (2008). Style versus substance: Multiple roles of language power in persuasion. *Journal of Applied Social Psychology, 38,* 37-60. doi: 10.1111/j.1559-1816.2008.00295.x

Spellman, B. A., & Tenney, E. R. (2010). Credible testimony in and out of court. *Psychonomic Bulletin & Review, 17,* 168-173. doi: 10.3758/PBR.17.2.168

Sporer, S. L. (1992). Post-dicting eyewitness accuracy: Confidence, decision-times and person descriptions of choosers and non-choosers. *European Journal of Social Psychology, 22,* 157-180. doi: 10.1002/ejsp.2420220205

Sporer, S. L. (1993). Eyewitness identification accuracy, confidence, and decision times in simultaneous and sequential lineups. *Journal of Applied Psychology, 78,* 22-33. doi: 10.1037/0021-9010.78.1.22

Sporer, S. L. (1994). Decision times and eyewitness identification accuracy in simultaneous and sequential lineups. In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 300-327). Cambridge: Cambridge University Press.

Sporer, S. L. (1997). The less travelled road to truth: Verbal cues in deception detection in accounts of fabricated and self-experienced events. *Applied Cognitive Psychology, 11,* 373-397. doi: 10.1002/(SICI)1099-0720(199710)11:5<373::AID-ACP461>3.0.CO;2-0

Sporer, S. L. (2004). Reality monitoring and the detection of deception. In P. A. Granhag & L. A. Stroemwall (Eds.), *Deception detection in forensic contexts* (pp. 64-102). Cambridge: Cambridge University Press.

Sporer, S. L. (in press). Deception and cognitive load: Expanding our horizon with a working memory model. *Frontiers in Psychology*.

Sporer, S. L., & Cutler, B. L. (2003). Identification evidence in Germany and the United States. In P. J. VanKoppen & S. D. Penrod (Eds.), *Adversarial versus inquisitorial justice: Psychological perspectives on criminal justice systems* (pp. 191-208). New York: Springer.

Sporer, S. L., Kaminski, K. S., Davids, M. C., & McQuiston, D. (2015). The verbal facilitation effect: Re-reading person descriptions as a system variable to improve identification performance. *Memory*. doi: 10.1080/09658211.2015.1106561

Sporer, S. L., & Küpper, B. (1995). Realitätsüberwachung und die Beurteilung des Wahrheitsgehaltes von Erzählungen: Eine experimentelle Studie [Reality monitoring and the judgment of the truthfulness of accounts: An experimental study]. *Zeitschrift für Sozialpsychologie, 26*(3), 173-193.

Sporer, S. L., Masip, J., & Cramer, M. (2014). Guidance to detect deception with the Aberdeen Report Judgment Scales: Are verbal content cues useful to detect false accusations? *American Journal of Psychology, 127*(1), 43-61.

Sporer, S. L., Penrod, S., Read, D., & Cutler, B. (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

Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: Determinants, consequences, accuracy, and functional significance. *Annual Review of Psychology, 66,* 519-545. doi: 10.1146/annurev-psych-113011-143831

Vicaria, I. M., Bernieri, F. J., & Isaacowitz, D. M. (2015). Perceptions of rapport across the life span: Gaze patterns and judgment accuracy. *Psychology and Aging, 30,* 396-406. doi: 10.1037/pag0000019

Weber, N., Brewer, N., Wells, G. L., Semmler, C., & Keast, A. (2004). Eyewitness identification accuracy and response latency: The unruly 10-12-second rule. *Journal of Experimental Psychology: Applied, 10,* 139-147. doi: 10.1037/1076-898x.10.3.139

Wells, G. L. (1984). The psychology of lineup identifications. *Journal of Applied Social Psychology, 14,* 89-103. doi: 10.1111/j.1559-1816.1984.tb02223.x

Wells, G. L., & Leippe, M. R. (1981). How do triers of fact infer the accuracy of eyewitness identifications: Using memory for peripheral detail can be misleading. *Journal of Applied Psychology, 66,* 682-687. doi: 10.1037/0021-9010.66.6.682

Wells, G. L., & Lindsay, R. C. L. (1983). How do people infer the accuracy of eyewitness memory? Studies of performance and a metamemory analysis. In S. Lloyd-Bostock & B. R. Clifford (Eds.), *Evaluating witness evidence: Recent psychological research and new perspectives* (pp. 41-55). Chichester: John Wiley & Sons.

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

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

Wells, G. L., Small, M., Penrod, S., Malpass, R. S., Fulero, S. M., & Brimacombe, C. A. E. (1998). Eyewitness identification procedures: Recommendations for lineups and photospreads. *Law and Human Behavior, 22,* 603-647. doi: 10.1023/a:1025750605807

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

Wixted, J. T., Mickes, L., Clark, S. E., Gronlund, S. D., & Roediger, H. L. (2015). Initial eyewitness confidence reliably predicts eyewitness identification accuracy. *American Psychologist, 70,* 515-526. doi: 10.1037/a0039510

Zebrowitz, L. A., Hall, J. A., Murphy, N. A., & Rhodes, G. (2002). Looking smart and looking good: Facial cues to intelligence and their origins. *Personality and Social Psychology Bulletin, 28,* 238-249. doi: 10.1177/0146167202282009

Zebrowitz, L. A., Voinescu, L., & Collins, M. A. (1996). “Wide-eyed” and “crooked-faced”: Determinants of perceived and real honesty across the life span. *Personality and Social Psychology Bulletin, 22,* 1258-1269. doi: 10.1177/01461672962212006

Tables

Table 1

Means and Standard Deviations of Ratings of Witness Traits for (1) Objectively and (2) Perceived Correct and Incorrect Identifications

	Incorrect ID		Correct ID		
Observer ratings	*M*	*SD*	*M*	*SD*	g_u [95% CI]
	Objective Identification Accuracy				
Likeability	4.79	0.47	4.76	0.63	-0.05 [-0.61, 0.50]
Trustworthiness	4.98	0.56	5.00	0.69	0.03 [-0.53, 0.58]
Knowledge	4.44	0.59	4.46	0.70	0.03 [-0.53, 0.59]
Confidence	4.38	0.50	4.50	0.66	0.20 [-0.36, 0.76]
	Perceived Identification Accuracy				
Likeability	4.72	0.56	4.81	0.55	0.17 [-0.41, 0.74]
Trustworthiness	4.82	0.66	5.09	0.59	0.43 [-0.16, 1.01]
Knowledge	4.25	0.65	4.57	0.61	0.50 [-0.09, 1.08]
Confidence	**4.12**	**0.57**	**4.63**	**0.52**	**0.94 [0.33, 1.54]**

Note. g_u = corrected standardized mean difference [and 95% confidence interval]; Bold values represent significant mean differences ($p < .05$) and effect sizes.

Table 2

Means and Standard Deviations of Observer Speech Style Ratings for Objectively Correct and Incorrect Identifications (N = 48) Presented Without (Identification Only) and With Descriptions (Identification and Descriptions)

	Total					Identification only					Identification and descriptions				
	Incorrect ID		Correct ID			Incorrect ID		Correct ID			Incorrect ID		Correct ID		
Observer ratings	*M*	*SD*	*M*	*SD*	g_u [95% CI]	*M*	*SD*	*M*	*SD*	g_u [95% CI]	*M*	*SD*	*M*	*SD*	g_u [95% CI]
Hedges	4.04	0.85	4.07	1.08	0.04 [-0.52, 0.59]	**4.45**	**0.99**	**3.96**	**1.35**	-0.41 [-0.97, 0.15]	3.63	1.05	4.19	1.01	0.54 [-0.03, 1.10]
Intensifiers	**2.78**	**0.47**	**3.13**	**0.60**	**0.63 [0.05, 1.20]**	3.08	0.74	3.39	0.95	0.35 [-0.21, 0.91]	**2.48**	**0.53**	**2.86**	**0.64**	**0.64 [0.07, 1.21]**
Hesitations	4.17	1.00	3.79	1.12	-0.35 [-0.91, 0.21]	4.35	1.31	3.60	1.33	-0.56 [-1.13, 0.01]	3.99	1.15	3.98	1.07	-0.01 [-0.57, 0.55]
Questioning forms	3.11	0.72	3.04	0.88	-0.08 [-0.64, 0.47]	3.40	1.06	3.14	1.10	-0.24 [-0.80, 0.32]	2.82	0.69	2.95	0.83	0.16 [-0.40, 0.72]
Gestures	2.82	0.60	3.24	1.18	0.44 [-0.12, 1.01]	3.28	0.82	3.42	1.34	0.12 [-0.44, 0.68]	**2.35**	**0.60**	**3.06**	**1.30**	**0.69 [0.11, 1.26]**
Formal grammar	4.49	0.82	4.48	0.92	-0.01 [-0.57, 0.54]	4.48	1.08	4.35	1.09	-0.11 [-0.67, 0.44]	4.50	0.91	4.60	0.96	0.11 [-0.45, 0.67]
Polite forms	5.01	0.60	4.85	0.79	-0.23 [-0.79, 0.33]	5.08	0.72	4.75	1.01	-0.37 [-0.94, 0.19]	4.94	0.76	4.95	0.68	0.01 [-0.54, 0.57]
Long, indirect, evasive answers	3.32	0.82	3.08	1.05	-0.24 [-0.80, 0.31]	**3.82**	**1.02**	**3.16**	**1.17**	**-0.60 [-1.17, -0.03]**	2.81	0.89	3.01	1.12	0.19 [-0.37, 0.75]
Speech rate	**3.37**	**0.58**	**3.70**	**0.49**	**0.61 [0.04, 1.18]**	**3.44**	**0.70**	**3.91**	**0.62**	**0.70 [0.12, 1.27]**	3.30	0.63	3.50	0.63	0.31 [-0.25, 0.87]
Voice Loudness	4.19	0.86	4.01	0.61	-0.25 [-0.81, 0.31]	4.38	0.93	4.18	0.86	-0.22 [-0.78, 0.34]	4.01	0.94	3.83	0.56	-0.23 [-0.78, 0.33]

Note. g_u = corrected standardized mean difference [and 95% confidence interval]; Bold values represent significant mean differences ($p < .05$) and effect sizes.

Table 3

Means and Standard Deviations of Observer Speech Style Ratings for Perceived Correct and Incorrect Identifications (N = 48) Presented Without (Identification Only) and With Descriptions (Identification and Descriptions)

	Total					Identification only					Identification and descriptions				
	Perceived Incorrect ID		Perceived Correct ID			Perceived Incorrect ID		Perceived Correct ID			Perceived Incorrect ID		Perceived Correct ID		
Observer ratings	*M*	*SD*	*M*	*SD*	g_u [95% CI]	*M*	*SD*	*M*	*SD*	g_u [95% CI]	*M*	*SD*	*M*	*SD*	g_u [95% CI]
Hedges	**4.44**	**0.99**	**3.82**	**0.88**	**-0.66 [-1.25, -0.07]**	**4.65**	**1.22**	**3.93**	**1.11**	**-0.61 [-1.20, -0.03]**	4.24	1.04	3.71	1.04	-0.50 [-1.08, 0.08]
Intensifiers	2.90	0.53	2.99	0.59	0.16 [-0.42, 0.73]	3.19	0.89	3.26	0.85	0.07 [-0.50, 0.65]	2.60	0.61	2.72	0.63	0.19 [-0.39, 0.77]
Hesitations	4.22	0.77	3.84	1.21	-0.35 [-0.93, 0.22]	4.21	1.21	3.84	1.45	-0.26 [-0.84, 0.31]	4.24	0.89	3.83	1.20	-0.36 [-0.94, 0.22]
Questioning forms	**3.47**	**0.90**	**2.84**	**0.62**	**-0.83 [-1.43, -0.23]**	**3.83**	**1.12**	**2.93**	**0.90**	**-0.90 [-1.50, -0.30]**	3.10	0.94	2.76	0.60	-0.45 [-1.03, 0.13]
Gestures	2.97	1.04	3.06	0.92	0.09 [-0.48, 0.67]	3.32	1.06	3.37	1.14	0.04 [-0.53, 0.62]	2.63	1.17	2.76	1.01	0.12 [-0.45, 0.70]
Formal grammar	4.21	0.90	4.65	0.82	0.51 [-0.07, 1.10]	4.13	1.13	4.59	1.02	0.43 [-0.15, 1.01]	4.29	0.89	4.71	0.93	0.45 [-0.13, 1.03]
Polite forms	4.79	0.79	5.01	0.63	0.31 [-0.27, 0.89]	4.85	0.94	4.96	0.86	0.12 [-0.45, 0.70]	4.74	0.81	5.07	0.63	0.46 [-0.12, 1.05]
Long, indirect, evasive answers	3.42	0.85	3.07	0.99	-0.36 [-0.94, 0.22]	3.69	1.03	3.37	1.19	-0.28 [-0.86, 0.29]	3.14	0.97	2.78	1.02	-0.36 [-0.94, 0.22]
Speech rate	3.45	0.55	3.59	0.56	0.24 [-0.34, 0.82]	3.61	0.80	3.71	0.64	0.14 [-0.44, 0.71]	3.29	0.56	3.47	0.67	0.27 [-0.30, 0.85]
Voice Loudness	4.08	0.77	4.11	0.74	0.05 [-0.53, 0.62]	4.29	0.90	4.27	0.90	-0.03 [-0.60, 0.55]	3.86	0.81	3.96	0.76	0.12 [-0.45, 0.70]

Note. g_u = corrected standardized mean difference [and 95% confidence interval]; Bold values represent significant mean differences ($p < .05$) and effect sizes.

Figures

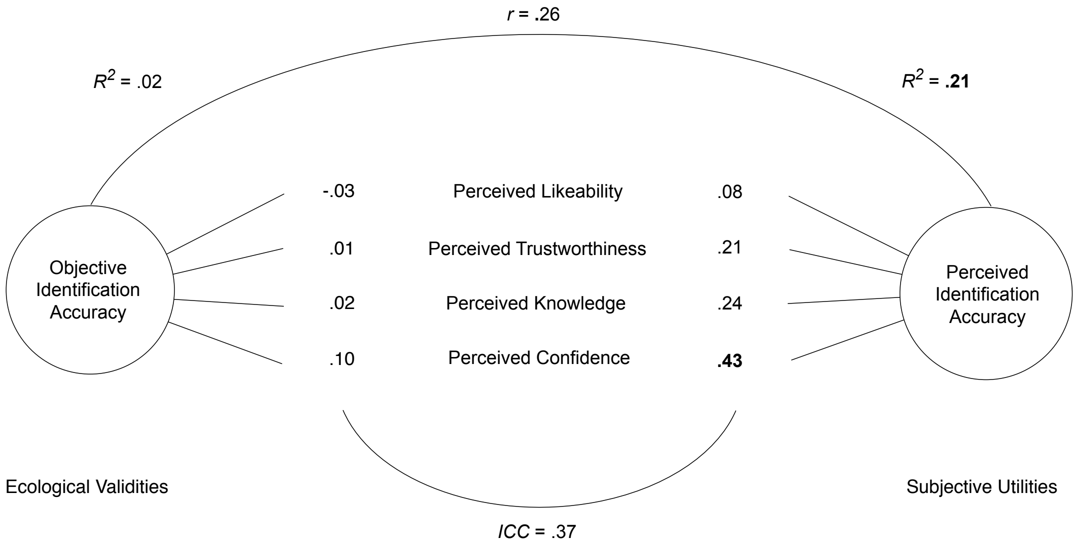

Figure 1. Brunswikian lens model of relationships between ratings of witness traits and objective and perceived identification accuracy (Pearson correlations). Bold values are significant at p < .05.

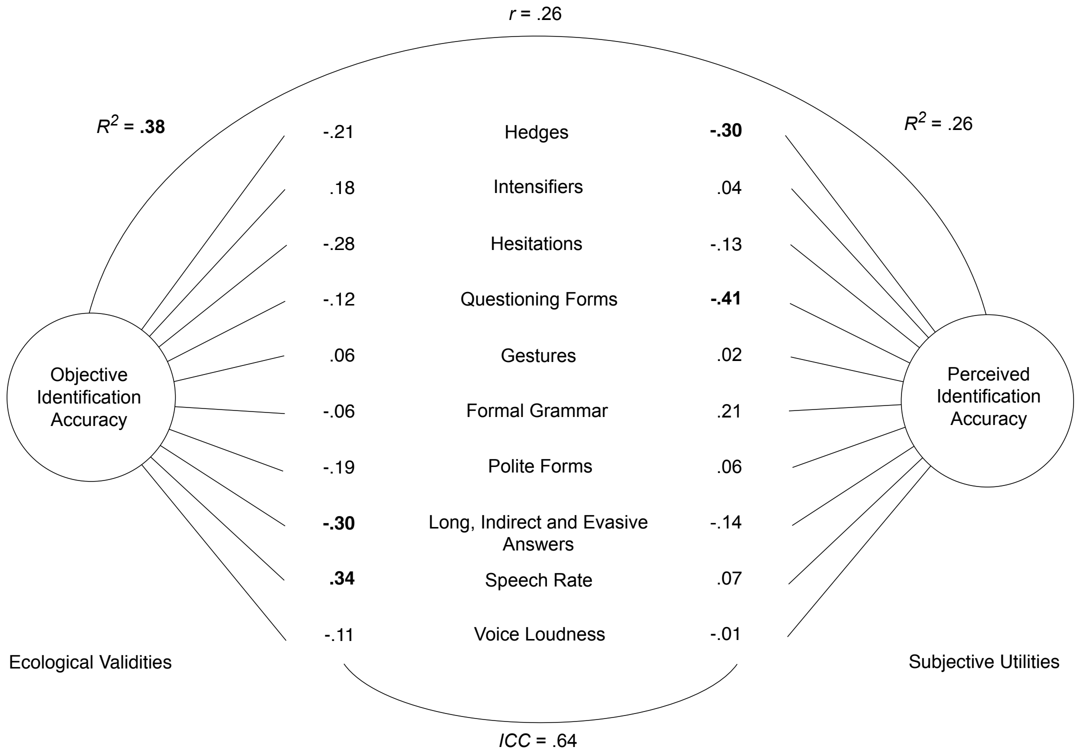

Figure 2. Brunswikian lens model of relationships between ratings of speech style and objective and perceived identification accuracy (Pearson correlations) in the identification-only condition. Bold values are significant at $p < .05$.

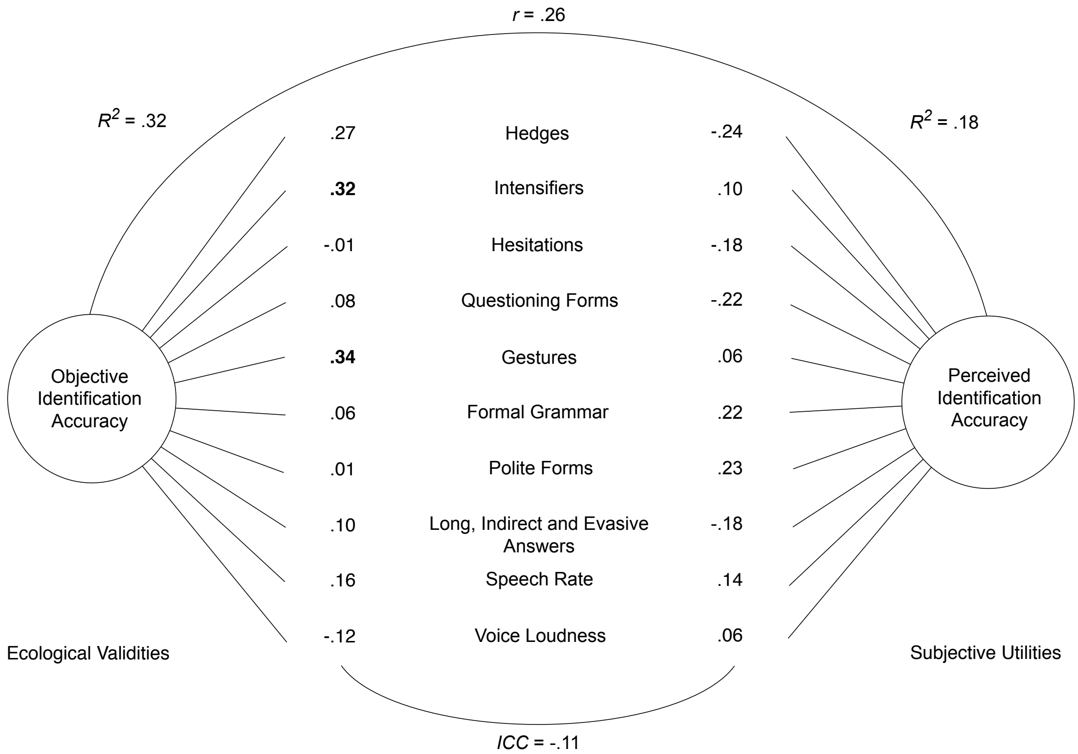

Figure 3. Brunswikian lens model of relationships between ratings of speech style and objective and perceived identification accuracy (Pearson correlations) in the identification plus description condition. Bold values are significant at $p < .05$.

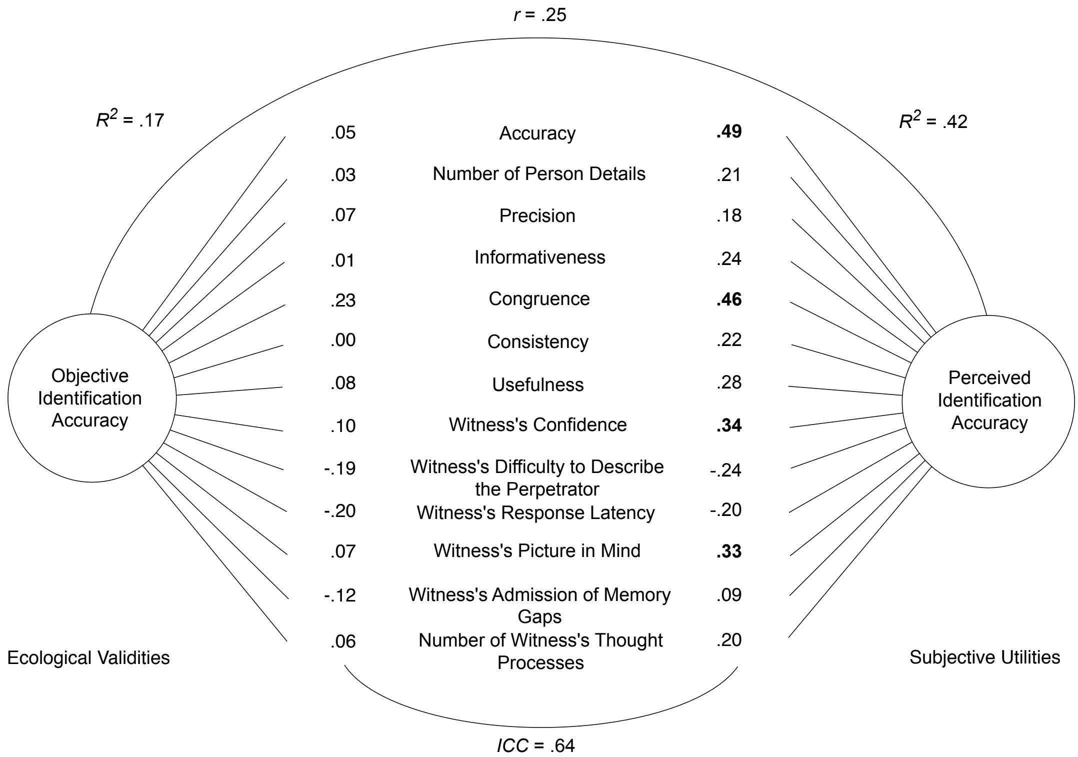

Figure 4. Brunswikian lens model of relationships between person description ratings and objective and perceived identification accuracy (Pearson correlations). Bold values are significant at p < .05.

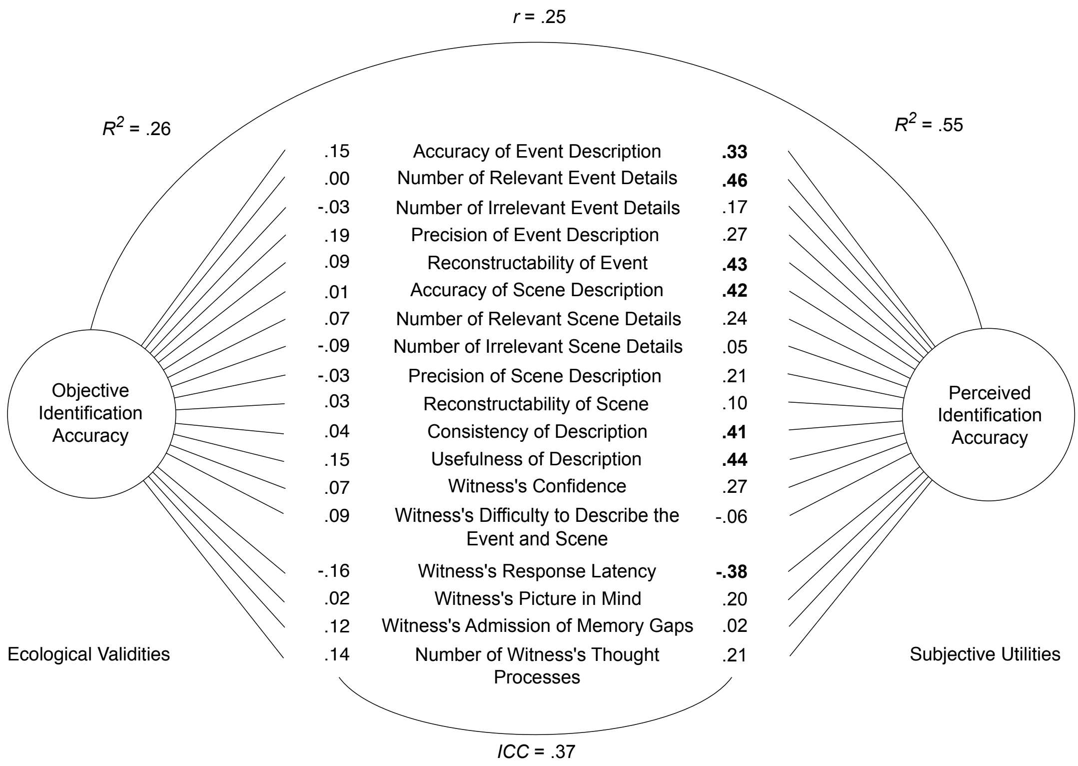

Figure 5. Brunswikian lens model of relationships between event and crime scene description ratings and objective and perceived identification accuracy (Pearson correlations). Bold values are significant at $p < .05$.

Appendix

Appendix A

Ratings of Person Description Qualities (Only in the Description Condition)

	Rating variables	Anchors
(1)	How *accurate* is the description?	*1 = Not accurate at all; 7 = Absolutely accurate*
(2)	How *many details* are described?	*1 = Very few; 7 = Many*
(3)	How *precise* are the details described?	*1 = Not precise at all; 7 = Very precise*
(4)	How i*nformative* is the description (i.e., is it possible to find a person in a crowd based on this description?)	*1 = Not informative at all; 7 = Very informative*
(5)	Did the description *match the appearance of the person identified* in the lineup?	*1 = No match at all; 7 = Perfect match*
(6)	Was the description *consistent*?	*1 = Not consistent at all; 7 = Very consistent*
(7)	How *useful* do you think the description is for the criminal investigation?	*1 = Not useful at all; 7 = Very useful*
(8)	How *confident* was the witness regarding his/her memory?	*1 = Not confident at all; 7 = Very confident*
(9)	How *difficult* was it for the witness to describe the perpetrator?	*1 = Very easy; 7 = Very difficult*
(10)	How *much time* did the witness need to answer the questions?	*1 = A very short time; 7 = A very long time*
(11)	Did the witness have a *good picture of the perpetrator in mind*?	*1 = No picture at all; 7 = Very good picture*
(12)	How often did the witness *admit memory failures*?	*1 = Never; 7 = Very often*
(13)	How often did the witness *explain his/her thoughts* while describing the perpetrator?	*1 = Never; 7 = Very often*

Appendix B

Ratings of Event and Crime Scene Description Qualities (Only in the Description Condition)

	Rating variables	Anchors
	Ratings of event descriptions only	
(1)	How *accurate* is the event description?	*1 = Not accurate at all; 7 = Absolutely accurate*
(2)	How *many relevant event details* are described?	*1 = Very few; 7 = Many*
(3)	How *many irrelevant event details* are described?	*1 = Very few; 7 = Many*
(4)	How *precise* are the event details described?	*1 = Not precise at all; 7 = Very precise*
(5)	To what extent is it possible to *reconstruct the event*?	*1 = Very difficult; 7 = Very easy*
	Ratings of crime scene descriptions only	
(6)	How *accurate* is the crime scene description?	*1 = Not accurate at all; 7 = absolutely accurate*
(7)	How *many relevant crime scene details* are described?	*1 = Very few; 7 = Many*
(8)	How *many irrelevant crime scene details* are described?	*1 = Very few; 7 = Many*
(9)	How *precise* are the crime scene details described?	*1 = Not precise at all; 7 = Very precise*
(10)	To what extent is it possible to *reconstruct the crime scene*?	*1 = Very difficult; 7 = Very easy*
	Ratings of both event and crime scene descriptions	
(11)	Were the descriptions *consistent*?	*1 = Not consistent at all; 7 = Very consistent*
(12)	How *useful* do you think the descriptions are for the criminal investigation?	*1 = Not useful at all; 7 = Very useful*
(13)	How *confident* was the witness regarding his/her memory?	*1 = Not confident at all; 7 = Very confident*
(14)	How *difficult* was it for the witness to describe the event and the crime scene?	*1 = Very easy; 7 = Very difficult*
(15)	How *much time* did the witness need to answer the questions?	*1 = A very short time; 7 = A very long time*
(16)	Did the witness have a *good picture of the event and the crime scene in mind*?	*1 = No picture at all; 7 = Very good picture*
(17)	How often did the witness *admit memory failures*?	*1 = Never; 7 = Very often*
(18)	How often did the witness *explain his/her thoughts* while describing the event and the crime scene?	*1 = Never; 7 = Very often*

Appendix C

Ratings of the Witness's Speech Style

	Rating variables
(1)[a]	The witness uses lots of *hedges*, that is, weakening words that do not allow a clear statement (e.g., "I think", "perhaps", "kind of").
(2)[a]	The witness uses lots of *intensifiers* to emphasize what he/she said (e.g., "very", "definitely", "totally").
(3)[a]	The witness uses lots of *hesitations* with or without filler words (e.g., "ok", "uh", "well").
(4)[a]	The witness uses lots of *questioning forms* (i.e., the use of a rising, questioning intonation).
(5)[a]	The witness uses lots of *gestures* to emphasize what he/she said.
(6)[a]	The witness uses an especially *formal grammar* (e.g., complete and grammatically correct sentences).
(7)[a]	The witness uses lots of *polite forms.*
(8)	The witness gives lots of *long, evasive and indirect answers*.
(9)	The witness speaks very *fast*.
(10)	The witness speaks very *loud*.

Note. [a] Items were adapted from Erickson et al. (1978) and O'Barr (1982); Endpoints of the scales were from *1 = Not at all* to *7 = Very much.*

Appendix D

Ratings of Perceived Witness Traits

Rating variables					
The witness appears to be…					
(1)[a]	Friendly	(11)[a]	Confident	(21)	Credible
(2)[a]	Respectful	(12)[a]	Well-spoken	(22)	Convinced
(3)[a]	Kind	(13)[a]	Relaxed	(23)	Extraverted
(4)[a]	Well-mannered	(14)[a]	Poised	(24)	Confused
(5)[a]	Pleasant	(15)[a]	Self-assured	(25)	Optimistic
(6)[a]	Trustworthy	(16)[a]	Informed	(26)	Reserved
(7)[a]	Truthful	(17)[a]	Logical	(27)	Spontaneous
(8)[a]	Dependable	(18)[a]	Educated	(28)	Attractive
(9)[a]	Honest	(19)[a]	Wise	(29)[b]	Calm
(10)[a]	Reliable	(20)[a]	Scientific	(30)[b]	Nervous
The witness seemed to…					
(31)[b]	... control his/her behavior.				
(32)[b]	... feel comfortable.				
(33)[b]	... organize his/her thoughts.				
(34)[b]	... admit memory failures.				
(35)[b]	... act natural.				
(36)[b]	... think hard.				
(37)	... be prepared for the interrogation.				

Note. [a] Items 1 to 20 were adapted from the *Witness Credibility Scale* (Brodsky et al., 2010); [b] Items 29 to 36 were adapted from the *Observed Witness Efficacy Scale* (Cramer et al., 2013); Endpoints of the scales were from *1 = Not at all* to *7 = Very much.*

Appendix E

Means, Standard Deviations and Internal Consistencies (Cronbach's alpha) of Observer Rating Scales (Likeability, Trustworthiness, Knowledge and Confidence) as well as Corrected Item-Total Correlations (CITC) of the Original Items

	Identification only			Identification and descriptions		
Scales and items	*M*	*SD*	*CITC*	*M*	*SD*	*CITC*
Likeability	4.78	0.63	*alpha* = .89	4.78	0.59	*alpha* = .91
Friendly[a]	5.15	0.68	.90	5.18	0.64	.84
Respectful[a]	5.23	0.69	.70	4.99	0.56	.72
Kind[a]	4.46	0.68	.77	4.54	0.61	.85
Well-mannered[a]	5.17	0.76	.70	5.15	0.68	.79
Pleasant[a]	4.86	0.76	.91	4.80	0.71	.85
Attractive	3.80	1.02	.48	4.01	0.98	.60
Trustworthiness	4.98	0.68	*alpha* = .95	4.99	0.72	*alpha* = .97
Trustworthy[a]	4.82	0.79	.86	4.86	0.78	.93
Truthful[a]	5.21	0.63	.82	5.11	0.75	.83
Dependable[a]	4.70	0.83	.88	4.76	0.78	.91
Honest[a]	5.51	0.56	.78	5.47	0.72	.86
Reliable[a]	4.77	0.83	.91	4.76	0.87	.92
Credible	4.89	0.88	.88	4.99	0.73	.91

Appendix E (continued)

	Identification only			Identification and descriptions		
Scales and items	*M*	*SD*	*CITC*	*M*	*SD*	*CITC*
Knowledge	4.47	0.75	*alpha* = .95	4.43	0.69	*alpha* = .95
Informed[a]	4.24	0.90	.85	4.35	0.77	.79
Logical[a]	4.71	0.89	.89	4.72	0.72	.88
Educated[a]	5.02	0.69	.77	4.89	0.70	.83
Wise[a]	4.97	0.71	.81	5.10	0.76	.82
Scientific[a]	3.95	0.94	.83	3.94	0.91	.81
Well-spoken[b]	4.42	0.80	.70	4.53	0.95	.83
Organized thoughts	4.56	0.95	.86	4.63	0.83	.91
Prepared for interrogation	3.84	1.10	.81	3.31	0.75	.63
Confidence	4.44	0.70	*alpha* = .93	4.44	0.59	*alpha* = .92
Confident[a]	4.51	0.89	.90	4.36	0.77	.81
Relaxed[a]	4.10	0.85	.82	4.16	0.74	.87
Poised[a]	4.48	0.72	.67	4.45	0.67	.84
Self-assured[a]	4.63	0.88	.81	4.58	0.71	.77
Calm	4.59	0.85	.49	4.61	0.83	.40
Feeling comfortable	4.11	0.79	.69	3.82	0.72	.87
Nervous (-)	4.53	0.95	.73	4.58	0.93	.73
Confused (-)	4.87	1.21	.76	5.54	0.86	.64
Convinced	4.35	1.25	.74	4.36	0.93	.75
Extraverted	4.03	0.68	.47	3.95	0.71	.38
Optimistic	4.68	0.87	.83	4.38	0.72	.68

Note. [a] These items were also part of this scale in Brodsky et al. (2010), the other items were added by us; [b] In Brodsky et al. (2010) this item was part of the Confidence scale; (-) The item was recoded to build the scale.

DISCUSSION

This dissertation reported three experiments investigating how to increase eyewitness identification accuracy and how to explain fact finders' judgmental processes when evaluating eyewitness identification decisions.

In Experiment 1, re-reading one's own person description prior to the identification task (Cutler, Penrod, & Martens, 1987; Cutler, Penrod, O'Rourke, & Martens, 1986; Sporer, 2007) has been shown to be a promising system variable to increase correct identification decisions. Challenging former research on the verbal overshadowing effect (VOE: Alogna et al., 2014; Schooler & Engstler-Schooler, 1990) the present results suggest beneficial effects of describing the perpetrator on identification accuracy.

Experiment 2 and 3 shed light on fact finders' judgmental processes when evaluating the accuracy of an identification decision. Subjective utilities (i.e., which cues do observers use to make their judgments and how do they interpret and weight them) and ecological validities (i.e., which cues are objectively related to identification accuracy; Brunswikian lens model: Brunswik 1956, 1965) of different eyewitness characteristics were contrasted. Experiment 2 focused on observers' ratings of several meta-memory variables and suggested the use of videotaped think-aloud protocols (Ericsson & Simon, 1993) as a fruitful method to make objectively valid indicators of identification accuracy more salient, and thus more perceivable and usable by observers. Experiment 3 demonstrated that observer judgments were also affected by indirect and more peripheral eyewitness characteristics (i.e., the subjective impression of certain witness traits, the attributed power of the witness's speech style as well as different description qualities). However, almost none of these indirect measures have been shown to be diagnostic of identification accuracy.

In the following sections the main results of each experiment are discussed. Practical recommendations to improve identification accuracy as well as to improve fact finders' judgment accuracy when evaluating identification decisions are derived. For a more detailed discussion see the discussion section of each experiment.

How to Increase Identification Accuracy: Beneficial Effects of Person Descriptions

Experiment 1 was aimed to investigate potentially beneficial effects of describing the perpetrator on subsequent identification performance. Especially, re-reading one's own person descriptions was investigated as a system variable to increase identification accuracy (cf. Cutler et al., 1987; Cutler et al., 1986; Sporer, 2007). Therefore, a no description control group, a description only group and a description plus re-reading group had to identify the perpetrator in a target-absent [TA] or a target-present [TP] lineup after a delay of two days or five weeks.

Results are in line with Anderson' s (1983) theory of an associative memory network, assuming that re-reading one's own descriptions re-activates the encoding context in which the target face is embedded resulting in better identification performance. A significant identification advantage for re-readers compared to the no description control group was observed, with the odds for re-readers to make a correct identification decision being more than three times as high as the odds for the control group. Especially, in TA lineups the odds of making a correct rejection were eight times higher for re-readers than for the control group. The robustness of these results was corroborated by meta-analytical findings demonstrating a significant mean weighted effect size (*OR* = 2.15, 95% CI [1.34, 3.44], *k* = 4; cf. Lipsey & Wilson, 2001) for correct identification decisions (i.e., hits and correct rejections), as well as for correct rejections in TA lineups (*OR* = 2.74, 95% CI [1.27, 5.92], *k* = 4).

In contrast to former research on the verbal overshadowing effect (Alogna et al., 2014; Schooler & Engstler-Schooler, 1990), the present results demonstrated that describing a perpetrator does *not* impair identification performance when a delay of more than two days was included which is more representative of real crime situations (cf. findings in Meissner & Brigham, 2001; Wells, Steblay, & Dysart, 2015) and when conservative description instructions were used that do not encourage the witness to generate false description details. No traditional VOE was found. Instead, the effect tended to be annihilated and even reversed. The odds of a correct identification decision were almost three times larger when witnesses were asked to give a description (i.e., both description groups were joined) compared to the control group, demonstrating a beneficial effect of verbalizing the perpetrator's physical appearance. As an explanation memory advantages due to early first retrieval attempts are assumed (cf. *the testing effect*: McDermott, Arnold, & Nelson, 2014; Roediger & Karpicke, 2006), which have been shown to be effective especially after long delays (Butler & Roediger, 2007).

In sum, the present results clearly contradict the instructions of the British Police and Criminal Evidence Act (Code D, p. 53), to prevent eyewitnesses to "be reminded of any photograph or description of the suspect or any given other indication as to the suspect's identity" prior to the identification procedure. Instead, based on the present findings eyewitnesses explicitly *should* be reminded of their prior description by re-reading it to activate the associative memory network in which the original target face is embedded (cf. Anderson, 1983). Moreover, it should be noted, that describing the perpetrator--as it is an indispensable police practice anyway--does *not* impair identification performance. However, practitioners should be warned not to encourage witnesses to guess or generate false description details. To conclude, the re-reading procedure is an easily applicable and useful police practice to increase identification accuracy that does not require special trainings or any additional resources.

How to Increase Observers' Judgment Accuracy: The Use of Videotaped Think-Aloud Protocols

As not only misidentifications per se, but also erroneous evaluations of identification decision have been shown to lead to judicial errors (cf. Garrett, 2011; *www.innocenceproject.org*), the following experiments asked for a deeper understanding of fact finders' judgment processes to explain their judgment accuracy.

Experiment 2 focused on observers' perception, interpretation and use of several meta-cognitive variables that have been shown to be related to identification accuracy in former studies using witnesses' self-reports (e.g., Brewer & Wells, 2006; Dunning & Stern, 1994; Sauerland & Sporer, 2007; Wells, 1984; Wixted, Mickes, Clark, Gronlund, & Roediger, 2015). Thus, observer-participants rated several variables measuring the perceived confidence, decision time, identification difficulty and perceptual basis as well as perceived decision processes of 48 choosers' identification decisions. It was examined if these variables as perceived by observers discriminate between correct and incorrect identification decisions and if observers use these cues appropriately to make their judgments. According to Leippe's (1994) validity-intuition model, judgment accuracy should increase if observers intuitively use highly valid cues and ignore invalid cues.

It was demonstrated that eyewitness identification protocols, either presented as written transcripts (Study 1) or videotapes (Study 2), contain valid indicators of identification accuracy that were perceivable for observers only when think-aloud protocols (cf. Ericsson & Simon, 1993) were used in contrast to retrospective reasoning statements. In particular, when videotaped think-aloud protocols were presented to observers (Study 2) ratings of perceived perceptual basis, perceived confidence, perceived automatic as well as perceived absolute decision processes were higher for correct than for incorrect identifications. In contrast, ratings of perceived decision time and difficulty and perceived deliberate and effortful decision processes were lower for correct than for incorrect identifications.

Similar results have been shown in former research on the "postdictive" value of these variables since many years, however focusing on witnesses' self-reports (e.g., Dunning & Stern, 1994; Sauerland & Sporer, 2007, 2009; Sporer, 1992, 1993, 1994; Sporer, Penrod, Read, & Cutler, 1995). The present results show, that videotaped think-aloud protocols allow *observers* to appropriately assess these variables inasmuch as they discriminate between correct and incorrect identifications even without asking eyewitnesses to retrospectively evaluate their identification decision. In line with Nisbett and Wilson's (1977) claim that people are unable to correctly explain causes for their decision afterwards as well as to the extent that witness self-reports are often biased (e.g., Steblay, Wells, Bradfield Douglass, 2014; Wells & Bradfield, 1998), think-aloud procedures seem to serve as an useful alternative to make discriminating cues more salient for observers. In the present experiment, videotaped think-aloud protocols have been shown to help observers to detect witnesses who are beating around the bush (i.e., witnesses who make effortful, deliberative and time-consuming decisions) and to distinguish them from those who rely on their memory and make absolute and automatic decisions.

As these effects are mainly restricted to the use of *videotaped* think-aloud protocols, it is assumed that videotapes of eyewitnesses verbalizing their thoughts during the identification task facilitate the perception of discriminating cues by displaying additional nonverbal and paraverbal cues (e.g., speech style characteristics, response latencies and facial expressions) that are not contained in literally transcribed think-aloud protocols. However, it is noteworthy that the present studies did not allow statistical comparisons of both conditions. Thus, effects of presentation medium have to be treated with caution.

However, the presence of valid indicators of identification accuracy alone is not sufficient to increase judgment accuracy. Instead, observers also have to be sensitive to these discriminative cues and have to interpret and use them appropriately. In both studies, observers used almost all of the investigated cues to make their judgments and interpreted them as theoretically expected. For example, in line with the interpersonal reality monitoring approach (Mitchell & Johnson, 2000; Sporer, 2004) an increased number of cognitive operations (i.e., deliberations, which are reflected in high response latencies and a perceived high difficulty to make a choice) was perceived as an indicator of an erroneous memory, whereas automatic decisions were more likely to be perceived as correct.

However, the Brunswikian analyses revealed that correspondence between ecological validities and subjective utilities, that is, an appropriate usage and weighting of valid and invalid cues, differed due to type of decision protocol. In the retrospective reasoning conditions of both studies correspondence between ecological validities and subjective utilities was poor. Although observers heavily relied on almost all cues to make their judgments, cues were non-diagnostic of identification accuracy (except for automatic and eliminative decision processes). Thus, with retrospective reasoning protocols, which usually are used to ask witnesses to explain their identification decision afterwards, an appropriate assessment of valid indicators of identification accuracy is problematic and thus correct observer judgments are unlikely.

In contrast, in the think-aloud conditions of both studies, correspondence between ecological validities and subjective validities was moderate to high. Especially when videotaped think-aloud protocols were used were there several cues that discriminated between correct and incorrect identifications, which were appropriately used and weighted by observers. As a result, in the videotaped think-aloud condition a marginal increase in observers' judgment accuracy (70.8%) was observed compared to the videotaped retrospective reasoning condition (56.3%). However, the difference did not reach statistical significance, which is probably due to the small sample size in Study 2. Unfortunately, when literally transcribed think-aloud protocols were used judgment accuracy was relatively poor (58.3% compared to 51.4% in the retrospective-reasoning condition). This finding is attributed to the overall low discriminative value of the investigated cues in this condition. Thus, even a theoretically appropriate use of these cues was not sufficient to increase observers' judgment accuracy.

In sum, videotaped think-aloud protocols of the original identification decision may serve as a promising method to increase fact finders' judgment accuracies by making valid indicators of identification accuracy more salient and thus more perceivable and usable by observers. Importantly, the present results demonstrate that witness self-report measures could be replaced by observer ratings to prevent fact finders' from relying on biased witness statements at trial (e.g., inflated confidence statements due to post-identification feedback effects: Steblay, Wells, & Bradfield Douglass, 2014; Bradfield Douglass & Jones, 2013). Future researchers are encouraged to replicate these findings and to optimize think-aloud instructions to be used with identification decisions. Perhaps improving think-aloud instructions would increase judgment accuracy not only when videotapes are used but also when transcripts are presented.

Finally, the present results support the common recommendation to videotape a witness's original identification decision (e.g., Sporer, 1992, 1993; Wells, Small, Penrod, Malpass, Fulero, & Brimacombe 1998). In line with the present findings, videotapes are needed to record the witnesses' nonverbal and paraverbal behavior during the identification, which has been shown to expose valid markers of identification accuracy. Additionally, videotapes are needed to record the lineup instructions given to the eyewitness, to preserve the actual appearance of the lineup as well as to uncover any suggestive police officers' behaviors (cf. Garrett, 2011; Sporer, 1992, 1993; Wells et al., 1998).

Persuasive Effects of Indirect Measures on Observer Judgments

In Experiment 2, it was assumed that videotapes of a witness's identification decision display several peripheral witness aspects (i.e., nonverbal and paraverbal behaviors) that contribute to affect observer judgments. This assumption is in line with typical two-process models of persuasion (e.g., Elaboration Likelihood Model: Petty & Cacioppo, 1986) that distinguish between persuasive effects of central and content-related message aspects on the one hand and peripheral message characteristics and heuristic decision rules on the other hand (e.g., length of the testimony, message delivery style, perceived witness credibility). As Experiment 2 primarily focused on content-related aspects of an identification decision, Experiment 3 investigated the persuasive impact of more peripheral aspects and simple heuristics of a witness's testimony on observer judgments of identification accuracy. In particular, observer ratings of certain witness traits, speech style characteristics and different easily accessible description qualities were examined.

Results demonstrated that observers heavily relied on their subjective impression of the overall witness's confidence (cf. Cutler, Penrod, & Stuve, 1988; Wells, Lindsay, & Ferguson, 1979) and slightly tended to be affected by the perceived witness's trustworthiness and knowledge (cf. *Witness Credibility Scale*: Brodsky, Griffin, & Cramer, 2010). However, none of these ratings was related to objective identification accuracy, except for confidence ratings in the think-aloud condition (i.e., in this condition higher perceived confidence was an indicator of a correct identification). This finding is comparable to the results of Experiment 2 and supports the assumption that think-aloud protocols seem to facilitate the assessment of valid indicators of identification accuracy.

Persuasive effects of witness speech style characteristics were only weak and depended on the presence of additional person and event descriptions. When additional descriptions were presented to observers, speech characteristics did not affect observer judgments at all, presumably due to the large persuasive effects, which were observed for several description qualities. In contrast, when identification statements were presented *without* descriptions, two characteristics of a powerless speech style (i.e., hedges and questioning forms: Erickson, Lind, Johnson, & O'Barr, 1978; O'Barr, 1982) have been shown to be influential. As former studies demonstrated that a powerless speech style is likely to be associated with an impression of a less credible and less competent witness (e.g., Hosman & Wright, 1987; Jules & McQuiston, 2013; Ruva & Bryant, 2004), observers intuitively associated a frequent use of hedges and questioning forms with an incorrect identification.

Although not used by observers, two ecologically valid speech characteristics could be found in identification protocols (i.e., few long, indirect and evasive answers and a high speech rate were related with a correct identification). Former research findings also showed that correct identifications are made faster (e.g., Sporer, 1992, 1993, 1994) and more automatically, that is, with fewer deliberations and with less eliminative thought processes (e.g., Dunning & Stern, 1994). Thus, witnesses' decision processes, which have been shown to be valid indicators of identification accuracy in Experiment 2, seem to be revealed in the witness's speech style as well. This supports the assumption that videotapes contribute to facilitate the assessment of some valid indicators of identification accuracy (i.e., decision processes and response latency and difficulty) by exposing additional peripheral cues (i.e., the witness's speech style).

Finally, large persuasive effects of several description qualities were observed, although none of these ratings discriminated between correct and incorrect identifications. Regarding person description qualities, an attributed good memory for the perpetrator as well as a perceived high confidence were intuitively associated with a good identification performance (cf. conclusions from Bell & Loftus, 1988, 1989; Wells & Leippe, 1981). Although courts often recommend to rely on person description quality when evaluating the accuracy of an identification decision (cf. Neil vs. Biggers, 1972; Sporer & Cutler, 2003), present results suggest that observer ratings of these qualities do not have any discriminative value.

In sum, Experiment 3 demonstrated that several peripheral aspects of an eyewitness's testimony that are not directly linked to the identification decision affect observers' judgments. However, observer ratings of these characteristics were non-diagnostic of identification accuracy (except for some speech style characteristics). Although Experiment 2 suggested the use of videotaped identification decisions to facilitate the perception of *valid* indicators of identification accuracy, Experiment 3 showed that many *invalid* aspects become visible as well. Thus, an inappropriately high weighting of those peripheral aspects might contribute to explain fact finders' low abilities to correctly evaluate identification performance (e.g., Beaudry, Lindsay, Leach, Mansour, Bertrand, Kalmet, 2015; Wells & Lindsay, 1983). Perhaps informing fact finders about these obstructive influences might help to increase the accuracy of their evaluations.

Limitations and Practical Implications

Based on this dissertation practical recommendations can be derived to increase eyewitness identification accuracy and to facilitate its evaluation. However, it should be noted, that the present findings are solely based on experimental laboratory data and online surveys. Thus, generalization to real cases is supposed to be limited. Therefore, the replication of the present findings in field studies and with real identification decisions is highly recommended.

Conclusions.

(1) Witnesses should be asked to re-read their own person description prior to the identification task.

(2) Think-aloud instructions should be optimized for their use in identification procedures to make valid indicators of identification accuracy more salient to observers and thus to increase their judgment accuracy. As a result, valid observer ratings could replace error-prone witness self-reports to judge identification accuracy.

(3) Identification decisions should always be videotaped for later evaluations.

(4) Fact finders should be made aware of factors that validly discriminate between correct and incorrect identifications. For these factors weighting guidelines should be developed and pilot-tested for their effectiveness. Moreover, fact finders should be informed about invalid markers of identification decisions that should ideally be ignored.

References

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

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior, 22,* 261-295. doi: 10.1016/S0022537183902013

Beaudry, J. L., Lindsay, R. C. L., Leach, A.-M., Mansour, J. K., Bertrand, M. I., & Kalmet, N. (2015). The effect of evidence type, identification accuracy, line-up presentation, and line-up administration on observers' perceptions of eyewitnesses. *Legal and Criminological Psychology, 20,* 343-364. doi: 10.1111/lcrp.12030

Bell, B. E., & Loftus, E. F. (1988). Degree of detail of eyewitness testimony and mock juror judgments. *Journal of Applied Social Psychology, 18,* 1171-1192. doi: 10.1111/j.1559-1816.1988.tb01200.x

Bell, B. E., & Loftus, E. F. (1989). Trivial persuasion in the courtroom: The power of (a few) minor details. *Journal of Personality and Social Psychology, 56,* 669-679. doi: 10.1037/0022-3514.56.5.669

Bradfield Douglass, A., & Jones, E. E. (2013). Confidence inflation in eyewitnesses: Seeing is not believing. *Legal and Criminological Psychology, 18,* 152-167. doi: 10.1111/j.2044-8333.2011.02031.x

Brewer, N., & Wells, G. L. (2006). The confidence-accuracy relationship in eyewitness identification: Effects of lineup instructions, foil similarity,

and target-absent base rates. *Journal of Experimental Psychology-Applied, 12*, 11-30. doi: 10.1037/1076-898x.12.1.11

Brodsky, S. L., Griffin, M. P., & Cramer, R. J. (2010). The witness credibility scale: An outcome measure for expert witness research. *Behavioral Sciences & the Law, 28,* 892-907. doi: 10.1002/bsl.917

Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. Berkeley: University of California Press.

Brunswik, E. (1965). *The conceptual framework of psychology*. Chicago, London: The University of Chicago Press.

Butler, A. C., & Roediger, H. L. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology, 19*, 514-527. doi: 10.1080/09541440701326097

Cutler, B. L., Penrod, S. D., & Martens, T. K. (1987). 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

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*(2), 113-134.

Cutler, B. L., Penrod, S. D., & Stuve, T. E. (1988). Juror decision-making in eyewitness identification cases. *Law and Human Behavior, 12,* 41-55. doi: 10.1007/BF01064273

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

Erickson, B., Lind, E. A., Johnson, B. C., & O' Barr, W. M. (1978). Speech style and impression-formation in a court setting: Effects of powerful and powerless speech. *Journal of Experimental Social Psychology, 14,* 266-279. doi: 10.1016/0022-1031(78)90015-X

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. Massachusetts: A Bradford Book.

Garrett, B. L. (2011). *Convicting the innocent: Where criminal prosecutions go wrong*. Cambridge, Massachusetts: Harvard University Press.

Hosman, L. A., & Wright, J. W. (1987). The effects of hedges and hesitations on impression formation in a simulated courtroom context. *Western Journal of Speech Communication, 51,* 173-188. doi: 10.1080/10570318709374263

Jules, S. J., & McQuiston, D. E. (2013). Speech style and occupational status affect assessments of eyewitness testimony. *Journal of Applied Social Psychology, 43,* 741-748. doi: 10.1111/jasp.12002

Leippe, M. R. (1994). The appraisal of eyewitness testimony. In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 385-418). Cambridge, England: Cambridge University Press.

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

McDermott, K. B., Arnold, K. M., & Nelson, S. M. (2014). The testing effect. In T. J. Perfect & D. S. Lindsay (Eds.), *The SAGE handbook of applied memory* (pp. 183-200). Los Angeles, CA: SAGE.

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

Mitchell, K. J., & Johnson, M. K. (2000). Source monitoring: Attributing mental experience. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 179-195). Oxford, UK: Oxford University Press.

Neil v. Biggers, 409 U.S. 188 (1972).

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84,* 231-259. doi: 10.1037//0033-295X.84.3.231

O'Barr, W. M. (1982). *Linguistic evidence: Language, power, and strategy in the courtroom*. San Diego: Academic Press.

Petty, R. E., & Cacioppo, J. T. (1986). *Communication and persuasion: Central and peripheral routes to attitude change*. New York: Springer.

Police and Criminal Evidence Act (1984). Code D: Code of practice for the identification of persons by police officers. Retrieved from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/253831/pace-code-d-2011.pdf

Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1,* 181-210. doi: 10.1111/j.1745-6916.2006.00012.x

Ruva, C. L., & Bryant, J. B. (2004). The impact of age, speech style, and question form on perceptions of witness credibility and trial outcome.

Journal of Applied Social Psychology, 34, 1919-1944. doi: 10.1111/j.1559-1816.2004.tb02593.x

Sauerland, M., & Sporer, S. L. (2007). Post-decision confidence, decision time, and self-reported decision processes as postdictors of identification accuracy. *Psychology, Crime & Law, 13,* 611-625. doi: 10.1080/10683160701264561

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

Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology, 22,* 36-71. doi: 10.1016/0010-0285(90)90003-m

Sporer, S. L. (1992). Post-dicting eyewitness accuracy: Confidence, decision-times and person descriptions of choosers and non-choosers. *European Journal of Social Psychology, 22,* 157-180. doi: 10.1002/ejsp.2420220205

Sporer, S. L. (1993). Eyewitness identification accuracy, confidence, and decision times in simultaneous and sequential lineups. *Journal of Applied Psychology, 78,* 22-33. doi: 10.1037/0021-9010.78.1.22

Sporer, S. L. (1994). Decision times and eyewitness identification accuracy in simultaneous and sequential lineups. In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 300-327). Cambridge: Cambridge University Press.

Sporer, S. L. (2004). Reality monitoring and the detection of deception. In P. A. Granhag & L. A. Stroemwall (Eds.), *Deception detection in forensic contexts* (pp. 64-102). Cambridge: Cambridge University Press.

Sporer, S. L. (2007). Person descriptions as retrieval cues: Do they really help? *Psychology, Crime & Law, 13,* 591-609. doi: 10.1080/10683160701253986

Sporer, S. L., & Cutler, B. L. (2003). Identification evidence in Germany and the United States. In P. J. vanKoppen & S. D. Penrod (Eds.), *Adversarial versus inquisitorial justice: Psychological perspectives on criminal justice systems* (pp. 191-208). New York: Springer.

Sporer, S. L., Penrod, S., Read, D., & Cutler, B. (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., Wells, G. L., & Bradfield Douglass, A. (2014). The eyewitness post identification feedback effect 15 years later: Theoretical and policy implications. *Psychology, Public Policy, and Law, 20*, 1-18. doi: 10.1037/law0000001

Wells, G. L. (1984). The psychology of lineup identifications. *Journal of Applied Social Psychology, 14*, 89-103. doi: 10.1111/j.1559-1816.1984.tb02223.x

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

Wells, G. L., & Leippe, M. R. (1981). How do triers of fact infer the accuracy of eyewitness identifications: Using memory for peripheral detail can be misleading. *Journal of Applied Psychology, 66,* 682-687. doi: 10.1037/0021-9010.66.6.682

Wells, G. L., & Lindsay, R. C. L. (1983). How do people infer the accuracy of eyewitness memory? Studies of performance and a metamemory analysis. In S. Lloyd-Bostock & B. R. Clifford (Eds.), *Evaluating witness evidence: Recent psychological research and new perspectives* (pp. 41-55). Chichester: John Wiley & Sons.

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

Wells, G. L., Small, M., Penrod, S., Malpass, R. S., Fulero, S. M., & Brimacombe, C. A. E. (1998). Eyewitness identification procedures: Recommendations for lineups and photospreads. *Law and Human Behavior, 22,* 603-647. doi: 10.1023/a:1025750605807

Wells, G. L., Steblay, N. K., & Dysart, J. E. (2015). Double-blind photo lineups using actual eyewitnesses: An experimental test of a sequential versus simultaneous lineup procedure. *Law and Human Behavior, 39,* 1-14. doi: http://dx.doi.org/10.1037/lhb0000096

Wixted, J. T., Mickes, L., Clark, S. E., Gronlund, S. D., & Roediger, H. L. (2015). Initial eyewitness confidence reliably predicts eyewitness identification accuracy. *American Psychologist, 70*, 515-526. doi: 10.1037/a0039510

DEUTSCHE ZUSAMMENFASSUNG

Wird eine Person Zeuge oder Opfer einer Straftat und berichtet diese der Polizei, ist es üblich, dass zunächst nach einer Beschreibung des Tathergangs sowie einer detaillierten Täterbeschreibung gefragt wird. Nach der Ermittlung eines Tatverdächtigen folgt häufig ein Identifizierungsverfahren. Hierzu wird der/die Zeuge/in zu einer Wahlgegenüberstellung gebeten oder es wird ihm/ihr eine Lichtbildvorlage präsentiert, die das Foto des/der Tatverdächtigen neben einer Reihe von Fotos ähnlich aussehender Personen zeigt. Aufgabe des/der Zeugen/in ist es anzugeben, ob sich der/die Täter/in unter den gezeigten Personen befindet oder nicht.

Häufig sind genau diese Identifizierungsentscheidungen von Zeugen/innen fehlerhaft. Basierend auf Angaben des „Innocence Projects" (*www.innocenceproject.org*), einer amerikanischen Organisation, die Wiederaufnahmeverfahren von fälschlich verurteilten Personen anstrebt, um diese durch DNA-Analysen zu entlasten, haben falsche Identifizierungsentscheidungen in einer Mehrzahl von Fällen zu Justizirrtümern beigetragen (vgl. auch Garrett, 2011, 2012). Die rechtspsychologische Forschung versucht daher seit vielen Jahren Faktoren zu ermitteln, die die Richtigkeit einer Identifizierungsentscheidung bedeutsam beeinflussen (z.B. National Academy of Sciences, 2014; Wells & Olson, 2003). Es lassen sich hierbei *Schätz-* und *Systemvariablen* unterscheiden (Wells, 1978). Diese wirken auf einem zeitlichen Kontinuum, beginnend bei der Wahrnehmung und dem Behalten bis hin zum Abruf der Information, auf die Erinnerungsleistung eines/r Zeugen/in ein und stehen somit mit der Identifizierungsleistung in Zusammenhang (Sporer, 2008;

Sporer & Sauerland, 2008). Schätzvariablen sind solche Faktoren, die sich nur nachträglich einschätzen lassen und sich auf die situativen Bedingungen während der Tat (z.B. Sichtverhältnisse, Dauer des Verbrechens, Distanz zum Täter) sowie auf Merkmale des Täters/der Täterin (z.B. Alter, ethnische Zugehörigkeit) und des/der Zeugen/in selbst (z.B. Aufmerksamkeit) beziehen. Im Gegensatz dazu umfassen Systemvariablen alle Faktoren, die vom Rechtssystem kontrollierbar und somit im Rahmen des Strafprozesses modifizierbar sind, um die Identifizierungsrichtigkeit zu erhöhen (z.B. Art und Durchführung der Gegenüberstellung). In der vorliegenden Dissertation wird das erneute Lesen der eigenen Personenbeschreibung vor einem Identifizierungsverfahren als eine Systemvariable untersucht, um die Richtigkeit der Identifizierungsentscheidung zu erhöhen (Experiment 1).

Es sind jedoch nicht allein Falschidentifizierungen, die als Ursache für Justizirrtümer herangezogen werden müssen, sondern auch die fälschlichen Beurteilungen von Identifizierungsentscheidungen durch Richter und Laienrichter (sowie im adversarischen Rechtssystem durch Geschworene; vgl. Garrett, 2011). Auf dieser Beurteilungsebene (vgl. Sporer, 2007a) ergibt sich die Frage nach validen Kriterien, die zur Bewertung einer Identifizierungsaussage herangezogen werden können und daher Aufschluss über die Richtigkeit der Identifizierung geben (sog. Beurteilungsvariablen: Sporer, 1993; Sporer & Sauerland, 2008).

In den verschiedenen Rechtssystemen existieren teils unterschiedliche Empfehlungen darüber, welche Faktoren bei der Beurteilung einer Identifizierungsaussage Beachtung finden sollten (vgl. Neil vs. Biggers, 1972; Meurer, Sporer, & Rennig, 1990; Semmler, Brewer, & Bradfield Douglass, 2012; Sporer & Cutler, 2003). Auf Forschungsebene wurde die Validität dieser Faktoren häufig untersucht, wobei Zusammenhänge zwischen verschiedenen meta-kognitiven Aspekten einer Zeugenaussage und der Identifizierungsrichtigkeit festgestellt werden konnten (z.B. *subjektive Sicherheit eines/r Zeugen/in nach der Identifizierung*: Brewer & Wells, 2006; Sporer, Penrod, Read, & Cutler, 1995; Wixted, Mickes, Clark, Gronlund, & Roediger, 2015; *berichtete Entscheidungsprozesse*: Dunning & Stern, 1994; Sauerland & Sporer, 2007; Wells, 1984). Ebenso stehen nonverbale Aussageaspekte wie *Entscheidungszeiten* (z.B. Dunning & Perretta, 2002; Smith, Lindsay, Pryke, 2000; Sporer, 1992, 1993, 1994), aber auch objektive *Maße der Quantität und Qualität von Personenbeschreibungen* (z.B. Meissner, Sporer, & Susa, 2008; Sporer, 1992) mit der Identifizierungsleistung im Zusammenhang und können zur Beurteilung herangezogen werden.

Einige wenige Studien untersuchten auch die Beurteilung von Identifizierungsaussagen selbst. Diese konzentrierten sich auf die Fähigkeit von Urteilern/innen, zwischen richtigen und falschen Identifizierungen zu unterscheiden, wobei die berichteten Urteilsrichtigkeiten meist bei einem Zufallsniveau von 50% lagen (z.B. Beaudry, Lindsay, Leach, Mansour, Bertrand, & Kalmet, 2015; Brigham & Bothwell, 1983; Reardon & Fisher, 2011; Wells & Lindsay, 1983). Zudem zeigte sich, dass Urteiler/innen sich vor allem auf die subjektive Sicherheit von Zeugen/innen stützen (z.B. Wells, Lindsay, & Ferguson, 1979) und insgesamt nur über ein begrenztes Wissen über Faktoren verfügen, welche die Richtigkeit einer Identifizierungsentscheidung beeinflussen (z.B. Benton, Ross, Bradshaw, Thomas, & Bradshaw, 2006; Desmarais & Read, 2011).

Als mögliche Erklärung für die beobachteten geringen Urteilsrichtigkeiten liegt es daher nahe, dass Urteiler/innen bei der Bewertung von Identifizierungsaussagen invalide Kriterien heranziehen, während sie die Bedeutung tatsächlich valider Indikatoren unterschätzen (Leippe, 1994; Lindsay, 1994; Semmler et al., 2012). Um die Urteilsrichtigkeit bei der Bewertung von Identifizierungsaussagen zu erhöhen, beschäftigt sich die vorliegende Dissertation daher mit der Frage, welche Kriterien von Urteilern/innen zur Bewertung einer Identifizierungsaussage herangezogen werden, wie diese von ihnen interpretiert und gewichtet werden und ob diese Merkmale tatsächlich valide Indikatoren für eine korrekte Identifizierungsentscheidung darstellen. Experiment 2 konzentriert sich dabei auf meta-kognitive Aussageaspekte, die in Identifizierungsprotokollen enthalten sind und sich auf die Identifizierungsentscheidung des/der Zeugen/in beziehen. In Experiment 3 werden hingegen indirekte Aussageaspekte (z.B. das non- und paraverbale Verhalten des/der Zeugen/in sowie bestimmte Merkmale von Personen- und Tathergangsbeschreibungen) untersucht, die in keinem direkten Zusammenhang mit der Identifizierungsentscheidung stehen.

Experiment 1: Das erneute Lesen von Personenbeschreibungen als Systemvariable zur Erhöhung der Identifizierungsrichtigkeit

In Experiment 1 wurde untersucht, inwiefern Personenbeschreibungen gezielt als Systemvariable (Wells, 1978) eingesetzt werden können, um die Identifizierungsrichtigkeit eines/r Zeugen/in zu erhöhen. Basierend auf Andersons (1983) assoziativer Netzwerktheorie des Gedächtnisses wird ein Stimulus (z.B. das Gesicht des

Täters/der Täterin) nie alleine im Gedächtnis gespeichert, sondern stets zusammen mit externalen (z.B. Umgebung) und internalen (z.B. Emotionen des/der Zeugen/in) Kontextinformationen enkodiert und verarbeitet. Es wird angenommen, dass die enkodierten Informationen in Form eines assoziativen Netzwerks im Gedächtnis vorliegen und daher über beliebige Pfade miteinander verbunden sind und über diese auch aktiviert und abgerufen werden können. Kann die Erinnerung an das Gesicht des Täters/der Täterin beispielsweise nicht direkt abgerufen werden, sollte diese jedoch über assoziierte Kontextinformationen aktiviert werden können. Basierend auf ersten vielversprechenden Befunden von Cutler, Penrod, O'Rourke und Martens (1996) und Sporer (2007b) wird das erneute Lesen der eigenen, zuvor angefertigten Personenbeschreibung vor dem Identifizierungsverfahren in diesem Experiment als eine Form der Kontextwiederherstellung eingesetzt, um die Erinnerung an das Gesicht des Täters/der Täterin zu aktivieren und somit die Wahrscheinlichkeit einer richtigen Identifizierungsentscheidung zu erhöhen.

Insgesamt steht die Annahme eines solchen positiven Beschreibungseffekts im Gegensatz zu früheren Forschungsarbeiten zum verbalen Überlagerungseffekt („verbal overshadowing effect“: Alogna et al., 2014; Schooler & Engstler-Schooler, 1990), in denen ein negativer Effekt der Anfertigung einer Personenbeschreibung auf die nachfolgende Identifizierungsleistung postuliert wurde. Jedoch scheint das Auftreten des verbalen Überlagerungseffekts an bestimmte experimentelle Bedingungen geknüpft zu sein. Er tritt demnach vor allem dann auf, wenn ein kurzes Zeitintervall von nur wenigen Minuten zwischen Beschreibung und Identifizierung liegt und wenn Personen gebeten werden, eine sehr ausführliche Beschreibung des Täters/der Täterin (über 5 Minuten) abzugeben, was das Nennen falscher Personendetails offensichtlich provoziert. Werden hingegen realistischere Behaltensintervalle (d.h. mehrere Stunden bis Tage) verwendet, so zeigt sich, dass sich der Effekt verringert, verschwindet oder gar in einen gegenteiligen Effekt, den so genannten verbalen Erleichterungseffekt („verbal facilitation effect“), umkehrt (Alogna et al., 2014; Finger & Pezdek, 1999; Meissner & Brigham, 2001).

Erklärungen für den verbalen Erleichterungseffekt lassen sich aus Craik und Lockharts (1972) Theorie der Verarbeitungstiefe ableiten, die eine bessere Erinnerungsleistung postuliert, je tiefer ein Stimulus enkodiert wurde (z.B. durch das Hinzufügen selbst generierter semantischer Assoziationen während des Beschreibens). Zudem sollte ein zeitnaher erster Abruf der zu erinnernden Information diese im Gedächtnis konsolidieren und somit den späteren Abruf erleichtern (vgl. McDermott, Arnold, & Nelson, 2014; Roediger & Karpicke, 2006).

Unter der Verwendung realistischer Behaltensintervalle sowie angemessener Beschreibungsinstruktionen wurde in der vorliegenden Studie daher ein positiver Effekt der Anfertigung einer Personenbeschreibung auf die Identifizierungsleistung angenommen. Insgesamt wurden 208 Teilnehmer/innen in zwei Experimenten mit unterschiedlichen Behaltensintervallen (Experiment 1: 2 Tage; Experiment 2: 5 Wochen) untersucht. Unter Verwendung von Lichtbildvorlagen mit Täterabsenz (TA) und Täterpräsenz (TP) wurde die Identifizierungsleistung einer Kontrollgruppe, die keine Täterbeschreibung abgab, einer reinen Beschreibungsgruppe und einer Beschreibungsgruppe, die ihre Beschreibung vor der Identifizierung erneut lesen durfte, verglichen.

Übereinstimmend mit den Hypothesen, zeigten sich mehr korrekte Identifizierungsentscheidungen in der Gruppe, die ihre Beschreibung vor der Identifizierung erneut lesen durfte, als in der Kontrollgruppe (*OR* = 3.40). Besonders bei der Verwendung von TA-Lichtbildvorlagen war die Zahl korrekter Zurückweisungen in der Experimentalgruppe deutlich höher als in der Kontrollgruppe (*OR* = 8.14). Die Robustheit des Effekts ließ sich durch mehrere kleine Metaanalysen (korrekte Identifizierungsentscheidungen bei TA- und TP-Lichtbildvorlagen: *OR* = 2.15, *k* = 4; korrekte Zurückweisungen bei TA-Lichtbildvorlagen: *OR* = 2.74, *k* = 4) sowie durch eine Replikation des Effekts mit einer anderen Zielperson stützen. Folglich scheint das erneute Lesen der zuvor selbst angefertigten Personenbeschreibung als nützlicher Abrufreiz zu fungieren, der entsprechende Gedächtnispfade aktiviert, um die Erinnerung an den/die Täter/in abrufen zu können (vgl. Theorie eines assoziativen Gedächtnisnetzwerks: Anderson, 1983).

Wie erwartet, ließ sich ein traditioneller verbaler Überlagerungseffekt (Alogna et al., 2014; Schooler & Engstler-Schooler, 1990) in der vorliegenden Studie nicht nachweisen. Stattdessen konnte eine Überlegenheit der beiden Beschreibungsgruppen gegenüber der Kontrollgruppe in Form einer erhöhten Zahl an korrekten Identifizierungsentscheidungen beobachtet werden (*OR* = 2.89). Dies spricht dafür, dass der verbale Überlagerungseffekt bei realistischen Behaltensintervallen nicht auftritt und sich sogar umkehrt.

Zusammenfassend ist anzunehmen, dass Personenbeschreibungen in der polizeilichen Praxis, in der das Zeitintervall zwischen Beschreibung und Identifizierung in der Regel 13 bis 14 Tage beträgt (vgl. Feldstudie von Wells, Steblay, & Dysart, 2015), keinen negativen Effekt auf die

Identifizierungsleistung eines/r Zeugen/in ausüben und diese stattdessen sogar tendenziell erhöhen. Zudem können Personenbeschreibungen gezielt eingesetzt werden, um die Wahrscheinlichkeit einer richtigen Identifizierungsentscheidung zu steigern, ohne zusätzliche polizeiliche Ressourcen oder Training erforderlich zu machen.

Experiment 2: Beurteilung von Identifizierungsaussagen: Werden angemessene Kriterien verwendet?

Die bisherige Forschung hat gezeigt, dass Urteiler/innen nur begrenzt in der Lage sind, die Richtigkeit von Identifizierungsaussagen korrekt einzuschätzen (z.B. Wells & Lindsay, 1983). Um mögliche Empfehlungen zur Erhöhung der Urteilsrichtigkeit ableiten zu können, ist es daher notwendig, eine umfassende Einsicht in die zugrundeliegenden Urteilsprozesse zu gewinnen.

In Experiment 2 wurde im Rahmen des Brunswikschen Linsenmodells (Brunswik, 1956, 1965) untersucht, welche Kriterien Urteiler/innen zur Bewertung der Richtigkeit einer Identifizierungsaussage heranziehen („*Benutzte Hinweisreize*“), und ob diese Kriterien valide Indikatoren für die Richtigkeit der Identifizierung darstellen („*Ökologische Validität*“). Um die beobachtete Urteilsrichtigkeit zu erklären, wurde analysiert, inwiefern eine Übereinstimmung zwischen den verwendeten Kriterien und deren Validität vorliegt. Eine hohe Übereinstimmung sollte auftreten, wenn Urteiler/innen valide Indikatoren entsprechend interpretieren und angemessen gewichten, invalide Indikatoren hingegen vernachlässigen (Leippe, 1994; Semmler et al., 2012). Je höher diese Übereinstimmung ausfällt, desto höher ist die zu erwartende Urteilsrichtigkeit.

Um zu vermeiden, dass Urteiler/innen sich lediglich auf Selbstauskünfte von Zeugen/innen stützen, die häufig fehleranfällig sind (z.B. durch Feedbackeffekte: Steblay, Wells, & Bradfield Douglass, 2014; Wells & Bradfield, 1998), wurden die Teilnehmer/innen in der vorliegenden Studie gebeten, verschiedene Aussageaspekte, die sich in der Vergangenheit als valide Beurteilungsvariablen herausgestellt haben (z.B. subjektive Sicherheit, Entscheidungsprozesse, Entscheidungszeit und Schwierigkeit) basierend auf dem Identifizierungsprotokoll eines/r Zeugen/in *selbst* einzuschätzen.

Um die Entscheidungsprozesse und Gedanken eines/r Zeugen/in während der Identifizierung für die Urteiler/innen deutlicher sichtbar zu machen, wurden Protokolle des lauten Denkens (Ericsson & Simon, 1993) verwendet und mit üblicherweise verwendeten retrospektiven Urteilsbegründungen der Zeugen/innen verglichen. Basierend auf Nisbett und Wilsons (1977) Annahme, dass Personen nicht in der Lage seien, ihre Entscheidungen nachträglich angemessen zu begründen und diese lediglich rechtfertigen, wird angenommen, dass Protokolle des lauten Denkens die Erfassung und folglich den Gebrauch valider Beurteilungskriterien erleichtern sollten.

In zwei Studien wurden N = 288 (Studie 1) und N = 96 (Studie 2) Teilnehmer/innen je eins von insgesamt 48 Identifizierungsprotokollen zur Beurteilung präsentiert. Die verwendeten Identifizierungsprotokolle wurden in einer Pilotstudie erstellt und enthielten entweder eine korrekte Wahl der Täterin aus einer TP-Lichtbildvorlage oder eine falsche Wahl aus einer TA-Lichtbildvorlage. Die Identifizierungsentscheidungen der Zeugen/innen wurden in Form von Protokollen des lauten Denkens oder mit einem

Protokoll der nachträglichen Entscheidungsbegründung präsentiert. In Studie 1 wurden wörtliche Transkripte der Identifizierungsentscheidung verwendet, wohingegen in Studie 2 das entsprechende Videomaterial gezeigt wurde. Es wurde angenommen, dass Videos im Vergleich zu wörtlichen Transkripten zusätzliche non- und paraverbale Hinweisreize liefern würden (z.B. Zögern, Mimik, Sprachstil), die Aufschluss über die zu beurteilenden Kriterien sowie die Richtigkeit der Identifizierung geben. Somit wurde erwartet, dass valide Beurteilungskriterien in Studie 2 deutlicher für die Urteiler/innen sichtbar werden sollten als in Studie 1.

Es konnte gezeigt werden, dass basierend auf den Ratings der Urteiler/innen eine Reihe valider Indikatoren in den Identifizierungsprotokollen nachzuweisen waren, ohne dass die Zeugen/innen explizit nach ihrer Einschätzung gefragt wurden. Dies zeigte sich vor allem dann, wenn Videos der Protokolle des lauten Denkens verwendet wurden. Wie in bisherigen Forschungsarbeiten mit Selbstauskünften von Zeugen/innen berichtet wurde, gingen korrekte Identifizierungen im Vergleich zu falschen Identifizierungen mit einer höheren wahrgenommenen subjektiven Sicherheit des/der Zeugen/in (vgl. Sporer et al., 1995), einer geringeren Entscheidungszeit und Schwierigkeit (vgl. Dunning & Perretta, 2002) und einer wahrgenommenen besseren Wahrnehmungsbasis einher. Ebenso wurden die kognitiven Prozesse des/der Zeugen/in bei einer korrekten Identifizierungsentscheidung als automatischer und absoluter sowie als weniger abwägend und aufwändig eingeschätzt (vgl. Dunning & Stern, 1994; Wells, 1984). Wurden hingegen Protokolle einer nachträglichen Entscheidungsbegründung präsentiert, konnten unabhängig vom Präsentationsmedium nahezu keine signifikanten

Zusammenhänge zwischen den wahrgenommenen Aussageaspekten und der Identifizierungsrichtigkeit festgestellt werden.

Hinsichtlich der benutzten Hinweisreize, stützen sich die Urteiler/innen in beiden Studien unabhängig von der Art des verwendeten Protokolls stark auf fast alle der untersuchten Kriterien, um die Richtigkeit der Identifizierung zu beurteilen, und interpretierten diese übereinstimmend mit den aufgestellten theoretischen Annahmen (z.B. dem interpersonalen Realitätsüberwachungsansatz: Johnson & Raye, 1981; Mitchell & Johnson; Sporer, 2004). Eine adäquate Gewichtung der verwendeten Kriterien konnte jedoch nicht in allen Bedingungen gezeigt werden. In beiden Bedingungen, in denen Protokolle einer nachträglichen Entscheidungsbegründung verwendet wurden, überschätzten die Urteiler/innen die ökologische Validität der Kriterien stark. Folglich gaben sie den Kriterien, die tatsächlich keine Indikatoren für die Richtigkeit der Identifizierung darstellten, ein zu großes Gewicht, was sich in geringen Urteilrichtigkeiten von 51.4% (Studie 1) und 56.3% (Studie 2) niederschlug.

Wurden hingegen Protokolle des lauten Denkens verwendet, ergab sich eine vergleichsweise hohe Übereinstimmung zwischen den verwendeten Kriterien und deren ökologischer Validität. Besonders in Studie 2, in der Videos der Protokolle des lauten Denkens präsentiert wurden, diskriminierten die genutzten Kriterien zwischen korrekten und inkorrekten Identifizierungen und wurden von den Urteilern/innen entsprechend gewichtet. Folglich zeigte sich hier eine tendenziell höhere Urteilsrichtigkeit von 70.8%.

Zusammenfassend scheint die Verwendung von Videos von Protokollen des lauten Denkens ein vielversprechender Ansatz zu sein, um die Urteilsrichtigkeit bei der Bewertung von Identifizierungsaussagen zu erhöhen. Valide Indikatoren zur Beurteilung einer Identifizierungssaussage werden für Urteiler/innen sichtbar und somit nutzbar und interpretierbar gemacht. Somit könnten Videos von Protokollen des lauten Denkens eine Alternative zu der Verwendung potentiell verfälschter meta-kognitiver Selbsteinschätzungen von Zeugen/innen bei der Bewertung von Identifizierungsaussagen bieten.

Experiment 3: Die Verwendung indirekter Maße zur Beurteilung von Identifizierungsaussagen

Basierend auf typischen Zwei-Prozess-Theorien der Persuasion (z.B. Modell der Elaborationswahrscheinlichkeit: Petty & Cacioppo, 1986) kann angenommen werden, dass nicht nur inhaltliche Aussageaspekte die Beurteilung einer Identifizierungsaussage beeinflussen, sondern auch einfache Heuristiken und periphere, indirekte Merkmale einen Einfluss ausüben, die nicht direkt mit der Identifizierungsentscheidung eines/r Zeugen/in in Verbindung stehen (z.B. Aussehen, non- und paraverbales Verhalten des/der Zeugen/in).

Frühere Studien haben gezeigt, dass Zeugen/innen, deren Sprachstil als besonders überzeugungskräftig wahrgenommen wurde, insgesamt auch positivere Eigenschaften (z.B. Intelligenz, Kompetenz, subjektive Sicherheit und Glaubwürdigkeit) zugesprochen wurden als Zeugen/innen, deren Sprachstil als schwach eingestuft wurde (Erickson, Lind, Johnson, & O'Barr, 1978; Hosman & Wright, 1987; O'Barr, 1982; Ruva & Bryant, 2004; Schooler, Gerhard, & Loftus, 1986). Zudem scheint die Wahrnehmung bestimmter Personeneigenschaften untereinander stark korreliert zu sein und mit sozialen Urteilen (z.B. Glaubwürdigkeitsurteile, Strafzumessung) zusammenzuhängen (z.B. Garcia & Griffitt, 1978; Pryor & Buchanan, 1984; Zebrowitz, Hall, Murphy, & Rhodes, 1996). In Experiment 3 wurde daher untersucht, inwiefern der wahrgenommene Sprachstil eines/r Zeugen/in (vgl. „powerful speech“: Erickson et al., 1978; O’Barr, 1982) sowie der subjektive Gesamteindruck, den sich Urteiler/innen von einem/r Zeugen/in bilden (vgl. „Witness Credibility Scale“: Brodsky, Griffin, & Cramer, 2010), die Beurteilung einer Identifizierungsaussage beeinflussen.

Zusätzlich wurden Zusammenhänge zwischen verschiedenen Aspekten der eingeschätzten Quantität und Qualität von Personen- und Tathergangsbeschreibungen und der Beurteilung der Identifizierungsrichtigkeit betrachtet. In Übereinstimmung mit Annahmen des interpersonalen Realitätsüberwachungsansatzes (Mitchell & Johnson, 2000; Sporer, 2004) legen frühere Studienergebnisse nahe, dass Urteiler/innen sich bevorzugt auf einfache heuristische Merkmale wie die Anzahl an nebensächlichen Beschreibungsdetails stützen, um auf die Erinnerungsleistung des/der Zeugen/in und somit auf die Richtigkeit der Identifizierung zu schließen (Bell & Loftus, 1988, 1989; Wells & Leippe, 1981).

Parallel zu Experiment 2 wurde im Rahmen des Brunswikschen Linsenmodells (Brunswik, 1956, 1965) getestet, ob die untersuchten indirekten Merkmale zwischen korrekten und inkorrekten Identifizierungen diskriminieren, welche dieser Merkmale von den Urteilern/innen verwendet und wie diese im Beurteilungsprozess interpretiert und gewichtet werden.

Insgesamt zeigten die Ergebnisse, dass die Urteiler/innen sich bei der Bewertung der Identifizierungsrichtigkeit auf eine Reihe indirekter

Aussagemerkmale bezogen. Jedoch diskriminierten die verwendeten Merkmale nicht zwischen korrekten und inkorrekten Identifizierungen und wurden somit unangemessen stark gewichtet. Valide Aussagemerkmale ließen sich lediglich im wahrgenommenen Sprachstil eines/r Zeugen/in finden, wurden jedoch von den Urteilern/innen nicht als solche erkannt und daher nicht zur Urteilsfindung herangezogen.

Eine Identifizierungsentscheidung wurde schließlich eher dann als korrekt eingestuft, wenn Urteiler/innen den/die Zeugen/in als besonders selbstsicher wahrnahmen (vgl. Brodsky et al., 2010; Wells et al., 1979). Ebenso stützten sich die Urteiler/innen stark auf eine Reihe von Beschreibungsmerkmalen, die ein gutes Gedächtnis für den/die Täterin sowie eine hohe Sicherheit bezüglich der eigenen Erinnerung widerspiegelten. Im Gegensatz zu Befunden von Bell und Loftus (1988, 1989) und Wells und Leippe (1981) war jedoch nicht die reine Anzahl an erinnerten Details entscheidend, sondern vielmehr inhaltliche Aspekte wie eine konsistente und nachvollziehbare Schilderung der Tat. Persuasive Effekte vereinzelter wahrgenommener Sprachcharakteristika konnten nur dann beobachtet werden, wenn die Identifizierungsaussage alleine, d.h. ohne zusätzliche Personen- und Tathergangsbeschreibungen präsentiert wurde. Eine wahrgenommene fragende Betonung sowie die Verwendung von relativierenden Ausdrücken (z.B. „ich vermute" oder „vielleicht") wurden dabei als Merkmale eines schwachen Sprachstils interpretiert und somit als Indikatoren für eine falsche Identifizierung herangezogen (vgl. Erickson et al., 1978; Jules & McQuiston, 2013; O'Barr, 1982). Tatsächlich fanden sich jedoch für keine der von den Urteilern/innen *verwendeten* Merkmale Zusammenhänge mit der Richtigkeit der Identifizierung.

Zwar hat Experiment 2 gezeigt, dass die Präsentation von Videos einer Identifizierungsaussage die Wahrnehmung valider Aussagemerkmale scheinbar erleichtert, jedoch sprechen die Ergebnisse aus Experiment 3 dafür, dass unabhängig von der Art des präsentierten Entscheidungsprotokolls zusätzlich invalide periphere Aussagemerkmale sichtbar gemacht werden, die die Urteilsfindung ebenso beeinflussen. Eine unangemessene Gewichtung invalider peripherer Aussagemerkmale kann daher zur Erklärung der allgemein eher niedrigen Urteilsrichtigkeit bei der Bewertung von Identifizierungsaussagen herangezogen werden.

Praktische Implikationen

Zusammenfassend lassen sich anhand der vorliegenden Dissertation Anregungen für weitere Forschungsarbeiten sowie verschiedene praktische Empfehlungen ableiten, um die Identifizierungsrichtigkeit eines/r Zeugen/in zu erhöhen, sowie die Urteilsrichtigkeit bei der Bewertung von Identifizierungsaussagen zu verbessern:

(1) Zeugen/innen sollten gebeten werden, vor dem Identifizierungsverfahren ihre zuvor angefertigte Beschreibung des Täters/der Täterin erneut zu lesen.

(2) Instruktionen zum lauten Denken sollten für den polizeilichen Gebrauch optimiert und evaluiert werden.

(3) Zeugen/innen sollten während ihrer Identifizierungsentscheidung „laut denken“ und dabei gefilmt werden, um valide Beurteilungskriterien für Urteiler/innen sichtbar und somit nutzbar zu machen.

(4) Basierend auf den Ergebnissen der verschiedenen Brunswikschen Linsenmodelle können Richtlinien zur Gewichtung valider

Beurteilungskriterien entwickelt und getestet werden. Ebenso sollten Urteiler/innen über invalide Beurteilungskriterien informiert werden.

Literatur

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

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior, 22,* 261-295. doi: 10.1016/S0022537183902013

Beaudry, J. L., Lindsay, R. C. L., Leach, A.-M., Mansour, J. K., Bertrand, M. I., & Kalmet, N. (2015). The effect of evidence type, identification accuracy, line-up presentation, and line-up administration on observers' perceptions of eyewitnesses. *Legal and Criminological Psychology, 20,* 343-364. doi: 10.1111/lcrp.12030

Bell, B. E., & Loftus, E. F. (1988). Degree of detail of eyewitness testimony and mock juror judgments. *Journal of Applied Social Psychology, 18,* 1171-1192. doi: 10.1111/j.1559-1816.1988.tb01200.x

Bell, B. E., & Loftus, E. F. (1989). Trivial persuasion in the courtroom: The power of (a few) minor details. *Journal of Personality and Social Psychology, 56,* 669-679. doi: 10.1037/0022-3514.56.5.669

Benton, T. R., Ross, D. F., Bradshaw, E., Thomas, W. N., & Bradshaw, G. S. (2006). Eyewitness memory is still not common sense: Comparing jurors, judges and law enforcement to eyewitness experts. *Applied Cognitive Psychology, 20*, 115-129. doi: 10.1002/acp.1172

Brewer, N., & Wells, G. L. (2006). The confidence-accuracy relationship in eyewitness identification: Effects of lineup instructions, foil similarity,

and target-absent base rates. *Journal of Experimental Psychology-Applied, 12*, 11-30. doi: 10.1037/1076-898x.12.1.11

Brigham, J. C., & Bothwell, R. K. (1983). The ability of prospective jurors to estimate the accuracy of eyewitness identifications. *Law and Human Behavior, 7*, 19-29. doi: 10.1007/BF01045284

Brodsky, S. L., Griffin, M. P., & Cramer, R. J. (2010). The witness credibility scale: An outcome measure for expert witness research. *Behavioral Sciences & the Law, 28,* 892-907. doi: 10.1002/bsl.917

Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. Berkeley: University of California Press.

Brunswik, E. (1965). *The conceptual framework of psychology*. Chicago, London: The University of Chicago Press.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671-684. doi:10.1016/S0022-5371(72)80001-X

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*(2), 113-134.

Desmarais, S. L., & Read, J. D. (2011). After 30 years, what do we know about what jurors know? A meta-analytic review of lay knowledge regarding eyewitness factors. *Law and Human Behavior, 35*, 200-210. doi: 10.1007/s10979-010-9232-6

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

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

Erickson, B., Lind, E. A., Johnson, B. C., & O' Barr, W. M. (1978). Speech style and impression-formation in a court setting: Effects of powerful and powerless speech. *Journal of Experimental Social Psychology, 14,* 266-279. doi: 10.1016/0022-1031(78)90015-X

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. Massachusetts: A Bradford Book.

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

Garcia, L. T., & Griffitt, W. (1978). Impact of testimonial evidence as a function of witness characteristics. *Bulletin of the Psychonomic Society, 11,* 37-40. doi: 10.3758/BF03336759

Garrett, B. L. (2011). *Convicting the innocent: Where criminal prosecutions go wrong*. Cambridge, Massachusetts: Harvard University Press.

Garrett, B. L. (2012). Eyewitnesses and exclusion. *Vanderbilt Law Review, 65*, 451-506.

Hosman, L. A., & Wright, J. W. (1987). The effects of hedges and hesitations on impression formation in a simulated courtroom context. *Western Journal of Speech Communication, 51,* 173-188. doi: 10.1080/10570318709374263

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88*, 67-85. doi: 10.1037//0033-295X.88.1.67

Jules, S. J., & McQuiston, D. E. (2013). Speech style and occupational status affect assessments of eyewitness testimony. *Journal of Applied Social Psychology, 43,* 741-748. doi: 10.1111/jasp.12002

Leippe, M. R. (1994). The appraisal of eyewitness testimony. In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 385-418). Cambridge, England: Cambridge University Press.

Lindsay, R. C. L. (1994). Expectations of eyewitness performance: Juror's verdicts do not follow from their beliefs. In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 362-384). Cambridge, England: Cambridge University Press.

McDermott, K. B., Arnold, K. M., & Nelson, S. M. (2014). The testing effect. In T. J. Perfect & D. S. Lindsay (Eds.), *The SAGE handbook of applied memory* (pp. 183-200). Los Angeles, CA: SAGE.

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

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

Meurer, D., Sporer, S. L., & Rennig, C. (1990). Der Beweiswert von Personenidentifizierungen: Auf dem Weg von alltagspsychologischen Erfahrungssätzen zu empirisch überprüfbaren Fragestellungen [The probative value of eyewitness identifications: From common sense psychology to empirically testable hypotheses]. In D. Meurer & S. L. Sporer (Eds.), *Zum Beweiswert von Personenidentifizierungen* (pp. 1-18). Marburg, Germany: N. G. Elwert.

Mitchell, K. J., & Johnson, M. K. (2000). Source monitoring: Attributing mental experience. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 179-195). Oxford, UK: Oxford University Press.

National Academy of Sciences (2014). *Identifying the culprit: Assessing eyewitness identification*. Washington, D. C.: The National Academies Press.

Neil v. Biggers, 409 U.S. 188 (1972).

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84,* 231-259. doi: 10.1037//0033-295X.84.3.231

O'Barr, W. M. (1982). *Linguistic evidence: Language, power, and strategy in the courtroom*. San Diego: Academic Press.

Petty, R. E., & Cacioppo, J. T. (1986). *Communication and persuasion: Central and peripheral routes to attitude change*. New York: Springer.

Pryor, B., & Buchanan, R. W. (1984). The effects of a defendant's demeanor on juror perceptions of credibility and guilt. *Journal of Communication, 34,* 92-99. doi: 10.1111/j.1460-2466.1984.tb02176.x

Reardon, M. C., & Fisher, R. P. (2011). Effect of viewing the interview and identification process on juror perceptions of eyewitness accuracy. *Applied Cognitive Psychology, 25*, 68-77. doi: 10.1002/acp.1643

Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives* *on Psychological Science, 1,* 181-210. doi: 10.1111/j.1745-6916.2006.00012.x

Ruva, C. L., & Bryant, J. B. (2004). The impact of age, speech style, and question form on perceptions of witness credibility and trial outcome. *Journal of Applied Social Psychology, 34,* 1919-1944. doi: 10.1111/j.1559-1816.2004.tb02593.x

Sauerland, M., & Sporer, S. L. (2007). Post-decision confidence, decision time, and self-reported decision processes as postdictors of identification accuracy. *Psychology, Crime & Law, 13,* 611-625. doi: 10.1080/10683160701264561

Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology, 22,* 36-71. doi: 10.1016/0010-0285(90)90003-m

Schooler, J. W., Gerhard, D., & Loftus, E. F. (1986). Qualities of the unreal. *Journal of Experimental Psychology-Learning Memory and Cognition, 12*, 171-181. doi: 10.1037//0278-7393.12.2.171

Semmler, C., Brewer, N., & Bradfield Douglass, A. (2012). Jurors believe eyewitnesses. In B. L. Cutler (Ed.), *Conviction of the innocent: Lessons from psychological research* (pp. 185-209). Washington, DC: American psychological Association.

Smith, S. M., Lindsay, R. C. L., & Pryke, S. (2000). Postdictors of eyewitness errors: Can false identifications be diagnosed? *Journal of Applied Psychology, 85,* 542-550. doi: 10.1037//0021-9010.85.4.542

Sporer, S. L. (1992). Post-dicting eyewitness accuracy: Confidence, decision-times and person descriptions of choosers and non-choosers.

European Journal of Social Psychology, 22, 157-180. doi: 10.1002/ejsp.2420220205

Sporer, S. L. (1993). Eyewitness identification accuracy, confidence, and decision times in simultaneous and sequential lineups. *Journal of Applied Psychology, 78,* 22-33. doi: 10.1037/0021-9010.78.1.22

Sporer, S. L. (1994). Decision times and eyewitness identification accuracy in simultaneous and sequential lineups. In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 300-327). Cambridge: Cambridge University Press.

Sporer, S. L. (2004). Reality monitoring and the detection of deception. In P. A. Granhag & L. A. Stroemwall (Eds.), *Deception detection in forensic contexts* (pp. 64-102). Cambridge: Cambridge University Press.

Sporer, S. L. (2007a). Evaluating witness evidence: The fallacies of intuition. In C. Engel & F. Strack (Eds.), *The impact of court procedure on the psychology of judicial decision making* (pp. 111-150). Baden-Baden, Germany: Nomos Verlag.

Sporer, S. L. (2007b). Person descriptions as retrieval cues: Do they really help? *Psychology, Crime & Law, 13,* 591-609. doi: 10.1080/10683160701253986

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., & Cutler, B. L. (2003). Identification evidence in Germany and the United States. In P. J. vanKoppen & S. D. Penrod (Eds.), *Adversarial versus inquisitorial justice: Psychological perspectives on criminal justice systems* (pp. 191-208). New York: Springer.

Sporer, S. L., Penrod, S., Read, D., & Cutler, B. (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

Sporer, S. L., & Sauerland, M. (2008). Personenidentifizierung [Person identification]. *Forensische Psychiatrie, Psychologie und Kriminologie, 1,* 1-9. doi: 10.1007/s11757-008-0057-9

Steblay, N., Wells, G. L., & Bradfield Douglass, A. (2014). The eyewitness post identification feedback effect 15 years later: Theoretical and policy implications. *Psychology, Public Policy, and Law, 20*, 1-18. doi: 10.1037/law0000001

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

Wells, G. L. (1984). The psychology of lineup identifications. *Journal of Applied Social Psychology, 14,* 89-103. doi: 10.1111/j.1559-1816.1984.tb02223.x

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

Wells, G. L., & Leippe, M. R. (1981). How do triers of fact infer the accuracy of eyewitness identifications: Using memory for peripheral detail can be misleading. *Journal of Applied Psychology, 66,* 682-687. doi: 10.1037/0021-9010.66.6.682

Wells, G. L., & Lindsay, R. C. L. (1983). How do people infer the accuracy of eyewitness memory? Studies of performance and a metamemory analysis. In S. Lloyd-Bostock & B. R. Clifford (Eds.), *Evaluating witness evidence: Recent psychological research and new perspectives* (pp. 41-55). Chichester: John Wiley & Sons.

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

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

Wells, G. L., Steblay, N. K., & Dysart, J. E. (2015). Double-blind photo lineups using actual eyewitnesses: An experimental test of a sequential versus simultaneous lineup procedure. *Law and Human Behavior, 39,* 1-14. doi: http://dx.doi.org/10.1037/lhb0000096

Wixted, J. T., Mickes, L., Clark, S. E., Gronlund, S. D., & Roediger, H. L. (2015). Initial eyewitness confidence reliably predicts eyewitness identification accuracy. *American Psychologist, 70,* 515-526. doi: 10.1037/a0039510

Zebrowitz, L. A., Hall, J. A., Murphy, N. A., & Rhodes, G. (2002). Looking smart and looking good: Facial cues to intelligence and their origins. *Personality and Social Psychology Bulletin, 28,* 238-249. doi: 10.1177/0146167202282009

ERKLÄRUNG

Ich erkläre: Ich habe die vorgelegte Dissertation selbstständig 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.

____

Kristina S. Kaminski