How Do They Know What They Know?

The Case of Eyewitnesses

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Philosophie des Fachbereichs 06

der Justus-Liebig-Universität Gießen

vorgelegt von

Melanie Sauerland

aus

Bonn

2007

Dekan:	Prof. Dr. Joachim C. Brunstein
1. Berichterstatter:	Prof. Siegfried L. Sporer, Ph.D.
2. Berichterstatter:	Prof. Dr. Joachim Stiensmeier-Pelster
Tag der Disputation:	05. Juli 2007

Dank

Während der Erstellung dieser Arbeit haben mich eine ganze Reihe von Personen unterstützt, denen ich an dieser Stelle ganz herzlich danken möchte.

Mein Dank gilt meinem Betreuer Prof. Siegfried L. Sporer, der es mir ermöglicht hat, eines der interessantesten Themen der Rechtspsychologie zu bearbeiten. Bedanken möchte ich mich auch bei meinem Zweitgutachter Prof. Dr. Joachim Stiensmeier-Pelster.

Für die gute Zusammenarbeit, viele lange Gespräche, die unermüdliche Beratung und Unterstützung danke ich meinen KollegInnen Maike Breuer, Jürgen Gehrke und Tanja Stucke.

Mein besonderer Dank gilt Gary Wells, der mich während eines Besuchs seines Labors in Iowa im März 2006 so herzlich empfangen und bei sich aufgenommen hat. Er nahm sich sehr viel Zeit, um mich bezüglich der Auswertung des vierten Experimentes zu beraten.

Hervorzuheben ist Christine Ernst, die sich eigentlich inzwischen mit ganz anderen Sachen beschäftigt, für ihr stetes Interesse an der Thematik und ihr immer offenes Ohr.

Franziska Holub, Elisabeth Baumgartner, Sabine Bloss, Sarah Blum, Verena Bohn, Kerstin Bullmann, Maike Davids, Claudia Dill, Lisa German, Sarah Jasmin Günthert, Miriam Herder, Markus Jonitz, Morten Kaletsch, Miriam Kirschner, Norman Koch, Alana Krix, Silke Leifheit, Nadine Lipke, Isabelle Sauerbier, Alexandra Schmoranzer, Patricia Waschk und Jürgen Weber sei für die Unterstützung bei der Datenerhebung gedankt.

Weiterhin möchte ich Tanja Basseler, Anne Brinkmann, Cornelia Glagla, Merle Kurpjuweit, Andrea Reinke, Birgitt Sauerland, Tanya Sauerland und Volker Sauerland danken, die die vorliegende Arbeit in sehr unterschiedlicher, aber immer wesentlicher Hinsicht beratend und/oder unterstützend begleitet haben.

Gießen, im Juli 2007 *Melanie Sauerland*

ABSTRACT

This dissertation reports 4 Experiments that are concerned with the evaluation of eyewitness identification testimony. In Experiment 1, three target description groups were assessed in order to test different theoretical accounts regarding the relationship between identification performance and target description (verbal overshadowing): non-describers, describers, and rereaders (describers with rereading of the description before the identification task). Rereaders less frequently chose somebody from the lineup than the other two groups, lending support to the decision criterion shift approach (Clare & Lewandowsky, 2004). In Experiment 2, post-decision confidence, decision time, and self-reported decision processes were used as postdictors of identification accuracy. Using a decision rule including highly confident and fast choosers led to more correct classifications than either variable alone. Unexpectedly, self-reported decision processes were neither associated with identification accuracy for choosers nor for nonchoosers. In Experiment 3, combinations of post-decision confidence, decision time, and Remember-Know-Familiar (RKF) judgments were evaluated as postdictors of identification accuracy in a field experiment with ten targets and a very large sample. Fast and confident choosers were highly accurate. Including the RKF judgment did not lead to higher correct classification rates. Participants' self-reported (estimated) decision times also proved to be a postdictor of identification accuracy. Postdicting nonchoosers' identification performance by forming three homogeneous groups of nonchoosers failed, although there were differences with regard to confidence measures and decision times. Experiment 4 tested the usefulness of multiple lineup decisions (portrait face, body, bag, and profile face) for the assessment of identification testimony for nine different targets. Performance in the four different lineup types was not associated with each other, lending support to the idea that multiple lineups can serve as independent sources of evidence. Compared to foil choices and lineup rejections, target/suspect choices were most diagnostic of guilt. The portrait face lineup alone and its combination with the body lineup were most diagnostic for target/suspect choices.

To conclude, the present studies suggest that both decision times and post-decision confidence should be collected at the time of identification and be combined in order to assess identification accuracy. Investigators need to be aware though, that there is no postdictive value of nonchooser's estimates. Furthermore, there seem to be no negative effects of target descriptions on identification accuracy when there is a sufficient interval between description and identification, as there is in real cases. Finally, the data speak for the application of multiple lineups with regard to suspect/target choices as a procedure to avoid false identifications, whereas the benefit of multiple lineups for lineup rejections and foil choices seems to be limited. Future studies should address how many and which specific lineup types should be used. In real cases, the results for the assessment (decision times, confidence, decision processes) and control variables (target description, multiple lineups) examined in the present dissertation may vary from those that we obtain in laboratory or field studies. Reasons could be, for example, awareness of the severe consequences of false identifications and false rejections or the stress level at encoding or recognition. It would be interesting to collect data on these issues in real cases so they can be compared to the data obtained in laboratory/field studies. Undoubtedly, it would be a great contribution to the field of identification if data were collected even where DNA samples exist, so that identification accuracy could actually be assessed in real cases.

Table of Contents

INTRODUCTION

When a person witnesses a crime or becomes the victim of a crime, he or she will be asked to give a description of the perpetrator and later will be presented with a live lineup or a photospread. On the one side, eyewitnesses are capable of giving valuable testimony, on the other hand, history has taught us that eyewitness identification evidence is not always reliable. A famous case of misidentification is the one of Jennifer Thompson. In 1984, the 22-year old college student was assaulted and raped by a man who had broken into her apartment. In order to get a better view of him, she lured him into the illuminated parts of the apartment. She was determined to do everything that would enable her to later give a good description of the man and to identify him so he could be convicted and pay for what he had done to her. In the identification procedure she identified a man named Cotton as the offender. Only little later, police was given a hint by a prison inmate who reported that a person named Poole had confessed the offence while he had served time with him. Consequently, Jennifer Thompson was presented with the lineup one more time but she declared that she had never seen Poole before and that she was absolutely confident that Cotton was the man. The police believed Jennifer and it was not until 1995 that Cotton was exonerated by DNA analysis. The analyses also provided evidence that Poole committed the offence. Even though Jennifer had been very confident about her decision and in spite of the fact that she had been presented with the actual offender, she erred (for coverage on the case, see www.truthinjustice.org/positive_id.htm; www.pbs.org/wgbh/pages/frontline/shows/dna). The case of Jennifer Thompson is by far no exception. An investigation of cases in which biological evidence was kept and analyzed when DNA analysis became available demonstrated that eyewitness testimony was involved in most cases of wrong conviction (Scheck, Neufeld, & Dwyer, 2000; Wells, Small, Penrod, Malpass, Fulero, & Brimacombe, 1998). Nevertheless, DNA samples do not always exist and then eyewitness identification testimony may be the only evidence available. The importance of eyewitness identifications for investigating and prosecuting crimes is still undoubted. Research on the psychology of eyewitness identification began in the late 1970s and has since identified numerous estimator and system variables (Wells, 1978) that can have an influence on the accuracy of identification decisions. System or control variables are those over which the criminal justice system normally has control, such as instruction to witnesses or lineup composition. Estimator variables can be further broken down into situational variables which can only be explored post hoc (e.g., lighting conditions) and assessment variables (Sporer, 1993) that may be used to assess individual witnesses' decision making processes.

One well-studied system variable is the description of the target. Any eyewitness identification task such as a live lineup or a photospread is usually preceded by a description of the perpetrator provided by an eyewitness. Although this seems to be straightforward, some research has challenged the idea that this process is unproblematic: Numerous studies have shown that the very process of describing a target face can have negative effects on identification performance, that is, a verbal overshadowing effect (VOE) can occur (see the meta-analysis by Meissner & Brigham, 2001; Meissner, Sporer, & Schooler, 2007). Different theoretical explanations have been developed to account for this phenomenon. Experiment 1 further examined the mental processes involved. In particular, I investigated the influence that describing a target freely and with open-ended questions has on identification performance, and also how rereading of this description prior to the identification task affects identification performance. To ensure high ecological validity, a 1-week delay was inserted between witnessing the crime and target description on the one side and the identification task on the other side.

Assessment or postdiction variables are those that may be used to retroactively assess individual witnesses' decision making accuracy. The most widely used assessment variables are post-decision confidence (e.g., Sporer, Penrod, Read, & Cutler, 1995) and decision times (e.g., Sporer, 1992, 1993, 1994; Weber, Brewer, Wells, Semmler, & Keast, 2004). As the case of Jennifer Thompson showed, eyewitness confidence is not resilient against mistakes and research demonstrated that confidence can be influenced by feedback given by the investigator (Semmler, Brewer, & Wells, 2004; Wells & Bradfield, 1998; Wells, Olson, & Charman, 2003). On the other hand, post-decision confidence has shown to be useful when assessed right after the identification and when only the decisions of witnesses who made a positive identification decision (choosers) were considered (Sporer et al., 1995). Other methods which further explore the decision processes (Dunning & Stern, 1994) have also been explored lately (Brewer, Palmer, McKinnon, & Weber, 2005; Caputo & Dunning, 2005; Kneller, Memon, & Stevenage, 2001). Although decision time and post-decision confidence have been combined successfully in the past (Sporer, 1992; Weber, Brewer, Wells, Semmler, & Keast, 2004), most research on assessment variables, or postdictors, focused on one or another of these variables in isolation. It was the aim of Experiment 2 to determine the usefulness of combinations of post-decision confidence, decision time, and self-reported decision processes for distinguishing between accurate and inaccurate identification decisions. Additionally, I intended to explore the usefulness of a judgment of the state of awareness and the existence of recollective experience via a Remember-Know-Familiar (RKF) judgment. However, the operationalization of that measure failed such that there was not a satisfactory number of Know answers, thus not allowing any meaningful comparisons. Consequently, the wording of the corresponding items was revised for Experiment 3, an extensive field study with 10 target persons and 720 participants. In this study, post-decision confidence, decision time, and RKF judgments were combined. By the inclusion of 10 targets in Experiment 3, greater stimulus generalizability was ensured and allowed for internal replication (Wells & Windschitl, 1999).

Associations of identification accuracy with post-decision confidence (e.g., Sporer et al., 1995; Weber & Brewer, 2003) and decision time (e.g., Dunning & Stern, 1994; Sporer, 1992, 1993; Weber & Brewer, 2006; Weber et al., 2004) exist for positive identification decisions (choosers), but not for negative identification decisions (nonchoosers). Until now, no assessment variable that is capable of making valid postdictions about nonchoosers' decisions has been identified. Therefore, another aim of Experiments 2 and 3 was to test the usefulness of nonchoosers' self-reports about their decision processes as a postdictor of lineup rejection accuracy with two different methodologies.

Even though higher correct classification rates are obtained when postdictors are combined (Brewer & Wells, 2006; Sporer, 1992; Weber et al., 2004), the question of how to proceed with those witnesses who do not meet the set criteria with regard to, for example, decision time and post-decision confidence remains unanswered. Obviously, their evidence cannot be discarded per se. A method applied by Lindsay and colleagues (Lindsay, Wallbridge, & Drennan, 1987; Pryke, Lindsay, Dysart, & Dupuis, 2004) may be useful for all witnesses. Rather than presenting witnesses with one lineup, Lindsay et al. (1987) carried out two independent lineups, one in which the face of the target was presented, a second one in which a piece of clothing that the target wore during the crime was to be identified. The authors then estimated the diagnosticity ratios (Wells & Lindsay, 1980; Wells & Turtle, 1986). The diagnosticity ratio (DR) is defined as the ratio of correct and incorrect decisions. When the term equals 1, the lineup is neither diagnostic of the guilt nor of the innocence of the suspect. These results so far speak in favor of multiple lineups (Lindsay et al., 1987; Pryke et al., 2004). Nevertheless, altogether, the effect has been shown for only three targets (Lindsay et al., 1987; Pryke et al., 2004) and hence its generalizability remains unclear. Experiment 4, which reports different data of the field study (see Experiment 3), aimed at replicating the previous finding with nine targets and a large sample. Furthermore we wanted to contrast the view that multiple lineups are independent of one another with the notion that performance in blank (TA) lineups has predictive value for TP lineups (Wells, 1984).

References

Brewer, N., Palmer, M., McKinnon, A., & Weber, N. (2005, March). Decision processes underlying eyewitness identifications. Paper presented at the Annual Meeting of the American Psychology-Law Society, in Phoenix, AZ.

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.

Caputo, D. D., & Dunning, D. (2005). Eyewitness accuracy and decision processes using sequential lineups. Manuscript submitted for publication.

Dunning, D. &. Stern, L. B. (1994). Distinguishing accurate from inaccurate eyewitness identification via inquiries about decision processes. Journal of Personality and Social Psychology, 67, 818-835.

Kneller, W., Memon, A., & Stevenage, S. (2001). Simultaneous and sequential lineups: Decision processes of accurate and inaccurate eyewitnesses. Applied Cognitive Psychology, 15, 659-671.

Lindsay, R. C. L., Wallbridge, H., & Drennan, D. (1987). Do the clothes make the man? An exploration of the effect of lineup attire on eyewitness identification accuracy. Canadian Journal of Behavioural Science, 19, 464-478.

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

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 (pp 3-34). Mahwah, NJ: Lawrence Erlbaum & Associates.

Public Broadcasting Service. Aufgerufen im November 2006 von www.pbs.org/wgbh/pages/frontline/showes/dna/.

Pryke, S., Lindsay, R. C. L., Dysart, J. E., & Dupuis, P. (2004). Multiple independent identification decisions: A method of calibrating eyewitness identifications. Journal of Applied Psychology, 89, 73-84.

Scheck, B., Neufeld, P., & Dwyer, J., (2000). Actual innocence: When justice goes wrong and how to make it right. New York: Doubleday.

Semmler, C., Brewer, N., & Wells, G. L. (2004). Effects of postidentification feedback on eyewitness identification and nonidentification confidence. Journal of Applied Psychology, 89, 1-13.

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.

Sporer, S. L. (1993). Eyewitness identification accuracy, confidence and decision times in simultaneous and sequential lineups. Journal of Applied Psychology, 78, 22-33.

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). New York: Cambridge University Press.

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

Truth in Justice. Aufgerufen im November 2006 von http:// www.truthinjustice.org/positive_id.htm/.

Weber, N., & Brewer, N. (2003). The effect of judgment type and confidence scale on confidence-accuracy calibration in face recognition. Journal of Applied Psychology, 88, 490-499.

Weber, N., & Brewer, N. (2006). Positive versus negative face recognition decisions: Confidence, accuracy, and response latency. Applied Cognitive Psychology, 20, 17-31.

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.

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

Wells, G. L. (1984). The psychology of lineup identifications. Journal of Applied Social Psychology, 14, 89-103.

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.

Wells, G. L., & Lindsay, R. C. L. (1980). On estimating the diagnosticity of eyewitness nonidentifications. Psychological Bulletin, 88, 776-784.

Wells, G. L., Olson, E. A., & Charman, S. D. (2003). Distorted retrospective eyewitness reports as functions of feedback and delay. Journal of Experimental Psychology: Applied, 9, 42-52.

Wells, G. L., Small, M., Penrod, S. D., 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.

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

Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling and social psychology experimentation. Personality and Social Psychology Bulletin, 25, 1115-1125.

EXPERIMENT 1

Person Descriptions and Person Identifications: Verbal Overshadowing or Recognition Criterion Shift?

Any eyewitness identification task such as a live lineup or a photospread is usually preceded by a description of the perpetrator provided by an eyewitness. However, numerous studies have shown that the very process of describing a target face can have negative effects on identification performance, that is, a verbal overshadowing effect (VOE) can occur (see the meta-analysis by Meissner & Brigham, 2001; Meissner, Sporer, & Schooler, 2007). Different theoretical explanations have been proposed to account for this phenomenon. The present study examines the mental processes involved. In particular, we investigated the influence that describing a target freely and with open-ended questions can have on identification performance, and also how a witness's rereading of his or her own description prior to the identification task affects identification performance.

Verbal Overshadowing

The VOE, that is, decreased recognition performance in persons describing a face compared to non-describers was investigated in a series of six experiments by Schooler and Engstler-Schooler (1990). Since this initial set of experiments the body of research on the VOE has grown vastly (see Meissner & Brigham, 2001; Meissner, Sporer, & Susa, in press). While some researchers succeeded in replicating the VOE in face recognition experiments (e.g., Dodson, Johnson, & Schooler, 1997; Fallshore & Schooler, 1995; Ryan & Schooler, 1998; Schooler, Ryan, & Reder, 1996), others failed to replicate the VOE (e.g., Clifford, 2003; Meissner, Brigham, & Kelly, 2001; Memon & Bartlett, 2002; Tunnicliff & Clark, 1999; Yu & Geiselman, 1993) or even reported verbal facilitation effects (Cutler, Penrod, & Martens, 1987; Krafka & Penrod, 1985). In these studies, however, verbal and visual context reinstatement variables were manipulated together so that it remained unclear, which manipulation the positive effects have to be ascribed to. In a meta-analysis across 15 research articles with a total of 29 effect size comparisons ($\underline{N}$ = 2018), Meissner and Brigham (2001) found a small, yet significant, verbal overshadowing effect ($\underline{Z}_r$ = -0.12), demonstrating that participants who described a target face were 1.27 times more likely to later make an identification error when compared to non-describers. Furthermore, post-description delay and type of description instruction were found to moderate this relationship (Meissner & Brigham, 2001). Specifically, overshadowing effects were more likely to occur when the identification task immediately followed the description task, and when participants were given an elaborative description instruction, as opposed to a standard (free recall) description instruction. A comparison of studies using immediate or short delays ($\leq$ 10 min) with those using long delays ($\geq$ 30 min) revealed that the long delay influenced only participants in the no-description control condition. These participants showed a significant degree of forgetting compared to a short delay, whereas participants who had previously described the target face showed no change in performance across the delay conditions. As an explanation for these findings the authors suggest that a preservation of the memory trace across the extended post-description delay occurs due to verbalization.

However, an exception to these findings are those of Finger and Pezdek (1999, Experiment 3), where the description condition showed better identification performance after a retention interval of 24 minutes compared to an immediate testing description group. When compared to an immediate testing control group, there was no difference. As an explanation for release from verbal overshadowing, Meissner and Brigham (2001) suggested that differences in performance resulted from memory decay across the delay for the control condition. Unfortunately, as Finger and Pezdek (1999) did not include a post-description delay for the control group, this hypothesis cannot be investigated with their data (Meissner & Brigham, 2001). Another exception to the effect of post-description delay, as postulated by Meissner and Brigham (2001), are the findings by Schooler and Engstler-Schooler (1990, Experiment 5) where a significant VOE was observed even after a retention interval of two days.

There are three major theoretical explanations of the VOE. The first approach, originally termed transfer-inappropriate retrieval (Schooler, Fiore, & Brandimonte, 1997), but subsequently renamed transfer-inappropriate processing shift (TIPS; Schooler, 2002), suggests that the activation of verbal processes, involved in providing a face description inhibits subsequent non-verbal processes considered primarily responsible for face recognition. However, these verbal processes are not assumed to alter the original memory of the face. Schooler et al. (1997) hypothesized that in verbal processing of faces the emphasis lies on the featural information whereas in visual processing configural information is crucial. Hence, after describing a face participants are involved in a verbal (featural) mode of processing faces, and attempt to recognize a face by referring to the verbal instead of the visual memory trace. According to TIPS, VOE is not attributed to excessive reliance on a memory representation corresponding to verbalization. Rather, verbal recall is hypothesized to disrupt the successful application of nonreportable processes omitted in the initial verbal retrieval. Further support for the TIPS was provided by Dodson, Johnson, and Schooler (1997) and Brown and Lloyd-Jones (2002, 2003).

Another assumption made by the TIPS account is that the original memory only temporarily becomes inaccessible instead of being permanently altered by verbalization. Support for this assumption was found in several studies reporting release from verbal overshadowing. For example, Finger and Pezdek (1999, Experiment 3) found even an increase in identification accuracy after a retention interval between description and identification task of only 24 minutes (see also Finger, 2002).

Whether the VOE constitutes a temporal or a permanent interference with an eyewitness's memory of a face is of utmost practical importance in criminal investigations. While it is a standard procedure for police officials to interview eyewitnesses after a crime was committed and ask for a description of the perpetrator the probability of an immediately following identification task with a photospread or live lineup is highly unlikely. Even a time interval of two days between the witness's statement and an identification task as in Schooler and Engstler-Schooler (1990, Experiment 5) appears to be the exception rather than everyday practice (see Behrman & Richards, 2005; Valentine, Pickering, & Darling, 2003). Thus, a nonpermanent effect of verbalizing visual memories would have little practical relevance for identification procedures in police investigations in real cases.

Another theoretical explanation of the VOE, known as retrieval-based interference (RBI), suggests that the VOE arises from an alteration of the original memory trace caused by verbalization. The RBI was first introduced by Meissner, Brigham, and Kelley (2001) following a study in which they manipulated the amount and elaboration of people's verbalization. In an earlier study, Finger and Pezdek (1999, Experiment 1) had already compared the VOE after an elaborate verbalization (using the cognitive interview) and a standard verbalization of a previously seen photograph. In the elaborate interview participants recalled significantly more correct, incorrect and subjective details and also performed significantly less accurate in the identification task than participants in the standard interview condition. Meissner et al. (2001) partly replicated these results. By using three different instruction types they altered people's response criterion for descriptions. Specifically, when participants were explicitly instructed to provide a detailed and extensive description of the perpetrator and were even encouraged to guess (forced recall), subsequent identification accuracy significantly decreased both immediately and 30 min after the verbalization. Thus, unlike Finger and Pezdek (1999, Experiment 3), no release from verbal overshadowing was found after a similar retention interval. However, note that Meissner et al. (2001) used forced recall instructions in this condition. In contrast, when participants were discouraged from guessing and asked to only tell what they were sure they remembered correctly (warning condition) no VOE occurred. In this condition, identification performance improved in both immediate and delay conditions, relative to the control and forced recall conditions (Meissner et al., 2001). In another condition in which participants were given standard description instructions (free recall), identification accuracy did not differ significantly from the control condition, that is, no VOE occurred. This instructional bias effect was replicated in several studies (Meissner, 2002; MacLin, Tapscott, & Malpass, 2002) and has been found to be persistent after delays of 30 min or 1 week. As guessing also evokes more inaccurate details, Meissner et al. (2001) argued that these inaccuracies interfere with the original memory of the face, thus causing a higher error rate in the identification task. These results obviously challenge the TIPS account and strongly suggest that the visual memory trace is permanently altered by erroneous verbalization.

A novel account of the VOE, first introduced by Clare and Lewandowsky (2004), could possibly explain the results of those studies in which no apparent negative effect of verbalization was found. Clare and Lewandowsky (2004) pointed out that previous research has left open two major issues involving (a) the types of responses witnesses can make during identification and (b) the nature of the lineup. Obviously it should make a difference if participants merely have the option to choose from a lineup (forced-choice procedure) or can also reject a lineup (optional-choice procedure). Furthermore, only presenting target-present (TP) lineups in an experiment does not create a realistic scenario of the situation in a crime investigation because it does not allow to adequately assess the rate of false identifications.

Therefore, for the sake of ecological validity any experiment should also include target-absent (TA) lineups, along with optional-choice instructions. Consequently, participants need to decide not only who the perpetrator is but also whether the lineup must be rejected entirely (Clare & Lewandowsky, 2004). This decision requires a response criterion, such that witnesses say "not present" when no lineup member matches their memory or make an identification if a particular face in the lineup meets the response criterion. The placement of that response criterion is likely to influence identification performance: With a conservative criterion, people might rather not choose anyone from the lineup, whereas with a liberal criterion, identification attempts might increase (Clare & Lewandowsky, 2004). Criterion effects are pervasive and have frequently been observed in other memory paradigms that permit optional choice, especially when there is a trade-off between quantity and accuracy (Koriat & Goldsmith, 1994; Koriat, Goldsmith, & Pansky, 2000).[1] Therefore, criterion shifts may also be a contributing factor to the VOE if verbalization raises people's response criterion. As one possible explanation for a criterion shift following verbalization Clare and Lewandowsky (2004) proposed that people monitor their descriptive ability the same way they monitor their per-

[1]Note, however, that identification decisions are binary and therefore no such trade-off is possible.

formance during other memory tasks (Brigham & Pressley, 1988; Koriat & Goldsmith, 1996; Schraw, 1998).

In the identification context with a previous description task, people's inexperience with providing descriptions of faces implies that they might find the task rather difficult and that they are unlikely to have an appropriate reference against which they can compare their own description (Clare & Lewandowsky, 2004). Clare and Lewandowsky (2004) hypothesized that those two factors combined may make people unsure about the quality of their provided description, which in turn might lower their tendency to choose someone from a lineup. According to this assumption, the criterion shift to be expected after verbalization is a cautious shift as people become more cautious in their actions after experiencing their lacking ability to satisfyingly describe the perpetrator. Thus, Clare and Lewandowsky (2004) reasoned that in an experimental design including both TA and TP lineups the rate of correct identification decisions for describers should decrease in TP lineups (increase of false rejections) but increase in TA lineups (decrease of false alarms). Accordingly, no VOE should occur in forced-choice methodologies in which the decision of choosing or not choosing is reduced to one of choosing among alternatives (Clare & Lewandowsky, 2004). These assumptions clearly differentiate the criterion shift account from the competing TIPS and RBI accounts, as it focuses on choosing rates rather than considering only accuracy rates (TIPS) or suggesting that most errors should consist of false identifications (RBI).

As the distinction between optional-choice and forced-choice, and in consequence the use of TA lineups, has largely been ignored in previous research on the VOE (note that all studies included in the meta-analysis by Meissner & Brigham [2001] used TP lineups only), Clare and Lewandowsky (2004) assumed that people's response criterion may have contributed in unknown ways to existing experimental outcomes. However, it should be noted that in the initial set of experiments by Schooler and Engstler-Schooler (1990) the optional-choice methodology (yet, no TA lineups) was used and no criterion effect was found. Furthermore, Meissner (2002, Experiment 1) used TA and TP lineups but did not find a criterion shift in any of the description conditions.

In a series of experiments accounting for the factors of optional-choice and TA lineups, Clare and Lewandowsky (2004) reexamined the possibility of criterion effects in verbal overshadowing. The results of Experiment 1 clearly supported the criterion shift account: In an optional-choice identification task participants who had previously described the target significantly less often chose someone from a lineup, which in turn lead to a decrease of decision accuracy in the TP condition, but also to increased identification accuracy in the TA condition (i.e., correct rejections), relative to non-describers.

In order to further test their response criterion shift interpretation of Experiment 1, Clare and Lewandowsky (2004) applied the forced-choice methodology with TP lineups only in Experiment 2. They argued that, with the option of lineup rejection no longer available, participants in the verbalization condition should no longer make more errors than the control group if a response criterion shift was actually taking place. Supporting the criterion shift account, it turned out that describers did not perform less accurately than non-describers in the identification task when they were forced to choose someone from a lineup. Based on the findings in their study, the authors argued, that unlike the TIPS and RBI accounts, the criterion shift explanation can simultaneously account for (a) the results of those experiments in which standard description instructions and forced-choice identifications were used and no VOE occurred, (b) a large VOE in optional-choice TP lineups, and (c) the beneficial effect of verbalization with optional-choice TA lineups.

In summary, the results of both the meta-analysis (Meissner & Brigham, 2001) and the more recent study by Clare and Lewandowsky (2004) suggest that more than one process may be responsible for the observed variations in the VOE. The effect observed after standard verbalization may be based on a response criterion shift, while the impact of an elaborative description on identification performance appears to be due to either a change in processing style or an alteration of the original memory trace.

In the present study, three description groups were assessed: non-describers, describers only (describers without rereading the description prior to the identification task), and rereaders (describers with rereading of the description immediately before the identification task). A rereading group was included in order to put some participants back into a verbal mode. If this leads to inappropriate processing, then the VOE should be particularly high in this group.[2] In any realistic case where a witness describes the target, it is very unlikely that a suspect is found and that a lineup is constructed within less than 24 hours (see Behrman & Richards, 2005; Valentine et al., 2003). To ensure ecological validity, we therefore allowed for a 1-week delay between the description of the target and the identification task. In order to analyze the data with regard to the discussed decision criterion shift due to target descriptions (Clare & Lewandowsky, 2004), we included both TA and TP lineups. As standard description instructions (see Meissner & Brigham, 2001) were used, no VOE as caused by a change in processing style (TIPS) or an alteration of the original memory trace (RBI) was expected. Instead, we expected a response criterion shift (Clare & Lewandowsky, 2004). Specifically, we expected that describers shift their criterion towards a more conservative direction leading to a higher rate of lineup rejections in the identification task relative to the control group. As no results or theories regarding the duration of this cautious shift existed to date, there were two possible outcomes: (1) If the response criterion shift is permanent, all describers should reject the lineup equally often; (2) however, if the criterion shift is temporary, the description only group should choose equally often as non-describers and only rereaders should show a larger degree of lineup rejections. Furthermore, as both TA and TP lineups were included, the tendency of rereaders (respectively describers in general) to not choose a person from a lineup (as expected in the hypothesis above) should lead to less correct identifications (hits) in TP lineups but also to an increase of correct rejections in TA lineups.

---

[2]Note, however, describing and rereading are not the same; describing a person is the effort to verbally retrieve a visually encoded stimulus, whereas rereading the description merely reactivates the previously encoded verbal memory trace. One assumption is that both processes have a similar effect due to the VOE. Yet, the possibility that rereading has an associative effect in terms of context reinstatement should not be ignored.

Context Reinstatement

Verbalization apparently can have negative implications for identification accuracy by "overshadowing" the original memory trace, altering it, or producing a cautious shift in people's response criterion. However, contrary to the verbal overshadowing accounts, the use of person descriptions as a means of context reinstatement prior to the identification task may also lead to memory facilitation (e.g., Cutler, Penrod, & Martens, 1987; Cutler, Penrod, O'Rourke, & Martens, 1986). Cutler et al. (1987) provided context reinstatement cues by conducting an interview using the mnemonic instructions of the cognitive interview (see Geiselman, Fisher, MacKinnon, & Holland, 1985), providing snapshots of the crime scene, the victim, and another person involved, and by having the participants reread their written description of the incident and the perpetrator. Additionally, the target's disguise and presence of a weapon were manipulated. A significant interaction between disguise and context reinstatement was observed. That is, only when the target's face was difficult to see in the encoding situation was subsequent identification performance improved by the context reinstatement cues.

Cutler et al. (1986) conducted an extensive study in order to determine which of different context reinstatement methods was the most effective and under which circumstances. They found that only the rereading of a description about the incident had a significant effect on the identification decision, however, only in interaction with other factors. The first interaction found was between rereading and the retention interval: Those participants who did not reread their descriptions showed impairment of identification performance after a retention interval of one month compared to a retention interval of one week, whereas those participants who reread their own description performed almost equally well after both retention intervals. The second interaction was found between rereading and target presence in the lineup. In the TA condition, rereading was associated with better identification performance, while in the TP condition, rereading was associated with decreased identification performance. However, this effect only occurred when the offender was disguised and thus the encoding situation was non-optimal for the observer. While these results have been interpreted primarily on the basis of a possibly facilitative effect of context reinstatement it should be noted that they can also be explained with Clare and Lewandowsky's (2004) criterion shift account: While the memory preserving effect due to rereading across the one month retention interval only occurred in the target-disguise condition also a general tendency of rereaders to not choose (55%) when compared to non-rereaders (40%) was observed. Sporer (in press) compared describers only to describers with rereading and found a nonsignificant tendency toward a facilitative effect of rereading (51.7% vs. 36.0% accuracy in the description only condition) in the expected direction. However, Sporer's (in press) experiment may have lacked statistical power to detect a context reinstatement effect.

Thus, while several lines of research have tried to produce a memory facilitating effect by including the rereading of the target description, the results suggest the effect to be unstable. Possibly, other effects of describing the to be identified person and rereading the description prior the identification task, such as verbal overshadowing or response criterion shifts may counteract the impact context reinstatement may have under the given circumstances. Therefore, in the present study we included the rereading of one's own target description prior to the identification task to examine these rival views. According to the principle of context reinstatement, rereaders were expected to experience memory facilitation due to context reinstatement cues provided in their own person descriptions relative to the control group and the description only group. Therefore, rereaders should perform better in the identification task than the other groups.

Relationship between Quantity and Quality of Descriptions and Identification Accuracy

The significance of person descriptions for assessing eyewitness identification accuracy became apparent when the U.S. Supreme Court specified the accuracy of a witness' description of the criminal as one of five factors to be considered in the evaluation of identification evidence (Neil v. Biggers, 1972). The practical importance of person descriptions is evident from the discussion on their utility to assess ("postdict") the accuracy of a given identification (Sporer, 1992b, 1996; Wells, 1985; for a recent meta-analytic review, see Meissner, Sporer, & Susa, in press). Two aspects of person descriptions can be distinguished: description accuracy (usually defined as the number of correct descriptors divided by the number of correct plus incorrect descriptors), and description quantity (the total number of descriptors, irrespective of accuracy). In actual criminal cases, the accuracy of person descriptions cannot be established as this requires knowledge of the true perpetrator. Description quantity, however, can be ascertained by the number of descriptors or features mentioned which may or may not be related to identification accuracy (see Sporer, 1996). Quantity, and indirectly also accuracy of descriptions, is likely to be influenced by the type of instruction given, viz. standard instructions vs. elaborative description instructions, which have been an important moderator of the VOE (Meissner & Brigham, 2001).

Correlations between Description Accuracy and Identification Accuracy

While some studies observed positive correlations between description accuracy and identification accuracy (e.g., Meissner et al., 2001; Wogalter, 1996), other studies did not find an association between the two variables (Goldstein, Johnson, & Chance, 1979; Pigott & Brigham, 1985; Pigott, Brigham, & Bothwell, 1990; Grass & Sporer, 1991; see Sporer, 1996). In a recent meta-analysis, Meissner, Sporer, and Susa (in press) synthesized the reported point-biserial correlations between various aspects of description quality and quantity and identification accuracy. Across k = 32 hypothesis tests of the relationship between description accuracy and identification accuracy with N = 2973 participants, the weighted mean effect size was r = .14 (p < .001), with CIs of .11 and .18.

Sporer (in press) found a significant interaction between identification accuracy and choosing with accurate nonchoosers (M = 5.5) reporting more correct descriptors than inaccurate nonchoosers (M = 3.3). No effect was found for choosers (M = 4.2 vs. 4.6). Additionally, a series of experiments conducted by Meissner and colleagues (Meissner et al., 2001; Meissner, 2002) consistently demonstrated a significant negative association between the number of incorrect descriptors provided and identification accuracy while no such association was found for the number of correct descriptors given. Finger and Pezdek (1999) found that inaccurate identifiers reported more incorrect description details (M = 4.0) than accurate identifiers (M = 2.1) when only a 10-minute delay between description and identification task was inserted (Experiment 1). However, no such differences were found when longer delays between description and identification were used (Experiment 2: 1-hour; Experiment 3: 24 minutes).

Number of Features Mentioned and Identification Accuracy

Among other studies, Sporer (1992b) observed a positive correlation between the number of descriptors and identification accuracy (r = .28). However, other studies failed to find such an association (Franzen & Sporer, 1994b; Pigott et al., 1990; Wells, 1985; see Sporer, 1996). The meta-analysis by Meissner, Sporer, and Susa (in press) analyzed 33 studies with N = 2578 participants that examined the relationship between the number of features mentioned (description quantity) and identification accuracy. A weighted mean r = -.04, ns, with CIs of -.08 and .00, was found. Thus, the number of features mentioned seems to be unrelated to identification accuracy.

As Meissner, Sporer, and Susa (in press) noted, the results of studies on description accuracy and quantity are difficult to compare as the various authors used different operationalizations of description accuracy (e.g., some analyzed only facial descriptors while others included bodily descriptors or estimates of height and weight). Also, studies varied considerably with respect to the methodological rigor with which descriptions were assessed. For example, some analyses were carried out by establishing clear criteria for scoring and reporting high interrater agreement using Pearson's r, while others used single raters or established agreement simply by consensus of raters. Finally, some studies reported only few descriptive elements with little variation across participants while others contained lengthy descriptions that varied considerably. Of course, to the extent that description quantity and accuracy are not precisely measured or show very little variation across participants we cannot expect substantial correlations between these measures and identification accuracy.

The relationship between description quantity and identification accuracy will again be tested in the present study, using an elaborate coding scheme for description accuracy and quantity. Participants who gave more elaborate person descriptions in terms of quantity of details could be expected to experience a VOE in form of a change in processing style (TIPS; Schooler, 2002) or an alteration of the original memory trace (RBI; Meissner et al., 2001). If an alteration of the original memory trace (RBI) takes place, then both description groups should be equally affected by description elaborateness and accuracy. Specifically, elaborate describers of both description groups would be expected to show lower identification accuracy due to a form of VOE. Furthermore, a negative relation between description accuracy and identification accuracy would be expected for both groups.

On the other hand, if a change in processing style (TIPS) takes place, one would expect no effect of description accuracy or elaborateness on identification accuracy for describers only, as it would not be expected to last over an interval of 1 week. However, a change in processing style might be reactivated by rereading. Therefore, according to TIPS, elaborately describing rereaders would be expected to show lower identification accuracy rates than less elaborately describing rereaders. Additionally, rereaders who make an incorrect identification decision should have reported more incorrect descriptors. Consequently, description accuracy was expected to be positively correlated with identification accuracy for rereaders.

Method

Participants

One hundred and forty four individuals (72 males and 72 females; age 16 to 53, Mdn = 23 years) completed this experiment. Most participants were psychology majors (59%) who received course credit for their participation. Other participants were students with other areas of study (26%) and persons of various occupations (15%). They were randomly assigned to the conditions and tested individually. Half of the participants were tested at the Free University Berlin and half at the Justus Liebig University Giessen, Germany.

Design

A 3 (description: no description vs. description only vs. description with rereading) x 2 (target presence: TP vs. TA) between-subjects design was used. Dependent variables were identification accuracy and choosing rate. For the two description groups we measured description quantity and quality and their association with identification accuracy.

Stimulus Film

The stimulus film was taken from an earlier study by Sporer and Franzen (1994a). The film showed the theft of an expensive pair of sunglasses in an optometrist’s store. Altogether, five amateur actors (one woman and four men) participated in the film, which lasted 6 minutes and 30 seconds. The target person could be seen for 18 seconds. A close-up showing the target's head and shoulders in half-profile lasted for about 2 seconds. For the remaining time the target was filmed from a distance of several meters, where his head and whole torso could be seen. The actual theft took about 30 seconds (looking at the sunglasses on the rack, taking a pair and putting it in the pocket). The content of the film can be described as follows:

> A young female (optician) is standing behind the counter, polishing glasses. One after another, three male customers enter the store. When a fourth customer comes in, the optician asks him to wait for a moment until she has served the other customers. While she is taking care of the other customers, the fourth customer walks up to a rack, takes a pair of sunglasses from the shelve and puts it in his pocket. At that time he can be seen in the background while the third customer stands in the foreground at the counter waiting for the optician. He then leaves the store with the words: "This is taking too long, I'll come back later." After the third customer has paid and is about to leave the shop, the optician discovers the theft.

Photo Lineup

Each lineup consisted of six frontal 6 x 9 cm photographs simultaneously displayed on the computer screen at a color depth of 16.7 million colors, that is, 32 bit and a resolution of 1024 x 768 pixels, depicting six male individuals who all fit the general description of the target person as determined by a pilot study (effective size = 4.7). The men all wore the same sweater (different from clothing in the film) and each picture had been taken in the same windowless room with the same illumination and in front of the same wall.

The photos were arranged in two rows of three pictures each. For half of the participants, the target photograph was present (TP), for the other half it was replaced with an innocent foil (TA). Target position as well as distractor position were completely balanced to appear at any of the six positions an equal number of times.

Procedure

In the present experiment, an effort was made to achieve a high ecological validity by (a) inserting a retention interval of one week between description and identification task, and (b) including TA lineups equally often as TP lineups as well as a "not present" option following unbiased lineup instructions. In line with common police practice, we collected person descriptions first via free report followed by open-ended questions.

Before and in-between the separate parts of the experiment the participants were given thorough instructions on the computer screen how to respond to the questions and which keys to use. All instructions and lineup presentation were programmed with SuperLab 1.75 (www.cedrus.com).

Participants were informed that they were taking part in an experiment concerned with witness statements. The advertisement for the experiment displayed the question "Are You A Good Witness?", in order to appeal to peoples' ambition and curiosity but did not explicitly mention the topic of person identification. Participants were tested individually. Before viewing the video they were asked to watch the film closely and pay attention to every detail. Afterwards, participants completed a 30-minute filler task consisting of 40 general knowledge questions. In the following, participants of the experimental groups were instructed to give a detailed written description of the crime they had witnessed earlier in the film. Participants were asked to imagine they were making a witness statement for a real police investigation. This description was to be a free report consisting only of the information the participants remembered by themselves. Subsequently, participants were asked to answer eight open-ended questions concerning the crime on another sheet. The same procedure was followed for the description of the culprit. First, participants were asked to describe the target with their own words, emphasizing precision ("the description should be precise enough for another person to be able to recognize the culprit in a crowd"). Then, twelve open questions concerning the target's appearance followed (see Appendix). The description instructions were in line with the standard description instructions as used in previous studies (e.g., Finger & Pezdek, 1999; Meissner & Brigham, 2001).

The identification task was scheduled exactly one week later. Half of those participants who had provided a target description a week earlier were given the opportunity to reread their free description of the perpetrator as well as their answers to the specific questions concerning the physical appearance of the thief which they had provided. Before the identification task, participants indicated their pre-decision confidence regarding the accuracy of their identification decision on an eleven-point scale ranging from 0% to 100% (with intervals marked in 10% steps: 0%, 10%, ..., 100%). Subsequently, participants were asked to identify the culprit on the computer screen. Participants were advised that the culprit might or might not be present in the lineup. Decision time was measured automatically via SuperLab. After giving a rating of recollective experience, post-decision confidence ratings concerning the identification decision were assessed on an eleven-point scale ranging from 0% to 100% for all participants.

Coding of the Person Description Details

Two raters received training in which they individually rated all details named in ten cases and discussed the discrepancies with each other and the trainer. Subsequently, all descriptions were coded for correct, incorrect or subjective details (cf. Finger & Pezdek, 1999). Subjective details contained descriptors of personality traits or impressions the three raters could not objectively agree on as correct or incorrect (e.g., “looked tired”, “unfriendly face”) and were excluded from further analyses. Interrater correlations were computed for clothing, body, facial and other (e.g., posture, nationality) details separately for free report and the open questions. Due to the non-ambiguity of the coding system the mean interrater correlations after Fisher's Z-transformation of free descriptions and specific questions was equally high for body, facial, and other details, rs = .99. The mean interrater correlation after Fisher's Z-transformation for clothing details was r = .86 for free reports and r = .99 for open-ended questions. Estimates of age, height and weight of the target were excluded from analysis due to a lack of standard for coding "accuracy" (e.g., is a response of "20-25 years" to be considered accurate when a person is 25 years old?).

For data analysis, description scores combining free descriptions and specific questions were formed in addition to the four categories listed above. When a participant named the same detail in the free report and the following questions, this repetition was excluded from calculations. The subsequent analyses were carried out with the total number of details (correct + incorrect details = description quantity). Additionally, description accuracy scores[3] were computed (description accuracy = correct details /[correct + incorrect details]). We chose this rather time-consuming method to evaluate description accuracy and quantity because this allowed us to analyze data differently than in most other VOE studies. Specifically, we wanted to analyze data not only according to the description condition participants were allocated to but also with regard to the actual descriptive features of the descriptions. Although participants in the description conditions all received the same instructions, individual differences may lead to differences in description elaborateness (quantity) and description accuracy which in turn may have an influence on the processes believed to underlie the VOE.

---

[3]Using description accuracy as a predictor for identification accuracy has been criticized in previous research (Sporer, 1996; Wells, 1985) because the accuracy coefficient does not differentiate between more and less detailed descriptions. A description consisting of only one correct detail will obviously have a higher accuracy score (100%) than a description consisting of 10 details with 2 of them wrong (80%). However, the minimum of total descriptors provided by participants in our experiment was 7 (M = 16.8; SD = 4.1), which made analyses with description accuracy appropriate and justifiable.

Results

An alpha level of alpha = .05 was used for all inferential analyses. Cohen's (1988) d and f are reported as measures of effect size for ANOVAs, and Cramer's V and phi are reported for nonparametric analyses of 3 x 2 and 2 x 2 contingency tables, respectively.

In the following, we first report descriptive results for the whole sample. Subsequently, we look at the influence of the description conditions on identification accuracy and choosing behavior, followed by analyses of the relationship between description accuracy and quantity and identification outcomes.

Identification Accuracy

Table 1 displays the distribution of identification decisions for TA and TP lineups in the three experimental conditions. Altogether, 47.2% of the 144 participants made a correct decision. Identification accuracy differed significantly for the 60 choosers (33.3%) from the 84 nonchoosers (57.1%), $chi^2(1, N = 144) = 7.96$, $p = .007$, $phi = -.24$. Identification accuracy for TA lineups (66.7%) was higher than for TP lineups (27.8%), $chi^2(1, N = 144) = 21.85$, $p < .001$, $phi = -.39$.

Target Descriptions

Table 2 displays the means and standard deviations of accuracy rates and the number of descriptors separately for free reports, and for free reports and open-ended questions combined. The mean number of descriptors in the free report was M = 10.40 (SD = 3.75, Min = 4, Max = 20) and M = 14.05 for the open-ended questions (SD = 3.49, Min = 6, Max = 23). On average, open-ended questions increased the number of details reported by M = 6.38 (SD = 2.33), after eliminating descriptors mentioned in both. There were no differences in description accuracy for free report (M = 70.12%, SD = 17.51) and open-ended questions (M = 69.28%, SD = 11.75), $t(95) = .66$, $p = .512$, $d = -0.07$. The mean description accuracy for free report and open-ended questions together, after eliminating duplicates, was M = 69.68% (SD = 11.64).

Table 1

Distribution of Identification Decisions (in %) in Target-Absent and Target-Present Lineups

	Target-absent (n = 72)		Target-present (n = 72)			Total sample (N = 144)	
	Choosers	Nonchoosers	Choosers		Nonchoosers		
Condition	False alarm	Correct rejection	Hit	Foil identification	False rejection	Mean accuracy	Mean choosing rates
No description	33.3	66.7	29.2	25.0	45.8	47.9	43.8
(n = 48)	(8)	(16)	(7)	(6)	(11)	(23)	(21)
Description only	37.5	62.5	33.3	33.3	33.3	47.9	52.1
(n = 48)	(9)	(15)	(8)	(8)	(8)	(23)	(25)
Rereading	29.2	70.8	20.8	8.3	70.8	45.8	29.2
(n = 48)	(7)	(17)	(5)	(2)	(17)	(22)	(14)
Total sample	33.3	66.7	27.8	22.2	50.0	47.2	41.7
(N = 144)	(24)	(48)	(20)	(16)	(36)	(68)	(60)

Note. Figures enclosed in parentheses represent absolute frequencies.

Table 2

Means and Standard Deviations of Total Descriptors and Proportion Correct in Free Reports and in Free Reports and Open-Ended Questions Combined (N = 96).

	Free report				Free report and open-ended questions			
	Total number descriptors		Proportion correct		Total number descriptors		Proportion correct	
	M	SD	M	SD	M	SD	M	SD
Face total	4.52	2.05	.72	.23	7.01	2.01	.67	.17
Hair	2.77	1.10	.69	.25	4.18	1.15	.67	.17
Face holistic	.35	.58	.55	.47	1.11	.69	.55	.44
Eyes	.56	.68	.80	.38	.58	.69	.77	.39
Nose	.15	.48	.87	.32	.19	.53	.78	.38
Skin	.25	.44	.92	.28	.28	.45	.89	.32
Face other	.44	.69	.86	.33	.66	.86	.84	.33
Body	.59	.61	.35	.47	1.54	.91	.39	.40
Clothes	5.28	2.97	.74	.24	8.23	2.95	.78	.15
Total	10.40	3.75	.70	.18	16.78	4.06	.70	.12

For descriptors concerning the face only (holistic facial descriptors, hair, eyes, nose, skin, and other features) the mean number descriptors named across both free report and open-ended questions was M = 7.01 (SD = 2.01). Of these, a mean of M = 4.52 (SD = 2.05) descriptors had already been named in the free report. Thus, on average, 41.78% of the descriptors referred to the face of the perpetrator, most of which concerned hair style and color (59.63%).

Additionally, descriptors of age, weight, and height occurred with a mean of M = 2.25 (SD = 0.52), and subjective details with a mean of M = 1.27 (SD = 1.16), across both free report and open-ended questions. However, for the reasons explained earlier these latter details were not included in any of the following analyses.

Effects of Describing and Rereading

Effects on Identification Accuracy

Our main hypotheses were based on the question of whether identification performance was affected by the different description conditions. Identification accuracy did not differ as a function of the three conditions (no description: 47.9%; description only: 47.9%; rereading: 45.8%), $chi^2(2, N = 144) = .06$, $p = .973$, Cramer's V = .02, that is, no VOE in the traditional sense was found, nor did the results confirm our expectations concerning a context reinstatement effect, or an effect of memory preservation due to describing the target (see Table 1).

Effects on Choosing Rates

Table 1 above also displays the distribution of choosers and non-choosers in the three conditions. First, comparing describers (describers only and rereaders; 39.6%) to non-describers (43.8%), showed no effect on choosing rates, $chi^2(1, N = 144) = .13$, $p = .724$, $phi = -.03$. Secondly, comparing rereaders to non-rereaders showed that rereaders significantly less often chose a person from the lineup (29.2%) than non-rereaders (47.9%), $chi^2(1, N = 144) = 4.63$, $p = .033$, $phi = -.18$. Choosing rates in the rereading condition

(29.2%) also differed from those in the description only condition (52.1%), $\underline{chi}^2$(1, N = 96) = 5.23, p = .037, phi = -.23. These results indicate a cautious shift in the rereading condition.

Quality and Quantity of Descriptions and their Relationships to Identification Accuracy

Analyses of variance and correlational analyses were conducted to examine the associations between different aspects of person descriptions with identification accuracy and choosing behavior.

Three 2 x 2 x 2 unweighted means ANOVAs with choice (choosers vs. nonchoosers), description condition (description only vs. rereading), and decision outcome (correct vs. incorrect) as classification variables and description accuracy, total number of descriptors and number of false descriptors as dependent variables were computed. The main effect for choosing became significant for description accuracy, F(1, 88) = 4.60, p = .035, d = 0.11, and number of false descriptors, F(1, 88) = 5.08, p = .027, d = 0.12, but not for total number of descriptors, F(1, 88) = .88, p = .352, d = 0.05. Specifically, choosers gave more accurate descriptions (M = 70.06%) and less false descriptors (M = 4.81) than nonchoosers (M = 65.27%; M = 6.08). The main effects of decision outcome and description condition were nonsignificant for the three ANOVAs, Fs ≤ 2.62, ds ≤ .09. All three ANOVAs revealed significant interactions of Decision Outcome and Description Condition (description accuracy: F(1, 88) = 5.29, p = .024, f = .25; total number of descriptors: F(1, 88) = 3.99, p = .049, f = .21; number of false descriptors: F(1, 88) = 9.14, p = .003, f = .32).

Figures 1 and 2 display the interactions of Description Condition with Identification Accuracy for description accuracy and number of false descriptors. There was a simple main effect of identification accuracy within describers only for description accuracy, F(1, 88) = 7.56, p = .007, d = 0.14, and for number of false descriptors, F(1, 88) = 8.33, p = .005, d = 0.15. That is, within describers only, inaccurate identifiers gave less accurate descriptions (M = 65.86%) than correct identifiers (M = 74.89%). Likewise, within describers only, inaccurate identifiers reported more incorrect descriptors (M = 5.76) than accurate identifiers (M = 3.91). Within rereaders, no significant differences were found for description accuracy, F(1, 88) = .03, p = .855, d = 0.00, or number of false details reported, F(1, 88) = .38, p = .541, d = 0.03. The simple main effects of identification accuracy for total number of descriptors were nonsignificant for describers only, F(1, 88) = 1.38, p = .244, d = 0.06, and rereaders, F(1, 88) = 1.15, p = .286, d = 0.06.

Analogous ANOVAs with description accuracy of facial descriptions, total number of facial descriptors and number of false facial descriptors as classification variables were carried out. No significant main effects for choice, decision outcome, and description condition were found, Fs ≤ .2.06, |d|s ≤ .08. The only significant interaction was between Decision Outcome and Description Condition for number of incorrect facial descriptors, F(1, 88) = 6.25, p = .014, f = .27. That is, within describers only, incorrect identifiers (M = 2.48) reported more incorrect facial descriptors than correct identifiers (M = 1.74), F(1, 88) = 4.33, p = .040, d = 0.11. No significant difference was found within rereaders, F(1, 88) = .55, p = .461, d = 0.04.

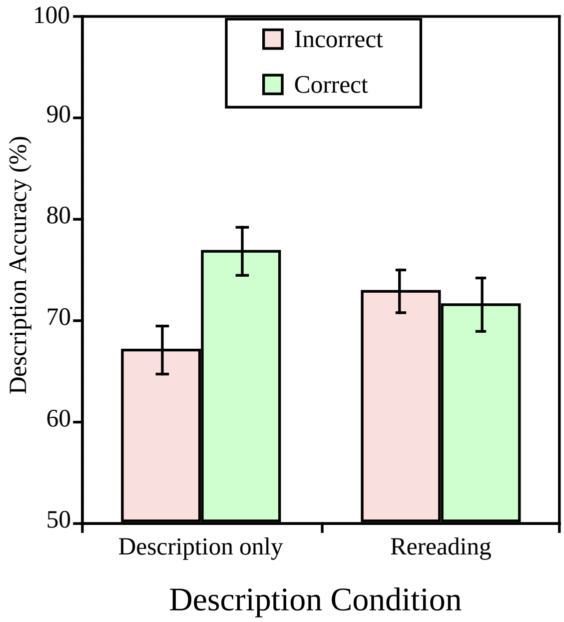


Figure 1. Description accuracy (%) of correct and incorrect lineup decisions with or without rereading.

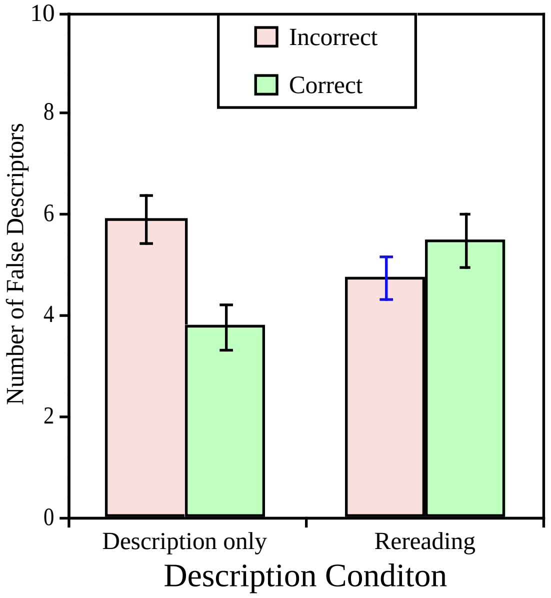


Figure 2. Number of false descriptors in correct and incorrect lineup decisions with or without rereading.

These interactions between Decision Outcome and Description Condition were nonsignificant for accuracy of facial descriptors, $F(1, 88) = 2.74$, $p = .101$, $f = .18$, and total number of facial descriptors, $F(1, 88) = 2.44$, $p = .122$, $f = .17$.

Correlational Analyses of Description Accuracy and Identification Accuracy

Overall description accuracy (i.e., referring to all descriptors, not just facial descriptors) and description quantity did not correlate significantly, $r(94) = -.16$, $p = .877$. Table 3 displays the correlations of description accuracy, description quantity as well as number of correct and incorrect details with identification accuracy for all describers and separately for describers only and rereaders. Additionally, associations are reported not only for all descriptors but also for facial descriptors. In the following, we will concentrate on the different result patterns of describers only and rereaders with regard to all descriptors named. Results for facial descriptors and the whole sample can be obtained from the Table.

Positive associations between description accuracy and identification accuracy were found for the description only group for open-ended questions, $r(46) = .41$, $p = .004$, and the combination of free report and open-ended questions, $r(46) = .39$, $p = .006$. No significant results emerged for the rereading group, all $|r|$s $\leq .04$, ns.

Correlational Analyses of Description Quantity, Correct and Incorrect Details and Identification Accuracy

No significant associations between identification accuracy and description quantity were found, |r|s ≤ .28, ns, with correlations for describers only tending to be negative, correlations for rereaders tending to be positive. Likewise, no significant associations between identification accuracy and number of correct details were found, |r|s ≤ .26, ns. However, for describers only, significant negative associations between number of incorrect details named and identification accuracy emerged for free report, $r(46) = -.30$, $p = .038$, open-ended questions, $r(46) = -.49$, $p < .001$, as well as the combination of both, $r(46) = -.46$, $p = .001$. No such associations were found for rereaders, |r|s ≤ .12, ns.

Correlational Analyses of Description Accuracy and Choosing

Table 4 displays the correlations of description accuracy, description quantity as well as number of correct and incorrect details named with choosing for all describers and separately for describers only and rereaders. Additionally, associations are reported not only for all descriptors but also for facial descriptors. As for identification accuracy, we will concentrate on the different result patterns of describers only and rereaders with regard to all descriptors named. Results for facial descriptors and the whole sample can be obtained from the Table.

For rereaders, choosing and description accuracy were positively associated for free report, $r(46) = .41$, $p = .004$, open-ended questions, $r(46) = .29$, $p = .047$, and the combination of both, $r(46) = .37$, $p = .011$. That is, witnesses who reread their low accuracy descriptions tended to not choose, while accurately describing rereaders tended to choose. No associations between choosing and description accuracy were found for describers only, |r|s ≤ .07, ns, with all the correlations tending to be negative.

Table 3

Correlations of Identification Accuracy with Description Accuracy, Description Quantity, Number of Correct Details, and Number of Incorrect Details

	All describers (N = 96)	Description only (n = 48)	Rereading (n = 48)
Description accuracy			
Free report			
All descriptors	.14	.25°	.04
Facial descriptors	.14	.23	.08
Open questions			
All descriptors	.22*	.41**	.02
Facial descriptors	.19°	.30*	.07
Free report and open questions[a]			
All descriptors	.18°	.39**	-.03
Facial descriptors	.12	.29*	-.06
Description quantity (correct and incorrect details)			
Free report			
All descriptors	.11	-.08	.28°
Facial descriptors	.08	-.08	.22
Open questions			
All descriptors	-.08	-.20	.04
Facial descriptors	.00	-.00	.00
Free report and open questions[a]			
All descriptors	-.01	-.18	.16
Facial descriptors	-.01	-.12	.10

Table 3 (continued)

	All describers (N = 96)	Description only (n = 48)	Rereading (n = 48)
Number correct details			
Free report			
All descriptors	.17	.07	.26°
Facial descriptors	.12	.04	.19
Open questions			
All descriptors	.06	.06	.05
Facial descriptors	.13	.21	.04
Free report and open questions[a]			
All descriptors	.10	.07	.14
Facial descriptors	.06	.08	.03
Number incorrect details			
Free report			
All descriptors	-.07	-.30*	.12
Facial descriptors	-.05	-.26°	.10
Open questions			
All descriptors	-.23*	-.49**	-.01
Facial descriptors	-.18°	-.30*	-.05
Free report and open questions[a]			
All descriptors	-.16	-.46**	.09
Facial descriptors	-.10	-.35*	.11

°$p \leq .10$; *$p \leq .05$; **$p \leq .01$.

[a]Features mentioned in both free report and open-ended questions were only coded once.

Correlational Analyses of Description Quantity, Correct and Incorrect Details and Choosing

No significant associations between choosing and description quantity with regard to all descriptors were found, $|r|s \leq .24$, ns. Likewise, no significant associations between identification accuracy and number of correct details were found, $|r|s \leq .22$, ns. However, for rereaders, significant negative associations between number of incorrect details and choosing emerged for free report, $r(46) = -.31$, $p = .032$, open-ended questions, $r(46) = -.35$, $p = .016$, and the combination of both for rereaders, $r(46) = -.31$, $p = .035$. No significant results were obtained for describers only, $|r|s \leq .09$, ns.

Table 4

Correlations of Choosing with Description Accuracy, Description Quantity, Number of Correct Details, and Number of Incorrect Details

	All describers (N = 96)	Description only (n = 48)	Rereading (n = 48)
Description accuracy			
Free report			
All descriptors	.22*	-.01	.41**
Facial descriptors	.11	-.17	.28°
Open questions			
All descriptors	.09	-.07	.29*
Facial descriptors	.10	-.06	.21
Free report and open questions[a]			
All descriptors	.16	-.05	.37*
Facial descriptors	.11	-.03	.25°
Description quantity (correct and incorrect details)			
Free report			
All descriptors	-.11	-.20	.02
Facial descriptors	-.00	-.10	.00
Open questions			
All descriptors	-.19°	-.15	-.24°
Facial descriptors	-.20°	-.35*	-.03
Free report and open questions[a]			
All descriptors	-.10	-.13	-.05
Facial descriptors	-.05	-.19	.14

Table 4 (continued)

	All describers (N = 96)	Description only (n = 48)	Rereading (n = 48)
Number correct details			
Free report			
All descriptors	.02	-.17	.22
Facial Descriptors	.07	-.12	.24°
Open questions			
All descriptors	-.12	-.18	-.05
Facial descriptors	-.10	-.30*	.09
Free report and open questions[a]			
All descriptors	.00	-.14	.17
Facial descriptors	.04	-.14	.26°
Number incorrect details			
Free report			
All descriptors	-.23*	-.09	-.31*
Facial descriptors	-.12	.02	-.18
Open questions			
All descriptors	-.17	.02	-.35*
Facial descriptors	-.16	-.12	-.15
Free report and open questions[a]			
All descriptors	-.19	-.02	-.31*
Facial descriptors	-.14	-.08	-.13

°$p \leq .10$; *$p \leq .05$; **$p \leq .01$.

[a]Features mentioned in both free reports and open-ended questions were only coded once.

Discussion

The major aim of this study was to investigate the influence of describing a target-thief seen in a film and the influence of rereading one's description prior to a lineup on identification performance and choosing rates. We considered a number of possible theoretical approaches with partly opposing predictions. Extensive care was taken to ensure ecological validity. First, the identification task was postponed until a week after both seeing the film and providing a written description of the target. Second, both TP and TA lineups were used. Third, the description instructions used were similar to the ones used in real criminal proceedings (free recall followed by a few open-ended questions to elicit additional details). The results of our experiment emphasize the need to reconsider the effect that person descriptions can have on eyewitness identification performance.

We found no impact of person descriptions and rereading on identification performance but only on choosing rates under certain conditions. No VOE as postulated by a change in processing style (TIPS, Schooler et al., 1997; Schooler, 2002) or by an alteration of the original memory trace (RBI; Meissner et al., 2001) was found. This was in line with previous research that used standard description instructions (Meissner & Brigham, 2001). Apparently, only the inclusion of a forced recall condition allows for a more direct test of RBI and TIPS (Meissner et al., 2001).

Also, our results yielded no support for a memory facilitation effect through context reinstatement by rereading of one's target description as found in some earlier studies (Cutler et al., 1986; Cutler et al., 1987). Instead, consistent with our hypothesis, our results support the recognition criterion shift approach as suggested by Clare and Lewandowsky (2004). The fact that control and experimental conditions did not differ in identification accuracy supports the assumption that the standard description instruction (or free recall) is a reasonably "safe" way to obtain a person description with regard to the VOE (cf., Meissner & Brigham, 2001), without reducing the description quantity to a minimum as observed with warning recall instructions (e.g., Meissner, 2002).

Our results are in line with other studies that also did not find a VOE after post-description delays of 24 hours or longer (e.g., Clifford, 2003; Memon & Rose, 2002; Yu & Geiselman, 1993) and the meta-analysis by Meissner and Brigham (2001) which detected a VOE only for elaborate description instructions. To our knowledge, there is only the exception of Schooler and Engstler-Schooler's study (1990, Experiment 5) who did use a retention interval of 48 hours and found a VOE. Yet, most studies examining the VOE used only short post-description delays of less than 24 hours (cf., Meissner & Brigham, 2001). Therefore, there is a clear need of studies with longer post-description delays in order to address the questions of the durability and permanence that target descriptions can have on identification performance. Also, we need to pay closer attention to the type of description instructions and the way description quantity and accuracy are measured. After all, the probability that a lineup is carried out immediately after the description is very low as documented in archival analyses of real criminal cases (Behrman & Richards, 2005; Sporer, 1992a; Valentine et al., 2003; van Koppen & Lochun, 1997). Hence, the ecological validity of previously reported findings is arguable.

With regard to the criterion shift, our results differed for rereaders and non-rereaders. Specifically, there were more nonchoosers within rereaders than within non-rereaders (description only and control conditions). As no such effect was found for describers in general, we have to assume that the recognition criterion shift only occurs if (a) the identification task immediately follows the description task, as assessed in the study by Clare and Lewandowsky (2004), or (b) the description is reactivated before the identification task, for example by rereading it, as it was the case in the present study. Our results in this regard are in line with the underlying processes proposed by Clare and Lewandowsky (2004). According to Clare and Lewandowsky, a criterion shift occurs due to the perceived difficulty of the description task combined with the presumed inadequacy of the description itself. Especially the latter of these factors is not confined to the situation immediately after the description task but can also be applied to the situation when participants reread their own descriptions. Concerning the duration of the criterion shift, our results indicate the effect to only prevail temporarily, for less than a week, as it appeared for rereaders only.

Similar to the present study, Sporer (in press) had applied a 1-week delay and used the same stimulus materials. He found a nonsignificant tendency of rereaders to be more accurate in their identification on TA and TP lineups than describers only but no effect of the experimental manipulations on choosing. Note, however, that Sporer (in press) did not include a no description control group, against which identification accuracy usually is contrasted in VOE studies.

The second aim of our study concerned the relationship of different aspects of person descriptions with identification accuracy and choosing behavior. The lack of a relationship between description accuracy and quantity suggests that these two aspects need to be considered separately. Differences in description quantity can also be interpreted as differences in elaboration, not evoked by varying instructions but most likely by the participants' individual differences in their description criterion or ability. Participants in the present study provided rather detailed target descriptions (descriptive details: M = 16.8; SD = 4.1) when compared to the findings in archival analyses (Sporer, 1992a, M = 9.7; van Koppen & Lochun, Mdn = 8) or staged event studies (e.g., Lindsay et al., 1994, M = 7.6; Sporer, 1992b, M = 5.5). Also, the rate of descriptors referring to the perpetrator’s face was higher (42%) than the rate found in the archival analysis by Sporer (1992a; 30%). Different from the findings by van Koppen and Lochun (1997), the majority of the facial descriptors in our study were correct (67%). Yet, again in concordance with the findings of the archival analysis by Sporer (1992a), most of the facial descriptors referred to hair style and color (60%). Although description accuracy coefficients have been criticized in the past (Sporer, 1996; Wells, 1985) because they do not differentiate between more and less complete descriptions, we believe that the minimum number of total descriptors provided by our participants was sufficiently large (Min = 7) to make analyses with description accuracy appropriate.

Even though the accuracy of a person description cannot be assessed in actual cases (see Wells, 1985) it has been listed as one of the criteria to be considered in the evaluation of identification evidence by the U.S. Supreme Court (Neil v. Biggers, 1972). Yet, analyses of description accuracy can be useful in order to find out more about the decision processes underlying identification decisions with regard to target descriptions. In line with previous studies (e.g., Meissner et al., 2001; Meissner 2002; see also the meta-analysis by Meissner, Sporer, & Susa, in press), we found a positive association between description accuracy and identification accuracy along with a negative association between the number of false details and identification accuracy. However, here this effect was limited to the description only condition. These results are neither in line with the RBI approach nor with the TIPS approach. According to the RBI approach, associations for both description groups, not just for one, would have been expected. According to the TIPS approach, one would expect the observed relationships for rereaders but not describers only. At first sight it seems puzzling, that descriptions that are reread prior to the identification task should have less impact on identification performance than descriptions that are not reread. However, within rereaders, description accuracy was positively associated with choosing, supporting the idea that lineup rejections might be due to the perceived inadequacy of the description (Clare & Lewandowsky, 2004). Apparently, rereaders somehow became aware of the accuracy or inaccuracy of their descriptions. Comparing the observed associations with rereaders and with describers only, it seems that the positive association between description accuracy and identification accuracy within describers only changed to an association with choosing behavior for rereaders. Possibly, the inclusion of two repeated descriptions--once shortly after observing the target, and again before the identification after the 1-week delay--may have resulted in a correlation between identification accuracy and number of descriptors as reported by Meissner (2002, Experiment 2).

When we consider the description-only group by itself, the present study showed higher correlations between various measures of description accuracy and identification accuracy than previous studies, thus questioning findings that have reported null-findings. We attribute these higher correlations to our careful attempts to operationalize and measure face and person descriptions. In Meissner, Sporer, and Susa's (in press) meta-analysis, higher correlations were also observed when substantial effort was invested into measuring description quality more rigorously. Correlations were also found to be higher when there were longer description-identification delays than when the recognition test was conducted shortly after the description phase.

Although much care was taken to ensure ecological validity in the present study in form of a complex filmed scenario of 6 min length, a 1-week delay, the use of TP and TA lineups, and the description procedure, generalizability of the findings may be limited as we used only a single target. Stimulus sampling through the use of multiple targets is desirable for two reasons: First, to ensure stimulus generalizability and to allow for internal replication (Wells & Windschitl, 1999). Second, previous research discussed the role of facial distinctiveness/typicality not only regarding face recognition performance (e.g., Shapiro & Penrod, 1986) but also concerning the relationship between target descriptions and identification (Wells, 1985). This issue should be addressed in future research.

In summary, our results help to understand the decision processes underlying eyewitness identifications in connection with person descriptions. First, a retention interval of one week between description and identification appears to be sufficient to eliminate the negative effects of person descriptions on identification accuracy as observed in other studies (Dodson et al., 1997; Fallshore & Schooler, 1995; Ryan & Schooler, 1998; Schooler et al., 1996; Schooler & Engstler-Schooler, 1990). For police practice, this means that no change in proceedings is necessary in most cases as the presentation of a lineup rarely happens within less than one week after the description (Behrman & Richards, 2005; Valentine et al., 2003). Second, a rereading task did not lead to a VOE, but to a criterion shift (Clare & Lewandowsky, 2004), however without affecting identification accuracy. Furthermore, an association between description accuracy and choosing emerged for rereaders. Interestingly, rereaders seemed to be aware of the quality of their descriptions and this affected their choosing behavior. In contrast, an association between description accuracy and identification accuracy materialized for describers only. Apparently, when describers reread their own descriptions, the memory effect of the description which lead bad describers to be bad identifiers shifts to a decision criterion effect which lead bad describers to become nonchoosers. Whether one or which of these two effects is (more) desirable is left for further studies to investigate. Specifically, the (non-)existence of an effect of rereading on identification (in-)accuracy needs to be reexamined before further conclusions can be drawn.

References

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.

Brigham, J. C., & Pressley, M. (1988). Cognitive monitoring and strategy choice in younger and older adults. Psychology and Aging, 3, 249-257.

Brown, C., & Lloyd-Jones, T. J. (2002). Verbal overshadowing in a multiple face presentation paradigm: Effects of description instruction. Applied Cognitive Psychology, 16, 873-885.

Brown, C. & Lloyd-Jones, T. J. (2003). Verbal overshadowing of multiple face and car recognition: Effects of within- versus across-category verbal descriptions. Applied Cognitive Psychology,17, 183-201.

Clifford, B. R. (2003). The verbal overshadowing effect: In search of a chimera. In M. Vanderhallen, G. Verwaeke, P. J. van Koppen, & J. Goethals (Eds.), Much ado about crime. Chapters on psychology and law (pp. 151-162). Brussel: Politeia.

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

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

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.

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 Behaviour, 1, 113-134.

Dodson, C. S., Johnson, M. K., & Schooler, J. W. (1997). The verbal overshadowing effect: Why descriptions impair face recognition. Memory and Cognition, 25, 129-139.

Fallshore, M., & Schooler, J. W. (1995). The verbal vulnerability of perceptual expertise. Journal of Experimental Psychology: Learning, Memory, and Cognition, 21, 1608-1623.

Finger, K., & Pezdek, K. (1999). The effect of the cognitive interview on face recognition accuracy: Release from verbal overshadowing. Journal of Applied Psychology, 84, 340-348.

Finger, K. (2002). Mazes and music: Using perceptual processing to release verbal overshadowing. Applied Cognitive Psychology, 16, 887-896.

Franzen, S., & Sporer, S. L. (1994a). Personenverwechslungen durch irreführende Rekonstruktionsbilder: Zum Einfluss nachträglicher Informationen und der Wiederherstellung des Wahrnehmungskontextes. [Person mixups as a function of misleading composites: On the influence of postevent information and context reinstatement]. In S. L. Sporer & D. Meurer (Eds.), Die Beeinflussbarkeit von Zeugenaussagen (pp. 207-236). Marburg: N.G. Elwert.

Franzen, S., & Sporer, S. L. (1994b). Personenverwechslungen und Möglichkeiten ihrer Vermeidung: Können Augenzeugen durch Visualisierung gegen den Einfluss von irreführenden Rekonstruktionsbildern immunisiert werden? [Person mixups and possible countermeasures: Can eyewitnesses be inoculated against misleading composites through visualization?]. In S. L. Sporer & D. Meurer (Eds.), Die Beeinflussbarkeit von Zeugenaussagen (pp. 237-283). Marburg: N.G. Elwert.

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

Goldstein, A. G., Johnson, K. S., & Chance, J. (1979). Does fluency of face description imply superior face recognition? Bulletin of the Psychonomic Society, 13, 15-18.

Grass, E., & Sporer, S. L. (1991, March). Zur Vorhersage von Identifizierungsleistungen durch weitere Aussagen von Zeugen. [Correct or false? Post-dicting eyewitness identification accuracy from verbal statements]. Paper presented at the 33rd annual meeting of Experimental Psychologists in Giessen.

Koriat, A., & Goldsmith, M. (1994). Memory in naturalistic and laboratory contexts: Distinguishing the accuracy-oriented and quantity-oriented approaches to memory assessment. Journal of Experimental Psychology: General, 123, 297-315.

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. Psychological Review, 103, 490-517.

Koriat, A., Goldsmith, M., & Pansky, A. (2000). Toward a psychology of memory accuracy. Annual Review of Psychology, 51, 481-537.

Krafka, C., & Penrod, S. D. (1985). Reinstatement of context in a field experiment on eyewitness identification. Journal of Personality and Social Psychology, 49, 58-69.

MacLin, O. H., Tapscott, R. L., & Malpass, R. S. (2002). The development of a computer system to collect descriptions of culprits. Applied Cognitive Psychology, 16, 937-945.

Meissner, C. A. (2002). Applied aspects of the instructional bias effect in verbal overshadowing. Applied Cognitive Psychology, 16, 911-928.

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

Meissner, C. A., Brigham, J. C., & Kelley, C. M. (2001). The influence of retrieval processes in verbal overshadowing. Memory and Cognition, 29, 176-186.

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 (pp 3-34). Lawrence Erlbaum & Associates.

Meissner, C. A., Sporer, S. L., & Susa, K. J. (in press). A theoretical review and meta-analysis of the description-identification relationship in memory for faces. European Journal of Cognitive Psychology.

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

Memon, A., & Rose, R. (2002). Identification abilities of children: Does a verbal description hurt face recognition? Psychology, Crime, & Law, 8, 229-242.

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

Pigott, M., & Brigham, J. C. (1985). Relationship between accuracy of prior description and facial recognition. Journal of Applied Psychology, 70, 547-555.

Pigott, M. A., Brigham, J. C., & Bothwell, R. K. (1990). A field study of the relationship between quality of eyewitnesses' descriptions and identification accuracy. Journal of Police Science and Administration, 17, 84-88.

Ryan, R. S., & Schooler, J. W. (1998). Whom do words hurt? Individual differences in susceptibility to verbal overshadowing. Applied Cognitive Psychology, 12, 105-125.

Schooler, J. W. (2002). Verbalization produces a transfer inappropriate processing shift. Applied Cognitive Psychology, 16, 989-997.

Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. Cognitive Psychology, 22, 36-71.

Schooler, J. W., Fiore, S. M., & Brandimonte, M. A. (1997). At a loss from words: Verbal overshadowing of perceptual memories. In D. L. Medin (Ed.), The psychology of leaming and motivation (pp. 293-334). Academic Press: San Diego, CA.

Schooler, J. W., Ryan, R. S., & Reder, L. (1996). The costs and benefits of verbally rehearsing memory for faces. In D. Herrmann, M. Johnson, C. McEvoy, C. Hertzog, & P. Hertel (Eds.), Basic and applied memory: New findings (pp. 51-65). Hillsdale, NJ: Erlbaum.

Schraw, G. (1998). Promoting general metacognitive awareness. Instructional Science, 26, 113-125.

Sporer, S. L. (1992a, March). An archival analysis of person descriptions. Paper presented at the biennial meeting of the American Psychology-Law Society in San Diego, CA.

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

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

Sporer, S. L. (in press). Person descriptions as retrieval cues: Do they really help? Psychology, Crime, & Law.

SuperLab Pro (v. 1.75). [Computer software]. Cedrus Corporation. Phoenix, Arizona.

Tunnicliff, J. L., & Clark, S. E. (1999, unpublished raw data). Cited in Meissner, C. A., & Brigham, J. C. (2001). A meta-analysis of the verbal overshadowing effect in face identification. Applied Cognitive Psychology, 15, 603-616.

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

van Koppen, P. J., & Lochun, S. K. (1997). Portraying perpetrators: The validity of offender descriptions by witnesses. Law and Human Behavior, 21, 661-685.

Wells, G. (1985). Verbal descriptions of faces from memory: Are they diagnostic of identification accuracy? Journal of Applied Psychology, 70, 619-626.

Wogalter, M. S. (1996). Describing faces from memory: Accuracy and effects on subsequent recognition performance. Proceedings of the Human Factors Society, 40, 536-540.

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

Appendix

Open-ended questions asked after free report was given

1. How old do you think the culprit was?
2. How tall was the culprit in cm?
3. Describe the culprit's build!
4. Describe the culprit's clothing!
5. Describe the culprit's hair color!
6. Describe the culprit's hairdo!
7. Describe the culprit's face shape!
8. Which of the culprit's special features caught your eye?
9. Did the culprit wear headgear? If yes, what kind?
10. Did the culprit wear glasses? If yes, what did they look like?
11. Did the culprit have a beard? If yes, what did it look like?
12. Did the culprit speak in a dialect or did he have an accent? If yes, which?

EXPERIMENT 2

Post-Decision Confidence, Decision Time, and Self-Reported Decision Processes as Postdictors of Identification Accuracy

Eyewitness identifications are of immense importance for investigating and prosecuting crimes. Research on the psychology of eyewitness identification began in the late 1970s and has since identified numerous estimator and system variables (Wells, 1978) that can have an influence on the accuracy of identification decisions. Estimator variables can be further broken down into situational variables which can only be explored post hoc (e.g., lighting conditions) and assessment variables (Sporer, 1993) that may be used to assess individual witnesses' decision making processes (e.g., confidence and decision time). In contrast, system or control variables are those over which the criminal justice system normally has control. These include lineup test factors such as instruction to witnesses as well as lineup composition.

Although decision time and post-decision confidence have been combined successfully in the past (Sporer, 1992; Weber, Brewer, Wells, Semmler, & Keast, 2004), most research on assessment variables, or postdictors, focused on one or another of these variables in isolation. It is our aim to determine the usefulness of combinations of post-decision confidence, decision time, and self-reported decision processes for distinguishing between accurate and inaccurate identification decisions. Hereby, we also hope to learn more about the underlying decision processes. A particular strength of the present study is that special care was taken to ensure ecological validity by using a long, complex stimulus film and a retention interval of one week.

Response Latencies as Markers of Accuracy

A series of studies have consistently demonstrated a negative relationship between response time and identification accuracy for positive identifications (choosers) but not for lineup rejections (nonchoosers; Dunning & Stern, 1994; Dunning & Perretta, 2002; Kneller, Memon, & Stevenage, 2001; Smith, R. C. L. Lindsay, & Pryke, 2000; Sporer, 1992, 1993, 1994; Weber et al., 2004). Dunning and Perretta (2002) attempted to establish an empirical, absolute time boundary that best discriminates between correct and false choices.

Across four studies, they found that this time boundary was located between 10 and 12 s. Witnesses who made their identification within 10 s were largely accurate, with an average accuracy rate across the four studies of 87%. Slower choosers were accurate in about 50% of the cases. Dunning and Perretta (2002) concluded that time boundaries are invariant across experimental conditions.

However, other studies failed to replicate the 10-12 s "rule" (Weber et al., 2004; Brewer, Caon, Todd, and Weber, 2006). Weber et al.'s (2004) peaks varied from 5 s to 29 s, and for only two of six samples did the identified peaks include the 10-12 s window postulated by Dunning and Perretta (2002). Additionally, the high accuracy rates obtained previously for fast choosers (Dunning & Perretta, 2002) could not be replicated but stayed as low as 67%. When confidence was taken into account in addition to response time, higher classification rates were obtained within the 10 s window (88%) for rather confident witnesses (confidence estimates of 90% or 100%) than for less confident witnesses (confidence estimates of 0%-80%; 54%). Similar results were found for the empirically established time boundary in this study (84% vs. 63%). These data suggest that response latencies should not be considered in isolation but jointly with other assessment variables, in particular, confidence.

Brewer et al. (2006) provided another test of the 10-12 s rule. In two experiments retention interval (0 vs. 15 vs. 30 minutes) and nominal lineup size (4 vs. 8 vs. 12 persons) were manipulated separately. The experiments replicated the finding that correct choosers made their decisions faster than incorrect choosers and no differences were found for non-choosers. The manipulation of retention interval and lineup size both influenced response times and time boundaries, with short retention intervals and small lineup size leading to faster decisions and earlier optimum time boundaries than longer retention intervals and larger nominal lineup size. None of the identified time boundaries was within the 10-12 s interval postulated by Dunning and Perretta (2002). The different retention interval conditions of Experiment 1 produced relatively low pre-time boundary accuracy of no more than 55%. In contrast, Experiment 2 produced high pre-time boundary accuracy rates in a number of conditions. Thus, Dunning and Perretta's (2002) claim for an invariant time boundary was contradicted by the results of Weber et al. (2004) and Brewer et al. (2006) who failed to find a consistent time boundary across conditions. Brewer et al. (2006) explained their results with reference to cognitive research on recognition memory (Ratcliff, 1978; Vickers & Packer, 1982) which showed that similarity, that is, discriminability between distractors and a target, is an important predictor of recognition performance where low discriminability leads to an increase of latencies. This account calls into question the idea of an absolute optimum time boundary and points to the possible influence that estimator and control variables have on response times.

To sum up, past research established a negative relationship between identification accuracy and response time for choosers, but not nonchoosers. We expected to replicate these findings in the present study. Additionally, we determined the empirical time boundary for our data in order to contribute to the current discussion on absolute time boundaries (Dunning & Peretta, 2002; Weber et al., 2004).

Confidence-Accuracy Relationship

Meta-analyses reported small to moderate correlations between post-decision confidence and accuracy of r = .25, (Bothwell, Deffenbacher, & Brigham, 1987) and r = .28 (Sporer, Penrod, Read, & Cutler, 1995). However, practically relevant conditions were identified under which die CA relationship is more reliable. In line with Deffenbacher’s (1980) optimality hypothesis, Bothwell et al. (1987) found that longer target exposure duration (longer than approximately 75 s) lead to a higher CA relation. Sporer et al. (1995) conducted separate analyses for choosers and nonchoosers and found that the CA relationship was considerably higher for choosers (r = .37) than for nonchoosers (r = .12).

Furthermore, work by D. S. Lindsay and Read (D. S. Lindsay, Nilsen, & Read, 2000; D. S. Lindsay, Read, & Sharma, 1998; Read, D. S. Lindsay, & Nicholls, 1998) provided evidence that the CA relationship (following immediately after the identification, not in the courtroom) can be more substantial when tested across heterogeneous conditions rather than in the homogenous conditions commonly used in previous studies. For example, D. S. Lindsay et al. (2000) obtained highly significant CA correlations of r = .55 (whole sample) and r = .68 (choosers) when tested under heterogeneous conditions (i.e., combining good and poor viewing conditions) whereas homogeneous (i.e., computing data for good and poor viewing conditions separately) conditions lead to lower correlations comparable to those observed in other studies. The rationale behind this approach is that under homogeneous conditions differences in witnesses’ ability to identify the target are only attributable to individual differences in attention, motivation, etc. as well as unintended procedural variations. Hence, in a given condition, participants who make a correct decision in the identification task are likely to have only a slightly better ability to identify the target than do participants who respond incorrectly, thus producing a weak CA correlation.

We expected to replicate the positive relationship between identification accuracy and post-decision confidence for choosers in the present study. No such relation was expected for nonchoosers. Additionally, rather than setting an arbitrary decision rule for more or less confident witnesses, we attempted to establish an empirical confidence boundary that best discriminates between correct and false choices analogous to Dunning and Perretta's (2002) time boundary. To our knowledge, this procedure has not been applied to post-decision confidence in any previous study.

Self-Reported Decision Processes

Across four studies, Dunning and Stern (1994) asked their witnesses to provide information about their decision processes. They found that eyewitnesses reporting automatic recognition (e.g., "the face just popped out at me") were more likely to be accurate in identification performance than those reporting eliminative decision processes (e.g., "I compared the photos to each other to narrow the choices"). These results are in line with Wells' (1984) distinction between relative and absolute judgment strategies, with erroneous witnesses tending to make use of relative decision strategies and accurate witnesses tending to rely on absolute decision strategies.

Some studies supported Dunning and Stern's findings (Caputo & Dunning, 2005; Kneller et al., 2001), whereas another study failed to find differences between accurate and inaccurate decisions via self-reported decision processes (Brewer, Gordon, & Bond, 2000). Furthermore, Kneller et al. (2001) did show the effect when analyzing simultaneous and sequential lineups together, but not when analyzing them separately. Some studies used target-present lineups only (Brewer et al., 2000; Dunning & Stern, 1994, Experiments 1 to 3), which does not allow for an appropriate test of the chooser vs. nonchooser distinction. Clearly, further research is needed in order to clarify the findings.

In the current study, we explored self-reported decision processes of choosers, combining the procedures that differed across previous studies (Dunning & Stern, 1994; Kneller et al., 2001; see Table 1). We expected identification accuracy to be positively associated with automatic recognition and to be negatively associated with eliminative decision strategies. As automatic recognition is described as a fast process, we expected it to be negatively associated with decision time. Eliminative decision strategies, on the other hand, are considered as time consuming and deliberate, and were therefore expected to be positively associated with decision time. Furthermore, if eyewitnesses derive their post-decision confidence ratings from metacognitive processes regarding the identification process, then post-decision confidence should be negatively associated with eliminative decision strategies, and positively associated with automatic recognition processes. Additionally, witnesses might use the time taken for an identification decision as a cue for the ease of the decision. Therefore, decision time and post-decision confidence were expected to be negatively associated.

Previous research has consistently failed to show associations of postdictors and identification accuracy for nonchoosers. Nevertheless, the assessment of nonchoosers' decisions is valuable for exonerating innocent suspects. We addressed this issue by developing a different set of decision process items for nonchoosers in order to explore their usefulness for postdicting identification accuracy of nonchoosers. The items belonged to one of two groups that we termed memory match strategy and comparative judgment strategy (see Table 2). Memory match strategy items were designed to describe the failure to find a

Table 1

Choosers' Decision Process Items with Factor Loadings for Principal Components Analysis (N = 70)

Items	Eliminative decision strategy	Automatic recognition
I just recognized him, I cannot explain why[ab]	.25	.48
His face just popped out at me[a]	-.14	**.82**
When I first saw the person I identified in the lineup, I instantly knew that it was him	-.52	**.70**
I compared the photos to each other in order to narrow the choices[a]	**.76**	-.02
I first eliminated the ones definitely not him, then chose among the rest[a]	**.68**	-.06
He was the closest person to what I remember, but not exact[a]	**.66**	-.03

Note. Loadings of |.59| or larger were significant (see Stevens, 2002). The Item that fell below this cut-off was deleted. [a]From Dunning & Stern (1994). [b]Modified from Kneller et al. (2001).

Table 2

Nonchoosers' Decision Process Items with Factor Loadings for Principal Components Analysis (N = 122)

Items	Memory match strategy	Comparative judgment strategy
I couldn't identify anyone, as the target was not in the lineup.	**.87**	-.14
Had the target been present in the lineup, I probably would have recognized him.	**.85**	-.18
I tried to match each photograph to my memory of the target person's face, but none of the faces matched this image. [a]	**.62**	.15
I compared each face shown with the others to narrow down the selection.[a]	.14	**.70**
I first eliminated those pictures that definitely weren't him but could eventually not decide on one of the other pictures.	-.04	**.84**
One of the faces resembled the target closest, but I could not decide on him.	-.15	**.68**
Because the faces were quite similar to each other, I could not make a decision.	-.41	**.49**

Note. Loadings of |.46| or larger were significant (see Stevens, 2002). [a]Modified from Kneller et al. (2001). All other items were new items.

match between the members of the lineup and the memory for the target. Comparative judgment strategy items describe the process of comparing the foils to each other, rather than with the memory of the target. We predicted that the memory match strategy would be positively associated with identification accuracy whereas the comparative judgment strategy would be negatively associated with identification accuracy. Furthermore, we expected memory match to be negatively associated with decision time and positively associated with post-decision confidence. The associations were expected to be opposite in sign for comparative judgment strategies.

To sum up, the major aim of the present research was to combine three postdictors of identification accuracy, namely decision time, post-decision confidence, and self-reported decision processes. We expected that the combination of the postdictors would lead to a higher number of correct classifications than each of them by itself. Additionally we wished to test the usefulness of self-reported decision process strategies again for choosers, but particularly also for nonchoosers. Finally, we wanted to gain more insight into the differing processes underlying choices and nonchoices by scrutinizing the interrelationships between confidence, decision time, and self-reported decision strategies.

Method

Participants

One-hundred-and-ninety-two (96 male, 96 females; age 19 to 52, Mdn = 24 years) individuals completed this experiment. Most participants were psychology majors (62%) who received course credit for their participation. Other participants were students with other areas of study and persons of various occupations. They were randomly assigned to the experimental conditions and tested individually. Participants were tested at the Free University of Berlin and at the Justus Liebig University of Giessen, Germany.

Materials

Stimulus Film

The stimulus film was taken from an earlier study by Franzen and Sporer (1994). The film showed the theft of an expensive pair of sunglasses in an optometrist’s store. Altogether five amateur actors (one woman and four men) participated in the film, which lasted 6 minutes and 30 seconds. The target person could be seen for 18 seconds. A close-up showing the target's head and shoulders in half-profile lasted for about 2 seconds. For the remaining time the target was filmed from a distance of several meters, where his head and whole torso could be seen. The actual theft took about 30 seconds (looking at the sunglasses on the rack, taking a pair and putting it in the pocket).

Photo Lineup

Each lineup consisted of six frontal 6 x 9 cm photographs simultaneously displayed on the computer screen at a color depth of 16.7 million colors, that is, 32 bit and a resolution of 1024 x 768 pixels, depicting six male individuals who all fit the general description of the target person as determined by a pilot study (effective size = 4.7). The men all wore the same sweater (different from clothing in the film) and each picture had been taken in the same windowless room with the same illumination and in front of the same wall.

The photos were arranged in two rows of three pictures each. For half of the participants, the target photograph was present (TP), for the other half it was replaced with an innocent foil (target absent, TA). Target position as well as distractor position were completely balanced to appear at any of the six positions an equal number of times.

Procedure

The experiment was conducted on two different days. For all parts of the experiment the participants were given thorough instructions on the computer screen how to respond to the questions and which keys to use. All instructions and the lineup presentation were programmed with SuperLab 1.75 (www.cedrus.com). Participants were informed that the experiment dealt with witness statements, but the topic of person identification was not mentioned on the first day of the experiment. Each participant was tested individually. Participants were told that they would be shown a short video and then were asked to watch the film closely and pay attention to every detail. After viewing the film participants completed a 30-minutes filler task consisting of 40 general knowledge questions. Participants were then asked to give a description of the witnessed crime and of the culprit.

The lineup procedure was scheduled one week later. Prior to lineup presentation, participants indicated their pre-decision confidence regarding the accuracy of their identification decision on an 11-point scale ranging from 0% to 100% (with intervals marked in 10% steps: 0%, 10%, ..., 100%). Subsequently, participants were asked to identify the culprit from the lineup. Participants were advised that the culprit might or might not be present. Decision time was measured automatically via SuperLab.[1] Subsequently, post-decision confidence ratings concerning the identification decision were again assessed on an 11-point scale. Hereafter, the decision process items were rated on 7-point Likert scales

[1] After the identification task, choosers were asked to make a Remember-Know (-Familiar) judgment (Tulving, 1985). However, the operationalization of the variable failed to elicit a satisfactory number of Know answers, thus not allowing any meaningful comparisons. Therefore, the results will not be discussed.

ranging from 1 (not at all) to 7 (absolutely so). Depending on a participant's response to the lineup (choice vs. no choice), they received different questions regarding their decision process (see Tables 1 and 2).

Results

An alpha level of .05 was used for all inferential analyses. Effect sizes, that is, Cohen’s (1988) f, are reported for ANOVAs for interactions and for tests with more than 1 df in the numerator; d is reported for comparisons of two means, and phi is reported for nonparametric analyses of 2 x 2 contingency tables.

We first present results for identification accuracy. Subsequently, we look at the relationships between identification accuracy and the postdictors, first for each of them individually, then their combination.

Identification Accuracy

Altogether, 44.8% of the identification decisions were correct. Identification accuracy differed significantly for choosers (n = 70) and nonchoosers (n = 122), $chi^2(1, N = 192) = 9.75$, $p = .001$, phi = 0.23. Nonchoosers were accurate in their decision more often (53.3%) than choosers (30.0%). Thus, not surprisingly, identification accuracy for target-absent lineups was higher (67.7%) than for target-present lineups (21.9%), $chi^2(1, N = 192) = 40.78$, $p < .001$, phi = 0.46.

Decision Time

Due to a significantly positively skewed response latency distribution, one extreme value was winsorized on either side of the distribution (Winer, 1971) and inferential analyses were conducted on log-transformed data. Means and standard deviations are reported as backtransformed values.

A 2 x 2 ANOVA with choice (choosers vs. nonchoosers) and decision outcome (correct vs. incorrect) as classification variables and decision times as dependent variable revealed a significant main effect for decision outcome, $F(1, 188) = 9.25$, $p = .003$, $d = -0.44$. Accurate witnesses made their decision faster ($M = 16.5$ s) than inaccurate witnesses ($M = 22.7$ s). Neither the main effect for choice, $F(1, 188) = 0.24$, $p = .623$, $d = 0.07$, nor the interaction, $F(1, 188) = 0.56$, $p = .455$, $f = .05$, were significant. Means and standard deviations for identification accuracy and all postdictor variables and their intercorrelations are shown in Tables 3 to 5 for choosers, nonchoosers, and the total sample, respectively.

The correlation between decision time and identification accuracy was significantly negative for choosers, $r(68) = -.29$, $p = .017$, and marginally significant for nonchoosers, $r(120) = -.18$, $p = .054$. Therefore, time boundary analyses were performed for choosers only. As previously done by Dunning and Perretta (2002) and Weber et al. (2004), we computed a series of chi^2 statistics on the 2 (accuracy: correct vs. incorrect) x 2 (time boundary: faster or equal vs. slower) contingency tables with the time boundaries set at each integer value from 1 s (i.e., 1 s, 2 s, 3 s, etc.). The decision time that produced the greatest chi^2 value was identified as the time boundary that optimally discriminated between correct and incorrect decisions.

The plot of chi^2 values by time boundaries is presented in Figure 1. The optimum time boundary for choosers was at 18 s, $chi^2(1, N = 70) = 8.96$, $p = .006$, $phi = 0.36$, with another peak at 20-21 s, $chi^2(1, N = 70) = 8.96$, $p = .004$, $phi = 0.36$. The proportion of correct decisions made within the 18 s time boundary was 52.0%. The proportion of wrong decisions made after 18 s was 82.2%. The proportion of correct decisions made below the 20-21 s time boundary was 48.4%. The proportion of wrong decisions made after 21 s was 86.6%.

Following Dunning and Perretta (2002), we additionally worked backward from long time windows in order to find a boundary beyond which most (i.e., at least 85%) witnesses were inaccurate. We found that 91.7% (11 out of 12) of the decisions were inaccurate when they were made after 35 s had passed.

Confidence

A 2 x 2 ANOVA with choice (choosers vs. nonchoosers) and decision outcome (correct vs. incorrect) as classification variables and pre-decision confidence as dependent variable revealed a significant main effect for choice, $F(1, 188) = 7.47$, $p = .007$, $d = -0.41$, but not for accuracy, $F(1, 188) = 0.08$, $p = .772$, $d = 0.04$. Nonchoosers showed higher pre-decision confidence ratings (66.0%) than choosers (57.6%). The interaction did not reach significance, $F(1, 188) = 3.54$, $p = .061$, $f = .11$.

The correlation between pre-decision confidence and identification accuracy was significantly negative for nonchoosers, $r(120) = -.18$, $p = .047$, but non-significant for choosers, $r(68) = .11$, $p = .375$, and the whole sample, $r(190) = -.02$, $p = .794$.

Table 3

Means, Standard Deviations, and Intercorrelations of Postdictors of Identification Accuracy for Choosers (N = 70)

	M	SD	Pre-decision confidence	Post-decision	Decision time	Automatic recognition	Eliminative strategy
Accuracy	30.0%	46.2	.11	.28*	-.29*	.17	-.06
Pre-decision confidence	56.6%	21.0		.41**	.24*	.48**	.08
Post-decision confidence	50.7%	24.3			-.21	.73**	-.25*
Decision time[a]	21.5 s	1.9				-.15	.43**
Automatic recognition[b]	5.7	1.1					-.33**
Eliminative decision strategy[b]	5.1	1.4					

*p < .05, two-tailed. **p < .01, two-tailed.

[a]All analyses were conducted with log-transformed values. The mean and standard deviations of decision times were backtransformed from log.

[b] Means of items on 7-point Likert scales from Table 1.

Table 4

Means, Standard Deviations, and Intercorrelations of Postdictors of Identification Accuracy for Nonchoosers (N = 122)

	M	SD	Pre-decision confidence	Post-decision	Decision time	Memory match	Comparative judgment
Accuracy	53.3%	50.1	-.18*	.09	-.18	.04	-.11
Pre-decision confidence	65.7%	18.6		.42**	.12	.43**	-.12
Post-decision confidence	65.1%	25.9			-.52**	.75**	-.33**
Decision time[a]	18.7 s	2.0				-.44**	.42**
Memory match strategy[b]	5.4	1.2					-.19*
Comparative judgment strategy[b]	4.0	1.5					

*p < .05, two-tailed. **p < .01, two-tailed.

[a]All analyses were conducted with log-transformed values. The mean and standard deviations of decision times were backtransformed from log.

[b] Means of items on 7-point Likert scales from Table 2.

Table 5

Means, Standard Deviations, and Intercorrelations of Postdictors of Identification Accuracy for the Whole Sample (N = 192)

	M	SD	Pre-decision confidence	Post-decision confidence	Decision time
Accuracy	44.8%	49.9	-.02	.20*	-.23*
Pre-decision confidence	62.4%	19.9		.45**	.01
Post-decision confidence	59.8%	26.2			-.43**
Decision time[a]	19.7 s	2.0			

*$p < .05$, two-tailed. **$p < .01$, two-tailed.

[a]All analyses were conducted with log-transformed values. The mean and standard deviations of decision times were backtransformed from log.

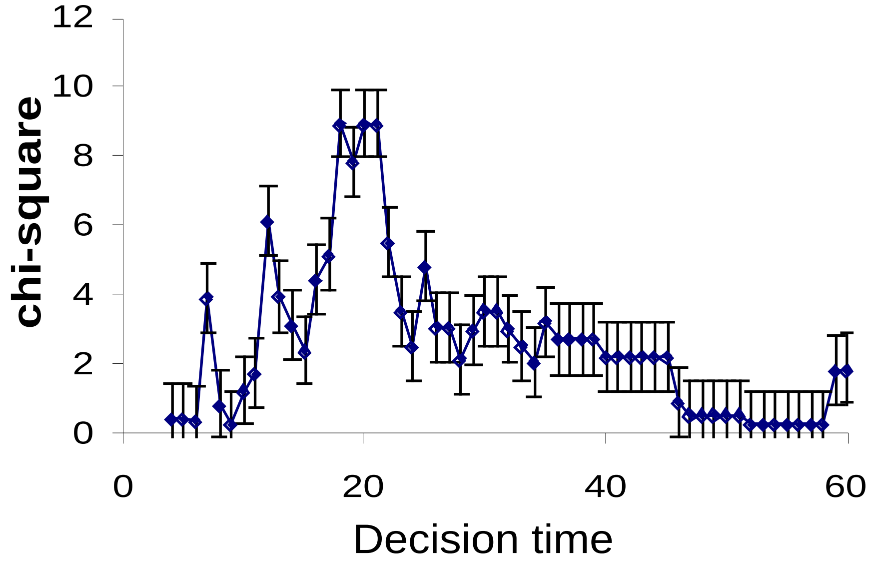


Figure 1. Plot of chi^2 values (and 95% CI) by time boundaries.

A 2 x 2 ANOVA with choice (choosers vs. nonchoosers) and decision outcome (correct vs. incorrect) as classification variables and post-decision confidence as dependent variable revealed significant main effects for decision outcome, $F(1, 188) = 5.83$, $p = .017$, $d = 0.35$, and choosing, $F(1, 188) = 8.04$, $p = .005$, $d = -0.41$. Accurate participants made higher confidence ratings ($M = 64.1\%$) than inaccurate participants ($M = 54.5\%$). Choosers were significantly less confident about their decision ($M = 53.6\%$) than nonchoosers ($M = 64.9\%$). The interaction did not reach significance, $F(1, 188) = 1.58$, $p = .210$, $f = .09$.

As expected, the correlation between post-decision confidence and identification accuracy was significant for choosers, $r(68) = .28$, $p = .020$, but not for nonchoosers, $r(120) = .09$, $p = .329$.

Analogous to the time boundary analyses performed by Dunning and Perretta (2002), chi^2 analyses on 2 (accuracy: correct vs. incorrect) x 2 (post-decision confidence boundary: more vs. less or equally) contingency tables were computed. The plot of chi^2 values by post-decision confidence boundaries is presented in Figure 2. Curiously, the optimum post-decision confidence boundary here is at 50%, chi^2(1, N = 70) = 8.27, p = .007, phi = .34. Of the 42 participants who were 50% or more confident about their choice, 18 (42.9%) were correct. Participants indicating lower post-decision confidence were wrong in 25 out of 28 cases (89.3%).

When we adapted Weber et al.'s (2004) cut off point of 90% post-decision confidence, 3 out of the 7 choosers (42.9%) who were 90-100% confident were correct. Participants indicating lower post-decision confidence were wrong in 45 out of 63 cases (71.4%).

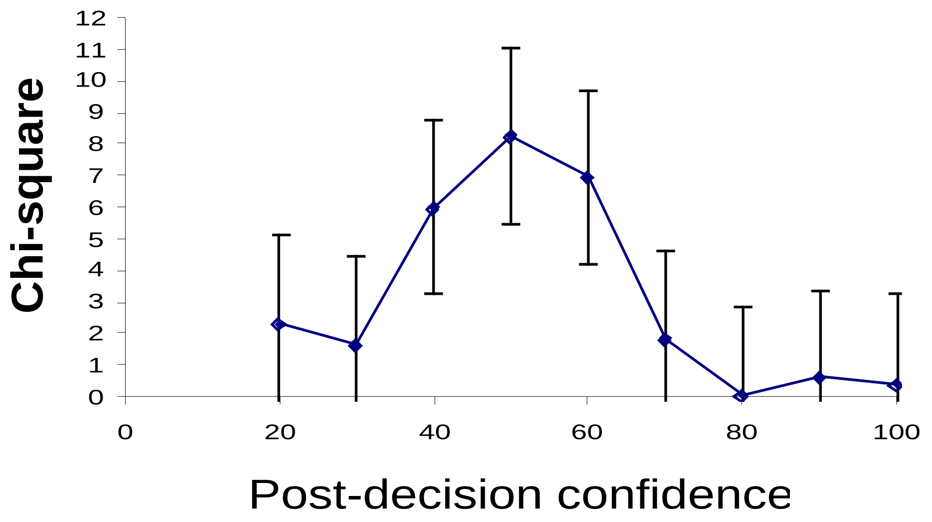


Figure 2. Plot of chi^2 values (and 95% CI) by post-decision confidence boundaries.

Contrary to our hypotheses, the association between decision time and post-decision confidence of choosers failed to reach significance, r(68) = -.21, p = .075, although the direction of the correlation was as expected. For nonchoosers, the association was highly reliable, r(120) = -.52, p < .001.

Self-Reported Decision Processes of Choosers

Following Dunning and Stern (1994) decision process items were combined based on factor analyses conducted separately for choosers and nonchoosers. For choosers, automatic recognition items loaded on one factor (Cronbach's alpha = .55), eliminative decision strategy items loaded on another factor (alpha = .68). Automatic recognition accounted for 41.92% of variance (eigen value = 2.1), eliminative decision strategy for 22.17% of variance (eigen value = 1.1). Loadings of the items are shown in Table 1. The two scales were computed by averaging the corresponding items into one score. Unexpectedly, correlational analyses showed no significant associations between identification accuracy and self-reports of automatic recognition, or eliminative decision strategy. When looking at target-present lineups only, the expected association was found for automatic recognition, $r(37) = .37$, $p = .021$, but not eliminative decision strategy, $r(37) = -.18$, $p = .255$, although the correlation was in the expected direction.[1]

Table 3 displays means and standard deviations of postdictors as well as their intercorrelations. As expected, post-decision confidence was positively associated with automatic recognition, $r(68) = .73$, $p < .001$, and negatively associated with eliminative decision strategies, $r(68) = -.25$, $p = .035$. Also as predicted, decision time and eliminative decision strategy were positively associated, $r(68) = .43$, $p < .001$. Contrary to our hypotheses, the association between decision time and automatic recognition failed to reach significance, although the direction of the correlation was as expected, $r(68) = -.15$, $p = .230$.

---

[1]No separate analyses for target-absent lineups are possible as choosers are always incorrect in target-absent lineups.

Self-Reported Decision Processes of Nonchoosers

The factor analysis of nonchoosers' self-reported decision process items revealed that the corresponding items loaded on two factors (see Table 2). Three items loaded on memory match strategy (alpha = .73), and four items loaded on comparative judgment strategy (alpha = .64). Memory match strategy accounted for 34.94% of variance (eigen value = 2.4), comparative judgment strategy for 22.96% of variance (eigen value = 1.6). Unexpectedly, correlational analyses of nonchoosers' self-reported decision processes showed no significant associations of identification accuracy with memory match strategies or comparative judgment strategies (see Table 4).[2]

Yet, as expected, memory match strategy was positively associated with post-decision confidence, $r(120) = .75$, $p < .001$, and negatively associated with decision time, $r(120) = -.44$, $p < .001$. Comparative judgment strategies were negatively associated with post-decision confidence, $r(120) = -.33$, $p < .001$, and positively associated with decision time, $r(120) = .42$, $p < .001$.[3]

Combination of Identification Postdictors

Due to the lack of associations between self-reported decision processes and identification accuracy, these assessment variables were not combined with the other two. Thus, decision time and post-decision confidence remained for combination. Table 6 shows identifica-

---

[2] No separate analyses for target-present and target-absent lineups were carried out, as nonchoosers are always correct in target-absent, and always incorrect in target-present lineups.

[3] As displayed in Table 2, the last item loaded relatively high on both factors, although the loading on memory match strategy was non-significant.

tion accuracy for combinations of fast vs. slow decisions and high vs. low post-decision confidence. Higher correct classification rates were obtained when confidence and decision time were combined than when decision time or confidence were looked at individually. As aforementioned, choosers who decided within the first 18 s were correct in 52.0% of the cases and choosers who were 50% or more confident about their choice were correct in 42.9% of the cases. Combining both measures, that is, fast (≤ 18 s) and confident choosers (≥ 50%), were correct in 64.7% of their decisions. Choosers who made their decisions slower than within 18 s and with less confidence erred in 95.0% of the cases.

Table 6

Identification Accuracy for Postdictor Combinations

	Decision time				Total	
Post-decision	Fast		Slow			
Confidence	n	% correct	n	% correct	n	% correct
0 - 40%	8	25.0	20	5.0	28	10.7
50 -100%	17	64.7	25	28.0	42	42.9
Total	25	52.0	45	17.8	70	30.0

Discussion

This study investigated the usefulness of post-decision confidence, decision time, and decision processes for postdicting identification accuracy. Relatively low accuracy rates were obtained for the total sample and particularly for choosers. Additionally, surprisingly few participants (36.5%) chose someone from the lineup although in 50% of our lineups the target was present. Thus, the identification task in the present study proved to be a difficult one. One possible, though speculative, explanation is that the one-week interval between witnessing the event and the identification task might have resulted in the low choosing rate, which resulted in the low hit rate in target-present lineups. Another potential explanation concerns the stimulus film, which was rather long compared to those used in most other studies, and the course of action which did not draw much attention to the culprit. Additionally, the culprit was seen for a relatively short time, and the lineup photo was shot on a different day (which is typical for real criminal cases), with the target looking different from the film with respect to his hairstyle and not as clean shaven.

As expected, eyewitness identification accuracy was positively associated with post-decision confidence and negatively associated with decision time for choosers, thus replicating the findings of Sporer et al.'s (1995) meta-analysis as well as Sporer's studies with decision times (Sporer, 1992, 1993, 1994). No such associations were found for nonchoosers (although the Choice by Outcome interaction was not significant). Nevertheless, there was a reliable negative association between pre-decision confidence and identification accuracy, indicating that nonchoosers low in pre-decision confidence were more likely to be accurate in the identification task. That is, those participants who rated their capability of recognizing the target as low and therefore made no choice, turned out to be more accurate than those who had more confidence in their capability of recognizing the target and made no choice. No association between pre-decision confidence and identification accuracy was found for choosers.

The time boundary analyses revealed a cut-off point at 18 s. Therefore, as was the case for Weber et al. (2004), the notion of an absolute time boundary at 10-12 s (Dunning & Perretta, 2002) was not supported by our data. Speculatively, the elevated difficulty of the identification task in the present study may have lead to longer decision times and therefore caused the relatively long optimum time boundary. At any rate, these data also question the invariance claim of the 10 to 12 s rule.

For the first time, we also established an optimum boundary for post-decision confidence ratings. At 50%, the optimum post-decision confidence boundary was surprisingly low. For example, Weber et al. (2004) previously adopted an arbitrary cut-off point of 90% post-decision confidence, indicating that the cut-off point for post-decision confidence would be expected to be much higher.

Contrary to our hypotheses, self-reported decision processes were not associated with identification accuracy for choosers or nonchoosers. For target-present lineups only, accurate choices were associated with automatic recognition whereas inaccurate choices tended to be associated with eliminative decision strategies. These findings are in line with other studies that demonstrated the postulated effects for target-present lineups (Caputo & Dunning, 2005, Experiment 1; Dunning & Stern, 1994, Experiments 1 to 3) although the negative association between identification accuracy and eliminative decision strategies failed to reach significance here. However, there is also one study that did not find the postulated associations for target-present lineups (Brewer et al., 2000). Nevertheless, others demonstrated the expected relationships for both target-present and target-absent lineups (Dunning & Stern, 1994, Experiment 4; Caputo & Dunning, 2005; Kneller et al., 2001). So far, the research results remain contradictory and further research is needed to shed more light on this issue.

Our finding, that automatic recognition is associated with identification accuracy for target-present lineups but not for the combination of target-present and target-absent lineups demonstrates once again that not including target-absent lineups in an identification study seriously questions the generalizability of the obtained findings. If a postdictor, specifically, self-reported decision processes, has postdictive value for target-present lineups only, its practical applicability is at best limited.

Some interesting intercorrelations between the postdictors were found. As hypothesized, choosers responded as if they were deriving their post-decision confidence from their decision process and to a limited degree from the time they took for the decision. Choosers who perceived the decision process as primarily automatic, indicated high post-decision confidence. Unexpectedly, automatic recognition was not significantly associated with faster decision making. As predicted, those choosers who described their decision process as mainly eliminative, indicated lower post-decision confidence and took longer for their decision. Shorter decision times tended to be associated with higher post-decision confidence for choosers.

Our attempt to identify a postdictor that is capable of postdicting identification accuracy of nonchoosers failed. Contrary to the hypothesis, neither the memory match strategy nor the comparative judgment strategy were postdictive of identification accuracy. The only other study that examined self-reported decision processes of nonchoosers (Kneller et al., 2001) did not separately analyze the data for choosers and nonchoosers. Thus, it seems that self-reported decision process, like post-decision confidence and decision time, is not a valid postdictor of identification accuracy. However, conclusions must not be made prematurely, as the present study is one of very few that assessed self-reported decision processes of nonchoosers.

Since the expected associations of self-reported decision processes and identification accuracy were not found, the intended combination of all three postdictors was not carried out. Combining only post-decision confidence and decision time (i.e., looking at fast and confident participants) led to 64.7% correct classifications, whereas considering decision time or post-decision confidence by themselves only led to 52.0% and 42.9% correct classifications, respectively. Although these numbers do not seem to be too impressive, the low overall accuracy of choosers in our sample (30.0%) must be considered. Furthermore, applying these postdictors showed a strong capability in marking inaccurate identification decisions. For example, how likely was it, that a slow and not so confident witness was correct? Choosers who decided slowly and with little confidence, that is, outside the established boundaries, erred in 95.0% of the cases.

In spite of the lack of an association between decision process and identification accuracy for nonchoosers, all hypotheses with regard to the associations between postdictors were supported by the data. Those nonchoosers, who employed primarily memory match strategies, showed high post-decision confidence and were fast decision makers. On the contrary, comparative judgment strategies were associated with low post-decision confidence and slower decision times. In addition, nonchoosers also responded as if they strongly relied on decision time when making their post-decision confidence ratings.

It is notable that the associations within the individual postdictors were not only obtained for choosers but also for nonchoosers. These results speak in favor of the idea that all witnesses, choosers and nonchoosers, derive their confidence and decision process reports from each other and from their decision time. This highlights that it is important to bring the different research areas and processes together. In combination they may be more useful in assessing eyewitness identification decisions.

Yet, postdictors failed to be associated with identification accuracy for nonchoosers. As Sporer et al. (1995) pointed out, the heterogeneity within the group of nonchoosers might be responsible for the lack of association. There may be different types of nonchoosers who reject a lineup for different reasons (e.g., because their memory is poor, or because they are very cautious not to indict somebody erroneously). Dividing nonchoosers into more homogeneous subgroups might be one way to address this issue.

References

Bothwell, R. K., Deffenbacher, K. A., & Brigham, J. C. (1987). Correlation of eyewitness accuracy and confidence: Optimality hypothesis revised. Journal of Applied Psychology, 72, 691-695.

Brewer, N., Caon, A., Todd, C., & Weber, N. (2006). Eyewitness identification accuracy and response latency. Law and Human Behavior, 30, 31-50.

Brewer, N., Gordon, M., & Bond, N. (2000). Effect of photoarray exposure duration on eyewitness identification accuracy and processing strategy. Psychology, Crime and Law, 6, 21-32.

Caputo, D. D., & Dunning, D. (2005). Eyewitness accuracy and decision processes using sequential lineups. Manuscript submitted for publication.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Deffenbacher, K. A. (1980). Eyewitness accuracy and confidence. Can we infer anything about their relationship? Law and Human Behavior, 4, 243-260.

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.

Dunning, D. &. Stern, L. B. (1994). Distinguishing accurate from inaccurate eyewitness identification via inquiries about decision processes. Journal of Personality and Social Psychology, 67, 818-835.

Franzen, S., & Sporer, S. L. (1994). Personenverwechslungen und Möglichkeiten ihrer Vermeidung: Können Augenzeugen durch Visualisierung gegen den Einfluss von irreführenden Rekonstruktionsbildern immunisiert werden? [Person mixups and possible countermeasures: Can eyewitnesses be inoculated against misleading composites through visualization?] In S. L. Sporer, & D. Meurer (Hrsg.), Die Beeinflussbarkeit von Zeugenaussagen (pp. 237-283). Marburg: N.G. Elwert.

Kneller, W., Memon, A., & Stevenage, S. (2001). Simultaneous and sequential lineups: Decision processes of accurate and inaccurate eyewitnesses. Applied Cognitive Psychology, 15, 659-671.

Lindsay, D. S., Nilsen, E., & Read, J. D. (2000). Witnessing-condition heterogeneity and witnesses' versus investigators' confidence in the accuracy of witnesses' identification decisions. Law and Human Behavior, 24, 685-697.

Lindsay, D. S., Read, J. D. & Sharma, K. (1998). Accuracy and confidence in person identification: The relationship is strong when witnessing conditions vary widely. Psychological Science, 9, 215 – 218.

Ratcliff, R. (1978). A theory of memory retrieval. Psychological Review, 85, 59-108.

Read, J. D., Lindsay, D. S., & Nicholls, T. (1998). The relation between confidence and accuracy in eyewitness identification studies. In C. P. Thomspson, D. J. Hermann, J. D. Read, D. Bruce, D. G. Payne, & M. P. Toglia (Eds.), Eyewitness memory: Theoretical and applied perspectives (pp. 107-130). Mahwah, NJ: Lawrence Erlbaum Associates.

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.

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.

Sporer, S. L. (1993). Eyewitness identification accuracy, confidence and decision times in simultaneous and sequential lineups. Journal of Applied Psychology, 78, 22-33.

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). New York: Cambridge University Press.

Sporer, S. L., Penrod, S., Read, J. 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.

SuperLab Pro (v. 1.75). [Computer software]. Cedrus Corporation. Phoenix, Arizona.

Stevens, J. P. (2002). Applied multivariate statistics for the social sciences (4th ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

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

Vickers, D., & Packer, J. (1982). Effects of alternating set for speed or accuracy on response time, accuracy and confidence in a unidimensional discrimination task. Acta Psychologica, 50, 179-197.

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.

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

Wells, G. L. (1984). The psychology of lineup identifications. Journal of Applied Social Psychology, 14, 89-103.

Winer, B. J. (1971). Statistical principles in experimental design. London: McGraw-Hill.

Fast and Confident: Postdicting Eyewitness Identification Accuracy in a Field Study

The importance of eyewitness identifications for investigating and prosecuting crimes is undoubted. However, cases of miscarriage of justice in which eyewitness identifications played a key role (Scheck, Neufeld, & Dwyer, 2000; Wells, Small, Penrod, Malpass, Fulero, & Brimacombe, 1998) call for procedures that allow the assessment of identification decisions. The focus of the present study therefore is on assessment variables which may be used to retroactively assess individual witnesses' decision making accuracy. The most widely used assessment variables are post-decision confidence (e.g., Sporer, Penrod, Read, & Cutler, 1995) and decision time (e.g., Sporer, 1992a, 1993, 1994; Weber, Brewer, Wells, Semmler, & Keast, 2004). Other methods which further explore different types of decision processes (Dunning & Stern, 1994) have also been explored lately (Brewer, Palmer, McKinnon, & Weber, 2005; Caputo & Dunning, 2005; Kneller, Memon, & Stevenage, 2001; Sauerland & Sporer, in press).

It is the first aim of the present research to examine the combined utility of three assessment variables for postdicting identification accuracy in a field study. Specifically, we study the relationship between response time and identification accuracy, focusing on the time boundary that best discriminates correct from incorrect decisions. We also analyze the confidence-accuracy (CA) relationship via correlation and calibration measures. Additionally, we examine the relationship between confidence and identification accuracy, analogous to the time boundary analysis, focusing on the confidence boundary that best discriminates correct from incorrect decisions. Finally, the usefulness of a judgment of the state of awareness and the existence of recollective experience is explored via Remember-Know-Familiar (RKF) judgments (e.g., Tulving, 1985; Conway, Gardiner, Perfect, Anderson, & Cohen, 1997; Gardiner, Ramponi, & Richardson-Klavehn, 2002). Post-decision confidence and decision time each by themselves have been researched extensively in the past, but have only rarely been combined (Brewer & Wells, 2006; Sauerland & Sporer, in press; Sporer, 1992a; Weber et al., 2004). There is now some, albeit unpublished research reporting RKF data for identifications (Brewer, Palmer et al. 2005; Davids, 2006; Penrod, 2006), but combinations with other postdictors are the exception (Davids, 2006). We expected combinations of postdictors to lead to more correct classifications than each postdictor by itself.

Until now, no assessment variable that is capable of making valid postdictions about decisions made by nonchoosers has been identified. A postdictor for nonchoosers' identification accuracy could be a valuable instrument to exonerate innocent suspects. The second aim of the study was therefore to test the usefulness of nonchoosers' self-reports about their decision processes as a postdictor of lineup rejection accuracy.

Furthermore, the present study addresses one shortcoming that affects most identification studies. Although the need for a sufficient participant sample sizes seems obvious to most, the need to sample stimuli appears to be far less straightforward (Wells & Windschitl, 1999). Whereas some studies have reported results for two targets (Brewer, Caon, Todd, & Weber, 2006, Experiment 2; Brewer & Wells, 2006; D. S. Lindsay, Read, & Sharma, 1998), very few studies have used more (Brigham, 1990). Studies that did use two targets found that the results differed with regard to identification accuracy, choosing rates in target-present (TP) and target-absent (TA) lineups (Brewer, Caon et al., 2006, Experiment 2; Brewer & Wells, 2006), optimum time boundaries (Brewer, Caon et al., 2006, Experiment 2), the CA correlations (D. S. Lindsay et al., 1998), and the CA calibrations (Brewer & Wells, 2006). These results show, that general conclusions based on one target could be misleading. Therefore, we included ten targets in the present study.

Even though a wealth of research on identification accuracy exists using laboratory staged events or filmed scenarios, only a small number of field experiments have been conducted to date (e.g., Brigham, van Verst, & Bothwell, 1986; Krafka & Penrod, 1985; Yarmey, 2004). Nevertheless, field experiments have some potential to provide a more differentiated understanding of processes underlying eyewitness identification (Cutler & Penrod, 1995), compared to laboratory research. In order to enhance the ecological validity and generalizability of postdiction research, data were collected in a natural setting, in the pedestrian zone of a small university town.

Decision Time-Accuracy Relationship

The negative relationships between response time and identification accuracy for choosers (people who make a positive identification) but not for nonchoosers (people who reject the lineup) are well established (Dunning & Stern, 1994; Dunning & Perretta, 2002; Kneller et al., 2001; Sauerland & Sporer, in press; S. M. Smith, R. C. L. Lindsay, & Pryke, 2000; Sporer, 1992a, 1993, 1994; Weber et al., 2004; Weber & Brewer, 2006). Recently, researchers have tried to establish a decision time boundary within which correct decisions are highly likely, and therefore incorrect decisions highly unlikely. When splitting the time to decision into three groups (1-15 s, 16-30 s, > 30 s), S. M. Smith et al. (2000) found that eyewitnesses making their choice within the first 15 s were correct in 70% of the cases, whereas accuracy rates in the other intervals dropped to 43% and 18%, respectively. Dunning and Perretta (2002) postulated that witnesses making a positive identification within 10 to 12 s were highly likely to be correct and that these time boundaries were invariant across experimental conditions. The authors based their conclusion on four studies, all of which showed that the best discrimination between correct and incorrect choosers was roughly at 10 to 12 s and termed this cut-off point optimum time boundary. Witnesses who made their identification within 10 s were largely accurate, with an average accuracy rate across the four studies of 87%. Slower choosers were accurate in about 50% of the cases. Response times of correct and incorrect nonchoosers did not differ. In an attempt to explain their results, the authors argued that correct identifications tended to be made via an automatic cognitive process whereas incorrect identification decisions were made via a process of elimination (see also Dunning & Stern, 1994). In contrast to deliberate decisions, which are known to be conscious and effortful, automatic decisions are described as fast and unconscious. Therefore, Dunning and Perretta (2002) claimed that automatic processes should not be influenced by external circumstances, that is, changes in the context of the identification procedure. The invariability of the optimum time boundary found in their experiments was therefore taken as evidence for the automaticity of accurate positive identifications.

Contrary to the findings of Dunning and Perretta (2002), Weber et al. (2004) found optimum time boundaries varying between 5 s to 29 s across four studies using six targets. Only for two of the six samples did the identified optimum time boundaries include the 10-12 s window postulated by Dunning and Perretta (2002). Furthermore, the accuracy rates were considerably lower (67%). When confidence was taken into account in addition to response time, much higher classification rates were obtained within the 10 s window (88%) for rather confident witnesses (90% or 100% confident) than for less confident ones (0% to 80% confident; 54%). These data suggest that response latencies should not be considered in isolation but jointly with other assessment variables, in particular, confidence.

Also contrary to Dunning and Perretta's conclusions, Brewer, Caon et al. (2006) found that manipulations of retention interval (0 vs. 15 vs. 30 minutes) and lineup size (4 vs. 8 vs. 12 persons) both influenced response times and time boundaries. Short retention intervals and smaller lineup size lead to faster decisions and earlier optimum time boundaries than longer retention intervals and larger nominal lineup size. None of the identified time boundaries were within the 10-12 s interval postulated by Dunning and Perretta (2002). Thus, it seems that the claim of an invariant time boundary (Dunning & Perretta, 2002) cannot be upheld. However, the critical results were exclusively obtained by using one stimulus film from Weber et al. (2004). In the present study, we aimed at analyzing empirical time boundaries using a different research paradigm, a field experiment. Whereas we cannot rule out the possibility that the optimum time boundary here is at 10-12 s, we did not expect this. Furthermore, we expected to replicate the negative correlation between decision times and identification accuracy for choosers but not nonchoosers.

Even though the measurement of decision times is highly desirable, practitioners are often confronted with the lack of such measures. As a possible resort in cases where the actual decision times are not available, we asked our participants to give an estimation of their decision time. First results on this were obtained by Davids (2006), who combined estimated decision times with "easiness of identification" and "willingness to swear an oath on the testimony made" in a factor analysis. The factor was negatively correlated with measured decision times and positively correlated with identification accuracy for two targets. In order to test the usefulness of estimated decision times as a substitute of measured decision times, we reran all analyses that were carried out for measured decision times for estimated decision times. We expected a positive correlation between the two measures. Furthermore, we expected both measures to elicit similar result patterns with regard to their postdictive value.

Confidence-Accuracy Relationship

Meta-analyses reported small to medium point-biserial correlations between postdecision confidence and accuracy of $\underline{r}$ = .25, (Bothwell, Deffenbacher, & Brigham, 1987) and $\underline{r}$

= .28 (Sporer et al., 1995). Importantly, Sporer et al.'s (1995) separate analyses for choosers and nonchoosers revealed that the CA relationship was considerably higher for choosers (r = .37) than for nonchoosers (r = .12).

However, assessing the CA relationship via point-biserial correlations has been criticized in recent research (Brewer, 2006; Brewer & Wells, 2006; Juslin, Olsson, & Winman, 1996; Weber & Brewer, 2006) as reliance on point-biserial correlations can result in discounting of informative CA relations (Brewer, 2006), for example due to a lack of variation in confidence or accuracy. Moreover, the correlation coefficient does not give guidance on how to interpret a single confidence rating. Consequently, some researchers have chosen a different approach to determine the CA relationship, namely by computing the calibration between confidence and accuracy. Whereas at first there were primarily face recognition studies using calibration measures (Cutler & Penrod, 1989; Olsson, Juslin, & Winman, 1998; Weber & Brewer, 2003, 2004, 2006) the body of research using eyewitness identification paradigms is now growing (e.g., Brewer, Keast, & Rishworth, 2002; Brewer & Wells, 2006; Juslin et al., 1996; Olsson & Juslin, 1999). In line with results on CA correlations, the general result pattern shows a better calibration for choosers than for nonchoosers (Brewer & Wells, 2006; Brewer, Keast et al., 2002; Weber & Brewer, 2003, 2006).

The present field experiment examined the CA relation, using both correlation and calibration for the identification task. We expected the CA relation to be stronger for choosers than for nonchoosers. For the combination with decision times and Remember-Know-familiar judgments, we sought to establish an empirical confidence boundary that best discriminates between correct and false choices analogous to Dunning and Perretta’s (2002) time boundary.

Remember-Know(-Familiar) Judgments

The third assessment variable included in the present study was a judgment of recollective experience, that is, Tulving's (1985) Remember-Know (RK) distinction proposed in the context of dual-process theories of recognition memory. The RK paradigm has been used primarily in the cognitive literature on word recognition (e.g., Conway & Dewhurst, 1995; Gardiner, 1988; Gardiner & Java, 1990, 1991; Perfect, Mayes, Downes, & Van Eijk, 1996; Rajaram & Geraci, 2000). Typically, at study participants were exposed to a list of words. At test, participants indicated whether a particular word had been presented before (i.e., "old") or not (i.e., "new"). After positive ("old") responses, participants made a Remember or Know judgment. Remembering was defined as "the ability to become consciously aware again of some aspect of what happened or what was experienced at the time the word was presented" (Gardiner, 1988, p. 311) whereas Knowing was defined as "recognition that the word was in the booklet but the inability to collect consciously anything about its actual occurrence or what happened or what was experienced at the time of its occurrence" (Gardiner, 1988, p. 311). Later, these two categories were supplemented with a Guess option (RKG), the reason being that the two judgments could not be considered statistically independent because participants had to make a Know judgment when they failed to give a Remember judgment (Rajaram, 1996). In the RKG paradigm, participants were asked to give a Guess response when they have some other reason than Remembering or Knowing to believe, or suspect, that a test item was encountered in the study list. Some studies additionally used a Familiar option, which lies between Know and Guess responses (e.g., Conway et al., 1997; Wright & Sladden, 2003). Conway et al. (1997) explained the Familiar option as follows: "It may be however, that you did not remember a specific instance, nor do you know the answer. Nevertheless the alternative you have selected may seem or feel more familiar than any of the other alternatives.” (p. 398).

Tulving (1985) postulated Remember responses to be associated with autonoetic consciousness and episodic memory, that is, recollective memories. In contrast, Know responses were supposed to involve noetic consciousness and semantic memory, that is, nonrecollective memories. A wealth of studies has led to the conclusion that there are some experimental manipulations that influence Remember responses whereas others only have an impact on Know responses (Gardiner, 2001).

Since Tulving's original proposal, other models have been developed in an effort to account for RK data. Dual-process theories of recognition memory assume recollection to be an all-or-none retrieval process (high-threshold process; e.g., Yonelinas, 1994). Specifically, recollection is defined as a threshold process which is accompanied by retrieval of qualitative information about a studied event (Yonelinas, 2001). In contrast to dual-process theories of recognition memory, one-dimensional signal detection models (Donaldson, 1996; Hirshman & Master, 1997) propose that RK judgments are simply equivalent to confidence statements (Donaldson, 1996; Dunn, 2004). Nevertheless, many studies showed that the patterns of results were different for RK versus confidence statements (e.g., Gardiner & Java, 1990; Mäntylä, 1997; Parkin & Walter, 1992; Rajaram, 1993; but see Perfect et al., 1996, for an exception, and Dunn, 2004, for a discussion).

Finally, Rotello, Macmillan, and Reeder (2004) proposed a two-dimensional model of RK judgments. They combined Tulving’s idea that Remembering and Knowing reflected two distinct forms of memory (global familiarity and specific recollection) with the signal detection approach. This approach differentiates between accuracy and response bias and accounts for the confidence level with which responses are made. According to the model, stimuli vary in global memory strength (familiarity) and in the strength of their more specific features or characteristics. "Old" items are more familiar (i.e., show stronger global memory strength) and activate more contextual and semantic information (i.e., showing more specific memory strength) than "new" items. When deciding whether an item is "new" or "old", both memory strengths are summed. Remember-Know judgments divide "old" responses into those for which specific memory strength is relatively strong (Remember responses) and those for which specific memory strength is relatively weak (Know responses), in comparison to global memory strength. Accordingly, Remember and Know judgments are both based on an assessment of global and specific memory strength. The authors concluded that Remember responses neither were simply high confidence old decisions nor did they result from a high-threshold process, as assumed by the dual-process model. According to Rotello et al. (2004), those people who make a Remember judgment perceive greater specific memory strength compared to global memory strength. The opposite is assumed for Know responses. Translated to eyewitness identification, choosers who give a Remember response after making their identification decision should accordingly be more accurate than those giving Know responses.

Another way of looking at the decision processes concerns the distinction between automatic and deliberate decision processes proposed by Dunning and Stern (1994). They hypothesized witnesses reporting automatic decision processes (i.e., “just knowing”) to be more likely to make a correct decision than those making a deliberate decision. Comparing this approach to the Remember-Know paradigm, we would expect Know answers to be associated with automatic processes and higher accuracy than the more deliberate Remember answers. Dunning and Stern's account has gained support in some studies (Caputo & Dunning, 2005; Dunning & Stern, 1994; Kneller et al., 2001), but in another study self-reported decision processes were unable to distinguish between accurate and inaccurate decisions (Brewer, Gordon, & Bond, 2000).

With regard to Remember-Know judgments in the context of identification accuracy, people have hypothesized about the superiority of one or the other, but until now, not a single study has been published. The results of the few unpublished studies on the topic show mixed results: Supporting the assumptions made in the Remember-Know literature, Brewer, Palmer et al. (2005) found significantly higher accuracy for Remember than for Know answers. On the contrary, supporting Dunning and Stern's (1994) notion, Davids (2006) who used Remember, Know, and Familiar options, found a tendency of Know, not Remember answers to be associated with identification accuracy.

To summarize, rival hypotheses exist about the superiority of Remember and Know judgments for postdicting identification accuracy and the data are inconclusive. Only few studies exist on the issue to date and in the present study we sought to provide more data on these rival hypotheses. Specifically, we used Remember, Know, and Familiar options. A Guess option was not included, because we hold the view that this option would undermine the ecological validity in the context of person identifications. Courts certainly would have reservations to prosecute a person on the basis of a witness who justified his or her lineup decision on nothing more than a good “guess”. Both, Remember and Know responders were expected to show better identification performance than participants giving a Familiar response.

Nonchoosers’ Decision Processes

Previous research has consistently shown that associations of identification accuracy with post-decision confidence (e.g., Sporer et al., 1995; Weber & Brewer, 2003) and decision time (e.g., Dunning & Stern, 1994; Sporer, 1992a, 1993; Weber et al., 2004, Weber & Brewer, 2006) are statistically reliable for positive identification decisions, but not for negative identification decisions. In an attempt to explain this observation Sporer et al. (1995) argued that the group of nonchoosers was too heterogeneous. Specifically, they proposed that rejecting a lineup can indicate, first, that the witness was absolutely sure that the target was absent, second that the witness believed that one person in the lineup was the target but that he/she was not confident enough to choose this person. Thirdly, the witness may have had no memory for the event. In the present study, we asked nonchoosers to indicate one of three reasons why they rejected the lineup after their decision. The options included that (a) the target had not been among them ("absent”), (b) they thought the target had been among them, but that he/she had not been confident enough to choose him/her ("low confidence”), or (c) he/she did not know whether the target had been among them or not ("no memory”). In doing so, we hoped to create three subgroups of nonchoosers who differ in their accuracy. Specifically, we expected participants in the "absent" group to show higher identification accuracy than those in the two other groups. The "low confidence" group was expected to be less accurate than the "no memory" group for the following reasons: The "no memory" group should be accurate in their decisions in about 50% of the cases (the base rate for target absence/presence). Lower accuracy than chance was expected for participants who reported that they believed the target may have been present but that they had not been confident enough to make a choice. Analogous, we expected confidence to be highest and decision times to be shortest for the absent group, followed by the "no memory" and "low confidence" groups. Table 1 gives a summary of all our hypotheses.

Table 1

Summary of Hypotheses

Variable	Hypothesis
Decision times	Negative correlation with identification accuracy for choosers but not nonchoosers
Estimated decision times	Result patterns paralleling those for measured decision times
	Positive correlation with measured decision times
Confidence	Positive correlation with identification accuracy for choosers but not nonchoosers
	Better calibration for choosers than nonchoosers
Remember-Know-Familiar judgments	Rival hypotheses: Rotello et al. (2005): higher accuracy for Remember than Know answers; Dunning and Stern (1994): higher accuracy for automatic (Know) than deliberate (Remember) answers
Combination of postdictors	Higher correct classification rates for the combination of all three postdictors than for each postdictor by itself
Nonchoosers' decision processes	Accuracy, confidence, and decision times (reversed) in descending order: "absent" group, "no memory" group, "low confidence" group

Method

Participants

Nine-hundred-and-forty-four persons (479 female, 465 male) were approached in the pedestrian zone in a small university town in Germany. Of these, 720 (360 male, 360 female; age 15 to 84, Mdn = 33.5 years) agreed to take part in our study. Most participants were students with various majors (27.2%). Note that this is representative for this city, with 75,000 inhabitants and 23,000 students. Other participants worked in academic (16.5%) or other occupations (36.2%), were high school pupils (8.5%), retired (6.8%), housewives/-husbands (4.2%) or unemployed (0.6%).

Manipulated Variables

In a 10 x 2 design, target person (targets 1 through 10) and target presence (present vs. absent) were independent variables, with an equal number of cases in each cell. Identification accuracy was the dependent variable, with pre- and post-decision confidence, decision time, and estimated decision time as postdictor variables. In addition, a judgment of recollective experience for choosers, and a judgment of decision process for nonchoosers was collected. For choosers, it was randomly manipulated whether there were Remember and Know options or Remember, Know, and Familiar options for the judgment of recollective experience.

Photo Lineup

Each lineup consisted of six frontal 9 x 13 cm photographs mounted on a 30 x 33.5 cm display board depicting six individuals that were arranged in two rows of three pictures each and numbered 1 to 6. Each lineup member wore different long sleeve clothing. Jewelry, eyeglasses and hair accessories were taken off and hair was worn loose. During their encounter with the participant, target persons wore different clothing from what they were wearing on their lineup picture. All lineup pictures were taken outdoors in front of the same wall.

Target persons were eight female and two male psychology major students (age 20 to 37, Mdn = 20 years) who received course credit for data collection.

For half of the participants, the target photograph was present (TP) in the lineup, for the other half it was substituted by a replacement (TA). For practical reasons, in this field study, the target as well as the replacement always appeared on position 3. All foils, including the replacement, fit the general descriptions of each of the target persons as determined by a pilot study with N = 55 mock witnesses (effective sizes, determined as Tredoux's Es were between 5.14 and 6.76; Tredoux, 1998, 1999).

Procedure

Ten persons were involved in the present study both as stimulus persons (targets) and as interviewers, alternating their roles. Data collection took place in a town different from their home university. Data were collected in pairs of two persons, one serving as target, one as the interviewer. After data of eight participants were collected, the roles were changed. Every interviewer collected data of 8 participants with each of the other targets. Interviewers underwent three training sessions in which the interaction with the participants as both target and interviewer, time taking as well as controlling interaction time was practiced.

Specifically, the target asked a passer-by for directions to a certain location in the pedestrian zone of a university town. The conversation was scheduled to last between 15 s and 60 s. Hereafter, the target walked into the shown direction. To ensure that the target was no longer seen by the passer-by after the interaction, targets only approached people whose walking direction was opposite of the location asked for.

The other person, the interviewer, watched the situation from a distance and recorded the interaction time with a stopwatch. Thirty seconds after the interaction between target and participant was terminated the interviewer approached the participant and explained that the true nature of the preceding interaction was a study on face recognition. If consent to participate was given, passers-by were handed a questionnaire. After indicating their pre-decision confidence on an 11-point scale ranging from 0% to 100% (with intervals marked in 10% steps: 0%, 10%, ..., 100%), participants were handed the display board with the lineup. To ensure the blindness of the interviewer, target-present and -absent lineups were placed into differently colored envelopes, with interviewers being unaware of their contents. The experimental design sheet informed interviewers which envelope to use for which participant. Thus, the interviewer pulled the display board out of the envelope and handed it to the participant without looking at it. This procedure ensured that the interviewers could only see the display board from the back, but not the pictures, that is, whether the target was present or not.

On the answer sheet, participants were advised that they would be asked to identify the person who had asked them for directions. The instructions clearly stated that the person might or might not be present in the lineup. Identification decision time was measured unobtrusively via stopwatches which the interviewers held in their hands but were covered by a shopping bag.

After the identification task, choosers were asked to make a RK or RKF judgment. The RKF judgment instructions followed those provided by Conway et al. (1997) but were adapted for the identification procedure:

During the identification

A ... I had the person, his/her manner of speaking, the situation, or what I was thinking when he/she asked me for directions before my mental eye.

B ... I just knew that it was the face of the person, who asked me for directions.

C ... the face just felt somewhat familiar, so that I chose it.

In the RK condition, the third response was omitted. Hereafter, choosers indicated their post-decision confidence on an 11-point scale ranging from 0% to 100%.

The following procedure was adopted for nonchoosers: Immediately after the identification, nonchoosers were first asked for their post-decision confidence rating (0%-100%) and then made a judgment about their decision process. To gain further insight into the decision processes of nonchoosers they were given the following question:

You did not make a choice. Which of the following statements applies best to you?

A The person was not among them.

B I think the person was among them, but I was not confident enough to choose him/her.

C I did not know whether the person was among them or not.

Finally, participants were asked, how long they thought they had seen the target person, and how long they estimated it had taken them to make a decision. Estimations were made in s with no intervals allowed.

After participants had completed all items, they were thanked and handed a card with the researchers' web address where the results of the study would be available after approximately three months.

Results

An alpha level of <u>alpha</u> = .05 was used for all inferential analyses. Cohen's (1988) <u>d</u> and <u>f</u>, are reported as measures of effect size for ANOVAs, and Cramer's <u>V</u> and <u>phi</u> are reported for nonparametric analyses of 3 x 2 and 2 x 2 contingency tables, respectively.

Identification accuracy varied across targets from 41.7% to 70.8% (M = 56.9%, SD = 49.5). Of the 720 participants 410 (56.9%) made a correct identification decision. Nonchoosers ($n = 284$) were more accurate in their decisions (70.1%) than choosers ($n = 436$; 48.4%), $chi^2(1, N = 720) = 32.96$, $p < .001$, $phi = 0.21$. Accuracy did not differ for TA (55.3%) and TP (58.6%) lineups, $chi^2(1, N = 720) = .816$, $p = .408$, $phi = .03$[1].

In the following, we first look at the relationships between identification accuracy and the each postdictor individually. For each postdictor, a 2 x 2 x 10 ANOVA with choice (choosers vs. nonchoosers), decision outcome (correct vs. incorrect), and target (target 1 vs. ... vs. target 10) as independent variables was computed and the results are reported in the following section. Subsequently, the postdictors are combined. Descriptive statistics of decision time measures and decision confidence measures as well as their intercorrelations can be found in Table 2.

Postdicting Identification Accuracy from Assessment Variables

Pre-Decision Confidence

For pre-decision confidence there only was a main effect of decision outcome, $F(1, 680) = 16.91$, $p < .001$, $d = 0.32$. Participants making a correct identification decision showed higher pre-decision confidence ratings (M = 64.4%, 95% CI = 62.2 – 66.6%) than participants making an incorrect identification decision (M = 56.9%, 95% CI = 54.3 – 59.5%).

[1]Results for the individual targets regarding identification accuracy and choosing rates, as well as diagnosticity can be obtained from the Appendix.

Table 2

Means, Standard Deviations, and Intercorrelations of Identification Accuracy, Postdictors of Identification Accuracy, Estimated Viewtime, and Estimated Decision Time for Choosers (n = 436)

	M	SD	95% CI	Pre-decision confidence	Post-decision confidence	Decision time	Viewing time	Estimated viewing time	Estimated decision time
Identification accuracy	48.4%	50.0	43.7-53.1%	.16**	.39**	-.30**	.08	.10*	-.21**
Pre-decision confidence	60.6%	23.6	58.4-62.8%		.46**	-.13**	-.01	-.04	-.16**
Post-decision confidence	60.6%	27.1	58.1-63.1%			-.33**	-.01	.05	-.31**
Decision time[1]	9.4 s	2.5	8.6-10.2 s				-.04	-.04	.49**
Viewing time	28.5 s	10.8	27.5-29.5 s					.27**	.05
Estimated viewing time[1]	19.4 s	2.3	17.9-21.0 s						.35**
Estimated decision time[1]	9.3 s	3.0	8.4-10.3 s						

*p ≤ .05, two-tailed. **p ≤ .01, two-tailed.

[1]The mean and standard deviations were backtransformed from the logarithmic values used for the inferential analyses.

Decision Time

Due to significant positive skewness and kurtosis in the response latency distribution, one extreme value was winsorized (Winer, 1971) and inferential analyses were conducted on log-transformed data (i.e., log base 10).

There were significant main effects of decision outcome, $F(1, 680) = 21.37$, $p < .001$, $d = -0.35$, and choice, $F(1, 680) = 8.88$, $p = .003$, $d = -0.23$. Specifically, accurate participants decided faster ($M = 8.6$ s, 95% CI = 7.9 - 9.5 s) than inaccurate ones ($M = 12.1$ s, 95% CI = 11.1 - 13.3 s), and choosers decided faster ($M = 9.4$ s, 95% CI = 8.6 - 10.2 s) than non-choosers ($M = 11.0$ s, 95% CI = 10.0 - 12.2 s). The main effects were modified by the expected interaction between Decision Outcome and Choice, $F(1, 680) = 6.26$, $p = .013$, $f = .10$. There was a simple main effect of decision outcome within choosers, $F(1, 680) = 43.90$, $p < .001$, $d = -0.49$, but not within nonchoosers, $F(1, 680) = 0.53$, $p = .465$, $d = -0.06$. Within choosers, accurate decisions were made faster ($M = 7.3$ s, 95% CI = 6.3 - 7.9 s) than inaccurate ones ($M = 12.4$ s, 95% CI = 10.9 - 13.8 s). Within nonchoosers, there was no difference between accurate ($M = 11.0$ s, 95% CI = 9.5 - 12.1 s) and inaccurate ($M = 12.8$ s, 95% CI = 9.7 - 14.1 s) decisions with regard to decision time.

Furthermore, the interaction between Decision Outcome and Target was significant, $F(9, 680) = 2.30$, $p = .015$, $f = .17$. Note, however, that the predicted Choice by Accuracy interaction was not qualified by an interaction with Target, $F(9, 680) = 0.49$, $p = .884$, $f = .08$.

Post-Decision Confidence

There was a significant main effect for decision outcome, $F(1, 680) = 37.50$, $p < .001$, $d = 0.47$. Accurate participants were more confident ($M = 67.8\%$, 95% CI = 65.3% - 70.3%) than inaccurate ones ($M = 52.7\%$, 95% CI = 49.8% - 55.5%). The main effect was modified by the expected interaction between Decision Outcome and Choice, $F(1, 680) = 11.83$, $p = .001$, $f = .15$. There was a significant simple main effect of accuracy within choosers, $F(1, 680) = 66.70$, $p < .001$, $d = 0.63$, but not within nonchoosers, $F(1, 680) = 2.74$, $p = .098$, $d = .13$. Specifically, within choosers, accurate identifiers were more confident ($M = 70.7\%$, 95% CI = 67.1 – 74.3%) than inaccurate ones ($M = 50.1\%$, 95% CI = 46.7 - 53.5%). Within nonchoosers, accurate ($M = 63.7\%$, 95% CI = 60.2 - 67.3%) and inaccurate identifiers ($M = 58.0\%$, 95% CI = 52.1 – 63.8%) did not differ significantly with regard to post-decision confidence.

Estimated Decision Time

Due to significant positive skewness and kurtosis in the estimated decision time distribution, three extreme values on both sides of the distribution were winsorized (Winer, 1971) and inferential analyses were conducted on log-transformed data.

Estimated decision time and the measured times were positively associated for choosers, $r(434) = .49$, $p < .001$, and nonchoosers, $r(282) = .46$, $p < .001$. The estimated and measured means did not differ significantly from one another for choosers (Ms = 9.3 s and 9.4 s), $t(435) = -.164$, $p = .870$, $d = -0.02$, or nonchoosers (Ms = 10.3 s and 11.0 s), $t(283) = -1.16$, $p = .246$, $d = -0.11$. Figure 1 displays the means of correct and incorrect choosers and nonchoosers with regard to actual and estimated decision time.

There was a significant main effect of decision outcome, $F(1, 680) = 5.50$, $p = .019$, $d = -0.18$. Correct decisions were estimated to have been made faster ($M = 9.0$ s, 95% CI = 7.8 – 9.7 s) than incorrect ones ($M = 11.1$ s, 95% CI = 10.0 – 12.6 s). This main effect was modified by an interaction between Decision Outcome and Choice, $F(1, 680) = 5.42$, $p = .020$, f = .09. There was a significant simple main effect of accuracy within choosers. Correct choosers estimated their decision time as shorter ($M = 7.7$ s, 95% CI = 6.6 – 8.9 s) than did incorrect choosers (M = 11.7 s, 95% CI = 10.1 – 13.5 s), F(1, 680) = 15.94, p < .001, d = -0.31. No such difference was found between correct (M = 10.6 s, 95% CI = 9.1 – 12.3 s) and incorrect nonchoosers (M = 10.6 s, 95% CI = 8.3 – 13.5 s), F(1, 680) = 0.00, p = .992, d = 0.00.

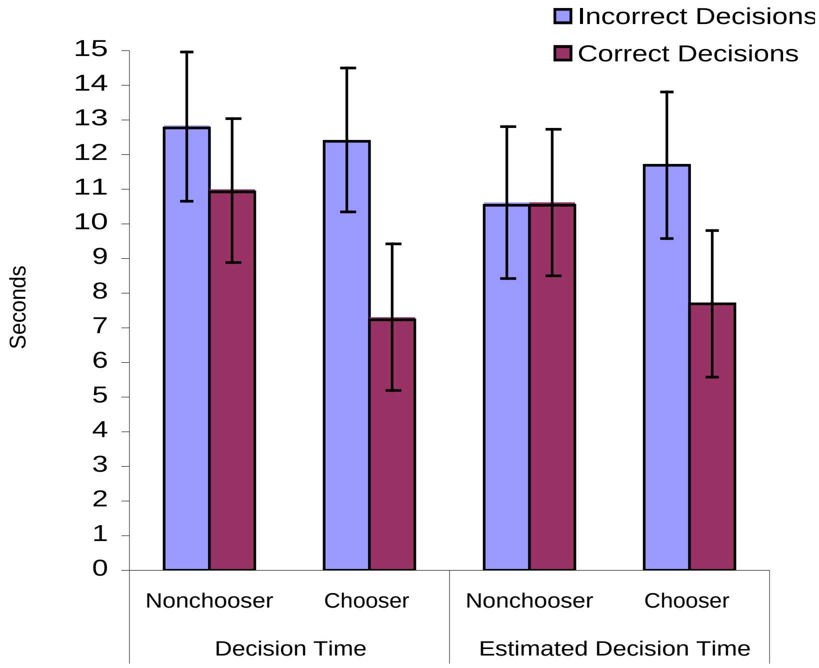


Figure 1. Means (and CIs) of actual vs. estimated decision time as a function of choice and decision outcome.

Calibration of Confidence Measures

For confidence measures, in addition to ANOVAs, calibrational analyses were carried out in order to establish the postdictive value of these measures.

Pre-Decision Confidence

The pre-decision confidence-accuracy correlation did not differ between choosers, $r(434) = .16$, $p = .001$, and nonchoosers, $r(282) = .16$, $p = .006$, $z = 0.11$. Therefore, calibration analyses with regard to pre-decision confidence were carried out for the whole group of participants. Following Brewer and Wells (2006) the 11 confidence categories were collapsed into five categories (i.e., 100% and 90%, 80% and 70%, 60% and 50%, 40% and 30%, 20%, 10% and 0%) to provide more stable estimates of proportion correct for each confidence category. To create the CA calibration curve, the proportion correct for each of the collapsed categories was plotted against the weighted mean confidence for that category. Figure 2 displays the CA calibration curve, which shows only very little slope, indicating poor calibration. The overall calibration was $C = .024$. The curve displays strong underconfidence in the first two categories (0 - 20% and 30 - 40%), good calibration in the third category (50 - 60%), and strong overconfidence in the last two categories (70 - 80% and 90 - 100%), leading to an overall O/U of .042. The resolution statistic showed only little capability to discriminate between correct and incorrect identification decisions ($NRI = .041$). The NRI statistic is directly comparable to eta^2 (see Baranski & Petrusic, 1994; Yaniv, Yates, & J. E. K. Smith, 1991). As eta^2 is directly related to Cohen's f, cutoffs for small, moderate, and large NRI values can be derived from the .10, .25, and .40 cutoffs for f, specifically, the values are .010, .059, and .138 (see Brewer & Wells, 2006).

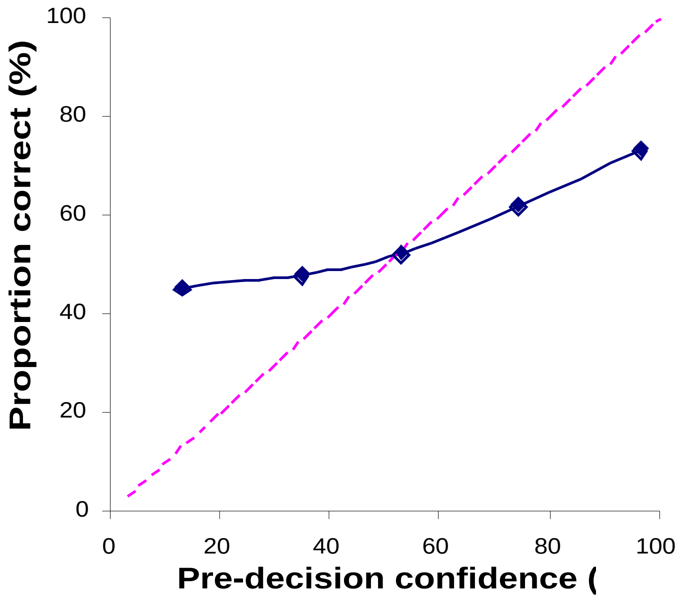
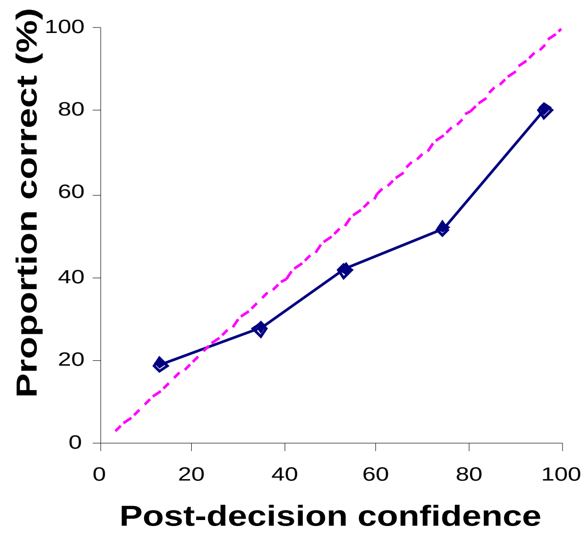


Figure 2. Pre-decision confidence-identification accuracy calibration curves for choosers and nonchoosers combined.

Post-Decision Confidence

As expected, the CA correlation was highly significant for choosers, $r(434) = .39$, $p < .001$, but non-significant for nonchoosers, $r(282) = .09$, $p = .137$.

Figure 3 displays the CA calibration curve for choosers and nonchoosers. The calibration curve for choosers generally follows the slope of the identity line. Choosers showed an overall calibration of C = .026. The curve displays some underconfidence in the first category (0 - 20%), fairly good calibration in the second category (30 - 40%), and strong overconfidence in the last three categories at the higher end of the scale (50 - 100% confident), leading to an overall O/U of .122. The resolution statistic showed high capability to discriminate between correct and incorrect identification decisions for choosers, NRI = 0.174.



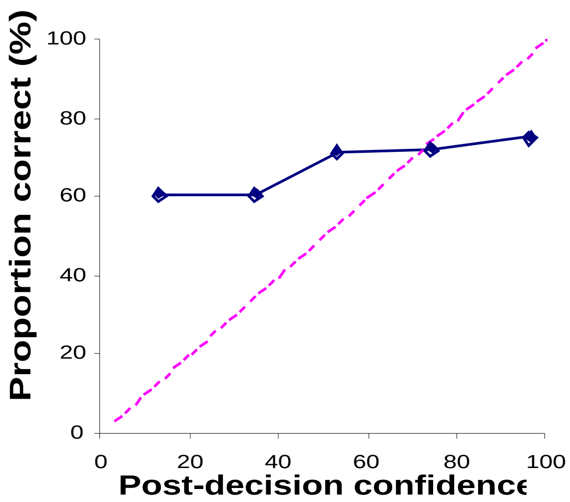


Figure 3. Chooser's (top panel) and nonchooser's (lower panel) post-decision confidence-identification accuracy calibration curves.

For nonchoosers, the calibration line almost parallels the x axis, C = .059. Nonchoosers were highly underconfident in the first three categories, fairly well calibrated in the fourth category and overconfident in the fifth confidence category, leading to an overall O/U of -.077. The resolution index indicates low capability to discriminate between correct and incorrect nonchoosers, NRI = .030.

RKF Judgment

The usefulness of the RK(F) judgment for postdicting identification accuracy was tested via two separate three-way frequency analyses for the RK and the RKF condition to develop a hierarchical log-linear model of RK(F) judgment, decision outcome, and target. For the RKF condition, stepwise selection by simple deletion of effects produced a model that included the two-way association between RKF judgment and Decision Outcome as well as the two-way association between Decision Outcome and Target, $chi^2(36, N = 204) = 37.52$, $p = .399$. Post-hoc chi^2-analyses revealed that identification accuracy for Remember ($M = 47.1\%$, 95% CI = 35.2 - 59.1%) and Know answers ($M = 59.5\%$, 95% CI = 48.4 - 70.6%) did not differ significantly, $chi^2(1, N = 149) = 2.28$, $p = .089$, $phi = -.12$, whereas there was a significant difference in accuracy of Know and Familiar answers ($M = 34.5\%$, 95% CI = 21.6 - 47.5%), $chi^2(1, N = 134) = 8.08$, $p = .004$, $phi = .26$.

For the RK condition, stepwise selection by simple deletion of effects produced a model that included the two-way association between RK judgment and Target, $chi^2(20, N = 232) = 21.74$, $p = .356$, indicating that the distribution of Remember and Know judgments differed as a function of target (but not accuracy: R: 42.4% accurate; K: 54.4%).

Combination of Postdictors

The previous analyses showed that there were associations between decision times and confidence measures with identification accuracy for choosers. However, in a given case, it is not clear how fast, or how confident is “enough” in order to evaluate the identification evidence as accurate. Therefore, decision time boundaries (Dunning & Perretta, 2002) were computed based on the 2 (accuracy: correct vs. incorrect) x 2 (time boundary: faster or equal vs. slower) contingency tables with the time boundary set at each integer value (i.e., 1 s, 2 s, etc.). The decision time that produced the greatest chi^2 value was identified as the time boundary that optimally discriminated between correct and incorrect choosers. An analogous analysis was carried out for post-decision confidence based on the 2 (accuracy: correct vs. incorrect) x 2 (post-decision confidence boundary: more or equal vs. less confident) contingency tables. Based on the results of the boundary analysis, accuracy within and outside the optimum boundary were computed for each target. We then computed the mean accuracy rates across targets. In a second step, we determined how many choosers made accurate decisions, provided they responded within both optimum boundaries, only in one or the other, or outside both optimum boundaries.

Boundary Analyses

Decision Time. As expected, the decision time-accuracy (LA) relation was significant for choosers, $r(434) = -.30$, $p < .001$, but not for nonchoosers, $r(282) = -.05$, $p = .451$. Thus, time boundary analyses were performed only on choosers' data. The plot of chi^2 values by time boundaries is presented in Figure 4. The optimum time boundary here was at 6 s, $chi^2(1, N = 436) = 50.19$, $p < .001$, $phi = -.34$. The mean proportion of correct decisions made before the optimum time boundary was 68.8% and 36.4% after the optimum time boundary.

Post-Decision Confidence. The plot of chi^2 values by post-decision confidence boundaries is presented in Figure 5. The mean proportion of correct decisions made within the optimum post-decision confidence boundary was 77.5% and 38.3% outside the optimum post-decision boundary.

Estimated Decision Time. The estimated decision time-accuracy relationship was significant for choosers, $r(434) = -.21$, $p < .001$, but not for nonchoosers, $r(282) = .01$, $p = .931$. The plot of chi^2 value by estimated time boundaries is presented in Figure 6. The optimum estimated time boundary here was at 3 s, $chi^2(1, N = 436) = 17.52$, $p < .001$, $phi = -.20$. The mean proportion of correct decisions made before the optimum estimated time boundary was 66.1% and 43.7% after the optimum estimated time boundary.

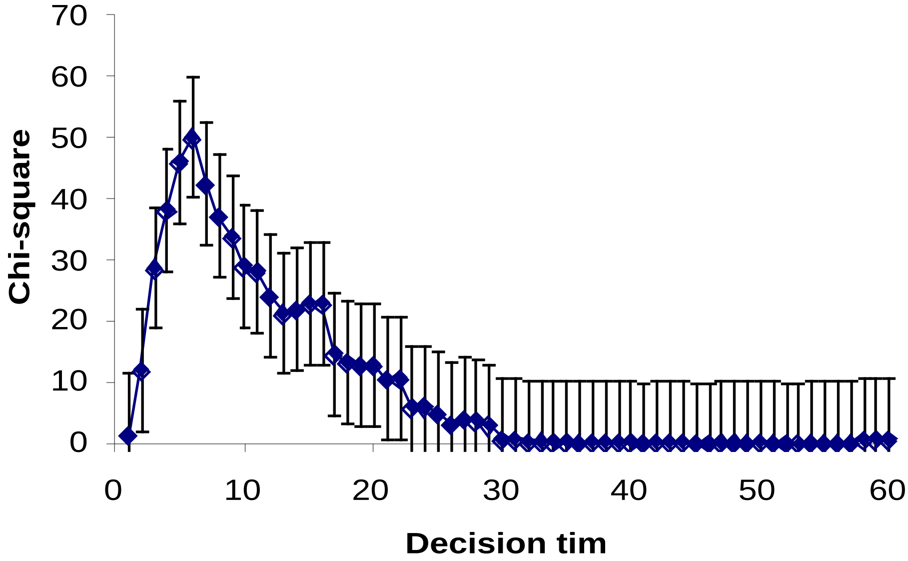


Figure 4. Plot of chi^2 values (and 95% CI) by decision time boundary.

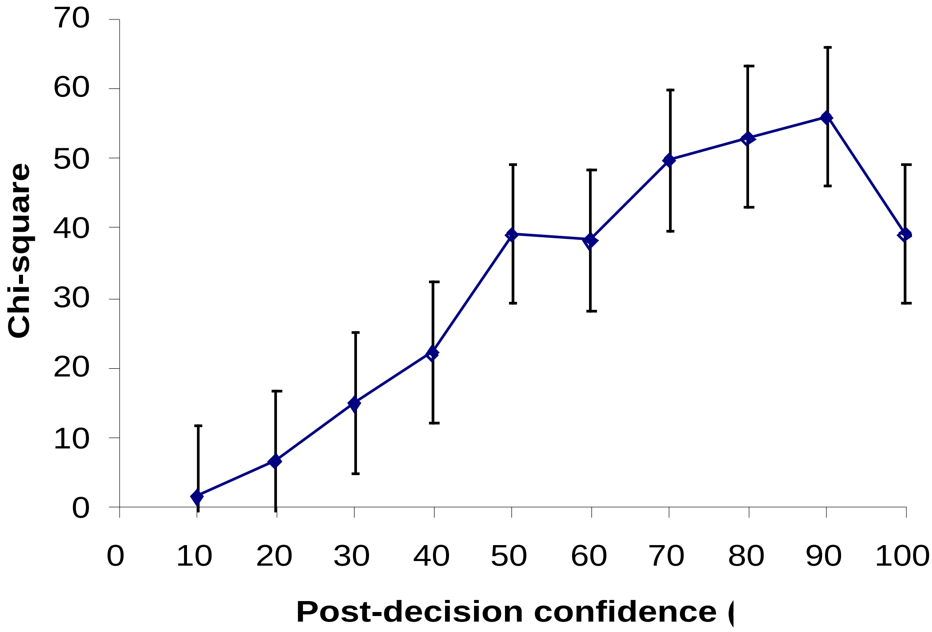


Figure 5. Plot of chi^2 values (and 95% CI) by post-decision confidence boundary.

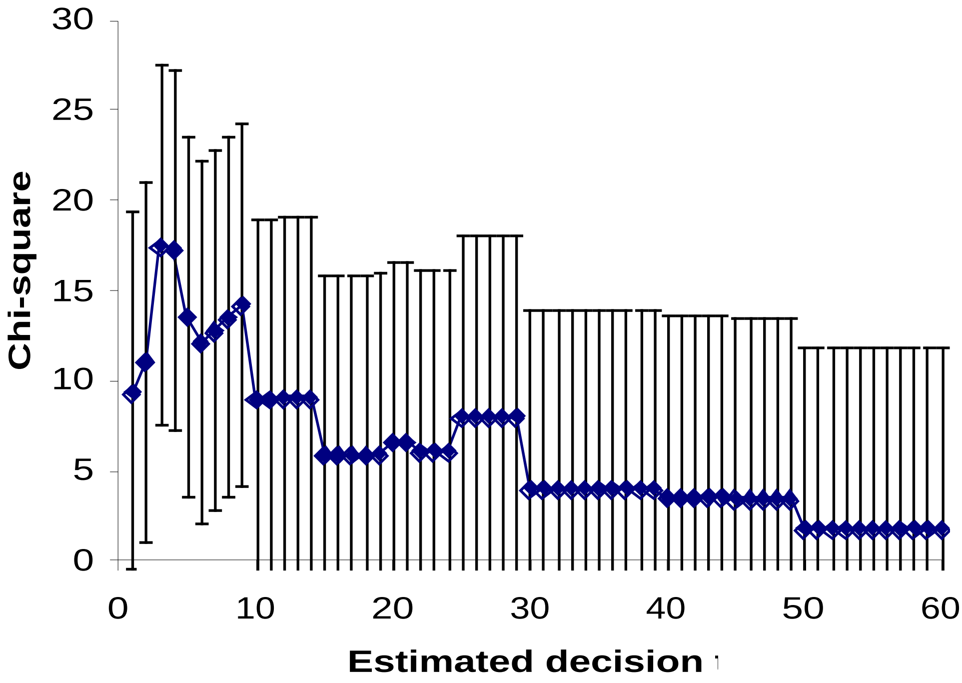


Figure 6. Plot of chi^2 values (and 95% CI) by estimated decision time boundary.

Combining Postdictors According to Boundary Results

In the next step, postdictors were combined according to the boundary analyses results. For the RKF condition, R and K vs. F judgments were also included. Remember and Know judgments were combined as there was no significant difference in identification accuracy for these two groups. Table 3 shows identification accuracy for postdictor combinations. The columns labeled % correct contain what proportion of participants was correct in this condition. For example, of the non-confident fast choosers, 56.8% of the 88 participants made a correct decision. The columns labeled M contain the unweighted mean for this cell across all 10 targets. This was slightly less (55.4%) for the example.

Choosers who decided within the first 6 s and gave a post-decision confidence rating of 90% or 100% were on average (computed across targets) correct in 97.2% of the cases. For choosers who were less than 90% confident and slower than 6 s, accuracy was 31.8% on average.

Table 3

Identification Accuracy of Choosers for Combinations of Postdictors as a Function of RK(F) Condition

Condition	Decision time Fast (≤ 6 s)			Slow (> 6 s)			Total		
RK and RKF conditions together	n	% correct	M*	n	% correct	M*	n	% correct	M*
Post-decision confidence									
0- 80%	88	56.8	55.4	245	31.8	31.8	333	38.4	38.3
95% CI		46.4-67.2			26.0-37.6			33.2-43.6	
90-100%	57	96.5	97.2	46	60.9	60.4	103	80.6	77.5
95% CI		91.4-100			46.7-75.1			72.9-88.3	
Total	145	72.4	68.8	291	36.4	36.4	436	48.4	57.0
95% CI		65.1-79.7			30.9-41.9			43.7-53.1	

Table 3 (continued)

	Decision time						Total		
	Fast (≤ 6 s)			Slow (> 6 s)					
RKF condition only	n	% correct	M*	n	% correct	M*	n	% correct	M*
Low post-decision confidence (0-80%)									
Remember/Know	25	48.0	52.5	70	32.9	32.2	95	36.8	35.6
95% CI		28.0–68.0			21.8-44.0			27.1-46.5	
Familiar	9	44.4	47.6	44	31.8	27.5	53	34.0	35.7
95% CI		10.0-78.8			17.9-45.7			21.1-46.9	
Total	34	47.1	47.4	114	32.5	31.6	148	35.8	36.2
95% CI		29.4-64.7			23.7-41.1			28.0-43.6	

	Decision time						Total		
	Fast (≤ 6 s)			Slow (> 6 s)					
RKF condition only	n	% correct	M*	n	% correct	M*	n	% correct	M*
High post-decision confidence (90-100%)									
Remember/Know	31	96.8	97.8	23	65.2	67.5	54	83.3	78.1
95% CI		90.5-100			45.3-85.1			73.3-93.3	
Familiar	1	100	100	1	0.0	0	2	50.0	50.0
95% CI		100-100			0-0			0-100	
Total	32	96.9	97.8	24	62.5	65.8	56	82.1	77.3
95% CI		90.5-100			41.6-83.3			71.8-92.5	
High and low confidence	66	71.2	71.8	138	37.7	37.5	204	48.5	48.1
95% CI		60.2-82.2			29.6-45.8			41.6-55.4	

* Mean accuracy calculated separately as the unweighted mean for this cell across all 10 targets (see text).

Replacing decision time with estimated decision time, that is, looking at those participants who estimated their decision as 3 s or faster and who were also 90% or 100% confident, produced an accuracy rate of 96.7%. For choosers who were less than 90% confident and estimated their decision time as slower than 3 s, accuracy was 37.6% on average.

Including the RKF judgment in the RKF condition (N = 204) led to the following results: Those participants who made a Remember or Know judgment were on average accurate in 52.7% of the identifications. Looking only at decision time in this group (within the first 6 s) led to 71.8% correct identifications, and using only post-decision confidence as a postdictor (90 or 100%) to 77.3%. Combining the three postdictors resulted in 97.8% correct identifications. However, combining only decision time and post-decision confidence also resulted in 97.8% correct decisions; thus RKF judgments did not improve postdictions. For the other two combinations (post-decision confidence and RKF judgments; decision time and RKF judgments) accuracy rates were at 78.1% and 74.7%, respectively.

Combining estimated decision times with post-decision confidence and RKF judgment produced an accuracy rate of 96.9% for those participants who estimated their decision time as 3 s or faster, were 90% or 100% confident and gave a Remember or Know judgment.

For the former analyses it might be objected, that post-hoc criterion setting does not help to postdict the results in future samples. Therefore, we split our sample into two random subsamples wherby 50% of the choosers of each target were randomly assigned to one of two groups. We performed all analyses again for one subsample and then postdicted identification accuracy for the second subsample. The optimum time boundary for the subsample was at 6 s, $chi^2(1, N = 218) = 25.68$, $p < .001$, $phi = -.34$. The proportion of correct decisions made before the optimum time boundary was 71.6% and 35.4% after the optimum time boundary. The optimum post-decision confidence boundary for the subsample was at 70%, $chi^2(1, N = 218) = 27.67$, $p < .001$, $phi = .36$. The proportion of correct decisions made within the optimum post-decision confidence boundary was 68.9% and 28.7% outside the optimum post-decision confidence boundary. The optimum estimated time boundary was at 4 s, $chi^2(1, N = 218) = 9.06$, $p = .003$, $phi = -.20$. The proportion of correct decisions made before the optimum estimated time boundary was 67.4% and 42.4% after the optimum estimated time boundary.

Hierarchical log-linear model of RK(F) judgment, decision outcome, and target produced a model that only included decision outcome for the RK, $\underline{chi}^2(38, \underline{N} = 118) = 42.86$, $\underline{p} = .270$, and the RKF conditions, $\underline{chi}^2(58, \underline{N} = 100) = 70.42$, $\underline{p} = .127$. Therefore, RK(F) judgments were not included for the combination of postdictors.

Choosers who decided within the first 6 s and gave a post-decision confidence rating between 70 and 100% were on average across targets correct in 82.9% of the cases. Choosers who were less confident and slower were on average 23.2% accurate. Replacing decision time with estimated decision time produced an accuracy rate of 81.0%. Choosers who estimated their decision as slower and who were less confident were on average accurate in 36.2% of the cases. Thus, with regard to combining (estimated) decision times and post-decision confidence, we found comparable results when the optimum boundaries were estimated for one subsample and then used to postdict identification accuracy of another subsample. RKF judgment had no postdictive value.

To summarize, we successfully postdicted identification accuracy of choosers by using post-decision confidence and decision time as postdictors. The RKF judgment did not improve postdiction beyond decision time and post-decision confidence. Estimated decision times also served as a useful postdictor. Note, howerver that postdiction was successful only for a small number of participants in the sample: fast and confident choosers.

Nonchoosers' Decision Processes

Most of the results reported so far referred to choosers' decisions only. In the following, we report results with regard to nonchoosers' decision processes. Descriptive statistics of decision time, decision confidence measures, and decision process measures of nonchoosers as well as their intercorrelations can be found in Table 4. Of the 284 nonchoosers, 204 simply stated "The person was not among them" (absent group), 39 responded "The target was among them, but I was not confident enough" ("low confidence" group), and 41 selected the option “I did not know whether the person was among them or not“ ("no memory" group). However, the three groups did not differ with regard to decision outcome, $\underline{chi}^2(2, \underline{N} = 284) = 2.13$, $\underline{p} = .345$, Cramer's V = .09, although the "absent" group showed a tendency to greater accuracy (72.5%) than the other two groups (64.1% vs. 63.4%).

Three 3 x 2 ANOVAs with the decision process groups ("absent" vs. "low confidence" vs. "no memory") and decision outcome (correct vs. incorrect) as independent variables and pre-decision confidence, decision times, and post-decision confidence as dependent variables were computed. There were main effects of decision process on pre-decision confidence, $\underline{F}(2, 278) = 19.80$, $\underline{p} < .001$, $\underline{f} = .38$, decision times, $\underline{F}(2, 278) = 6.82$, $\underline{p} = .001$, $\underline{f} = .22$, and post-decision confidence, $\underline{F}(2, 278) = 36.08$, $\underline{p} < .001$, $\underline{f} = .51$. Post hoc $\underline{t}$-tests showed that the "absent" group decided faster ($\underline{M} = 9.7$ s), with higher pre-decision decision confidence ($\underline{M} = 66.8\%$), and higher post-decision confidence ($\underline{M} = 69.9\%$) than both the "low confidence" ($\underline{M}$s = 14.0 s, 56.4%, 45.6%) and "no memory" groups ($\underline{M}$s = 16.3 s, 43.4%, 40.7%), all $|\underline{t}|\text{s} \geq 2.56$, $|\underline{d}|\text{s} \geq 0.42$. The "low confidence" and "no memory" groups only differed with regard to their pre-decision confidence ratings ($\underline{M} = 56.4\%$ vs. 43.4%, respectively), $\underline{t}(78) = 2.63$, $\underline{p} = .010$, $\underline{d} = 0.31$. The main effect of decision outcome was significant for pre-decision confidence only, $\underline{F}(1, 278) = 6.02$, $\underline{p} = .015$, $\underline{d} = .29$. Specifically, pre-decision confidence ratings were higher for accurate (58.5%) than inaccurate (49.9%) non-choosers. The decision process by decision outcome interaction was non-significant for all three variables. There were no associations of identification accuracy with post-decision confidence or decision times, all $|\underline{r}|\text{s} \leq .21$, $\underline{p}\text{s} \geq .183$.

Table 4

Means, Standard Deviations, and Intercorrelations of Postdictors of Identification Accuracy for Nonchoosers as a Function of Reasons Given for Not Choosing

	M	SD	95% CI	2	3	4	5	6	7
	"Target was not among them" (n = 204)								
1 Accuracy	72.5%	44.7	66.4-78-6%	.11	.02	-.05	-.09	.03	.05
2 Pre-decision confidence	66.8%	21.4	63.9-69.7%		.38**	-.06	.03	.03	-.10
3 Post-decision confidence	69.9%	23.7	66.6-73.2%			-.18*	-.05	.12	-.08
4 Decision time	9.7 s	2.3	8.7-10.9 s				-.02	.21**	.45**
5 Viewing time	27.9 s	10.4	26.4-29.3 s					.24**	.03
6 Estimated viewing time	20.0 s	2.6	17.6-22.7 s						.48**
7 Estimated decision time	9.0 s	3.0	7.7-10.5 s						

Table 4 (continued)

	M	SD	95% CI	2	3	4	5	6	7
	“Low confidence” (n = 39)								
1 Accuracy	64.1%	48.6	48.8-79.4%	.24	.12	-.17	.13	-.00	.05
2 Pre-decision confidence	56.4%	22.9	49.2-63.6%		.39*	-.09	.06	.04	-.14
3 Post-decision confidence	45.6%	21.7	38.8-52.4%			-.22	.13	-.03	-.24
4 Decision time	14.0 s	2.4	10.7-18.4 s				.18	.03	.46**
5 Viewing time	27.0 s	9.9	23.8-30.2 s					.51**	.50**
6 Estimated viewing time	19.5 s	2.4	14.7-25.7 s						.52**
7 Estimated decision time	14.2 s	2.6	10.4-19.3 s						

Table 4 (continued)

	M	SD	95% CI	2	3	4	5	6	7
	“No memory” (n = 41)								
1 Accuracy	63.4%	48.8	48.5-78.3%	.22	.19	.21	.21	-.02	-.15
2 Pre-decision confidence	43.4%	21.3	36.8-49.8%		.21	.14	.08	.25	.12
3 Post-decision confidence	40.7%	22.4	33.8-47.6%			.17	.32	.30	-.25
4 Decision time	16.3 s	2.1	13.1-20.3 s				.33*	.15	.36*
5 Viewing time	30.2 s	10.9	26.7-33.6 s					.41**	-.04
6 Estimated viewing time	20.0 s	2.7	14.7-27.2 s						.10
7 Estimated decision time	14.5 s	2.9	10.4-20.3 s						

* $p \leq .05$; ** $p \leq .01$.

Discussion

In the present study three major research questions were addressed: First, we analyzed the postdictive value of the combination of three assessment variables for choosers. Second, the usefulness of estimated decision times as a postdictor of identification accuracy was examined. Third, we tested the utility of self-reported decision process of nonchoosers as assessment variable. The issues of stimulus sampling, generalizability as well as ecological validity were addressed by carrying out a large scale field study with 720 participants and ten targets.

Previous research to test the usefulness of assessment variables has usually looked at post-decision confidence, decision times or decision processes only in isolation. In contrast, we sought to test their postdictive value relative to each other as well as in combination. We emphasize that the postdictive values of these assessment variables is limited to their measurement in the immediate context of the identification (Sporer, 1993), that is, uncontaminated by postidentification feedback or other factors distorting their postdictive value.

A number of findings from the literature were replicated in the present field experiment: As in other studies, significant positive CA and negative LA associations were found only for choosers but not for nonchoosers (Dunning & Stern, 1994; Dunning & Perretta, 2002; S. M. Smith et al., 2000; Sporer, 1993, 1994; Sporer et al., 1995; Weber et al., 2004; Weber & Brewer, 2006), resulting in the expected interaction between Choice and Decision Outcome (Sporer, 1994). For choosers, the observed point-biserial correlations are considered large for confidence ($\underline{r}$ = .39) and medium to large for decision times ($\underline{r}$ = -.30; Cohen, 1988, p. 81). Calibrational analyses replicated the finding that there is high capability to discriminate accurate from incorrect choosers but not nonchoosers via post-decision confidence (Brewer & Wells, 2006; Brewer, Keast et al., 2002; Weber & Brewer, 2003, 2004, 2006).

We know of no study where pre-decision confidence was calibrated. Here, calibration measures of pre-decision confidence showed only little capability to discriminate accurate from inaccurate witnesses. This is consistent with many correlational studies on the pre-decision confidence accuracy relation (e.g., Cutler & Penrod, 1988; Cutler, Penrod, O'Rourke, & Martens, 1986; Cutler, Penrod, & Martens, 1987; Gwyer & Clifford, 1997; Sporer, 1992a). Nevertheless, despite low calibration we did find a significant positive correlation between pre-decision confidence and identification accuracy not only for choosers but also for non-choosers, and hence for the whole sample. This is consistent with the results of Sporer (1993). Over a wide range of abilities and knowledge, broad agreement between self-assessed and objective knowledge has been found (Ackerman, Beier, & Bowen, 2002; Perfect, 2004). However, when confronted with a lineup, witnesses normally have had no experience with this task and they do not know what the identification task will be like (see Perfect, 2004; Perfect & Hollins, 1996). This may be one reason for the lack of reliability of the association between pre-decision confidence and identification accuracy. Specifically, witnesses usually are not told the number of lineup members, the perspective from which the pictures were taken (frontal, 3/4 pose or profile), or how similar the foils will be to the suspect (e.g., Cutler & Penrod, 1988; Cutler Penrod, & Martens, 1987; Cutler, Penrod, O'Rourke, & Martens, 1986; Gwyer & Clifford, 1997; Sporer, 1992a, 1993). Considering these circumstances, it is not surprising that pre-decision confidence ratings often lack an association with identification accuracy. In the present study, we informed witnesses that they would be asked to make an identification from a lineup consisting of six portrait (frontal) pictures. Of course, the number of lineup members should not be disclosed when the sequential lineup method is used (R. C. L. Lindsay & Wells, 1985; Sporer, 1993). Nevertheless, information about the angle and quality of the pictures could help witnesses to make better calibrated estimates of pre-decision confidence.

We successfully combined post-decision confidence and decision time via boundary analyses which resulted in correct classifications of 97.2% of the fast and confident choosers. To overcome the argument of post-hoc criterion setting, we split our sample of choosers into two halves and postdicted accuracy of the first subsample by using the boundaries established in the second subsample. The obtained results were comparable to those observed for the whole sample with somewhat lower but yet high correct classification rates. Although there was an association between RK(F) judgment and identification accuracy in the RKF condition, the RKF judgment did not further improve classification rates. Furthermore, no postdictive value of RKF judgments was found when we tried to postdict accuracy from the results obtained from subsample 1 to subsample 2. Although we obtained a high number of correctly classified choosers within the established boundaries, it should be clear that our postdiction rate holds only for a small subgroup of participants, viz. n = 57 fast and confident choosers, given there were 436 choosers in our total sample of 720. Replacing decision times with estimated decision times led to comparable results, with a correct classification rate of 96.7%, which, however, applied to an even smaller number of participants (n = 24). Admittedly, there was also a number of choosers outside one or both boundaries that were accurate.

Boundary analyses for decision times and post-decision confidence revealed optimum boundaries of 6 s and 90%. This is inconsistent to the 10 – 12 s rule established by Dunning and Perretta (2002). Sauerland and Sporer (in press) found optimum boundaries of 18 s and 50% confidence. One difference between the present and the Sauerland and Sporer (in press) study was the retention interval between witnessing the event and the identification task. Sauerland and Sporer used a 1-week retention interval, whereas in the present study it was only 30 s. Furthermore, Sauerland and Sporer (in press) used laboratory methodology with a filmed event. Likewise, Brewer, Caon et al. (2006) found shorter optimum time boundaries for participants who were tested immediately (13 s) than for those who were tested after 15 min (36 s) or 20 min (35 s). Weber et al. (2004) found optimum time boundaries that varied widely between 5 s and 28 - 29 s across four laboratory studies with retention intervals of 15 - 20 min between a filmed event and testing. With the growing number of studies on the topic, a meta-analysis could shed light on those variables that have an impact on the optimum time and confidence boundaries. As of now, the optimum boundaries seem to vary from study to study, thus making it difficult to derive recommendations for practice. Furthermore, in real cases, decision times and confidence may vary numberwise from those that we obtain in laboratory or field studies, for example due to the awareness of the severe consequences of false identifications and false rejections. It would be interesting to collect decision times and confidence ratings in real cases so they can be compared to the descriptive data obtained in laboratory/field studies. Undoubtedly, it would be a great contribution to the field of identification if data were collected even where DNA samples exist, so that identification accuracy could actually be assessed in real cases.

There was a tendency of Know judgments to be associated with more accurate identifications than Remember judgments in both the RK and the RKF condition (cf., Davids, 2006), thus supporting the assumptions made by Dunning and Stern (1994). On the contrary, we found no support for Rotello et al.'s (2004) arguments. However, only in the RKF condition did a significant association of RKF judgment and identification accuracy emerge: Witnesses who made a Familiar judgment were significantly less often accurate in their identification decision than those who made a Know judgment. When looking at the distribution of correct and incorrect decisions in the RKF and the RK conditions, it seems that many participants who would have made a Know judgment in the RK condition tended to make a Familiar judgment in the RKF condition, resulting in higher accuracy rates of Know judgments than Familiar judgments in the RKF condition. Remember judgments also tended to be more accurate in the RKF than in the RK condition. According to these results it is not so crucial whether a witness Remembers or Knows that a person is that target, but whether s/he only thought the face was familiar. Practitioners should be very cautious with identification evidence given on the grounds that the face simply looked more familiar than the other ones (see Behrman & Richards, 2005).

In practice, measured decision times may not always be available. Therefore, the second aim of the study was to analyze the utility of estimated decision times as a postdictor of identification accuracy. An estimation of the time taken for the identification may then be an alternative (albeit crude) way of gaining information about the witness's decision process. The fact that estimated and measured decision times were highly positively associated for the whole sample legitimates our procedure. Altogether, the pattern of results for estimated decision times paralleled the one for measured decision times, even though the effect sizes were somewhat smaller. Specifically, the association between estimated decision times and identification accuracy was significant for choosers but not nonchoosers resulting in the Choice by Decision Outcome interaction. Although we know of no study that tested the usefulness of estimated decision times after longer time intervals, we are aware that the association with identification accuracy may be distorted due to feedback (cf., Semmler, Brewer, & Wells, 2004; Wells & Bradfield, 1998; Wells, Olson, & Charman, 2003). Wells and colleagues (Wells & Bradfield, 1998, Experiment 1; Wells, Olson et al., 2003) found that participants receiving confirming feedback estimated their decision time as being significantly longer than those not receiving feedback. No such effect was found for disconfirming feedback. Thus, decision times estimated retroactively by a witness at a later interrogation or in the courtroom may not have any probative value. The effect of retention interval and feedback on the decision time estimation with and without feedback should be tested more thoroughly in future studies.

The third aim of the present study was to identify a postdictor for nonchoosers. It was expected that those nonchoosers who indicated that they rejected the lineup because "The target was absent" would be more accurate than those who simply indicated that “The target was present, but I was not confident enough“ or who “did not know whether the person was present“. There was only a non-significant trend in this direction (Cramer's V = .09). However, as expected, the "absent" group made their decisions faster and with higher pre- and post-decision confidence. The only difference that emerged between the "low confidence" and the "no memory" group was for pre-decision confidence. Contrary to the hypothesis, the "low confidence" group showed higher pre-decision confidence than the "no memory" group.

An explanation for the lack of association between the postdictors and identification accuracy other than the heterogeneity of nonchoosers may be that choosers’ and nonchoosers’ identification decision possesses asymmetric features (Weber & Brewer, 2004). Whereas choosers’ decisions are likely to be based on a match between their memory for the target and one of several faces shown in a lineup (Sporer, 1993), nonchoosers’ decisions are based on a failure to match their memory to any of the faces in the lineup. Weber and Brewer (2006) argued that confidence ratings of nonchoosers are not based on confidence for the most likely match (face), because then one would expect the CA relationship of nonchoosers to be similar to the one observed for choosers, which is not the case. Instead, the authors suggested that confidence of nonchoosers reflected the average match between all lineup faces and the target. This hypothesis could be tested by either asking nonchoosers about their confidence with regard to each lineup member (as done by Sporer, 1993, with sequential lineups), or by using one-person lineups (i.e., showups). Following Weber and Brewer's (2006) reasoning, the associations between postdictors and identification accuracy should be observed when confidence and decision time measures apply to only one lineup member. Unfortunately, many studies with showups either do not report the CA correlation at all, or if they do, they do not differentiate between choosers and nonchoosers (e.g., Beal, Schmitt, & Dekle, 1995; Gonzales, Ellsworth, & Pembroke, 1993; R. C. L. Lindsay, Pozzulo, Craig, Lee, & Corbers, 1997; Yarmey, Yarmey, & Yarmey, 1996). Dysart, R. C. L. Lindsay, and Dupuis (2006) found a significant CA correlation for both choosers ($\underline{r}$ = .32) and nonchoosers ($\underline{r}$ = .19), supporting Weber and Brewer's (2006) hypothesis. Dysart, R. C. L. Lindsay, MacDonald, and Wicke (2002) reported neither a significant CA association for choosers nor for nonchoosers when asking bar patrons (many of which were intoxicated with alcohol) to make an identification from a showup.

Some caveats need to be discussed. First, although ecological validity was enhanced by using a natural setting, there was no crime scenario in the present study. Thus, it is unlikely that participants experienced the same arousal level as witnesses of a real crime would. However, it is equally unlikely that participants in studies with filmed crimes do. The well known inverted-U relationship between arousal and memory suggests that identification accuracy should be higher for events that have a medium arousal level, but should be lower in extremely arousing situations, as witnessing serious crimes may be. Quite on the contrary, recent research demonstrated that arousal with regard to life events has an impact on a feeling of remembering, the vividness of the memory (Sharot, Delgado, & Phelps, 2004), and perceived accuracy (Talarico & Rubin, 2003) whereas was actual accuracy unaffected. Yet again, others (Kensinger & Schacter, 2006) found that negative arousal affected not only vividness, but also accuracy. The data base on the issue seems to be inconclusive at this time and further studies are needed to find out about the effect that arousal has on identification accuracy, as well as on confidence and RKF judgments.

A second caveat in the present study is the short (30 s) interval between the event and the identification task. In real cases, there are normally days, if not months between witnessing the event and the identification (Behrman & Richards, 2005). It is likely, that the memory for the target would have decreased after a longer retention interval. However, leaving a longer interval would have made it extremely difficult to track the participants and would most likely have led to a large number of drop outs. Nonetheless, the main body of laboratory studies in the identification field uses time intervals of 30 min or less. We believe that there should be at least an interval of one day between event and identification when there are no severe practical reasons that speak against this.

To summarize, we successfully combined decision times and post-decision confidence and found that fast and confident choosers are highly accurate. Additionally, estimated decision times also qualified as postdictor of identification accuracy. The issue of how to proceed with slower and less confident choosers, however, remains unsolved. For future studies, we emphasized the contributive value that real cases could have. Our attempt to identify a postdictor of nonchoosers' performance failed. Studies that look at nonchoosers' showup decisions could help us find out more about nonchoosers' decision processes. In conclusion, we are optimistic that further research on system and assessment variables along the lines presented here should increase the probative value of eyewitness identification evidence, and reduce miscarriages of justice.

References

Ackerman, P. L., Beier, M. E., & Bowen, K. R. (2002). What we really know about our abilities and our knowledge. Personality and Individual Differences, 33, 587-605.

Baranski, J. V., & Petrusic, W. M. (1994). The calibration and resolution of confidence in perceptual judgments. Perception & Psychophysics, 55, 412-428.

Beal, C. R., Schmitt, K. L., & Dekle, D. J. (1995). Eyewitness identification of children: Effects of absolute judgments, nonverbal response options, and event encoding. Law and Human Behavior, 19, 197-216.

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.

Bothwell, R. K., Deffenbacher, K. A., & Brigham, J. C. (1987). Correlation of eyewitness accuracy and confidence: Optimality hypothesis revised. Journal of Applied Psychology, 72, 691-695.

Brewer, N. (2006). Uses and abuses of eyewitness identification confidence. Legal and Criminological Psychology, 11, 3-23.

Brewer, N., Caon, A., Todd, C., & Weber, N. (2006). Eyewitness identification accuracy and response latency. Law and Human Behavior, 30, 31-50.

Brewer, N., Gordon, M., & Bond, N. (2000). Effect of photoarray exposure duration on eyewitness identification accuracy and processing strategy. Psychology, Crime & Law, 6, 21-32.

Brewer, N., Keast, A., & Rishworth, A. (2002). The confidence-accuracy relationship in eyewitness identification: The effects of reflection and disconfirmation on correlation and calibration. Journal of Experimental Psychology: Applied, 8, 44-56.

Brewer, N., Palmer, M., McKinnon, A., & Weber, N. (2005, March). Decision processes underlying eyewitness identifications. Paper presented at the Annual Meeting of the American Psychology-Law Society, in Phoenix, AZ.

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.

Brigham, J. C. (1990). Target person distinctiveness and attractiveness as moderator variables in the confidence-accuracy relationship in eyewitness identifications. Basic and Applied Social Psychology, 11, 101-115.

Brigham, J. C., Van Verst, M., & Bothwell, R. K. (1986). Accuracy of children's eyewitness identifications in a field setting. Basic and Applied Social Psychology, 7, 295-306.

Caputo, D. D., & Dunning, D. (2005). Eyewitness accuracy and decision processes using sequential lineups. Manuscript submitted for publication.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Conway, M. A., & Dewhurst, S. A. (1995). The self and recollective experience. Applied Cognitive Psychology, 9, 1-19.

Conway, M. A., Gardiner, J. M., Perfect, T., Anderson, S. J., & Cohen, G. M. (1997). Changes in memory awareness during learning: The acquisition of knowledge by psychology undergraduates. Journal of Experimental Psychology: General, 126, 393-413.

Cutler, B. L., & Penrod, S. D. (1988). Improving the reliability of eyewitness identification: Lineup construction and presentation. Journal of Applied Psychology, 73, 281-290.

Cutler, B. L., & Penrod, S. D. (1989). Moderators of the confidence-accuracy correlation in face recognition: The role of information processing and base-rates. Applied Cognitive Psychology, 3, 95-107.

Cutler, B. L., & Penrod, S. D. (1995). Mistaken identity. Cambridge: Cambridge University Press.

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.

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 Behaviour, 1, 113-134.

Davids, M. S. C. (2006). Knowing what you know: Meta-memory and the effects of verbal overshadowing on eyewitness identification testimony. Unpublished master's thesis, University of Giessen, Germany.

Donaldson, W. (1996). The role of decision processes in remembering and knowing. Memory & Cognition, 24, 523-533.

Dunn, J. C. (2004). Remember-know: A matter of confidence. Psychological Review, 111, 524-542.

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.

Dunning, D. &. Stern, L. B. (1994). Distinguishing accurate from inaccurate eyewitness identification via inquiries about decision processes. Journal of Personality and Social Psychology, 67, 818-835.

Dysart, J. E., Lindsay, R. C. L., & Dupuis, P. R. (2006). Show-ups: The critical issue of clothing bias. Applied Cognitive Psychology, 20, 1009-1023.

Dysart, J. E., Lindsay, R. C. L., MacDonald, T. K., & Wicke, C. (2002). The intoxicated witness: Effects of alcohol on identification accuracy from showups. Journal of Applied Psychology, 87, 170-175.

Gardiner, J. M. (1988). Functional aspects of recollective experience. Memory and Cognition, 16, 309-313.

Gardiner, J. M. (2001). Episodic memory and autonoetic consciousness: A first person approach. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences, 356, 1351-1361.

Gardiner, J. M., & Java, R. I. (1990). Recollective experience in word and nonword recognition. Memory and Cognition, 18, 23-30.

Gardiner, J. M., & Java, R. I. (1991). Forgetting in recognition memory with and without recollective experience. Memory and Cognition, 19, 617-623.

Gardiner, J. M., Ramponi, C., & Richardson-Klavehn, A. (2002). Recognition memory and decision processes: A meta-analysis of remember, know, and guess responses. Memory, 10, 83-98.

Gonzalez, R., Ellsworth, P. C., & Pembroke, M. (1993). Response biases in lineups and showups. Journal of Personality and Social Psychology, 64, 525-537.

Gwyer, P., & Clifford, B. R. (1997). The effects of the cognitive interview on recall, identification, confidence and the confidence/accuracy relationship. Applied Cognitive Psychology, 11, 121-145.

Hirshman, E., & Master, S. (1997). Modeling the conscious correlates of recognition memory: Reflections on the remember-know paradigm. Memory & Cognition, 25, 345-351.

Juslin, P., Olsson, N., & Winman, A. (1996). Calibration and diagnosticity of confidence in eyewitness identification: Comments on what can be inferred from the low confidence-accuracy correlation. Journal of Experimental Psychology: Learning, Memory, and Cognition, 22, 1304-1316.

Kensinger, E. A., & Schacter, D. L. (2006). Reality monitoring and memory distortion: Effects of negative, arousing content. Memory & Cognition, 34, 251-260.

Kneller, W., Memon, A., & Stevenage, S. (2001). Simultaneous and sequential lineups: Decision processes of accurate and inaccurate eyewitnesses. Applied Cognitive Psychology, 15, 659-671.

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

Lindsay, D. S., Read, J. D. & Sharma, K. (1998). Accuracy and confidence in person identification: The relationship is strong when witnessing conditions vary widely. Psychological Science, 9, 215–218.

Lindsay, R. C. L., Pozzulo, J. D., Craig, W., Lee, K., & Corber, S. (1997). Simultaneous lineups, sequential lineups, and showups: Eyewitness identification decisions of adults and children. Law and Human Behavior, 21, 391-404.

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

Mäntylä, T. (1997). Recollections of faces: Remembering differences and knowing similarities. Journal of Experimental Psychology: Learning, Memory, and Cognition, 23, 1203-1216.

Olsson, N., & Juslin, P. (1999). Can self-reported encoding strategy and recognition skill be diagnostic of performance in eyewitness identification? Journal of Applied Psychology, 84, 42-49.

Olsson, N., Juslin, P., & Winman, A. (1998). Realism of confidence in earwitness versus eyewitness identification. Journal of Experimental Psychology: Applied, 4, 101-118.

Parkin, A. J., & Walter, B. M. (1992). Recollective experience, normal aging, and frontal dysfunction. Psychology and Aging 7, 290-298.

Penrod, S. D. (2006, March). Eyewitness Choosing and Guessing. Paper presented at the Annual Meeting of the American Psychology-Law Society, in Tampa, FL.

Perfect, T. J. (2004). The role of self-rated ability in the accuracy of confidence judgements in eyewitness memory and general knowledge. Applied Cognitive Psychology, 18, 157-168.

Perfect, T. J., & Hollins, T. S. (1996). Predictive feeling of knowing judgements and postdictive confidence judgements in eyewitness memory and general knowledge. Applied Cognitive Psychology, 10, 371-382.

Perfect, T. J., Mayes, A. R., Downes, J. J., & Van Eijk, R. (1996). Does context discriminate recollection from familiarity in recognition memory? Quarterly Journal of Experimental Psychology, A49, 797-813.

Rajaram, S. (1993). Remembering and knowing: Two means of access to the personal past. Memory and Cognition, 21, 89-102.

Rajaram, S. (1996). Perceptual effects on remembering: Recollective processes in picture recognition memory. Journal of Experimental Psychology: Learning, Memory and Cognition, 22, 365-377.

Rajaram, S., & Geraci, L. (2000). Conceptual fluency selectively influences knowing. Journal of Experimental Psychology: Learning, Memory and Cognition, 26, 1070-1074.

Rotello, C. M., Macmillan, N. A., & Reeder, J. A. (2004). Sum-difference theory of remembering and knowing: A two-dimensional signal-detection model. Psychological Review, 111, 588-616.

Scheck, B., Neufeld, P., & Dwyer, J., (2000). Actual innocence: When justice goes wrong and how to make it right. New York: Doubleday.

Sauerland, M., & Sporer, S. L. (in press). Post-decision confidence, decision time, and self-reported decision processes as postdictors of identification accuracy. Psychology, Crime, & Law.

Sharot, T., Delgado, M. R., & Phelps, E. A. (2004). How emotion enhances the feeling of remembering. Nature Neuroscience, 7, 1376-1380.

Semmler, C., Brewer, N., & Wells, G. L. (2004). Effects of postidentification feedback on eyewitness identification and nonidentification confidence. Journal of Applied Psychology, 89, 1-13.

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.

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

Sporer, S. L. (1993). Eyewitness identification accuracy, confidence and decision times in simultaneous and sequential lineups. Journal of Applied Psychology, 78, 22-33.

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). New York: Cambridge University Press.

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

Tabachnick, B. G., & Fidell, L. S. (2006). Using multivariate statistics. New York: Pearson.

Talarico J. M., & Rubin, D. C. (2003). Confidence, not consistency, characterizes flashbulb memories. Psychological Science, 14, 455-461.

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

Weber, N., & Brewer, N. (2003). The effect of judgment type and confidence scale on confidence-accuracy calibration in face recognition. Journal of Applied Psychology, 88, 490-499.

Weber, N., & Brewer, N. (2004). Confidence-accuracy calibration in absolute and relative face recognition judgments. Journal of Experimental Psychology: Applied, 10, 156-172.

Weber, N., & Brewer, N. (2006). Positive versus negative face recognition decisions: Confidence, accuracy, and response latency. Applied Cognitive Psychology, 20, 17-31.

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.

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.

Wells, G. L., Olson, E. A., & Charman, S. D. (2003). Distorted retrospective eyewitness reports as functions of feedback and delay. Journal of Experimental Psychology: Applied, 9, 42-52.

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.

Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling and social psychology experimentation. Personality and Social Psychology Bulletin, 25, 1115-1125.

Winer, B. J. (1971). Statistical principles in experimental design. London: McGraw Hill.

Wright, D. B., & Sladden, B. (2003). An own gender bias and the importance of hair in face recognition. Acta Psychologica, 114, 101-114.

Yaniv, I., Yates, J. F., & Smith, J. E. K. (1991). Measures of discrimination skill in probabilistic judgment. Psychological Bulletin, 110, 611-617.

Yarmey, A. D. (2004). Eyewitness recall and photo identification: A field experiment. Psychology, Crime & Law, 10, 53-68.

Yarmey, A. D., Yarmey, M. J., & Yarmey, A. L. (1996). Accuracy of eyewitness identifications in showups and lineups. Law and Human Behavior, 20, 459-477.

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. Journal of Experimental Psychology: Leaming, Memory, and Cognition, 20, 1341-1354.

Yonelinas, A. P. (2001). Consciousness, control, and confidence: The 3 cs of recognition memory. Journal of Experimental Psychology: General, 130, 361-379.

Appendix

Table A1

Identification Accuracy (in %) and Choosing Rates (in %) as a Function of the Target Person and Target Presence

Target	1	2	3	4	5	6	7	8	9	10	Total
Identification accuracy (%)											
TA and TP	43.1	70.8	65.3	45.8	62.5	58.3	41.7	50.0	61.1	70.8	56.9
	(31)	(51)	(47)	(33)	(45)	(42)	(30)	(36)	(44)	(51)	(410)
TP only	41.7	80.6	75.0	38.9	63.9	72.2	33.3	41.7	66.7	72.2	58.6
	(15)	(29)	(27)	(14)	(23)	(26)	(12)	(15)	(24)	(36)	(211)
TA only	44.4	61.1	55.6	52.8	61.1	44.4	50.0	58.3	55.6	69.4	55.3
	(16)	(22)	(20)	(19)	(22)	(16)	(18)	(21)	(20)	(25)	(199)

Table A1 (continued)

Target	1	2	3	4	5	6	7	8	9	10	Total
Choosing rate (%)											
TA and TP	62.5	61.1	63.9	61.1	58.3	69.4	52.8	52.8	63.9	59.7	60.6
	(45)	(44)	(46)	(44)	(42)	(50)	(38)	(38)	(46)	(43)	(436)
TP only	69.4	83.3	83.3	75.0	77.8	83.3	55.6	63.9	83.3	88.9	76.4
	(25)	(30)	(30)	(27)	(28)	(30)	(20)	(23)	(30)	(32)	(275)
TA only	55.6	38.9	44.4	47.2	38.9	55.6	50.0	41.7	44.4	30.6	44.7
	(20)	(14)	(16)	(17)	(14)	(20)	(18)	(15)	(16)	(11)	(161)

Note. Figures enclosed in parentheses represent absolute frequencies. TA = target-absent. TP = target-present.

Table A2

CA Correlations, Decision Time-Accuracy Correlations, Post-Decision Confidence Boundary (in %), and Time Boundary (in s) as a Function of the Target Person (Choosers only)

Target	1	2	3	4	5	6	7	8	9	10	Total
Mean post-decision confidence (%)	57.1	63.4	60.9	54.8	62.6	64.4	56.8	52.1	63.0	68.8	60.6
Mean decision time (s)[1]	10.1	9.1	8.6	13.7	9.2	7.9	9.0	11.1	10.2	6.8	9.4
CA relation (r)	.43**	.57**	.38**	.29°	.39*	.19	.47**	.02	.56**	.43**	.39**
LA relation (r)	-.25	-.40**	-.27	-.14	-.40**	-.13	-.25	.01	-.45**	-.43**	-.30**
Post-decision confidence boundary (%)	80	60	70	100	90	20	90	ns[2]	70	90	90
Time boundary (in s)	4	6	5	ns[2]	6	5 - 6	5	ns[2]	14 - 16	10 - 12	6

[1]The means of decision times were backtransformed from the logarithmic values used for the inferential analyses. [2]When there was no significant association between the postdictor and identification accuracy, no boundary was computed.

*$p \leq .05$, two-tailed. **$p \leq .01$, two-tailed.

EXPERIMENT 4

The Application of Multiple Lineups in a Field Study

When attempting to establish if a suspect actually is the culprit, investigators can frequently rely on DNA analysis. However, DNA samples often do not exist, and then eyewitness identification testimony may be the only evidence available. For the past 30 years, research in this field has dealt with numerous factors that can have either a positive or negative impact on identification performance (see Wells & Olson, 2003). Estimator variables are those that cannot be controlled by the criminal justice system (e.g., lighting conditions; Wells, 1978). The focus of the present study are system or control variables (Wells, 1978), over which the criminal justice system normally has control (e.g., instructions to witnesses and lineup composition). Specifically, we tested the usefulness of multiple lineups in order to determine whether a particular identification decision is correct or incorrect. Note, that here, multiple lineup presentation does not indicate the repeated presentation of one lineup to a witness but the presentation of several lineups depicting different aspects of a person (face, body, clothing) to one witness.

Furthermore, the present study addresses one shortcoming that affects most identification studies. Although the need for a sufficient participant sample sizes seems obvious to most, the need to sample stimuli appears to be far less straightforward (Wells & Windschitl, 1999). Whereas some studies have reported results for two targets (Brewer, Caon, Todd, & Weber, 2006, Experiment 2; Brewer & Wells, 2006; D. S. Lindsay, Read, & Sharma, 1998), very few studies have used more (Brigham, 1990). Studies that did use two targets have reported different results for different targets (Brewer et al., 2006, Experiment 2; Brewer & Wells, 2006; D. S. Lindsay et al., 1998). This shows, that general conclusions based on one target could be misleading. Therefore, we included nine targets in the present study.

Even though a wealth of research on identification accuracy exists using laboratory staged events or filmed scenarios, only a small number of field experiments have been conducted to date (e.g., Brigham, Van Verst, & Bothwell, 1986; Krafka & Penrod, 1985; Yarmey, 2004). Nevertheless, field experiments have some potential to provide a more differentiated understanding of processes underlying eyewitness identification (Cutler & Penrod, 1995), compared to laboratory research. In order to enhance the ecological validity and generalizability of postdiction research, data were collected in a natural setting, in the pedestrian zone of a small university town.

MULTIPLE LINEUPS

In an identification procedure, the witness is commonly asked to identify the face of the target, as this seems to be the most prominent part of a person to be recognized. However, other aspects of a person such as the body, voice, gait or posture may also be subject to recognition. This could be particularly true in cases where the target's face could not be seen clearly or only for a short period of time. Drawing from the principle of encoding specifity (Tulving & Thomson, 1973), several studies have attempted to improve identification accuracy by providing different aspects of a person in one lineup (e.g., Cutler & Penrod, 1988; Cutler, Penrod, & Martens, 1987; Cutler & Fisher, 1990; Egan, Pittner, & Goldstein, 1977; Melara, Dewitt-Rickards, & O'Brien, 1989). Although some studies reported positive effects, the meta-analysis by Cutler, Berman, Penrod, and Fisher (1994) reported only small effect sizes, concluding that, averaged across studies, the presentation of different aspects of a person produced only trivial effects on identification accuracy.

Pryke, Lindsay, Dysart, and Dupuis (2004) argued that one reason for the lack of improvement in identification performance when presenting different aspects of a person in one lineup (cf., Cutler et al., 1994) may be that the face overwhelmed the other aspects of the person. Specifically, witnesses may make a choice on the basis of the face, while discounting inconsistencies of body or voice. Therefore, the authors argued that different aspects of a person should be presented independently, so that one aspect could not be overwhelmed by another. A clear advantage of such a procedure is that independent evidence can be obtained from one witness. For example, the probability that a witness identifies a suspect by chance out of a six person lineup is 1/6. When there are two witnesses, the probability that both identify the same person merely by chance is 1/36, and 1/216 with three witnesses. In many cases, there is only one witness which brings us back to a guessing probability of 1/6. However, the probability that this witness identifies the suspect by chance can be reduced to 1/36 by presenting a second, say body lineup, and to 1/216 when a third lineup (e.g., voice) is added.

Lindsay, Wallbridge, and Drennan (1987) presented witnesses with two independent lineups, one in which the face of the target was to be identified, and a second one in which a piece of clothing that the target wore during the crime was to be identified. Note, that the clothing worn in the facial lineup was non-identical to the clothing worn during the crime. In order to establish the probative value of these identifications, the authors computed the diagnosticity ratio (Wells & Lindsay, 1980; Wells & Turtle, 1986). The diagnosticity ratio (DR) is defined as the ratio of correct to incorrect decisions. For target/suspect choices this refers to the ratio of hits (target choices) in target-present (TP) lineups and false alarms (suspect choices) in target-absent (TA) lineups. When the term equals 1, the lineup is neither diagnostic of the guilt nor of the innocence of the suspect. For nonchoosers the DR is the ratio of correct rejections in TA lineups and false rejections in TP lineups. DRs for foil choices can be established by dividing the proportion of foil identifications in TA lineups by the proportion of foil identifications in TP lineups (Wells, personal communication, March 2006). Another way to present the results is to look at the proportion of suspects that would be guilty if the procedure (i.e., multiple lineups) was applied to a large number of cases. To estimate this value, assumptions about the base rate of TA and TP lineups have to be made (Wells & Lindsay, 1980). According to Wells (as cited by Pryke et al., 2004), a prior probability near 50% is not unreasonable in real world cases. Thus, we set equal base rates (50%) of innocent versus guilty suspects in the lineup, so that the percentage guilty value is determined by the ratio of correct identifications to all attempted identifications. For lineup rejections, the percentage guilty value is determined by the ratio of false rejections to all rejections, and for foil choices by the ratio of foil identifications in TP to all foil identifications. Whereas the percentage guilty value should ideally approach 100 for target/suspect choices, the value for lineup rejections and foil choices should ideally approach 0.

Designation of an Innocent Suspect

One issue in analyzing eyewitness identification data is the designation of an innocent suspect. At least five methods of doing so can be found in the literature: First, the foil who resembles the target most according to a pilot study is selected to be the innocent suspect (e.g., Kneller, Memon, & Stevenage, 2001; Sporer, 1992, 1993, 2007). Second, one of several foils which are rated as similar to the target is randomly selected (e.g., Brewer et al., 2006; Clare & Lewandowsky, 2004; Juslin, Olsson, & Winman, 1996; Krafka & Penrod, 1985). Third, experienced policemen create the photo-spreads and select the innocent suspect (e.g., Brigham et al., 1986; Fleet, Brigham, & Bothwell, 1987; Olsson & Juslin, 1999). Fourth, the foil chosen most often is the innocent suspect (Pryke et al., 2004). Finally, Wells (personal communication, March 2006) argued that all foils could potentially be the innocent suspect and therefore recommended to compute the average across all foils. Indeed, choosing rates for a single foil can vary widely and therefore may have a great impact on the results. Therefore, the decision was made to analyze the data according to the averaging approach. In order to carry out the analyses, it was necessary that the foils in all TA lineups were the same. To make our results comparable to those of Pryke et al. (2004), results for the foil chosen most often in the portrait lineup will also be reported. By definition, the procedure applied by Pryke et al. (2004) results in smaller or equal DRs for portrait face lineups than the averaging method. However, for the remaining lineups one method or the other may lead to higher DR estimates.

Previous Research on Multiple Lineups

Lindsay et al. (1987) reported that the DR for face and clothing lineups was 13.60 whereas it was 1.91 for face lineups only. Thus, when a witness chose a face and then chose the matching clothing originally worn by that person in a second independent lineup, it was

13.6 times more likely that the identified person actually was the target than that it was an innocent suspect. However, was a witness presented with a facial lineup only, it was only 1.9 times more likely that the identified person actually was the target than that it was an innocent suspect.

To study the usefulness of multiple lineups more extensively, Pryke et al. (2004) presented witnesses with three or four independent lineups in two experiments. In Experiment 1, witnesses made independent decisions for simultaneous face and body lineups. A voice lineup was presented sequentially. The DR was computed separately for each of the lineups and for all combinations. The DR of target choices from the face lineup was notable (3.63) but rather low for the voice lineup (1.12) and the body lineup (0.65). The DR for choosing the face, body, and voice of the same person was 11.00. In other words, it was 11 times more likely that a chosen lineup member actually was the target when the witness picked the same person out of all three lineups than it was that the chosen person was the innocent suspect. The DR for the combination of the face and voice lineups was considerably smaller (3.52) when compared to that of all three lineups but also when compared to the DR of 13.60 obtained for face and clothing lineups by Lindsay et al. (1987). Pryke et al. (2004) discussed the possibility that clothing lineups were generally more diagnostic than voice lineups. Therefore, in Experiment 2, a clothing lineup was added and all lineups were presented sequentially in order to test the usefulness of multiple lineups in this presentation mode. Again, the DR of choices from the face lineup was considerable (4.67), but low for the voice lineup (0.38). The DR of the body lineup was 2.01 and 5.00 for the clothing lineup, replicating the notion that clothing lineups are more diagnostic than voice lineups. For some lineup combinations (face, body, and clothes; face, voice, and clothes; face, body, voice, and clothes) multiple identifications of the same person only occurred for TP but not TA lineups, causing undefined DRs due to division by zero. In these cases, estimation of guilt was 100%. Clearly, Experiment 2 replicated the success of multiple lineup administrations.

So far, the effect has been demonstrated for only three different targets (R. C. L. Lindsay et al., 1987; Pryke et al., 2004), and hence its generalizability remains unclear. In the present field study, witnesses responded to body, accessory (shopping bag) and two facial lineups. In addition to a portrait face lineup, witnesses were presented with a profile (90° angle) face lineup.

A profile face lineup was presented in addition to the portrait face lineup for the following reasons. Early face recognition studies suggested that recognition was easier from 3/4 view than from frontal or profile view (e.g., Patterson & Baddeley, 1977; Woodhead, Baddeley, & Simmonds, 1979, Experiment 2), while others failed to find such an effect (e.g., Davies, Ellis, & Shepherd, 1978, Experiment 2). In a thorough review of the literature Liu and Chaudhuri (2002) pointed to the impact of the amount of angular rotation between learning and testing. Specifically, when faces are presented frontally or in profile view in the study phase, the 3/4 pose represents the least amount of angular rotation at test, unless study and test stimuli are presented in the same pose. When the amount of angular rotation was separated from the effects of learning or test views, there was only little evidence for a 3/4 view advantage per se (Liu & Chaudhuri, 2002).

In eyewitness identification studies, participants are presented with moving persons during the learning phase, and the face of the target may be viewed in different poses. Our targets were instructed to show their faces both frontally and from the side while interacting with the participants. However, we expected that participants would not look at the target's face for the whole duration of the interaction. Thus, some participants may have seen the target primarily from the front, others more from the side. We expected that identification for participants would be easier from frontal or profile view, depending on the view that had dominated during the interaction. Therefore, we included both frontal and profile face view lineups in our study.

In addition, we included an accessory (bag) lineup, rather than a clothing lineup, because we thought it more likely that a culprit leaves an accessory like a bag than a piece of clothing at the scene of crime. Therefore a bag may be more likely to be available for constructing a lineup.

If multiple lineups really are independent sources of evidence, we would not expect to find a significant correlation between the performance in the individual lineups. This is contrary to the finding that performance in blank lineups (lineups that are known to be TA), are predictive of subsequent performance in TA and TP lineups (Wells, 1984). Generally, this notion presupposes that there is a positive correlation of identification accuracy decisions across lineup tasks. To our knowledge, the only study to have investigated this issue was Wells' (1984) study, however only for the sequence of TA (blank) facial lineups followed by TA or TP facial lineups. In order to test the usefulness of blank lineups for different types of lineups (i.e., portrait, body, bag, profile), witnesses were presented with two TA and two TP lineups, which were counterbalanced across participants. When the portrait face lineup was shown in TP mode, the body lineup was also shown as TP, the other two as TA, and vice versa. Of course, such a procedure is only meaningful if participants are explicitly instructed for each lineup the stimulus to be recognized may or may not be present. Please note, that the presentation of TA and TP lineups to the same person is accompanied by the fact that the DR can only be computed for those lineups that were shown in the same target presence mode.

If blank lineups are predictive of performance in other types of lineups, participants who correctly rejected the TA (blank) lineups should show higher performance in the TP lineups compared to those who falsely chose someone from TA lineups. If, however, performance in different lineup types is independent, there should be no predictive value of blank lineups. Furthermore, we expected two lineups (portrait face and body/profile face and accessory) to be more diagnostic than just the portrait or profile facial lineups by themselves.

METHOD

Participants

Six-hundred-forty-eight participants (324 male, 324 female; age 15 to 84, *Mdn* = 33 years) agreed to take part in this study when they were approached in the pedestrian zone in a small university town in Germany. Participants were representative of the general population of this university town which has 75,000 inhabitants and 23,000 students. They worked in numerous professions (36.3%), were students (27.5%), academics (15.9%), high school pupils (8.5%), retired (6.8%), housewives/-husbands (4.5%), or unemployed (0.6%).

Photo Lineups

Target persons were seven female[1] and two male psychology major students (age 20 to 37, Mdn = 20 years) from a university in a different town who received course credit for data collection.

Materials consisted of photos of faces (neck up), bodies (shoulders down), and bags. Lineups were mounted on a 30 x 33.5 cm display board and consisted of six pictures that were numbered 1 to 6 and arranged in two rows of three pictures each.

For facial photographs, jewelry, eyeglasses and hair accessories were taken off and hair was worn loose. On the body photographs, each lineup member wore different long sleeve clothing. The head was blotched out with an opaque oval covering the head and neck. The target shopping bag was a white cotton bag (55 x 56 cm), depicting a hand-drawn cow on a colorful background. The foil bags were similarly white and also had a colorful picture and/or writing on them. Portrait face and shopping bag lineup pictures were 9 x 13 cm photographs, body lineups 8.5 x 15 cm and profile pictures 9 x 9.5 cm in size.

For practical reasons, the target as well as the replacement photos always appeared at the same position, that is, position 3 for the portrait face lineups, position 4 for body lineups, position 2 for bag lineups, and position 5 for profile face lineups.

---

[1]Actually, 10 targets were involved. However, after the data collection, a mistake in the TA body lineup of one female target was detected, so that the data collected for this target was dismissed. Participants for whom this target was the interviewer, not the target, remained in the analysis.

Lineup Construction

Potential lineup members were recruited from a university town different from the investigation site based on their physical appearance resembling the respective targets. Each potential lineup member was photographed outside in front of the same wall of a dining hall. For the portrait face lineup, all foils, including the replacement, fit the general description of the target persons (Wells, Rydell, & Seelau, 1993) as determined by a pilot study with N = 55 mock witnesses (effective sizes, determined as Tredoux's Es were between 5.14 and 6.76; Tredoux, 1998, 1999). Following Pryke et al.'s (2004) procedure, little effort was made to select foils on the basis of similarity to features other than facial as witnesses usually do not provide detailed descriptions about the target's body (Sporer, 1996). Also, from a practical perspective, it is unlikely that the police engage in the time consuming procedure of matching foils to the suspect with regard to more than general similarity unless there are some unusual features, which was not the case for our targets (see Pryke et al., 2004).

In order to select foil bags and a replacement, participants of the pilot study ranked 11 cotton bags according to their likeness to the target bag. The bag with the highest ranking was chosen as replacement, the succeeding five as foils. Following Pryke et al.'s (2004) procedure for the clothing lineup, the six bags in the TA lineup were randomly assigned to lineup members. Random assignment of bags was performed for each individual target.

Procedure

Nine stimulus persons were involved in the present study. Data collection took place in a town different from their home university. Data were collected by pairs of two persons, one serving as target, one as interviewer. After data of eight participants were collected, the roles were changed. Thus, every interviewer collected data of 8 participants with each of the other persons.

The stimulus persons underwent three training sessions in which the interaction with the participants as both target and interviewer, time taking as well as controlling interaction time was practiced. During data collection, target persons wore different clothing from what they wore on their body lineup picture.

Specifically, the target asked a passer-by for directions to a certain location in the pedestrian zone of a university town. The conversation was scheduled to last between 15 s and 60 s. To ensure that the target was no longer seen by the passer-by after the interaction, targets only approached people whose walking direction was opposite of the location asked for.

The other person, the interviewer, watched the situation from a distance and recorded the interaction time with a stopwatch. Thirty seconds after the interaction between target and participant had terminated the interviewer approached the participant and explained that the true nature of the preceding interaction was a study on face recognition. If consent to participate was given, passers-by were handed a questionnaire. First, participants indicated their pre-decision confidence with regard to identifying the previously seen person from a 6-person portrait face lineup on an 11-point scale ranging from 0% to 100% (with intervals marked in 10% steps: 0%, 10%, ..., 100%). Then, participants were handed the display board with the portrait face lineup. To ensure the blindness of the interviewer, TP and TA lineups were placed into differently colored envelopes. The experimental design told interviewers which envelope to use for which participant. Thus, the interviewer pulled the display board out of the envelope and gave it to the participant so that he/she could only see the display board from the back, but not the pictures, that is, whether the target was present or not.

On the answer sheet, participants were advised that they would be asked to identify the person who had asked them for directions and that the person might or might not be present in the lineup. The instructions clearly stated that the person might or might not be present in the lineup. After the identification decision was made, participants indicated their post-decision confidence, again on an 11-point scale ranging from 0% to 100%.

Participants were not informed about the subsequent lineups until the post-decision confidence rating for the portrait face lineup was completed. After the portrait face lineup, the body, bag, and profile face lineups followed, always in this sequence. Participants were warned that the target may or may not be present for every lineup. After each lineup, participants were again asked to indicate their post-decision confidence. Pre-decision confidence was not collected in order to keep the amount of time needed to fill in the questionnaire within limits. For half of the participants, the portrait face and body lineups were TP, and the bag and profile face lineups TA, and vice versa.

After participants had completed all items, they were thanked and handed a card with the researchers' web address where some results of the study would be available after approximately three months.

RESULTS

First, descriptive statistics for the four lineup types and the correlations of identification performance and choosing rates between the lineups are reported. Second, we present DRs for the four lineup types individually (ignoring other decisions) followed by those for multiple lineup decisions[2].

Identification Accuracy and Choosing

An alpha level of .05 was used for all inferential analyses. Cramer's V and phi are reported as effect sizes for nonparametric analyses of 3 x 2 and 2 x 2 contingency tables, respectively. Hits and false alarms across targets can be obtained from Table 1. Table 2 shows the means and correlations of identification performance and choice across the four different lineup types. Identification accuracy for portrait face lineups was 57.7% (TA: 54.9%; TP: 60.5%), for body lineups 22.5% (TA: 26.2%; TP: 18.8%), for bag lineups 35.2% (TA: 59.0%; TP: 11.4%), and for profile face lineups 30.2% (TA: 32.8%; TP: 28.7%). Performance in TA and TP lineups differed significantly for body, $\chi^2(1, N = 648) = 5.09$, $p = .015$, $\phi = -0.09$, and bag lineups, $\chi^2(1, N = 648) = 160.48$, $p < .001$, $\phi = -0.50$, but not for the two facial lineups, χ^2s$(1, N = 648) \leq 2.05$, ps $\geq .088$, ϕs $\leq |0.06|$.

---

[2]Results for the individual targets can be obtained from the Appendix.

Table 1

Hits and False Alarms (%) Across Targets (N = 648)

		Frontal face	Body	Bag	Profile
Frontal face	Hits	60.5	18.8	11.4	28.7
	False alarms (average)[a]	7.5	12.3	6.8	11.4
	False alarms (designated suspect)[b]	19.7	4.3	7.1	7.1

[a]Innocent suspect identifications were computed as the mean of all positive identifications in TA lineups. [b]Foil chosen most often in the TA portrait face lineup was designated to be the innocent suspect.

Table 2

Means and Correlations of Identification Performance and Choosing Across Four Different Lineup Types (N = 648)

Lineup type			Body	Bag	Profile
	M (%)	SD		Identification accuracy	
Frontal face	57.7	49.4	.10**	.02	.03
Body	22.5	41.8		-.06	-.06
Bag	35.2	47.8			.04
Profile	30.2	46.0			
	M (%)	SD		Choosing	
Frontal face	61.4	48.7	.15**	.07	.22**
Body	75.5	43.1		.24**	.09*
Bag	39.5	48.9			.05
Profile	74.1	43.9			

*$p \leq .05$, two-tailed. **$p \leq .01$, two-tailed.

For all four lineup types, identification accuracy of choosers was significantly lower than identification accuracy of nonchoosers, all chi^2s(1, N = 648) ≥ 30.33, ps < .001, phis ≤ -0.22 (see Table 2). Specifically, in the portrait face lineup choosers' accuracy was 49.2%, nonchoosers' accuracy was 71.2%. In the body lineup, choosers made a correct decision in 12.5% of the cases, nonchoosers in 53.5% of the cases. In the bag lineup, choosers were accurate in 14.5% of the cases, nonchoosers in 48.7% of the cases. For the profile face lineup the numbers were 19.4% and 61.3%, respectively.

For identification accuracy, there was only one significant, albeit low correlation, which was the one between performance in portrait face and body lineups, r(646) = .10, p = 009. For choice, there were significant positive correlations of the body lineup with all other lineups (portrait: r(646) = .15, p < .001, bag: r(646) = .24, p < .001; profile: r(646) = .09, p = .025). There also was a significant correlation between the two facial lineups, r(646) = .22, p < .001.

Predicting Lineup Performance from Blank Lineups

Performance in blank (TA) lineups did not correlate with performance in TP lineups, r(646) = .00, p = .975. It made no difference, whether the portrait face and body lineups were blank, r(322) = -.07, p = .219, or the bag and the profile face lineups did not contain the suspect, r(322) = .02, p = .731.

Looking only at the two facial lineups neither lead to significant correlations, r(646) = -.05, p = .185, and it made no difference whether the portrait face lineup was blank, r(322) = .01, p = .823, or the profile face lineup, r(322) = .04, p = .513.

Table 3

Diagnosticity Ratios and % Guilty for Individual and Multiple Lineups Aross Nine Targets (N = 648)

	Target/suspect choice			Lineup rejection		Foil choice	
	DR (average)[a]	% guilty	DR (designated suspect)[b]	DR	% guilty	DR	% guilty
			Specific lineup modes				
Portrait	8.12	89	3.06	2.47	29	2.61	28
Body	1.53	60	4.35	1.15	47	1.26	44
Bag	1.67	62	1.54	0.95	51	1.55	39
Profile	2.52	72	4.65	1.58	39	1.33	43
			Multiple lineups				
Portrait	7.61	88	3.12	2.45	29	2.86	26
Body	0.58	37	-.-[c]	0.53	65	1.23	45
Portrait and body	10.43	91	2.86	2.52	28	1.77	36
Bag	1.49	60	1.26	0.79	56	1.58	39
Profile	2.49	71	4.47	1.61	38	1.33	43
Bag and profile	3.00	75	8.00	1.57	39	1.33	43

[a]Innocent suspect identifications were computed as the mean of all positive identifications in TA lineups. [b]Foil chosen most often in the TA portrait face lineup was designated to be the innocent suspect. [c]No DR because, by defintion, there are no accurate body identifications other than those that are also accurate portrait face identifications as the designated suspect is the person who was most frequently chosen from the portrait face lineup.

Diagnosticity of Lineups

Another way to look at these dependencies is to calculate the conditional probabilities that the body, bag, and profile lineup decisions are accurate, given that the first decision regarding the frontal face lineup was accurate or inaccurate, and given that the first decision was a choice or nonchoice. Diagnosticity was estimated for target/suspect choices, lineup rejections, and foil choices. Additionally, the percentage of suspects that would be guilty if the procedure (i.e., multiple lineups) was applied to a large number of cases was estimated as explained above.

DRs and % guilty for individual and multiple lineups for the total sample can be found in Table 3. The columns labeled DR (average) contain the DR as obtained when suspect identifications were computed as the mean of all positive identifications in TA lineups. The % guilty columns show estimates of the percentage of suspects that would be guilty if multiple lineups were applied to a large number of cases. For target/suspect choices the column labeled DR (designated suspect) contain DR estimates when the foil chosen most often in the TA portrait face lineup was designated to be the innocent suspect. By definition, the latter procedure results in smaller or equal DRs for portrait face lineups than the averaging method. However, for the remaining lineup types one method or the other may lead to higher DR estimates.

Diagnosticity as a Function of Specific Lineup Mode

In the top part of Table 3, results for target/suspect choices, lineup rejections and foil choices for decisions in each of the four lineup modes are reported. Note that for these results, the decisions made in the other lineup modes were not considered.

Overall, the DRs for target/suspect choices were higher than those for lineup rejections and foil choices. For all lineup decisions (rejections, target/suspect choices, foil choices), the portrait face lineup was most diagnostic.

Target/suspect choices. For target/suspect choices, performance in profile face lineups was larger than 2 (DR = 2.52, 72% guilty), even though much lower than for the portrait face lineup (DR = 8.12, 89% guilty). Body (DR = 1.53, 60% guilty) and bag lineups (DR = 1.67, 62% guilty) showed only weak diagnosticity. When designating the foil chosen most often in the TA portrait lineup to be the innocent suspect, estimates of DR were reduced substantially for the portrait face lineup (as expected by definition; DR = 3.06) but enhanced for the body (DR = 4.35) and profile face lineup (DR = 4.65).

Lineup rejections and foil choices. For lineup rejections and foil choices, the % guilty value should be low if the DR is high, indicating a high probability that the suspect is innocent.

For lineup rejections (DR = 2.47, 29% guilty), the portrait face lineup was the only one that was diagnostic, while performance for body (DR = 1.15), bag (DR = 0.95) and profile face (DR = 1.58) was only weakly or non-diagnostic of innocence. Similarly, for foil choices (DR = 2.61, 28% guilty), the portrait face lineup was the only one that was diagnostic, while performance for body (DR = 1.26), bag (DR = 1.55) and profile face (DR = 1.33) was only weakly or non-diagnostic of innocence.

Multiple Lineup Decisions

In the bottom part of Table 3, results for multiple lineup decisions are reported. For each participant, two lineups were TA and two TP, so that combined DRs could be computed only for combinations of the respective two lineups. The lines in the multiple lineups section of Table 3 that indicate DRs for only one lineup mode, for example, portrait face, were computed by eliminating those cases in which the same lineup member had additionally been chosen in the lineup that was presented in the same target presence mode (in this case the body lineup). Therefore, the values may differ from those presented in the specific lineup mode section of the table.

There are no direct tests to compare DRs against chance level. We can, however, test whether the result pattern obtained statistically significantly deviates from the null hypothesis that the proportions in each cell are the same. Our hypothesis was that suspect identifications/lineup rejections in addition to the facial identifications/rejections would be highly diagnostic of guilt. Thus, for target/suspect choices we computed $\underline{chi}^2$-values of the proportion of witnesses who did not identify the (innocent) suspect's face, identified just the (innocent) suspect's face, or the (innocent) suspect's face and an additional aspect in TA and TP lineups. For lineup rejections, we computed $\underline{chi}^2$ of the proportion of witnesses who did not reject the (innocent) suspect's face, rejected just the (innocent) suspect's face, or the (innocent) suspect's face and an additional aspect in TA and TP lineups. The results of these analyses will be reported after the DRs in the corresponding results section.

Target/suspect choices. Overall, portrait face lineups as well as the combination of the portrait face lineup with the body lineup were highly diagnostic of guilt. As expected, the combinations of portrait face and body lineups (DR = 10.43, 91% guilty) as well as bag and profile face lineups (DR = 3.00, 75% guilty) were more diagnostic than each lineup by itself. The portrait face lineup (DR = 7.61, 88% guilty) and the profile face lineup (DR = 2.49, 71% guilty) themselves were also diagnostic. Performance in body and bag lineups was poor (DRs ≤ 1.49). Designating the foil chosen most often in the TA portrait face lineup to be the innocent suspect reduced diagnosticity for the combination of the portrait face and body lineups to DR = 2.86, but increased diagnosticity for the combination of the bag and profile lineups to DR = 8.00.

Table 4 presents the proportion of witnesses who did not identify the suspect's face, identified just the suspect's face (portrait or profile), or the suspect's face and the second aspect of the person (body or bag, respectively) in TA and TP lineups. The results are displayed for the two methods of establishing the number of false alarms (average vs. designating the foil chosen most often in the TA portrait face lineup to be the innocent suspect). Non-identifications in portrait face and body lineups were diagnostic of innocence for the averaging (DR = 2.34, 30% guilty) and designated suspect method (DR = 2.03, 33% guilty). Identification of the portrait only was more diagnostic for the averaging (DR = 7.43, 88% guilty) than the designated suspect method (DR = 3.12, 76% guilty). Identification of the portrait and the body was more diagnostic for the averaging method (DR = 10.43, 91% guilty) than just the identification of the portrait, but not for the designated suspect method

Table 4

Suspect Choices for Portrait Face plus Body Lineups and Profile Face plus Bag Lineups as a Function of Target Presence (N = 648)

	Frequencies			Proportions-diagnosticity		
Condition	None	1	2	None	1	2
	Portrait face and body lineups					
TP	128	156	40	.40	.48	.12
TA (average)[a]	299.2	21.0	3.8	.92	.06	.01
DR				2.34[c]	7.43	10.43
% Guilty				30	88	91
TP	128	156	40	.40	.48	.12
TA (designated suspect)[b]	260	50	14	.80	.15	.04
DR				2.03[c]	3.12	2.86
% Guilty				33	76	74

Table 4 (continued)

Condition	Frequencies None	Frequencies 1	Frequencies 2	Proportions-diagnosticity None	Proportions-diagnosticity 1	Proportions-diagnosticity 2
	Profile face and bag lineups					
TP	231	85	8	.71	.26	.02
TA (average)[a]	287.2	34.2	2.7	.89	.11	.01
DR				1.24[c]	2.49	3.00
% Guilty				45	71	75
TP	231	85	8	.71	.26	.02
TA (designated suspect)[b]	304	19	1	.94	.06	.00
DR				1.32[c]	4.47	8.00
% Guilty				43	82	89

[a]Innocent suspect identifications were computed as the mean of all positive identifications in TA lineups. [b]Foil chosen most often in the TA portrait face lineup was designated to be the innocent suspect. [c]Diagnosticity of non-identifications is the ratio of lineup rejections in TA lineups to lineup rejections in the TP condition.

(DR = 2.86, 74% guilty). The pattern of frequencies generating these DRs differed from chance expectations for the averaging method, $chi^2(2, N = 648) = 200.90$, $p < .001$, Cramer's V = 0.56, and the designated suspect method, $chi^2(2, N = 648) = 111.97$, $p < .001$, Cramer's V = 0.42. However, when excluding nonchoosers, the effect became non-significant for the averaging, $chi^2(1, N = 648) = 0.27$, $p = .603$, phi = -0.02, and designated suspect method, $chi^2(1, N = 648) = 0.06$, $p = .806$, phi = -0.01.

Non-identifications in profile face and bag lineups were weakly diagnostic of innocence for the averaging (DR = 1.24, 45% guilty) and designated suspect method (DR = 1.32, 43% guilty). Identification of the profile only was diagnostic for the averaging (DR = 2.49, 71% guilty) and designated suspect method (DR = 4.47, 82% guilty). Identification of the profile and the bag was more diagnostic for the averaging method (DR = 3.00, 75% guilty) and the designated suspect method (DR = 8.00, 89% guilty) than just the identification of the profile. The pattern of frequencies generating these DRs differed from chance expectations for the averaging method, $chi^2(2, N = 648) = 30.18$, $p < .001$, Cramer's V = 0.22, and the designated suspect method, $chi^2(2, N = 648) = 57.29$, $p < .001$, Cramer's V = 0.30. Again, when excluding nonchoosers, the effect became non-significant for the averaging, $chi^2(1, N = 648) = 0.01$, $p = .920$, phi = 0.00, and designated of suspect method, $chi^2(1, N = 648) = 0.29$, $p = .590$, phi = 0.02.

Lineup rejections. Overall, portrait face lineups as well as their combination with body lineups were somewhat diagnostic of innocence. Contrary to expectations, the combination (DR = 2.52, 28% guilty) did not exceed the value of the portrait face lineups alone (DR = 2.45, 29% guilty). Performance for body, bag, and profile lineups as well as the combination of the latter two was poor (DRs ≤ 1.61).

Table 5 presents the proportion of witnesses who did not reject the suspect's face, rejected just the suspect's face (portrait or profile), or the suspect's face and the second aspect of the person (body or bag, respectively) in TA and TP lineups. Identifications in portrait face and body lineups were not diagnostic of innocence (DR = 0.58, 63% guilty). Rejections of the portrait only were diagnostic (DR = 2.45, 29% guilty). Rejections of the portrait and the body were not more diagnostic (DR = 2.52, 28% guilty) than just the

Table 5

Lineup Rejections for Portait Face plus Body Lineups and Profile Face plus Bag Lineups as a Function of Target Presence (N = 648)

	Frequencies			Proportions-diagnosticity		
Condition	None	1	2	None	1	2
	Portrait face and body lineups					
TP	252	49	23	.78	.15	.07
TA	146	120	58	.45	.37	.18
DR				0.58	2.45	2.52
% Guilty				63	29	28
	Profile face and bag lineups					
TP	259	23	42	.80	.07	.13
TA	221	37	66	.68	.11	.20
DR				0.85	1.61	1.57
% Guilty				54	38	39

rejection of the portrait. The pattern of frequencies generating these DRs differed from chance expectations, $chi^2(2, N = 648) = 73.18$, $p < .001$, Cramer's V = 0.34. However, when excluding choosers, the effect became non-significant, $chi^2(1, N = 648) = 0.01$, $p = .920$, phi = 0.00.

Identifications in profile face and bag lineups were not diagnostic of innocence (DR = 0.85, 54% guilty). Rejections of the profile (DR = 1.61, 38% guilty) as well as rejections of the profile and the bag (DR = 1.57, 39% guilty) were weakly diagnostic. Nevertheless, the pattern of frequencies generating these DRs differed from chance expectations, $chi^2(2, N = 648) = 11.61$, $p = .003$, Cramer's V = -0.13. Again, when excluding choosers, the effect became non-significant, $chi^2(1, N = 648) = 0.01$, $p = .920$, phi = 0.00.

Foil choices. The only lineup that seemed to be somewhat diagnostic for foil choices was the portrait face lineup (DR = 2.86). Performance in all other lineups was poor (DRs ≤ 1.77).

DISCUSSION

The aim of the present field experiment was to test the usefulness of multiple lineups with a very large sample of participants in a natural setting. Stimulus sampling was emphasized by using nine different targets. Besides reporting DRs for target/suspect choices we also analyzed DRs for lineup rejections and foil choices. This allows for a more complete view on the utility of multiple lineups, also in cases where no choice or foil choices are made. In previous studies, no (Lindsay et al., 1987) or limited information were reported for lineup rejections and foil choices (Pryke et al., 2004).

Compared to foil choices and lineup rejections, target/suspect choices were most diagnostic of guilt. The portrait face lineup alone and its combination with the body lineup were most diagnostic for target/suspect choices.

Lineup rejections showed some capability of establishing innocence, but the DRs obtained here were much lower than those obtained for target/suspect choices. Again, the portrait face lineup and its combination with the body lineup were most diagnostic of innocence. However, the combination of the two lineups did not increase diagnosticity. The diagnosticity of multiple foil choices was acceptable for portrait face lineups but limited for all other lineups or combinations with DRs ≤ 1.77.

Taken together, our results on target/suspect choices support the results of former studies with multiple lineups (Lindsay et al., 1987; Pryke et al., 2004). These studies reported higher diagnosticity of target/suspect choices for multiple lineups when compared to single lineup decisions, but only for a total of three targets.

We included a second face lineup, a profile lineup, because during the interaction witnesses may have seen the target primarily from the side, so that an identification from a profile face picture may be easier. This reasoning is supported by research that reported improved recognition when there was no or little angular rotation between encoding and recognition (Liu & Chaudhuri, 2002). Generally, performance in the profile face lineup was low (30.5%) and much lower than in the portrait face lineup (57.7%). Yet, 12.2% of the participants made an incorrect identification decision in the portrait face, but a correct decision in the profile face lineup. Furthermore, depending on the method of establishing the innocent suspect, the profile lineup by itself as well as its combination with the bag lineup showed high diagnosticity for target/suspect choices, sometimes exceeding those values obtained for the portrait face and its combination with the body lineup.

Particularly the use of a clothing lineup had proved useful in previous studies, not by itself but when combined with other lineups (Lindsay et al., 1987; Pryke et al., 2004, Experiment 2). In the present study, we used a shopping bag as clothing in a broader sense, presuming that such an item would be more likely to be left behind by a culprit than a piece of garment and would therefore be available for lineup construction. Similar to previous results on clothing lineups, the utility of the bag lineup was limited by itself, but its combination with the profile face lineup boosted diagnosticity for target/suspect choices. However, for lineup rejections and foil choices the combination of the profile face lineup with the bag lineup did not increase diagnosticity. One possible explanation for this finding is, that after the testing, quite a few participants spontaneously mentioned to the interviewer that they had not noticed the target carry a bag and therefore had rejected the lineup. Apparently, these participants had no memory for the bag. Research on the weapon focus (Steblay, 1992) has demonstrated that witnesses do pay a lot of attention to some objects (weapons) that a target carries, however, at the cost of lower face identification performance. This does not seem to be the case for less salient objects, such as a bag (cf. Pickel, 1998, 1999). This is also compatible to the social norm of keeping eye contact while talking to another person. Providing an "I don't know" option may be a way of eliminating witnesses who have no memory for an item and thus may reduce deflation of lineup rejection DRs as observed here (all bag DRs for lineup rejections here were smaller than 1).

The present results direct our attention to an important methodological issue in eyewitness identification research: the designation of the innocent suspect in TA lineups. As observed, the results can differ widely depending on the method that is selected to establish innocent foil choices. While the averaging method resulted in larger DRs in most cases, this was not always the case. For example, for multiple lineups, DRs for the profile face lineup and its combination with the bag lineup were larger for the designated suspect method than for the averaging method. We believe that both methods and their reasoning have their legitimization. Researchers must be cautious when deciding for one or the other method and we believe it most appropriate to report more than one measure (for methods of establishing innocent suspect choices other than the ones used here, see e.g., Brewer et al., 2006; Clare & Lewandowsky, 2004; Fleet et al., 1987; Olsson & Juslin, 1999; Sporer, 1993, 2007).

Performance in the four different lineup types was not associated with each other. This result supports the idea that multiple lineups can serve as independent sources of evidence. In accordance with this result, performance in blank lineups was not associated with performance in TP lineups, as would have been expected from Wells' study (1984). However, there were a number of methodological differences between the present and Wells' study (1984). While Wells (1984) used blank portrait face lineups to predict TP or TA portrait face lineups, our blank lineups were different lineup types compared to those that were predicted. Specifically, we used portrait face and body lineups to predict the outcome of bag and profile face lineups and vice versa. Possibly, the usefulness of blank lineups is limited to portrait face lineups or at least to the same lineup mode and viewing angle. For example, blank profile face lineups may only be predictive of profile face lineups but not of body lineups. Also, correlations may be higher if two lineup modes are either TA or TP, respectively. This could be tested in future studies.

Furthermore, in order to compute DRs, we used identical distractors in the blank and TP lineups, although they were presented on different lineup positions. This is different from Wells' (1984) study and may have mislead witnesses in a way that resulted in an impact on their choosing behavior: In lineups 2, 3, and 4, witnesses may have recognized some distractors presented before and thus may have assumed that all lineup members were identical, resulting in the same choosing behavior as in lineup 1. Supporting this notion, our data on choosing in the body and profile face lineups suggest that witnesses were not so much influenced by the presence of the target when they made a choice but by the fact of whether or not they had made a choice before.

Despite the strengths of the present study, namely the large sample, multiple targets and the natural setting, there are some caveats that need to be discussed. First, although ecological validity was enhanced by using a natural setting, there was no crime scenario in the present study. Thus, it is unlikely that participants experienced the same arousal level as witnesses of a real crime would. However, it is equally unlikely that participants in studies with filmed crimes do. Yet, even if the accuracy of identification decisions was influenced by the arousal level, this does not necessarily need to have an impact on DRs. However, future studies should investigate this issue.

A second caveat here is the short (30 s) interval between the event and the identification task. In real cases, there are normally days, if not months between witnessing the event and the identification (Behrman & Richards, 2005). It is likely, that the memory for the target would have decreased after a longer retention interval. Yet again, this need not influence the obtained DRs. This is another issue that could be investigated in future studies.

Altogether, as there is only a very limited number of studies on multiple lineup decisions, only little is known about conditions that have a positive or a negative impact on the DRs. Future research may dig deeper into this topic. It should also be explored further how many and which specific lineup types should be used. Asking witnesses from which perspective they saw the target could aid on deciding for a portrait or profile face lineup.

To conclude, what can be said about the application of this novel control variable, multiple lineups, in real cases? The data speak for the application of multiple lineups with regard to suspect/target choices as a procedure to avoid false identifications, whereas the benefit of multiple lineups for lineup rejections and foil choices seems rather limited. Yet, although the combination of lineups for target/suspect choices shows higher diagnosticity than looking at just a single lineup, there are also cases where participants correctly identify the target from a facial lineup but not from another lineup. Thus, multiple identifications should be viewed as stronger evidence than a single identification (cf. Pryke et al., 2004). However, single lineup evidence must not be discarded per sé.

References

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.

Brewer, N., Caon, A., Todd, C., & Weber, N. (2006). Eyewitness identification accuracy and response latency. Law and Human Behavior, 30, 31-50.

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.

Brigham, J. C. (1990). Target person distinctiveness and attractiveness as moderator variables in the confidence-accuracy relationship in eyewitness identifications. Basic and Applied Social Psychology, 11, 101-115.

Brigham, J. C., Van Verst, M., & Bothwell, R. K. (1986). Accuracy of children's eyewitness identifications in a field setting. Basic and Applied Social Psychology, 7, 295-306.

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

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

Cutler, B. L., & Fisher, R. P. (1990). Live lineups, videotaped lineups, and photoarrays. Forensic Reports, 3, 439-448.

Cutler, B. L., & Penrod, S. D. (1988). Improving the reliability of eyewitness identification: Lineup construction and presentation. Journal of Applied Psychology, 73, 281-290.

Cutler, B. L., & Penrod, S. D. (1995). Mistaken identity. Cambridge: Cambridge University Press.

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.

Davies, G., Ellis, H., & Shepherd, J. (1978). Face recognition accuracy as a function of mode of representation. Journal of Applied Psychology, 63, 180-187.

Egan, D., Pittner, M., & Goldstein, A. G. (1977). Eyewitness identification. Photographs vs. live models. Law and Human Behavior, 1, 199-206.

Fleet, M. L., Brigham, J. C., & Bothwell, R. K. (1987). The confidence-accuracy relationship: The effects of confidence assessment and choosing. Journal of Applied Social Psychology, 17, 171-187.

Juslin, P., Olsson, N., & Winman, A. (1996). Calibration and diagnosticity of confidence in eyewitness identification: Comments on what can be inferred from the low confidence-accuracy correlation. Journal of Experimental Psychology: Learning, Memory, and Cognition, 22, 1304-1316.

Kneller, W., Memon, A., & Stevenage, S. (2001). Simultaneous and sequential lineups: Decision processes of accurate and inaccurate eyewitnesses. Applied Cognitive Psychology, 15, 659-671.

Krafka, C., & Penrod, S. D. (1985). Reinstatement of context in a field experiment on eyewitness identification. Journal of Personality and Social Psychology, 49, 58-69.

Lindsay, D. S., Read, J. D. & Sharma, K. (1998). Accuracy and confidence in person identification: The relationship is strong when witnessing conditions vary widely. Psychological Science, 9, 215–218.

Lindsay, R. C. L., Wallbridge, H., & Drennan, D. (1987). Do the clothes make the man? An exploration of the effect of lineup attire on eyewitness identification accuracy. Canadian Journal of Behavioural Science, 19, 464-478.

Liu, C. H., & Chaudhuri, A. (2002). Reassessing the 3/4 view effect in face recognition. Cognition, 83, 31-48.

Melara, R. D., DeWitt-Rickards, T. S., & O´Brien, T. P. (1989). Enhancing lineup identification accuracy. Two codes are better than one. Journal of Applied Psychology, 74, 706-713.

Olsson, N., & Juslin, P. (1999). Can self-reported encoding strategy and recognition skill be diagnostic of performance in eyewitness identification? Journal of Applied Psychology, 84, 42-49.

Patterson, K. E., & Baddeley, A. D. (1977). When face recognition fails . Journal of Experimental Psychology: Human Learning and Memory, 3, 406-417.

Pickel, K. L. (1998). The influence of context on "weapon focus effect". Law and Human Behaviour, 23, 299-311.

Pickel, K. L. (1999). Unusualness and threat as possible causes of "weapon focus". Memory, 6, 277-295.

Pryke, S., Lindsay, R. C. L., Dysart, J. E., & Dupuis, P. (2004). Multiple independent identification decisions: A method of calibrating eyewitness identifications. Journal of Applied Psychology, 89, 73-84.

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.

Sporer, S. L. (1993). Eyewitness identification accuracy, confidence and decision times in simultaneous and sequential lineups. Journal of Applied Psychology, 78, 22-33.

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

Sporer, S. L. (2007). Person descriptions as retrieval cues: Do they really help? Psychology, Crime, & Law.

Steblay, N. M. (1992). A meta-analytic review of the weapon focus effect. Law and Human Behavior, 16, 413-424.

Tredoux, C. G. (1998). Statistical inference on measures of lineup fairness. Law and Human Behavior, 22, 217-237.

Tredoux, C. G. (1999). Statistical considerations when determining measures of lineup size and lineup bias. Applied Cognitive Psychology, 13, 9-26.

Tulving, E., & Thomson, D. M. (1973). Encoding specifity and retrieval processes in episodic memory. Psychological Review, 80, 352-373.

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

Wells, G. L. (1984). The psychology of lineup identifications. Journal of Applied Social Psychology, 14, 89-103.

Wells, G. L., & Lindsay, R. C. L. (1980). On estimating the diagnosticity of eyewitness nonidentifications. Psychological Bulletin, 88, 776-784.

Wells, G. L., & Olson, E. A. (2003). Eyewitness testimony. Annual Review of Psychology, 54, 277-295.

Wells, G. L., Rydell, S. M., & Seelau, E. P. (1993). The selection of distractors for eyewitness lineups. Journal of Applied Psychology, 78, 835-844.

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

Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling and social psychology experimentation. Personality and Social Psychology Bulletin, 25, 1115-1125.

Woodhead, M. M., Baddeley, A. D., & Simmonds, D. C. V. (1979). On training people to recognize faces. Ergonomics, 22, 333-343.

Yarmey, A. D. (2004). Eyewitness recall and photo identification: A field experiment. Psychology, Crime & Law, 10, 53-68.

Appendix

Table A1

Hits and False Alarms (%) for Nine Targets (ns = 72)

		Target								
		1	2	3	4	5	6	7	8	9
Frontal	Hits	41.7	80.6	75.0	38.9	63.9	72.2	33.3	66.7	72.2
face	False alarms (average)[a]	9.3	6.5	7.4	7.9	6.5	9.3	8.3	7.4	5.1
	False alarms (designated suspect)[b]	38.9	11.1	19.4	25.0	16.7	22.2	19.4	16.7	8.3
Body	Hits	5.6	13.9	11.1	41.7	13.9	22.2	2.8	19.4	38.9
	False alarms (average)[a]	11.6	10.6	13.0	12.5	10.2	14.8	13.4	12.5	12.0
	False alarms (designated suspect)[b]	0.0	2.8	5.6	11.1	5.6	2.8	0.0	8.3	2.8
Bag	Hits	11.1	2.8	8.3	2.8	5.6	22.2	13.9	19.4	16.7
	False alarms (average)[a]	3.7	4.6	6.5	9.3	8.8	5.6	8.3	7.9	6.9
	False alarms (designated suspect)[b]	2.8	8.3	5.6	19.4	8.3	2.8	5.6	8.3	2.8
Profile	Hits	2.8	27.8	16.7	19.4	47.2	58.3	16.7	36.1	33.3
	False alarms (average)[a]	12.0	10.6	11.1	12.5	11.6	11.1	11.1	11.1	11.1
	False alarms (designated suspect)[b]	2.8	8.3	5.6	19.4	8.3	2.8	5.6	8.3	2.8

[a]Innocent suspect identifications were computed as the mean of all positive identifications in TA lineups. [b]Foil chosen most often in the TA portrait face lineup was designated to be the innocent suspect.

Table A2

Diagnosticity Ratios (DRs) and Guilt Estimates for Lineup Rejections, Target/Suspect Choices and Foil Choices for Individual Lineup Decisions

		Target								
	Estimated	1	2	3	4	5	6	7	8	9
		Target/suspect choice								
Portrait face	DR (average)[a]	4.50	12.44	10.13	4.94	9.86	7.80	4.00	9.00	14.18
	% guilty	82	93	91	83	91	89	80	90	93
	DR (designated suspect)[b]	1.07	7.25	3.86	1.56	3.83	3.25	1.71	4.00	8.67
Body	DR (average)[a]	0.48	1.31	0.86	3.34	1.36	1.50	0.21	1.55	3.23
	% guilty	33	57	46	77	58	60	17	61	76
	DR (designated suspect)[b]	-	5.00	2.00	3.75	2.50	7.99	-	2.33	14.00
Bag	DR (average)[a]	3.00	0.60	1.28	0.30	0.64	4.00	1.67	2.46	2.40
	% guilty	75	38	56	23	39	80	63	71	71

Table A2 (continued)

		Target								
	Estimated	1	2	3	4	5	6	7	8	9
		Target/suspect choice								
	DR (designated suspect)[b]	4.00	0.34	1.49	0.14	0.67	7.99	2.50	2.33	6.01
Profile	DR (average)[a]	0.23	2.61	1.50	1.55	4.08	5.25	1.50	3.25	3.00
	% guilty	19	72	60	61	80	84	60	76	75
	DR (designated suspect)[b]	0.25	3.34	-	2.33	16.99	20.99	6.01	-	2.00

Table A2 (continued)

	Estimated	Target 1	2	3	4	5	6	7	8	9
						Lineup rejection				
Portrait face	DR	1.45	3.96	3.33	2.11	2.75	2.66	1.13	3.33	6.25
	% guilty	41	20	23	32	27	27	47	23	14
Body	DR	1.00	1.44	0.67	2.25	2.33	0.36	1.16	2.25	0.91
	% guilty	50	41	60	31	30	73	46	31	52
Bag	DR	0.97	1.04	0.96	0.76	0.85	1.09	1.00	0.91	0.95
	% guilty	51	49	51	57	54	48	50	52	51
Profile	DR	1.43	2.16	1.09	1.00	1.58	4.01	1.99	1.99	1.20
	% guilty	41	32	48	50	39	20	33	33	45

Table A2 (continued)

		Target								
	Estimated	1	2	3	4	5	6	7	8	9
		Foil choice								
Portrait face	DR	2.00	14.00	5.33	1.31	2.80	5.00	2.25	2.67	1.83
	% guilty	33	7	16	43	26	17	31	27	35
Body	DR	1.09	1.05	1.40	1.59	0.88	1.88	1.00	1.08	2.36
	% guilty	48	49	42	39	53	35	50	48	30
Bag	DR	2.67	1.00	1.40	1.43	1.36	2.00	1.38	2.13	1.88
	% guilty	27	50	42	41	42	33	42	32	35
Profile	DR	0.93	1.15	1.26	1.35	2.08	2.00	1.00	1.41	1.71
	% guilty	52	47	44	43	32	33	50	41	37

Note. Dashes indicate undefined values because the frequency in the TA condition only was zero for target/suspect choices. [a]Innocent suspect identifications were computed as the mean of all positive identifications in TA lineups. [b]Foil chosen most often in the TA portrait face lineup was designated to be the innocent suspect.

Table A3

Diagnosticity Ratios (DRs) of Lineup Selections in Combinations for Lineup Rejections, Target/Suspect Choices and Foil Choices

		Target								
	Estimated	1	2	3	4	5	6	7	8	9
		Target/suspect choice								
Portrait face	DR (average)[a]	4.75	12.00	12.00	5.45	10.91	6.00	4.24	8.31	9.33
	% guilty	83	92	92	85	92	86	81	89	90
	DR (designated suspect)[b]	1.07	8.00	4.80	2.00	5.00	2.71	1.71	6.00	7.00
Body[c]	DR (average)[a]	0.50	0	0.25	3.14	0.63	0.19	0.21	0.25	0.50
	% guilty	33	0	20	76	39	16	18	20	33
Portrait face and body	DR (average)[a]	0.00	15.00	4.50	4.00	6.00	42.00	0.00	12.00	36.00
	% guilty	0	94	82	80	86	98	0	92	97
	DR (designated suspect)[b]		5.04	1.44	0.99	1.44	6.84		2.04	11.88

Table A3 (continued)

		Target								
	Estimated	1	2	3	4	5	6	7	8	9
		Target/suspect choice								
Bag	DR (average)[a]	3.43	0.60	1.09	0.33	0.38	3.00	1.41	2.40	2.31
	% guilty	77	38	52	25	27	75	59	71	70
	DR (designated suspect)[b]	4.10	0.46	1.21	0.17	0.64	11.99	2.39	3.13	7.49
Profile	DR (average)[a]	0.24	2.61	1.43	1.68	4.36	4.91	1.30	3.27	3.00
	% guilty	19	72	59	63	81	83	57	77	75
	DR (designated suspect)[b]	0.28	3.43	-	2.40	16.96	23.15	5.80	-	2.20
Bag and Profile	DR (average)[a]	0		2.00	0.00	2.00	9.00	6.00	3.00	3.00
	% guilty	0		67	0	67	90	86	75	75
	DR (designated suspect)[b]			-		1.01	-	-	-	-

Table A3 (continued)

		Target								
	Estimated	1	2	3	4	5	6	7	8	9
		Lineup rejection								
Portrait face	DR	1.38	3.67	6.50	1.75	2.40	3.25	1.00	2.80	17.00
	% guilty	42	21	13	36	29	24	50	26	6
Body	DR	0.75	0.33	0.13	1.33	1.33	0.33	0.44	1.00	0.25
	% guilty	57	75	89	43	43	75	69	50	80
Portrait face and body	DR	1.67	3.67	1.75	5.00	3.33	1.50	1.00	6.00	2.67
	% guilty	38	21	36	17	23	40	50	14	27

Table A3 (continued)

		Target								
	Estimated	1	2	3	4	5	6	7	8	9
		Lineup rejection								
Bag	DR	0.91	0.80	0.73	0.69	0.69	0.80	0.87	0.71	0.82
	% guilty	52	56	58	59	59	56	54	59	55
Profile	DR	3.00	3.00	0.33	1.00	1.67	4.00	2.33	2.50	1.00
	% guilty	25	25	75	50	38	20	30	29	50
Bag and Profile	DR	1.17	2.00	1.38	1.00	1.50	4.00	1.67	1.75	1.40
	% guilty	46	33	42	50	40	20	38	36	42

Table A3 (continued)

		Target								
	Estimated	1	2	3	4	5	6	7	8	9
		Foil choice								
Portrait face	DR	2.11	-	6.00	1.10	2.20	4.75	2.83	13.00	1.50
	% guilty	32	0	14	48	31	17	26	7	40
Body	DR	1.09	1.00	1.26	1.50	0.76	1.82	1.04	1.20	2.18
	% guilty	48	50	44	40	57	35	49	45	31
Portrait face and body	DR	1.00	2.00	4.00	2.00	-	-	0.50	0.60	-
	% guilty	50	33	20	33	0	0	67	63	0

Table A3 (continued)

		Target								
	Estimated	1	2	3	4	5	6	7	8	9
		Foil choice								
Bag	DR	2.33	1.00	1.38	2.00	1.14	2.00	1.55	2.50	1.63
	% guilty	30	50	42	33	47	33	39	29	38
Profile	DR	0.89	1.15	1.24	1.67	1.83	2.00	1.05	1.47	1.57
	% guilty	53	47	45	38	35	33	49	41	39
Bag and Profile	DR	-		1.50	0.40	-	2.00	0.50	1.00	-
	% guilty	0		40	71	0	33	67	50	0

Note. Dashes indicate undefined values because the frequency in the TA condition only was zero for target/suspect choices or because the TP condition only was zero for lineup rejections or filler choices. Blank cells indicate that thte frequency in both the present and absent conditions was zero. [a]Innocent suspect identifications were computed as the mean of all positive identifications in TA lineups. [b]Foil chosen most often in the TA portrait face lineup was designated to be the innocent suspect. [c]No DR (designated suspect) because, by defintion, there are no accurate body identifications other than those that are also accurate portrait face identifications as the designated suspect is the person who was most frequently chosen from the portrait face lineup.

DISCUSSION

This dissertation reported 4 experiments that are concerned with the evaluation of eyewitness identification testimony. In order to increase ecological validity even in a laboratory setting, a 1-week interval was inserted between witnessing the crime and identification in Experiments 1 and 2. Experiments 3 and 4, which report different data of a large scale field study, emphasized stimulus sampling by including 10 targets (9 in Experiment 4). Experiment 1 challenged previous research findings regarding the relationship between identification performance and target description by not only testing describers and non-describers, but also a group of rereaders who reread their target description before the identification task. Experiments 2 and 3 suggested that investigators combine postdictors rather than looking at each of them individually when assessing identification decisions. Experiment 4 showed that multiple lineups can be beneficial as a procedure to avoid false identifications with regard to suspect/target choices, whereas the benefit of multiple lineups for lineup rejections and foil choices seems rather limited. In the following, the main results of each experiment will be discussed. For a more comprehensive discussion, see the discussion sections of each Experiment.

Verbal Overshadowing

Experiment 1 assessed three target description groups in order to test different theoretical accounts regarding the relationship between identification performance and target description: non-describers, describers, and rereaders (describers with rereading of the description before the identification task). No verbal overshadowing effect (VOE) as postulated by a change in processing style (TIPS, Schooler, Fiore, & Brandimonte, 1997; Schooler, 2002) or by an alteration of the original memory trace (RBI; Meissner, Brigham, and Kelley, 2001) was found. Also, our results yielded no support for a memory facilitation effect through context reinstatement by rereading of one's target description (Cutler, Penrod, & Martens, 1987; Cutler, Penrod, O'Rourke, & Martens, 1986). Instead, consistent with our hypothesis, our results support the recognition criterion shift approach as suggested by Clare and Lewandowsky (2004).

With regard to the duration of the criterion shift, the present results support the assumption that the recognition criterion shift only occurs if (a) the identification task immediately follows the description task, as assessed in the study by Clare and Lewandowsky (2004), or (b) the description is reactivated before the identification task, for example by rereading it, as it was the case in the present study.

Looking at retention intervals between description and identification, we found that there is only one study (Schooler & Engstler-Schooler, 1990, Experiment 5) who used a retention interval of 48 hours and found a VOE. Yet, most studies examining the VOE used only short post-description delays of less than 24 hours (cf., Meissner & Brigham, 2001). Therefore, there is a clear need of studies with longer post-description delays in order to address the questions of the durability and permanence that target descriptions can have on identification performance. After all, the probability that a lineup is carried out immediately after the description is very low as documented in archival analyses of real criminal cases (Behrman & Richards, 2005; Sporer, 1992a; Valentine, Pickering, & Darling, 2003; van Koppen & Lochun, 1997). Hence, the ecological validity of previously reported findings is arguable.

Postdicting Choosers' Identification Performance

In Experiments 2 and 3, post-decision confidence and decision times were shown to be useful postdictors of choosers' identification accuracy. This is in line with previous findings (e.g., Dunning & Perretta, 2002; Sporer, Penrod, Read, & Cutler, 1995; Sporer, 1992b, 1993, 1994; Weber, Brewer, Wells, Semmler, & Keast, 2004). A decision rule including highly confident and fast participants led to more correct classifications than either variable alone in both studies, suggesting that investigators should look at decision times and confidence in combination and not individually. Although we obtained a high number of correctly classified choosers within the established boundaries, it should be clear that our postdiction rate holds only for a small subgroup of participants, fast and confident choosers. The issue of how to proceed with slower and less confident choosers, however, remains unsolved.

Additional postdictors, such as self-reported decision processes (Experiment 2) and Remember-Know-Familiar (RKF) judgments (Experiment 3) did not lead to higher correct classification rates. Significant associations between postdictors (Experiment 2) highlight the need to consider the different processes jointly as, in combination, they may be more useful in assessing identification decisions.

Postdicting Nonchoosers' Identification Performance

Experiments 2 and 3 also dealt with nonchoosers' decision processes. Using a different procedure in both experiments, nonchoosers made statements about their decision processes. In either case, no associations between self-reported decision processes and identification accuracy were found. Thus, it seems that self-reported decision processes, like post-decision confidence and decision time, are not a valid postdictor of nonchoosers' identification performance.

An explanation for the lack of association between the postdictors and identification accuracy of nonchoosers may be that choosers' and nonchoosers' identification decisions possess asymmetric features (Weber & Brewer, 2004). Whereas choosers' decisions are likely to be based on a match between their memory for the target and one of several faces shown in a lineup, nonchoosers' decisions are based on a failure to match their memory to any of the faces in the lineup. Weber and Brewer (2006) argued that confidence ratings of nonchoosers are not based on confidence for the most likely match (face), because in this case, one would expect the CA relationship of nonchoosers to be similar to the one observed for choosers, which is not the case. Instead, the authors suggested that confidence of nonchoosers reflected the average match between all lineup faces and the target. This hypothesis could be tested by either asking nonchoosers about their confidence with regard to each lineup member (as done by Sporer, 1993, with sequential lineups) or by using one-person lineups (i.e., showups). Following Weber and Brewer's (2006) reasoning, the associations between postdictors and identification accuracy should be observed when confidence and decision time measures apply to only one lineup member. Unfortunately, many studies with showups either did not report the CA correlation at all, or, if they did, they did not differentiate between choosers and nonchoosers (e.g., Beal, Schmitt, & Dekle, 1995; Gonzales, Ellsworth, & Pembroke, 1993; Lindsay, Pozzulo, Craig, Lee, & Corbers, 1997; Yarmey, Yarmey, & Yarmey, 1996). Dysart, Lindsay, and Dupuis (2006) found a significant CA correlation for both choosers ($\underline{r}$ = .32) and nonchoosers ($\underline{r}$ = .19), supporting Weber and Brewer's (2006) hypothesis.

Assessing Identification Testimony via Multiple Lineup Decisions

Finally, Experiment 4 tested the usefulness of multiple lineup decisions with regard to different lineup types (portrait, body, bag, profile) for the assessment of identification testimony regarding nine different targets.

Utility of Blank Lineups to Predict Performance in Other Lineup Types

Performance in the four different lineup types was not associated with each other, supporting the idea that multiple lineups can serve as independent sources of evidence. Accordingly, performance in blank lineups was not associated with performance in TP lineups, as would have been expected from Wells' study (1984). However, there were a number of methodological differences between the present and Wells' study (1984). While Wells (1984) used blank portrait face lineups to predict TP or TA portrait face lineups, our blank lineups were different lineup types compared to those that were predicted. Specifically, we used portrait face and body lineups to predict the outcome of bag and profile face lineups and vice versa. Possibly, the usefulness of blank lineups is limited to portrait face lineups or at least to the same lineup mode and viewing angle. For example, blank profile face lineups may only be predictive of profile face lineups but not of body lineups. Also, correlations may be higher if two lineup modes are either TA or TP, respectively. This could be tested in future studies.

Furthermore, in order to compute DRs, we used identical distractors in the blank and TP lineups, although they were presented on different lineup positions. This is different from Wells' (1984) study and may have mislead witnesses in a way that resulted in an impact on their choosing behavior: In lineups 2, 3, and 4, witnesses may have recognized some distractors presented before and thus may have assumed that all lineup members were identical, resulting in the same choosing behavior as in the first lineup, the portrait lineup. Supporting this notion, our data on choosing in the body and profile face lineups suggest that witnesses were not so much influenced by the presence of the target when they made a choice but by the fact of whether or not they had made a choice before.

Diagnosticity of Multiple Lineups

For target/suspect choices, the portrait face lineup alone and its combination with the body lineup were most diagnostic. Lineup rejections showed some capability of establishing innocence, but the DRs obtained here were much lower than those obtained for target/suspect choices. Again, the portrait face lineup and its combination with the body lineup were most diagnostic of innocence, but the combination of the two lineups failed to increase diagnosticity. The diagnosticity of multiple foil choices was acceptable for portrait face lineups but limited for all other lineups or combinations. Taken together, our results on target/suspect choices support the results of former studies with multiple lineups (Lindsay, Wallbridge, & Drennan, 1987; Pryke, Lindsay, Dysart, & Dupuis, 2004).

The results indicate that multiple lineups can be applied as a procedure to avoid false identifications with regard to suspect/target choices, whereas the benefit of multiple lineups for lineup rejections and foil choices seems rather limited. Yet, although the combination of lineups for target/suspect choices shows higher diagnosticity than looking at just a single lineup, there are also cases where participants correctly identify the target from a facial lineup but not from another lineup. Thus, multiple identifications should be viewed as stronger evidence than a single identification (cf. Pryke et al., 2004). However, single lineup evidence must not be discarded per se.

To conclude, the present studies suggest that both decision times and post-decision confidence should be collected at the time of identification and be combined in order to assess identification accuracy. Investigators need to be aware though, that there is no postdictive value of nonchooser's estimates. Furthermore, these data ought to be collected at the time of identification in order to prevent distortion by feedback or media (see Semmler, Brewer, & Wells, 2004; Wells & Bradfield, 1998; Wells, Olson, & Charman, 2003). The finding that confidence ratings that are made in the court room should not serve as evidence entails the need for some changes in police work and the judiciary. From the results of empirical research we can derive, that questions about confidence regarding the identification decision should not be admitted at court. Instead, a video of the identification or at least a written protocol which includes the decision time and post-decision confidence of the witness should be provided. There seem to be no negative effects of target descriptions on identification accuracy when there is a sufficient interval between description and identification, as there is in real cases. Finally, the data speak for the application of multiple lineups with regard to suspect/target choices as a procedure to avoid false identifications. Before multiple lineups are applied in real cases, however, further studies are necessary, in order to specify how many and which particular lineup types should be used.

In real cases, the results for the assessment variables (decision times, confidence, decision processes) and control variables (target description, multiple lineups) examined in the present dissertation may vary from those that we obtain in laboratory or field studies. Reasons could be, for example, awareness of the severe consequences of false identifications and false rejections or the stress level at encoding or recognition. It would be interesting to collect data on these issues in real cases so they can be compared to the data obtained in laboratory/field studies. Undoubtedly, it would be a great contribution to the field of identification if data were collected even where DNA samples exist, so that identification accuracy could actually be assessed in real cases.

References

Beal, C. R., Schmitt, K. L., & Dekle, D. J. (1995). Eyewitness identification of children: Effects of absolute judgments, nonverbal response options, and event encoding. Law and Human Behavior, 19, 197-216.

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.

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

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.

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 Behaviour, 1, 113-134.

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.

Dysart, J. E., Lindsay, R. C. L., & Dupuis, P. R. (2006). Show-ups: The critical issue of clothing bias. Applied Cognitive Psychology, 20, 1009-1023.

Gonzalez, R., Ellsworth, P. C., & Pembroke, M. (1993). Response biases in lineups and showups. Journal of Personality and Social Psychology, 64, 525-537.

Lindsay, R. C. L., Pozzulo, J. D., Craig, W., Lee, K., & Corber, S. (1997). Simultaneous line-ups, sequential lineups, and showups: Eyewitness identification decisions of adults and children. Law and Human Behavior, 21, 391-404.

Lindsay, R. C. L., Wallbridge, H., & Drennan, D. (1987). Do the clothes make the man? An exploration of the effect of lineup attire on eyewitness identification accuracy. Canadian Journal of Behavioural Science, 19, 464-478.

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

Meissner, C. A., Brigham, J. C., & Kelley, C. M. (2001). The influence of retrieval processes in verbal overshadowing. Memory and Cognition, 29, 176-186.

Pryke, S., Lindsay, R. C. L., Dysart, J. E., & Dupuis, P. (2004). Multiple independent identification decisions: A method of calibrating eyewitness identifications. Journal of Applied Psychology, 89, 73-84.

Schooler, J. W. (2002). Verbalization produces a transfer inappropriate processing shift. Applied Cognitive Psychology, 16, 989-997.

Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. Cognitive Psychology, 22, 36-71.

Schooler, J. W., Fiore, S. M., & Brandimonte, M. A. (1997). At a loss from words: Verbal overshadowing of perceptual memories. In D. L. Medin (Ed.), The psychology of leaming and motivation (pp. 293-334). Academic Press: San Diego, CA.

Semmler, C., Brewer, N., & Wells, G. L. (2004). Effects of postidentification feedback on eyewitness identification and nonidentification confidence. Journal of Applied Psychology, 89, 1-13.

Sporer, S. L. (1992a, March). An archival analysis of person descriptions. Paper presented at the biennial meeting of the American Psychology-Law Society in San Diego, CA.

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

Sporer, S. L. (1993). Eyewitness identification accuracy, confidence and decision times in simultaneous and sequential lineups. Journal of Applied Psychology, 78, 22-33.

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). New York: Cambridge University Press.

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

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

van Koppen, P. J., & Lochun, S. K. (1997). Portraying perpetrators: The validity of offender descriptions by witnesses. Law and Human Behavior, 21, 661-685.

Weber, N., & Brewer, N. (2004). Confidence-accuracy calibration in absolute and relative face recognition judgments. Journal of Experimental Psychology: Applied, 10, 156-172.

Weber, N., & Brewer, N. (2006). Positive versus negative face recognition decisions: Confidence, accuracy, and response latency. Applied Cognitive Psychology, 20, 17-31.

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.

Wells, G. L. (1984). The psychology of lineup identifications. Journal of Applied Social Psychology, 14, 89-103.

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.

Wells, G. L., Olson, E. A., & Charman, S. D. (2003). Distorted retrospective eyewitness reports as functions of feedback and delay. Journal of Experimental Psychology: Applied, 9, 42-52.

Yarmey, A. D., Yarmey, M. J., & Yarmey, A. L. (1996). Accuracy of eyewitness identifications in showups and lineups. Law and Human Behavior, 20, 459-477.

DEUTSCHE ZUSAMMENFASSUNG

Die Bedeutung von Personenidentifizierungen durch Augenzeugen für die Ermittlungsarbeit und die strafrechtliche Verfolgung von Straftaten ist nach wie vor unbestritten, auch wenn inzwischen häufig auf DNA-Analysen zurückgegriffen werden kann. Die Forschung auf diesem Gebiet beschäftigt sich seit über 30 Jahren mit Faktoren, die einen positiven oder negativen Einfluss auf die Identifizierungsleistung haben können. Diese werden Schätz- und Systemvariablen genannt (Wells, 1978). Schätzvariablen können unterteilt werden in Situationsvariablen, die nur post hoc untersucht werden können (z.B. Lichtverhältnisse), und Beurteilungsvariablen (Sporer, 1993), die angewendet werden, um den Entscheidungsprozess von Zeugen zu evaluieren. System- oder Kontrollvariablen hingegen sind solche Variablen, die vom Rechtssystem normalerweise kontrolliert werden können. Sie beinhalten Testfaktoren, wie z.B. die Instruktionen für den Zeugen und die Zusammensetzung der Gegenüberstellung. Eine vielfach untersuchte Systemvariable ist die Beschreibung der Zielperson durch einen Augenzeugen. In der vorliegenden Dissertation wurden einerseits System-/Kontrollvariablen (Experimente 1 und 4) sowie Beurteilungsvariablen untersucht (Experimente 2 und 3).

Experiment 1: Personenbeschreibungen und Personenidentifizierung: Verbale Überlagerung oder Verschiebung des Wiedererkennenskriteriums?

Jeder Identifizierungsaufgabe geht normalerweise eine Personenbeschreibung durch einen Augenzeugen voraus. Obwohl dieses Vorgehen offenkundig zu sein scheint, gibt es einige Untersuchungen, die dieses Vorgehen in Frage stellen. So wurde gezeigt, dass der Personenbeschreibungsprozess negative Auswirkungen in Form eines verbalen Überlagerungseffekts auf die Identifizierungsleistung haben kann (Meissner & Brigham, 2001), d.h. dass Probanden, die die Zielperson beschrieben, schlechtere Identifizierungsleistungen erbrachten als solche, die dies nicht taten.

Es gibt drei theoretische Ansätze, um dieses Phänomen zu erklären: Der Ansatz des transfer-inappropriate processing shift (TIPS; Schooler, 2002) geht davon aus, dass die verbalen Prozesse, die während einer Personenbeschreibung ablaufen, die darauf folgenden non-verbalen Prozesse, die für das Wiedererkennen eines Gesichtes notwendig sind, hemmen. Es wird aber keine Veränderung der ursprünglichen Gedächtnisspur angenommen. Der Ansatz der retrieval-based interference (RBI; Meissner, Brigham, & Kelley, 2001) postuliert im Gegensatz dazu, dass der verbale Überlagerungseffekt aus einer Veränderung der Gedächtnisspur resultiert. Der Ansatz von Clare und Lewandowsky (2004) hingegen konzentrierte sich auf mentale Prozesse. Demnach überwachen Zeugen ihre Fähigkeit, eine Person zu beschreiben, genauso, wie sie dies bei anderen Gedächtnisaufgaben tun (Brigham & Pressley, 1988; Koriat & Goldsmith, 1996; Schraw, 1998). Aufgrund der relativen Unerfahrenheit hinsichtlich des Beschreibens von Personen ist es wahrscheinlich, dass diese Aufgabe als schwierig empfunden wird, und dass es keine Vorerfahrungen gibt, mit denen die Zeugen ihre eigene Beschreibung vergleichen können (Clare & Lewandowsky, 2004). Aufgrunddessen nehmen Clare und Lewandowsky (2004) an, dass sich Zeugen hinsichtlich der Qualität ihrer Beschreibung unsicher sind, was dazu führen kann, dass sie die Lichtbildvorlage eher zurückweisen, also Nichtwähler werden. Es wird also eine Verschiebung des Entscheidungskriteriums in die konservative Richtung erwartet.

Anders als von den Ansätzen zur verbalen Überlagerung angenommen, ist es aber auch denkbar, dass Verbalisierung durch Wiederherstellung des Wahrnehmungskontextes positive Effekte auf das Wiedererkennen hat (z.B. Cutler, Penrod, & Martens, 1987; Cutler, Penrod, O'Rourke, & Martens, 1986; Davids, 2006; Sporer, 2007). Die Forschungsergebnisse hierzu sind jedoch uneinheitlich. Möglicherweise beeinträchtigen andere Effekte, wie der verbale Überlagerungseffekt und die Verschiebung des Entscheidungskriteriums, die positiven Effekte der Wiederherstellung des Wahrnehmungskontextes.

Experiment 1 zielte darauf ab, die verschiedenen Erklärungsansätze gegeneinander zu testen. Es wurden drei Gruppen getestet: keine Personenbeschreibung (Kontrollgruppe), nur

Personenbeschreibung und Personenbeschreibung mit nochmaligem Lesen der eigenen Beschreibung (Wiederlesen). Das Lesen der eigenen Beschreibung unmittelbar vor der Identifizierungsaufgabe sollte dazu dienen, die Probanden wieder in den verbalen Modus zu versetzen. Wenn dies zu ungeeigneter Verarbeitung führt, sollte der verbale Überlagerungseffekt in dieser Gruppe besonders deutlich zu Tage treten. Um die ökologische Validität der Untersuchung zu erhöhen, lag zwischen der Beschreibung und der Identifizierung ein Intervall von einer Woche (s. Behrman & Richards, 2005; Valentine, Pickering, & Darling, 2003). Um die Hypothese der Kriteriumsverschiebung (Clare & Lewandowsky, 2004) untersuchen zu können, erhielten 50% der Probanden eine Lichtbildvorlage mit (target-present; TP), 50% eine Lichtbildvorlage ohne Zielperson (target-absent; TA). Da Standardinstruktionen verwendet wurden (vgl. Meissner & Brigham, 2001), wurde kein verbaler Überlagerungseffekt, verursacht durch ungeeignete Verarbeitung (TIPS) oder eine Veränderung der Gedächtnisspur, (RBI) erwartet. Stattdessen wurde eine Kriteriumsverschiebung (Clare & Lewandowsky, 2004) postuliert. Konkret wurde für die Beschreibungsgruppen eine konservative Kriteriumsverschiebung angenommen, die im Vergleich zur Kontrollgruppe zu einer höheren Zurückweisungsrate führen sollte. Da es bisher keine Befunde hinsichtlich der Dauer dieser Kriteriumsverschiebung gibt, waren zwei Resultate denkbar: (1) Wenn die Kriteriumsverschiebung dauerhaft ist, sollten alle Beschreiber die Lichtbildvorlage gleichhäufig zurückweisen; (2) wenn die Kriteriumsverschiebung aber vorübergehend ist, dann sollte die Gruppe, die nur beschreibt (ohne Wiederlesen), ein ähnliches Wahlverhalten zeigen wie die Kontrollgruppe. Wiederleser sollten hingegen die Lichtbildvorlage häufiger zurückweisen als die anderen beiden Gruppen. Dies würde für TA Lichtbildvorlagen zu einer höheren Rate korrekter Entscheidungen (richtige Zurückweisung), für TP Lichtbildvorlagen zu einer geringeren Rate korrekter Entscheidungen (Treffer) führen.

Ausgehend vom Prinzip der Wiederherstellung des Wahrnehmungskontextes wurde erwartet, dass Wiederleser eine bessere Identifizierungsleistung aufweisen als die anderen beiden Gruppen.

Die Ergebnisse zeigten keinen Einfluss der Beschreibung oder des Wiederlesens auf die Identifizierungsleistung, wohl aber auf das Wahlverhalten: Wiederleser wiesen die

Gegenüberstellung signifikant häufiger zurück als die Kontrollgruppe und Beschreiber ohne Wiederlesen. Diese Ergebnisse sprechen für eine Kriteriumsverschiebung bei den Wiederlesern (Clare & Lewandowsky, 2004). Hinsichtlich der Dauer der Kriteriumsverschiebung ist aufgrund der vorliegenden Ergebnisse anzunehmen, dass diese entweder nur dann auftritt, wenn (1) die Identifizierungsaufgabe unmittelbar auf die Beschreibungsaufgabe folgt (vgl. Clare & Lewandowsky, 2004), oder (2) die Beschreibung vor der Identifizierungsaufgabe reaktiviert wird, beispielsweise durch Wiederlesen.

Das Ausbleiben eines verbalen Überlagerungseffektes wurde bereits in anderen Untersuchungen, bei denen zwischen Beschreibung und Identifizierung ein Zeitintervall von mehr als 24 Stunden lag, berichtet (z.B. Clifford, 2003; Memon & Rose, 2002; Yu & Geiselman, 1993). Dies deutet darauf hin, dass es einen Bedarf an Untersuchungen mit längeren (ökologisch valideren) post-Beschreibungsintervallen gibt, um den Fragen nach der Dauer und Umkehrbarkeit der Auswirkungen von Personenbeschreibungen auf die Identifizierungsleistung weiter nachzugehen.

Experiment 2: Subjektive Sicherheit, Entscheidungszeit und selbstberichtete Entscheidungsprozesse als Beurteilungsvariablen für die Richtigkeit von Identifizierungsaussagen

Beurteilungsvariablen sind Variablen, die dazu dienen, eine Identifizierungsaussage nachträglich hinsichtlich ihrer Richtigkeit zu bewerten. Die am häufigsten verwendeten Beurteilungsvariablen sind subjektive Sicherheit nach der Identifizierung und Entscheidungszeit. Bei der Untersuchung dieser Variablen ist die Unterscheidung zwischen Personen, die die Lichtbildvorlage zurückweisen (Nichtwähler) und solchen, die eine Person aus der Lichtbildvorlage auswählen (Wähler) von großer Bedeutung. In der Literatur zeigte sich für Wähler ein recht robuster positiver Zusammenhang zwischen subjektiver Sicherheit und Identifizierungsleistung (z.B., Sporer, Penrod, Read, & Cutler, 1995) sowie ein negativer Zusammenhang zwischen Entscheidungszeit und Identifizierungsleistung (z.B., Sporer, 1992a, 1993, 1994; Weber, Brewer, Wells, Semmler, & Keast, 2004). Keine solchen Zusammenhänge wurden für Nichtwähler gefunden. Die Ergebnisse zur Entscheidungszeit werden nun vermehrt im Zusammenhang mit Befunden zum Entscheidungsprozess von Augenzeugen betrachtet: Wähler, die von automatischem Wiedererkennen sprachen, trafen deutlich häufiger eine korrekte (und schnelle) Entscheidung als solche, die von eliminativen Prozessen während des Wiedererkennens berichteten (Caputo & Dunning, 2005; Dunning & Stern, 1994; Kneller, Memon, & Stevenage, 2001).

Obwohl Entscheidungszeit und subjektive Sicherheit nach der Identifizierung in der Vergangenheit bereits erfolgreich kombiniert wurden (Sporer, 1992a; Weber et al., 2004), befasst sich der überwiegende Anteil der Forschung mit nur einer Beurteilungsvariablen. Ziel von Experiment 2 war es zu überprüfen, ob die Kombination von drei Beurteilungsvariablen dazu führt, dass mehr Identifizierungsentscheidungen korrekt klassifiziert werden, als wenn nur eine Beurteilungsvariable herangezogen wird. Bei den untersuchten Beurteilungsvariablen handelte es sich um die subjektive Sicherheit, Entscheidungszeit und selbstberichtete Entscheidungsprozesse.

Weiterhin befasste sich Experiment 2, das als Laboruntersuchung durchgeführt wurde, mit Entscheidungsprozessen von Nichtwählern. Vorangehende Untersuchungen zeigten relativ einheitlich, dass Zusammenhänge der Identifizierungsleistung mit der subjektiven Sicherheit nach der Identifizierung (z.B. Sporer et al., 1995; Weber & Brewer, 2003) und der Entscheidungszeit (z.B. Dunning & Stern, 1994; Sporer, 1992a, 1993; Weber et al., 2004, Weber & Brewer, 2006) für negative Identifizierungsentscheidungen (Nichtwähler) nicht gelten. Bisher wurde keine Beurteilungsvariable gefunden, die ein valides Urteil hinsichtlich der Richtigkeit von Identifizierungsentscheidungen von Nichtwählern erlaubt. Gleichwohl ist die Beurteilung der von Nichtwählern gemachten Identifizierungsaussagen zur Entlastung unschuldiger Verdächtiger von erheblicher Bedeutung. Daher wurden in der vorliegenden Untersuchung speziell für Nichtwähler Fragen zum Entscheidungsprozess konzipiert, um deren Brauchbarkeit als Beurteilungsvariable zu überprüfen.

Wie erwartet konnten durch die Kombination von subjektiver Sicherheit und Entscheidungszeit mehr Wähler korrekt klassifiziert werden als bei Berücksichtigung von nur einer Beurteilungsvariablen. Hinsichtlich der selbstberichteten Entscheidungsprozesse ergaben sich weder für Wähler noch für Nichtwähler Zusammenhänge mit der Identifizierungsleistung. Weiterhin wurde das Ergebnis repliziert, dass es für Nichtwähler keine Zusammenhänge zwischen subjektiver Sicherheit, Entscheidungszeit und Identifizierungsleistung gibt.

Die Entscheidungsprozesse von Wählern und Nichtwählern wurden in Experiment 3 weiter untersucht.

Experiment 3: Schnell und sicher: Die Nachhersage von Identifizierungsaussagen in einer Felduntersuchung

Experiment 3 hatte, wie Experiment 2, die Zielsetzung, Beurteilungsvariablen zur Evaluation von Identifizierungsentscheidungen zu kombinieren. Weiterhin wurden die Entscheidungsprozesse von Wählern und Nichtwählern weiter untersucht. Im Vergleich zu Experiment 2 wurden die folgenden Änderungen vorgenommen. (1) Zur Erhöhung der ökologischen Validität wurde Experiment 3 als Feldexperiment durchgeführt. (2) Im Sinne der internen Replikation der Ergebnisse wurde mit 10 anstatt mit nur einer Zielperson gearbeitet (s. Wells & Windschitl, 1999). (3) Ursprünglich sollte in Experiment 2 noch eine weitere Beurteilungsvariable, ein Urteil hinsichtlich des Bewusstseinszustandes und der Existenz einer Erinnerungserfahrung (Erinnern-Wissen-Vertraut sein Urteil), einbezogen werden. Die Operationalisierung schlug jedoch fehl und wurde daher für Experiment 3 geändert. (4) Selbstberichtete Erinnerungsprozesse wurden nicht erhoben.

Hinsichtlich des Erinnern-Wissen-Vertraut sein Urteils argumentierten Rotello, Macmillan, und Reeder (2004), dass Personen, die ein Erinnern-Urteil fällen (Erinnerung an bestimmte Merkmale der Person), über eine größere spezifische Gedächtnisstärke verfügen. Im Vergleich dazu gehen Wissen-Urteile eher mit einer globalen (unspezifischen) Gedächtnisstärke einher. Ausgehend von diesen Annahmen würde erwartet, dass Wähler, die ein Erinnern-Urteil abgeben, häufiger eine richtige Entscheidung treffen als solche, die ein Wissen-Urteil abgeben.

Ein anderer Ansatz zur Betrachtung von Entscheidungsprozessen wird von Dunning und Stern (1994) vertreten. Sie unterscheiden zwischen automatischen und deliberaten Entscheidungsprozessen. Dunning und Stern (1994) postulieren, dass Zeugen, die von automatischen Entscheidungsprozessen berichten, häufiger korrekte Identifizierungsentscheidungen treffen als Zeugen, die eher deliberate Entscheidungen treffen. Setzt man diesen Ansatz mit dem Erinnern-Wissen Paradigma in Bezug, dann würden wir erwarten, dass Wissen Antworten mit automatischen Entscheidungsprozessen und höherer Richtigkeit in Zusammenhang stehen als die deliberaten Erinnern Antworten. Dunning und Sterns (1994) Ansatz fand einige Unterstützung (Caputo & Dunning, 2005; Dunning & Stern, 1994; Kneller et al., 2001). In einer anderen Untersuchung gelang es jedoch nicht, anhand von selbstberichteten Entscheidungsprozessen richtige von falschen Entscheidungen zu unterscheiden (Brewer, Gordon, & Bond, 2000).

Es gibt also rivalisierende Hypothesen bezüglich der Überlegenheit von Erinnern- und Wissens-Urteilen in Bezug auf die Klassifikation von Identifizierungsaussagen und die Datenlage ist uneinheitlich. Bislang gibt es nur wenige Untersuchungen zu diesem Thema. Experiment 3 sollte unter anderem dazu dienen, zusätzliche Daten diesbezüglich zu liefern. Es wurde erwartet, dass sowohl Personen, die Erinnern, als auch solche, die Wissen angeben, im Durchschnitt bessere Identifizierungsleistungen erbringen würden als Personen, die Vertraut sein angeben.

Weiterhin beschäftigte sich Experiment 3 mit der Beurteilung von Identifizierungsaussagen, die von Nichtwählern gemacht wurden, bediente sich aber einer anderen Vorgehensweise als Experiment 2. Sporer et al. (1995) argumentierten, dass die Gruppe der Nichtwähler möglicherweise eine große Heterogenität aufweist. Daher wurden in der vorliegenden Untersuchung die Nichtwähler anhand der von ihnen gemachten Aussagen in drei Gruppen aufgeteilt: (1) Nichtwähler, die davon überzeugt waren, dass die Zielperson in der Lichtbildvorlage nicht anwesend war ("abwesend"), (2) Nichtwähler, die sich zu unsicher waren, um eine Wahl zu treffen und ("unsicher"), und (3) Nichtwähler, die keine Erinnerung hatten und daher keine (positive) Entscheidung treffen wollten ("keine Erinnerung"). Es wurde erwartet, dass in der abwesend Gruppe mehr korrekte Entscheidungen getroffen werden als in den beiden anderen Gruppen. Weiterhin wurde erwartet, dass die abwesend Gruppe ihre Entscheidungen schneller, sowie mit höherer subjektiver Sicherheit vor und nach der Identifizierungsentscheidung trifft. Für die unsichere und die keine Erinnerung Gruppe wurde erwartet, dass erstere die geringste Identifizierungsleistung erbringt, längere Entscheidungszeiten braucht und niedrigere subjektive Sicherheiten angibt.

Wie in Experiment 2 wurden subjektive Sicherheit und Entscheidungszeit für Wähler erfolgreich miteinander kombiniert. Hierdurch wurde eine sehr hohe Anzahl der Entscheidungen korrekt klassifiziert. Zwar zeigte sich ein Zusammenhang zwischen dem Erinnern-Wissen-Vertraut sein Urteil und der Identifizierungsleistung, jedoch trug dies nicht zu einer höheren Anzahl korrekter Klassifizierungen bei, als es bereits durch die beiden anderen Beurteilungsvariablen der Fall war.

Es wurde erwartet, dass Nichtwähler der abwesend Gruppe in ihrer Identifizierungsentscheidung häufiger richtig lagen als Nichtwähler der anderen beiden Gruppen. Es zeigte sich ein nicht-signifikanter Trend in diese Richtung. Die abwesend Gruppe war aber wie erwartet hinsichtlich ihrer Entscheidungen schneller und sicherer. Zwischen den anderen beiden Gruppen ergab sich nur hinsichtlich der subjektiven Sicherheit vor der Identifizierung ein signifikanter Unterschied. Allerdings war die unsichere Gruppe entgegen den Erwartungen vor der Identifizierung signifikant sicherer als die keine Erinnerung Gruppe.

Neben mangelnder Homogenität innerhalb der Gruppe der Nichtwähler gibt es noch eine weitere mögliche Erklärung für die nichtgefundenen Zusammenhänge zwischen Identifizierungsleistung und Beurteilungsvariablen bei Nichtwählern. So verlaufen die Entscheidungsprozesse von Wählern und Nichtwählern vermutlich asymmetrisch (Weber & Brewer, 2004). Während die Entscheidungen von Wählern auf einer Passung zwischen dem Gedächtnis für die Zielperson und einem Gesicht in der Lichtbildvorlage basiert, basiert die Entscheidung von Nichtwählern darauf, dass eben solch eine Passung nicht festgestellt werden kann. Weber und Brewer (2006) argumentierten, dass die Angaben zur subjektiven Sicherheit von Nichtwählern wohl nicht auf der subjektiven Sicherheit hinsichtlich des Gesichtes mit der besten Passung basieren. In diesem Fall würde man nämlich, ähnlich wie bei Wählern, einen

Zusammenhang zwischen subjektiver Sicherheit und Identifizierungsleistung erwarten. Dies ist aber nicht der Fall. Weber und Brewer (2006) schlossen daher, dass die Angaben zur subjektiven Sicherheit von Nichtwählern die durchschnittliche Passung für alle Personen in der Lichtbildvorlage widerspiegelt. Diese Hypothese könnte entweder durch Abfragen der subjektiven Sicherheit für jedes einzelne Gesicht in einer Lichtbildvorlage überprüft werden (Sporer, 1993) oder durch den Gebrauch von Lichtbildvorlagen mit nur einer Person (showups). Folgt man Weber und Brewers (2006) Argumentation, dann sollten sich dann Zusammenhänge zwischen Beurteilungsvariablen und Identifizierungsleistung ergeben.

Experiment 4: Die Anwendung multipler Gegenüberstellungen in einer Felduntersuchung

Nach wie vor bleibt die Frage, wie man mit Nichtwählern und weniger sicheren oder langsameren Wählern umgehen soll, unbeantwortet. Eine neue Methode zur Beurteilung von Identifizierungsaussagen wurde von Lindsay, Wallbridge und Drennan (1987) eingeführt und von Pryke, Lindsay, Dysart und Dupuis (2004) weiterentwickelt. In diesen Untersuchungen nahmen die Versuchspersonen nicht nur eine Identifizierung aus einer Portraitvorlage vor, sondern unabhängig davon auch aus einer Bekleidungsvorlage (Lindsay et al., 1987) oder Körper- und Stimmvorlage (Pryke et al., 2004). Auf diese Art und Weise konnten von einem Zeugen/einer Zeugin mehrere voneinander unabhängige Aussagen gewonnen werden und die Ratewahrscheinlichkeit deutlich verringert werden. Die bisherigen Ergebnisse (Lindsay et al., 1987; Pryke et al., 2004) sprechen dafür, dass die wiederholte Identifizierung einer Person aus verschiedenen Lichtbildvorlagen wahrscheinlicher ist, wenn es sich hierbei tatsächlich um die Zielperson handelt und nicht um einen unschuldigen Verdächtigen. Mit solchen multiplen Lichtbildvorlagen beschäftigte sich Experiment 4.

Der beschriebene Effekt wurde bisher nur für drei Zielpersonen gezeigt (Lindsay et al., 1987; Pryke et al., 2004). Um eine bessere Generalisierbarkeit der Ergebnisse zu gewährleisten, wurden in der vorliegenden Untersuchung neun Zielpersonen einbezogen. Den Probanden wurden Portrait-, Körper-, Accessoire- (Einkaufstasche) und Profilvorlagen vorgelegt. Wenn multiple Lichtbildvorlagen tatsächlich unabhängig voneinander sind, so würden wir keine signifikante Korrelation hinsichtlich der Leistung in den verschiedenen Lichtbildvorlagen erwarten. Dies steht jedoch im Gegensatz zu dem Befund, dass leere Lichtbildvorlagen (blank lineups; Lichtbildvorlagen die bekanntermaßen TA sind) einen Vorhersagewert für die Leistung in folgenden Gegenüberstellungen haben (Wells, 1984). Dies setzt nämlich eben eine solche Korrelation zwischen den Leistungen in den einzelnen Lichtbildvorlagen voraus. Meines Wissens ist Wells' (1984) Studie die einzige, die sich bisher mit diesem Thema auseinandergesetzt hat, allerdings nur für die Sequenz von TA Portraitvorlagen, gefolgt von TA oder TP Portraitvorlagen. Um die Nützlichkeit leerer Lichtbildvorlagen für unterschiedliche Arten von Gegenüberstellungen (Portrait, Körper, Tasche, Profil) zu überprüfen, wurden jeder Versuchsperson je zwei TA (Portrait und Körper oder Profil und Tasche) und TP Vorlagen vorgelegt.

Wenn leere Lichtbildvorlagen einen Vorhersagewert hinsichtlich der Leistung in anderen Arten von Gegenüberstellungen haben, dann sollten Versuchspersonen, die die leeren (TA) Gegenüberstellungen korrekt zurückweisen, in TP Gegenüberstellungen bessere Leistungen erbringen als solche, die hierzu nicht in der Lage sind. Wenn die Leistung in den unterschiedlichen Gegenüberstellungsarten jedoch unabhängig voneinander ist, dann sollten die leeren Gegenüberstellungen keinen prädiktiven Wert hinsichtlich der Leistung in TP Lichtbildvorlagen haben. Weiterhin wurde erwartet, dass zwei Identifizierungsentscheidungen (Portrait und Körper/Profil und Tasche) über eine höhere Diagnostizität verfügen als nur die Identifizierungsentscheidung hinsichtlich des Gesichts (Portrait oder Profil).

Es zeigte sich, dass die Wahl der Zielperson bzw. deren Austauschperson (in der Praxis ein unschuldig Verdächtigter) im Vergleich zu Zurückweisungen und zur Wahl eines Distraktors (bekanntermaßen unschuldig) die höchste Diagnostizität erreichte. Die Portraitvorlage und deren Kombination mit der Körpervorlage erreichten die höchste Diagnostizität für die Wahl der Zielperson/Austauschperson. Zurückweisungen sprachen zu einem gewissen Maße für Unschuld, allerdings waren die ermittelten Diagnostizitäten weitaus geringer als die für die Wahl der Zielperson/deren Austauschperson. Auch hier erreichten die Portraitvorlage und deren Kombination mit der Körpervorlage die höchste Diagnostizität, jedoch führte die Kombination nicht zu einer Erhöhung derselben. Die Wahl eines Distraktors war akzeptabel für Portraitvorlagen, aber sehr begrenzt für alle anderen Vorlagen oder deren Kombinationen. Insgesamt unterstützen die vorliegenden Ergebnisse die Befunde vorangehender Untersuchungen (Lindsay et al., 1987; Pryke et al., 2004).

Die Identifizierungsleistungen in den vier unterschiedlichen Lichtbildvorlagen standen nicht miteinander in Zusammenhang. Dies spricht dafür, dass multiple Lichtbildvorlagen als voneinander unabhängige Beweisstücke verwendet werden können. Demgemäß wurden in der vorliegenden Untersuchung im Gegensatz zu Wells (1984) keine Anhaltspunkte für die Nützlichkeit von leeren Lichtbildvorlagen gefunden. Möglicherweise ist die Nützlichkeit leerer Lichtbildvorlagen auf Portraits oder Lichtbildvorlagen desselben Körperteils beschränkt. So könnten leere Profilvorlagen prädiktiven Wert für Profilvorlagen haben, nicht jedoch für Körpervorlagen.

Zusammenfassend kann aufgrund der vorliegenden Daten geschlossen werden, dass sowohl Entscheidungszeit als auch subjektive Sicherheit bei der Identifizierung erhoben und zur Beurteilung der Identifizierungsaussage miteinander kombiniert werden sollten. Ermittlungsbeamte sollten sich allerdings dessen bewusst sein, dass für Nichtwähler keine Aussage aufgrund dieser Variablen getroffen werden kann. Zudem sollten diese Daten unmittelbar während/nach der Identifizierung erhoben werden, um eine Beeinflussung durch Feedback oder Medien zu verhindern (s. Semmler, Brewer, & Wells, 2004; Wells & Bradfield, 1998; Wells, Olson, & Charman, 2003). Die Erkenntnis, dass die Beurteilung der subjektiven Sicherheit im Gerichtssaal aus psychologischer Sicht keinerlei Beweiswert hat, bringt die Notwendigkeit von Veränderungen der Poliziearbeit und des Rechtswesens mit sich. Aus den Ergebnissen der empirischen Forschung lässt sich ableiten, dass vor Gericht keine Fragen zur subjektiven Sicherheit zugelassen werden sollten. Stattdessen sollte möglichst ein Video von der Identifizierung gemacht werden, mindestens jedoch ein Protokoll erstellt werden, das die subjektive Sicherheit und Entscheidungszeit des Zeugen zum Zeitpunkt der Identifizierung festhält. Einen negativen Effekt von Personenbeschreibungen auf die Identifizierungsleistung scheint es nicht zu geben, wenn zwischen Beschreibung und Identifizierung ausreichend Zeit liegt. Dies ist in der Realität der Fall. Die Daten sprechen weiterhin für die Anwendung multipler Lichtbildvorlagen bezüglich Wahlen der Zielperson/des Verdächtigen, um falsche Identifizierungen zu vermeiden. Hinsichtlich Distraktor-Wahlen und Zurückweisungen hingegen sind multiple Lichtbildvorlagen nach derzeitigen Erkenntnissen weniger brauchbar. Bevor die Anwendung multipler Lichtbildvorlagen jedoch in die Praxis umgesetzt wird, sind weitere Studien erforderlich, die untersuchen, wie viele und welche spezifischen Lichtbildvorlagen zu verwenden sind.

In realen Fällen könnten die Ergebnisse bezüglich der hier untersuchten Beurteilungsvariablen (subjektive Sicherheit, Entscheidungszeit, Entscheidungsprozesse) und Kontrollvariablen (Personenbeschreibung, multiple Lichtbildvorlagen) von jenen, die wir in Labor- oder Feldstudien erhalten, abweichen. Gründe hierfür könnten beispielsweise die Konsequenzen der Entscheidung oder Stress bei der Enkodierung/dem Wiedererkennen sein. Es wäre interessant, Daten in realen Fällen zu erheben, so dass diese mit Feld-/Laboruntersuchungen verglichen werden könnten. Zweifelsohne wäre es ein großer Beitrag zur Identifizierungsforschung, wenn Daten auch dann erhoben werden würden, wenn DNA Spuren vorliegen, so dass die Identifizierungsleistung in realen Fällen überprüft werden könnte.

References

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.

Brewer, N., Gordon, M., & Bond, N. (2000). Effect of photoarray exposure duration on eyewitness identification accuracy and processing strategy. Psychology, Crime and Law, 6, 21-32.

Brigham, J. C., & Pressley, M. (1988). Cognitive monitoring and strategy choice in younger and older adults. Psychology and Aging, 3, 249-257.

Caputo, D. D., & Dunning, D. (2005). Eyewitness accuracy and decision processes using sequential lineups. Manuscript submitted for publication.

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

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.

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 Behaviour, 1, 113-134.

Davids, M. S. C. (2006). Knowing what you know: Meta-memory and the effects of verbal overshadowing on eyewitness identification testimony. Unpublished master's thesis, University of Giessen, Germany.

Dunning, D. &. Stern, L. B. (1994). Distinguishing accurate from inaccurate eyewitness identification via inquiries about decision processes. Journal of Personality and Social Psychology, 67, 818-835.

Kneller, W., Memon, A., & Stevenage, S. (2001). Simultaneous and sequential lineups: Decision processes of accurate and inaccurate eyewitnesses. Applied Cognitive Psychology, 15, 659-671.

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. Psychological Review, 103, 490-517.

Lindsay, R. C. L., Wallbridge, H., & Drennan, D. (1987). Do the clothes make the man? An exploration of the effect of lineup attire on eyewitness identification accuracy. Canadian Journal of Behavioural Science, 19, 464-478.

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

Meissner, C. A., Brigham, J. C., & Kelley, C. M. (2001). The influence of retrieval processes in verbal overshadowing. Memory and Cognition, 29, 176-186.

Pryke, S., Lindsay, R. C. L., Dysart, J. E., & Dupuis, P. (2004). Multiple independent identification decisions: A method of calibrating eyewitness identifications. Journal of Applied Psychology, 89, 73-84.

Rotello, C. M., Macmillan, N. A., & Reeder, J. A. (2004). Sum-difference theory of remembering and knowing: A two-dimensional signal-detection model. Psychological Review, 111, 588-616.

Schooler, J. W. (2002). Verbalization produces a transfer inappropriate processing shift. Applied Cognitive Psychology, 16, 989-997.

Schraw, G. (1998). Promoting general metacognitive awareness. Instructional Science, 26, 113-125.

Semmler, C., Brewer, N., & Wells, G. L. (2004). Effects of postidentification feedback on eyewitness identification and nonidentification confidence. Journal of Applied Psychology, 89, 1-13.

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

Sporer, S. L. (1993). Eyewitness identification accuracy, confidence and decision times in simultaneous and sequential lineups. Journal of Applied Psychology, 78, 22-33.

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). New York: Cambridge University Press.

Sporer, S. L. (2007). Person descriptions as retrieval cues: Do they really help? Psychology, Crime, & Law.

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

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

Weber, N., & Brewer, N. (2003). The effect of judgment type and confidence scale on confidence-accuracy calibration in face recognition. Journal of Applied Psychology, 88, 490-499.

Weber, N., & Brewer, N. (2004). Confidence-accuracy calibration in absolute and relative face recognition judgments. Journal of Experimental Psychology: Applied, 10, 156-172.

Weber, N., & Brewer, N. (2006). Positive versus negative face recognition decisions: Confidence, accuracy, and response latency. Applied Cognitive Psychology, 20, 17-31.

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.

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

Wells, G. L. (1984). The psychology of lineup identifications. Journal of Applied Social Psychology, 14, 89-103.

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.

Wells, G. L., Olson, E. A., & Charman, S. D. (2003). Distorted retrospective eyewitness reports as functions of feedback and delay. Journal of Experimental Psychology: Applied, 9, 42-52.

Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling and social psychology experimentation. Personality and Social Psychology Bulletin, 25, 1115-1125.

ERKLÄRUNG

Ich erkläre: Ich habe die vorgelegte Dissertation selbständig und nur mit den Hilfen angefertigt, die ich in der Dissertation angegeben habe. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten oder nicht veröffentlichten Schriften entnommen sind, und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht.

Gießen, im März 2007

____

Melanie Sauerland