

# Fast and confident: postdicting eyewitness identification accuracy in a field study

Citation for published version (APA):

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

## Document status and date:

Published: 01/01/2009

## DOI:

[10.1037/a0014560](https://doi.org/10.1037/a0014560)

## Document Version:

Publisher's PDF, also known as Version of record

## Document license:

Taverne

## Please check the document version of this publication:

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

[Link to publication](#)

## General rights

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

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

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

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

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

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

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

Melanie Sauerland and Siegfried L. Sporer  
University of Giessen

The combined postdictive value of postdecision confidence, decision time, and Remember-Know-Familiar (RKF) judgments as markers of identification accuracy was evaluated with 10 targets and 720 participants. In a pedestrian area, passers-by were asked for directions. Identifications were made from target-absent or target-present lineups. Fast (optimum time boundary at 6 seconds) and confident (optimum confidence boundary at 90%) witnesses were highly accurate, slow and nonconfident witnesses highly inaccurate. Although this combination of postdictors was clearly superior to using either postdictor by itself these combinations refer only to a subsample of choosers. Know answers were associated with higher identification performance than Familiar answers, with no difference between Remember and Know answers. The results of participants' post hoc decision time estimates paralleled those with measured decision times. To explore decision strategies of nonchoosers, three subgroups were formed according to their reasons given for rejecting the lineup. Nonchoosers indicating that the target had simply been absent made faster and more confident decisions than nonchoosers stating lack of confidence or lack of memory. There were no significant differences with regard to identification performance across nonchooser groups.

*Keywords:* postdecision confidence, decision time, remember-know distinction, eyewitness identification, multiple targets

The importance of eyewitness identifications for investigating and prosecuting crimes is undoubted. However, cases of miscarriages 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 is on assessment variables that may be used to retroactively assess individual witnesses' decision making accuracy. The most widely used assessment variables are postdecision *confidence* (e.g., Sporer, Penrod, Read, & Cutler, 1995) and *decision time* (e.g., Sporer, 1992, 1993, 1994; Weber, Brewer, Wells, Semmler, & Keast, 2004). Other approaches that explore different types of decision processes via self-reports have also been explored lately (Dunning & Stern, 1994; Kneller, Memon, & Stevenage, 2001; Sauerland & Sporer, 2007).

It is the first aim of the present research to examine the combined utility of three assessment variables for postdicting identification accuracy using live interactions in a field setting. Specifically, we study the relationship between response times 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 study 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., Conway, Gardiner, Perfect, Anderson, & Cohen, 1997; Gardiner, Ramponi, & Richardson-Klavehn, 2002; Tulving, 1985).

Postdecision confidence and decision times each by themselves have been researched extensively in the past, but have only been combined in a few studies (Brewer & Wells, 2006; Sauerland & Sporer, 2007; Sporer, 1992; Weber et al., 2004). Furthermore, combinations with other postdictors are rare. 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.

---

Melanie Sauerland and Siegfried L. Sporer, Department of Psychology, University of Giessen, Germany.

We thank Elisabeth Baumgartner, Verena Bohn, Claudia Dill, Lisa German, Sarah Jasmin Günther, Markus Jonitz, Morten Kaletsch, Alana Krix, Isabell Sauerbier, Alexandra Schmoranzler, Patricia Waschk, and Jürgen Weber for their help with data collection. We are also grateful to Maike C. S. Davids for preparing the lineups with PhotoShop. This paper is part of Melanie Sauerland's doctoral dissertation conducted at Giessen University. The research was supported by a grant from the Deutsche Forschungsgemeinschaft (German Science Foundation) to the second author.

Correspondence concerning this article should be addressed to Melanie Sauerland, Department of Psychology, University of Giessen, Otto-Behagel-Str. 10F, 35394 Giessen, Germany. E-mail: melanie.sauerland@psychol.uni-giessen.de

Furthermore, the present study addresses one shortcoming that affects most identification studies. Although the need for 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). These studies found that the results differed between targets 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 a single target could be misleading. Therefore, we included 10 targets in the present study.

Data were collected from passers-by in the pedestrian zone of a small university town. With targets asking for directions, participants were not aware that they would later be asked to make an identification (i.e., an incidental learning situation without the involvement of strong emotions). Thus, the study simulates cases where the fact that a crime is taking place is not apparent to potential witnesses. As the stimulus event was a live event, not a film, the variation in stimuli does not only entail differences in appearance between targets to be recognized. Rather, the variation in stimuli represents a complex construct of different stimulus persons and encoding conditions with regard to interaction times with participants, angle of view, background context, and so forth.

### Decision Time-Accuracy Relationship

The negative relationships between response times 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; Kneller et al., 2001; Sauerland & Sporer, 2007; Sporer, 1992, 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 (Dunning & Perretta, 2002; Smith, Lindsay, & Pryke, 2000). When splitting the time to decision into three groups (1–15 seconds, 16–30 seconds, >30 seconds), Smith et al. (2000) found that eyewitnesses making their choice within the first 15 seconds 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 seconds 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 seconds and termed this cut-off point optimum time boundary. Witnesses who made their identification within 10 seconds 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, Dunning and Perretta (2002) argued that correct identifications tended to be made via an automatic cognitive process whereas incorrect identification decisions were made more deliberately via a process of elimination (see also Dunning & Stern, 1994). Deliberate decisions that use an eliminative decision strategy are considered to be conscious and effortful. In contrast, 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 to vary between 5 seconds to 29 seconds across four studies using six targets. Their accuracy rates were considerably lower (67%) than those found by Dunning and Perretta (2002). When confidence was taken into account in addition to response time, much higher classification rates were obtained within the 10 second 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 nominal lineup sizes lead to faster decisions and earlier optimum time boundaries than longer retention intervals and larger lineup sizes. None of the identified time boundaries were within the 10 to 12 second interval postulated by Dunning and Perretta (2002).

In the present study, we aimed at providing further evidence regarding the invariance of the "10–12 second rule" postulated by Dunning and Perretta. To do so, we analyzed empirical time boundaries for 10 different targets who interacted with a representative sample of interviewees in short interactions of various lengths at different locations. 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, police do not (yet) routinely collect those. As a possible resort in cases where the actual decision times are not available, we asked our participants to give an estimate of their decision time. To test the usefulness of estimated decision times as a substitute of measured decision times, we conducted analyses on estimated decision times that were parallel to those conducted with measured decision times. We expected a positive correlation between the two measures, and a negative relationship with identification accuracy for choosers but not for nonchoosers. Furthermore, we expected both measures to elicit similar result patterns with regard to their postdictive value.

### Confidence-Accuracy Relationship

Meta-analyses on the relationship between postdecision confidence and identification accuracy reported small to medium point-biserial correlations of  $r = .25$  (Bothwell, Deffenbacher, & Brigham, 1987) and  $r = .28$  (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 (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, because of a lack of variation in confidence or accuracy. Furthermore, 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 et al., 2002; Weber & Brewer, 2003, 2006).

The present study examined the CA relationship, 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, 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. In a typical RK study, first participants typically are exposed to a list of stimuli (often words). At test, participants indicate whether a particular stimulus had been presented before. Upon positive responses, a Remember or Know judgment follows. Remembering corresponds to conscious recollection with regard to the moment when the stimulus was presented. Knowing refers to recognition of the word in the absence of conscious recollection.

Later, Guess and Familiar options were added. Guess responses (Rajaram, 1996) were supposed to be given when participants had some other reason than Remembering or Knowing to assume that a stimulus was encountered in the study list. Familiar responses (e.g., Conway et al., 1997; Wright & Sladden, 2003) are assumed to lie between Know and Guess responses, indicating that one alternative seems more familiar than any other, in the absence of Remembering or Knowing (Conway et al., 1997). To test this, half of our participants

making a positive identification were randomly assigned to an answer form that contained a Familiar option in addition to the Remember and Know options. A Guess option was not included, as this option would undermine the representativeness of the experimental task to courtroom situations. Courts certainly would have reservations to convict a person based on a witness who justified his or her lineup decision as a good "guess."

Since Tulving's original proposal, other models have been developed in an effort to account for RK data, including dual-process theories of recognition memory (Yonelinas, 1994) and one-dimensional signal detection models (Donaldson, 1996; Hirshman & Master, 1997). Some of these models assume RK judgments to be equivalent to confidence statements (Donaldson, 1996), others found differences between them (e.g., Mäntylä, 1997; see Dunn, 2004, for a discussion). 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. Translating their ideas to eyewitness identification, choosers who give Remember responses decision should be more accurate than those giving Know responses.

However, when we contrast this view with the automatic and deliberate decision processes postulated by Dunning and Stern (1994), we would expect witnesses reporting automatic decision processes (i.e., "just knowing") to make more correct, and faster decisions than those making deliberate decisions.

To summarize, rival hypotheses exist about the superiority of Remember and Know judgments for postdicting identification accuracy, with existing (mostly unpublished) data being inconclusive (see Table 1). We expected both, Remember and Know responders to show better identification performance than participants giving a Familiar response.

### Nonchoosers' Decision Processes

As noted above, previous research has consistently shown that associations of identification accuracy with postdecision confidence and decision times 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 or she was not confident enough to choose this person. Third, the witness may have had no memory for the event.

In the present study, we asked nonchoosers to indicate why they rejected the lineup after their decision. The options included that (a) the target had not been among them ("absent"), (b) he or she thought the target had been among them, but that he or she had not been confident enough to choose him or her ("low confidence"), or (c) he or 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

**Table 1**  
*Summary of Postdictor Variables, Analyses Conducted, and Outcomes Predicted (and Obtained) for Choosers and Nonchoosers*

Variables	Analysis	Choosers	Nonchoosers
Predecision confidence	Correlation with identification accuracy	Zero or small positive <sup>†</sup>	Zero or small positive <sup>†</sup>
Decision times	Correlation with identification accuracy	Negative <sup>†</sup>	Zero <sup>†</sup>
	Correlation with estimated decision times	Positive <sup>†</sup>	Positive <sup>†</sup>
	Correlation with confidence	Negative <sup>†</sup>	Negative <sup>†</sup>
Estimated decision times	Correlation with identification accuracy	Negative <sup>†</sup>	Zero <sup>†</sup>
Postdecision confidence	Correlation with identification accuracy	Positive <sup>†</sup>	Zero <sup>†</sup>
	Calibration	Good <sup>†</sup>	Poor <sup>†</sup>
RKF judgments	Mean accuracy	(a) R > K (Rotello et al., 2004)	NA <sup>a</sup>
		(b) K > R (Dunning & Stern, 1994)	
		(c) (R, K) > F <sup>†</sup>	
Combination of postdictors	Classification of identification accuracy	Combination > individual postdictors <sup>†</sup>	NA <sup>a</sup>
Nonchoosers' decision processes	Mean identification accuracy	NA <sup>a</sup>	Absent > <sup>†</sup> No memory > Low confidence
	Mean post-decision confidence	NA <sup>a</sup>	Absent > <sup>†</sup> No memory > Low confidence
	Mean decision times	NA <sup>a</sup>	Absent < <sup>†</sup> No memory < Low confidence

<sup>a</sup> NA: Not applicable.  
<sup>†</sup> Prediction supported by our data.

“absent” group to show higher identification accuracy than those in the two other groups. Furthermore, we expected the “low confidence” group 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 and a preview of the results obtained.

**Method**

*Participants*

There were 944 persons (479 female, 465 male) that 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. Of the Participants 27.2% were students with various majors. 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 × 2 design, target person (Targets 1 through 10) and target presence (present vs. absent) were independent variables, with an equal number of cases, and an equal number of males and

females, in each cell. Identification accuracy was the dependent variable, with pre- and postdecision confidence, decision times, and estimated decision times 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 × 13 cm photographs mounted on a 30 × 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, 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 (95% CI = 5.10–5.17), and 6.76 (95% CI = 6.75–6.77; 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, alternating roles after data of eight participants were collected. Every interviewer collected data of eight 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 seconds and 60 seconds. 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 to the location asked for.

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 predecision 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, TP and TA 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 that 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 or her manner of speaking, the situation, or what I was thinking when he or 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 option was omitted. Hereafter, choosers indicated their postdecision 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 postdecision 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. Estimates were made in seconds 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 3 months.

### Results

An *alpha* level of .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  $\chi$  are reported for nonparametric analyses of  $3 \times 2$  and  $2 \times 2$  contingency tables, respectively.

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,  $\chi^2(1, N = 720) = .82, p = .408, \Phi = .03$ .

In the following, we first look at the relationships between identification accuracy and each postdictor individually. Descriptive statistics of decision time measures and decision confidence measures as well as their intercorrelations can be found in Table 2 for choosers and nonchoosers. The results generally replicate the well known finding that the associations with identification accuracy are significant for choosers but not for nonchoosers. Nevertheless, the intercorrelations of the postdictors with each other were comparable when looking at choosers and nonchoosers.

Table 2 also contains the correlations when only choices of the replacements were counted as false identifications. The correlations showed a very similar pattern to those when all false choices were included. Only the association between identification accuracy and estimated decision-time became nonsignificant,  $r(240) = -.11, p = .104$ .

For each postdictor, a  $2 \times 2$  ANOVA with choice (choosers vs. nonchoosers) and decision outcome (correct vs. incorrect) as independent variables was computed and the results are re-

Table 2

Means, SDs, and Intercorrelations of Identification Accuracy and Postdictors of Identification Accuracy for Choosers ( $N = 436$ ) and Nonchoosers ( $N = 284$ )

	<i>M</i>	<i>SD</i>	95% CI	Predecision confidence	Postdecision confidence	Decision time	Estimated decision time
<b>Choosers</b>							
Identification accuracy	48.4%	50.0	43.7–53.1%	.16**	.39**	-.30**	-.21**
Identification accuracy <sup>a</sup>	87.2%	33.5	82.9–91.4%	.13*	.34**	-.17**	-.11
Predecision confidence	60.6%	23.6	58.4–62.8%		.46**	-.13**	-.16**
Postdecision confidence	60.6%	27.1	58.1–63.1%			-.33**	-.31**
Decision time <sup>b</sup>	9.4 seconds	2.5	8.6–10.2 seconds				.49**
Estimated decision time <sup>b</sup>	9.3 seconds	3.0	8.4–10.3 seconds				
<b>Nonchoosers</b>							
Identification accuracy	70.1%	45.9	64.7–75.4%	.16**	.09	-.05	.01
Predecision confidence	62.0%	23.1	59.3–64.7%		.45**	-.12*	-.13*
Postdecision confidence	62.4%	26.1	59.3–65.4%			-.23**	-.19**
Decision time <sup>b</sup>	11.0 seconds	2.3	10.0–12.2 seconds				.46**
Estimated decision time <sup>b</sup>	10.2 seconds	3.0	9.0–11.6 seconds				

<sup>a</sup> Measures with only choices of the replacement treated as false identifications ( $n = 242$ ). <sup>b</sup> The means and standard deviations were backtransformed from the logarithmic values used for the inferential analyses.

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

ported in the following section. Subsequently, the postdictors are combined.<sup>1</sup>

*Postdicting Identification Accuracy From Assessment Variables*

*Predecision confidence.* For predecision confidence there only was a main effect of decision outcome,  $F(1, 716) = 17.23, p < .001, d = 0.31$ . Participants making a correct identification decision showed higher predecision 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%).

*Decision times.* Because of 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). Means are reported for back-transformed values.

There were significant main effects of decision outcome,  $F(1, 716) = 20.50, p < .001, d = -0.34$ , and choice,  $F(1, 716) = 6.96, p = .009, d = -0.20$ . Specifically, accurate participants decided faster ( $M = 8.6$  seconds, 95% CI = 7.9–9.5 seconds) than inaccurate ones ( $M = 12.1$  seconds, 95% CI = 11.1–13.3 seconds), and choosers decided faster ( $M = 9.4$  seconds, 95% CI = 8.6–10.2 seconds) than nonchoosers ( $M = 11.0$  seconds, 95% CI = 10.0–12.2 seconds). The main effects were modified by the expected significant interaction between Decision Outcome and Choice,  $F(1, 716) = 11.24, p = .001, f = .10$ . There was a simple main effect of decision outcome within choosers,  $F(1, 716) = 43.90, p < .001, d = -0.50$ , but not within nonchoosers,  $F(1, 716) = 0.53, p = .465, d = -0.05$ . Within choosers, accurate decisions were made faster ( $M = 7.3$  seconds; 95% CI = 6.3–7.9 seconds) than inaccurate ones ( $M = 12.4$  seconds; 95% CI = 10.9–13.8 seconds). Within nonchoosers, there was no difference between accurate ( $M = 11.0$  seconds; 95% CI = 9.5–12.1 seconds) and inaccurate ( $M = 12.8$  seconds; 95% CI = 9.7–14.1 seconds) decisions with regard to decision time.

*Postdecision confidence.* There was a significant main effect for decision outcome,  $F(1, 716) = 40.94, p < .001, d = 0.48$ . 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,  $F(1, 716) = 15.48, p < .001, f = .15$ . There was a significant simple main effect of accuracy within choosers,  $F(1, 716) = 75.49, p < .001, d = 0.65$ , but not within nonchoosers,  $F(1, 716) = 2.35, p = .126, d = 0.12$ . 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 postdecision confidence.

*Estimated decision times.* Because of 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 times 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

<sup>1</sup> For each postdictor, the analyses were also conducted with target as an additional factor. The only interaction that became significant was the one between Decision Outcome and Target for decision time,  $F(9, 680) = 2.30, p = .015, f = .17$ . The predicted Choice by Accuracy interaction, however, was not qualified by an interaction with target. Similar conclusions were reached when sequential logistic regression analyses (see Tabachnick & Fidell, 2007) were run with the data of choosers: Although entering target as an additional predictor did significantly improve the model (just as the ANOVA interaction with decision outcome had shown), higher order interactions with target did not further improve the models. For ease of comparison with past research (e.g., Weber & Brewer, 2004), ANOVA results are reported as in those studies.

means did not differ significantly from one another for choosers ( $M_s = 9.3$  seconds and  $9.4$  seconds),  $t(435) = -.164$ ,  $p = .870$ ,  $d = -0.02$ , or nonchoosers ( $M_s = 10.3$  seconds and  $11.0$  seconds),  $t(283) = -1.16$ ,  $p = .246$ ,  $d = -0.11$ . 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,  $F(1, 716) = 6.55$ ,  $p = .011$ ,  $d = -0.19$ . Correct decisions were estimated to have been made faster ( $M = 9.0$  seconds, 95% CI =  $7.8$ – $9.7$  seconds) than incorrect ones ( $M = 11.1$  seconds, 95% CI =  $10.0$ – $12.6$  seconds). This main effect was modified by an interaction between Decision Outcome and Choice,  $F(1, 716) = 7.30$ ,  $p = .007$ ,  $f = .10$ . There was a significant simple main effect of accuracy within choosers. Correct choosers estimated their decision times as shorter ( $M = 7.7$  seconds; 95% CI =  $6.6$ – $8.9$  seconds) than did incorrect choosers ( $M = 11.7$  seconds; 95% CI =  $10.1$ – $13.5$  seconds),  $F(1, 716) = 19.57$ ,  $p < .001$ ,  $d = -0.33$ . No such difference was found between correct ( $M = 10.6$  seconds; 95% CI =  $9.1$ – $12.3$  seconds) and incorrect nonchoosers ( $M = 10.6$  seconds; 95% CI =  $8.3$ – $13.5$  seconds),  $F(1, 716) = 0.01$ ,  $p = .930$ ,  $d = 0.01$ .

### Calibration of Confidence Measures

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

**Predecision confidence.** The predecision 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 predecision 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 cate-

gory. 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, & 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).

**Postdecision confidence.** As expected, the postdecision confidence-accuracy (CA) correlation was significant for choosers,  $r(434) = .39$ ,  $p < .001$ , but nonsignificant 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,  $NRI = 0.174$ .

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  $\chi^2$ -analyses. In the RKF condition, there was a significant effect of RKF judgment on identification accuracy,  $\chi^2(2, N = 204) = 8.16$ ,  $p = .017$ ,  $Cramer's V = .20$ . Post hoc comparisons of the three groups with a Bonferroni-adjusted alpha of  $.05/3 = .017$  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 =  $21.6$ – $47.5\%$ ),  $\chi^2(1, N = 134) = 8.08$ ,  $p = .004$ ,  $\phi = .26$ .

In the RK condition, there was no significant effect,  $\chi^2(1, N = 232) = 3.35$ ,  $p = .087$ ,  $\phi = -.12$ . Accuracy with Remember judgments ( $M = 42.4\%$ ; 95% CI =  $33.3$ – $51.4\%$ ) tended to be somewhat lower than with Know judgments ( $M = 54.4\%$ ; 95% CI =  $45.1$ – $63.7\%$ ).

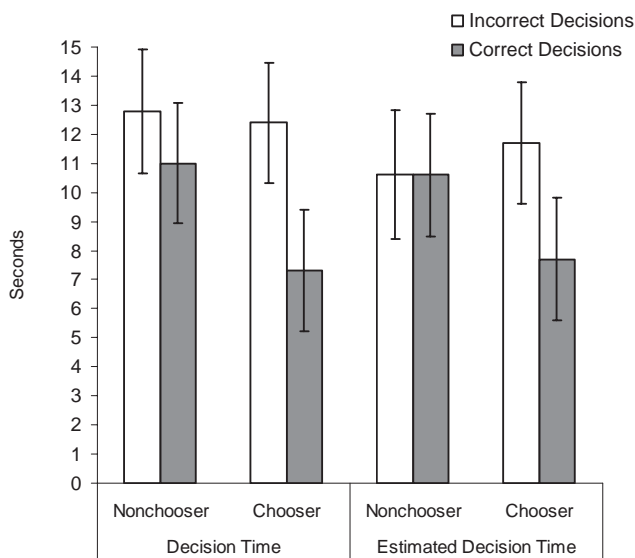


Figure 1. Means (and confidence intervals) of actual versus estimated decision times as a function of choice and decision outcome.



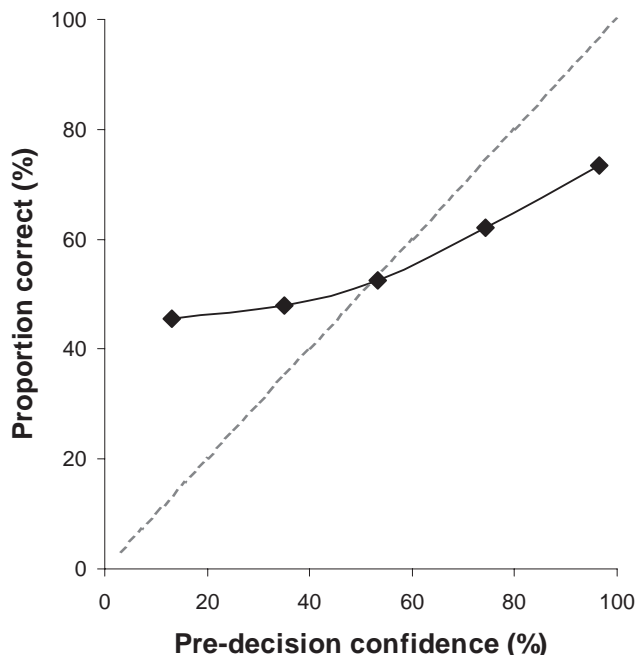


Figure 2. Predecision confidence-identification accuracy calibration curves for choosers and nonchoosers combined.

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” to evaluate the identification evidence as accurate. Therefore, decision time boundaries (Dunning & Perretta, 2002) were computed based on the 2 (accuracy: correct vs. incorrect) × 2 (time boundary: faster or equal vs. slower) contingency tables with the time boundary set at each integer value (i.e., 1 second, 2 second, 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 postdecision confidence. Based on the results of the boundary analyses, accuracy within and outside the optimum boundaries 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. The plots of  $\chi^2$ -values by boundaries are presented in Figures 4, 5, and 6.

*Boundary analyses.* As the decision time-accuracy (LA) relation was significant for choosers but not for nonchoosers (see Table 2) time boundary analyses were performed only on choosers’ data. The optimum time boundary was at 6 seconds,  $\chi^2(1, N = 436) = 50.19, p < .001, \phi = -.34$ . The mean proportion of correct decisions made before the optimum time boundary was 68.8% and 36.4% afterward.

The optimum postdecision confidence boundary was at 90%,  $\chi^2(1, N = 436) = 55.95, p < .001, \phi = .36$ . The mean proportion of correct decisions made within the optimum postdecision confidence boundary was 77.5%, and 38.3% below.

The *estimated* decision time-accuracy relationship was significant for choosers,  $r(434) = -.21, p < .001$ , but not for nonchoosers,  $r(282) = .01, p = .931$ , thus replicating the pattern of relationships for measured decision times. The optimum estimated time boundary was at 3 seconds,  $\chi^2(1, N = 436) = 17.52, p < .001, \phi = -.20$ . 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.

*Combining postdictors according to boundary results.* In the next step, decision times and postdecision confidence were combined according to the boundary analyses results. As the predicted difference between R and K judgments did not emerge, RK(F) judgments were not included. Table 3 shows identification accuracy for postdictor combinations. The columns labeled % correct contain the proportion of participants who were correct in this condition. For example, 56.8% of the 88 nonconfident fast choosers made a correct decision. The columns labeled  $M^a$  contain the unweighted mean for this cell across all 10 targets. This was slightly less (55.4%) for this cell.

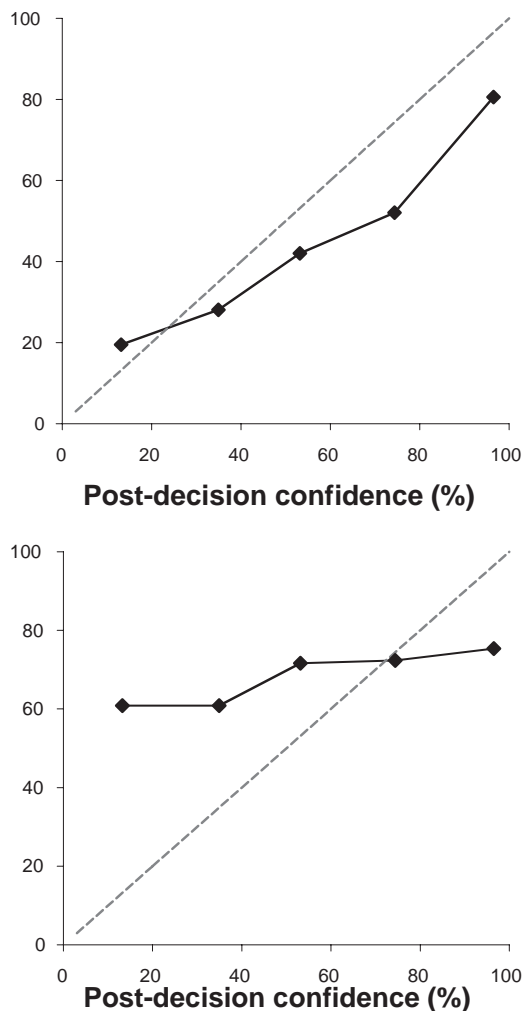


Figure 3. Choosers’ (top panel) and nonchoosers’ (lower panel) postdecision confidence-identification accuracy calibration curves.

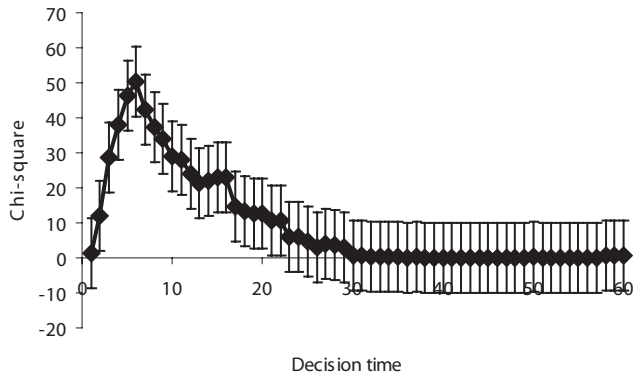


Figure 4. Plot of  $\chi^2$ -values (and 95% CI) by decision time boundary.

On average (computed across targets), choosers who decided within the first 6 seconds and gave a postdecision confidence rating of 90% or 100% ( $n = 57$ ) were correct in 97.2% of the cases ( $n = 55$ ). For less confident and slower choosers ( $n = 245$ ), accuracy was 31.8% on average ( $n = 78$ ). Please note, that altogether there were only 57 such highly confident and fast choosers and 245 slow and nonconfident choosers. Altogether there were 156 (73.9%) choosers outside one or both boundaries who were accurate.

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

*Internal replication of classification analysis.* Using the average boundaries across 10 targets may overestimate the accuracy of classifications based on these boundaries. Therefore, we computed boundaries for each target individually. The results can be found in Table 4. We then used the time and postdecision confidence boundaries for each target individually to predict identification accuracy for the remaining targets. The results are shown in Table 5. For example, in Column 1 of Table 4 the decision time boundary for Target 1 was at 4 seconds, the postdecision confidence boundary at 80%. Now, looking at the first column of Table 5, the results can be read as follows: Of those choosers regarding Targets 2 to 10,

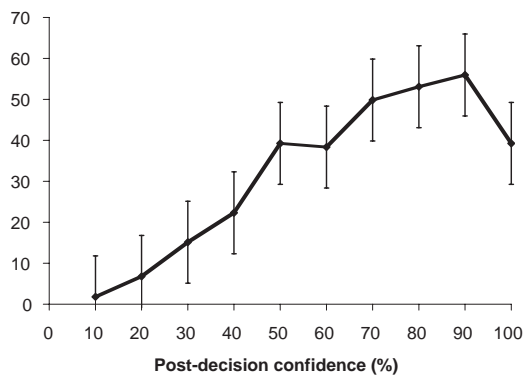


Figure 5. Plot of  $\chi^2$ -values (and 95% CI) by postdecision confidence boundary.

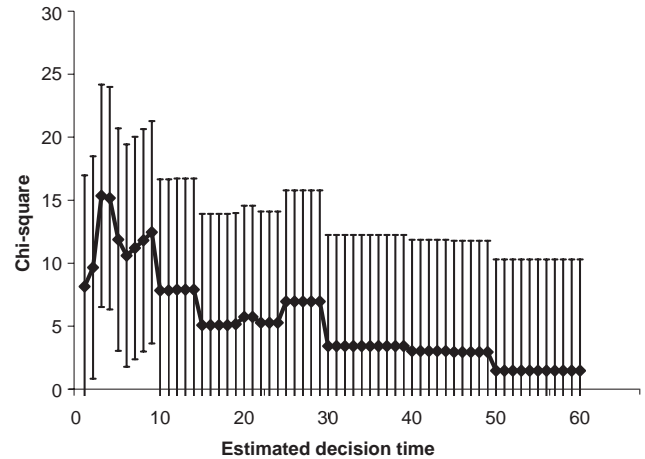


Figure 6. Plot of  $\chi^2$ -values (and 95% CI) by estimated decision time boundary.

who decided within 4 seconds, 78% made a correct decision ( $n = 85$ ). Of those choosers taking more than 4 seconds to make a decision 42% were accurate ( $n = 306$ ). Of the choosers making a decision with 80% confidence or more 74% ( $n = 133$ ) were correct, of those choosers making a decision with less than 80% confidence, 38% were correct ( $n = 258$ ). Choosers regarding Targets 2 to 10 making their decision within both boundaries computed for Target 1 were accurate in 90% of the cases ( $n = 52$ ), those making their decision outside both boundaries in 35% of the cases ( $n = 225$ ).

As in the previous analyses, combining both postdictors led to more accurate classifications than each by itself. On average, 84% (unweighted; 74% weighted) of the identifications were accurate, when both boundaries of one target were used as a basis. When witnesses were outside both boundaries, only 30% (unweighted; 31% weighted) were accurate.

Another way of internal replication could be to split the sample into two random subsamples whereby 50% of the choosers of each target were randomly assigned to one of two groups. We performed all analyses first for one subsample and then postdicted identification accuracy for the second subsample. The optimum time boundary for Subsample 1 was at 6 seconds,  $\chi^2(1, N = 218) = 25.68, p < .001, \phi = -.34$ . The proportion of correct decisions made before the optimum time boundary was 71.6% and 35.4% afterward. The optimum postdecision confidence boundary for Subsample 1 was at 70%,  $\chi^2(1, N = 218) = 35.26, p < .001, \phi = .40$ . The proportion of correct decisions made within the optimum postdecision confidence boundary was 68.9% and 28.7% outside the optimum postdecision confidence boundary. The optimum *estimated* time boundary was at 4 seconds,  $\chi^2(1, N = 218) = 9.06, p = .003, \phi = -.20$ . The proportion of correct decisions made before the optimum estimated time boundary was 67.4% and 42.4% afterward.

Choosers who decided within the first 6 seconds and made a postdecision confidence rating between 70% and 100% showed an average accuracy rate across targets of 82.9%. Slow and nonconfident choosers were on average only 23.2% accurate. Replacing decision times with estimated decision times produced an accuracy

Table 3  
Identification Accuracy of Choosers for Combinations of Confidence and Decision Times

Postdecision confidence	Decision time								
	Fast (≤6 s)			Slow (>6 s)			Fast and slow		
	<i>n</i>	% correct	<i>M</i> <sup>a</sup>	<i>n</i>	% correct	<i>M</i> <sup>a</sup>	<i>n</i>	% correct	<i>M</i> <sup>a</sup>
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	
0–100%	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	

<sup>a</sup> Mean accuracy calculated separately as the unweighted mean for this cell across all 10 targets (see text).

rate of 81.0%. Nonconfident choosers who estimated their decision as slower were on average 36.2% accurate.

Subsequently, the boundaries found for Subsample 1 were applied to Subsample 2. The proportion of correct decisions made within the 6 seconds optimum time boundary was 73.2% and 37.4% afterward,  $\chi^2(1, N = 218) = 24.59, p < .001, \phi = .34$ . The proportion of correct decisions made within the 70% optimum postdecision confidence boundary was 64.6% and 36.9% outside the optimum postdecision confidence boundary,  $\chi^2(1, N = 218) = 16.49, p < .001, \phi = -.28$ . The proportion of correct decisions made before the 4 second optimum estimated time boundary was 66.7% and 43.7% afterward,  $\chi^2(1, N = 218) = 8.24, p = .004, \phi = .19$ . In Subsample 2, choosers who lay within both optimum boundaries were on average across targets accurate in 81.0% of the cases. Slow and nonconfident choosers were on average 36.2% accurate. Applying the boundary of estimated decision times produced an accuracy rate of 78.4%. Choosers who estimated their decision as slower and who were less confident were on average accurate in 60.4% of the cases. Thus, with regard to combining postdictors we found comparable results when the optimum boundaries were computed for one subsample and then employed

to postdict identification accuracy of another subsample, thus successfully cross-validating our findings.

*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 times, decision confidence, and decision process measures of nonchoosers can be found in Table 6. 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,  $\chi^2(2, N = 284) = 2.13, p = .345, Cramer's V = .09$ , although the "absent" group showed a tendency to greater accuracy (72.5%) than the other two groups (64.1% and 63.4%, respectively).

Three 3 × 2 ANOVAs with the decision process groups ("absent" vs. "low confidence" vs. "no memory") and decision outcome (correct vs. incorrect) as independent variables and predecision confi-

Table 4  
CA Correlations, Decision Time-Accuracy Correlations, Postdecision Confidence Boundaries (in %), and Time Boundaries (in Seconds) as a Function of the Target Persons (N = 436 Choosers Only)

Variables/relationships	Targets										Total
	1 ( <i>n</i> = 45)	2 ( <i>n</i> = 44)	3 ( <i>n</i> = 46)	4 ( <i>n</i> = 44)	5 ( <i>n</i> = 42)	6 ( <i>n</i> = 50)	7 ( <i>n</i> = 38)	8 ( <i>n</i> = 38)	9 ( <i>n</i> = 46)	10 ( <i>n</i> = 43)	
Mean postdecision 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 (seconds) <sup>a</sup>	10.1	9.1	8.6	13.7	9.2	7.9	9.0	11.1	10.2	6.8	9.4
CA relation ( <i>r</i> )	.43**	.57**	.38**	.29 <sup>o</sup>	.39*	.19	.47**	.02	.56**	.43**	.39**
LA relation ( <i>r</i> )	-.25	-.40**	-.27	-.14	-.40**	-.13	-.25	.01	-.45**	-.43**	-.30**
Postdecision Confidence boundary (%)	80	60	70	100	90	20	90	ns <sup>b</sup>	70	90	90
Time boundary (in seconds)	4	6	5	ns <sup>b</sup>	6	5–6	5	ns <sup>b</sup>	14–16	10–12	6

<sup>a</sup> The means of decision times were backtransformed from the logarithmic values used for the inferential analyses. <sup>b</sup> When there was no significant association between the postdictor and identification accuracy, no boundary was computed.

\* *p* < .05, two-tailed. \*\* *p* < .01, two-tailed. <sup>o</sup> *p* < .10, two-tailed.

This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

**Table 5**  
*Identification Accuracy (%) of the Ten Targets for Postdictors and Their Combinations, Predicting From the Optimum Boundary of One Target to the Remaining Targets (Choosers Only)*

Postdictors	Targets										Unw./Weighted
	1	2	3	4	5	6	7	8	9	10	Mean
<b>Decision time</b>											
Within boundary	78	69	72	— <sup>a</sup>	71	73	75	— <sup>a</sup>	54	55	68/65
Outside boundary	42	35	37	— <sup>a</sup>	37	36	40	— <sup>a</sup>	33	35	37/37
<b>Postdecision confidence</b>											
Within boundary	74	59	65	82	80	49	81	— <sup>a</sup>	65	79	70/64
Outside boundary	38	32	32	41	38	24	40	— <sup>a</sup>	35	38	35/37
<b>Time and confidence</b>											
Within both boundaries	90	78	86	— <sup>a</sup>	96	73	98	— <sup>a</sup>	69	85	84/74
Outside both boundaries	35	27	28	— <sup>a</sup>	32	20	34	— <sup>a</sup>	28	32	30/31

<sup>a</sup> No boundary was computed as the relationship of this postdictor with identification accuracy was nonsignificant for this target (see Table 4).

dence, decision times, and postdecision confidence as dependent variables were computed. There were significant main effects of decision process on predecision confidence,  $F(2, 278) = 19.80, p < .001, f = .38$ , decision times,  $F(2, 278) = 6.82, p = .001, f = .22$ , and postdecision confidence,  $F(2, 278) = 36.08, p < .001, f = .51$ . Post hoc  $t$  tests showed that the “absent” group decided faster ( $M = 9.7$  seconds), with higher predecision confidence ( $M = 66.8\%$ ), and higher postdecision confidence ( $M = 69.9\%$ ) than both the “low confidence” ( $M_s = 14.0$  seconds, 56.4%, 45.6%) and “no memory” groups ( $M_s = 16.3$  seconds, 43.4%, 40.7%), all  $|t|s \geq 2.56, |d|s \geq 0.42$ . The “low confidence” and “no memory” groups only differed with regard to their predecision confidence ratings ( $M = 56.4\%$  vs. 43.4%, respectively),  $t(78) = 2.63, p = .010, d = 0.31$ . The main effect of decision outcome was significant for predecision confidence only,  $F(1, 278) = 6.02, p = .015, d = .29$ . Specifically, predecision confidence ratings were higher for accurate (58.5%) than inaccurate (49.9%) nonchoosers. The decision process by decision outcome interaction was nonsignificant for all three variables, all  $F_s \leq 1.36$ .

**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 by themselves and in combination. Second, the usefulness of witnesses’ estimated decision times as a postdictor of identification accuracy was examined. Third, we explored the utility of self-reported decision processes of nonchoosers as an assessment variable. Emphasis was placed on stimulus sampling and variability of encoding conditions by carrying out a large scale study with 720 participants and 10 targets in a field setting.

Previous research to test the usefulness of assessment variables has often looked at postdecision 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

**Table 6**  
*Means, SDs, and 95% CIs of Postdictors of Identification Accuracy for Nonchoosers as a Function of Reasons Given for Not Choosing (N = 284)*

Reason given	M	SD	95% CI
<b>“Target was not among them” (n = 204)</b>			
Accuracy	72.5%	44.7%	66.4–78.6%
Predecision confidence	66.8%	21.4%	63.9–69.7%
Postdecision confidence	69.9%	23.7%	66.6–73.2%
Decision time	9.7 s	2.3 s	8.7–10.9 s
Estimated decision time	8.9 s	3.0 s	7.7–10.4 s
<b>“Low confidence” (n = 39)</b>			
Accuracy	64.1%	48.6%	48.8–79.4%
Predecision confidence	56.4%	22.9%	49.2–63.6%
Postdecision confidence	45.6%	21.7%	38.8–52.4%
Decision time	14.0 s	2.4 s	10.7–18.4 s
Estimated decision time	14.2 s	2.6 s	10.5–19.1 s
<b>“No memory” (n = 41)</b>			
Accuracy	63.4%	48.8%	48.5–78.3%
Predecision confidence	43.4%	21.3%	36.8–49.8%
Postdecision confidence	40.7%	22.4%	33.8–47.6%
Decision time	16.3 s	2.1 s	13.1–20.3 s
Estimated decision time	14.5 s	2.9 s	10.5–20.1 s

This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

context of the identification (Sporer, 1993), that is, uncontaminated by postidentification feedback or other factors distorting their postdictive value (see Douglass & Steblay, 2006). Thus, the findings do not apply to later questioning of witnesses or in-court ("dock") identifications. Hence, either witness response forms should be used by a blind administrator who measures decision times inconspicuously, or the identification test should be administered by a computer program.

As in previous studies, significant positive CA and negative LA associations were found only for choosers but not for nonchoosers (e.g., Dunning & Perretta, 2002; Sporer et al., 1995; Weber et al., 2004), resulting in the expected interaction between Choice and Decision Outcome (Sporer, 1994). For choosers, the observed point-biserial correlations are considered large for confidence ( $r = .39$ ) and medium to large ( $r = -.30$ ) for decision times (Cohen, 1988, p. 81). Calibration analyses replicated the finding that there is high capability to discriminate accurate from incorrect choosers but not nonchoosers via postdecision confidence (Brewer & Wells, 2006; Brewer et al., 2002; Weber & Brewer, 2003, 2004, 2006). The observed overconfidence, particularly with high levels of confidence, appears well documented (Juslin et al., 1996; Olsson & Juslin, 1999; Weber & Brewer, 2004).

We know of no study where predecision confidence was calibrated. Here, calibration measures of predecision confidence showed only little capability to discriminate accurate from inaccurate witnesses. This is consistent with many correlational studies showing low or zero predecision CA relationships (e.g., Cutler & Penrod, 1988; Cutler, Penrod, O'Rourke, & Martens, 1986; Cutler, Penrod, & Martens, 1987; Gwyer & Clifford, 1997; Sporer, 1992). Despite low calibration we did find a significant positive correlation between predecision confidence and identification accuracy not only for choosers but also for nonchoosers, and hence for the whole sample.

Over a wide range of abilities and knowledge domains, 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 type of 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 predecision confidence and identification accuracy in many identification studies. 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), whether or not the target's appearance has changed (see Charman & Wells, 2007; Read, 1995), or how similar the foils will be to the suspect (e.g., Cutler & Penrod, 1988; Cutler et al., 1986, 1987; Gwyer & Clifford, 1997; Sporer, 1992, 1993). Considering these circumstances, it is not surprising that predecision 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 number, angle and quality of the pictures could help witnesses to make better calibrated estimates of predecision confidence.

### *Benefits and Limitations of Combining Postdictors*

We successfully combined postdecision confidence and decision times via boundary analyses that resulted in correct classifications of 97.2% of the fast and confident choosers who made correct identifications, and 68.2% correct classifications of slow and nonconfident choosers who made incorrect identifications, thus creating an opportunity to exonerate innocent suspects. The latter finding is consistent with Sauerland and Sporer (2007) who observed that slow and nonconfident participants were inaccurate in 95% of the cases. However, in the present study these two subgroups of choosers ( $n = 57$  and  $n = 245$ , respectively) comprise only a small subsample ( $n = 302$ ; 69.3%) of the 436 choosers. Altogether, there were 156 choosers outside one or both boundaries who were accurate. Note, however, that choosers who fulfilled only one of the two criteria were accurate in 55.4% (fast but nonconfident) and 60.4% (confident but slow) of the cases, demonstrating the importance of combining both variables rather than looking at each of them individually.

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 only slightly lower but yet high correct classification rates.

Additionally, we postdicted identification accuracy using the obtained boundaries from one target for all other targets. The result pattern paralleled the one observed for the whole sample although the correct classification rates varied as a function of target. Here, heterogeneous as well as homogeneous aspects across targets can be noted. It speaks for the homogeneity part of our results, that the Decision Outcome by Choice interaction was not modified by higher order interactions with target for any postdictor. Furthermore, there is a consistent result pattern indicating that the combination of both postdictors results in higher accurate classification rates than the use of a single postdictor.

On the other hand, heterogeneity across targets was also notable. Classification rates differed as a function of target, sometimes widely. These differences did not follow an easily discernible pattern. For example, they cannot simply be explained by target distinctiveness or attractiveness (which we also measured in an additional pilot study not reported here), nor as a function of the effective lineup sizes. We return to this point at the end of our discussion.

Boundary analyses for decision times and postdecision confidence revealed optimum boundaries of 6 seconds and 90%. This is inconsistent with the 10 to 12 second rule established by Dunning and Perretta (2002). Sauerland and Sporer (2007) found optimum boundaries of 18 seconds and 50% confidence. One difference between the present and the Sauerland and Sporer (2007) 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 seconds. Weber et al. (2004) reported optimum time boundaries that varied widely between 5 seconds and 29 seconds across four laboratory studies with retention intervals of 15 to 20 minutes between a filmed event and testing. Likewise, Brewer et al. (2006) found shorter optimum time boundaries for participants who were tested immediately (13-sec boundary) than for those who were tested

after 15 minutes (36 seconds) or 20 minutes (35 seconds). However, in that study, nominal lineup size (4 vs. 8 vs. 12) was also varied. Obviously, a larger set size at test will require longer inspection, and hence decision times. Therefore, we believe that estimates of time boundaries are only meaningful relative to nominal (or perhaps functional or effective) lineup size, thus dividing decision times by lineup members.

### *Estimated Decision Times*

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 estimate of the time taken for an 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 expected 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 because of feedback (cf., Semmler, Brewer, & Wells, 2004; Wells & Bradfield, 1998; Wells, Olson, & Charman, 2003; see the meta-analysis by Douglass & Steblay, 2006). Wells and colleagues (Wells & Bradfield, 1998, Experiment 1; Wells et al., 2003) found that participants receiving confirming feedback estimated their decision times 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 estimate with and without feedback should be tested more thoroughly in future studies. Although double-blind testing (Wells & Luus, 1990) could eliminate the effect of immediate feedback, there would still be the hazard of media influences. Therefore, video recordings of the identification procedure (or computerized administration) seem to be necessary to preserve as much information as possible about the decision making processes later to be used as assessment variables to evaluate the evidentiary value of the decision.

### *Converging Evidence From Archival Studies*

With the growing number of studies on the topic, we hope to gain more insight on how differences across experiments such as stimulus persons, lineup size and type, retention interval, and so forth affect optimum time and confidence boundaries. As of now, the optimum boundaries seem to vary from study to study, thus making it difficult to derive concrete recommendations for practice. From archival analyses in the United Kingdom we know that fast decisions are more likely to result in the identification of a suspect (87%) than average or slow decisions (38% vs. 31%;

Valentine, Pickering, & Darling, 2003), indicating that differences in decision speeds are associated with differences in identification outcomes in real cases.

Likewise, Behrman and Richard's (2005) archival analysis in the United States showed that decisions described by police officers as being made "without hesitation" or "immediately" happened more often when the suspect was identified than a foil. Whereas hesitant witnesses may leave jurors believing the witness was not very certain, and therefore not reliable, will jurors and other fact finders infer that an identification is accurate when the decision was made without hesitation and immediately? This is an interesting empirical issue that needs to be explored.

As Behrman and Richard also observed, the (what we may call) "spontaneous decision effect" was strongest for cases with substantial external evidence compared to minimal or no external evidence, thus supporting our and other findings that fast decisions are more likely to be correct. Interestingly, not only researchers have noticed the informative value of decision times, but also practitioners and legal scholars as far back as 1838 (Henke, 1838; see Sporer, 1982, 2008). Yet, we still need to know more about the distribution of decision times and postdecision confidence in real cases where the decision outcome is established unequivocally, for example, by DNA evidence.

According to the Remember-Know research tradition (e.g., Rotello et al., 2004) we would have expected choosers who gave a Remember response to be more accurate than those giving Know responses. Note, however, that as for CA relations, RKF judgments for identification decisions may not be the same with memory for details (Ibabe & Sporer, 2004). In the present study, we found no significant differences between Remember and Know responses but if at all there was a tendency in the opposite direction in both the RK and the RKF condition, which may be indirect support for Dunning and Stern's (1994) theorizing about automatic decision processes.

Yet, only in the RKF, not the RK condition did a significant association between RKF judgments and identification accuracy emerge: Witnesses who made a Familiar judgment were significantly less often accurate in their identification decisions 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 likely to be associated with accurate decisions in the RKF than in the RK condition. Hence, it does not appear crucial whether a witness judges that he or she Remembers or Knows that a person is the target, but whether he or she indicates the face to be Familiar only.

Fact finders should be very cautious with identification evidence given on the grounds that the face simply looked more familiar than the other ones (see the archival analysis by Behrman & Richards, 2005). If our results are replicated, explicitly giving a familiar option might be one way of eliminating witnesses who would make a wrong decision. However, our results should be interpreted with caution as the wording of the RKF judgments employed may have been ambiguous. The prompts given did not specify that viewing the target face in the lineup evoked the feeling of Remembering/Knowing/Familiarity but could also be inter-

preted by a participant that he or she remembered/knew/felt familiar with the target or the situation *before* reaching a decision.

### *Nonchoosers' Decision Processes*

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 nonsignificant trend in this direction (*Cramer's V* = .09). However, as expected, the "absent" group made their decisions faster and with higher pre- and postdecision confidence. The only difference that emerged between the "low confidence" and the "no memory" group was for predecision confidence. Contrary to the hypothesis, the "low confidence" group showed higher predecision confidence than the "no memory" group.

One may wonder why people would indicate "I thought the target was present, but I was not confident enough to choose him/her" rather than making a choice with low confidence. However, in another unpublished study Sauerland (2007) asked nonchoosers to give a free report on why they had not made a choice. Many participants uttered that they thought the target may have been present but that they were not sure enough to make a choice. Judging from their comments, participants set themselves a certain threshold that kept them from making a choice with low confidence (see Clark, 2003).

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 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 (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 using a sequential lineup procedure (as done by Sporer, 1993), 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, Lindsay, and Dupuis (2006) found a significant CA correlation for both choosers ( $r = .32$ ) and nonchoosers ( $r = .19$ ), supporting Weber and Brewer's (2006) hypothesis. Dysart, 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.

### *Limitations and Benefits of the Field Study Methodology*

Some caveats of the present findings need to be discussed. First, although we used a live setting, we used no crime scenario. Thus, it is unlikely that participants experienced the same arousal level as witnesses of a real crime would. Therefore, the results of the present study may not apply to cases with strong emotional involvement (see Deffenbacher, Bornstein, Penrod, & McGorty, 2004). Nonetheless, there are also many real cases that involve incidental observation, where it is not clear at the time of witnessing an event that a crime is taking place, let alone that a particular person will have to be identified. Furthermore, one has to bear in mind that it is rather unlikely that in the majority of studies with filmed crimes participants experience high levels of arousal, compared to real crime witnesses.

Recent research on memory of items versus source memory reported that item memory was less impaired by arousal than source memory (Kensinger & Schacter, 2006; Mather, Mitchell, Raye, Novak, Greene, & Johnson, 2006). For an eyewitness identification setting, this could mean that confidence ratings and RKF judgments may differ as a function of arousal level, possibly decreasing their relation with identification accuracy. Future studies should address this issue.

The lack of any consequences of the identification procedure is also a shortcoming of this and many other identification studies. However, the few studies that have used serious consequences tend to show that performance may actually be worse, rather than better with serious consequences (Leippe, Wells, & Ostrom, 1978; Malpass & Devine, 1980).

A second caveat in the present study is the short (30 seconds) retention 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 minutes or less. Perhaps, RKF judgments may be more conclusive when longer retention intervals are used. We believe that identification studies should employ an interval of at least one day between event and identification when there are no severe practical reasons that speak against this.

With respect to postdicting identification performance, Read (1995) and D. S. Lindsay et al. (1998) have stressed the advantages of field experiments (as well as the use of multiple targets and longer retention intervals). In field experiments, because of the restricted control over the situation, the variability in target-participant interaction is likely to produce different levels of encoding and hence variation in the strengths of memory traces that in turn will increase the variability in correct/incorrect decisions. Low CA and LA correlations may have been observed in the past because of the lack of variability in encoding conditions in these

studies that may have had limited variability either in decision outcome, or confidence, or both.

As we used a live event, the variation in targets entailed a number of complex variations in stimulus and encoding conditions that in sum led to the present results. It is this variation, however, that makes our results more generalizable. In real cases, no two interactions will ever be exactly alike, and criminals are likely to vary in their distinctiveness. As an analog, in the present experiment we used 10 different targets varying in gender, distinctiveness, attractiveness, and many other attributes, thus satisfying the requirement of stimulus sampling emphasized by Wells and Windschitl (1999). Both from laboratory studies on face recognition (see Shapiro & Penrod, 1986; Valentine, 1991) and from field experiments (Brigham, 1990) we know that target distinctiveness is related to recognition performance.

Sampling participants from passersby in a small city also assured a broad range in the age distribution that ranged from 15 to 84 years (with an *SD* of 15.3), which is much broader than in most studies published.

Noting these advantages of the field methodology used is not meant to be a critique of controlled laboratory studies that have tried to disentangle the determinants of eyewitness performance. Obviously, staged event methodologies can be used both in laboratory and in field settings, and better control (e.g., of exposure time) using laboratory/video film methodologies is necessary to isolate specific causal factors.

### Conclusion

To summarize, we successfully combined decision times and postdecision confidence and found that fast and confident choosers were highly accurate. Slow and nonconfident witnesses were mostly inaccurate. Our data do neither suggest that identification evidence given by fast and confident witnesses should always be relied upon, nor that less confident or slower choosers should be discarded although the chances of these witnesses making a false identification are much higher than for fast, confident choosers. Therefore, fact finders should be very cautious when considering the evidence of slow or nonconfident witnesses and try very hard to gather additional evidence apart from the identification decision. Estimated decision times also qualified as a postdictor of identification accuracy. For future studies, we emphasize the need to find out more about the distribution of decision times and postdecision confidence in real cases where the decision outcome can be independently established, for example, by DNA analysis. Although our attempt to identify a postdictor of nonchoosers' performance failed, distinguishing between different types of nonchoosers may be promising. 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., 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., & 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.
- Charman, S. D., & Wells, G. L. (2007). Eyewitness lineups: Is the appearance change instruction a good idea? *Law and Human Behavior, 31*, 3–22.
- Clark, S. E. (2003). A memory and decision model for eyewitness identification. *Applied Cognitive Psychology, 17*, 629–654.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- 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., & 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.
- Deffenbacher, K. A., Bornstein, B. H., Penrod, S. D., & McGorty, E. K. (2004). A meta-analytic review of the effects of high stress on eyewitness memory. *Law and Human Behavior, 28*, 687–706.
- Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition, 24*, 523–533.
- Douglass, A., & Steblay, N. (2006). Memory distortion in eyewitnesses: A meta-analysis of the post-identification feedback effect. *Applied Cognitive Psychology, 20*, 859–869.
- 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., 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.
- Henke, E. (1838). *Handbuch des Criminalrechts und der Criminalpolitik. Vierter Teil* [Handbook of criminal law and criminal politics. Pt. 4]. Berlin: Nicolai.
- Hirshman, E., & Master, S. (1997). Modeling the conscious correlates of recognition memory: Reflections on the remember-know paradigm. *Memory & Cognition, 25*, 345–351.
- Ibabe, I., & Sporer, S. L. (2004). The influence of question form on the accuracy and confidence of memories for central and peripheral action and descriptive details of an event. *Applied Cognitive Psychology, 18*, 711–726.
- 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). Amygdala activity is associated with the successful encoding of item, but not source, information for positive and negative stimuli. *Journal of Neuroscience, 26*, 2564–2570.
- Kneller, W., Memon, A., & Stevenage, S. (2001). Simultaneous and sequential lineups: Decision processes of accurate and inaccurate eyewitnesses. *Applied Cognitive Psychology, 15*, 659–671.
- Leippe, M. R., Wells, G. L., & Ostrom, T. M. (1978). Crime seriousness as a determinant of accuracy in eyewitness identification. *Journal of Applied Psychology, 63*, 345–351.
- 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.
- Malpass, R. S., & Devine, P. G. (1980). Realism and eyewitness identification research. *Law and Behavior, 4*, 347–358.
- Mäntylä, T. (1997). Recollections of faces: Remembering differences and knowing similarities. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 1203–1216.
- Mather, M., Mitchell, K. J., Raye, C. L., Novak, D. L., Greene, E. J., & Johnson, M. K. (2006). Emotional arousal can impair feature binding in working memory. *Journal of Cognitive Neuroscience, 18*, 614–625.
- 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.
- 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.
- Rajaram, S. (1996). Perceptual effects on remembering: Recollective processes in picture recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 365–377.
- Read, J. D. (1995). The availability heuristic in person identification: The sometimes misleading consequences of enhanced contextual information. *Applied Cognitive Psychology, 9*, 91–122.
- 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.
- Sauerland, M. (2007). Why do nonchoosers reject lineups? Unpublished raw data.
- Sauerland, M., & Sporer, S. L. (2007). Post-decision confidence, decision time, and self-reported decision processes as postdictors of identification accuracy. *Psychology, Crime, & Law, 13*, 611–625.
- 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.
- Shapiro, P., & Penrod, S. (1986). A meta-analysis of facial identification studies. *Psychological Bulletin, 100*, 139–156.
- 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. (1982). A brief history of the psychology of testimony. *Current Psychological Reviews, 2*, 323–340.
- 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. (2008). Lessons from the origin of eyewitness testimony research in Europe. *Applied Cognitive Psychology, 22*, 737–757.
- 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. (2007). *Using multivariate statistics*. New York: Pearson.
- 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. (1985). How many memory systems are there? *American Psychologist, 40*, 385–398.
- Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion and race in face recognition. *Quarterly Journal of Experimental Psychology, 43A*, 161–204.
- 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 confi-

- dence 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., & Luus, C. A. E. (1990). Police lineups as experiments: Social methodology as a framework for properly conducted lineups. *Personality and Social Psychology Bulletin*, 16, 106–117.
- 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., & 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., 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: Learning, 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.

Received March 21, 2007

Revision received October 6, 2008

Accepted October 14, 2008 ■

### Members of Underrepresented Groups: Reviewers for Journal Manuscripts Wanted

If you are interested in reviewing manuscripts for APA journals, the APA Publications and Communications Board would like to invite your participation. Manuscript reviewers are vital to the publications process. As a reviewer, you would gain valuable experience in publishing. The P&C Board is particularly interested in encouraging members of underrepresented groups to participate more in this process.

If you are interested in reviewing manuscripts, please write APA Journals at [Reviewers@apa.org](mailto:Reviewers@apa.org). Please note the following important points:

- To be selected as a reviewer, you must have published articles in peer-reviewed journals. The experience of publishing provides a reviewer with the basis for preparing a thorough, objective review.
- To be selected, it is critical to be a regular reader of the five to six empirical journals that are most central to the area or journal for which you would like to review. Current knowledge of recently published research provides a reviewer with the knowledge base to evaluate a new submission within the context of existing research.
- To select the appropriate reviewers for each manuscript, the editor needs detailed information. Please include with your letter your vita. In the letter, please identify which APA journal(s) you are interested in, and describe your area of expertise. Be as specific as possible. For example, “social psychology” is not sufficient—you would need to specify “social cognition” or “attitude change” as well.
- Reviewing a manuscript takes time (1–4 hours per manuscript reviewed). If you are selected to review a manuscript, be prepared to invest the necessary time to evaluate the manuscript thoroughly.