Journal of Applied Research in Memory and Cognition 1 (2012) 96–103

Contents lists available at SciVerse ScienceDirect

Journal of Applied Research in Memory and Cognition

journal homepage: www.elsevier.com/locate/jarmac

Individual differences in face identification postdict eyewitness accuracy

Markus Bindemann[a,*], Chennelle Brown[b], Tiffany Koyas[b], Andrew Russ[a]

[a] School of Psychology, University of Kent, UK
[b] Department of Psychology, University of Essex, UK

ARTICLE INFO

Article history:
Received 2 September 2011
Received in revised form 18 January 2012
Accepted 2 February 2012
Available online 15 February 2012

Keywords:
Eyewitness identification
Face perception
Individual differences
Postdictors

ABSTRACT

Eyewitnesses frequently mistake innocent suspects for the culprits of an observed crime, and such misidentifications have caused the wrongful convictions of many innocent people. This study attempted to establish the accuracy of individual eyewitnesses by assessing their ability to process unfamiliar faces. Observers viewed a staged crime and later tried to select the culprit from an identity lineup. This was followed by a face test that provides a laboratory analogue to lineup identifications. We found that this face test could determine the reliability of individual witnesses when a positive eyewitness identification had been made. Importantly, this was possible based on the specific response that a witness had made and without prior knowledge of whether the culprit was actually present in the lineup. These findings demonstrate that individual differences in face processing provide a potential instrument for *postdicting* eyewitness accuracy and for preventing miscarriages of justice.

1. Introduction

All justice systems rely heavily on eyewitnesses to identify the perpetrators of crime but this process is highly fallible. Eyewitness misidentifications are, in fact, the leading cause of convictions of innocent people (Scheck, Neufeld, & Dwyer, 2003; Wells, Memon, & Penrod, 2006). Many eyewitnesses do, however, make perfectly accurate person identifications. In criminal investigations and in court, the difficulty arises in differentiating accurate from inaccurate eyewitnesses.

Many of the factors that reduce eyewitness accuracy have now been identified (for reviews, see Wells & Olson, 2003; Wilcock, Bull, & Milne, 2008). Few factors have been found, however, that can be used directly to determine if a witness has made an accurate person identification for a critical event. Some recent investigations demonstrate that decision times and post-decision confidence can provide useful post hoc indicators to assess identification accuracy (see, e.g., Brewer, Weber, Clark, & Wells, 2008; Sauerland & Sporer, 2007, 2009; Sporer, 1992), but a complete method for determining the reliability of a specific witness remains elusive. Here we argue that eyewitness identification should depend directly on a factor that has so far only received limited attention in this context, namely the a priori ability of an eyewitness to process unfamiliar faces (see Wilmer et al., 2010; Zhu et al., 2010). This is based on the observation that eyewitness identification is, fundamentally, a perceptual task that requires the encoding, storage and subsequent recognition of a person's appearance. Provided that a perpetrator's face is not covered or disguised, the human face affords the most reliable means for such visual processing (Burton, Wilson, Cowan, & Bruce, 1999). The ability to process faces is therefore an important component of accurate eyewitness identification.

In some sense, this has been acknowledged widely already. It is well known, for example, that eyewitnesses are better at recognizing faces of their own race than those of another ethnic group (Meissner & Brigham, 2001; Shapiro & Penrod, 1986). However, investigations into the other-race effect and other aspects of face processing that impinge on eyewitness identification are typically based on broad comparisons, across groups of observers, rather than the assessment of individual abilities (see, e.g., Hancock, Bruce, & Burton, 2000; Johnston & Edmonds, 2009; Wells et al., 2006).

This lack of focus on individual differences in face processing is perhaps unsurprising. Considering, for example, the biological and social importance of the human face (see, e.g., Bruce & Young, 1998), it seems only natural to assume that most observers should possess the ability to process this class of stimuli and be capable therefore of remembering a face from a crime scene. This assumption is certainly justified when it comes to the identification of familiar faces, which is remarkably accurate (see, e.g., Bahrick, Bahrick, & Wittlinger, 1975; Burton et al., 1999; Liu, Seetzen, Burton, & Chaudhuri, 2003). On the other hand, there is increasing evidence that *unfamiliar* face identification is characterized by a broad distribution in ability between individual observers (Burton, White, & McNeill, 2010; Darling, Martin, Hellman, & Memon, 2009; Russell,

* Corresponding author at: School of Psychology, University of Kent, CT2 7NP, UK. Tel.: +44 01227 823087; fax: +44 01227 827030.
E-mail address: m.bindemann@kent.ac.uk (M. Bindemann).

2211-3681/$ – see front matter © 2012 Society of Applied Research in Memory and Cognition. Published by Elsevier Inc. All rights reserved.
doi:10.1016/j.jarmac.2012.02.001

Duchaine, & Nakayama, 2009; Woodhead & Baddeley, 1981). For example, in a simple face-matching task, in which observers have to decide if pairs of faces depict the same person or two different people, individual performance ranges from close-to-chance to perfect (Burton et al., 2010). Recognition memory for unfamiliar faces reveals a similar distribution of ability, with individual *d*' scores ranging from 0.5 to 6.8 for old-new decisions to previously seen and unseen faces (Woodhead & Baddeley, 1981).

The range of individual abilities in unfamiliar face processing is therefore substantial and there is now a growing body of research linking these differences to eyewitness accuracy. There is evidence, for example, that the self-reported face recognition skill of a person is associated positively with identification accuracy for the culprit of a staged crime (Olsson & Juslin, 1999). These reports were obtained after the identification test, which raises the possibility that recognition performance may have affected a person's self-assessment (see, e.g., Sporer, Penrod, Read, & Cutler, 1995). However, further research has shown a similar link when self-report is obtained prior to tests of face perception (Russell et al., 2009).

Other research has uncovered even more direct links between individual face perception ability and eyewitness accuracy. Hosch (1994) explored the relationship between eyewitness accuracy and performance in the Benton Facial Recognition Test (BFRT; Benton, Sivan, Hamsher, Varney, & Spreen, 1983), in which observers have to find the matches for a series of targets in a simultaneous lineup of six faces. In an extensive series of experiments, eyewitness identification accuracy was strongly and consistently related to BFRT performance. This was found for live encounters and video-based eyewitness scenarios, for identification lineups in which the culprit could be present or absent, and across short and extended delays (ranging from 40 min to 8 weeks). These findings have since been replicated independently (Geiselman, Tubridy, Blumkin, et al., 2001) and extended to live encounters in which observers experience levels of stress that are comparable to those in actual life-threatening crime events (Morgan et al., 2007).

These studies therefore provide converging evidence that individual differences in face perception ability are an important factor in eyewitness identification. These findings suggest further that the accuracy of individual witnesses could be assessed by measuring these differences, and could then be used to determine the weight that should be given to a person's testimony in criminal investigations and judicial proceedings (Geiselman, Tubridy, Barroso, et al., 2001). So far, however, these studies have focused on establishing *general* associations between eyewitness accuracy and face recognition ability. It remains unresolved if the accuracy of individual witnesses can be determined also from the *specific* response that a person has made to a police identity parade. Consider that the presence of a target cannot be established clearly in these instances, as a lineup may contain the sought-after culprit for a crime or simply an innocent suspect. It is therefore difficult to determine which type of lineup – with a target present or absent – an eyewitness is exposed to. This problem is compounded further as (1) both types of lineup can yield "absent" and "present" responses, and (2) a present response to a target-present lineup can indicate the correct identification of a culprit but also the misidentification of an innocent lineup member. To establish if an individual's face processing skills can be used as an index of his/her eyewitness accuracy, it is therefore essential to determine if it is possible to do so depending on the type of response that a witness has made. The aim of this study was to extend previous investigations in this important direction.

For this purpose, observers viewed a video of a staged theft, in line with other studies in this field (e.g., Geiselman, Tubridy, Blumkin, et al., 2001; Hosch, 1994; Olsson & Juslin, 1999). After a short delay, this was followed by an identity lineup in which the perpetrator could be present or absent. Observers had to decide if the perpetrator was in the lineup and if so, indicate who he is. This was then followed by an extended face identification test that mimics simultaneous police identity lineups (see, e.g., Bruce et al., 1999; Megreya & Burton, 2006). In each trial of this test, observers studied a target face and then had to determine its presence in a subsequent ten-person lineup. Crucially, this 1-in-10 test produces large individual differences in face identification performance (Megreya & Burton, 2006, 2008). Here we therefore sought to determine if this variation can provide an index of eyewitness accuracy for the observed theft.

These two tasks share key components so we provide distinct labels in advance to avoid confusion. The first task involving viewing of the crime video will be referred to as the *eyewitness test* and the associated identity lineups as the *perpetrator-present* and *perpetrator-absent* conditions. To distinguish these conditions from the subsequent *face test*, the conditions of this task will be referred to as *target-present* and *target-absent* lineups.

2. Experiment 1

2.1. Method

2.1.1. Participants

Eighty student and non-student volunteers (52 female, 28 male) participated in this experiment. The participants had a mean age of 25.5 years (SD = 10.6 years). All reported normal or corrected-to-normal vision.

2.1.2. Stimuli and procedure

2.1.2.1. Eyewitness test. In this experiment, participants first watched a video of a staged crime, which lasted for 1 min 6 s. In this video, the victim enters a waiting room, sits, and places his rucksack and coat on an empty seat. The perpetrator then enters and also sits down. A short time later, a third person fetches the victim, who leaves his belongings behind, from the waiting room. The perpetrator then searches the victim's rucksack for valuables, removes an item, and leaves.

Following the screening of the video, observers were given a filler task that involved a visual search for letters and numbers and took approximately 5 min to complete. Upon completion of the filler task, half of the participants were shown an identity lineup of ten faces comprising a photograph of the culprit and nine similar looking foils, in the perpetrator-present condition. The remaining participants were given a lineup of ten foil faces in the perpetrator-absent condition. All foil faces were taken from the Glasgow University Face Database (Burton et al., 2010) and were chosen by the experimenters to be of similar age and general appearance to the perpetrator. All participants were told explicitly that the perpetrator could be present or absent in these lineups and had to decide whether he was there, and if so, indicate who he is.

2.1.2.2. Face test. In the second phase of the experiment, participants were presented with forty trials of the 1-in-10 task (Bruce et al., 1999). In each trial of this task, observers were first presented with a single target face, which was a frame of high-quality video footage depicting a full-face image with a neutral expression at a size of 4.1 cm (*W*) × 5.3 cm (*H*). Observers were asked to study this target face until confident they could identify it from a subsequent lineup. The target was then instantly replaced by an identity lineup, consisting of ten full-face photographs at a size of 3.5 cm (*W*) × 4.5 cm (*H*), taken with a studio camera. At this point, observers had to indicate if the target was present in the lineup, and if so, identify who it is. Each observer was given twenty target-present and twenty target-absent trials in randomized order. A different target identity was used in each of these trials. There was

Table 1
Identification accuracy in the perpetrator-present and perpetrator-absent conditions of the eyewitness test as a function of performance (%) on the face test in Experiment 1. Parentheses show 95% confidence intervals.

Face test		Eyewitness test					
		Perpetrator-present			Perpetrator-absent		
		Correct witnesses $N=9$	Incorrect witnesses $N=31$	C – I	Correct witnesses $N=26$	Incorrect witnesses $N=14$	C – I
Target-present	Hits	75.6 [60.9,90.3]	60.8 [54.4,67.2]	14.7	66.5 [60.8,72.2]	66.8 [59.3,74.3]	−0.2
	Misses	13.3 [−2.9,29.5]	22.3 [15.3,29.3]	−8.9	23.5 [18.4,28.6]	18.9 [9.5,28.3]	4.5
	Misids	11.1 [2.5,19.7]	16.9 [12.7,21.1]	−5.8	10.0 [5.5,14.5]	14.3 [9.0,19.6]	−4.3
Target-absent	CRs	61.7 [36.3,87.1]	57.7 [48.7,66.7]	3.9	74.2 [66.5,81.9]	53.9 [39.3,68.5]	20.3

no time limit for any aspect of this task to encourage best-possible accuracy (for further details, see Bruce et al., 1999).

2.2. Results

2.2.1. Eyewitness test

To analyse this experiment, performance on the eyewitness test was first broken down into two groups of participants, corresponding to the observers who were either given a perpetrator-present or a perpetrator-absent lineup. These two groups were then split further into participants who made a correct or an incorrect identification decision. For the perpetrator-present condition, nine observers (22.5%) correctly identified the culprit from the lineup, while thirty-one observers (77.5%) made an incorrect response, by picking an innocent foil (35.0%) or by making an absent response (42.5%). For the perpetrator-absent condition, twenty-six observers (65.0%) responded correctly that the culprit was missing from the lineup, while fourteen observers (35.0%) incorrectly identified a foil face as the culprit.

2.2.2. Face test accuracy for the perpetrator-present and perpetrator-absent conditions

For the four groups of eyewitnesses (perpetrator-present correct/incorrect, perpetrator-absent correct/incorrect) performance on the face test was then calculated. Specifically, for the target-present trials of the face test, accuracy was analysed by calculating hits (identification of the correct face from the lineup), misidentifications (identification of a foil) and misses (the incorrect response that the target is absent). For the target-absent condition, performance was broken down into correct rejections (the correct response that the target is absent) and false positives (the false claim that the target is present). These two measures are inversely proportional so only correct rejections were used in the analysis. Hits, misidentifications, misses and correct rejections were then compared separately for the two perpetrator-present conditions (correct vs. incorrect eyewitnesses) and the two perpetrator-absent conditions (correct vs. incorrect eyewitnesses).

Table 1 illustrates accuracy on the eyewitness test as a function of mean performance on the face test. For the perpetrator-present eyewitness condition, these data show that the accurate observers of the eyewitness test generated more hits and correct rejections on the face test, and fewer target misses and misidentifications than inaccurate eyewitnesses. A series of independent-samples t-tests showed that these differences were reliable in hits, $t(38)=2.18$, $p<0.05$, $d=0.71$, but not in misses, $t(38)=1.21$, $p=0.24$, $d=0.39$, misidentifications, $t(38)=1.35$, $p=0.19$, $d=0.44$, and correct rejections, $t(38)=0.39$, $p=0.70$, $d=0.13$.

Table 1 also illustrates accuracy in the perpetrator-absent condition as a function of performance on the face test. The accurate eyewitnesses of this condition generated more correct rejections than inaccurate eyewitnesses, $t(38)=2.87$, $p<0.01$, $d=0.93$, while performance was more evenly matched in hits, $t(38)=0.05$, $p=0.96$, $d=0.02$, misses, $t(38)=0.98$, $p=0.33$, $d=0.32$, and misidentifications, $t(38)=1.22$, $p=0.23$, $d=0.40$.

2.2.3. Individual eyewitness accuracy and face test performance

Individual eyewitness accuracy was then plotted directly against performance on the face test. For this purpose, the observers were first divided into choosers and non-choosers, according to whether they had selected a person from the identity lineups of the eyewitness test or not (see Sporer, 1992; Havard & Memon, 2009). These two groups were then split further, into correct and incorrect choosers and non-choosers, according to their accuracy as an eyewitness. Thus, observers were classified as correct choosers if they had identified the right person from the perpetrator-present lineup, or as incorrect choosers if they selected the wrong person in the presence or absence of the culprit. Similarly, observers were categorized as correct non-choosers if they rightly noted the absence of the culprit from the perpetrator-absent lineup, or as incorrect non-choosers if they made an absent response to the perpetrator-present lineup.

The correct and incorrect choosers/non-choosers were then grouped according to their performance on the face test. For choosers, this was achieved by determining the proportion of correct eyewitnesses that achieved a set *minimum* percentage of hits on the face test. This analysis essentially determines whether the percentage hits that a person scores on the face test can distinguish whether an individual has previously made a correct or an incorrect person identification in the eyewitness test. Similarly, for non-choosers the proportion of correct eyewitnesses was compared against the percentage of correct rejections on the face test, to explore whether this measure can distinguish observers who had previously made a correct or an incorrect absent response in the eyewitness test. These data are provided in Table 2. In addition, we also supply cumulative probabilities of these measures by pooling data across descending accuracy bins, from 100% downwards. These cumulative measures amplify the probability of being a good witness but are interesting in that they provide a smoother function than the regular proportions (see Fig. 1 and Table 2).

This analysis shows that the probability of being a good witness can be established well for choosers. For this group, the percentage of hits on the face test correlated well with the standard probability, $r=0.699$, $p<0.05$, and also the cumulative probability of being a good witness, $r=0.861$, $p<0.01$. By contrast, the percentage of correct rejections on the face test provides a poorer index of eyewitness accuracy for non-choosers, as this measure did not correlate reliably with the standard probability, $r=0.490$, $p=0.18$, or with the cumulative probability of being a good witness, $r=-0.210$, $p=0.59$.

2.3. Discussion

This experiment sought to investigate whether it is possible to determine the accuracy of an eyewitness from his/her ability to process unfamiliar faces. As expected, overall performance was poor

Table 2
Estimates of eyewitness accuracy for choosers and non-choosers in Experiment 1. The standard and cumulative probabilities of being a good eyewitness are based on the percentage hits (for choosers) and correct rejections (for non-choosers) on the face test.

Choosers					Non-Choosers				
Face test Hits %	*N* (correct/total)	Probability Good witness	Cum. *N* (correct/total)	Cum. probability Good witness	Face test CRs %	*N* (correct/total)	Probability Good witness	Cum. *N* (correct/total)	Cum. probability Good witness
100	1/1	1.00	1/1	1.00	100	1/4	0.25	1/4	0.25
90–99	2/2	1.00	3/3	1.00	90–99	8/8	1.00	9/12	0.75
80–89	1/5	0.20	4/8	0.50	80–89	4/5	0.80	13/17	0.76
70–79	3/9	0.33	7/17	0.41	70–79	3/5	0.60	16/22	0.73
60–69	1/11	0.09	8/28	0.29	60–69	5/6	0.83	21/28	0.75
50–59	0/4	0.00	8/32	0.25	50–59	2/7	0.29	23/35	0.66
40–49	0/2	0.00	8/34	0.24	40–49	2/4	0.50	25/39	0.64
30–39	1/2	0.50	9/36	0.25	30–39	1/2	0.50	26/41	0.63
20–29	0/1	0.00	9/37	0.24	20–29	0/2	0.00	26/43	0.60
10–19	–	–	–	–	10–19	–	–	–	–
0–9	–	–	–	–	0–9	–	–	–	–

in the eyewitness test and in the face test. In the eyewitness test, more than three quarters of observers failed to identify the culprit in the perpetrator-present lineup and more than a third identified an innocent foil in his absence. Similarly, individual accuracy in the face test ranged from 10% to 100% of hits in the target-present condition and from 20% to 100% of correct rejections with target-absent displays. Importantly, however, we also found that eyewitness accuracy was related to performance on the face test. This association was already apparent from the group means, which revealed that the eyewitnesses who were able to identify the perpetrator were also more likely to identify the targets from the lineups of the face test, while the good witnesses of the perpetrator-absent group showed a similar advantage in correct rejections on target-absent trials.

The association between these two tasks is even more striking when these data are broken down further according to individual performance on the face test. In this analysis, the face test could not provide an index of eyewitness accuracy for non-choosers, who had not made a person identification in the eyewitness test. However, we found a reliable association between performance on the face test and those instances in which an eyewitness identification had previously been made (see Table 2 and Fig. 1). Moreover, within this group of observers, the individuals who achieved 90% of hits or higher on the face test were also always accurate eyewitnesses, whereas eyewitness accuracy declined with lower scores. This shows that the face test can provide a useful index of individual eyewitness accuracy.

This is an important finding but is based on a limited sample, as only nine of the forty witnesses in the perpetrator-present condition could identify the target. While this reinforces the general difficulty of eyewitness identification, we therefore sought to replicate these findings in a further experiment. For this purpose, we tested a larger group of observers with the same eyewitness test and a shorter version of the face test in Experiment 2.

3. Experiment 2

3.1. Method

3.1.1. Participants

One hundred and eighty-five students (149 female, 36 male) participated in this experiment. The participants had a mean age of 19.6 years (SD = 4.1 years). All reported normal or corrected-to-normal vision and none had participated in the preceding experiment.

3.1.2. Stimuli and procedure

Testing was conducted in four separate groups of 40–50 observers in a university lecture theatre. The stimulus materials were projected onto a screen. The seating of the participants was

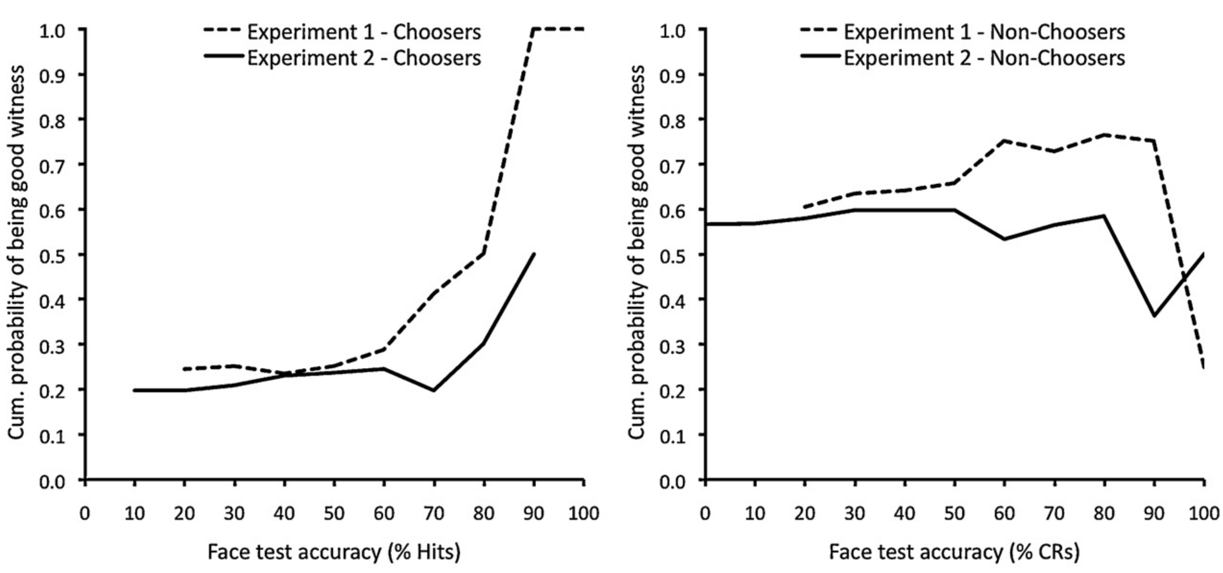

Fig. 1. The cumulative probabilities of being a good eyewitness for choosers (left graph) and non-choosers (right graph) as a function of performance on the face test in Experiment 1 and 2.

Table 3
Identification accuracy in the perpetrator-present and perpetrator-absent conditions of the eyewitness test as a function of performance (%) on the face test in Experiment 2. Parentheses show 95% confidence intervals.

Face test		Eyewitness test					
		Perpetrator-present			Perpetrator-absent		
		Correct witnesses $N = 17$	Incorrect witnesses $N = 77$	C – I	Correct witnesses $N = 56$	Incorrect witnesses $N = 35$	C – I
Target-present	Hits	57.1 [48.8,65.4]	51.6 [47.4,55.8]	5.5	50.5 [45.6,55.4]	52.3 [46.3,58.3]	-1.8
	Misses	27.6 [19.6,35.6]	32.1 [28.4,35.8]	−4.5	32.5 [27.1,37.9]	30.9 [24.4,37.4]	1.6
	Misids	15.3 [8.7,21.9]	16.4 [13.0,19.8]	−1.1	17.0 [13.1,20.9]	16.9 [12.5,21.3]	0.1
Target-absent	CRs	42.9 [32.8,53.0]	55.1 [49.9,60.3]	−12.2	60.0 [55.0,65.0]	49.4 [42.6,56.2]	10.6

arranged so that all had a clear view of the screen but the seats either side of each person remained empty. The participants recorded their responses on pre-printed answer sheets.

3.1.2.1. Eyewitness test. In the eyewitness test, participants watched the same crime video as in Experiment 1. Following the screening of the video, observers were given 3 min to complete the filler task (visual search for letters and numbers). The participants were then either shown a perpetrator-present or a perpetrator-absent lineup for the crime video. All participants were told explicitly that the perpetrator could be present or absent in these lineups and had to decide whether he was there, and if so, indicate who he is. All testing was conducted on the same day and the presentation of the perpetrator-absent and -present lineups was counterbalanced (APPA) across groups.

3.1.2.2. Face test. The face test consisted of ten target-present and ten target-absent trials of the 1-in-10 task (Bruce et al., 1999). We selected the lineups that yielded the highest error rates in Experiment 1 to provide a sensitive test for individual differences. On each trial of the face test, observers were first presented with a single target face, which was projected onscreen for 6 s. This presentation time was derived from a pilot study and reflects the average duration for which observers study these stimuli under self-paced viewing conditions. Observers were then shown a ten-face identity lineup for 30 s and had to decide if the target was present in the lineup, and if so, identify who it is. The stimuli were presented in the same fixed random order to all participants.

3.2. Results

3.2.1. Eyewitness test and face test

In the eyewitness test, seventeen observers (18.1%) correctly identified the culprit from the lineup in the perpetrator-present condition, while seventy-seven observers (81.9%) made an incorrect response, by picking an innocent foil (36.2%) or by making an absent response (45.7%). For the perpetrator-absent condition, fifty-six observers (61.5%) responded correctly that the culprit was missing from the lineup, while thirty-five observers (38.5%) incorrectly identified a foil face.

The correct and incorrect observers in the perpetrator-present and -absent condition of the eyewitness test were then compared further according to their performance on the face test (see Table 3). As in Experiment 1, this analysis shows that the accurate eyewitnesses in the perpetrator-present condition generated more hits on the subsequent face test, but fewer target misses, misidentifications, and correct rejections than inaccurate eyewitnesses. Independent-samples t-tests showed that these differences were reliable in correct rejections, $t(92) = 2.00$, $p < 0.05$, $d = 0.42$, but not in hits, $t(92) = 1.13$, $p = 0.33$, $d = 0.24$, misses, $t(92) = 1.02$, $p = 0.69$, $d = 0.21$, and misidentifications, $t(92) = 0.27$, $p = 0.59$, $d = 0.06$. In the perpetrator-absent condition, the accurate eyewitnesses also generated more correct rejections on the face test than inaccurate eyewitnesses, $t(89) = 2.56$, $p < 0.05$, $d = 0.54$, while performance was evenly matched in hits, $t(89) = 0.45$, $p = 0.93$, $d = 0.10$, misses, $t(89) = 0.39$, $p = 0.36$, $d = 0.08$, and misidentifications, $t(89) = 0.36$, $p = 0.28$, $d = 0.08$.

3.2.2. Individual eyewitness accuracy and face test performance

In the analysis of most importance, the eyewitness accuracy of choosers and non-choosers was plotted against their performance on the face test (see Fig. 1 and Table 4). For choosers, the percentage of hits on the face test correlated well with the standard probability, $r = 0.825$, $p < 0.01$, and also the cumulative probability of being a good witness, $r = 0.712$, $p < 0.05$. In non-choosers, on the other hand, the percentage of correct rejections on the face test did not correlate reliably with the standard probability, $r = 0.382$, $p = 0.25$, or with the cumulative probability of being a good witness, $r = -0.554$, $p = 0.08$.

3.3. Discussion

This experiment replicates the important aspects of Experiment 1 with a larger sample of participants. Overall performance was poor again on the eyewitness test and the subsequent face test, but eyewitness accuracy was associated clearly with individual performance in face identification. As in Experiment 1, the face test could not distinguish between correct and incorrect non-choosers. However, the choosers who had made a correct eyewitness identification were also more likely to score highly on the face test than incorrect choosers. This provides further evidence that individual differences in face identification can provide a useful index of eyewitness accuracy.

4. General discussion

This study sought to investigate whether it is possible to determine the accuracy of eyewitnesses from their ability to process unfamiliar faces. In line with previous findings, overall performance was poor in the eyewitness test (e.g., Hosch, 1994; Geiselman, Tubridy, Blumkin, et al., 2001; Morgan et al., 2007; Olsson & Juslin, 1999) and in the face test (Bruce et al., 1999; Megreya & Burton, 2006, 2008), but the face test also revealed a very broad distribution in ability across participants. Importantly, we found that these individual differences on the face test were clearly related to eyewitness accuracy.

These findings converge with other attempts that have used tests of face perception to assess the reliability of eyewitnesses (Geiselman, Tubridy, Blumkin, et al., 2001; Hosch, 1994; Morgan et al., 2007). Each of these studies has placed emphasis on different factors, such as the delay between exposure to a culprit and subsequent identification (Hosch, 1994), the stress experienced during exposure to a culprit (Morgan et al., 2007), or the difficulty of identity parades (Geiselman, Tubridy, Blumkin, et al., 2001). These studies are unified, however, on the basis that a person's face

Table 4
Estimates of eyewitness accuracy for choosers and non-choosers in Experiment 2. The standard and cumulative probabilities of being a good eyewitness are based on the percentage hits (for choosers) and correct rejections (for non-choosers) on the face test.

Choosers					Non-Choosers				
Face test Hits %	*N* (correct/total)	Probability Good witness	Cum. *N* (correct/total)	Cum. probability Good witness	Face test CRs %	*N* (correct/total)	Probability Good witness	Cum. *N* (correct/total)	Cum. probability Good witness
100	–	–	–	–	100	1/2	0.50	1/2	0.50
90–99	1/2	0.50	1/2	0.50	90–99	3/9	0.33	4/11	0.36
80–89	2/8	0.25	3/10	0.30	80–89	10/13	0.77	14/24	0.58
70–79	1/10	0.10	4/20	0.20	70–79	8/15	0.53	22/39	0.56
60–69	5/17	0.29	9/37	0.24	60–69	10/21	0.48	32/60	0.53
50–59	4/18	0.22	13/55	0.24	50–59	14/17	0.82	46/77	0.60
40–49	3/15	0.20	16/70	0.23	40–49	6/10	0.60	52/87	0.60
30–39	1/12	0.08	17/82	0.21	30–39	3/5	0.60	55/92	0.60
20–29	0/3	0.00	17/85	0.20	20–29	0/3	0.00	55/95	0.58
1–19	0/1	0.00	17/86	0.20	1–19	0/2	0.00	55/97	0.57
0–9	–	–	–	–	0–9	1/2	0.50	56/99	0.57

perception skill appears to be related to his/her aptitude as a witness. The current study extends these findings in an important way, as previous investigations have stopped short of assessing whether the accuracy of a witness can be determined from the specific identification response that is made to a lineup. This is a complicated issue considering that a lineup identification can reflect an accurate identification of the perpetrator or the identification of an innocent lineup member in the presence or absence of the culprit. Similarly, an "absent" response by a witness can reflect the correct knowledge that the perpetrator is missing from a lineup or the failure to spot the perpetrator when he is, in fact, there.

The current study extends previous research by demonstrating that eyewitness accuracy can be determined under these precise conditions. Specifically, in both experiments we found a reliable association between the instances in which a positive eyewitness identification had been made (i.e., in choosers) and the percentage of correct person identifications (hits) on the subsequent face test. In contrast to these successes, the face test could not provide an index of eyewitness accuracy for lineup rejections, in which no person identification was made (i.e., in non-choosers).

A similar discrepancy has been found with other postdictors of eyewitness accuracy, such as post-decision confidence and response times, which also appear to be limited to choosers (see, e.g., Sauer, Brewer, Zweck, & Weber, 2010; Sporer et al., 1995). These findings might reflect some fundamental differences in the processes that underlie lineup identifications and rejections. The former require a match between a person's memory of a culprit and *one* of the faces comprising an identity lineup (Sporer, 1993). The latter, on the other hand, reflect decisions that are based on the failure to find such a match in *all* of the lineup faces (see Sauerland & Sporer, 2009; Weber & Brewer, 2006). Laboratory studies of face identification already demonstrate that these processes are dissociable (see, e.g., Megreya & Burton, 2007; Vokey & Read, 1992), and a correlational analysis of hits and correct rejections in the face test indicates that the abilities to make lineup identifications and rejections are also not associated here (Experiment 1: $r = 0.090$, $p = 0.43$; Experiment 2: $r = 0.055$, $p = 0.46$).

It is noteworthy that the absence of such a correlation also provides some evidence against an explanation for the present findings in terms of a response bias, whereby the correct choosers might have simply been more likely to select a person from the identity lineups than incorrect choosers. Accordingly, one might have expected the correct choosers also to report more false alarms (as indexed by fewer correct rejections), but the correlational analysis of hits and correct rejections shows that this was clearly not the case. This notion is undermined further by an inspection of face test performance for the choosers with the highest hit rates. In Experiment 2, for example, the two choosers who achieved the highest percentage hits (90%) produced 80% and 50% of correct rejections on target-absent trials, respectively. These percentage scores are not only notably different from each other but also above the group mean of 43% for all correct choosers (see Table 3).

While the absence of a correlation between lineup identifications and rejections might seem puzzling (see Megreya & Burton, 2007), research on eyewitness identification points to a possible explanation for this phenomenon. This is based on the observation that lineup rejections might reflect several different response categories that are often grouped together under this umbrella term (see Sauerland, Sagana, & Sporer, in press; Sauerland & Sporer, 2009; Sporer, 1992). Some of these responses may be based on the explicit knowledge that a target is indeed absent from a lineup. Such rejections might also be made, however, when there is less certainty regarding a correct decision. For example, absent responses could also occur when witnesses suspect the target to be in a lineup but feel too uncertain to make an identification, or when witnesses reject a lineup because they simply do not know whether a target is present or not (Sauerland & Sporer, 2009).

These distinctions appear to be promising for postdicting the eyewitness accuracy of non-choosers. When identity *showups* are used to assess eyewitness accuracy, dividing non-choosers' responses in this manner improves decision confidence and response times as postdictors of eyewitness accuracy (Sauerland et al., in press; Sauerland & Sporer, 2009). In the current study, we employed identity *lineups* and followed the conventional response options that are offered to observers in laboratory studies with these stimuli (see, e.g., Bruce et al., 1999; Megreya & Burton, 2007, 2008). As a consequence, we are unable to distinguish between different absent decisions. However, it is conceivable that eyewitness identification accuracy can be postdicted also for non-choosers when their responses on the eyewitness and the face test are broken down accordingly. Knowing whether an absent response reflects a correct (or an incorrect) lineup rejection is also of importance for the criminal justice system, for example, for establishing a suspect's innocence (see, e.g., Clark & Wells, 2008; Wells & Olson, 2002). This is therefore a worthwhile issue for further research (see Sauerland et al., in press).

5. Practical application

We believe that our findings highlight a procedure that could be of important application. Eyewitness testimonies often provide the primary leads in criminal investigations and juries strongly believe this evidence in court (Cutler, Penrod, & Stuve, 1988; Kassin & Neumann, 1997). However, despite its widespread use, a wealth of evidence indicates that this process is unreliable because eyewitnesses are often inaccurate (e.g., Scheck et al., 2003; Wells et al., 1998). The current study provides a potential solution to this problem by demonstrating a method that can determine the accuracy of eyewitness identifications from a person's ability to process unfamiliar faces. The long version of our face test (Experiment 1) takes approximately 30 min to complete and can be administered quite easily. The outcome of this test can then be compared with a set of probabilities, such as those provided in Fig. 1, to deduce the likely accuracy of a witness. This procedure could therefore provide some simple rules for professionals in the criminal justice system for evaluating the identification accuracy of individual eyewitnesses.

We describe this possibility to demonstrate that the precise postdiction of eyewitness identification accuracy is becoming increasingly feasible. However, our approach is clearly reductionist as it ignores the wide range of factors that can affect eyewitness accuracy (see, e.g., Wells & Olson, 2003; Wilcock et al., 2008), and measuring face identification ability alone is clearly also not sufficient. In Experiment 1, for example, we could only postdict eyewitness accuracy with *certainty* for 8% of choosers, who achieved 90–100% hits on the face test, and these probabilities were even lower with the shorter face test in Experiment 2.

Our findings therefore provide proof that face identification ability is a useful postdictor of eyewitness accuracy but the implementation of such a procedure in criminal investigations is certainly premature. We fully anticipate that a combination of postdictors is likely to improve these prospects further (see, e.g., Charman & Cahill, in press; Sauerland & Sporer, 2007, 2009; Sporer, 1992) but these possibilities require meticulous testing first. The implementation of postdictors of eyewitness accuracy within the criminal justice system is also precluded by additional issues, as such methods could raise legal objections on the basis that they might prejudice jurors' verdicts. Experiments in mock-trial settings suggest that such prejudicial effects can be avoided when jurors are briefed about the conditions in which postdictors are most effective (Geiselman, Tubridy, Barroso, et al., 2001). These mock-trial studies differ from actual court cases but raise hopes that such obstacles can be overcome (Wiener, Krauss, & Lieberman, 2011). Finally, the practical implementation of a postdictor test would also require training of expert operators and strict guidelines to ensure it is applied with an unbiased procedure. In the current study, the shorter face test of Experiment 2 resulted in fewer hits and appeared to reduce the possibility of detecting accurate eyewitnesses. How the implementation of postdictor variables is standardized might therefore be crucial to their success. These are issues that await solution.

References

Bahrick, H. P., Bahrick, O. O., & Wittlinger, R. P. (1975). Fifty years of memory for names and faces: A cross-sectional approach. *Journal of Experimental Psychology: General*, *104*, 54–75.

Benton, A. L., Sivan, A. B., Hamsher, K., Varney, d. S., & Spreen, N. R. O. (1983). Facial recognition: Stimulus and multiple choice pictures. In *Contributions to neuropsychological assessment*. New York: Oxford University Press.

Brewer, N., Weber, N., Clark, A., & Wells, G. L. (2008). Distinguishing accurate from inaccurate eyewitness identifications with an optional deadline procedure. *Psychology, Crime & Law*, *14*, 397–414.

Bruce, V., Henderson, Z., Greenwood, K., Hancock, P. J. B., Burton, A. M., & Miller, P. (1999). Verification of face identities from images captured on video. *Journal of Experimental Psychology: Applied*, *5*, 339–360.

Bruce, V., & Young, A. W. (1998). *In the eye of the beholder: The science of face perception*. Oxford University Press.

Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow face matching test. *Behavior Research Methods*, *42*, 286–291.

Burton, A. M., Wilson, S., Cowan, M., & Bruce, V. (1999). Face recognition in poor-quality video: Evidence from security surveillance. *Psychological Science*, *10*, 243–248.

Charman, S. D., & Cahill, B. S. Witnesses' memories for lineup fillers postdict their identification accuracy. *Journal of Applied Research in Memory and Cognition*, in press.

Clark, S. E., & Wells, G. L. (2008). On the diagnosticity of multiple-witness identifications. *Law and Human Behavior*, *32*, 406–422.

Cutler, B. L., Penrod, S. D., & Stuve, T. E. (1988). Jury decision making in eyewitness identification cases. *Law and Human Behavior*, *12*, 41–56.

Darling, S., Martin, D., Hellman, J. H., & Memon, A. (2009). Some witnesses are better than others. *Personality and Individual Differences*, *47*, 369–373.

Geiselman, R. E., Tubridy, A., Barroso, A., McClean, R., Mozafarian, M., & Zoumberakis, H. (2001). Benton facial recognition test scores: Their effect on juror verdicts. *American Journal of Forensic Psychology*, *19*, 89–97.

Geiselman, R. E., Tubridy, A., Blumkin, R., Schroppel, T., Turner, L., Yoakum, K., et al. (2001). Benton facial recognition test scores: Index of eyewitness accuracy. *American Journal of Forensic Psychology*, *19*, 77–88.

Hancock, P. J. B., Bruce, V., & Burton, A. M. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Sciences*, *4*, 330–337.

Havard, C., & Memon, A. (2009). The influence of face age on identification from a video lineup: A comparison between older and younger adults. *Memory*, *17*, 847–859.

Hosch, H. (1994). Individual differences in personality and eyewitness identification. In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 328–347). New York: Cambridge University Press.

Johnston, R. A., & Edmonds, A. J. (2009). Familiar and unfamiliar face recognition: A review. *Memory*, *17*, 577–596.

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

Liu, C. H., Seetzen, H., Burton, A. M., & Chaudhuri, A. (2003). Face recognition is robust with incongruent image resolution: Relationship to security images. *Journal of Experimental Psychology: Applied*, *9*, 33–41.

Megreya, A. M., & Burton, A. M. (2006). Unfamiliar faces are not faces: Evidence from a matching task. *Memory & Cognition*, *34*, 865–876.

Megreya, A. M., & Burton, A. M. (2007). Hits and false positives in face matching: A familiarity-based dissociation. *Perception & Psychophysics*, *69*, 1175–1184.

Megreya, A. M., & Burton, A. M. (2008). Matching faces to photographs: Poor performance in eyewitness memory (without the memory). *Journal of Experimental Psychology: Applied*, *14*, 364–372.

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

Morgan, C. A., Hazlett, G., Baranoski, M., Doran, A., Southwick, S., & Loftus, E. (2007). Accuracy of eyewitness identification is significantly associated with performance on a standardized test of face recognition. *International Journal of Law & Psychiatry*, *30*, 213–223.

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

Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic Bulletin & Review*, *16*, 252–257.

Sauer, J., Brewer, N., Zweck, T., & Weber, N. (2010). The effect of retention interval on the confidence-accuracy relationship for eyewitness identification. *Law and Human Behavior*, *34*, 337–347.

Sauerland, M., Sagana, A., & Sporer, S. L. Assessing nonchoosers' eyewitness identification accuracy from photographic showups by using confidence and response times. *Law and Human Behavior*, in press.

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.

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

Scheck, B., Neufeld, P., & Dwyer, J. (2003). *Actual innocence*. New York: Random House.

Shapiro, P., & Penrod, S. (1986). A meta-analysis of facial identification studies. *Psychological Bulletin*, *100*, 139–156.

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

Vokey, J. R., & Read, J. D. (1992). Familiarity, memorability, and the effect of typicality on the recognition of faces. *Memory & Cognition*, *20*, 291–302.

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

Wells, G. L., Memon, A., & Penrod, S. D. (2006). Eyewitness evidence: Improving its probative value. *Psychological Science in the Public Interest*, *7*, 45–75.

Wells, G. L., & Olson, E. A. (2002). Eyewitness identification: Information gain from incriminating and exonerating behaviors. *Journal of Experimental Psychology: Applied*, *8*, 155–167.

Wells, G. L., & Olson, E. A. (2003). Eyewitness testimony. *Annual Review of Psychology*, *54*, 227–295.

Wells, G. L., Small, M., Penrod, S., Malpass, R. S., Fulero, S. M., & Brimacombe, C. A. E. (1998). Eyewitness identification procedures: Recommendations for lineups and photospreads. *Law & Human Behavior*, *22*, 1–39.

Wiener, R. L., Krauss, D. A., & Lieberman, J. D. (2011). Mock jury research: Where do we go from here? *Behavioural Sciences and the Law*, *29*, 467–479.
Wilcock, R., Bull, R., & Milne, R. (2008). *Witness identification in criminal cases*. Oxford University Press.
Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Williams, M., Loken, E., et al. (2010). Human face recognition ability is specific and highly heritable. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 5238–5241.
Woodhead, M. M., & Baddeley, A. D. (1981). Individual differences and memory for faces, pictures, and words. *Memory & Cognition*, *9*, 368–370.
Zhu, Q., Song, Y., Hu, S., Li, X., Tian, M., Zhen, Z., et al. (2010). Heritability of the specific cognitive ability of face perception. *Current Biology*, *20*, 137–142.