Behavioral Sciences and the Law
Behav. Sci. Law **31**: 624–636 (2013)
Published online 9 September 2013 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/bsl.2082

Witnesses' Blindness for their Own Facial Recognition Decisions: A Field Study

Anna Sagana*, Melanie Sauerland, Ph.D.† and Harald Merckelbach†

In a field study, we examined choice blindness for eyewitnesses' facial recognition decisions. Seventy-one pedestrians were engaged in a conversation by two experimenters who pretended to be tourists in the center of a European city. After a short interval, pedestrians were asked to identify the two experimenters from separate simultaneous six-person photo lineups. Following each of the two forced-choice recognition decisions, they were confronted with their selection and asked to motivate their decision. However, for one of the recognition decisions, the chosen lineup member was exchanged with a previously unidentified member. Blindness for this identity manipulation occurred at the rate of 40.8%. Furthermore, the detection rate varied as a function of similarity (high vs. low) between the original choice and the manipulated outcome. Finally, choice manipulations undermined the confidence–accuracy relation for detectors to a greater degree than for blind participants. Stimulus ambiguity is discussed as a moderator of choice blindness.
Copyright © 2013 John Wiley & Sons, Ltd.

INTRODUCTION

People are confronted with hundreds of decisions daily. Some are of minor significance, whereas others have the power to cause profound changes to one's life. Yet, recent experiments demonstrate that people are sometimes poor in monitoring their decision outcomes if their choices are manipulated. Johansson et al. (2005) dubbed this phenomenon choice blindness. In their first demonstration of this phenomenon, participants were asked to select the more attractive alternative from pairs of female faces. Subsequently, they were given the selected face and were asked to explain the reasons underlying their choice. On three of the 15 trials, however, participants were presented with the opposite of the chosen picture. Surprisingly, the overwhelming majority (87%) of the manipulated trials went undetected right after the manipulation (i.e., concurrently).

Choice blindness has been documented with visual (Johansson, Hall, & Sikström, 2008), auditory (Sauerland, Sagana, & Otgaar, 2012), gustatory, and olfactory stimuli (Hall et al., 2010). Furthermore, stimulus similarity has been shown to moderate the

*Correspondence to: Anna Sagana, Section Forensic Psychology, Department of Clinical Psychological Science, Maastricht University, P.O. Box 616, 6200 MD Maastricht, the Netherlands. E-mail: anna.sagana@maastrichtuniversity.nl
†Department of Clinical Psychological Science, Maastricht University

Copyright © 2013 John Wiley & Sons, Ltd.

effect (Hall et al., 2010; Sauerland et al., 2012), although some studies have found no such effect (Johansson et al., 2005).

Despite the apparent robustness of choice blindness in the laboratory, one might wonder whether the effect would occur when poor monitoring of choices could have serious consequences. To address this question, Sauerland et al. (2013b) examined blindness for one's own history of norm-violating behaviors. Participants were asked to indicate the frequency with which they had displayed certain transgressive behaviors in the past on a four-point Likert scale (i.e., never, seldom, sometimes, often). After either a 10-min or a 1-week delay, participants were interviewed about four of these items. However, their answers to two target items had been increased or decreased by two full scale points. Of the 65 participants, 25% were blind to at least one of the two manipulations after a 10-min delay. This percentage increased to 45% with the 1-week interval. Analogous to earlier research on choice blindness which indicates that the effect is prominent minutes after the original choice (Hall, Johansson, & Strandberg, 2012; Hall et al., 2013; Johansson et al., 2005), these findings suggest that choice blindness is not solely memory-based. Indeed, it is quite unlikely that participants forgot how many times they committed certain transgressions. Yet, the fact that the longer interval led to increased blindness implies that forgetting plays a role, although it cannot account for the phenomenon *per se*. Importantly, Sauerland et al. (2013b) study shows that choice blindness also occurs with stimulus material that is of personal significance. In a similar tenor, it has recently been shown that people can be blind to changes of reports concerning their own moral attitudes (Hall et al., 2012, 2013).

Apart from memory failure and stimulus similarity, ambiguity has been proposed as a moderating factor for choice blindness (Merckelbach, Jelicic, & Pieters, 2011). Ambiguity refers to vagueness and imprecision in the environment that allows for multiple interpretations (Sloman, Fernbach, & Hagmayer, 2010). Accordingly, choice blindness occurs whenever decisions pertain to subjective experiences. Evaluative judgments, such as facial attractiveness (Johansson et al., 2008, 2005) or food preferences (Hall et al., 2010), are accompanied by an inherent ambiguity. This is thought to facilitate choice blindness. Following this line of reasoning, one would predict that an objective decision that lacks ambiguity, as for example a facial recognition decision, would make target manipulations much easier to detect.

The application of choice blindness in an eyewitness setting, however, exhibits parallels with the misinformation paradigm (for reviews, see Ayers & Reder, 1998; Loftus, 2005). In this paradigm, witnesses' memory trace is contaminated by introducing misleading information, thus lowering the accuracy of the reports for a witnessed event. When looking at identification decisions, the presentation of the manipulated target could serve as the suggestive information that hinders the memory of the original perpetrator. Thus, one could argue that the present line of research actually portrays different manifestations of the same effect. However, for choice blindness, memory strength does not seem to be the cause, but rather one of the various moderators of the effect. Thus, choice blindness seems to be a broader construct that incorporates suggestion and misleading information.

Another objection could be that blindness for one's facial recognition decision reflects nothing more than people's poor ability to match faces. It is a consistent finding that people have little capacity in matching faces, even when they perform a task that has no memory component (Bruce et al., 2001; Bruce et al., 1999; Jenkins & Burton, 2011). But a degree of consistency with the original decision, albeit incorrect, is to be expected on a re-evaluation. However, choice blindness exemplifies the

Copyright © 2013 John Wiley & Sons, Ltd.

DOI: 10.1002/bsl

failure to detect a discrepancy between intention and outcome, and hence to maintain a reasonable decision consistency over time. Even if propensity to consistency is not to be expected (e.g., Moore & Haggard, 2006), one would predict the same degree of inconsistency across manipulated and non-manipulated decisions. By contrast, after a manipulation, inconsistency increases substantially. Specifically, for facial attractiveness participants display higher consistency (93%) for non-manipulated faces than for manipulated ones (56%), when they were asked to perform a second round of choices (Johansson, Hall, & Chater, 2012). Similarly, Sauerland et al. (2013b) report absolute consistency for their non-manipulated items compared with 17.9% shifted answers for the manipulated items.

Given the apparent deviation of choice blindness from effects such as misinformation or simple forgetting, we aimed to investigate whether choice blindness is relevant for eyewitness facial recognition decisions in a field study. We hypothesized that the short duration and the various distracters inherent to real-life interactions would introduce ambiguity to the task, resulting in high levels of blindness despite immediate confrontation with the manipulated items. Earlier research on change blindness (i.e., difficulty in noticing changes to visual scenes when these are accompanied by some other visual disturbance) supports our assumption. Simons and Levin (1998) demonstrated that 46.6 % of their participants failed to notice changes in the identity of a pedestrian in a real-world interaction. Furthermore, we wanted to test again whether similarity between the originally identified suspect and the manipulated outcome would affect blindness levels. We expected high similarity between targets and manipulated stimuli to introduce an additional source of ambiguity, resulting in increased levels of blindness. In the present, only target-present lineups were used and participants were not given the option to reject the lineup (forced choice). We realize that this procedure imposes limitations on the generalization of our findings to standard lineup situations. Nevertheless, we opted for it as we were, at this stage, mainly interested in determining whether the effect would arise. Note that other phenomena, such as the feedback effect, have been studied using a similar approach (Wells & Bradfield, 1998, 1999).

METHOD

Participants

Seventy-one pedestrians (32 men, 39 women) participated in the study (M_{age} = 36.2 years, SD_{age} = 15.4, age range: 18–64). Participants worked in the private sector (16.9%), public sector (16.9%), and academia (8.5%), were students (26.2%), retired: (4.2%), or worked in other professions (27.3%). Participation was voluntary and no incentives were granted. The study was approved by the university's standing ethical board.

Materials

Lineups

A total of four target-present photo lineups were created. Target persons were four female undergraduate psychology students aged 21–24 years who participated as part of a bachelor course. The lineups consisted of six 8.7 × 7.6 cm frontal face photos;

Copyright © 2013 John Wiley & Sons, Ltd.

DOI: 10.1002/bsl

presented in two rows and three pictures on a 26.0 × 20.0 cm display board. All foils fit the general descriptions of each of the targets, as determined by a pilot study with $N = 32$ mock witnesses (effective sizes, determined as Tredoux's E values ranged from 4.92 to 5.75; Tredoux, 1998; Tredoux, 1999).

Similarity

To test the similarity between the originally identified photo and the photos presented as manipulations, a pilot study ($N = 26$) was performed. In this study, all members of each lineup were paired with each other and rated with regard to their similarity on a seven-point Likert scale (1 = not similar at all, 7 = highly similar). The least and most similar pictures for each lineup member were selected to be presented as the manipulated outcome, creating the two similarity conditions. The averaged means for the low and high similarity pairs were $M = 4.11$ (95% CI [3.9, 4.3]) and $M = 2.98$ (95% CI [2.8, 3.2]), respectively [$t(46) = 10.71$, $p < .001$, $d = 3.16$].

Post-test questionnaire

To examine whether participants had noticed our manipulation and refrained from revealing it, we administered a post-test questionnaire (adjusted from Johansson et al., 2008). First, participants were asked if they had any remarks, if they had encountered any problems, and, if yes, what was the nature of these problems. Participants were then misinformed that the current study employed two conditions: one in which their choice was manipulated and one where this was not the case. Participants had to indicate to which condition they thought they had been assigned. If participants said that they belonged to the former condition, they were asked to indicate how many times and which lineup they had noticed to be manipulated. These answers were counted as retrospective detection.

Design

Similarity (low vs. high) between the target picture and the picture presented as manipulation was varied in a one-factorial between-subjects design. Detection rates were measured concurrently and retrospectively. Concurrent detection refers to detection immediately after the presentation of the manipulated photograph. Retrospective detection additionally includes detection reported in the post-test questionnaire. These definitions of concurrent and retrospective detection are in line with earlier research in the field (Johansson et al., 2008).

Procedure

Data were collected in groups of three, two persons acting as targets and one as experimenter. In total, four different students acted as targets and three as experimenters. The combination of targets was counterbalanced to avoid recognition effects caused by differences in distinctiveness, resulting in six different target combinations. All but one experimenter (who gathered data from three participants) collected data from four participants with each target pair. Since the level of blindness did not differ as a function

Copyright © 2013 John Wiley & Sons, Ltd.

DOI: 10.1002/bsl

of the experimenter [all $x^2(2) \leq 2.67$, $p \geq .263$, Cramer's V ≤ 1.94) or as a result of the different target combinations (all $|t|(69) \leq 0.12$, $p \geq .902$, $d \leq 0.09$), we will not discuss these factors any further.

Data collection took place in the centre of a western European city. Specifically, two targets pretending to be tourists approached a random pedestrian asking for directions. The targets had predetermined roles mainly to help the experimenter to specify later which of the targets the first and second facial recognition tasks were referring to. The primary target would be the first one to talk to the pedestrian and to lead the conversation. The secondary target interacted at a lower level with the pedestrian and was responsible for keeping track of the time. The conversation was scheduled to last between 30 and 60 s. Then, the two targets walked in the direction the pedestrian had indicated. Each target acted both as primary and as secondary target. Equivalence test analysis revealed that recognition accuracy (defined by hits) did not differ as a function of target [all $t(32–35) \leq 1.17$, $p \geq .116$, $\delta = 0.30$].

The experimenter watched the situation from a distance. Around 40 s after the conversation had come to an end, s/he approached the pedestrian and explained that the two young women asking for directions were actually students conducting a study on eyewitnesses' facial recognition decisions. If consent to participate was given, the experimenter presented the lineup of the primary target. Participants were given unlimited time to make a decision, but were not allowed to reject the lineup. Subsequently, they rated how confident they felt about their recognition decision on an 11-point scale ranging from 0% to 100%. The same procedure was followed for the secondary target. Participants were then asked to write down a short description of the event. This was done to introduce a short interval before participants were confronted with their recognition decision. No time limit was imposed for this task. However, the duration of this interval was not measured. Finally, participants were presented with the photograph of the person they had identified in the primary and the secondary lineup and were sequentially asked to motivate their decisions. However, the recognition made for the secondary target was always manipulated, leading participants to end up with a different photograph from the one they had selected. We refrained for manipulating both targets as we first wanted to make sure that we gained participants' trust and attention and that they were familiar with the task. The manipulated photo was either of high or of low similarity with the original choice, depending on the condition the participant was assigned to. Subsequently, participants were again asked to provide a confidence estimate for their earlier recognition. Due to a procedural slip (neglect to present the scale), however, confidence data after the manipulation were unavailable for the first 19 participants, leaving 52 participants for these confidence analyses. Finally, participants were handed the post-test questionnaire. Upon completion participants were thanked and fully debriefed.

RESULTS

An alpha level of .05 was used for all inferential analyses. For comparisons of means, we report Cohen's d and f (Cohen, 1988) as measures of effect size. For 2×2 and for 2×3 contingency tables, φ and Cramer's V are reported, respectively.

Copyright © 2013 John Wiley & Sons, Ltd.

DOI: 10.1002/bsl

Blindness for One's Facial Recognition Decisions

Overall, concurrent detection was 31.0% ($n = 22$). An additional 28.2% ($n = 20$) reported detection in the post-test questionnaire. Thus, in total, 59.2% ($n = 42$) of the participants detected our manipulation retrospectively.

As expected, participants in the low similarity condition were more likely to concurrently detect the manipulation (48.6%) than were participants in the high similarity condition (13.9%) {$x^2(1, N = 71) = 9.98$, $p = .002$, $\varphi = 0.38$, 95% CI [0.17, 0.58]}. Likewise, for retrospective detection, participants in the low similarity condition were more likely to detect the manipulation (74.3%) than participants in the high similarity condition (44.4%) {$x^2(1, N = 71) = 6.54$, $p = .016$, $\varphi = 0.30$, 95% CI [0.09, 0.50]}. Figure 1 displays detection rates in the two similarity conditions.

Recognition Accuracy, Post-decision Confidence and Confidence–accuracy (CA) Relation

Regardless of detection status, participants displayed low levels of recognition accuracy across the two targets ($M = 46.4\%$, SD = 39.0), perhaps due to the field setting and the short exposure to the targets. We were most interested, however, in whether blind participants differed from participants who detected the manipulation (detectors) in terms of recognition accuracy. Recognition performance of blind participants and concurrent detectors did not differ for the manipulated lineup {$x^2(1, N = 71) = 1.46$, $p = .226$, $\varphi = 0.14$, 95% CI [−0.11, 0.37]}. For retrospective detection, detectors were more accurate (45.2%) than blind participants (20.7%) {$x^2(1, N = 71) = 4.53$, $p = .044$, $\varphi = 0.25$, 95% CI [0.16, 0.45]}. No such differences in accuracy between detectors and blind participants were observed for the non-manipulated lineups [all $x^2(1, N = 71) \leq 0.13$, $p \geq .715$, $\varphi \leq 0.04$]. Thus, we cannot conclude an absolute accuracy or memory strength advantage of detectors over blind participants.

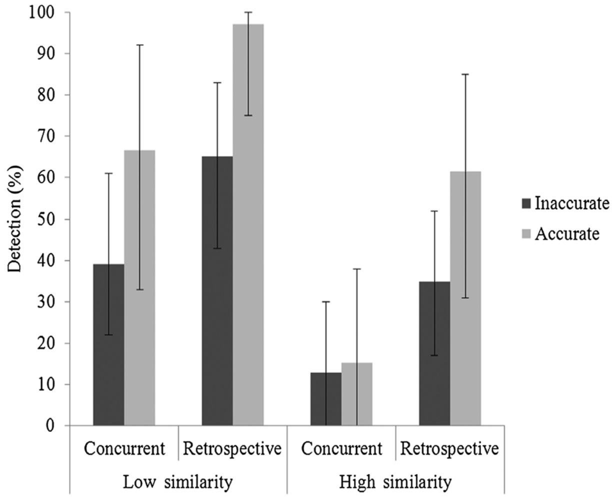

Figure 1. Mean concurrent and retrospective detection rates for accurate and inaccurate participants in the low and high similarity conditions. Error bars represent 95% confidence intervals.

Copyright © 2013 John Wiley & Sons, Ltd.
Behav. Sci. Law 31: 624–636 (2013)
DOI: 10.1002/bsl

As to post-recognition confidence, participants in general (i.e., regardless of detection status) displayed low levels of confidence before ($M = 57.5\%$, SD = 21.8) and after the manipulation ($M = 48.7\%$, SD = 20.6). In a next step, we conducted two mixed ANOVAs with concurrent and retrospective detection (detectors vs. blind participants) as between-subject variables, and manipulation status (manipulated vs. non-manipulated lineups) and test time (before vs. after the manipulation) as within-subject factors. We used this method for a more in-depth analysis of the interplay between manipulation and blindness and confidence. Note that only $n = 52$ participants were included for this analysis.

The results for concurrent and retrospective detection were analogous. The three-way interaction effect among detection, manipulation status and test time was significant, for concurrent [$F(1,51) = 16.26$, $p < .001$, $f = 0.33$] and retrospective detection [$F(1,51) = 5.47$, $p = .023$, $f = 0.35$]. Consequently, we performed separate ANOVAs for detectors and blind participants, with manipulation status and test time as within-subject factors. For concurrently and retrospectively blind participants, there were significant main effects of manipulation status on post-decision confidence [$F(1,35) = 16.83$, $p < .001$, $d = 0.57$; $F(1,19) = 10.77$, $p = .004$, $d = 0.62$]. Confidence rates were lower for manipulated lineups ($M = 44.7\%$, SD = 20.1; $M = 37.5\%$, SD = 24.1) than for non-manipulated ones ($M = 58.2\%$, SD = 24.7; $M = 53.5\%$, SD = 27.1). All other effects were non-significant for blind participants [all $F(1,35) \leq 0.79$, $p \geq .396$, $d \leq 0.04$]. For detectors, there was a significant interaction of manipulation status and test time for concurrent [$F(1,15) = 9.70$, $p = .007$, $f = 0.43$] and retrospective detection [$F(1,32) = 9.14$, $p = .005$, $f = 0.65$]. Analyses of the simple main effects revealed that for the manipulated lineups, confidence estimates given before the manipulation were higher ($M = 55.0\%$, SD = 26.3; $M = 55.3\%$, SD = 23.1) than estimates provided after the manipulation ($M = 21.9\%$, SD = 29.9; $M = 36.9\%$, SD = 28.7). This was true for concurrent [$F(1,15) = 11.09$, $p = .005$, $d = 1.17$] and retrospective detection [$F(1,31) = 9.36$, $p = .005$, $d = 0.71$]. No such effect was found for the non-manipulated lineups for concurrent, $F(1,15) < 0.01$, $p = 1.00$, $d < 0.01$, or retrospective detection [$F(1,31) = 0.04$, $p = .845$, $d = 0.04$].

Finally, we examined whether the CA relationship differed for retrospectively blind participants vs. detectors. Given the small sample size, statistically preferable procedures, such as calibration, could not be performed. Therefore, we used point-biserial correlations to examine the CA relationship. Note that a point-biserial correlation of $r_{pb} = .37$ is considered large, $r_{pb} = .24$ is moderate, and $r_{pb} = .10$ is small (Cohen, 1988). Table 1 sumarizes the obtained correlations for the manipulated and the non-manipulated lineups separately for detectors and blind participants. Here, we will only discuss the effects after the manipulation, which is most interesting for the current study. For detectors, confidence measured after the manipulation was negatively associated with recognition accuracy performance for the manipulated lineup {$r(30) = -.23$, $p = .208$, 95% CI [−.53, .09]}. For the non-manipulated lineup, however, the CA correlation was positive and large {$r(30) = .36$, $p = .040$, 95% CI [.31, .64]}. The difference between the two CA correlations was significant (Fisher's $Z = -2.5$, $p = .012$). For blind participants, the effects were large for both the manipulated lineup {$r(18) = .40$, $p = .083$, 95% CI [−.22, .80]} and the non-manipulated lineup {$r(18) = .53$, $p = .015$, 95% CI [.15, .82]}. The two correlations did not differ (Fisher's $Z = -0.49$, $p = .624$). Overall, the results indicate that our manipulations had a more detrimental effect on the CA relation for detectors than for blind participants.

Copyright © 2013 John Wiley & Sons, Ltd.

DOI: 10.1002/bsl

Table 1. Point-biserial confidence–accuracy (CA) correlations for manipulated and non-manipulated lineups and 95% confidence intervals (CIs)

	Manipulated lineup		Non-manipulated lineup	
	r_{pb}	95% CI	r_{pb}	95% CI
Confidence measured before manipulation				
Detectors ($n = 42$)	.10	−.20–.39	.42**	.15–.65
Blind ($n = 29$)	.54**	.19–.77	.48**	.17–.74
Confidence measured after manipulation				
Detectors ($n = 32$)	−.23	−.53–.09	.36*	.03–.64
Blind ($n = 20$)	.40	−.22–.80	.53*	.15–.82

*$p < 0.05$, **$p < 0.01$.
Note. Due to a procedural error, confidence data after the manipulation were unavailable for 19 participants, leaving 52 participants for these analyses.
Before the manipulation, the differences between the CA correlations of detectors and blind participants were not significant (all Fisher's $Z \leq 1.53$, $p \geq .126$). After the manipulation, the difference between the two CA correlations was significant for detectors (Fisher's $Z = -2.5$, $p = .012$), but not for blind participants (Fisher's $Z = -0.49$, $p = .624$).

DISCUSSION

The current study examined blindness for one's eyewitness facial recognition decisions in a real-life setting. A subsidiary aim was to assess the role of similarity between the original choice and the manipulated outcome as a moderator of the effect. As expected, a large portion of participants failed to notice our manipulation concurrently (68.6%) and retrospectively (40.8%). Furthermore, blindness was facilitated by high similarity between target and manipulated outcome.

In line with our hypothesis, the current study shows that blindness for facial recognition decisions can occur at an alarmingly high level. Importantly, the effect was found for a broad age range and under realistic encoding conditions, enhancing the ecological validity of our findings. Nevertheless, there are some important differences between the present field study and real-world conditions. Specifically, due to ethical considerations, participants did not witness a real crime. Furthermore, they were not allowed to reject the lineup. Also, the intervals between the witnessed event, the facial recognition decision, and the confrontation with the manipulated outcome were much shorter than in a real-life case. Yet the testing conditions were less artificial than those in the typical laboratory experiment. In Sagana, Sauerland, and Merckelbach's (2013; experiments 2a–c) laboratory studies, concurrent detection rates ranged from 66.7% to 75.0%, while retrospective detection varied from 94.4% to 100% when the manipulated outcome was presented immediately after the recognition task. In the current field study, the concurrent detection rate was 31.4% and retrospective detection was 59.2%. These relatively low detection rates might be attributable to the turbulence and dynamics of the real-life setting. Note that when Sagana et al. (2013; experiment 3) inserted a 48-hour interval between the recognition and the presentation of the manipulated outcome, concurrent detection decreased to 31.7% and retrospective detection to 60.6%, approaching the rates obtained in the present study. Thus, although we might not be able to draw direct parallels to real-world lineup identifications, the present study suggests, at a minimum, that blindness for one's facial recognition decisions may be more prominent in real life than in the

Copyright © 2013 John Wiley & Sons, Ltd.

DOI: 10.1002/bsl

laboratory. Furthermore, this convergence speaks to the validity of laboratory simulations and underlines the importance of conducting field research.

Our findings also support the hypothesis that when an element of ambiguity (i.e., everyday circumstances, longer interval) is introduced to a forced-choice recognition task, blindness for one's facial recognition decisions becomes more prominent. Alternatively, one could argue that the high blindness rates found in the current study result from the low accuracy rates for the manipulated lineup. However, earlier studies have demonstrated that people can be choice blind after short intervals (Hall et al., 2012; Johansson et al., 2005) and for decisions related to autobiographical events (Merckelbach et al., 2011; Sauerland et al., 2013b). Additionally, the current detection rates are comparable to prior findings (Sagana et al., 2013), where recognition accuracy exceeded 87%. Thus, weak memory of the original target or simply forgetting the selected face cannot fully account for the effect. Note that our recognition accuracy rates are also in accordance with a recent field study (Horry et al., 2012) that summarized data of over 1,000 real lineups administered in England. Only 39.0% resulted in a suspect identification. It is unclear, however, how many of these lineups actually included the perpetrator.

Conversely, the low accuracy–high blindness effect fits well with the ambiguity explanation for choice blindness (Merckelbach et al., 2011). The poor memory for the original target may have caused vagueness and imprecision while performing the forced-choice recognition task, which in turn may have resulted in an increased feeling of ambiguity in our participants, hence fostering blindness. This is not to imply that accurate participants were immune to the effect. Among accurate participants, 38.5% were concurrently blind (see Figure 1). The ambiguity hypothesis is further supported by the findings regarding the similarity of the chosen and the presented photograph in manipulated lineups. The high similarity between these photographs may have worked as an additional source of ambiguity, leading to increased blindness rates for that condition, compared with the low similarity condition. When combining more than one source of ambiguity, the magnitude of the effect increases dramatically. Consequently, in the high similarity condition, inaccurate participants detected the manipulation only at a 13% rate concurrently and 35% retrospectively.

As for post-decision confidence, blind participants were less confident than detectors for the later to-be-manipulated lineup. This suggests that low confidence may facilitate blindness. Note that the overall low confidence and accuracy rates for both targets may reflect the difficulty of the recognition task. Therefore, it may be the increased ambiguity caused by the difficulty of the task that caused more blindness. Another explanation is that participants who actually wanted to reject the lineup indicated very low confidence, hence decreasing the average confidence rates. The current design only allows us to speculate about the role of confidence as a moderator of blindness. Future studies relying on target-absent and target-present lineups in fully counterbalanced designs could shed more light on this role.

Also related to post-decision confidence is the finding that our manipulation affected its relationship with recognition accuracy. As discussed, the manipulation undermined post-decision confidence. However, blind participants, in contrast to detectors, displayed strong CA relations both before and after the manipulation. The more prominent negative influence of the manipulation for detectors compared with blind participants implies differences in the decision-making processes of the

Copyright © 2013 John Wiley & Sons, Ltd.

DOI: 10.1002/bsl

two groups. A plausible explanation for this pattern could be that detectors felt confused about the target for which they were to provide confidence ratings. Alternatively, the manipulation might have affected participants' memory for the event. This is in line with other studies that found long-lasting manipulation effects. For example, in a follow-up, participants had to make a choice for the same stimuli as in the initial first round. The results showed that participants either chose the alternative that had been presented to them in the manipulated trial (Johansson, Hall, & Gärdenfords, 2011) or adjusted their responses to match the direction of the manipulation (e.g., positive vs. negative; Merckelbach et al., 2011; Sauerland et al., 2013b).

Given the analogies between the choice blindness paradigm and the *post hoc* misinformation paradigm, we assume that explanations suggested for the latter may apply to blindness phenomena as well. This is not to imply that the choice blindness and misinformation are equivalent phenomena, even though underlying mechanisms may be similar. Thus, we suspect an interplay between ambiguous conditions and the recollection of the original memory trace to be responsible for the effect. With certain degrees of ambiguity, the manipulated outcome may perform as an anchor to reconstruct the original memory trace of the participants, resulting in a failure to detect the manipulation. However, a direct test of this hypothesis was beyond the scope of this study. Nevertheless, these findings demonstrate a poor introspective capacity and lack of insight into one's decision-making strategies (Johansson et al., 2008; Sauerland et al., 2013a).

There are several limitations to our study that warrant discussion. First, we applied a forced-choice face recognition task and we relied exclusively on target-present lineups. This procedure deviates from a typical eyewitness identification. Therefore, the generalizability of our findings to real-world investigations is limited. In real lineup procedures, participants would (or at least should) be warned that the perpetrator may or may not be present and should be given the option to reject the lineup (Wells et al., 1998). Nonetheless, a face recognition task is suited to explore the decision-making processes of eyewitnesses. A second limitation, which is related to the forced-choice decisions, concerns the low confidence rates observed in our study. In real cases, it is unlikely that identifications accompanied by equally low confidence ratings would be accepted as evidence in court. Thirdly, no criminal act was staged. We cannot exclude the possibility that high stress levels of an emotionally intense event would have affected the blindness rates. On the one hand, stress could impair the memory of the target (Deffenbacher et al., 2004; Schwabe & Wolf, 2010), leading to increased ambiguity when identifying the target and thus more blindness. On the other hand, stress could facilitate encoding of the witnessed event and facilitate memory for the target (Christianson, 1992; Smeets et al., 2009), leading participants to be less willing to accept the manipulation. Fourthly, only a short retention interval was inserted between the event and the recognition task. In real cases, intervals would generally be longer (Horry et al., 2012; Shermer, Rose, & Hoffman, 2011). Still, a longer interval would cause the memory trace to fade, producing more ambiguity and probably higher rates of blindness (e.g., Sagana et al., 2013). Finally, we did not determine whether the manipulation had a long-lasting effect. This is difficult to implement in a field study. Nevertheless, such results could provide valuable insight regarding the impact of choice blindness on meta-cognition, long after the manipulation took place.

Copyright © 2013 John Wiley & Sons, Ltd.

DOI: 10.1002/bsl

Moving to the practical relevance of the study, our findings have important implications for lineup administrators and law enforcement bodies. Indeed, a lineup administrator could mistakenly write down the wrong identification decision. Given the present findings, one would expect a large proportion of eyewitnesses to fail to detect such a mistake. One may object that the records of identification decisions are unlikely to be incorrect or that witnesses usually sign their names next to the photograph of the identified suspect. Considering, though, how often professionals make procedural errors, we suspect that such a scenario is not all that unrealistic. As a matter of fact, errors made by forensic scientists, such as mistakes in testing procedures (erroneous matches of bullets, footprints and fingerprints) together with police misconduct, are a major source of miscarriages of justice (Saks & Koehler, 2005). Having said that, it is worth mentioning that both the first and the second author, while presenting our research findings at different occasions, were alerted to cases in which a switch in the identification decision had occurred.

Apart from the lineup decisions, the effect of choice blindness can be relevant for identifications pertaining to events that take place many years before they are investigated, such as war crimes. Here, eyewitnesses often have to compare their memory of the perpetrator with a photograph of that person from the distant past and a suspect who is much older (e.g., John Demjanjuk's case; Loftus & Ketcham, 1991). Finally, our findings may inspire new research on eyewitnesses' reports. Specifically, it is known from previous studies that people are relatively poor at recognizing fabricated details in reports of their own autobiographical memories (Barclay & Wellman, 1986; Merckelbach, Wessel, & Horselenberg, 1997). In the present study, we found that eyewitnesses are poor at monitoring the outcome of their identification decisions minutes after their choice. Hence, it seems reasonable to assume that witnesses would fail to detect surreptitious changes in their description of a witnessed crime due to choice blindness. This issue warrants further study and, more generally, it would be interesting to explore how typical misinformation effects and choice blindness are related to each other (e.g., whether choice blindness sets the stage for misinformation effects).

ACKNOWLEDGEMENTS

The authors would like to thank Tim Foitzik, Senne Heugen, Sanne Rooimans, Nicole Schroers, Marlene Stier, Lilith Westwood and Emma Zwaveling for their help in data collection.

REFERENCES

Ayers, M. S., & Reder, L. M. (1998). A theoretical review of the misinformation effect: Predictions from an activation-based memory model. *Psychonomic Bulletin & Review*, *5*, 1–21. doi:10.3758/bf03209454

Barclay, C. R., & Wellman, H. M. (1986). Accuracies and inaccuracies in autobiographical memories. *Journal of Memory and Language*, *25*, 93–103. doi:10.1016/0749-596x(86)90023-9

Bruce, V., Henderson, Z., Greenwood, K., Hancock, P., Burton, A. M., & Miller, P. (1999). Verification of face identities from images captured on video. *Journal of Experimental Psychology. Applied*, *5*, 339–360. doi:10.1037/1076-898X.5.4.339

Bruce, V., Henderson, Z., Newman, C., & Burton, A. M. (2001). Matching identities of familiar and unfamiliar faces caught on CCTV images. *Journal of Experimental Psychology. Applied*, *7*, 207–218. doi:10.1037/1076-898x.7.3.207

Copyright © 2013 John Wiley & Sons, Ltd.

DOI: 10.1002/bsl

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

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

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

Hall, L., Johansson, P., & Strandberg, T. (2012). Lifting the Veil of Morality: Choice Blindness and Attitude Reversals on a Self-Transforming Survey. *PLoS ONE*, *7*, e45457. doi:10.1371/journal.pone.0045457

Hall, L., Johansson, P., Tärning, B., Sikström, S., & Deutgen, T. (2010). Magic at the marketplace: Choice blindness for the taste of jam and the smell of tea. *Cognition*, *117*, 54–61. doi:10.1016/j.cognition.2010.06.010

Hall, L., Strandberg, T., Pärnamets, P., Lind, A., Tärning, B., & Johansson, P. (2013). How the polls can be both spot on and dead wrong: Using choice blindness to shift political attitudes and voter intentions. *PLoS ONE*, *8*, e60554. doi:10.1371/journal.pone.0060554

Horry, R., Memon, A., Wright, D. B., & Milne, R. (2012). Predictors of eyewitness identification decisions from video lineups in England: A field study. *Law and Human Behavior*, *36*, 257–265. doi:10.1037/h0093959

Jenkins, R., & Burton, A. M. (2011). Stable face representations. *Philosophical Transactions of the Royal Society B*, *366*, 1671–1683. doi:10.1098/rstb.2010.0379

Johansson, P., Hall, L., & Chater, N. (2012). Preference change through choice. In R. J. Dolan & T. Sharot (Eds.), Neuroscience of preference and choice: Cognitive and neural mechanisms (pp. 121–141). New York: Academic Press.

Johansson, P., Hall, L., & Gärdenfords, P. (2011). Choice blindness and the non-unitary nature of the human mind. *Behavioral and Brain Sciences*, *34*, 28–29. doi:10.1017/SO140525X10002591

Johansson, P., Hall, L., & Sikström, S. (2008). From change blindness to choice blindness. *Psychologia: An International Journal of Psychology in the Orient*, *51*, 142–155. doi:10.2117/psysoc.2008.142

Johansson, P., Hall, L., Sikström, S., & Olsson, A. (2005). Failure to detect mismatches between intention and outcome in a simple decision task. *Science*, *310*, 116–119. doi:10.1126/science.1111709

Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory*, *12*, 361–366. doi:10.1101/lm.94705

Loftus, E., & Ketcham, K. (1991). Ivan the Terrible: John Demjanjuk. In Witness for the defense: The accused, the eyewitness, and the expert who puts memory on trial. (pp. 210–242). New York, NY US: St Martin's Press.

Merckelbach, H., Jelicic, M., & Pieters, M. (2011). Misinformation increases symptom reporting - a test - retest experiment. *Journal of the Royal Society of Medicine Short Reports*, doi:10.1258/shorts.2011.011062

Merckelbach, H., Wessel, I., & Horselenberg, R. (1997). The accuracy of autobiographical memory: A replication of Barclay & Wellman (1986). *Behavioural and Cognitive Psychotherapy*, *25*, 103–111. doi:10.1017/s1352465800018336

Moore, J., & Haggard, P. (2006). Commentary on 'How something can be said about telling more than we can know: On choice blindness and introspection'. *Consciousness and Cognition: An International Journal*, *15*, 693–696. doi:10.1016/j.concog.2006.09.003

Sagana, A., Sauerland, M., & Merckelbach, H. (2013). "This is the person you selected": About being blind for one's own eyewitness identification decision. Manuscript submitted for publication.

Saks, M. J., & Koehler, J. J. (2005). The coming paradigm shift in forensic identification science. *Science*, *309*, 892–895. doi:10.1126/science.1111565

Sauerland, M., Sagana, A., & Otgaar, H. (2012). Theoretical and legal issues related to choice blindness for voices. *Legal and Criminological Psychology*, Advance online publication. doi:10.1111/j.2044-8333.2012.02049.x

Sauerland, M., Sagana, A., Siegmann, K., Merckelbach, H., & Jenkins, R. (2013a). These two are different. Yes, they're the same: Choice blindness for facial identity. Manuscript submitted for publication.

Sauerland, M., Schell, J., Collaris, J., Reimer, N., Schneider, M., & Merckelbach, H. (2013b). "Yes, I have sometimes stolen bikes": Blindness for norm-violating behaviors and implications for suspect interrogations. *Behavioral Sciences & the Law*, *31*, 239–255. doi:10.1002/bsl.2063

Schwabe, L., & Wolf, O. T. (2010). Learning under stress impairs memory formation. *Neurobiology of Learning and Memory*, *93*, 183–188. doi:http://dx.doi.org/10.1016/j.nlm.2009.09.009

Shermer, L. O. N., Rose, K. C., & Hoffman, A. (2011). Perceptions and credibility: Understanding the nuances of eyewitness testimony. *Journal of Contemporary Criminal Justice*, *27*, 183–203. doi:10.1177/1043986211405886

Simons, D. J., & Levin, D. T. (1998). Failure to detect changes to people during a real-world interaction. *Psychonomic Bulletin & Review*, *5*, 644–649.

Sloman, S. A., Fernbach, P. M., & Hagmayer, Y. (2010). Self-deception requires vagueness. *Cognition*, *115*, 268–281. doi:http://dx.doi.org/10.1016/j.cognition.2009.12.017

Smeets, T., Wolf, O. T., Giesbrecht, T., Sijstermans, K., Telgen, S., & Joëls, M. (2009). Stress selectively and lastingly promotes learning of context-related high arousing information. *Psychoneuroendocrinology*, 1152–1161. doi:10.1016/j.psyneuen.2009.03.001

Copyright © 2013 John Wiley & Sons, Ltd.

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

Tredoux, C. G. (1999). Statistical considerations when determining measures of lineup size and lineup bias. *Applied Cognitive Psychology*, *13*, 9–26. doi:10.1002/(SICI)1099-0720

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

Wells, G. L., & Bradfield, A. L. (1999). Measuring the goodness of lineups: Parameter estimation, question effects, and limits to the mock witness paradigm. *Applied Cognitive Psychology*, *13*, 27–39. doi:0888-4080/99/SI0S27-13

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

Copyright © 2013 John Wiley & Sons, Ltd.

DOI: 10.1002/bsl