'This Is the Person You Selected': Eyewitnesses' Blindness for Their Own Facial Recognition Decisions

ANNA SAGANA*, MELANIE SAUERLAND and HARALD MERCKELBACH

Forensic Psychology Section, Department of Clinical Psychological Science, Maastricht University, Maastricht, The Netherlands

Summary: The aim of the current research was to identify conditions under which choice blindness in facial recognition decisions occurs. In five experiments, participants watched four mock-crime videos and made choices that were either evaluative (Experiment 1) or absolute in nature (Experiments 2a–c and 3). When participants were subsequently asked to motivate their choice, they were sometimes presented with choices they had not made. For evaluative decisions, concurrent (27%) and retrospective blindness rates (21%) were relatively low compared with previous studies. For absolute decisions, choice-blindness rates varied, depending on when exposure to the manipulated outcome took place (immediate: concurrent 32–35%, retrospective 0–6% [Experiments 2a–c]; 48 hours' delay: concurrent 68%, retrospective 39% [Experiment 3]). We argue that blindness for facial recognition decisions is more likely for evaluative decisions and for longer intervals between decision and manipulation and also for conditions of increased task complexity, which we interpret in terms of ambiguity. Copyright © 2014 John Wiley & Sons, Ltd.

INTRODUCTION

In many western countries, eyewitnesses first identify the suspect from a lineup at the police station and later confirm their decision during a court proceeding. Thus, at trial, eyewitnesses have to testify in support of their earlier decision. Consider though the situation where the lineup administrator (intentionally or unintentionally) wrote down a different decision than the one made by the witness. Would eyewitnesses notice a change in their identification decision when they are later interviewed about it during, for example, a court session? The case of Bernard Maughan (Wolchover, n.d.) is an example of such a case. Specifically, the eyewitness in this case made an identification by saying 'I think it's number six' in the presence of his solicitor. However, the administrator wrote down 'I think it's number seven' (who happened to be the suspect) and read back the utterance to the witness. At that point, neither the eyewitness nor the solicitor demurred. As a result, the official record mentioned a member other than the actual identified lineup member. Subsequently, the defendant was charged and released on bail. It was only on the appeal, 2 years later, that the solicitor of the defendant spotted the miscommunication while reviewing the identification tapes. This appears to be an incident of *choice blindness* in the setting of eyewitness identification.

Choice blindness refers to the difficulty people have in detecting manipulations of a choice they previously made and their tendency to confabulate introspective arguments for the very choice they did not make (Johansson, Hall, Sikström, & Olsson, 2005). In a first demonstration of the phenomenon, Johansson et al. (2005) asked participants to choose which of two female faces they found more attractive. After they had made a decision, participants were presented with their choice and were asked to explain the reasons behind their decision. Using a magic card trick, three of the 15 trials were manipulated such that participants actually ended up with the *non*-chosen face. Participants

were blind to 308 (87%) of the 354 manipulated pairs. However, when participants were given a description of the procedure and were asked whether they would have noticed such choice manipulations, the vast majority of them were confident that they would have done so. Johansson et al. (2005) termed this meta-cognitive bias *choice-blindness blindness*.

Inspired by these findings Sagana, Sauerland, and Merckelbach (2013) examined choice blindness for eyewitnesses' facial recognition decisions. Specifically, pedestrians in a European city were engaged in a conversation with two experimenters who pretended to be tourists. Shortly thereafter, the pedestrians were asked to identify the two experimenters from separate simultaneous photo lineups using a forced-choice recognition decision format. Subsequently, they were confronted with their selection and asked to motivate their decision. However, for the second target, the chosen lineup member was swapped with a previously *non*-identified member. Interestingly, 68.3% of the pedestrians failed to immediately report that they noticed the change (i.e., *concurrently*). Even after the end of the experiment in a post-test questionnaire (i.e., *retrospectively*), 39.8% of the participants were blind to this identity manipulation. Additionally, Sagana et al. (2013) reported that participants who made an accurate lineup decision were more likely to retrospectively detect the manipulation than participants who made an erroneous recognition decision. Hence, superior recognition performance was associated with higher detection rates.

Although blindness phenomena have been demonstrated repeatedly, its border conditions have been an underexplored facet. Investigating these conditions is important because it may shed new light on factors that affect choice blindness in settings other than preference, such as the eyewitness identification setting. Drawing from the negative association between recognition performance and blindness, the phenomenon may be conceptualized as a manifestation of weak memory. Indeed, memory deterioration for one's own decisions may foster blindness effects, as demonstrated by Sauerland, Schell, et al. (2013). Nevertheless, other studies showed that participants can be blind to changes in the outcome of decisions that they themselves made only a few minutes before the presentation of the manipulated outcome

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

(Hall, Johansson, & Strandberg, 2012; Hall et al., 2013; Johansson et al., 2005). Additionally, research on memory for choices suggests that people have an enhanced memory for a preferred option compared with a rejected one, even when the grounds for this decision are unclear (Dellarosa & Bourne, 1984; Mather, Shafir, & Johnson, 2000). Thus, it appears that memory decay cannot fully explain choice blindness.

Poor processing of the original targets (Johansson et al., 2005) or the manipulated items (Sauerland, Sagana, & Otgaar, 2013) has also been proven to be an insufficient determinant of choice blindness. Furthermore, participants' compliance (Johansson, Hall, & Sikström, 2008; Johansson et al., 2005; Sauerland, Sagana, et al., 2013), suggestibility (Merckelbach, Jelicic, & Pieters, 2011; Sauerland, Schell, et al., 2013), or tendency to react in socially desirable ways (Merckelbach et al., 2011; Sauerland, Sagana, et al., 2013) does not seem to modulate the phenomenon. The degree of similarity between the original target and manipulated item has sometimes (Hall, Johansson, Tärning, Sikström, & Deutgen, 2010; Sagana et al., 2013; Sauerland, Schell, et al., 2013), but not always (Johansson et al., 2005), been found to act as a moderating factor.

From another point of view, choice blindness is related to the constructive nature of preferences and self-persuasion (Johansson, Hall, & Gardenfors, 2011). Specifically, Johansson et al. (2011) argued that people have limited access to the reasons for their actions (also Nisbett & Wilson, 1977). As a consequence, people can come to like what they are told or suggested to like. Endorsing choices suggested by others may generate a degree of cognitive dissonance, which people want to overcome (Festinger, 1962; Henkel & Mather, 2007). Hence, they deliberate reasons to convince themselves about their assumed choice. Once the changed outcome has publicly been endorsed, people need to convince themselves that they, indeed, prefer that outcome. Relatedly, after people have accepted the manipulated choice, they may distort their memory for the selected outcome, through self-deception, in order to reduce dissonance.

Taken together, the literature on choice blindness shows that people have poor insight into their decisions or flexible preferences that are easily influenced by external cues. Either way, the factors that affect the occurrence of the phenomenon remain unclear. Nonetheless, the relevance to legal settings and the potential devastating consequences of the phenomenon for (innocent) defendants are evident. With these considerations in mind, the experiments in this paper constitute a first exploration of the borderline conditions under which blindness for facial recognition decisions materializes.

Sagana et al. (2013) demonstrated the relevance of choice blindness for eyewitness decisions. However, the employed field study approach allows little control over the specific conditions that foster blindness phenomena in the eyewitness setting. The difficulty of the matter is demonstrated in the reduced blindness rates for recognition decisions compared with previous findings using visual stimuli (Johansson et al., 2005, 2008). One reason for this discrepancy could be the use of absolute versus evaluative decisions. Eyewitnesses base—or should base—an identification decision on their memory of the individual rather than on evaluative judgements. This is not to imply that eyewitnesses do not make evaluative decisions in real life. In fact, in an archival analysis of persons' descriptions, Sporer (as cited by Tuckey & Brewer, 2003) found that about 5% of the descriptors referenced perceived personality characteristics. In principle though, eyewitness identifications in an experimental setting can be tested against a ground truth, and their outcome is absolute (i.e., correct or incorrect). Given the objective nature of identification decisions, one would intuitively expect that manipulations would readily be detected, implying little room for choice blindness. This is very much unlike for evaluative judgments about faces. A person may judge someone more or less sympathetic depending on their mood or on the context (Loewenstein & Small, 2007). Accordingly, uncertainty about the evaluative decision made at an earlier point may easily occur.

If a missing evaluative component is responsible for the lower blindness rates of Sagana et al. (2013), it is reasonable to anticipate increased levels of blindness when an evaluative element is introduced to an eyewitness recognition task. To test this hypothesis, we employed a typical eyewitness recognition paradigm with the difference that participants, instead of selecting the perpetrator from a lineup, were asked to indicate which of the actors seen in a mock crime they found most sympathetic (Experiment 1). Additionally, to explore factors that might be critical for the occurrence of choice blindness, we performed a series of three studies (Experiments 2a–c) where absolute decisions were required. Hence, in Experiments 2a–c, we employed a procedure similar to that in Experiment 1, with the difference that participants made a facial recognition decision rather than sympathy evaluations. We anticipated lower blindness rates for the absolute compared with evaluative decisions. Experiments 2a–c also differed in the number of manipulations, mask duration (interval between 50 and 500 milliseconds), and sample characteristics in order to further investigate conditions that facilitate or inhibit blindness. Finally, in a fifth study (Experiment 3), we examined whether increased retention intervals between the recognition decision and the presentation of the manipulated outcome would affect blindness rates. We considered this important as it is unlikely that the identification procedure and the confrontation with one's decision at trial take place on the same day (Shermer, Rose, & Hoffman, 2011). Thus, Experiment 3 employed a procedure similar to that in Experiment 2a with the exception that the manipulated outcome was presented 48 hours after the recognition task, which is a delay that adds to the ecological relevance of the study.

EXPERIMENT 1

Method

Participants

Thirty-four participants (19 men and 15 women) of various nationalities took part in the study ($M_{age} = 25.2$ years, $SD_{age} = 5.2$, range: 19–45). The majority (61%) were Maastricht University students studying different majors (law: 14.7%, psychology: 11.8%, and mental health: 11.8%), whereas 38.2% were employees (private sector: 23.5%, public sector:

2.9%, and other: 11.8%). All reported studies were approved by the standing ethical board of the Faculty of Psychology and Neuroscience of Maastricht University.

Materials

Stimulus films

Four stimulus video fragments showing minor offenses were displayed on a computer screen. The mean duration was 187 seconds (duration range: 160–214 seconds). Two of the fragments showed the theft of a wallet at a university cafeteria (Cafeteria 1 and Cafeteria 2). The third fragment depicted the theft of a wallet in a bar (Bar), and the fourth one was about an assault attempt at a bus stop (Bus). In each video fragment, there were four actors: a perpetrator, a victim, and two bystanders. Actors were different for each film. The gender of perpetrators and victims was counterbalanced across videos. In two videos (Cafeteria 1 and Bus), the perpetrator was a man and the victim a woman, whereas the reverse was true for the other two videos (Cafeteria 2 and Bar). Each target was in sight for a minimum of 76 seconds, with close-ups of 2–9 seconds. All targets were shown from frontal and side views.

Post-test questionnaire

To examine whether participants had noticed our manipulations but refrained from mentioning it, we administered a post-test questionnaire that was adjusted from Johansson et al. (2008). Participants were first asked whether they had noticed anything strange during the experiment, and when they responded affirmatively, they were invited to provide details. If they reported they noticed the manipulation, they were coded as retrospective detectors. Next, participants were given a description of an imaginary experiment in which some identifications were manipulated in such a way that the participant would end up with a photo that she or he did not choose. To test choice-blindness blindness (i.e., the meta-cognitive bias; Johansson et al., 2005), participants first had to indicate if they thought that they would notice such a change. Second, they were asked if they believed that we had carried out such a manipulation in the current experiment. If participants responded affirmatively to this question, they were again counted as retrospective detectors.

Design

In Experiment 1, we manipulated the preferred target in one of the four video fragments. Specifically, the decisions of the Bus or Cafeteria 2 fragments were manipulated. Detection rates were measured both *concurrently* and *retrospectively*. Concurrent detection refers to participants who immediately noticed the manipulation after it had taken place. Retrospective detection, next to concurrent detection, *additionally* includes those participants who reported in the post-test questionnaire that they had noticed the change. In accordance with Johansson et al. (2008), any positive answer to the questions about noticing anything strange or noticing a manipulation taking place during the experiment was counted as retrospective detection.

Procedure

All parts of the experiment were presented on a 20-in. computer screen at a resolution of 1024×768 pixels using DMDX Display Software (Forster & Forster, 2003). Participants were naïve to the actual purpose of the study and were tested individually.

After signing the informed consent form, participants were told that they would watch four video fragments and would act as eyewitnesses. Then, the first video was presented. After the end of the fragment, participants were shown a slide with photos of the actors they had seen during the video. Their task was to select which of these actors they found most sympathetic. Although the procedure is very similar to that in an identification experiment, participants' task was evaluative in nature.

For the three video fragments (Cafeteria 1, Cafeteria 2, and Bar) in which three actors were of one gender and only one actor of the other gender, we presented only the three actors of the same gender (in one row of three pictures). This was carried out to prevent an increased detection rate simply due to noticing the gender difference. For the fourth video, in which two male and two female actors were shown, all four actors were presented in two rows of two pictures.

After participants had made their choice, a 50-millisecond random pattern (i.e., mask) was displayed, followed by the presentation of the chosen actor. Now participants were asked to justify their decision. The same procedure was followed for all four videos that were presented either in the order Cafeteria 1, Bar, Cafeteria 2, and Bus or in the order Cafeteria 1, Bar, Bus, and Cafeteria 2. However, the decision for the video fragment presented last (either Bus or Cafeteria 2) was manipulated, and participants were confronted with a predetermined photo they had actually *not* selected. If participants' justifications included a comment indicating that the displayed photo did not correspond to their choice or that the program had made a 'mistake', they were classified as concurrent detectors. Detectors frequently also verbally informed the experimenter about the change. No significant differences in detection emerged between the two orders for concurrent, $x^2(1, N = 34) = 1.89$, $p = .25$, $phi = 0.24$, or retrospective detection, $x^2(1, N = 34) = 2.10$, $p = .21$, $phi = 0.25$. Therefore, we will not discuss this factor further. At the end of the experiment, participants were given the post-test questionnaire and were fully debriefed.

Results and discussion

The concurrent detection rate was 73.5% ($n = 25$), and retrospective detection was 79.4% ($n = 27$). On the whole, 20.6% ($n = 7$) of the participants were completely choice blind. Of these choice-blind participants, 85.7% ($n = 6$) said in the post-test questionnaire that they would be able to detect the manipulation. Thus, this group evidently exhibited choice-blindness blindness.

The observed concurrent and retrospective choice-blindness rates of the present study are relatively low compared with previous choice-blindness studies for visual stimuli (Johansson et al., 2005, 2008). This may be due to the use of a video fragment with a specific plot. Specifically, it seemed that participants based their preference on the role of the target

and their empathic feelings toward the victims and the bystanders. This is illustrated by the low preference rate for the perpetrators (3 of the 34 participants) compared with the victims and bystanders (91.2%). Thus, because their decision was often based on a clear category (e.g., victims), participants may have found it relatively easy to detect the manipulated outcome.

Most importantly and in contrast to our hypothesis, the choice-blindness rates observed in Experiment 1 are lower compared with those reported by Sagana et al. (2013; 31.7–60.2%). These findings do *not* speak to the idea that blindness rates depend on the decision type required (i.e., absolute vs. evaluative). The diversion in blindness rates across the two studies could be due to differences in methodology, as the present is a laboratory study, whereas Sagana et al. (2013) performed a field study. That is, the various distracters that are inherent to real-life interactions may have increased the complexity of the task, resulting in raised blindness rates (Sagana et al., 2013). To test this notion, we examined whether choice blindness would be less prominent if participants performed a procedure similar to Experiment 1, however with participants performing absolute facial recognition decision rather than evaluative decisions. To this end, we performed a series of three studies (Experiments 2a–c). In addition to varying decision type (compared with Experiment 1), Experiments 2a–c differed in the number of manipulations, mask duration, and sample characteristics in order to test other conditions that could confine blindness for facial recognition decisions.

EXPERIMENTS 2A–C

Method

In Experiments 2a–c, participants made forced-choice recognition decisions from target-present lineups. In Session 1, participants watched four video fragments and made facial recognition decision for each of the four targets depicted in the films (i.e., 16 targets in total). After each decision, a mask was presented, and subsequently, the photo of the selected lineup member was displayed on the screen. However, some decisions (either two [Experiment 2a] or four [Experiments 2b and c]) were manipulated and replaced with a non-selected lineup member. Twenty-four hours later, participants returned to the lab for a second session and were asked to identify all perpetrators and victims again (i.e., eight targets in total). Session 2 was introduced to investigate whether the manipulations had an impact on participants' future decisions.

For Experiments 2a–c, the same stimulus films, lineups, testing materials, and procedure were used. In Experiments 2a and b, we tested whether blindness for facial recognition decisions was affected by the number of manipulations and by mask duration. Specifically, we decreased the number of manipulations from four in Experiment 2a to two in Experiment 2b and increased mask duration from 50 to 500 milliseconds. This was to test whether a large number of manipulations as well as the short mask duration would alarm participants and hence increase detection.

In Experiment 2c, we investigated whether our results were affected by our heavy reliance on bachelor psychology students as participants. Students may just have waited for some kind of manipulation to occur while they were tested, or they might have read or heard about the choice-blindness effect during the course of their studies. Relatedly, the use of undergraduate student samples has been vigorously criticized among others as malleable to social influences (e.g., Sears, 1986). To rule out this possibility, Experiment 2c relied on participants who were currently not active in or had never been involved in academia. The procedure of Experiment 2c was analogous to that of Experiment 2b, with the exception that there was only one session. This seemed appropriate because we were mostly interested in the blindness rates in Session 1. Additionally, we asked participants in the post-test questionnaire how many manipulations they had detected in Experiment 2c.

Participants

In Experiments 2a and b, participants were 18 and 19 (26 women) students of Maastricht University, respectively ($M_{age} = 20.9$ years, $SD_{age} = 1.6$, range: 19–25). The majority was psychology students (73.0%), and one was a mental health student (2.7%), whereas the majors of the remaining students (24.3%) were not specified. In Experiment 2c, 20 (11 women) participants ($M_{age} = 41.5$ years, $SD_{age} = 16.2$, age range: 19–64) had varying professional backgrounds (private or public sector: 35%, freelance: 20%, household: 20%, and temporary employment: 25%). Participation was voluntary. Student participants received course credit in return, whereas for non-academics, no monetary or other incentive was granted. All participants were naïve to the purpose of the study and were tested individually.

Materials

Lineups
A total of 16 simultaneous, target-present photo lineups were created. To that end, head-and-shoulder photos were selected to match the description of the respective targets. Each lineup was presented on a 2×3 array format. Thus, each lineup included one target (i.e., perpetrator, victim, or bystander) and five distracters. The size of the photos, as presented on the computer screen, was $9.0\,cm \times 10.0\,cm$. Only target-present lineups were administered, and participants were not allowed to make rejections. We opted for this procedure because we wanted to test for the existence of choice blindness for lineup selections (rather than rejections). Note that other phenomena such as the feedback effect (Wells & Bradfield, 1998, 1999) or the own-race bias effect (for a complete review, see Meissner & Brigham, 2001) were originally studied using a similar approach.

Design
In Experiments 2a and b, we manipulated either four or two of the 16 lineups. Table 1 summarizes the manipulations per experiment. Detection rates were measured both concurrently and retrospectively.

Table 1. Overview of the order of film and lineup presentation and performed manipulations for Sessions 1 and 2 across all experiments

Experiment 1	Experiment 2a			Experiment 2b			Experiment 2c		Experiment 3	
Film order	Film order	Manipulated lineup	Presented lineup	Film order	Manipulated lineup	Presented lineup	Film order	Manipulated lineup	Film order	Presented and manipulated lineup
Session 1										
Cafeteria 1	Cafeteria 2	Bystander 2		Cafeteria 2			Cafeteria 2		Cafeteria 1	Thief[c]
Bar	Bus	Victim		Bus	Thief		Bus	Thief	Bar	Thief[d]
Bus[a]	Bar	Thief		Bar			Bar		Bus	Thief[c]
Cafeteria 2[b]	Cafeteria 1	Bystander 1		Cafeteria 1	Bystander 1		Cafeteria 1	Bystander 1	Cafeteria 2	Thief[d]
Session 2										
	Cafeteria 2		Thief / Victim	Cafeteria 2		Thief / Victim				
	Bus		Thief / Victim	Bus		Thief / Victim				
	Bar		Thief / Victim	Bar		Thief / Victim				
	Cafeteria 1		Thief / Victim	Cafeteria 1		Thief / Victim				

Note: For all films, lineups were presented in the order Thief, Victim, Bystander 1, and Bystander 2. In Experiment 1, participants made evaluative judgements of sympathy, comparing all actors appearing in one film. In Experiments 1 and 2c, there was only one session. In Experiments 1 and 3, different manipulations were carried out depending on the assigned experimental condition. In Experiments 2a and b, manipulations were carried out in Session 1. In Experiment 3, manipulations were carried out in Session 2.

[a]Manipulated in Condition 1.
[b]Manipulated in Condition 2.
[c]Manipulated in Condition 1.
[d]Manipulated in Condition 2

Procedure

The experimental procedure involved two sessions for Experiments 2a and b, but only one session for Experiment 2c. In Session 1, participants, after signing the informed consent form, were told that they would watch four video fragments and act as eyewitnesses. Then, the first video was presented. Participants were asked to identify all four targets (i.e., perpetrator, victim, and two bystanders) involved in separate simultaneous photo lineups immediately after the end of the first video. Participants were forced to make a choice. Participants selected a lineup member by pressing one of the numbers indicated underneath each of the pictures in the lineup. Participants had unlimited time to make their choice. After a decision was made, a mask was presented for either 50 milliseconds (Experiment 2a) or 500 milliseconds (Experiment 2b; Table 2). Subsequently, the photo of the selected lineup member was displayed on the screen, and participants were asked to motivate their choice on a paper sheet. Once all four identifications were completed, the next video was presented. The same procedure was followed for all videos and targets. However, for four or two of the 16 lineups, participants were confronted with a photo they had actually *not* selected. Specifically, participants were presented with the photo presented dextrorotatory second next to the original choice regardless the accuracy of the identification. For example, if participants selected number 1, they were presented with number 3, and if they selected number 6, number 2 was presented. The exact order of the video fragments, the lineup order presentation, and the manipulations can be found in Table 1.

Session 2 (Experiments 2a and b) took place 24 hours later. The aim of Session 2 was to investigate whether the manipulations would have an impact on participants' future decisions. To that end, we asked the participants to identify all perpetrators and victims again (eight lineups in total). The cover story informed participants that we were interested in the effects of different lineup presentation methods. Accordingly, the lineup photos in Session 2 were arranged in a single row rather than in a 2 × 3 matrix (as was the case in Session 1). To avoid participants reiterating the number they had chosen in Session 1, the order of the lineup members also differed between the two sessions. Given that only the perpetrators and the victims were presented in Session 2, there were only two previously manipulated lineups that carried over to the second session for Experiment 2a and only one for Experiment 2b (Table 1). This was performed to avoid confusion and cognitive overload in the participants. Because peripheral targets are associated with reduced

identification accuracy (Brown, 2003; Wessel, van der Kooy, & Merckelbach, 2000), we decided not to present any bystanders in Session 2. To make sure that participants were aware of which video fragment pertained to each lineup, they were presented with a sketch and a brief description of the plot before each lineup was shown. After completion of all eight identifications, participants were given the post-test questionnaire. Finally, participants were thanked for their participation and fully debriefed.

Results and discussion

Blindness for facial recognition decisions

Concurrent detection across Experiments 2a–c varied between 65.0% and 68.9%. For retrospective detection, percentages ranged between 94.4% and 100%. Thus, overall, more than 30% of the participants were concurrently blind, whereas retrospective blindness was largely absent. Note that we did not ask participants in Experiments 2a and b *how many* manipulations they had noticed. Thus, we were unable to determine how many participants detected only some of the manipulations and how many detected all manipulations. For Experiment 2c, concurrent detection for the first manipulation was at 80.0% ($n = 16$) and at 50.0% ($n = 10$) for the second one. Retrospective detection rate was for both manipulations 100% ($N = 20$). Table 2 summarizes the overall concurrent and retrospective detection rates across all experiments.

Recognition accuracy and identification shifts across sessions (Experiments 2a and b)

To explore the possibility that our findings were mainly due to memory strength attributable to the attention that participants paid to the stimulus videos, we examined the overall recognition accuracy in the two sessions. In Experiment 2a, recognition accuracy across all targets was high, with an average of $M = 76.9\%$ ($SD = 13.3$) in Session 1 (across 16 targets) and an average of $M = 90.8\%$ ($SD = 12.3$) in Session 2 (across eight targets). In Experiment 2b, accuracy across targets was somewhat lower, with an average of $M = 71.0\%$ ($SD = 13.7$) in Session 1 and $M = 84.0\%$ ($SD = 14.5$) in Session 2. These high rates are not surprising, given that participants had to make a choice (forced choice) and that all lineups were presented in the target-present mode (Steblay, 1997). The higher accuracy rate for Session 2 can be explained by the exclusive presentation of perpetrator and victim lineups in that session, whereas Session 1 also involved lineups for bystanders. Note that past research has

Table 2. Overview of the experimental design and overall percentage of concurrent and retrospective detections across all experiments

	Type of judgment	Performed choices	Performed manipulations	Retention interval	Sample	Overall detection rate (%)	
						Concurrent	Retrospective
Experiment 1	Evaluative	4	1	50 millisecond	Students	73.5	79.4
Experiment 2a	Absolute	16	4	50 millisecond	Students	66.7	94.4
Experiment 2b	Absolute	16	2	500 millisecond	Students	68.4	100
Experiment 2c	Absolute	16	2	500 millisecond	Non-academics	65.0	100
Experiment 3	Absolute	16	2	48 hours	Students	31.7	60.6

indicated significant differences in the identification performance regarding central (i.e., perpetrators) versus peripheral (i.e., bystanders) targets (Brown, 2003).

Furthermore, we tested whether the manipulation caused identification shifts from Session 1 (i.e., before the manipulation) to Session 2 (i.e., after the manipulation). For Experiment 2a, the average recognition accuracy for the two later-to-be-manipulated lineups for both sessions was $M = 86.1\%$ ($SD = 23.0$). For Experiment 2b, recognition accuracy was $M = 100\%$ for both sessions. Did participants alter their original choice at Session 2 as a result of the manipulation? For Experiment 2a, only two participants changed their identification decisions. However, none of the observed decision changes were consistent with the manipulated outcome. In Experiment 2b, there were no switches in identification decisions. Apparently, our manipulations did not affect participants' identification performance in Session 2.

In Experiments 2a–c, a substantial proportion of participants concurrently failed to detect the manipulation; however, we found next to no retrospective blindness. Thus, in contrast to what previous studies (Johansson et al., 2008; Sagana et al., 2013; Sauerland, Schell, et al., 2013) have found, blindness for eyewitness recognition decisions was limited to a concurrent level in Experiments 2a–c. Across the three experiments, we tried to rule out different explanations for the absence of retrospective blindness. In Experiment 2b, we showed that absence of retrospective blindness was not due to a relatively high proportion of manipulations or short mask duration that could have made our participants suspicious. Furthermore, the findings of Experiment 2c argue against the idea that the absence of retrospective choice blindness in Experiments 2a and 2b was due to sample characteristics such as prior knowledge, heightened alertness, or both.

Some may argue the absence of choice blindness in retrospective eyewitness judgments might be an artifact of the suggestive nature of the post-test questionnaire. However, this does not explain the vast differences in retrospective blindness found in our studies and previous choice-blindness studies (e.g., Johansson et al., 2005; Sagana et al., 2013; Sauerland, Schell, et al., 2013). Thus, in principle, the absence of choice blindness in the current experiments is good news as it suggests considerable preservation of the recognition decisions. From these results, we would be happy to accept that the phenomenon plays no (or little) role in this domain. However, it is unlikely that the identification procedure and the confrontation with one's decision at trial take place on the same day. In fact, even in 'speedy trial acts' as applied by the US judicial system, trials usually begin at least 90–120 days after the arrest of the suspect (Shermer et al., 2011). Therefore, we wondered what the extent of blindness would be for recognition decisions that were taken in the recent past. Will witnesses notice the identity change if they are confronted with the manipulated outcome not milliseconds (i.e., short-term memory) but hours or days later? Given the findings of Experiments 2a–c, we should not expect blindness phenomena. Conversely, Sauerland, Schell, et al. (2013) revealed that people can be blind for changes in decisions related to long-term autobiographical memory. Therefore, there may be a different cognitive process involved in long-term memory that could facilitate choice blindness.

To test these assumptions, we conducted a third experiment. In Experiment 3, we introduced a 48-hour interval between choosing a lineup member and being confronted with that very decision. After all, a delayed confrontation is more ecologically valid than an immediate one.

EXPERIMENT 3

Method

Experiment 3 employed the same video fragments and lineups as in the previous experiments. Likewise, the procedure was analogous to Experiment 2a with the exception that the manipulated outcome was presented 48 hours after the recognition task.

Participants

Fifty-two (15 men and 37 women) Maastricht University undergraduates and one university employee ($M_{age} = 20.8$, $SD_{age} = 3.1$, range: 17–35) took part in the study. Student participants had the following backgrounds: psychology (86.5%), mental health (9.6%), and law (1.9%). All participants were tested individually and received either course credit or a small monetary award in return for their participation. Participants were naïve to the purpose of the study.

Materials

Post-test questionnaire

Participants were given an adjusted version of the post-test questionnaire employed in our previous experiments. As in the previous experiments, participants were first asked whether they had noticed anything strange during the experiment, and if they responded affirmatively, they were invited to provide details. If participants provided details from which we could infer that they had noticed a manipulation, they were coded as retrospective detectors. In contrast to the previous experiments, participants were next (misleadingly) informed that the present study employed two experimental conditions, one in which some of the choices were manipulated (as in the actual experiment) and one control condition where this was not the case. Participants then had to indicate in which condition they believed they had been assigned to. If a participant reported that she or he believed to be in the manipulated condition, she or he also had to indicate how many times she or he had noticed a manipulation (Hall et al., 2010; Johansson et al., 2008).

Procedure

The experiment involved two sessions and a follow-up. In Session 1, participants watched the first video fragment and were then asked to identify all four actors from separate photo lineups, always in the order of perpetrator, victim, and two bystanders. Participants were not given the option to reject the lineup. The same procedure was followed for all the remaining video fragments.

The second session was held 48 hours later. In this session, participants were instructed to imagine that they were in a court room, with the victim and suspect present. The judge would now ask them to explain why they believed

that the person they had chosen from the perpetrator lineup (i.e., the suspect) in Session 1 was actually the perpetrator. The photo of the victim was presented to the participants before they were shown the picture of the suspect in order to provide a context and to avoid confusion between the four video fragments. Participants wrote down their reasons on a sheet of paper. If participants' statements indicated that they had noticed the change, these were coded as concurrent detection. The same procedure was followed for all four video fragments. However, for two of the four suspects, participants' choices were manipulated (Table 1). That is, they were presented with a predetermined lineup member different from the one they had chosen 2 days earlier. Once participants finished this task for all four videos, they were asked to fill in the post-test questionnaire. Upon the completion of Session 2, participants were thanked and fully debriefed.

To examine whether the manipulations would affect later decisions, we collected follow-up data. We hypothesized that participants would tend to choose the same lineup member that they had been shown during the manipulation. Three to five months after the original recognition task had taken place (i.e., Session 1), participants were contacted via e-mail and provided with a PowerPoint file that included all four perpetrator lineups. The lineup photos were presented in two rows of three pictures, but in a different order than at Session 1 so as to avoid that participant would remember not the face but the number they had selected at Session 1. Participants were asked to identify the four perpetrators again. Recall that two of these four lineups had been manipulated in Session 2.

Design
In a 2×2 between-subjects design, we varied which of the four perpetrator lineups were manipulated (Cafeteria 1 and Bar vs. Cafeteria 2 and Bus), as well as the order in which the video fragments were presented (Cafeteria 1–Bus–Bar–Cafeteria 2 vs. Cafeteria 2–Bar–Bus–Cafeteria 1). Participants were randomly assigned to the different conditions. Because the level of blindness for one's own recognition decisions (x^2's$(1) \leq 1.23$, p's $\geq .40$, *phi*s ≤ 0.15) and recognition accuracy for manipulated or non-manipulated lineups (x^2's$(1) \leq 8.90$, p's $\geq .17$, *phi*s ≤ 0.45) did not differ as a function of these two factors, we will not discuss them further.

Results

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

Blindness for facial recognition decisions
From the total of $2 * 52 = 104$ manipulated lineups, 33 (31.7%) were detected concurrently and 63 (60.6%) retrospectively. Thus, in 39.4% of the cases, participants were blind to the manipulations. Specifically, the first manipulation was concurrently detected by 23.1% ($n = 12$) of participants, and the second one by 40.3% ($n = 21$). Retrospective detection rates were 50.0% ($n = 26$) and 71.2%

($n = 37$). Hence, 28.8% ($n = 15$) of the participants detected one manipulated lineup concurrently, 17.3% ($n = 9$) detected both, and 53.8% ($n = 27$) were blind for the two manipulations. Furthermore, 40.4% ($n = 21$) of the participants detected one manipulated lineup retrospectively, 40.4% ($n = 21$) detected both, and 19.2% ($n = 10$) were blind for both manipulated lineups.

The blindness rates for the two manipulations suggest that the detection rate of the second manipulated lineup is inflated. This is understandable because participants probably become suspicious after the detection of the first manipulation (Johansson et al., 2005). To correct for this cascading effect, we discarded all second manipulated lineup detections when the first manipulation had been detected. Ninety-five manipulations remained for analyses. Of these, 24 (25.3%) were detected. This distribution differed significantly from the expected, Cochran's $Q(1) = 24.0$, $p < .001$. Subsequently, we ran all analyses for all 104 manipulations as well as for the corrected 95 that excluded cascading effects. By and large, the results for the 95 manipulated lineups were similar to those for the 104 lineups. Therefore, we will report the results for the full 104 manipulations.

Recognition accuracy
In Experiment 3, recognition decisions were not required for Session 2. Analogous to our earlier analyses, we compared recognition accuracy of the two later-to-be-manipulated perpetrator lineups only with the remaining two non-manipulated perpetrator lineups, given accuracy differences depending on the role of the target (e.g., perpetrators vs. bystanders; Brown, 2003). Accordingly, for the later-to-be-manipulated lineups ($M = 87.5\%$, $SD = 21.8$) and for non-manipulated ones ($M = 91.4\%$, $SD = 19.1$), accuracies did not differ, $t(51) = 0.94$, $p = .35$, $d = 0.14$. Thus, there was no presumptive advantage for any of the lineup groups that could account for the choice-blindness result.

Interestingly, participants who concurrently detected the first manipulation ($n = 12$) displayed greater recognition accuracy for the later-to-be-manipulated lineups ($M = 100\%$, $SD = 0.0$) than those who were blind ($M = 83.7\%$, $SD = 23.7$), $x^2(1, N = 52) = 5.20$, $p = .02$, $phi = 0.32$. Analogous results were obtained for the second manipulated lineup. Again, participants who concurrently detected the second manipulation displayed greater recognition accuracy for the later-to-be-manipulated lineups ($M = 95.2\%$, $SD = 15.0$) than those who did not ($M = 82.3\%$, $SD = 24.3$), $x^2(1, N = 52) = 4.50$, $p = .05$, $phi = 0.29$. No such effects were found for retrospective detection, x^2's$(1, N = 52) \leq 9.23$, p's $\geq .52$, phis ≤ 0.13. Furthermore, recognition accuracy for the later-to-be-manipulated lineups differed as a function of the number of detected manipulations (zero, one, or two detections), $x^2(2, N = 52) = 7.13$, $p = .03$, Cramer's $V = 0.37$. Post-hoc 2×2 chi-square tests revealed that participants who detected both manipulations concurrently ($n = 9$) were significantly more likely to be accurate ($M = 100\%$, $SD = 0.0$) than participants who were blind to both manipulations ($M = 80.4\%$, $SD = 24.9$), $x^2(1, N = 37) = 3.33$, $p = .02$, $phi = 0.36$. Participants who detected one manipulation did not differ significantly ($M = 93.3\%$, $SD = 17.6$) from either of the other two groups, x^2's$(1) \leq 2.01$, p's $\geq .08$, phis ≤ 0.27. No such effect was found for retrospective detection,

$x^2(2, N=52)=1.61$, $p=.45$, $phi=0.18$. Thus, concurrent detection seemed to be positively related to the memory strength of the original representation.

Recognition accuracy at follow-up

From our $N=52$ participants, only $n=14$ replied to our follow-up e-mail. Two of them (14.3%) had been blind to our manipulations. The overall recognition accuracy at follow-up was relatively low with $M=60.7\%$ ($SD=27.2$). For the manipulated lineups, recognition accuracy for decisions made prior to the manipulation (Session 1) was $M=89.3\%$ ($SD=21.3$) and for identifications made in the follow-up $M=39.3\%$ ($SD=35.0$). For the non-manipulated lineups, recognition accuracy rates at Session 1 were $M=89.3\%$ ($SD=21.3$) and $M=82.1\%$ ($SD=31.7$) at follow-up. Because of the small number of responses, no reliable statistical analysis could be performed.

Did the previous manipulations affect participants' new identification decision? From the $2*14=28$ manipulated lineups, the same person as the one presented at the manipulation was selected only once. For 18 (64.2%) lineup decisions, participants changed their selected lineup member to a different random alternative, whereas for nine lineups (32.1%), follow-up participants stuck to their original choice. Inspection of the two blind participants' responses revealed that for none of the lineups, they had selected the same person as presented at the manipulation. Thus, by and large, participants did not tend to select the lineup member suggested by the manipulation, although the manipulation seems to have undermined their accuracy. This finding is consistent with Experiments 2a and 2b, but it deviates from previous studies (Merckelbach et al., 2011; Sauerland, Schell, et al., 2013) that showed that blind participants tend to accept the manipulative misinformation at follow-up testing.

Comparison of choice blindness across experiments

Next, we compared the detection rates across experiments. Because neither concurrent nor retrospective detection differed across Experiments 2a–c, all x^2's$(1)\le0.55$, p's$\ge.29$, phis ≤0.10, we collapsed the data (total $N=57$). Pairwise comparisons between Experiment 1 (25 detectors of $N=34$) and Experiments 2a–c (40 detectors of $N=57$) revealed no differences in concurrent detection, $x^2(1, N=91)=0.01$, $p=.92$, $phi=0.03$. However, a significantly lower level of retrospective detection in Experiment 1 (27 detectors of $N=34$) compared with Experiments 2a–c (56 detectors of $N=57$) emerged, $x^2(1, N=91)=5.74$, $p=.01$, $phi=0.30$. These results partially support our hypothesis that an evaluative component is responsible for increased blindness levels. Furthermore, there was a significantly lower level of concurrent detection in Experiment 3 (23 detectors of $N=52$), compared with Experiment 1 (27 detectors of $N=34$) and Experiments 2a–c (40 detectors of $N=57$), all x^2's$(1, N=91–109)\ge6.02$, p's $\le.01$, phis ≥0.26. Retrospectively, there was a significantly lower level of detection for Experiment 3 (42 detectors out of $N=52$) compared with Experiments 2a–c (56 detectors of $N=57$), $x^2(1, N=109)=7.33$, $p=.01$, $phi=0.29$, but not compared with Experiment 1 (27 detectors of $N=34$). In line with our hypothesis, a longer interval between the identification task and the presentation of the manipulated choice led

to increased levels of blindness. Table 2 summarizes concurrent and retrospective detection rates for all experiments.

Discussion

The results of Experiment 3 indicate that eyewitnesses' blindness for facial recognition decisions is a non-trivial phenomenon. Specifically, 39% (41) of all 104 manipulations performed in Experiment 3 remained undetected. The effect becomes even more striking when inspecting the concurrent blindness rate of the first manipulation. Indeed, 76% of the participants failed to notice, or at least failed to report, the change concurrently. To put this in perspective, one should consider that in practice, eyewitnesses are initially presented with a lineup. If an identification is made, the suspect is often arrested and incarcerated until the trial. In most European countries, the regulation of pre-trial detention applies, meaning that the identified person can be kept incarcerated till the trial without actual proof of guilt (Detained without trial: Fair Trials International's response to the European Commissions Green Paper on detention, 2011). Thus, even if the eyewitness, after the original identification at police headquarters, detects the change retrospectively in court, the innocent suspect would have spent days if not months in prison. Thus, our findings demonstrate the relevance of the choice-blindness phenomenon for this setting.

GENERAL DISCUSSION

It was the aim of the presented research to explore the conditions under which blindness for facial recognition decisions would occur. To this end, we conducted five experiments. Specifically, we tested whether requiring evaluative in contrast to absolute decisions would induce blindness (Experiment 1). Our hypothesis that in an eyewitness setting, choice-blindness rates would depend on decision type was only partially supported. Although concurrently, there were no more blind participants for the evaluative compared with absolute decision condition, in retrospect, absolute decisions yielded lower blindness rates. Furthermore, choice blindness for evaluative decisions in the eyewitness setting was decreased, compared with other studies where participants performed evaluative judgements of visual stimuli (Johansson et al., 2005, 2008). In Experiments 2a–c, we employed simple recognition procedures and found moderate concurrent choice-blindness rates and virtually no retrospective choice blindness in student and non-student samples. To increase the ecological validity of our findings, we inserted a 48-hour interval between the presentation of the recognition task and the presentation of the manipulated outcome in Experiment 3. Under these conditions, participants were concurrently blind for 68.3% and retrospectively blind for 39.4% of the manipulated lineups (Table 2). Moreover, there was a tendency of participants who failed to notice the manipulation concurrently to be less accurate in Session 1 than detectors. Blind participants' memory traces of the targets were apparently less strong than those of detectors, a fact that might have hindered them in detecting manipulations. Similar findings have been reported for blind recognition decisions in a field study (Sagana et al., 2013). Nevertheless, in both studies, the lower recognition accuracy

for detectors compared with blind participants was absent for retrospective blindness rates, ruling out the possibility of an absolute memory advantage of detectors over blind participants.

Regarding our follow-up measure, the accuracy findings in our studies is compatible with those of previous studies that report an overall decrease in choice consistency for manipulated trials when participants are asked to perform their choices a second time (Johansson, Hall, & Chater, 2012). For example, Sauerland, Schell, et al. (2013, Experiment 2) reported a general tendency of follow-up participants to adjust their responses for the manipulated items so as to match the direction (e.g., positive vs. negative) of the manipulation. In our studies, recognition accuracy was not affected by the manipulation with short intervals (i.e., 24 hours; Experiments 2a and b). For longer intervals, recognition accuracy was diminished when a decision had been manipulated (3–5 months follow-up of Experiment 3), although our participants did not switch to the lineup member presented during the manipulation. One might argue that accuracy simply degraded as a result of normal memory trace decline, over time. However, if that interpretation was correct, the same pattern of results should have occurred for the non-manipulated lineups. This was not the case. Given the small number of responders in the follow-up, however, our findings should be read with appropriate caution.

We speculated that task ambiguity may explain why retrospective choice blindness was observed in Experiments 1 and 3 but did not occur in Experiments 2a–c. Ambiguity refers to vagueness and imprecision in the environment that allows for multiple interpretations (Sloman, Fernbach, & Hagmayer, 2010). In a study on witness statements, stimulus ambiguity was found to reduce accuracy and to increase the chances of reporting false memories (Klimesch, Schimke, & Pfurtscheller, 1993). Likewise, in Experiment 1, participants may have changed their sympathy evaluations. Given the absence of ground truth, such evaluations are inherently ambiguous. The identification decisions in Experiments 2a–c, however, did not require an evaluative judgment. Thus, the ambiguity inherent to evaluative tasks (Experiment 1) was absent by using more straightforward absolute (i.e., true or false) decisions (Experiments 2a–c). That ambiguity may help to set the stage for choice blindness is underlined by the higher levels of retrospective choice blindness in Experiment 1 than in Experiments 2a–c. Long retention intervals between the identification decision and the manipulated outcome would seem to be another way to introduce ambiguity. The findings of Experiment 3 support this line of reasoning. Additionally, in their field study, Sagana et al. (2013) attributed the relatively high blindness rates to the real-life interaction that increased the complexity of the situation. Thus, task complexity may be yet another way to introduce ambiguity. All things considered, ambiguity seems to be a moderator for the occurrence of choice blindness.

The tendency of participants who failed to notice the manipulation concurrently to be less accurate in their recognition decisions than detectors (also Sagana et al., 2013) could breed the belief that choice blindness in eyewitness settings is just a manifestation of memory degradation for the chosen alternative. In the introduction, we already provided reasons why memory decline is not sufficient to fully account for choice-blindness phenomena. Note that recognition accuracy for blind participants was 82.3%. Additionally, participants were willing to provide reasons not only for a non-selected alternative but also for an incorrect one. Taken together, our findings suggest that the interplay between task ambiguity and memory of the original trace might contribute to the effect.

At a theoretical level, blindness for facial recognition decisions seems a broad construct that integrates phenomena such as *misinformation* (Loftus, 2005; Loftus & Hoffman, 1989) and *post-identification feedback* (Wells & Bradfield, 1998). It is important, though, to make a distinction between the effects of a manipulation and those of blindness (i.e., acceptance of manipulation) in choice-blindness procedures. Effects mainly driven by the presentation of the manipulated outcome and that are not dependent on the acceptance of the manipulation should be considered as *consequences of the manipulation*. Such effects were illustrated by Sagana et al. (2013) in terms of diminished rates of post-decision confidence in both blind participants *and* detectors. In the present study, the descriptively lower accuracy rates for the manipulated compared with non-manipulated lineups in the follow-up of Experiment 3 might similarly reflect consequences of the manipulation. On the other hand, effects related to the internalization of the manipulation should be considered as *consequences of blindness*. This might be equivalent to the misinformation effect. Thus, the presentation of the manipulated outcome could lead participants to develop a detailed recollection of choosing the manipulated outcome. Evidence for this internalization comes from the typical finding in choice-blindness procedures that participants not only fail to detect the manipulation of their preferred outcome but also confabulate reasons for their non-choices (Hall et al., 2010; Johansson et al., 2005). As testing the precise precursors of choice blindness is beyond the scope of the present article, our findings do not allow for disentangling the factors responsible for choice blindness. Our findings only provide the existence proof of choice blindness in an eyewitness context. However, given the aforementioned analogies, we believe that future studies on choice blindness and eyewitness identification might profit from an explicit consideration of the misinformation literature.

Apart from extending choice blindness to the domain of eyewitness recognition decisions, our study has important implications for legal settings. A situation in which an original eyewitness identification is being swapped with a different outcome may not be that unrealistic, as the case of Bernard Maughan (Wolchover, n.d.) exemplifies. A similar case came to our attention when the second author, while presenting parts of this research at a conference, was approached by a person from the audience who had been confronted with a change in her identification decision. Specifically, in court, the identified suspect was seated in the audience, while the suspect's brother was standing trial. Luckily, the witness was not blind to the error. Thus, errors in the recording of eyewitness decisions do occur. Indeed, an increasing body of research is concerned with the impact of biases due to contextual influences in the forensic science

(e.g., Kassin, Dror, & Kukucka, 2013; Saks & Koehler, 2005). Additionally, recent findings suggest that lineup administrators' knowledge about the suspect's position in the lineup affects the evaluation of the lineup outcome in accordance to the expectancies of the administrator (Rodriguez & Berry, 2014). Our point is that the types of manipulations we describe in the present article do occur in real life. Our results furthermore show that it would be naïve to think that all eyewitnesses will be able to detect and correct such manipulations at a later stage.

Looking at the limitations of the present experiments, one obvious point is that our studies deviate from real-life procedures. Participants were confronted with the recognition task immediately after witnessing the mock crime, whereas the time between the crime and identification of the suspect usually varies from days to weeks (Behrman & Davey, 2001). Likewise, the time between making an identification decision and the trial can be relatively long (Charman & Wells, 2007). Nonetheless, if blindness for one's facial recognition decision is evident even when only days have passed, it is hard to believe that it would be absent when months or years have passed. An additional limitation is that the video fragments we used depicted rather minor crimes. If participants had seen a crime that would have been more emotionally provocative, then the pattern of results might have been different because of deeper processing of central features of the crime scene (Buchanan, 2007; LaBar & Cabeza, 2006). Furthermore, if participants had seen the crime in a more ecologically valid setting (i.e., live interaction), they might have been more resistant to choice-blindness manipulations (Sagana et al., 2013). On the other hand, we speculate that owing to shorter exposure times and numerous distracters inherent to real-life interaction, more ambiguity and therefore more blindness eyewitnesses would occur. Again, these issues warrant further studies.

Another limitation is that we only used target-present lineups, and participants were forced to choose one of the lineup members. We realize that this procedure makes generalizations to standard lineup situations less obvious. Clearly, eyewitnesses in real cases are (or at least should be) warned that the perpetrator may or may not be present, and they are given the option to reject a lineup (Malpass & Devine, 1981; Wells et al., 1998). Under these conditions, many eyewitnesses decide to reject the lineup (i.e., non-choosers; Horry, Memon, Wright, & Milne, 2012). In the present studies, we opted against rejections as we were mainly interested in examining whether and to what extent choice blindness occurs for positive recognition decisions. In future research, it would be interesting to examine whether blindness occurs for identifications made from target-absent lineups and for other lineup decisions (i.e., rejections). We hypothesize that non-choosers would be more resistant to manipulations than choosers as the level of comparison would not be about the specific facial characteristic of the identified suspect ('Did I identify person A or person B?') but rather about a more broad decision outcome ('Did I or did I not make an identification?'). Thus, we expect that non-choosers exhibit lower choice-blindness levels than both accurate and inaccurate choosers, but this requires empirical testing.

To conclude, the current experiments do not only replicate previous findings showing that choice blindness is relevant for eyewitness identification decisions but also advance our understanding of the circumstances that may be decisive for the occurrence of choice blindness. Foremost, our findings point at ambiguity as a key factor for the occurrence of choice blindness. Furthermore, we believe that our findings bear relevance to research on eyewitness identification decisions. Specifically, the choice-blindness paradigm has the potential to become a postdictor of eyewitnesses' accuracy. However, more research is warranted to establish whether being a detector safely denotes higher odds of accurate identification. Finally, our findings have important implications for the judicial system. We believe that research on choice blindness will add to our knowledge about the decision-making errors of eyewitnesses.

REFERENCES

Behrman, B. W., & Davey, S. L. (2001). Eyewitness identification in actual criminal cases: An archival analysis. *Law and Human Behavior*, 25, 475–491. doi:10.1023/A:1012840831846

Brown, J. M. (2003). Eyewitness memory for arousing events: Putting things into context. *Applied Cognitive Psychology*, 17, 93–106. doi: 10.1002/acp.848

Buchanan, T. W. (2007). Retrieval of emotional memories. *Psychological Bulletin*, 133, 761–779. doi: 10.1037/0033-2909.133.5.761

Charman, S., & Wells, G. L. (2007). Applied lineup theory. In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *The handbook of eyewitness psychology, vol II: Memory for people* (pp. 219–254). Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.

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

Dellarosa, D., & Bourne Jr, L. E. (1984). Decisions and memory: Differential retrievability of consistent and contradictory evidence. *Journal of Verbal Learning and Verbal Behavior*, 23, 669–682. doi: 10.1016/s0022-5371(84)90410-9

Detained without trial: Fair Trials International's response to the European Commissions Green Paper on detention. (2011). London, UK: Fair Trials International. http://ec.europa.eu/justice/newsroom/criminal/opinion/files/110510/fti_pre-trial_detention_report_en.pdf

Festinger, L. (1962). *A theory of cognitive dissonance.* Palo Alto, CA US: Stanford University Press.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, 116–124. doi: 10.3758/BF03195503

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

Henkel, L. A., & Mather, M. (2007). Memory attributions for choices: How beliefs shape our memories. *Journal of Memory and Language*, 57, 163–176. doi: 10.1016/j.jml.2006.08.012

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

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., & Gardenfors, P. (2011). Choice blindness and the non-unitary nature of the human mind. *Behavioral and Brain Sciences*, *34*, 28–29. doi: 10.1017/S0140525x10002591

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

Kassin, S. M., Dror, I. E., & Kukucka, J. (2013). The forensic confirmation bias: Problems, perspectives, and proposed solutions. *Journal of Applied Research in Memory and Cognition*, *2*, 42–52. doi: 10.1016/j.jarmac.2013.01.001

Klimesch, W., Schimke, H., & Pfurtscheller, G. (1993). Alpha frequency, cognitive load and memory performance. *Brain Topography*, *5*, 241–251. doi: 10.1007/BF01128991

LaBar, K. S., & Cabeza, R. (2006). Cognitive neuroscience of emotional memory. *Nature Reviews Neuroscience*, *7*, 54–64. doi: 10.1038/nrn1825

Loewenstein, G., & Small, D. A. (2007). The Scarecrow and the Tin Man: The vicissitudes of human sympathy and caring. *Review of General Psychology*, *11*, 112–126. doi: 10.1037/1089-2680.11.2.112

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

Loftus, E. F., & Hoffman, H. G. (1989). Misinformation and memory: The creation of new memories. *Journal of Experimental Psychology: General*, *118*, 100–104. doi: 10.1037/0096-3445.118.1.100

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

Mather, M., Shafir, E., & Johnson, M. K. (2000). Misrememberance of options past: Source monitoring and choice. *Psychological Science*, *11*, 132–138. doi: 10.1111/1467-9280.00228

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. doi: 10.1037//1076-8971.7.1.3

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

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*, 231–259.

Rodriguez, D. N., & Berry, M. A. (2014). The effect of line-up administrator blindness on the recording of eyewitness identification decisions.

Legal and Criminological Psychology, *19*, 69–79. doi: 10.1111/j.2044-8333.2012.02058.x

Sagana, A., Sauerland, M., & Merckelbach, H. (2013). Witnesses' blindness for their own facial recognition decisions: A field study. *Behavioral Sciences and the Law*, *31*, 624–636. doi: 10.1002/bsl.2082

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. (2013). Theoretical and legal issues related to choice blindness for voices. *Legal and Criminological Psychology*, *18*, 371–381. doi: 10.1111/j.2044-8333.2012.02049.x

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

Sears, D. O. (1986). College sophomores in the laboratory: Influences of a narrow data base on social psychology's view of human nature. *Journal of Personality and Social Psychology*, *51*, 515–530. doi: 10.1037/0022-3514.51.3.515

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

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

Steblay, N. M. (1997). Social influence in eyewitness recall: A meta-analytic review of lineup instruction effects. *Law and Human Behavior*, *21*, 283–297. doi: 10.1023/A:1024890732059

Tuckey, M. R., & Brewer, N. (2003). The influence of schemas, stimulus ambiguity, and interview schedule on eyewitness memory over time. *Journal of Experimental Psychology: Applied*, *9*, 101–118.

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: 10.1002/(SICI)1099-0720(199911)13:1+<S27::AID-ACP635>3.0.CO;2-M

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

Wessel, I., van der Kooy, P., & Merckelbach, H. (2000). Differential recall of central and peripheral details of emotional slides is not a stable phenomenon. *Memory*, *8*, 95–109. doi: 10.1080/096582100387641

Wolchover, D. (n.d.). Visual identification procedures under PACA code D Retrieved from http://www.davidwolchover.co.uk/docs/Visual%20Identification%20Procedures.doc