Contents lists available at ScienceDirect

Consciousness and Cognition

CrossMark

These two are different. Yes, they're the same: Choice blindness for facial identity

Melanie Sauerland [a,*], Anna Sagana [a], Kathrin Siegmann [a], Danitsja Heiligers [a], Harald Merckelbach [a], Rob Jenkins [b]

[a] Department of Clinical Psychological Science, Maastricht University, The Netherlands
[b] Department of Psychology, University of York, UK

ARTICLE INFO

ABSTRACT

We examined the manipulability of face identity judgements by combining a sorting task for unfamiliar faces with a standard test of choice blindness. In Experiment 1, 50 participants completed a sorting task and then justified grouping specific pairs of photos together or apart. On manipulated trials, the presented pairings were different from those the participants had actually produced. Detection rates for these identity manipulations were strikingly low (∼21%). Moreover, participants readily provided justifications for identity decisions that they had not made, typically referring to specific facial features. Experiment 2 was conducted along similar lines and confirmed that lower task difficulty and higher confidence in one's face identity judgements increase detection rates. We conclude that observers can easily be led to believe that they made identity judgements they did not make. As well as underscoring the fragility of unfamiliar face matching, our findings have implications for identity judgements in legal settings.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

Individual faces do not have a fixed appearance. They vary considerably due to fluctuations in the facial surface (e.g., expression, ageing), environmental conditions (e.g., lighting), and image capture device (e.g., camera or perceiver; Jenkins, White, Van Montfort, & Burton, 2011). Such image variability makes matching or recognising unfamiliar faces a difficult and error-prone task (Burton & Jenkins, 2011; Johnston & Edmonds, 2009). In a recent demonstration of this, Jenkins et al. (2011) asked participants to sort 40 unfamiliar face images by identity, so that different images of the same person were grouped together. What the participants didn't know was that only two identities were present in the set – 20 photos of Person A and 20 photos of Person B. Surprisingly, none of the participants arrived at the correct solution (i.e., two identities). Instead, they perceived an average of *seven* distinct identities in the set. In a follow-up experiment, participants who were familiar with the two identities straightforwardly sorted the photos into two sets, almost without error. Together, these results illustrate the transformative effect of familiarity on face identification: It is very difficult to map naturally varying photos of *unfamiliar* faces onto the correct identities, even though the same task is easy when the faces are *familiar* (Jenkins & Burton, 2011).

* Corresponding author at: Section Forensic Psychology, Department of Clinical Psychological Science, Faculty of Psychology and Neuroscience, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands.
E-mail address: melanie.sauerland@maastrichtuniversity.nl (M. Sauerland).

To date, studies of unfamiliar face matching have focused mainly on response-based measures of task difficulty such as error rate. Here we examine unfamiliar face matching in a confrontational setting designed to assess observers' commitment to their own identity decisions. This situation is potentially important for face identification in legal settings, where an observer's decision may be directly queried (Wells & Olson, 2003), and manipulation of the decision (whether deliberate or inadvertent) could have serious consequences.

Over the last decade, a number of studies have uncovered an intriguing phenomenon known as *choice blindness* (Johansson, Hall, Sikström, & Olsson, 2005) – apparent lack of insight into one's own decisions. In a typical choice blindness experiment, participants are asked to indicate their preference for one of two faces. Next, participants are confronted with their choice and asked to justify it. On critical trials, the participant's choice is secretly switched, so that the decision that the participant is asked to justify is opposite to the decision that he or she actually made. Choice blindness refers to the counterintuitive finding that only very few of these manipulations are detected at the time of presentation (concurrent detection). Indeed, participants readily report the reasoning behind decisions that they did not make. Even after they are informed about the manipulation, few report that they noticed anything untoward (retrospective detection). This surprising phenomenon appears to be robust across a wide range of stimuli and judgements (Hall, Johansson, Tärning, Sikström, & Deutgen, 2010; Johansson et al., 2005; Merckelbach, Jelicic, & Pieters, 2011; Sauerland, Sagana, & Otgaar, 2013).

In this paper, we examined the manipulability of identity judgements by combining a card-sorting task for unfamiliar faces with a standard choice blindness test across two experiments. After completing the card-sorting task, participants were asked to justify grouping specific pairs of photos together or apart. As in previous choice blindness manipulations, some of the presented pairings were markedly different from those that the participants had actually produced. We expected overall performance in the card-sorting task to be poor, in line with previous findings. However, our main interest was whether participants might fail to detect manipulations of their own decisions, demonstrating choice blindness for facial identity.

Experiment 2 is a conceptual replication of Experiment 1. The two experiments differ in task difficulty, which was lower in Experiment 2 by reducing the number faces and number of identities to be sorted by the participants. This way, we wanted to test whether lowered task difficulty reduces blindness effects. Additionally, the second experiment tested the idea that decisions in which subjects place much confidence make them relatively immune against blindness manipulations.

2. Experiment 1

2.1. Materials and methods

2.1.1. Participants

Fifty student participants (26 men, 24 women; M_{age} = 23.46, SD_{age} = 2.49) took part in the current study in exchange for course credit or a small reward.

2.1.2. Stimuli

Fifty face photographs were collected for the card-sorting task. The image set comprised 6–8 photos for each of seven Australian or British public figures (Fearne Cotton, Tess Daly, Jennifer Ellison, Johanna Griggs, Ada Nicodemou, Denise van Outen, Katie Ritchie) who were unknown to our Dutch participants (as confirmed in debriefing). All images were collected from the internet, using the individuals' names as Google Image search terms. We accepted only images that (i) exceeded 150 pixels in height, (ii) showed the face in roughly frontal aspect, and (iii) were free from occlusions. However, we did not constrain variability in facial factors (e.g., emotional expression), environmental factors (e.g., lighting conditions), and image factors (e.g., camera characteristics) that affect the appearance of a face photograph. All photos were colour-printed onto white card at a size of 38 mm wide × 50 mm high and laminated for use in the experiment. Copyright restrictions prevent us from reproducing the complete image set here, but they can be obtained from the first author. Fig. 1 shows representative examples.

2.1.3. Stimulus selection

The particular image pairs that were presented in each manipulation condition were pre-selected on the basis of two pilot studies. The purpose of the first pilot study was to ensure that all four manipulations could actually be performed. For example, in order to administer the Different–Same–Same condition (DS-Same; the conditions will be explained below), a participant would have to incorrectly assign two images of the same person to two different groups. To find image pairs that could give rise to the critical errors, 54 adults (19 women, 35 men) used a 10-point Likert scale (0 = not at all similar; 9 = highly similar) to rate the similarity of 22 image pairs in an online survey. For each pair, the raters also indicated whether or not the two images showed the same person.

The second pilot study allowed us to equate image pairs for distinctiveness and memorability. Equating pairs on these measures is essential to rule out alternative explanations of apparent choice blindness effects (e.g., Meissner, Brigham, & Butz, 2005; Vokey & Read, 1992; Wickham, Morris, & Fritz, 2000). Twenty-five adults (9 females, 16 males) rated 42 images for distinctiveness and memorability using a 10-point Likert scale. Two separate measures of distinctiveness were taken. For the first measure, participants simply indicated how distinctive they found a face. For the second one, participants indicated

Fig. 1. Can you sort these faces by identity? The solution is given in Appendix A1.

how difficult it would be to recognise the face in a crowd. Finally, to verify that the faces were not familiar to Dutch observers, participants were asked whether or not they could identify the faces. As expected, none of the faces was correctly identified by any of the participants.

For the DS-Different, DS-Same, and SD-Same conditions, two image pairs were selected. The most important criterion for selection was that they did not differ significantly in terms of distinctiveness and memorability. For the SD-Same pairs we selected two images of the *same* person that were (1) rated as quite similar, and (2) judged to be the same person by most raters. This was to ensure that participants could plausibly assign the two images to the same group, thus ensuring that the planned manipulation could actually be administered. Likewise, for the DS-Different and the DS-Same conditions, we selected image pairs that were rated (1) moderate to high in similarity, and (2) judged to be the same person by few raters. Table S1 in the supporting information available online summarises these ratings for the selected image pairs.

For the SD-Different condition, we did not pre-select any image pairs during pilot testing. This is because, based on earlier findings (Jenkins et al., 2011, Experiments 1 and 2), we did not expect the required merging errors to arise. As it turned out, in a test run with four participants, merging errors occurred several times, possibly because our image set contained seven identities rather than just two. We therefore seized the opportunity to manipulate these events too by selecting suitable image pairs for each participant at the time of testing.

2.1.4. Design

Four conditions of choice blindness were presented in a within-subjects design. In all four conditions, the experimenter's assertion about the participant's grouping of two images (*same identity* or *different identity*) was different from the participant's actual grouping. The conditions were differentiated by the direction of this manipulation (*same* to *different* judgement; or vice versa) and by the actual identity relation of the images (*same* or *different* identity). Table 1 summarises these conditions.

In the first manipulation condition, the participant *correctly* assigns two images to two different groups. However, in the subsequent interview, the experimenter asserts that the participant assigned the two images to the same group. To indicate the participant's judgement (*Different*), the experimenter's assertion (*Same*), and the actual identity relation (*Different*), we refer to this condition as the *DS-Different* condition.

In the second manipulation condition, the participant *incorrectly* assigns two images to two different groups, while the experimenter asserts that the participant had assigned the images to the same group (*DS-Same* condition). *SD-Same* and *SD-Different* conditions were created analogously.

Table 1
Overview of manipulation conditions and interview procedure in Experiment 1.

Trial	Participant judgment	Experimenter assertion	Actual identity relation	Condition	Manipulated trials (marked with ×)
	The two pictures are…	You said that these two pictures are…	The two pictures are…		
1	the same person	the same person	the same person	SS-Same	
2	different people	different people	different people	DD-Different	
3	different people	the same person	different people	DS-Different	×
4	different people	different people	the same person	DD-Same	
5	different people	the same person	the same person	DS-Same	×
6	the same person	the same person	the same person	SS-Same	
7	the same person	different people	the same person	SD-Same	×
8	different people	different people	the same person	DD-Same	
9	different people	the same person	different people	DS-Different	×
10	different people	different people	different people	DD-Different	
11	the same person	different people	different people	SD-Different	×
12	the same person	the same person	different people	SS-Different	

Concurrent and retrospective detection were the dependent variables for this task. A concurrent detection is recorded when a participant reported the manipulation at the time it was presented. A retrospective detection is recorded when a participant correctly distinguishes between manipulated and non-manipulated image pairs in the post-test questionnaire.

2.1.5. Procedure

Participants signed the informed consent form and were then asked to sort the 50 face images by identity, so that images of the same person were grouped together. Participants were free to create any number of groupings, with any number of cards in each group. No time limit was imposed for this task. Immediately following the card-sorting task, participants completed the Marlowe Crowne Social Desirability Scale (MCSDS; Crowne & Marlowe, 1960) and the Gudjonsson Compliance Scale (GCS; Gudjonsson, 1989).

Meanwhile, the experimenter inspected the participant's image groupings and extracted image pairs for the interview. Where possible, participants were confronted with twelve image pairs, five of which were manipulated. Two DS-Different trials were manipulated, because we deemed it to be of greatest forensic interest, because failure to detect this type of manipulation could lead to the incrimination of an innocent suspect.

Note that not all planned manipulations could be implemented for each participant, as the manipulations were dependent on participants' identity judgements. For example, if a participant grouped together images that had been earmarked for the *DS-Same* condition, this manipulation could not be implemented. Trials were presented in the same order for all participants to minimise risk of procedural error. The experimenter began the interview by asking the participant about two images that had been correctly assigned to the same group ("You said these two are the same person. Please explain your answer"). Manipulated and non-manipulated trials then followed in alternating order. Table 1 provides an overview of all trials.

In the subsequent post-test questionnaire, participants were (misleadingly) informed that they had been assigned to one of two possible conditions with equal probability: In one condition, some images had secretly been manipulated by the experimenter, so that the pairings presented at the interview differed from those produced by the participant. In the other condition, no such manipulation had taken place. Any participants who suspected that they were in the manipulation condition were shown all of the images intermixed on a display board, and were asked to indicate those for which they had noticed a change (retrospective detection). Finally, participants were debriefed.

2.2. Results

2.2.1. Card-sorting performance

Data from two participants, who created 24 and 32 identity groups from the 50 photographs, were excluded because they were outliers. The pile including the most pictures of one individual was considered that person's identity's pile. Any picture added to this pile that did not depict this individual would be considered a merge. Any additional pile including a picture of this individual would be considered a split. For the remaining 48 participants, the mean number of perceived identities was 10.91 (95% confidence interval [CI] 9.91; 11.91, $SD = 3.54$, range 6–19) – significantly higher than the seven identities that were actually present, $t(47) = 7.66$, $p < .001$, $d = 2.20$. Thus, on average, participants assigned photos of the same face to different identity groups (identity split errors; $M = 7.27$, 95% CI [6.04, 8.50], $SD = 4.23$) more often than they assigned photos of different faces to the same identity group (identity merge errors; $M = 2.88$, 95% CI [2.08, 3.67], $SD = 2.73$), $t(47) = 6.69$, $p < .001$, $d = 1.95$. In keeping with analyses of hits and false positives in face identification (e.g., Megreya & Burton, 2007), split errors and merge errors were not correlated ($ps \geqslant .170$; overall: $r = .20$, $p = .170$), with one exception (ID 6: $r = .49$, $p < .001$). The upper part of Table 2 summarises the frequency of these errors.

Table 2
Frequencies of identity splits and merges for identities in Experiments 1 and 2.

Experiment 1	ID 1	ID 2	ID 3	ID 4	ID 5	ID 6	ID 7
Number of splits							
M	0.75	0.56	2.25	1.60	0.79	0.75	0.56
95% CI	0.50; 1.00	0.38; 0.74	1.90; 2.60	1.24; 1.96	0.57; 1.01	0.51; 0.99	0.29; 0.83
SD	0.89	0.65	1.23	1.27	0.77	0.86	0.94
Minimum	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Maximum	3.00	2.00	7.00	5.00	3.00	4.00	4.00
Number of merges							
M	0.48	0.44	0.69	0.77	0.19	0.17	0.15
95% CI	0.28; 0.68	0.25; 0.63	0.47; 0.91	0.57; 0.97	0.08; 0.30	0.01; 0.33	0.05; 0.25
SD	0.71	0.68	0.78	0.69	0.39	0.56	0.36
Minimum	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Maximum	2.00	2.00	3.00	2.00	1.00	3.00	1.00

Experiment 2	ID 1		ID 2		ID 3		ID 4
Number of splits							
M	2.87		2.29		5.44		3.39
95% CI	2.70; 3.04		2.15; 2.44		5.17; 5.70		3.22; 3.55
SD	0.99		0.81		1.50		0.93
Minimum	1.00		1.00		2.00		2.00
Maximum	6.00		4.00		9.00		6.00
Number of merges							
M	0.36		1.06		0.47		0.60
95% CI	0.22; 0.50		0.83; 1.30		0.31; 0.63		0.42; 0.79
SD	0.78		1.33		0.91		1.04
Minimum	0.00		0.00		0.00		0.00
Maximum	4.00		7.00		6.00		6.00

Note. ID = identity, CI = confidence interval. Seven different identities were used in Experiment 1 ($N = 48$), four in Experiment 2 ($N = 124$).

2.2.2. Choice blindness

The mean number of implemented manipulations per participant was 3.96 ($SD = 0.80$), resulting in a total of 190 manipulations for analysis. All participants demonstrated choice blindness, by failing to detect at least one manipulation. Consistent with this, participants performed at chance when guessing whether they had been assigned to a manipulated group (46%) or a non-manipulated group (54%; chance = 50%, binomial $z = 0.43$, $p = .667$).

The mean concurrent detection rate across all participants was 20.53%, and the rate for retrospective detections (21.05%) was not much higher. The upper part of Table 3 shows the mean detection rates and 95% CIs for each manipulation trial.

As can be seen from the table, the highest detection rates were in the *SD-Same* condition ($M = 38.24\%$ for concurrent and retrospective detection). The detection rates for the *DS-Same* condition (both Ms = 21.21%) and the *DS-Different* condition ($M = 22.89\%$, merged across both *DS-Different* trials) were moderate. In contrast, not a single *SD-Different* manipulation was detected concurrently, and only one was detected retrospectively. As this condition was always discussed last, one might expect a relatively high detection rate due to cascading effects (Johansson et al., 2005). Instead, we found that participants were uniformly blind to their own identity merge errors.

Thus, participants were most likely to detect manipulations when they had correctly sorted two photographs that depicted the same person (*SD-Same*), while detection rates were remarkably low when participants had incorrectly sorted two different persons as being the same (*SD-Different*). Detection rates for correct and incorrect *different* assignments fell in between (*DS-Different* and *DS-Same* manipulations).

We would like to draw special attention to the *DS-Different* condition (detection rate $M = 22.89\%$, merged across both *DS-Different* trials). In this condition, correct separations of different identities were overruled by a challenge from the experimenter. From a legal perspective, this is perhaps the most important aspect of our findings, as failure to detect this type of manipulation could lead to the incrimination of an innocent suspect.

2.2.3. Queries of non-manipulated trials

In 25 of 301 cases (8.3%), participants also queried a non-manipulated assignment. That is, when asked to justify decisions that they had in fact made, they expressed doubt about those decisions. With the exception of a single *SS-Same* trial, spontaneous queries of non-manipulated trials always occurred in the *DD-Same* condition (20 cases) or *SS-Different* condition (4 cases), when participants had initially made an incorrect identity judgement. Presumably, *on average*, photos presented at interview looked more similar when they showed the same person than when they showed different people, which allowed participants to detect some of their errors. Note, however, that their original, incorrect grouping decisions may have been based on different image pairs than were presented at interview (e.g., Photo 1 of Person A and Photo 1 of Person B, versus Photo 1 of Person A and Photo 2 of Person B). Apparently some photos of Person A and Person B were easily confusable, whereas others were not.

Table 3
Number of manipulations carried out and detected per condition for Experiments 1 and 2.

Manipulation trial	Choice blindness condition	Confidence condition	# performed manipulations	Concurrent detection		Retrospective detection	
				M (%)	95% CI	*M* (%)	95% CI
Experiment 1							
1	DS-Different	n/a	45	31.11	17.58; 44.64	31.11	17.58; 44.64
2	DS-Same	n/a	33	21.21	7.26; 35.16	21.21	7.26; 35.16
3	SD-Same	n/a	34	38.24	21.90; 54.58	38.24	21.90; 54.58
4	DS-Different	n/a	39	12.82	2.33; 23.31	12.82	2.33; 23.31
5	SD-Different	n/a	39	0.00	–	2.56	0.00; 7.52
	Across trials		190	20.53	14.79; 26.27	21.05	15.25; 26.85
Experiment 2							
1	DS-Different	Low	88	30.68	21.04; 40.32	47.73	37.29; 58.17
		High	88	38.64	28.47; 48.81	51.14	40.7; 61.58
2	DS-Same	Low	80	22.50	13.35; 31.65	36.25	25.72; 46.78
		High	77	36.36	25.62; 47.10	53.25	42.11; 64.39
3	SD-Same	Low	44	36.36	22.15; 50.57	34.09	20.08; 48.10
		High	55	78.18	67.26; 89.10	72.73	60.96; 84.50
4	SD-Different	Low	22	18.18	2.06; 34.30	22.73	5.22; 40.24
		High	28	71.43	54.70; 88.16	53.57	35.10; 72.04
	SD-Different (non spec)	Low	7	14.29	0.00; 40.22	57.14	20.48; 93.80
		High	3	33.33	0.00; 86.67	66.67	13.33; 100.0
	Across trials		492	39.02	34.71; 43.33	48.37	43.96; 52.80

Note. n/a = not applicable. Experiment 1: *N* = 48, Experiment 2: *N* = 124. Non spec = trials for which confidence was not formally determined during piloting.

2.2.4. Justification of identity judgements

One striking aspect of choice blindness is how readily participants justify choices that they did not in fact make (Johansson et al., 2005). The upper part of Table 4 summarises the justifications that participants provided in the current study. These justifications are organised in six categories, according to the information they were based on: (i) specific feature of the face (e.g., eyes or mouth), (ii) global appearance of the face (e.g., heavy or light), (iii) apparent age of the person (i.e., young or old), (iv) character attribution (e.g., looks happy), (v) intuition (i.e., decision based on a gut feeling), and (vi) guessing (i.e., no basis for the decision). Frequencies for each category of justification were calculated separately for manipulated trials and non-manipulated trials. As can be seen from Table 4, the majority of justifications referred to a specific feature of the face, or to its global appearance. Remarkably, this was true for manipulated trials (in which the participant previously made a different decision) as well as for non-manipulated trials. Justifications referring to age, character attribution, intuition, and guessing were less frequent overall, but were somewhat more likely for manipulated trials than for non-manipulated trials. Presumably, confabulations that rely on specific physical similarities or differences are less satisfying when the visible evidence undermines them.

2.2.5. Compliance measures

There were no significant correlations between any of the compliance measures (MCSDS, GCS) and either concurrent or retrospective detection rates ($p \geqslant .505$ for all). Thus, we found no evidence that the observed choice blindness effects were driven by participants' compliance or willingness to please.

2.3. Discussion

In Experiment 1, we used a card-sorting task to investigate choice blindness in judgments of facial identity. In line with previous findings (Jenkins et al., 2011), overall performance on this task was poor. Participants reliably committed identity split errors, whereby photographs of the same individual were perceived as different individuals. Identity merge errors (perceiving photos of different individuals to show the same individual) were comparatively rare. Thus, the main difficulty for participants in this task was in integrating dissimilar images of an individual face. This overall pattern accords with the pattern reported by Jenkins et al. (2011).

More importantly for this study, choice blindness for identity decisions was remarkably prevalent. Manipulated trials were rarely detected (~21% detection rate for concurrent and retrospective detection). The highest and lowest detection rates were found for conditions where participants believed two pictures to depict the same person. When this assignment was correct (i.e., the two pictures actually depicted the same person), detection rate was high (*SD-Same*: highest detection rate). When this assignment was incorrect, however, detection rate was almost zero (*SD-Different*: lowest detection rate). The detection rates for *different* assignments, correct or incorrect, did not differ from each other (*DS-Different* and *DS-Same*).

Table 4

Raw frequencies and proportions for the six types of justification that participants provided in Experiments 1 and 2. For non-manipulated trials, participants were justifying decisions that they *had* actually made. On manipulated trials, participants were justifying decisions that they had *not* in fact made.

	Confidence condition	Justification						
		Specific feature	Global appearance	Age	Character attribution	Intuition	Guessing	Total
Experiment 1								
Non-manipulated trials								
SS-Same	n/a	86 (90.5%)	6 (6.3%)	2 (2.1%)	1 (1.1%)	0 (0.0%)	0 (0.0%)	95 (100%)
DD-Different	n/a	86 (89.6%)	9 (9.4%)	0 (0.0%)	1 (1.0%)	0 (0.0%)	0 (0.0%)	96 (100%)
DD-Same	n/a	46 (74.2%)	7 (11.3%)	4 (6.5%)	0 (0.0%)	1 (1.6%)	4 (6.5%)	62 (100%)
SS-Different	n/a	17 (73.9%)	1 (4.3%)	1 (4.3%)	0 (0.0%)	1 (4.3%)	3 (13.0%)	23 (100%)
Across trials	n/a	235 (85.1%)	23 (8.3%)	7 (2.5%)	2 (0.7%)	2 (0.7%)	7 (2.5%)	276 (100%)
Manipulated trials								
DS-Different	n/a	37 (56.9%)	7 (10.8%)	9 (13.8%)	0 (0.0%)	7 (10.8%)	5 (7.7%)	65 (100%)
DS-Same	n/a	17 (65.4%)	1 (3.8%)	1 (3.8%)	0 (0.0%)	6 (23.1%)	1 (3.8%)	26 (100%)
SD-Same	n/a	12 (57.1%)	2 (9.5%)	2 (9.5%)	0 (0.0%)	2 (9.5%)	3 (14.3%)	21 (100%)
SD-Different	n/a	30 (76.9%)	7 (17.9%)	0 (0.0%)	0 (0.0%)	1 (2.6%)	1 (2.6%)	39 (100%)
Across trials	n/a	96 (63.6%)	17 (11.3%)	12 (7.9%)	0 (0.0%)	16 (10.6%)	10 (6.6%)	151 (100%)
Experiment 2								
Non-manipulated trials								
DD-Different	Low	52 (98.1%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (1.9%)	53 (100%)
	High	58 (97.6%)	0 (0.0%)	2 (3.3%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	60 (100%)
	n/a	10 (83.3%)	2 (16.7%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	12 (100%)
DD-Same	Low	40 (93.0%)	2 (4.7%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (2.3%)	43 (100%)
	High	30 (100%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	30 (100%)
	n/a	34 (89.5%)	2 (5.4%)	1 (2.7%)	0 (0.0%)	1 (2.7%)	0 (0.0%)	38 (100%)
SS-Same	Low	12 (92.3%)	1 (7.7%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	13 (100%)
	High	53 (98.1%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (1.9%)	0 (0.0%)	54 (100%)
	n/a	55 (90.2%)	4 (6.6%)	0 (0.0%)	1 (1.6%)	0 (0.0%)	1 (1.6%)	61 (100%)
SS-Different	Low	9 (100%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	9 (100%)
	High	8 (100%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	8 (100%)
	n/a	58 (87.9%)	3 (4.5%)	1 (1.6%)	0 (0.0%)	1 (1.6%)	3 (4.5%)	66 (100%)
Across trials	n/a	419 (93.7%)	14 (3.1%)	4 (0.9%)	1 (0.2%)	3 (0.7%)	6 (1.3%)	447 (100%)
Manipulated trials								
DS-Different	Low	57 (93.4%)	4 (6.6%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	61 (100%)
	High	47 (87.0%)	4 (7.4%)	0 (0.0%)	0 (1.9%)	1 (1.9%)	2 (3.7%)	54 (100%)
DS-Same	Low	57 (91.9%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (1.6%)	4 (6.4%)	62 (100%)
	High	45 (91.8%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	4 (8.2%)	49 (100%)
SD-Same	Low	24 (85.7%)	2 (7.1%)	0 (0.0%)	0 (0.0%)	1 (3.6%)	1 (7.1%)	28 (100%)
	High	12 (100%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	12 (100%)
SD-Different	Low	17 (94.4%)	1 (5.6%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	18 (100%)
	High	7 (87.5%)	0 (0.0%)	1 (12.5%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	8 (100%)
SD-Same non spec	Low	6 (100%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	6 (100%)
	High	2 (100%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	2 (100%)
Across trials	n/a	274 (91.3%)	11 (3.7%)	1 (0.3%)	0 (0.0%)	3 (1.0%)	11 (3.7%)	300 (100%)

Moreover, participants readily offered justifications for decisions that they had not in fact made. These justifications were often strikingly specific, being based mainly on particular facial features (e.g., comparison of the eyes or mouth). Analysis of compliance measures found no evidence that participants were merely 'playing along' with the experimenter (see Merckelbach et al., 2011; Sauerland et al., 2013, for a similar conclusion). Together, these findings suggest genuine choice blindness for facial identity decisions, analogous to the effect previously reported using other stimuli and settings (e.g., Hall et al., 2010; Johansson et al., 2005; Merckelbach et al., 2011; Sauerland et al., 2013).

We find choice blindness in this context interesting for several reasons. First, one might expect participants not to make mistakes when grouping face photographs by identity. The task imposes minimal demands on memory, participants gave it their full attention, and there was no time pressure. As it turns out, this seemingly simple task attracts many errors, when the faces are unfamiliar (Jenkins et al., 2011). Similar difficulties are found consistently in other tasks involving identification of unfamiliar faces (Burton & Jenkins, 2011; Hancock, Bruce, & Burton, 2000). Previous studies have suggested that decisional ambiguity fosters choice blindness (e.g., Merckelbach et al., 2011; Sagana, Sauerland, & Merckelbach, 2013, 2014, 2015). The prevalence of choice blindness in the current study was likely exacerbated by the perceptual difficulty of the prescribed task.

Second, whether or not participants' identity decisions are correct, provided that they are *reasoned* decisions, one might expect a degree of commitment to them (Salti, El Karoui, Maillet, & Naccache, 2014). However, participants were just as willing to accept that they had made a different decision and readily supplied reasons for that decision when asked. This flexibility to accept a markedly different decision implies that grouping photographs by identity did not induce strong categorical boundaries (Beale & Keil, 1995; Rothstein, Henson, Treves, Driver, & Dolan, 2004), or cause the grouped images strongly to cohere into a single representation (Neumann, Schweinberger, & Burton, 2013) in this procedure. Indeed, the experimenter could extract photos from the groups formed by the participant and recombine them with photos from other groups, without the participant noticing.

This flexibility could reflect participants' low confidence in their own perceptual decisions, commensurate with the objective difficulty of the task. If the participant is merely *guessing* or more or less randomly deciding whether or not one photograph belongs with another, then it may not be surprising that the decision can easily be overruled. After all, the participant might equally have made a different guess. We examined this possibility in the next experiment.

3. Experiment 2

Our second experiment had two aims. First, we wanted to test whether assignments made with higher confidence would be detected at a higher rate than assignments made with lower confidence. Second, we wanted to lower task difficulty and examine how this affects detection rates. To this end, we reduced the number of faces that the participants had to sort from 50 to 25, and reduced the number of identities from seven to four.

3.1. Materials and methods

3.1.1. Participants

One-hundred-and-twenty-four participants (14 men, 110 women; $M_{age} = 21.67$, $SD_{age} = 3.82$) took part in exchange for course credit or a small reward.

3.1.2. Stimuli

Twenty-five face photographs were collected for the card-sorting task. The image set comprised 4–9 photos for each of four British public figures (Jennifer Ellison, Denise van Outen, Tess Daly, and Fearne Cotton) who were unknown to our Dutch participants (as confirmed in debriefing). Image collection and presentation was analogous to Experiment 1.

3.1.3. Stimulus selection

As in Experiment 1, manipulation photo pairs were chosen to match on distinctiveness and memorability. Additionally, for each manipulation type, we selected one or more photo pairs for which pilot participants displayed high or low confidence. More specifically, $N = 35$ participants (11 men, 24 women) rated the memorability and distinctiveness (two different measures) of 52 faces. Those pairs that matched most closely were selected as potential manipulation pairs. Following our experience in Experiment 1, we also matched photo pairs for the *SD-Different* condition. Additionally, two more groups of pilot participants ($N = 36$, 13 men, 23 women) indicated whether or not 48 photo pairs depicted the same or different individuals and how confident they were about this (from 0% to 100%). Confidence ratings for *different* responses were subsequently multiplied by -1 in order to differentiate between confidence in sameness vs. differentness. Those pairs that showed the greatest difference in confidence ratings, while still roughly matching in memorability and distinctiveness, were selected as pairs for the four manipulation types. The mean distinctiveness, memorability, and confidence ratings of the selected pairs are shown in Tables S2 and S3 in the supporting information available online.

3.1.4. Design

Assignment confidence (high vs. low) and choice blindness condition (*DS-Different* vs. *DS-Same* vs. *SD-Same* vs. *SD-Different*) were both manipulated. To prevent participants from becoming become overly suspicious, we limited the number of manipulations per participant to four. Therefore, we aimed at manipulating a high-confidence photo pair and a low-confidence photo pair for two of the four possible conditions of choice blindness. Which specific conditions of choice blindness were used was dependent on the assignments the participant made. Concurrent and retrospective detection were again the dependent variables.

3.1.5. Procedure

The procedure was analogous to Experiment 1, except that (i) participants now sorted only 25 faces instead of 50, and (ii) participants were interviewed about eight of their assignments in total, four of which were manipulated. The interview always began with a non-manipulated trial, with manipulation status alternating for the remaining trials. The non-manipulated trials comprised one *SS-Same*, *DD-Different*, *DD-Same*, and *SS-Different* trial each. When possible, the non-manipulated pairs were drawn from the pre-selected pairs that were not used for manipulation for that participant. When this was not possible, randomly selected pairs were used.

3.2. Results and discussion

3.2.1. Card-sorting performance

The mean number of perceived identities was 10.15 (95% CI [9.52, 10.78], *SD* = 3.55, range 3–19) – significantly higher than the four identities that were actually present, *t*(123) = 19.30, *p* < .001, *d* = 3.42. Thus, as in Experiment 1, participants assigned photos of the same face to different groups (identity split errors; *M* = 13.98, 95% CI [13.44, 14.52], *SD* = 3.05) more often than they assigned photos of different faces to the same group (identity merge errors; *M* = 2.50, 95% CI [1.91, 3.08], *SD* = 3.28), *t*(123) = 24.12, *p* < .001, *d* = 4.35. The two measures were negatively correlated for all identities, with *p*s varying between <.001 and .096 (overall: *r* = −.40, *p* < .001). The lower part of Table 2 summarises the frequency of these errors for each identity.

3.2.2. Choice blindness

Initially, we had aimed to test two different types of choice blindness manipulation with varying confidence conditions for each participant (i.e., within subjects). The choice blindness condition would vary between subjects. During data collection, however, it became apparent that strict adherence to this design would not be possible due to restrictions imposed by participants' assignments. More specifically, for many participants, a certain choice blindness manipulation would only be possible for the high confidence manipulation and not for the low confidence manipulation or vice versa. Furthermore, it turned out to be particularly difficult to perform *SD-Different* manipulations, because participants did not assign the selected photo pairs as necessary for this manipulation (i.e., assign two pictures as the same person when they were actually different persons). Two measures were taken to address these issues. First, if a high and a low confidence manipulation for two specific choice blindness conditions were not feasible within one participant, the experimenter would perform any other manipulation that was possible given the participant's assignments. Second, in order to obtain a critical number of *SD-Different* trials, the experimenter randomly chose *SD-Different* photo pairs in 10 cases. High and low confidence pairs were established based on remarks from previous participants. Given that high and low confidence was not formally established for these photo pairs, the results for these trials are reported separately.

Four-hundred-and-ninety-two manipulations were carried out (four manipulations for 122 participants, two manipulations for 2 participants). Let us first consider concurrent detection. In total, 192 of these manipulations were detected (39%; see lower part of Table 3). Participants concurrently detected 1.55 manipulations on average (95% CI [1.31, 1.78], *SD* = 1.33). Thirty-four participants detected none of the manipulations concurrently, 34 participants detected one, 24 participants detected two, 18 participants detected three, and 14 participants detected all four manipulations concurrently. Note that this concurrent detection rate is significantly higher than the one observed in Experiment 1, $\chi^2(1)$ = 20.94, *p* < .001, *phi* = .18.

In retrospect, 238 (48%; see lower part of Table 3) of the manipulations were detected, significantly more than concurrently (*M* = 1.93, 95% CI [1.70, 2.15], *SD* = 1.26), *t*(123) = 3.12, *p* = .002. Twenty-four participants detected none of the manipulations in retrospect, 16 detected one, 43 participants detected two, 27 participants detected three, and 14 participants detected all four manipulations retrospectively. Recall that in Experiment 1, all 48 participants had missed at least one of the manipulations. Unlike Experiment 1, most participants also realised that they had been assigned to a manipulated group (80.6%; chance = 50%, binomial *z* = 6.73, *p* < .001). Analogous to concurrent detection, the retrospective detection rate was significantly higher than the one observed in Experiment 1, $\chi^2(1)$ = 42.37, *p* < .001, *phi* = .25. These findings indicate that, as expected, when task difficulty decreases, detection rates increase. Nevertheless, the majority of manipulations went undetected.

Fig. 2 displays the detection rates of Experiment 1 and 2. Note that inferential comparisons of the high and low confidence conditions were not possible because, as explained above, the very challenging procedure resulted in a mixture of independent and dependent data. We therefore rely on condition means and 95% CIs when comparing conditions.

Inspection of Fig. 2 shows that, on a descriptive level, concurrent and retrospective detection rates differed as a function of assignment confidence for each choice blindness condition and hence overall. Specifically, detection rates were higher for high-confidence pairs than for low-confidence pairs. This finding confirms our hypothesis that higher confidence goes hand in hand with higher detection rates.

Comparing responses for high-confidence and low-confidence pairs, there are some notable differences. Consider the concurrent detection rates. When presented with a low-confidence pair, participants most strongly resisted challenges when their initial assignments had been correct (*DS-Different* and *SD-Same*). For high-confidence pairs, participants most strongly resisted challenges of faces that they thought to depict the same person, irrespective of this initial judgement being correct (*SD-Same*) or not (*SD-Different*). Strikingly, the low concurrent detection rates for the *SD-Different* condition in Experiment 1 (*M* = 0.00%) and *SD-Different* low confidence pairs in Experiment 2 (*M* = 18.18%) were complemented by one of the highest detection rates for *SD-Different* high confidence pairs (Experiment 2, *M* = 71.43%). For retrospective detection, the detection rate was again relatively low for the *SD-Different* condition across confidence levels, while higher detection rates were obtained for correct assignments of different identities (*DS-Different*; low confidence pairs) or same identities (*SD-Same*, high confidence pairs). This pattern suggests a possible interaction between assignment confidence and the type of manipulation. Although this assumption cannot be tested with the current data, it would be interesting to explore this further in future studies.
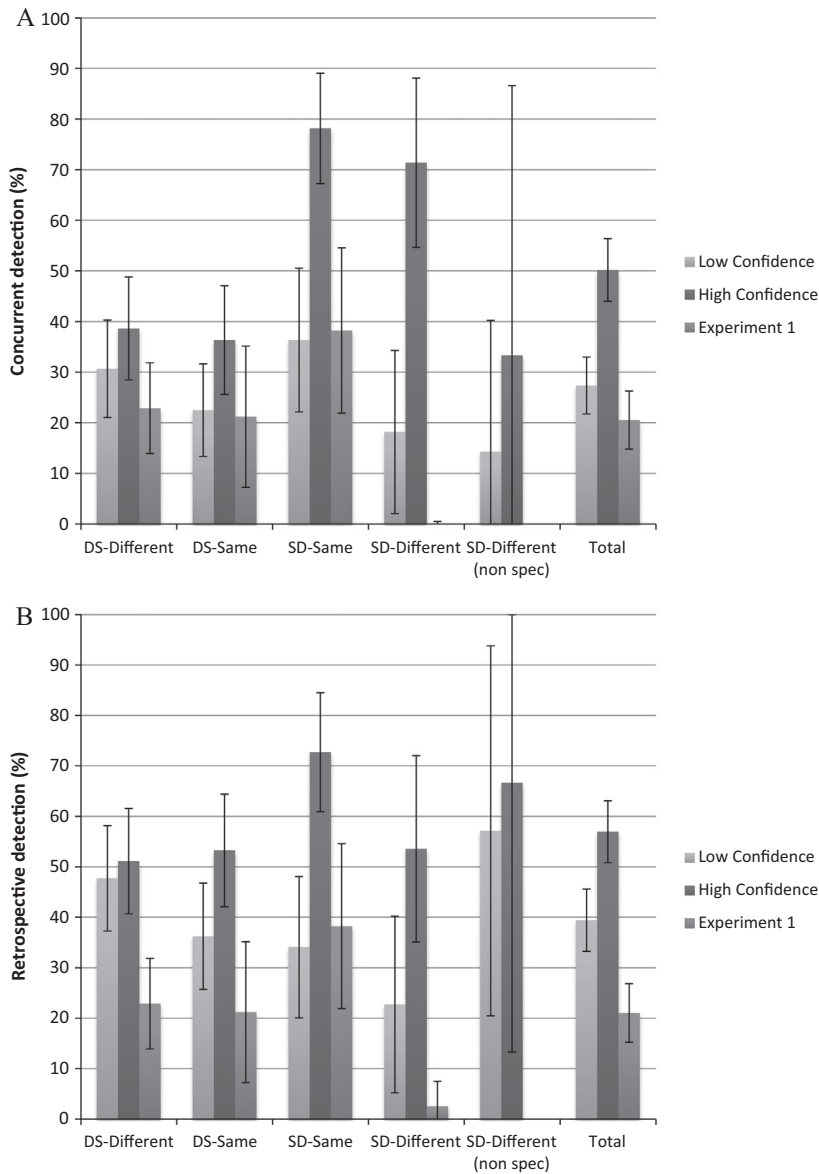
Fig. 2. Concurrent (A) and retrospective (B) detection rates (in %) and 95% confidence intervals for Experiments 1 and 2 (separately for high and low assignment confidence). Non spec = trials for which confidence was not formally determined during piloting.

3.2.3. Queries of non-manipulated trials

In 49 of 496 cases (9.9%), participants also queried a non-manipulated assignment. This is similar to the query rate in Experiment 1 (8.3%; $\chi^2(1) = 0.55$, $p = .271$, $phi = .03$). Also similar to Experiment 1, most spontaneous queries of non-manipulated trials concerned the *DD-Same* condition (26 cases) and *SS-Different* condition (13 cases). *DD-Different* trials (9 cases) and *SS-Same* trials (1 case) were queried less frequently, indicating that unwarranted queries again occurred primarily when participants had made an incorrect identity judgement.

3.2.4. Justification of identity judgements

Frequencies for each category of justification separately for manipulated trials and non-manipulated trials can be found in the lower part of Table 4. The overwhelming majority of justifications referred to a specific feature of the face, and this was again true for manipulated trials and non-manipulated trials (91.3% and 93.7%), followed by global facial appearance. With the exception of guesses in manipulated trials, other justifications occurred infrequently overall (⩽1.0%).

4. General discussion

Across two experiments we demonstrated that choice blindness in judgments of facial identity decisions is strikingly prevalent and that participants readily offer justifications for decisions that they had not in fact made. Experiment 1 established this phenomenon. Experiment 2 replicated and extended the main findings by confirming that the strength of the effect is affected by task difficulty and the level of confidence at which facial identity decisions were made: compared with Experiment 1, where task difficulty was relatively high, detection rates were higher (i.e., choice blindness was less prevalent) in Experiment 2, where the task at hand was easier. In Experiment 2, we also found that high confidence was associated with high detection rates (i.e., low prevalence of choice blindness).

One intriguing aspect of our findings is that participants seldom claimed to have been guessing, and nearly always offered specific reasons for their decisions (mainly similarity of facial features). Overall, these findings suggest surprisingly fragile insight into one's own judgments of other people's faces: Participants were not aware that their ostensible decisions (as claimed by the experimenter) were at odds with their actual decisions, and they were not aware that their justifications of these decisions were confabulated.

As well as extending choice blindness to a new area of cognition research (face identification), our findings also have implications for identification in legal settings. For example, the concurrent detection rate for *DS-Different* manipulations was low in both experiments (Experiment 1: 23%, Experiment 2: 35% across both confidence conditions). This means that participants could easily be manipulated into believing that they had interpreted two photos as showing the same person, even though they had actually classified them as different people. This particular form of choice blindness is potentially important for legal identification, as it suggests a specific route by which an innocent person could be incriminated. The growing use of surveillance cameras and CCTV in public spaces has fuelled increasing use of video images for identification purposes (Policy Department for Citizens' Rights and Constitutional Affairs, 2009; Wells, Allard, & Wilson, 2006). Eyewitnesses may be asked to match CCTV images of the perpetrator with a suspect, and the images can be presented in court as identification evidence (Valentine, 2006). Although the current sorting task may differ from forensic face identification in some respects (e.g. number of images involved), it is not necessarily more difficult. Unlike many forensic situations, all of the images in this task were of high quality, and were present when the decisions were made, meaning that no memory was involved in the task. For identification tasks that are more difficult, choice blindness may be more likely still. While earlier research has shown that observers perform poorly in identification tasks when the faces are unfamiliar (Bruce, Henderson, Newman, & Burton, 2001; Davies, 2000), the present study goes much further, by implying in addition that witnesses or police may be blind to their own identity matches. If a matching decision is poorly documented, due to procedural errors, misunderstandings or even deliberate manipulations (see Findley & Scott, 2006; Leo & Drizin, 2010), a switch of identities could easily go undetected. Considering how often other errors slip into forensic testing procedures (e.g., erroneous matches of bullets or fingerprints, police misconduct; see Ridolfi & Possley, 2010; Saks & Koehler, 2005), it would be interesting to estimate the frequency of identity switches. The clear implication of our current findings is that such switches could easily go undetected (see Loftus & Ketcham, 1991; Wagenaar, 1989, for relevant cases). Future studies should investigate other conditions that exacerbate or mitigate choice blindness, besides the confidence factor implicated here. For now, we show that observers can easily be led to believe that they made identity judgements that they did not make. As well as demonstrating the fragility of unfamiliar face matching, our findings have direct implications for identity and mistaken identity in legal settings.

Appendix A. Solution to Fig. 1

ABCDEFGCFG
EFGBCAFDEB
DABCAGDFCE
FCEGFEABGD
GEGFEACDBC

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.concog.2016.01.003.

References

Beale, J. M., & Keil, F. C. (1995). Categorical effects in the perception of faces. *Cognition, 57,* 217–239. http://dx.doi.org/10.1016/0010-0277(95)00669-X.

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. http://dx.doi.org/10.1037/1076-898x.7.3.207.

Burton, A. M., & Jenkins, R. (2011). Unfamiliar face perception. In A. J. Calder, G. Rhodes, M. H. Johnson, & J. V. Haxby (Eds.), *The Oxford handbook of face perception* (pp. 287–306). Oxford: Oxford University Press.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*, 349–354. http://dx.doi.org/10.1037/h0047358.

Davies, G. S. (2000). Closed-circuit television: How effective an identification aid? *British Journal of Psychology, 91*, 411–426. http://dx.doi.org/10.1348/000712600161907.

Findley, K., & Scott, M. (2006). The multiple dimensions of tunnel vision in criminal cases. *Wisconsin Law Review, 2006,* 291–398.

Gudjonsson, G. H. (1989). Compliance in an interrogative situation: A new scale. *Personality and Individual Differences, 10*, 535–540. http://dx.doi.org/10.1016/0191-8869(89)90035-4.

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. http://dx.doi.org/10.1016/j.cognition.2010.06.010.

Hancock, P. J., Bruce, V., & Burton, A. M. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Sciences, 4*, 330–337. http://dx.doi.org/10.1016/S1364-6613(00)01519-9.

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

Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition, 121*, 313–323. http://dx.doi.org/10.1016/j.cognition.2011.08.001.

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. http://dx.doi.org/10.1126/science.1111709.

Johnston, R. A., & Edmonds, A. J. (2009). Familiar and unfamiliar faces recognition: A review. *Memory, 17*, 577–596. http://dx.doi.org/10.1080/09658210902976969.

Leo, R. A., & Drizin, S. A. (2010). The three errors: Pathways to false confession and wrongful conviction. In G. D. Lassiter & C. A. Meissner (Eds.), *Police interrogations and false confessions: Current research, practice, and policy recommendations* (pp. 9–30). Washington, DC, US: American Psychological Association.

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.

Megreya, A. M., & Burton, A. M. (2007). Hits and false positives in face matching: A familiarity-based dissociation. *Perception & Psychophysics, 69*, 1175–1184. http://dx.doi.org/10.3758/BF03193954.

Meissner, C. A., Brigham, J. C., & Butz, D. A. (2005). Memory for own- and other-race faces: A dual-process approach. *Applied Cognitive Psychology, 19*, 545–567. http://dx.doi.org/10.1002/acp.1097.

Merckelbach, H., Jelicic, M., & Pieters, M. (2011). Misinformation increases symptom reporting – A test – Retest experiment. *Journal of the Royal Society of Medicine Short Reports, 2*. http://dx.doi.org/10.1258/ shorts.2011.011062.

Neumann, M. F., Schweinberger, S. R., & Burton, A. M. (2013). Viewers extract mean and individual identity from sets of famous faces. *Cognition, 128*, 56–63. http://dx.doi.org/10.1016/j.cognition.2013.03.006.

Policy Department for Citizens' Rights and Constitutional Affairs – European Parliament (2009). *A review of the increased use of CCTV and video-surveillance for crime prevention purposes in Europe* (Note No. PE 419.588). <http://www.statewatch.org/news/2009/apr/ep-study-norris-cctv-video-surveillance.pdf>.

Ridolfi, K. M., & Possley, M. (2010). *Preventable error: A report on prosecutorial misconduct in California 1997–2009.* Santa Clara, CA: Northern California Innocence Project.

Rothstein, P., Henson, R. N., Treves, A., Driver, J., & Dolan, R. J. (2004). Morphing Marilyn into Maggie dissociates physical and identity face representations in the brain. *Nature Neuroscience, 8*, 107–113. http://dx.doi.org/10.1038/nn1370.

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. http://dx.doi.org/10.1002/bsl.2082.

Sagana, A., Sauerland, M., & Merckelbach, H. (2014). "This is the person you selected": Eyewitnesses' blindness for their own facial recognition decisions. *Applied Cognitive Psychology, 753–764.* http://dx.doi.org/10.1002/acp.3062.

Sagana, A., Sauerland, M., & Merckelbach, H. (2015). The effect of choice reversals on blindness for identification decisions. *Psychology, Crime & Law.* http://dx.doi.org/10.1080/1068316X.2015.1085984. Advance online publication.

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

Salti, M., El Karoui, I., Maillet, M., & Naccache, L. (2014). Cognitive dissonance resolution is related to episodic memory. *PLoS ONE, 9*, e108579. http://dx.doi.org/10.1371/journal.pone.0108579.

Sauerland, M., Sagana, A., & Otgaar, H. (2013). Theoretical and legal issues related to choice blindness for voices. *Legal and Criminological Psychology, 18*, 371–381. http://dx.doi.org/10.1111/j.2044-8333.2012.02049.x.

Valentine, T. (2006). Forensic facial identification. In A. Heaton-Armstrong, E. Shepherd, G. Gudjonsson, & D. Wolchover (Eds.), *Witness testimony: Psychological, investigative and evidential perspectives* (pp. 267–285). Oxford: Oxford University Press.

Vokey, J. R., & Read, J. D. (1992). Familiarity, memorability, and the effect of typicality on the recognition of faces. *Memory & Cognition, 20*, 291–302. http://dx.doi.org/10.3758/BF03199666.

Wagenaar, W. A. (1989). *Identifying Ivan: A case study in legal psychology.* Cambridge, MA: Harvard University Press.

Wells, H., Allard, T., & Wilson, P. (2006). *Crime and CCTV in Australia: Understanding the relationship. Centre for Applied Psychology and Criminology.* Australia: Bond University.

Wells, G. L., & Olson, E. A. (2003). Eyewitness testimony. *Annual Review of Psychology, 54*, 277–295. http://dx.doi.org/10.1146/annurev.psych.54.101601.145028.

Wickham, L. H. V., Morris, P. E., & Fritz, C. O. (2000). Facial distinctiveness: Its measurement, distribution and influence on immediate and delayed recognition. *British Journal of Psychology, 91*, 99–123. http://dx.doi.org/10.1348/000712600161709.