



Received: 12 August 2014
Accepted: 06 November 2014

*Corresponding author: Markus Bindemann, School of Psychology, University of Kent, Canterbury CT2 7NP, UK
E-mail: m.bindemann@kent.ac.uk

Reviewing editor:
Peter Walla, Webster Vienna Private University, Austria; University of Newcastle, Australia

Additional information is available at the end of the article

COGNITIVE SCIENCE & NEUROSCIENCE | RESEARCH ARTICLE

Perceived ability and actual recognition accuracy for unfamiliar and famous faces

Markus Bindemann^{1*}, Janice Attard¹ and Robert A. Johnston¹

Abstract: In forensic person recognition tasks, mistakes in the identification of unfamiliar faces occur frequently. This study explored whether these errors might arise because observers are poor at judging their ability to recognize unfamiliar faces, and also whether they might conflate the recognition of familiar and unfamiliar faces. Across two experiments, we found that observers could predict their ability to recognize famous but not unfamiliar faces. Moreover, observers seemed to partially conflate these abilities by adjusting ability judgements for famous faces after a test of unfamiliar face recognition (Experiment 1) and vice versa (Experiment 2). These findings suggest that observers have limited insight into their ability to identify unfamiliar faces. These experiments also show that judgements of recognition abilities are malleable and can generalize across different face categories.

Subjects: Behavioral Neuroscience; Cognitive Psychology; Perception; Psychological Science; Visual Cognition; Visual Perception

Keywords: face recognition; familiar; unfamiliar; beliefs; overgeneralization

1. Introduction

Person identification routinely involves unfamiliar faces. This is important in applied settings. Eyewitnesses, for example, might observe an unknown perpetrator at a crime scene and may later attempt to identify this person in police investigations (Wells, Memon, & Penrod, 2006; Wells & Olsen, 2003). And security tasks such as passport control require the matching of a face photograph from an identity document to its bearer (Jenkins & Burton, 2008; Johnston & Bindemann, 2013). There is considerable evidence that person identification in these tasks is highly error-prone. In the eyewitness domain, for example, correct identifications are typically made only on between 60 and

ABOUT THE AUTHORS

Markus Bindemann, Janice Attard and Robert Johnston work in the School of Psychology at the University of Kent, Canterbury, UK. Their shared research interests lie in the theoretical study of face processing in cognitive psychology as well as in applied forensic recognition tasks, such as person identification at national borders and by eyewitnesses during criminal investigations.

PUBLIC INTEREST STATEMENT

Faces are of fundamental importance for person identification. However, while we appear to be very good at recognizing the faces of people that we know, mistakes in the identification of unfamiliar faces occur frequently. The identification of such faces is particularly important in applied forensic tasks, such as eyewitness identification or person identification at passport control. This study explored the idea that such errors might arise because people are poor at judging their own ability to recognize unfamiliar faces. Across two experiments, observers could predict their ability to recognize famous faces but did not have similar insight into their ability to identify unfamiliar faces. The results also suggest that people can conflate these different face processes.

80% of trials under optimized experimental conditions (see, e.g. Bruce et al., 1999; Megreya & Burton, 2008; Shapiro & Penrod, 1986). Identification accuracy can be lower still outside the laboratory. For example, a recent evaluation of police investigations in the UK showed that mistaken identifications of innocent line-up members occur in 41% of cases (Memon, Havard, Clifford, Gabbert, & Watt, 2011). A similar pattern has been found with face matching tasks (Johnston & Bindemann, 2013). Under optimized conditions, identification errors are made on 10–20% of trials (e.g. Bindemann, Avetisyan, & Blackwell, 2010; Bindemann, Avetisyan, & Rakow, 2012; Burton, White, & McNeill, 2010). Performance declines further when the to-be-compared photographs are taken months apart (Megreya, Sandford, & Burton, 2013), picture quality is degraded (Bindemann, Attard, Leach, & Johnston, 2013) or realistic identity documents are used (Bindemann & Sandford, 2011; Kemp, Towell, & Pike, 1997).

The source of these errors is now well understood. With limited familiarity, observers can only have an incomplete idea of how a particular face looks. This representation is tied to the circumstances in which the person was viewed and is therefore instance-bound (see, e.g. Burton, Jenkins, & Schweinberger, 2011; Hancock, Bruce, & Burton, 2000; Johnston & Edmonds, 2009). When further changes occur in the appearance of a face, for example, through variation in lighting (Longmore, Liu, & Young, 2008), expression (Bruce, 1982), or age (Megreya et al., 2013), it becomes increasingly difficult to associate this with the initial encounter of the same person.

This theoretical framework might explain why the identification of unfamiliar faces can be difficult. However, the existing evidence cannot account for why observers continue to make identification errors despite the difficulty of this task. For example, such errors could be avoided in experiments and real-life incidents by making “target absent” or “don’t know” responses. However, observers persist in making identification errors even when such response options are available (see, e.g. Bruce et al., 1999; Megreya & Burton, 2008; Memon et al., 2011; Weber & Perfect, 2012). The challenge of understanding the cause of these errors in applied settings, therefore, not only requires awareness of the extraneous factors that make unfamiliar face processing difficult (e.g. lighting, expression and age), but needs to explain why observers are *willing* to make incorrect identifications.

A potential explanation for this behaviour could be that observers are poor at judging their own ability to identify unfamiliar faces. This explanation is appealing considering that observers rarely receive feedback for identification. Consider an encounter with a person whom we have met only briefly before. If we fail to recognize this person in a subsequent encounter then there is no reason to assume that they have already been met. Consequently, without any corrective feedback, our inability to recognize unfamiliar faces might remain unchallenged. The absence of such challenges could sustain a belief that unfamiliar face identification is generally accurate. In line with this reasoning, it is notable that laboratory experiments on unfamiliar face identification typically do not provide feedback for accuracy. Thus, observers might be unaware of their poor performance in these tasks. In turn, when such feedback is provided, clear performance benefits are found (Alenezi & Bindemann, 2013; White, Kemp, Jenkins, & Burton, 2014).

In the absence of corrective feedback for unfamiliar face identification, observers might also draw on other sources to inform judgements of their identification ability. One possibility is that they overgeneralize their ability to recognize *familiar* faces, of family members, acquaintances, or famous people, to situations in which unfamiliar faces need to be identified (see Burton, 2013; Burton & Jenkins, 2011; Jenkins & Burton, 2011). Such familiar face recognition appears to be qualitatively different from the identification of unfamiliar faces (Megreya & Burton, 2006) and much more robust (see, e.g. Bahrnick, Bahrnick, & Wittlinger, 1975; Bruce, 1982; Burton, Wilson, Cowan, & Bruce, 1999). If observers are prone to confounding these processes, then this might, therefore, also affect ability judgements for unfamiliar faces. This possibility, that we might overgeneralize our ability to recognize familiar faces to their unfamiliar counterparts, has also been put forward as a potential explanation for errors in forensic person identification (Burton & Jenkins, 2011; Jenkins & Burton, 2011) but has not been examined so far. The current study sought to investigate these questions in two experiments.

2. Experiment 1

The aims of this experiment were twofold. Firstly, we sought to examine whether observers can predict their ability to recognize unfamiliar faces. For this purpose, we first asked observers to judge their abilities to recognize unfamiliar faces. We then tested the recognition of unfamiliar faces to determine whether a priori ability judgements predict task performance. In this test, participants attempted to select face targets from subsequent identity line-ups in an established laboratory test of unfamiliar face recognition (Bruce et al., 1999).

As we anticipated the relationship between perceived ability and task accuracy to be poor prior to the face test, it was important to assess whether observers can refine their ability judgements after feedback for performance has been provided. To explore this possibility, the difficulty of this test was manipulated by presenting the same face image as the initial target in the corresponding line-up, in the same-image condition, or by using two different images of the same identity for the initial target and its counterpart in the line-up in a more difficult different-image condition (see, e.g. Bruce, 1982; Jenkins & Burton, 2011; Longmore et al., 2008). These conditions were designed to induce a feeling of competence in unfamiliar face recognition in the same-image condition or of relative incompetence in the different-image condition. We then measured whether feedback for these different face conditions exerted distinct effects on subsequent judgements of face recognition ability. If such effects can be found, then observers should rate their face recognition ability more highly in the comparatively easy same-image condition than the different-image condition.

We also recorded ability judgements for family members and famous people prior to and after the face test to explore whether observers would conflate the recognition of these categories with that of unfamiliar faces. If this is the case, then ability judgements for these different categories should correlate prior to the face test. In addition, feedback for the unfamiliar face test should not only affect subsequent judgements to unfamiliar faces, but might also influence ability ratings to family and famous faces.

3. Method

3.1. Participants

Sixty undergraduate students (40 female) in the School of Psychology at the University of Kent, with a mean age of 21 years (range = 18 to 34), participated in the experiment for course credit. All reported normal or corrected-to-normal vision.

3.2. Stimuli

The stimuli consisted of a questionnaire and a recognition test. The questionnaire comprised four questions to assess observers' judgements of their perceived ability for recognizing family, famous and unfamiliar faces (How good do you think you are at recognizing the faces of your family/famous people/unfamiliar faces that you have only seen once before/unfamiliar faces that you have seen several times?). In response to these questions, participants rated their ability from "very bad" to "very good" on seven-point Likert scales.

The stimuli for the recognition test consisted of 40 trials of a line-up task. On each trial of this task, a single unfamiliar target face was presented in the screen centre and was followed by an identity line-up of 10 faces. The target and line-up faces were shown in greyscale on a white background, with a neutral expression, and in a frontal view. Each face image measured approximately 7 cm × 6.5 cm. The target face was present in half of the identity line-ups (20 trials) and absent in the others (20 trials). In addition, on target-present trials, either an identical face image was used for the initial target and for the corresponding image in the identity line-up, to create the same-image condition, or two separate images were used for the different-image condition (for more information, see Bruce et al., 1999).

3.3. Procedure

Participants were administered the questionnaire to rate their recognition abilities for familiar and unfamiliar faces. They were then allocated randomly to one of the two face recognition tests, which were displayed on a desktop computer. Each trial began with a 1-s fixation cross, which was followed by a face target. Observers were asked to study the targets until they felt that they could identify them from a subsequent line-up. The target faces were then replaced with an identity line-up which was displayed until a response was made. Participants were asked to decide whether the target is present in the line-up, and if so, to press the corresponding number key on a standard computer keyboard (e.g. “1” for face 1) or to press “A” if the target was absent. In the same-image condition, participants were advised that the identical image to the target face would be present in the corresponding line-up. Similarly, in the different-image condition, participants were informed that two different images of the same person would be used for the initial target and its counterpart in the subsequent line-up. In this way, each participant completed 20 target-present and 20 target-absent trials in a random order.

On completion of these tasks, participants were given on-screen feedback for their performance in the form of the percentage correct responses. In addition, participants in the same-image condition were told that they had performed very well in the recognition task, whereas participants in the different-image condition were told that they had not performed well. This feedback was administered to further strengthen participants’ belief in their respective recognition abilities, as generated by the face tests. The questionnaire was then completed again.

4. Results

4.1. A priori ability judgements

In a first step of the analysis, observers’ a priori ability ratings were analysed for each of the questionnaire items. The aim of this analysis was to explore whether observers would conflate the recognition of these categories. If so, then ability ratings should correlate for familiar (family and famous) and unfamiliar faces.

The data shows that these ratings were close to ceiling for family faces ($M = 6.53$, $SD = 0.75$), and were also higher for famous faces ($M = 5.30$, $SD = 1.15$) than unfamiliar faces that have been seen several times ($M = 5.00$, $SD = 0.96$) or only once ($M = 4.03$, $SD = 1.18$). The overall pattern of these ratings, therefore, corresponds to the relative familiarity of the different face categories. Of greater interest was whether these judgements would correlate across different face categories. Ability judgements for family faces correlated with famous faces, $r(58) = 0.322$, $p < 0.05$, but not with unfamiliar-seen-once faces, $r(58) = 0.191$, $p = 0.143$, and unfamiliar-seen-several-times faces, $r(58) = 0.166$, $p = 0.205$. By contrast, ability judgements for unfamiliar-seen-once and seen-several-times faces correlated strongly, $r(58) = 0.646$, $p < 0.001$. A correlation between famous and unfamiliar-seen-once faces was not found, $r(58) = 0.242$, $p = 0.063$, but ability judgements for famous faces also correlated with unfamiliar-seen-several-times faces, $r(58) = 0.292$, $p < 0.05$. Taken together, these results suggest that observers tend to associate their abilities to recognize familiar and unfamiliar faces to some extent, particularly for famous faces and unfamiliar faces that have been seen several times. For a summary of all correlations, see Table 1.

Table 1. Correlations of a priori ability judgements for family, famous and unfamiliar faces in Experiment 1

	Family	Famous	Unfamiliar—Once	Unfamiliar—Several
Family	–	$r = 0.322, p = 0.012$	$r = 0.191, p = 0.143$	$r = 0.166, p = 0.205$
Famous	*	–	$r = 0.242, p = 0.063$	$r = 0.292, p = 0.024$
Unfamiliar—Once	n.s.	n.s.	–	$r = 0.646, p = 0.000$
Unfamiliar—Several	n.s.	*	***	–

* $p < 0.05$.

*** $p < 0.001$.

4.2. A priori ability judgements and face recognition accuracy

The next step of the analysis explored the extent to which ability judgements predict performance on the face test. Specifically, we sought to examine whether such judgements to unfamiliar faces would correlate with recognition accuracy on the face test. For this purpose, we first analysed performance for the line-up tasks. This was calculated for target-present (correct identifications of the target) and target-absent trials (correct rejections of the line-up). In the same-image condition, 90.7% correct identifications (SD = 9.4) and 83.2% correct rejections (SD = 17.1) were recorded, compared to 58.3% correct identifications (SD = 17.9) and 59.7% correct rejections (SD = 25.4) for different-image trials. A 2 (same-image versus different-image condition) × 2 (correct identifications, correct rejections) ANOVA of this data showed a main effect of condition, $F(1,58) = 73.57, p < 0.001$, partial $\eta^2 = 0.56$, due to higher accuracy in the same-image condition. A main effect of line-up type, $F(1,58) = 0.80, p = 0.37$, partial $\eta^2 = 0.01$ and interaction were not found, $F(1,58) = 1.65, p = 0.20$, partial $\eta^2 = 0.03$. The recognition test was, therefore, effective in manipulating the difficulty of this task.

We also recorded considerable individual differences on both versions of the line-up task. In the same-image condition, for example, individual performance ranged from 63 to 100% accuracy (means of correct identifications and rejections), and from 33 to 85% for the different-image line-up displays. Broad individual differences were also evident in the initial ability judgements. For unfamiliar faces-seen-once, for example, these ratings ranged from one to seven on the seven-point scale. Despite this variation, these ability ratings correlated poorly with performance on the face test. For example, judgements for unfamiliar-seen-once or seen-several-times faces did not correlate with correct line-up identifications or correct rejections in the different-image condition. And in the same-image condition, only one of these correlations, of judgements for faces-seen-once and correct rejections, approached significance (for a summary of all correlations, see Table 2). This pattern also persisted when these ratings were combined for the two face test conditions, which showed no correlation between ability ratings and correct identifications or correct rejections for unfamiliar faces seen once, $r(58) = -0.219, p = 0.093$ and $r(58) = 0.186, p = 0.155$, or seen several times, $r(58) = -0.119, p = 0.365$ and $r(58) = 0.004, p = 0.978$. This indicates that observers were generally poor at predicting their actual ability to identify unfamiliar faces.

4.3. A posteriori ability judgements and face recognition accuracy

Considering that a priori ability ratings poorly predicted performance on the face test, it is important to establish whether such a relationship can be found at all. For this purpose, ability judgements

Table 2. Correlations of ability judgements and recognition accuracy in Experiment 1, prior to and after the face test

	Same-image condition		Different-image condition	
	Correct identifications	Correct rejections	Correct identifications	Correct rejections
<i>Before face test</i>				
Family	$r = 0.217, p = 0.248$	$r = -0.030, p = 0.876$	$r = 0.145, p = 0.445$	$r = 0.026, p = 0.890$
Famous	$r = 0.288, p = 0.122$	$r = 0.016, p = 0.933$	$r = -0.043, p = 0.823$	$r = 0.261, p = 0.163$
Unfamiliar—Once	$r = 0.109, p = 0.567$	$r = 0.359, p = 0.051$	$r = -0.310, p = 0.096$	$r = 0.259, p = 0.166$
Unfamiliar—Several	$r = 0.051, p = 0.787$	$r = -0.215, p = 0.253$	$r = -0.210, p = 0.265$	$r = 0.133, p = 0.484$
<i>After face test</i>				
Family	$r = 0.081, p = 0.670$	$r = 0.243, p = 0.195$	$r = 0.025, p = 0.896$	$r = 0.206, p = 0.275$
Famous	$r = 0.074, p = 0.697$	$r = 0.145, p = 0.444$	$r = 0.017, p = 0.927$	$r = 0.354, p = 0.055$
Unfamiliar—Once	$r = 0.037, p = 0.846$	$r = 0.567, p = 0.001^{***}$	$r = 0.128, p = 0.500$	$r = 0.327, p = 0.078$
Unfamiliar—Several	$r = 0.132, p = 0.485$	$r = 0.486, p = 0.007^{**}$	$r = 0.246, p = 0.191$	$r = 0.404, p = 0.027^*$

* $p < 0.05$.

** $p < 0.01$.

*** $p < 0.001$.

were also compared with recognition performance after feedback had been provided. At this stage, ability judgements related to identification performance to some extent, as judgements for unfamiliar-seen-several-times faces correlated with correct rejections in the same-image, $r(28) = 0.404$, $p < 0.05$, and the different-image condition, $r(28) = 0.486$, $p < 0.01$. In addition, such a correlation was also found with ability judgements for unfamiliar-seen-once faces and same-image displays, $r(28) = 0.567$, $p < 0.001$. However, none of the analogous correlations with correct identifications was reliable (see Table 2). In contrast, both correct rejections and correct identifications correlated strongly with ability judgements for unfamiliar-seen-once and seen-several-times faces when the data from the same-image and different-image conditions was combined, all $r_s \geq 0.558$, all $p_s \leq 0.001$.

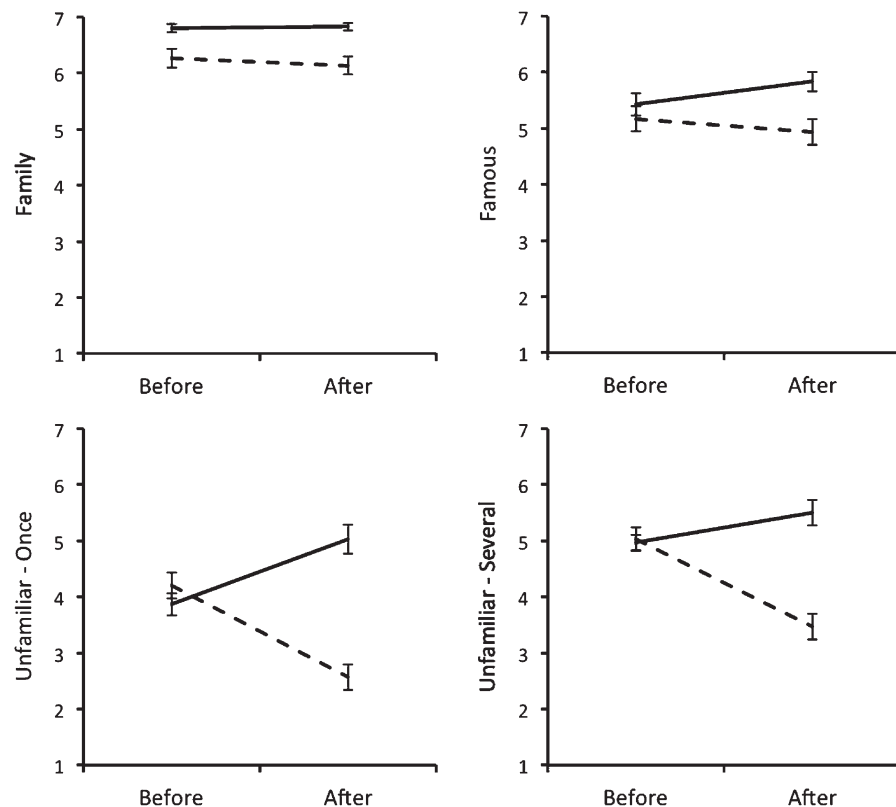
4.4. Change in ability ratings

Finally, to further assess whether observers tend to conflate the recognition of familiar and unfamiliar faces, we also explored the change in ability ratings prior to and after the face test more directly. Specifically, we sought to investigate whether the different conditions of the face test only affected a posteriori ability ratings for unfamiliar faces or whether this also influenced judgements to famous and family faces. For this analysis, we calculated the mean ratings for each of the questionnaire items (see Figure 1). These show that the face test did not affect observers' ratings to process family faces but influenced ratings for famous and unfamiliar faces. In these categories, ability ratings were matched evenly in the same- and different-image conditions prior to the face test but increased thereafter in the former and declined in the latter.

A series of four 2 (condition: same-image, different-image) \times 2 (time: before versus after the face test) ANOVAs for the questionnaire items did not show a main effect of time for unfamiliar-seen-once faces, $F(1,58) = 1.45$, $p = 0.23$, partial $\eta^2 = 0.02$, but a main effect of condition was found, $F(1,58) = 16.63$, $p < 0.001$, partial $\eta^2 = 0.22$, and an interaction between factors, $F(1,58) = 52.01$, $p < 0.001$, partial $\eta^2 = 0.47$. Analysis of simple main effects showed that ratings were evenly matched

Figure 1. Recognition ability ratings before and after the unfamiliar face test in Experiment 1 for family and famous faces, and for unfamiliar faces seen once and several times.

Notes: Solid line denotes same-image condition, dashed line denotes different-image condition. Vertical bars represent the standard error of the means.



across conditions prior to the face test, $F(1,58) = 1.05$, $p = 0.31$, partial $\eta^2 = 0.01$, but differed afterwards, $F(1,58) = 57.35$, $p < 0.001$, partial $\eta^2 = 0.33$, due to an improvement in the same-image condition, $F(1,58) = 18.06$, $p < 0.001$, partial $\eta^2 = 0.24$, and a decline in the different-image condition, $F(1,58) = 35.40$, $p < 0.001$, partial $\eta^2 = 0.38$.

A similar pattern was observed for ratings to unfamiliar-seen-several-times faces, which also showed a main effect of condition, $F(1,58) = 18.59$, $p < 0.001$, partial $\eta^2 = 0.24$, and time, $F(1,58) = 8.23$, $p < 0.01$, partial $\eta^2 = 0.12$, and an interaction between factors, $F(1,58) = 34.00$, $p < 0.001$, partial $\eta^2 = 0.37$. Once again, this interaction reflects the fact that ratings were evenly matched across conditions prior to the face test, $F(1,58) = 0.05$, $p = 0.82$, partial $\eta^2 = 0.00$, but differed thereafter, $F(1,58) = 48.97$, $p < 0.001$, partial $\eta^2 = 0.30$, due to an increase in the same-image condition, $F(1,58) = 4.39$, $p < 0.05$, partial $\eta^2 = 0.07$, and the reverse effect in the different-image condition, $F(1,58) = 37.85$, $p < 0.001$, partial $\eta^2 = 0.40$. These results, therefore, indicate that the recognition test influenced observers' perception of their recognition ability for unfamiliar faces.

The same analysis was conducted for judgements for famous and family faces. For famous faces, ANOVA did not show a main effect of time, $F(1,58) = 0.63$, $p = 0.43$, partial $\eta^2 = 0.01$, but of condition, $F(1,58) = 4.55$, $p < 0.05$, partial $\eta^2 = 0.07$, and an interaction between factors, $F(1,58) = 9.05$, $p < 0.01$, partial $\eta^2 = 0.14$. Again, analysis of simple main effects showed that ratings were evenly matched across conditions prior to the face test, $F(1,58) = 0.83$, $p = 0.37$, partial $\eta^2 = 0.01$, but differed thereafter, $F(1,58) = 9.42$, $p < 0.01$, partial $\eta^2 = 0.08$, due to an improvement in the same-image condition, $F(1,58) = 7.22$, $p < 0.01$, partial $\eta^2 = 0.11$. By contrast, performance was more comparable prior to and after the face test in the different-image condition, $F(1,58) = 2.46$, $p = 0.12$, partial $\eta^2 = 0.04$.

The same analysis also revealed a main effect of condition for ability ratings for family faces, $F(1,58) = 13.76$, $p < 0.001$, partial $\eta^2 = 0.19$, due to observers generally rating their recognition ability higher in the same-image condition. However, no main effect of time, $F(1,58) = 0.53$, $p = 0.47$, partial $\eta^2 = 0.01$, and no interaction were found, $F(1,58) = 0.20$, $p = 0.66$, partial $\eta^2 = 0.03$.

5. Discussion

This experiment showed that observers' judgement of their ability to process unfamiliar faces poorly predicts their accuracy in a recognition test for such faces. By contrast, such associations were found after participants had been given feedback for their recognition performance. This indicates that it is not generally impossible to find such correlations. Instead, these findings suggest that observers might not receive such feedback outside of the laboratory. As a consequence, observers might be poor initially at judging their own face recognition ability, with the possibility of improving awareness of ability after such feedback is provided. This notion is consistent with other recent studies, which have shown that feedback can enhance unfamiliar face identification (Alenezi & Bindemann, 2013; White et al., 2014).

We also investigated whether observers might conflate their ability judgements for unfamiliar faces with famous faces. We obtained some evidence for this, with an association in ability ratings between famous faces and unfamiliar-seen-several-times faces. In addition, observers also adjusted judgements of their face identification ability after the recognition test. As expected, performance was better in the same-image version of this test than the different-image condition, and observers subsequently rated their recognition abilities for unfamiliar faces according to the difficulty of these conditions. Remarkably, however, a similar pattern was also observed for famous faces. This provides additional evidence to suggest that observers conflate their abilities to recognize familiar and unfamiliar faces. These results therefore indicate that observers' judgement of their face recognition ability is malleable and can be altered after only a short recognition test.

The findings of this experiment raise the question of whether observers are generally poor at predicting their face recognition performance or whether this is confined to unfamiliar faces. In addition, the question also arises of whether observers only generalize their recognition

performance for unfamiliar faces to inform judgements of their recognition ability for famous faces, or whether the reverse effect is also found. These questions were explored in a further experiment.

6. Experiment 2

Experiment 1 suggests that observers are poor at estimating their ability to recognize unfamiliar faces. Moreover, the difficulty of an unfamiliar face recognition test can affect observers' judgement of their ability to recognize famous faces, which suggests that observers can conflate these processes. The next experiment sought to examine whether observers are also poor at predicting their ability to recognize familiar faces, and whether this relationship can be strengthened subsequently by providing feedback. Additionally, we examined whether performance from a recognition test for famous faces would, in turn, affect ability judgements for unfamiliar faces.

As in Experiment 1, observers' beliefs about their face recognition abilities were assessed with a set of four questions. Participants were then shown "current" or "before they were famous" (BTWF) faces of famous people in a recognition test. The latter manipulation, of using photographs of famous individuals before they became widely known, typically when they were children or adolescents, can be used to make the recognition of familiar faces more challenging (Russell, Duchaine, & Nakayama, 2009). Similar to the line-up tasks of Experiment 1, these conditions were designed to induce a feeling of competence in face recognition in the current condition or of relative incompetence in the BTWF condition. Observers then rated their recognition abilities again to determine how these judgements were influenced by the recognition test.

7. Method

7.1. Participants

Sixty undergraduate students (51 female) in the School of Psychology at the University of Kent, with a mean age of 20 years (range = 18 to 25), participated in this experiment for course credit. None had participated in Experiment 1. All reported normal or corrected-to-normal vision.

7.2. Stimuli

Observers' beliefs about their face recognition abilities were assessed with the same questionnaire as in Experiment 1, but the stimuli for the recognition test now consisted of photographs of 40 famous faces (A name list of these famous people can be viewed in Appendix A). Each face was shown in a frontal view at a size of 7 × 7 cm. Two photographs of each face were used, which consisted either of a recent photograph for the current condition or a photograph of the same person as a child or adolescent for the BTWF condition.

7.3. Procedure

Participants began the experiment by completing the questionnaire. They were then allocated randomly to one of the recognition tests, using current or BTWF faces. In both conditions, participants attempted to identify 40 famous people by name or a unique semantic description (e.g. a combination of nationality and occupation), which were displayed in a booklet at a rate of one face per page. An experimenter recorded participants' responses.

Upon completion of the face test, participants were informed of their recognition performance (in % accuracy). In the current condition, participants were also told they had performed well, whereas in the BTWF condition they were told they had performed poorly. This feedback was administered to strengthen the impression of good or bad recognition competence that we aimed to generate with the face test. Participants then completed the questionnaire for a second time. Finally, a familiarity check was administered, which consisted of a list of the names of the famous faces. Participants were asked to indicate which of these faces they knew.

8. Results

8.1. A priori ability judgements

As in Experiment 1, a priori ability ratings were analysed first for each of the questionnaire items to explore whether observers conflate recognition of the different face categories. Once again, these ratings were close to ceiling for family faces ($M = 6.58$, $SD = 0.65$), and were higher for famous faces ($M = 4.55$, $SD = 1.10$) than unfamiliar faces that have been seen several times ($M = 4.30$, $SD = 1.23$) or only once ($M = 3.42$, $SD = 1.37$). Ability judgements did not correlate between family and famous faces, $r(58) = 0.162$, $p = 0.217$, family and unfamiliar-seen-once faces, $r(58) = 0.066$, $p = 0.619$, and family and unfamiliar-seen-several-times faces, $r(58) = 0.204$, $p = 0.119$. By contrast, ability judgements for unfamiliar faces that had only been seen once or several times correlated strongly, $r(58) = 0.823$, $p < 0.001$. In addition, ability judgements for famous faces correlated with unfamiliar-seen-several-times faces, $r(58) = 0.304$, $p < 0.05$, but a correlation between famous and unfamiliar-seen-once faces was not found, $r(58) = 0.217$, $p = 0.095$. As in Experiment 1, these results, therefore, suggest that observers associate their abilities to recognize familiar and unfamiliar faces, particularly famous faces and unfamiliar faces that have been seen several times. For a summary of these correlations, see Table 3.

8.2. A priori ability judgements and face recognition accuracy

We then sought to determine whether these ability judgements predict performance on the famous face recognition test. We first analysed recognition performance. This analysis only included faces that observers knew, as indicated by the familiarity check. This led to the exclusion of 8.7% of trials for the current condition and 7.7% for the BTWF condition. Recognition accuracy for the remaining trials was at 77.3% for the current condition ($SD = 22.3$) and 28.1% for the BTWF condition ($SD = 13.4$). An independent-samples t-test showed that this difference was reliable, $t(58) = 10.36$, $p < 0.001$, $d = 2.68$. This indicates that the conditions were effective in manipulating the difficulty of this task.

To determine whether observers could predict their recognition accuracy, ability judgements were then correlated with individual performance. This revealed reliable correlations between recognition accuracy and observers ability judgements to recognize famous faces in the current and the BTWF condition, $r(28) = 0.472$, $p < 0.01$ and $r(28) = 0.392$, $p < 0.01$, respectively. In contrast, ability

Table 3. Correlations of a priori ability judgements for family, famous and unfamiliar faces in Experiment 2

	Family	Famous	Unfamiliar—Once	Unfamiliar—Several
Family	–	$r = 0.162$, $p = 0.217$	$r = 0.066$, $p = 0.619$	$r = 0.204$, $p = 0.119$
Famous	n.s.	–	$r = 0.217$, $p = 0.095$	$r = 0.304$, $p = 0.018$
Unfamiliar—Once	n.s.	n.s.	–	$r = 0.823$, $p = 0.000$
Unfamiliar—Several	n.s.	*	***	–

* $p < 0.05$.

*** $p < 0.001$.

Table 4. Correlations of ability judgements and recognition accuracy in Experiment 2, prior to and after the face test

	Before face test		After face test	
	Current condition	BTWF condition	Current condition	BTWF condition
Family	$r = 0.216$, $p = 0.252$	$r = 0.133$, $p = 0.485$	$r = 0.413$, $p = 0.023^*$	$r = 0.125$, $p = 0.509$
Famous	$r = 0.472$, $p = 0.009^{**}$	$r = 0.392$, $p = 0.032^*$	$r = 0.796$, $p = 0.000^{***}$	$r = 0.527$, $p = 0.003^{**}$
Unfamiliar—Once	$r = -0.018$, $p = 0.925$	$r = -0.157$, $p = 0.408$	$r = 0.101$, $p = 0.596$	$r = -0.127$, $p = 0.504$
Unfamiliar—Several	$r = -0.096$, $p = 0.613$	$r = 0.063$, $p = 0.742$	$r = 0.058$, $p = 0.762$	$r = -0.005$, $p = 0.980$

* $p < 0.05$.

** $p < 0.01$.

*** $p < 0.001$.

judgements for family and unfamiliar faces did not relate to recognition performance, all $r_s \leq 0.216$, $p_s \geq 0.252$ (for a summary of all correlations, see Table 4). This pattern persisted when these ratings were combined for the current and BTWF conditions. This analysis also revealed a correlation between recognition accuracy and ability judgements for famous faces, $r(58) = 0.274$, $p < 0.05$, but not for family or unfamiliar faces, all $r_s \leq 0.156$, $p_s \geq 0.232$.

8.3. A posteriori ability judgements and face recognition accuracy

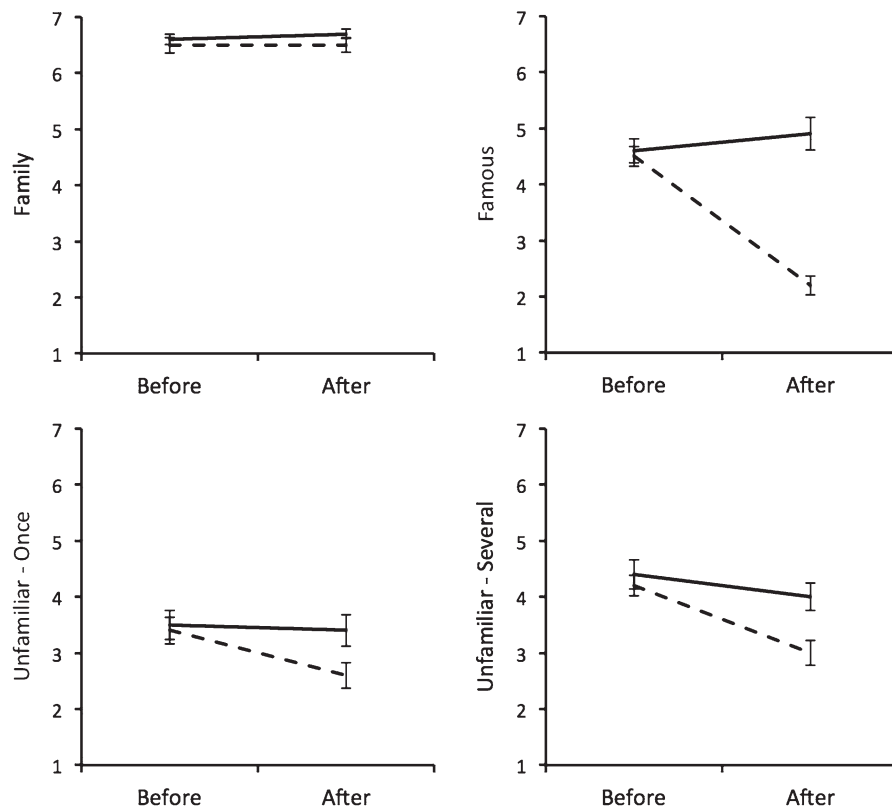
A similar pattern was obtained when the post-test ability judgements were compared with recognition accuracy. Once again, reliable correlations were found between accuracy and ability judgements to recognize famous faces in the current condition, $r(28) = 0.796$, $p < 0.001$, and the BTWF condition, $r(28) = 0.527$, $p < 0.01$. In addition, ability judgements for family faces also correlated with recognition accuracy in the current condition, $r(28) = 0.413$, $p < 0.05$. None of the other correlations reached significance, all $r_s \leq 0.127$, $p_s \geq 0.504$ (for a summary of correlations, see Table 4). This pattern persisted when these ratings were combined for the current and BTWF conditions, which also showed a correlation between recognition accuracy and ability judgements for famous faces, $r(58) = 0.880$, $p < 0.001$. In addition, this analysis also revealed a correlation between recognition accuracy and ability judgements for family faces, $r(58) = 0.255$, $p < 0.05$, and unfamiliar faces seen-several-times, $r(58) = 0.315$, $p < 0.05$, but not for unfamiliar faces seen-once, $r(58) = 0.229$, $p = 0.079$.

8.4. Change in ability ratings

Once again, the questionnaire ratings before and after the face test were also compared directly to explore further whether observers tend to conflate the recognition of familiar and unfamiliar faces. The cross-subject means of these ratings are shown in Figure 2. All ratings were initially matched across the two conditions of the face test. In the current condition, these ratings also appear comparable prior to and after the administration of the face test. By contrast, participants reported a substantial drop in ability in the BTWF condition after the face test. This was most pronounced for famous faces but seems to generalize to unfamiliar faces.

Figure 2. Recognition ability ratings before and after the famous face test in Experiment 2 for family and famous faces, and for unfamiliar faces seen once and several times.

Notes: Solid line denotes current condition, dashed line denotes BTWF condition. Vertical bars represent the standard error of the means.



To analyse these changes, four 2 (condition: current, BTWF) \times 2 (time: before versus after the face test) ANOVAs were conducted for the questionnaire items. For ratings for family faces, this analysis did not find a main effect of time, $F(1,58) = 0.00$, $p = 1.00$, partial $\eta^2 = 0.00$, or condition, $F(1,58) = 0.73$, $p = 0.40$, partial $\eta^2 = 0.01$, and no interaction between factors, $F(1,58) = 0.49$, $p < 0.49$, partial $\eta^2 = 0.01$. In contrast, main effects of condition, $F(1,58) = 23.58$, $p < 0.001$, partial $\eta^2 = 0.29$, time, $F(1,58) = 60.65$, $p < 0.001$, partial $\eta^2 = 0.51$, and an interaction were found for ratings for famous faces, $F(1,58) = 100.98$, $p < 0.001$, partial $\eta^2 = 0.64$. Analysis of simple main effects showed that ability ratings were matched across conditions at the start of the experiment, $F(1,58) = 0.01$, $p = 0.92$, partial $\eta^2 = 0.00$, but were lower in the BTWF than the current condition after the face test, $F(1,58) = 75.29$, $p < 0.001$, partial $\eta^2 = 0.39$. Whereas ability ratings for famous faces were constant throughout the experiment in the current condition, $F(1,58) = 2.56$, $p = 0.12$, partial $\eta^2 = 0.04$, they declined after the face test in the BTWF condition, $F(1,58) = 159.07$, $p < 0.001$, partial $\eta^2 = 0.73$.

A similar pattern emerged with the remaining questionnaire items. For unfamiliar faces seen once, a main effect of condition was not observed, $F(1,58) = 1.63$, $p = 0.21$, partial $\eta^2 = 0.03$, but a main effect of time, $F(1,58) = 11.67$, $p < 0.01$, partial $\eta^2 = 0.17$, and an interaction were found, $F(1,58) = 8.10$, $p < 0.01$, partial $\eta^2 = 0.12$. Analysis of simple main effects revealed comparable ability ratings across the face test conditions at the start of the experiment, $F(1,58) = 0.08$, $p = 0.78$, partial $\eta^2 = 0.00$, but not after the administration of the face test, $F(1,58) = 4.55$, $p < 0.05$, partial $\eta^2 = 0.04$. This reflects a decrease in ability ratings in the BTWF condition, $F(1,58) = 19.60$, $p < 0.001$, partial $\eta^2 = 0.25$, whereas these ratings were matched evenly in the current condition prior to and after the face test, $F(1,58) = 0.16$, $p = 0.68$, partial $\eta^2 = 0.00$.

Observers' ratings for unfamiliar faces seen several times revealed an identical pattern. Again, a main effect of condition was not present, $F(1,58) = 3.80$, $p = 0.06$, partial $\eta^2 = 0.06$, but a main effect of time, $F(1,58) = 29.96$, $p < 0.01$, partial $\eta^2 = 0.34$, and an interaction were found, $F(1,58) = 8.79$, $p < 0.01$, partial $\eta^2 = 0.13$. Ability ratings were matched across the current and BTWF conditions at the start of the experiment, $F(1,58) = 0.17$, $p = 0.68$, partial $\eta^2 = 0.00$, but differed after the face test, $F(1,58) = 9.44$, $p < 0.01$, partial $\eta^2 = 0.08$. This reflects a decrease in ability ratings in the BTWF condition, $F(1,58) = 35.61$, $p < 0.001$, partial $\eta^2 = 0.38$, whereas these ratings were more comparable in the current condition prior to and after the face test, $F(1,58) = 3.15$, $p = 0.08$, partial $\eta^2 = 0.05$.

9. Discussion

In this experiment, observers' a priori ability ratings to recognize famous faces correlated with recognition accuracy for current and BTWF faces. This indicates that observers have some insight into their ability to process familiar faces that translates into actual recognition performance. In addition, however, this experiment also provides further evidence that observers tend to conflate their perceived abilities to process famous and unfamiliar faces. As in Experiment 1, a priori judgements of recognition abilities correlated for famous and unfamiliar-seen-several-times faces. In addition, the different conditions of the famous face recognition test not only affected subsequent ability judgements for famous faces, but also produced a knock-on effect for unfamiliar faces. This indicates that observers generalized their recognition performance for famous faces to inform judgements of their recognition ability for unfamiliar faces.

10. General discussion

While the identification of unfamiliar faces is a difficult task (see, e.g. Bruce et al., 1999; Johnston & Bindemann, 2013; Memon et al., 2011), it remains unresolved why observers are prone to making identification errors. This study investigated a potential explanation for this phenomenon, by assessing whether observers are poor at judging their own ability to identify unfamiliar faces. We also explored whether observers might conflate the recognition of familiar and unfamiliar faces, by generalizing the ability to process one type of stimulus to the other.

Experiment 1 showed that ability judgements poorly predicted performance in a test of unfamiliar face recognition. Indeed, only one correlation, between ability judgements and correct line-up rejections in the same-image condition, approached significance. This condition was included here to provide a comparatively easy version of the face test, and to manipulate observers' perception of the difficulty of the task. Generally, however, the problem of unfamiliar face recognition in applied settings is the recognition of *different* instances of the same face (for an illustration, see, e.g. Jenkins & Burton, 2011). The same-image condition, therefore, provides only a poor proxy to the actual problem of unfamiliar face identification. Consequently, the rather moderate, and only, correlation between a priori ability judgements and correct line-up rejections in the same-image condition is also of limited interest here.

Considering that a priori ability judgements predicted unfamiliar face identification poorly, it is noteworthy that stronger correlations were obtained after the recognition test in Experiment 1. Moreover, we also found that observers could predict their recognition performance for famous faces in Experiment 2. These findings indicate that the a priori ability judgements for unfamiliar face identification do not fail to relate to actual performance in Experiment 1 because such associations cannot be found generally. Instead, these findings suggest that observers *initially* had limited insight into aspects of recognition ability that relate specifically to unfamiliar faces.

A possible explanation for this finding is that we rarely receive feedback for errors in unfamiliar face identification outside of the laboratory. As a consequence, observers might be poor at judging their own recognition ability. This notion is consistent with other recent studies, which have shown that accuracy is higher in unfamiliar face matching when performance feedback is administered (Alenezi & Bindemann, 2013; White et al., 2014). We also suggest that the presence of such feedback for familiar faces outside of the laboratory could explain why observers could predict their performance on the famous face recognition test. In social interaction, successful person identification is self-evident from the reaction of other people. Identification feedback for famous people might be even more explicit. Famous faces in the media are, for example, often accompanied by additional identity-related information, such as names and semantic information, to confirm recognition. If observers utilize this information to inform judgements of their own recognition ability, then one would expect to obtain a correlation between perceived and actual recognition ability for famous people (as in Experiment 2) but not for unfamiliar faces (as in Experiment 1).

Considering that observers should have a clearer notion of their ability to process famous than unfamiliar faces, we also wondered whether they might draw on the former to inform judgements of the latter. We obtained several lines of evidence for this. For example, while initial ability judgements did not correlate between famous and once-seen unfamiliar faces, they were associated with famous and unfamiliar-seen-several-times faces. While the identification of familiar faces appears to be qualitatively different from unfamiliar faces (Megreya & Burton, 2006), the correlation between these categories makes good sense when familiarity is viewed as a continuum. On this continuum, famous faces are not *as* familiar as family faces, whereas unfamiliar faces that have been seen several times are *more* familiar than once-seen unfamiliar faces. In the current experiments, famous and unfamiliar-seen-several-times faces, therefore, lie adjacent along the familiarity continuum *and* straddle the boundary between "familiar" and "unfamiliar" face recognition. We also note that initial ability judgements for these face categories were similar in both experiments (e.g. at 5.3 and 5.0 in Experiment 1 and 4.6 and 4.3 in Experiment 2 for famous and unfamiliar faces, respectively), which suggests further that observers might perceive their abilities to process these stimuli to be quite comparable.

A comparison of ability judgements prior to and after the recognition tests also indicates that observers tend to relate familiar and unfamiliar face recognition more generally. In both experiments, the face tests did not affect recognition ability ratings for family members, which were consistently close to ceiling. However, the unfamiliar face test influenced how observers viewed their ability to recognize famous faces in Experiment 1, by producing a decrease in these ability

ratings in the more difficult face test condition. Experiment 2 then revealed a similar pattern after an identification task for famous faces, whereby ability ratings were larger for unfamiliar faces after a relatively easy recognition test than in the more difficult condition.

While these changes in ability ratings indicate that observers tend to generalize the judgement of their recognition abilities across famous and unfamiliar faces, it is also notable that the largest changes in ability ratings were observed within face categories. This pattern converges with the correlations of the initial ability judgements for familiar and unfamiliar faces, which were present only for famous and unfamiliar-seen-several-times faces. Both sets of findings, therefore, suggest that observers conflate ability judgements for familiar and unfamiliar faces, but only do so partially.

We conducted these experiments to explore further why the identification of unfamiliar faces is so error-prone in experimental (e.g. Bruce et al., 1999; Burton et al., 2010; Megreya & Burton, 2006) and applied settings (e.g. Kemp et al., 1997; Memon et al., 2011). A range of factors have now been identified that can make this task difficult, but these focus primarily on extraneous influences that affect the appearance of a face. In addition, however, theories of unfamiliar face identification also need to explain why observers are willing to make (incorrect) identifications despite the difficulty of this task. The exploration of a priori ability ratings and their relationship with subsequent face recognition performance suggest that such errors might occur because observers have little insight into their own accuracy in this task. As a consequence, observers might be unaware of the likelihood that identification errors might be made and commit such mistakes more readily.

It is less clear from these data whether observers might also be prone to making identification errors because they overgeneralize their ability to process familiar faces to unfamiliar people. The current study revealed an association for a priori ability judgements to famous and unfamiliar-seen-several-times faces in both experiments and showed also that observers conflate these processes to some extent when they are given performance feedback for only one of these tasks (i.e. unfamiliar face identification in Experiment 1 and famous face recognition in Experiment 2). However, while such an effect was found in the overall ability ratings (see Figures 1 and 2), it was only partially evident from correlations of the initial ratings for the different face categories, as this effect was not present for famous and unfamiliar-seen-once faces (see Tables 1 and 3). In addition, we also found that a priori ability judgements for famous faces do not relate directly to unfamiliar face identification accuracy (see Table 1). Thus, it is not simply the case that people who *think* they are good at recognizing familiar faces are also less (or more) prone to making errors in unfamiliar face identification. The current experiments, therefore, suggest that the relationship between familiar and unfamiliar face recognition is one of *perceived ability* rather than *actual accuracy*.

We draw these conclusions with some obvious caveats. The current experiments are, for example, dependent on the tests that were used to measure face recognition performance. It is conceivable that other tests might reveal stronger links between observers' initial ability judgements and their recognition accuracy. Similarly, it is possible that better measures can be found to assess observers' judgement of their face recognition abilities than the simple scales that we have devised here. We also note that feedback for task performance (i.e. % accuracy) and verbal feedback (e.g. "You have performed/not performed well in this task") were confounded in the current study. Consequently, some participants might have perceived these feedback types to provide conflicting information, for example, when accuracy in the face test appears to be low but they are told to have done well.

As the feedback was provided specifically to induce a sense of relative competence or incompetence in facial recognition, we adopted this combined feedback approach for several reasons. One of these is that observers could have no advance knowledge of the performance level that constitutes good accuracy in the face tasks. The provision of a simple accuracy score, without additional

contextual information, might therefore provide insufficient feedback to observers about their face-processing competence. The face conditions also varied in difficulty in both experiments and it was difficult to determine in advance whether a percentage score in one condition constitutes better performance than in the other. It is unclear, for example, whether an observer with an accuracy of 70% in the more difficult different-image condition in Experiment 1 is better at face recognition than someone who achieves 80% in the easier same-image condition. For these reasons, we decided to supplement the percentage accuracy scores with verbal feedback for the questions under investigation here, and the results show that this combined approach was effective (see Figures 1 and 2). For future research, it would be interesting to determine which feedback type exerts greater influence on observers' ability ratings.

Such investigations could also examine whether feedback influences observers' beliefs about their face recognition abilities indirectly. It is possible, for example, that verbal feedback influences these beliefs via personality variables, such as failure-related action orientation, if negative feedback (e.g. "You have not performed well in this task") evokes anxiety or agitation (see Kuhl, 1994a, 1994b). While these might be interesting avenues for further research, it is notable that limited research continues to exist in first place regarding the accuracy of observers' judgement of their face perception abilities, both for familiar and unfamiliar faces. This is surprising given its potential applied value (e.g. for forensic identification tasks) and clinical relevance (e.g. for determining recognition impairments). Our study only provides a starting point here.

Funding

The authors received no direct funding for this research.

Author details

Markus Bindemann¹

E-mail: m.bindemann@kent.ac.uk

Janice Attard¹

E-mail: j.attard@kent.ac.uk

Robert A. Johnston¹

E-mail: r.a.johnston@kent.ac.uk

¹ School of Psychology, University of Kent, Canterbury CT2 7NP, UK.

Citation information

Cite this article as: Perceived ability and actual recognition accuracy for unfamiliar and famous faces, M. Bindemann, J. Attard & R.A. Johnston, *Cogent Psychology* (2014), 1: 986903.

References

- Alezei, H. M., & Bindemann, M. (2013). The effect of feedback on face-matching accuracy. *Applied Cognitive Psychology*, 27, 735–753. doi:10.1002/acp.2968
- Bahrick, H. P., Bahrick, O. O., & Wittlinger, R. P. (1975). Fifty years of memory for names and faces: A cross-sectional approach. *Journal of Experimental Psychology: General*, 104, 54–75. doi:10.1037/0096-3445.104.1.54
- Bindemann, M., Attard, J., Leach, A., & Johnston, R. A. (2013). The effect of image pixelation on unfamiliar-face matching. *Applied Cognitive Psychology*, 27, 707–717. doi:10.1002/acp.2970
- Bindemann, M., Avetisyan, M., & Blackwell, K.-A. (2010). Finding needles in haystacks: Identity mismatch frequency and facial identity verification. *Journal of Experimental Psychology: Applied*, 16, 378–386. doi:10.1037/a0021893
- Bindemann, M., Avetisyan, M., & Rakow, T. (2012). Who can recognize unfamiliar faces? Individual differences and observer consistency in person identification. *Journal of Experimental Psychology: Applied*, 18, 277–291. doi:10.1037/a0029635
- Bindemann, M., & Sandford, A. (2011). Me, myself and I: Different recognition rates for three photo-IDs of the same person. *Perception*, 40, 625–627. doi:10.1068/p7008
- Bruce, V. (1982). Changing faces: Visual and non-visual coding processes in face recognition. *British Journal of Psychology*, 73, 105–116. doi:10.1111/j.2044-8295.1982.tb01795.x
- Bruce, V., Henderson, Z., Greenwood, K., Hancock, P. J. B., Burton, A. M., & Miller, P. (1999). Verification of face identities from images captured on video. *Journal of Experimental Psychology: Applied*, 5, 339–360. doi:10.1037/1076-898X.5.4.339
- Burton, A. M. (2013). Why has research on face recognition progressed so slowly? The importance of variability. *The Quarterly Journal of Experimental Psychology*, 66, 1467–1485. doi:10.1080/17470218.2013.800125
- 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.
- Burton, A. M., Jenkins, R., & Schweinberger, S. R. (2011). Mental representations of familiar faces. *British Journal of Psychology*, 102, 943–958. doi:10.1111/j.2044-8295.2011.02039.x
- Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow face matching test. *Behavior Research Methods*, 42, 286–291. doi:10.3758/BRM.42.1.286
- Burton, A. M., Wilson, S., Cowan, M., & Bruce, V. (1999). Face recognition in poor-quality video: Evidence from security surveillance. *Psychological Science*, 10, 243–248. doi:10.1111/1467-9280.00144
- Hancock, P. J. B., Bruce, V., & Burton, A. M. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Sciences*, 4, 330–337. doi:10.1016/S1364-6613(00)01519-9
- Jenkins, R., & Burton, A. M. (2008). Limitations in facial identification: The evidence. *Justice of the Peace*, 172, 4–6.
- Jenkins, R., & Burton, A. M. (2011). Stable face representations. *Philosophical Transactions of the Royal Society B*, 366, 1671–1683. doi:10.1098/rstb.2010.0379
- Johnston, R. A., & Bindemann, M. (2013). Introduction to forensic face matching. *Applied Cognitive Psychology*, 27, 697–699. doi:10.1002/acp.2963
- Johnston, R. A., & Edmonds, A. J. (2009). Familiar and unfamiliar face recognition: A review. *Memory*, 17, 577–596. doi:10.1080/09658210902976969
- Kemp, R., Towell, N., & Pike, P. (1997). When seeing should not be believing: Photographs, credit cards and fraud. *Applied*

- Cognitive Psychology*, 11, 211–222. doi:10.1002/(SICI)1099-0720(199706)11:3<211:AID-ACP430>3.0.CO;2-O
- Kuhl, J. (1994a). A theory of action and state orientation. In J. Kuhl & J. Beckmann (Eds.), *Volition and personality: Action versus state orientation* (pp. 9–46). Göttingen: Hogrefe & Huber.
- Kuhl, J. (1994b). Action and state orientation: Psychometric properties of the action control scales (ACS-90). In J. Kuhl & J. Beckmann (Eds.), *Volition and personality: Action versus state orientation* (pp. 47–59). Göttingen: Hogrefe & Huber.
- Longmore, C. A., Liu, C. H., & Young, A. W. (2008). Learning faces from photographs. *Journal of Experimental Psychology: Human Perception & Performance*, 34, 77–100. doi:10.1037/0096-1523.34.1.77
- Megreya, A. M., & Burton, A. M. (2006). Unfamiliar faces are not faces: Evidence from a matching task. *Memory & Cognition*, 34, 865–876. doi:10.3758/BF03193433
- Megreya, A. M., & Burton, A. M. (2008). Matching faces to photographs: Poor performance in eyewitness memory (without the memory). *Journal of Experimental Psychology: Applied*, 14, 364–372. doi:10.1037/a0013464
- Megreya, A. M., Sandford, A., & Burton, A. M. (2013). Matching images of faces taken on the same day or months apart: The limitations of photo-ID. *Applied Cognitive Psychology*, 27, 700–706. doi:10.1002/acp.2965
- Memon, A., Havard, C., Clifford, B., Gabbert, F., & Watt, M. (2011). A field evaluation of the VIPER system: A new technique for eliciting eyewitness identification evidence. *Psychology, Crime, & Law*, 17, 711–729. doi:10.1080/10683160903524333
- Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic Bulletin & Review*, 16, 252–257. doi:10.3758/PBR.16.2.252
- Shapiro, P. N., & Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological Bulletin*, 100, 139–156. doi:10.1037/0033-2909.100.2.139
- Weber, N., & Perfect, T. J. (2012). Improving eyewitness identification accuracy by screening out those who say they don't know. *Law and Human Behaviour*, 36, 28–36. doi:10.1037/h0093976
- Wells, G. L., Memon, A., & Penrod, S. D. (2006). Eyewitness evidence: Improving its probative value. *Psychological Science in the Public Interest*, 7, 45–75. doi:10.1111/j.1529-1006.2006.00027.x
- Wells, G. L., & Olsen, E. A. (2003). Eyewitness testimony. *Annual Review of Psychology*, 54, 227–295. doi:10.1146/annurev.psych.54.101601.145028
- White, D., Kemp, R. I., Jenkins, R., & Burton, A. M. (2014). Feedback training for facial image comparison. *Psychonomic Bulletin and Review*, 21, 100–106. doi:10.3758/s13423-013-0475-3

Appendix A

A name list of the famous faces used in Experiment 2.

50 Cent	Beyoncé	Brad Pitt	Britney Spears
Christina Aguilera	Daniel Radcliffe	David Beckham	David Bowie
David Hasselhoff	Eminem	Enrique Iglesias	George Clooney
George Michael	Jennifer Aniston	Johnny Depp	Jude Law
Julia Roberts	Kate Middleton	Kate Moss	Katy Perry
Keira Knightley	Kelly Brook	Kristen Stewart	Kylie Minogue
Lady Gaga	Leonardo DiCaprio	Mariah Carey	Meryl Streep
Nicole Kidman	Owen Wilson	Renée Zellweger	Rihanna
Robbie Williams	Robert Downey Jr	Robin Williams	Sarah Jessica Parker
Simon Cowell	Snoop Dogg	Uma Thurman	Wayne Rooney



© 2014 The Author(s). This open access article is distributed under a Creative Commons Attribution (CC-BY) 3.0 license.

You are free to:

Share — copy and redistribute the material in any medium or format

Adapt — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made.

You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

No additional restrictions

You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

