Viewers base estimates of face matching accuracy on their own familiarity: Explaining the photo-ID paradox

Kay L. Ritchie[1,2], Finlay G. Smith[1], Rob Jenkins[2], Markus Bindemann[3], David White[4], & A.Mike Burton[1,2]

[1] School of Psychology, University of Aberdeen, Scotland, UK
[2] Department of Psychology, University of York, England, UK
[3] School of Psychology, University of Kent, England, UK
[4] School of Psychology, University of New South Wales, Australia

Corresponding author: A. Mike Burton, Department of Psychology, University of York, York, YO10 5DD, United Kingdom. Email: mike.burton@york.ac.uk

Keywords

face matching; face processing; familiarity; perceptual prediction

Acknowledgements

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n.323262, and from the Economic and Social Research Council, UK [ES/J022950/1].

Abstract

Matching two different images of a face is a very easy task for familiar viewers, but much harder for unfamiliar viewers. Despite this, use of photo-ID is widespread, and people appear not to know how unreliable it is. We present a series of experiments investigating bias both when performing a matching task and when predicting other people's performance. Participants saw pairs of faces and were asked to make a same/different judgement, after which they were asked to predict how well other people, unfamiliar with these faces, would perform. In four experiments we show different groups of participants familiar and unfamiliar faces, manipulating this in different ways: celebrities in experiments 1 to 3 and personally familiar faces in experiment 4. The results consistently show that people match images of familiar faces more accurately than unfamiliar faces. However, people also reliably predict that the faces they themselves know will be more accurately matched by different viewers. This bias is discussed in the context of current theoretical debates about face recognition, and we suggest that it may underlie the continued use of photo-ID, despite the availability of evidence about its unreliability.

Introduction

It is now well-established that matching two images of a face is very difficult for an unfamiliar viewer (Bruce et al, 1999; Burton & Jenkins, 2011; Johnston & Edmonds, 2009). Across a variety of tasks (e.g. line-ups and pair-wise matching), viewers are highly error-prone, even when images are shown in high quality, and for an unlimited time (Clutterbuck & Johnston, 2002, 2004; Megreya & Burton, 2006. 2007). Furthermore, matching a live person to a photo is just as error-prone (Kemp Towell & Pike, 1997; Davis & Valentine, 2009; Megreya & Burton, 2008). This results in a paradox. If people are so poor at matching faces to photos, why do they continue to be used extensively in identification documents? One possibility is that the professional checkers of photo-ID are better able to make a match than the general population. Although there is rather little evidence about the performance of professional ID-checkers, what is available suggests that this is not true: Burton et al (1999) tested police officers, and White et al (2014) tested passport officers. Both studies showed the same levels of performance in the professional groups and in untrained students. An alternative explanation for the continued use of photo-ID is that people do not know how unreliable it is, and that this results from their own experience of face recognition.

In this paper we examine the possibility that people are subject to a systematic bias in face perception which arises from their excellent performance with familiar faces. It is a commonplace observation that we are able to recognise familiar people over a huge range of conditions, for example we can recognise our family members over large changes in photographic conditions, over pose, health and age. In an experimental setting, Burton et al (1999) demonstrate how students familiar with their lecturers can recognise them even in very degraded CCTV images, while unfamiliar viewers are entirely unable to recognise these people. Here we suggest that our everyday experience with familiar face recognition – robust and accurate as it is – leads us falsely to believe that we are also good at *unfamiliar* face recognition.

There is growing evidence that the processes underlying familiar and unfamiliar face recognition are to some extent separate. It has been known for many years that unfamiliar faces are harder to remember in recognition memory tasks than familiar faces (Bruce, 1986; Ellis, 1981; Klatzky & Forest, 1984), and this may simply reflect quantitative differences in difficulty. However, there are a growing number of reports that suggest some level of perceptual dissociation between familiar and unfamiliar face processing (Hancock, Bruce & Burton, 2000; Schweinberger, Pickering, Burton & Kaufmann, 2002; Megreya & Burton, 2006, 2007). The proposal is that perception of unfamiliar faces is more reliant on image-based properties, and so generalises to novel instances poorly. In contrast, familiar face recognition incorporates abstractive representations which can be recruited to recognise novel instances of the person (Hancock et al, 2000; Burton & Jenkins, 2011).

In daily life, there is little opportunity to calibrate one's unfamiliar face recognition. If we see someone on the street on two consecutive days, and imagine them the same or different people, there is rarely an opportunity for corrective feedback. On the other hand, we can have many familiar face recognition events in a single day, and these are self-evidently successful. For example, in the workplace we may encounter dozens of familiar faces, and recognise each with ease – with immediate correction of occasional errors. We hypothesise that this success with familiar faces can lead to a bias to believe that all face recognition is easy, when in fact the psychological literature demonstrates that it is not.

In fact, the existence of a bias to over-generalise one's performance with familiar faces would be consistent with findings in many other fields. For example, Nickerson, Baddeley & Freeman (1987) demonstrated that people tend to over-estimate their own general knowledge in others: if they happened to know 'the island on which Napolean was born' then they tend to assume that others know this too. Similarly, if students know particular uncommon words, they tend to assume that others also know them (Hayes & Bajzek, 2008). An analogous effect in social psychology is referred to as the 'egocentric bias', or the 'perception of consensus' (Holmes, 1968; Ross, Greene & House, 1977; Krueger & Clement, 1994). Researchers consistently find that if we are given a choice of two possible actions, we tend to predict that others will choose the same action as us. Similar effects exist for people skilled in particular cognitive tasks – people with expertise in particular tasks (like operating electronic equipment) tend to assume that others will be able to learn these easily (Hinds, 1999).

The fact that these egocentric biases crop up across different areas of psychology makes it plausible that they will also be seen in face processing. If so, this may go some way to explaining why the distinction between familiar and unfamiliar faces is not always clearly drawn in the literature, despite some compelling evidence that perceptual processing of these two classes of faces is to some extent different. On a day to day basis, we may not be aware of the fact that unfamiliar face recognition is poor, and so our excellent recognition of familiar faces is taken to be the default.

In these experiments we compare people’s accuracy in matching pairs of familiar and unfamiliar faces. Following earlier work, we expect to see higher performance for familiar faces. However, we then ask participants to estimate how others might perform with these particular face pairs. If people over-generalise their good familiar face recognition, then we hypothesise that they will imagine that the faces they know will be better recognised by others – in other words they will fail to take into account their own familiarity with a face in judging the likely behaviour of others. If this prediction turns out to be true, then it offers a partial explanation for the continued use of photo-ID. In short, we know ourselves to be good at face recognition, and so this seems like a good way of identifying someone in a document. We present four experiments on this theme, using different participant groups, and different levels of familiarity.

Experiment 1. Matching celebrity faces

In the first experiment, we tested participants on a matching task using familiar and unfamiliar celebrity faces. In order to avoid any possible systematic differences between these sets of faces we ran the study bi-laterally in the UK and in Australia. Selecting photos of locally-familiar minor celebrities, we were therefore able to use each face as familiar (for participants in that person’s country) and unfamiliar (for participants in the other country), across the experiment. On each trial, participants were asked to indicate whether a pair of faces was the same or a different person. Following previous research (Clutterbuck & Johnston, 2002, 2004), we expect an advantage for familiar faces here, despite the fact that the task requirements are independent of familiarity.

This study also incorporated a prediction component. Having made a match judgement (same/different person) we then asked participants on each trial to predict how other viewers would perform. This first experiment was somewhat exploratory, and so we also asked whether the character of the putative ‘other viewers’ might influence responses. To examine this we asked half the participants to predict how ‘students in Germany’ might perform. The intention here was to bring to mind viewers in many ways similar to our participants (who were mostly students), but who would not know any of the minor celebrities shown. In contrast with these generic others, we asked half our participants to predict how well 'passport controllers' would perform. We anticipated that people would attribute greater matching powers to passport officers, and so this instruction invites participants to provide higher predicted scores.

Method

Participants

Thirty British participants (11 male; mean age: 22 years, range: 17-60 years) took part at the University of Aberdeen, UK. Thirty Australian participants (7 male; mean age: 18 years, range: 17-22 years) took part at the University of New South Wales, Australia.

Materials

Images of 16 British and 16 Australian local celebrities were taken from a UK/Australian database, developed for bi-lateral research. The celebrities were chosen to be familiar to participants in only one of the two countries. Mismatch foils were taken from an internet image search of the verbal description of the celebrity (e.g. "woman dark hair glasses", see figure 1B). Images showed head and shoulders, broadly full-facing, but unconstrained in terms of lighting and camera characteristics. Examples are shown in figure 1. The images were all presented on paper, as part of a booklet, at 5.5 x 8cm with two images per A4 page.

Figure 1. Example test items. A, 'Same person' trial. B, 'Different person' trial (A and B counterbalanced across participants). Half of participants were asked to predict how well passport controllers, and half how well students in Germany would perform. [Copyright restrictions prevent publication of the original images used in these experiments. Images

show two people who did not appear in the experiments, and have given permission for their images to be reproduced here.]

Procedure

Participants carried out a matching same/different task with 32 pairs of images, 16 familiar and 16 unfamiliar celebrities. Half the trials were matching, and half mismatching, with these counterbalanced so that each celebrity appeared in only one trial per participant (i.e. in a match trial for half the viewers, and a mismatch trial for the remaining viewers). The task was self-paced and completed with pen and paper. For each face pair, participants also indicated how many out of 100 other people would answer correctly. For half the participants these putative other people were 'passport controllers' (example in figure 1), while for the remaining participants these were 'students in Germany '.

Results and discussion

Post-experiment checks of the familiarity manipulation showed that British participants were familiar with an average of 13/16 of the British celebrities and 1/16 of the Australians. Australian participants' mean familiarity for Australian celebrities was 13/16, and for the British celebrities was 2/16 (see endnote).

Figure 2A shows mean performance by location of participant and celebrity (see note 1). A mixed design 2x2 ANOVA (face locality, within subjects x participant locality, between subjects) showed a main effect of face locality ($F(1,58) = 4.04$, $p = .049$, $\eta_p^2 = .07$) with participants performing slightly more accurately with the Australian than the UK faces overall. There was no main effect of participant location ($F(1,58) = .08$, $p = .778$, $\eta_p^2 < .01$). These main effects were qualified by a significant interaction ($F(1,58) = 96.07$, $p < .001$, $\eta_p^2 = .62$). Simple main effects showed an effect of face locality for both the UK ($F(1,58) = 30.36$, $p < .001$, $\eta_p^2 = .34$) and Australian ($F(1,58) = 69.74$, $p < .001$, $\eta_p^2 = .55$) participants.

These results are clear, and as expected. Despite no overall performance differences between UK and Australian viewers, each group was much more accurate in matching the faces of the people they knew. This cannot be due to inherent differences in the difficulty of the matching trials, because the same items were used as familiar and unfamiliar faces across the experiment.[1]

Figure 2B shows mean prediction data across all participants. A mixed design 2x2 ANOVA (familiar/unfamiliar x prediction type: passport controller/student) showed main effects of familiarity ($F(1,58) = 13.83$, $p < .001$, $\eta_p^2 = .19$), and prediction type ($F(1,58) = 17.99$, $p < .001$, $\eta_p^2 = .24$), and these were qualified by a significant interaction ($F(1,58) = 10.49$, $p = .002$, $\eta_p^2 = .15$). Simple main effects analyses showed an effect of familiarity for the passport controllers prediction ($F(1,58) = 24.20$, $p < .001$, $\eta_p^2 = .29$) but not the students in Germany prediction ($F(1,58) = 0.12$, $p = .730$, $\eta_p^2 < .01$).

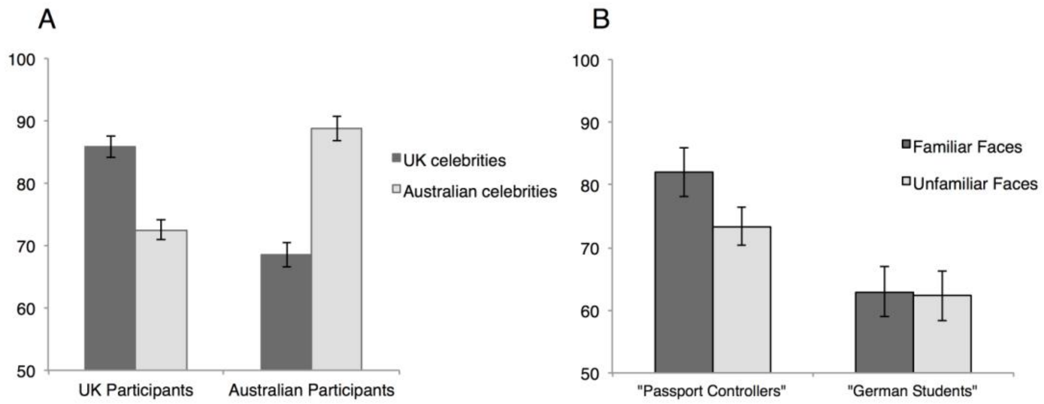

Figure 2. Mean responses for experiment 1. A. accuracy (% correct), B predictions of others' performance (% predicted correct). Error bars represent standard error of the mean.

These results clearly demonstrate that our participants predicted higher accuracy for passport controllers than for students in Germany. However, of more interest is the bias between familiar and unfamiliar faces for those passport controller predictions. Viewers apparently

[1] It is also possible to analyze performance data as sensitivity, computing d' from an analysis in which Hits correspond to correct 'same' trials, and False Alarms to incorrect 'different' trials. This analysis gives exactly the same pattern of performance. For UK participants d' = 2.15 and 1.35 for UK and Australian celebrities respectively; for Australian participants d' = 1.15 and 2.32 for UK and Australian celebrities respectively. 2x2 ANOVA shows no main effects of face locality ($F(1,58) = 2.68$, $p = .107$, $\eta_p^2 = .05$) or participant locality ($F<1$). However a significant interaction ($F(1,58) = 75.96$, $p < .001$, $\eta_p^2 = .58$) arises due to significant simple main effects of face locality for both UK ($F(1,58) = 25.06$, $p < .001$, $\eta_p^2 = .32$) and Australian ($F(1,58) = 53.58$, $p < .001$, $\eta_p^2 = .50$) participants. While this analysis confirms the standard cross-over interaction in matching performance, we shall subsequently only present accuracy, as this allows direct comparison with prediction data (e.g. figure 2).

believe that the faces that *they themselves* are familiar with, will be better matched by a third party, than faces that *they themselves* do not know. Furthermore, this is rather a large effect. As a reminder, this cannot be due to some matching items being inherently easier than others – because all items were familiar and unfamiliar equally often across the experiment. Instead, it may reflect an over-generalisation of perceivers' own performance – when they find faces easier to match, they assume that others will also find them easier to match.

This is a first suggestion of the perceptions which may lead to the passport paradox. There appears to be some support for the idea that we generalize our good familiar face recognition to all faces, and that this is not an accurate generalization. There is, however, a puzzle here in the attributions made to students in Germany. These putative students are attributed rather poor performance levels, and predictions for these people's performance do not discriminate between faces familiar and unfamiliar to our participants. Recall that the passport-controller/students-in-Germany condition is manipulated between subjects, and so this is not an effect of contrast, in which people are invited to draw a comparison between these target groups.

One possible explanation for the pattern of prediction data is that participants are actually aware of the influence of familiarity on face matching. They may assume that students in Germany, knowing none of the faces, will perform generally poorly. However, it is possible that they implicitly assumed that 'passport controllers' was a description of officers from their own country. We gave no indication of the nationality of these passport controllers, and so perhaps the effect of familiarity is due to an assumption that they will be familiar with the same set of faces as the viewers. So, the results of this first experiment are somewhat inconclusive. In order to examine the locus of the effect further, in the following experiment we compare participants' predictions about the accuracy of 'passport controllers in Germany' and 'students in Germany'. In this way, it is possible to eliminate any predictions based on an explicit model of familiarity effects, because both the putative 'other viewer' groups have the same nationality.

Experiment 2. Matching celebrities 2

This experiment is a follow-up to experiment 1, and replicates the basic design. However, because there were no differences between Australian and UK viewers, and because of the difficulties of running bi-lateral experiments, this study used only UK participants. The same set of UK/Australian minor celebrity faces were used as in experiment 1. The major difference was that in the 'predictions' component of each trial, half the participants were asked to judge the likely performance of 'passport controllers in Germany' and half 'students in Germany'.

Method

Participants

Twenty British participants took part in the study (10 male; mean age: 23 years, range: 19-30 years).

Procedure

The stimuli and experimental set up were identical to the previous experiment. The wording of the passport controllers prediction question was altered to eliminate any potential bias toward thinking of passport controllers in one's own country who would likely be familiar with the same identities as participants. This question was changed to read 'How many passport controllers in Germany would get this right /100'. All other details of the experiment remained unchanged. As with the previous experiment, familiarity of the matching pairs (familiar/unfamiliar) was manipulated within subjects, while the prediction question ('passport controllers in Germany' vs 'students in Germany') was manipulated between subjects, with participants being allocated to these two groups at random.

Results and discussion

Post-experiment familiarity checks showed that participants were familiar with an average of 14/16 of the British celebrities and 1/16 of the Australians.

Overall performance data showed the normal pattern, in that familiar faces were matched significantly more accurately than unfamiliar faces ($t(19) = 4.63$, $p < .001$, $\eta_p^2 = .53$, see figure 3A). Prediction data are shown in figure 3B. Mixed design 2x2 ANOVA (familiar/unfamiliar x passport controllers/students) showed a main effect of familiarity ($F(1,18) = 58.55$, $p < .001$, $\eta_p^2 = .76$), but no main effect of prediction type ($F(1,18) = .14$, $p = .713$, $\eta_p^2 = .01$) and no interaction ($F(1,18) = .15$, $p = .703$, $\eta_p^2 = .01$).

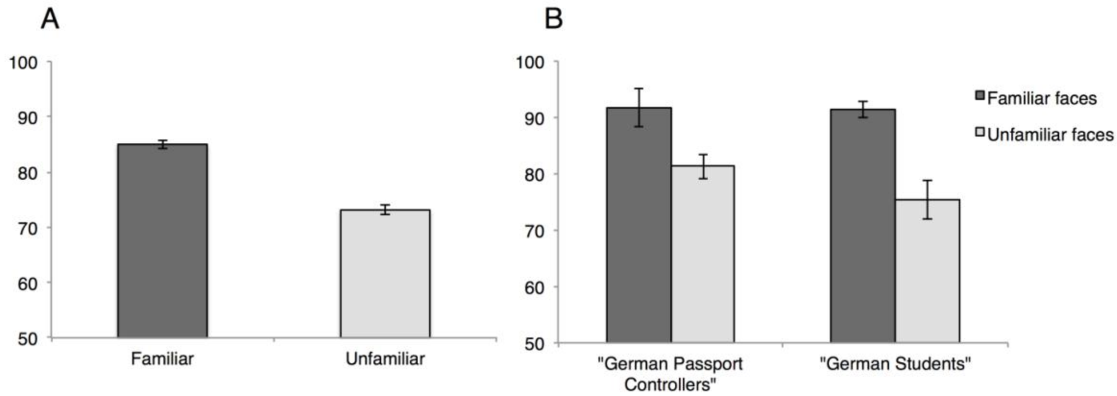

Figure 3. Mean responses for experiment 2. A. accuracy (% correct), B predictions of others' performance (% predicted correct). Error bars represent standard error of the mean.

Here, participants predicted that others would perform more accurately on a matching task for the identities with which they themselves were familiar. Furthermore, unlike experiment 1, in this case the familiarity effect was very clear for both types of 'other person' – passport controllers or students in Germany. This appears to reflect a genuine familiarity matching bias. The instructions make it clear the 'other viewers' would be unlikely to recognise the UK minor celebrities familiar to our participants.

While the predicted familiar face matching bias has now been evident in three out of four conditions across our two experiments, there remains a puzzle as to why predictions for 'students in Germany' would be so different in experiments 1 and 2. To explore this further we conducted a cross-experiment analysis comparing predictions for 'students in Germany' by British students in experiments 1 and 2. Figure 4 shows mean performance and prediction data for this comparison. For performance data, a mixed 2 factor ANOVA with Experiment as a between subjects factor, and familiarity as a within subjects factor revealed a significant main effect of familiarity ($F(1,23) = 35.82$, $p < .001$, $\eta p2 = .61$), but no main effect of experiment ($F(1,23) = .40$, $p = .533$, $\eta p2 = .017$), and no interaction ($F(1,23) = .560$, $p = .4621$, $\eta p2 = .02$). A separate ANOVA on predictions of others' performance showed main effects of Experiment and familiarity ($F(1,23) = 19.55$, $p < .001$, $\eta p2 = .46$; and $F(1,23) = 14.71$, $p = .001$, $\eta p2 = .39$, respectively). A clear significant interaction ($F(1,23) = 14.80$, $p < .001$, $\eta_p^2 = .39$) arises due to a large effect of familiarity in experiment 2 only.

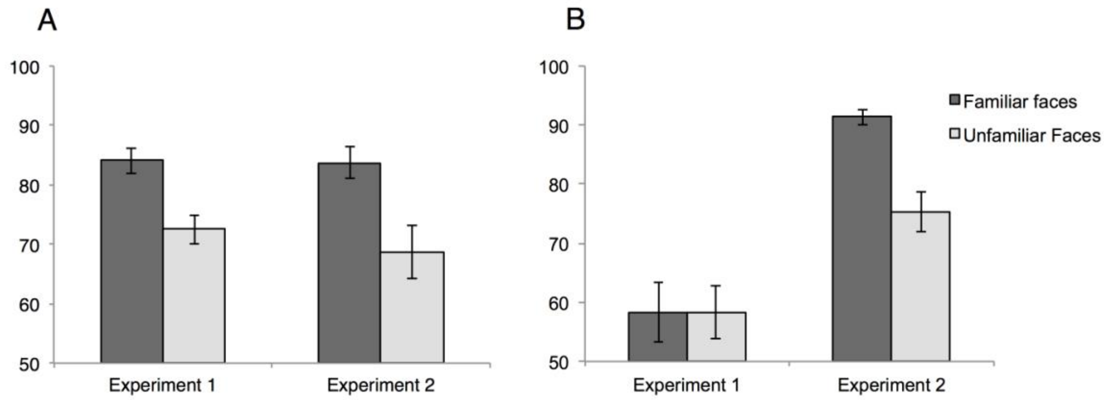

Figure 4. Responses for British participants predicting the performance of ‘students in Germany’ in experiments 1 and 2. A. accuracy (% correct), B predictions of others’ performance (% predicted correct). Error bars represent standard error of the mean.

From this analysis, it remains unclear why the predictions for ‘students in Germany’ are different across experiments. The performance difference for familiar and unfamiliar faces is robust across both experiments, and so there is no very obvious reason why prediction data should differ. This may reflect sampling differences between the experiments, or may suggest that this method of requesting participants to consider ‘other viewers’ does not produce very stable results. Across the two experiments, we have some accumulating evidence suggesting a bias in attribution of face matching: people tend to believe this to be easier for the faces they know – i.e. they fail to take into account the fact that they find these faces easier *because* they are familiar. However, there are some inconsistencies between the results of the two experiments, and so to investigate this further, we conducted experiments with different stimuli, and using different manipulations of familiarity.

Experiment 3. Matching celebrity faces of one’s own age group

In this experiment, we repeated the same basic design as experiment 1, in which two different groups of viewers are asked to match pairs of familiar and unfamiliar faces, as well as predicting how others would perform on the same test items. However, in this case, we asked all participants to predict how well “the general public” would perform – hence eliminating the ‘expertise’ manipulation used in experiments 1 and 2. In this study we manipulated familiarity by age group, choosing two sets of celebrities who would be familiar to an older and a younger participant group.

Method

Participants

Thirty-six participants took part. Eighteen were recruited for the ‘younger’ group (10 male; mean age: 24 years, range: 18-33 years) and eighteen for the ‘older’ group (7 male; mean age: 58 years, range: 50-77years). All participants were volunteers from the area local to the University of Aberdeen, UK.

Materials and procedure

Images of 16 ‘younger’ and 16 ‘older’ British celebrities were used. The celebrities were chosen to be familiar to younger and older participants respectively. Initially this selection was made by the experimenters, and was confirmed by sampling a small number of older and younger adults prior to the experiments (these people were not tested in the main experiment). This initial group was asked simply to name each of the celebrities, with younger adults correctly naming the younger celebrities but not the older celebrities, and vice versa. As in experiments 1 and 2, mismatch foils were taken from an internet image search of the verbal description of the celebrity. It was not possible to find colour photographs for each of the older celebrities, and so to keep the images within this experiment consistent, all images were presented in grayscale. Again, all image pairs were presented on paper, with order of young and old celebrities intermixed. Participants were asked to indicate whether each pair of images showed the same person or two different people, and to predict others’ performance on a trial-by-trial basis. The prediction question here was ‘how many people from the general public would get this right /100’.

Results and discussion

Figure 5A shows mean face matching performance by group. A mixed design 2x2 ANOVA (participant age, between subjects x face age, within subjects) showed no main effects (participant age $F(1,34) = 1.24$, $p = .274$, $\eta_p^2 = .04$; face age $F(1,34) = 3.52$, $p = .069$, $\eta_p^2 = .09$) but a significant interaction ($F(1,34) = 36.24$, $p < .001$, $\eta_p^2 = .52$). Simple main effects revealed an effect of face age for both younger ($F(1,34) = 8.58$, $p = .006$, $\eta_p^2 = .20$) and older participants ($F(1,34) = 31.18$, $p < .001$, $\eta_p^2 = .48$). This shows that, once again, participants performed the matching task more accurately with familiar than unfamiliar faces.

Figure 5B shows mean prediction data by group. A mixed design 2x2 ANOVA (participant age, between subjects x face age, within subjects) showed no main effects (participant age, $F(1,34) = 2.65$, $p = .113$, $\eta_p^2 = .07$; face age $F(1,34) = .85$, $p = .363$, $\eta_p^2 = .02$) but a significant interaction ($F(1,34) = 30.04$, $p < .001$, $\eta_p^2 = .47$). Simple main effects revealed an effect of face age for both younger ($F(1,34) = 20.51$, $p < .001$, $\eta_p^2 = .38$) and older participants ($F(1,34) = 10.38$, $p = .003$, $\eta_p^2 = .23$). In summary, participants predicted that others' performance would be more accurate for the faces with whom they themselves were familiar.

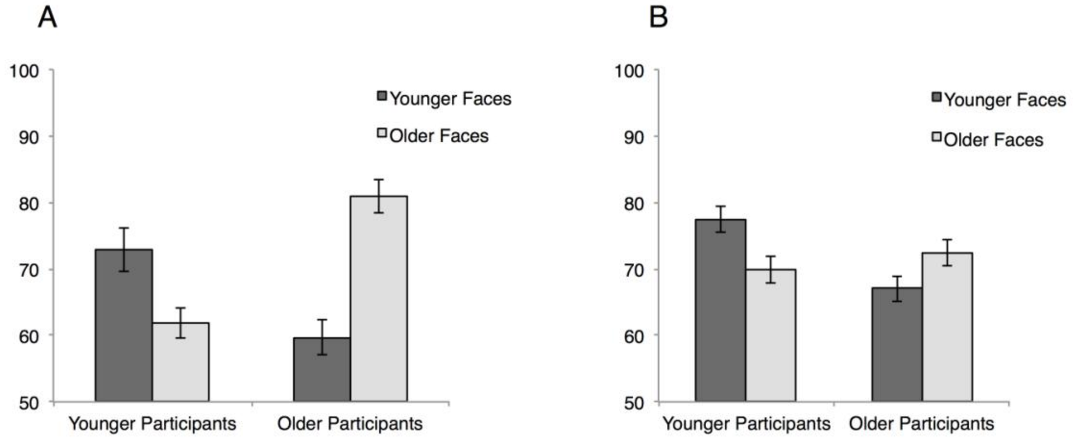

Figure 5. Mean responses for experiment 3. A. accuracy (% correct), B predictions of others' performance (% predicted correct). Error bars represent standard error of the mean.

Once again, this experiment demonstrates a clear bias to over-generalise one's own face matching-performance, on the basis of one's own familiarity. We have now observed this effect over three experiments, using very different target groups for prediction, and using different subject populations. Our claim rests on the notion of familiarity, but so far we have relied on celebrities to provide this. If people really do over-generalise their familiar face recognition, then we might expect this to arise from their experience with personal familiarity, rather than mass-media celebrities. In the final experiment, we used the faces of people personally familiar to different groups of participants.

Experiment 4. Matching personally familiar faces

This experiment examines the familiarity matching bias in personally familiar faces. It uses the same basic design as the experiments above: participants make same/different matching judgements to pairs of faces, and are asked to judge how others would perform on these items. However, in this case, our target images were the faces of staff from two different universities (Aberdeen and Glasgow). Pairs of these images were shown to students and staff in these two locations (i.e. people familiar with the targets from their own location), and to students and staff in a third location (University of Kent) who knew neither set. We expected to see the normal familiarity advantage in the matching test, and that viewers familiar with some of the target people would predict that others would perform more accurately on these faces.

Method

Participants

Twenty participants took part in each of three universities (Glasgow, 4 male; mean age: 31 years, range: 21-53 years; Aberdeen, 6 male; mean age: 24 years, range: 21-45 years; and Kent, 3 male; mean age: 23 years, range: 18-36 years). All were volunteers from the university community in those areas.

Materials and procedure

We collected two different photos for each of 32 target people: 16 in Aberdeen, 16 in Glasgow. These were taken from public sources (e.g. webpage photos) or from local databases of images for face recognition research. The two photos of each person were taken on different occasions, and all were used with permission.

As in previous experiments, we created two matching trials for each person: a match (two images of the same target) and a mismatch (a target with a foil). Once again, foils were taken from internet search using verbal descriptions. The images were presented in full colour as described above, i.e. each participant saw 32 pairs of faces and had to respond same/different to each pair. Across the experiment, each target person appeared equally often in a match and mismatch pairs, but these were counterbalanced such that participants saw each target person only once. Prior to the experiment, participants were told that we would also be using this test “with the general public around Europe”. For each trial, they were asked to predict how many people, out of 100, would answer correctly for that trial.

Results and discussion

Post-experiment checks of the familiarity manipulation showed that participants in Aberdeen were familiar with an average of 15/16 of the Aberdeen faces and 1/16 of the Glasgow faces.

Participants in Glasgow were familiar with an average of 13/16 of the Glasgow faces and 1/16 of the Aberdeen faces. Kent is over 500 miles from either university, and since all stimuli would be unfamiliar for these viewers, we did not ask them to provide familiar/unfamiliar ratings.

Figure 6A shows mean face matching performance by location of participant and target face. A mixed 3x2 ANOVA (participant location, between subjects x face location, within subjects) showed a main effect of participant location ($F(2,57) = 25.40$, $p < .001$, $\eta_p^2 = .47$), but no main effect of face location ($F(1,57) = .02$, $p = .888$, $\eta_p^2 < .01$). There was a significant interaction ($F(2,57) = 28.73$, $p < .001$, $\eta_p^2 = .50$) with simple main effects showing an effect of face location for participants in Aberdeen ($F(1,57) = 29.12$, $p < .001$, $\eta_p^2 = .34$) and in Glasgow ($F(1,57) = 28.34$, $p < .001$, $\eta_p^2 = .33$), but not Kent ($F(1,57) = .03$, $p = .863$, $\eta_p^2 < .01$). These results show a clear performance advantage for familiar faces, as in each of the experiments above. In the case of Kent participants, knowing none of the target faces, performance between the two subsets of faces was equivalent.

Figure 6B shows the mean prediction data by group. A mixed design 3x2 ANOVA (participant location, between subjects x face location, within subjects) showed main effects of participant location and face location ($F(2,57) = 3.55$, $p = .035$, $\eta_p^2 = .11$, and $F(1,57) = 10.58$, $p = .002$, $\eta_p^2 = .16$, respectively). These were qualified by a significant interaction ($F(2,57) = 15.02$, $p < .001$, $\eta_p^2 = .35$) with simple main effects showing an effect of face location for the participants in Aberdeen ($F(1,57) = 33.46$, $p < .001$, $\eta_p^2 = .37$), but no significant effect for the participants in Glasgow ($F(1,57) = 3.86$, $p = .054$, $\eta_p^2 = .06$) or Kent ($F(1,57) = 3.29$, $p = .075$, $\eta_p^2 = .05$).

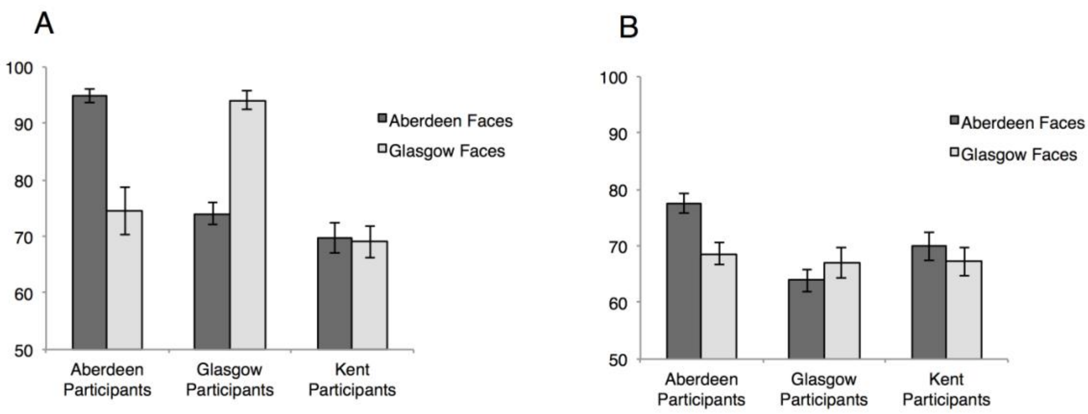

Figure 6. Mean responses for experiment 4. A. accuracy (% correct), B predictions of others' performance (% predicted correct). Error bars represent standard error of the mean.

Once again this experiment has produced evidence of a familiarity matching bias in judging the putative performance of other viewers. In this experiment, using personally familiar faces, there was clear evidence of the bias in one of the groups (the Aberdeen participants), while the Glasgow participants showed a trend in predicted direction that just failed to reach significance (NB testing uses two-tailed significance levels throughout this paper). This is interesting in the light of a non-significant trend in the data from Kent participants, favouring Aberdeen faces. Since these (Kent) participants know none of the faces, this suggests that Aberdeen faces are regarded as inherently slightly easier to match – a difference which could arise purely by chance. This context provides further support that the trend in the Glasgow participants is an interesting one.

General discussion

The results of these four experiments provide evidence for a potentially important bias in face recognition. First, we consistently show that people are better able to make pairwise same/different judgements to the faces they know, than those they do not. This replicates and extends previous research in the face recognition literature (e.g. Bruce et al, 2001; Clutterbuck & Johnston, 2002; Megreya & Burton, 2007). In these experiments we use the same pairwise task across a wide range of faces and participants, and find a completely consistent performance benefit for familiar faces across all studies.

In this context, we also provide evidence for a familiar face matching bias, which works to obscure the difficulty of unfamiliar face matching. In short, participants appear to believe that the faces they themselves know, are somehow easier to match, even by unfamiliar viewers. Support for this claim is present in each of the four rather different experiments. However, we should note that it is slightly nuanced. In experiment 1, it was present for only one type of putative 'other viewers', the passport officers, but not for others, the students in Germany. However, in experiment 2, both types of putative others were subject to the bias. This led us to the view that being specific about the 'other viewer' is likely to lead to more complex attributions, and so in experiments 3 and 4, we were deliberately more vague, telling participants only that we would be testing 'the general public'. In the final experiment, one of the groups (the Glasgow-based participants) did not show as strong an effect as the other group (the Aberdeen-based participants). When examining a novel hypothesis such as this, across very different experimental groups and stimuli, it is perhaps to be expected that not every condition will turn out an exact replica across studies. However, despite these qualifications, the pattern of the data is clear: in general, participants demonstrate the familiarity bias.

These results have both theoretical and practical importance. Theoretically, they lend support to the increasing evidence for some dissociation of processing between familiar and unfamiliar faces (e.g. Hancock et al, 2000; Megreya & Burton, 2007). We might ask *why* participants tend to over-estimate the ability of others to match their own familiarity. One suggestion is that familiar face processing is more fluent than unfamiliar face processing, relying less on image-based characteristics of the stimuli, and more on abstract representations (Bruce, 1986; Burton & Jenkins, 2011; Hancock et al, 2000). If this is true, then we might anticipate that the apparent effort required to make a match by the viewer is taken as a measure of the objective difficulty of the task. That would give participants a good reason to assume that others would also find these particular stimuli easier to process.

This explanation is consistent with accounts of more general egocentric biases (e.g. Kelley & Jacoby, 1996; Hinds, 1999). It seems to be a general property of cognition that we are poor at judging the difficulty that others will experience in performing a task. In particular, if we find a task easy, we tend to expect others to find it easy too. So, the phenomenon we have described here is certainly not one that is specific to faces. However, to our knowledge, this is the first demonstration of this egocentric bias in face perception, and it has clear implications for theoretical accounts of face recognition, which tend to conflate familiar and unfamiliar faces.

On more practical grounds, it is worthwhile considering whether the use of faces in photo-ID documents should continue. There is now a large body of evidence (briefly reviewed in the Introduction) which suggests that people are poor at unfamiliar face matching, but this seems to have had little impact on policy makers. The use of photo-ID is, if anything, increasing, in times of heightened security concerns. There seems to be a willingness to rely on face recognition, despite the clear behavioural evidence warning against it. We suggest that the bias uncovered in this paper goes some way to explaining this. In short, it is very difficult for people to believe that unfamiliar face recognition is poor – because they experience such fluent and accurate recognition of familiar faces. In parallel with the scientific community, policy makers have been slow to accept that processing of familiar and unfamiliar faces is, to some extent, separate. If this were to become more widely accepted, it would be clearer that tests of face identification need to be tailored to the particular setting in which they are used: in other words, recognising a family photo with ease, does not necessarily make it a good idea to verify identity with photos.

Endnote

There are two ways to analyse this type of data, each with advantages and disadvantages. We have presented an analysis by location, in which responses to all British and Australian faces are analysed for British and Australian viewers, regardless of whether post-experimental checks showed a particular face to have been recognized by an individual participant. This relies on the manipulation check, reported in the text, showing that viewers did indeed recognize most of their own, and very few of the opposite nationality faces. However, it is also possible to exclude items which were not recognized from one's own nationality or were recognized from the opposite nationality. This has the advantage that data are tailored individually to each participant, but the disadvantage that the post-experiment check itself is subject to some error (e.g., people sometimes claim to recognize the face of someone whom they actually do not know). In addition to the findings reported, we have also carried out the analysis in which unrecognized and falsely recognized items are excluded, across all three experiments where this arises (1, 2 and 4). In every case, the pattern of significance results is the same as those reported. Full details are available from the authors on request.

References

Bruce, V. (1986). Influences of familiarity on the processing of faces. *Perception, 15(4),* 387-397.

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*(4), 339–360.

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*(3), 207–218.

Burton, A. M. (2013). Why has research in face recognition progressed so slowly? The importance of variability. *Quarterly Journal of Experimental Psychology, 66*(8), 1467–1485.

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

Burton, A. M., Jenkins, R., & Schweinberger, S. R. (2011). Mental representations of familiar faces. *British Journal of Psychology, 102*(4), 943–958.

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.

Burton, A. M., White, D. & McNeill, A. (2010). The Glasgow Face Matching Test. *Behaviour Research Methods*, *42,* 286-291.

Clutterbuck, R., & Johnston, R. A. (2002). Exploring levels of face familiarity by using an indirect face-matching measure. *Perception, 31,* 985–994.

Clutterbuck, R., & Johnston, R. A. (2004). Matching as an index of face familiarity. *Visual Cognition, 11(7),* 857-869.

Davis, J. P., & Valentine, T. (2009). CCTV on trial: Matching video images with the defendant in the dock. *Applied Cognitive Psychology, 23*(4), 482–505.

Ellis H.D. (1981). Theoretical aspects of face recognition. In G.M. Davies, H.D. Ellis, and J.W. Shepherd (eds.) Perceiving and Remembering Faces, pp. 171–197. London: Academic Press.

Hancock, P. J. B., Bruce, V. & Burton, A. M. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Science, 4(9),* 330-337.

Hayes, J. R., & Bajzek, D. (2008). Understanding and reducing the knowledge effect:

Implications for writers. *Written Communication*, *25*(1), 104-118.

Hinds, P. J. (1999). The curse of expertise: The effects of expertise and debiasing methods on prediction of novice performance. *Journal of Experimental Psychology: Applied*, *5*(2), 205.

Holmes, D. S. (1968). Dimensions of projection. *Psychological Bulletin, 69*, 248-268.

Jenkins, R., & Burton, A. M. (2011). Stable face representations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1571), 1671–1683.

Johnston, R. A., & Edmonds, A. J. (2009). Familiar and unfamiliar face recognition: A review. *Memory*, *17*(5), 577–596.

Kelley, C. M., & Jacoby, L. L. (1996). Adult egocentrism: Subjective experience versus analytic bases for judgment. *Journal of Memory and language*, *35*(2), 157-175.

Kemp, R., Towell, N., & Pike, G. (1997). When seeing should not be believing: Photographs, credit cards and fraud. *Applied Cognitive Psychology*, *11*(3), 211–222.

Klatzky, R. L., & Forrest, F. H. (1984). Recognizing familiar and unfamiliar faces. *Memory and Cognition*, *12(1)*, 60–70.

Krueger, J. & Clement, R. W. (1994). The truly false consensus effect: An ineradicable and egocentric bias in social perception. *Journal of Personality and Social Psychology, 67(4),* 596-610.

Megreya, A. M., & Burton, A. M. (2006). Unfamiliar faces are not faces: Evidence from a matching task. *Memory & Cognition*, *34*(4), 865–876.

Megreya, A. M., & Burton, A. M. (2007). Hits and false positives in face matching: A familiarity-based dissociation. *Perception & Psychophysics*, *69*(7), 1175.

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*(4), 364–372.

Nickerson, R. S., Baddeley, A., & Freeman, B. (1987). Are people's estimates of what other people know influenced by what they themselves know?. *Acta Psychologica*, *64*(3), 245-259.

Ross, L., Greene, D., & House, P. (1977). The "false consensus effect": An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology, 13*, 279-301.

Schweinberger, S. R., Pickering, E. C., Burton, A. M., & Kaufmann, J. M. (2002). Human brain potential correlates of repetition priming in face and name recognition. *Neuropsychologia, 20(12),* 2057-2073.

White, D., Kemp, R.I., Jenkins, R., Matheson, M., & Burton, A.M. (2014). Passport Officers' errors in face matching. *PLoS one, 9,* 8, e103510.