



Familiar faces as islands of expertise

Peter J.B. Hancock^{*}

Psychology, Faculty of Natural Sciences, University of Stirling, Stirling FK9 4LA, United Kingdom

ARTICLE INFO

Keywords:

Familiar faces
Expertise
Face-space-R
Kin recognition
Face recognition

ABSTRACT

Most people recognise and match pictures of familiar faces effortlessly, while struggling to match unfamiliar face images. This has led to the suggestion that true human expertise for faces applies only to familiar faces. This paper develops that idea to propose that we have isolated 'islands of expertise' surrounding each familiar face that allow us to perform better with faces that resemble those we already know. This idea is tested in three experiments. The first shows that familiarity with a person facilitates identification of their relatives. The second shows that people are better able to remember faces that resemble someone they already know. The third shows that while prompting participants to think about resemblance at study produces a large positive effect on subsequent recognition, there is still a significant effect if there is no such prompt. Face-space-R (Lewis, 2004) is used to illustrate a possible computational explanation of the processes involved.

1. Introduction

"He looks just like his Dad" is a saying common enough to suggest that people can, or at least believe they can, spot family resemblances. In objective tests, however, our ability to do so is poorer than we might suppose. Performance is above chance but far from perfect (e.g. Alvergne et al., 2009; Dal Martello & Maloney, 2006; Nesse, Silverman, & Bortz, 1990). A key difference may be that we know the Dad in question, but the objective tests use faces that are unfamiliar to their participants. The 'islands of expertise' theory developed here posits that this familiarity is crucial: it enables us to see the likeness of relatives more clearly and to better process other faces that resemble those already known to us.

Studies of face matching and recognition have shown a radical difference in ability between familiar and unfamiliar faces. For example, Burton, Wilson, Cowan, and Bruce (1999) found that familiarity with the individuals shown in poor-quality video footage changed recognition performance from barely above chance to almost ceiling. The difference is qualitative; the task demands are quite different. For familiar faces, a matching study becomes a question of whether you can recognise that the images show the same person (they are both pictures of person X). The test phase of memory studies requires identification of the individual shown and deciding whether they were in the study set. The recognition process presumably involves comparing the new face with some form of stored representation. For unfamiliar faces, there is no existing representation to compare it with. Instead, participants tend to

resort to feature matching, looking for a particular mole or at the hairline. The more different the two faces are in terms of lighting, orientation etc., the harder the task becomes (Hancock, Bruce, & Burton, 2000; Johnston & Edmonds, 2009).

The differences between familiar and unfamiliar face processing led Megreya and Burton (2006) to claim that 'Unfamiliar faces are not faces'. Their evidence included the observation that performance on matching of inverted faces is strongly correlated with performance on upright faces for unfamiliar, but not familiar faces. The claim is arguably too strong. Unfamiliar faces do show effects that suggest they are indeed processed as whole faces, such as the part-whole illusion (Hole, 1994). Changing the bottom half of a face image affects the perception of the top half even for unfamiliar faces.

The question of the extent to which we are general face experts is hotly debated. It is argued that we gradually develop expertise in faces (Carey, 1992) and that this expertise accounts for phenomena such as the other race effect and inversion (Rhodes, Hayward, & Winkler, 2006). It has been likened to expertise in other domains such as birds and cars (Curby & Gauthier, 2009; Gauthier, Skudlarski, Gore, & Anderson, 2000). However, while phenomena such as the part-whole illusion suggest that there is some specialised processing for unfamiliar faces, Young and Burton (2018) argue that we are only truly expert with familiar faces. For example, Saether and Laeng (2008) found that the parents of identical twins, while able to identify their own children easily, were no better than average at distinguishing other sets of twins.

An implication of having true expertise only for familiar faces is that

^{*} Corresponding author.

E-mail address: pjbh1@stir.ac.uk.

we will each have ‘islands of expertise’, located in those regions of ‘face space’ (Valentine, 1991) corresponding to the faces that we know. The proposal in this paper is that, when an unfamiliar face is similar to a known face, it will inherit some of the associated expertise. That is, we should be better able to process faces that resemble someone we know. This proposal can account for some of the apparent expertise shown with unfamiliar faces. For example, the other-race effect can be explained because, by definition, we know fewer faces of the ‘other race’ and therefore have fewer islands of expertise in that region of face space. A given unfamiliar face is therefore less likely to look anything like a familiar face and reverts to relatively inexpert, unfamiliar face processing.

The idea is tested here in three experiments. The first considers our ability to identify relatives of a target face, depending on whether we are familiar with the target (i.e. do we know the Dad in question?). The second and third experiments look at effects in a face memory task of whether the faces to be remembered resemble previously known identities. Finally, a potential computational account is outlined, in terms of the Lewis Face-space-R model (Lewis, 2004).

2. Experiment 1

There has been a long study of human abilities to do kin recognition, partly driven by theories relating to paternal uncertainty. The argument is that babies might resemble their fathers more than their mothers, in order to reassure the father that it is indeed his child. Christenfeld and Hill (1995) reported that one year old children were indeed rated as resembling their father more than their mother. The downside of any such putative adaptation would be what happens if in fact the child resembles someone else. Bredart and French (1999) therefore argue that a baby should resemble both parents equally, on average. They failed to replicate the increased resemblance to fathers reported by Christenfeld and Hill (1995). Rather than rating resemblance, they simply asked participants to pick the parent from a line-up of three images. Given pictures of children aged 1, 3 or 5, Bredart and French found performance to be between 7% and 14% better than the chance performance of 33%. People still made mistakes more than half the time.

Using pictures of new born babies, Mclain, Setters, Moulton, and Pratt (2000) found an advantage for identifying the correct mother from a line-up of three images. Bressan and Grassi (2004) speculate that a possible reason for the different results in the three studies is that Christenfeld and Hill asked participants to rate resemblance on a scale from 1 to 10, while the other two asked them to identify the correct match. Bressan and Grassi therefore argue that strategies may differ between the two tasks, with matching being decided on a possibly idiosyncratic feature such as dimples, while overall similarity remains relatively low.

Bressan and Dal Martello (2002) argued that people’s perceived ability to see family likeness between parent and child is influenced by assumed knowledge of the relationship. In real life, parents and children are often seen together. Believing that they are thus related may affect how similar they appear. To test this idea, they asked for ratings of similarity between adult-infant pairs with information about relatedness as a controlled variable. Although the (spurious) information about relatedness did not affect accuracy, they found that it was the strongest predictor of rated similarity. Average similarity ratings were higher for non-related pairs incorrectly labelled as related than they were for genuinely related pairs labelled as unrelated. The finding has been repeated for a Japanese image set by Oda, Matsumoto-Oda, and Kurashima (2005), who got a larger effect size for their label than for genuine relatedness. So perhaps our perception that we can see family resemblances stems from already knowing the answer.

Set against that, however, are anecdotal accounts of people being stopped in the street by a stranger to be told, for example, ‘they must be the sister of X’. The key difference here may be that, by definition, the observer already knows person X, where the previous studies make a

point of ensuring that the participants do not know the people depicted in the photographs shown. The reason is not stated, but it would ensure that they don’t already know the answer. There has been no reported study of the effects of knowing the to-be-matched person, e.g. knowing the child to be matched to one of three unknown parents.

The hypothesis to be tested in this experiment is that it will be easier to identify the relative of a known face, because it ‘looks like’ the known person. Yet there are reasons to think that it may be harder, because of the phenomenon of categorical perception. This was demonstrated for faces by Beale and Keil (1995), who produced morphed sequences between two familiar or unfamiliar faces. They asked people to rate the similarity of pairs of images 20% apart in the morph sequence and found that there was a step change across the mid point for famous pairs, but not for unfamiliar ones. The notion is that a familiar face has a sharp decision boundary, outside of which is a different identity. Subsequent work has queried the lack of categorical perception for unfamiliar faces, with Levin and Beale (2000) suggesting that the categories can be learned quite quickly. Kikutani, Roberson, and Hanley (2008) found categorical perception only when unfamiliar face endpoints were given names, while Angeli, Davidoff, and Valentine (2008) suggested that details of experimental design and distinctiveness of the faces would affect the extent to which categorical perception is found with unfamiliar faces. All agree, however, that categorical perception with familiar faces is unequivocal and it therefore seems possible that someone unfamiliar with our faces would be *better* able to see similarities, as the *perceived* difference between the faces would be less. Similarly, Dwyer and Vladeanu (2009) argue that learning a face is explicitly aided by comparison with, and therefore differentiation from, similar faces. Again, this would suggest that familiarity with a face would make it easier to be sure that the relative is a different person but says nothing about identifying the resemblance.

Note, however, that the task is very different for those familiar with one of the faces and those who are not. When familiar with the parents, the task becomes identifying the people depicted and then deciding which of them the child looks more like, based on pre-existing knowledge of their appearance. When unfamiliar with any of those depicted, the participant must rely on identifying physical similarities between the faces. Clutterbuck and Johnston (2004) have shown that increasing familiarity with a face shortens the time required to match two images of the same face and propose that matching ability might be a useful index of familiarity. Here we have two different, but possibly related faces and it is therefore of interest to see what the effect of familiarity is on identifying family resemblances.

In this study, participants attempted to identify which of two target people was related to a third, who was in fact a first degree relative of one. They also rated the similarity of each target to the relative. The target faces were either members of staff or final year psychology students. The participants were first, final year or postgraduate students. The expectation was that the first year students, freshly arrived at the time of the study, would know very few of the staff and even fewer final year students, whereas the final year students would know most of the staff and many of their contemporaries. The prediction is that, being more familiar with the targets, the final year students would perform better on the matching task and give higher match ratings to related individuals. However, it is also plausible that the final year students would perform better for being more motivated, or simply better practised at doing psychology experiments. So a third group of students was added, postgraduates in the department. This postgraduate sample will know most of the staff, but few of the final year students. Thus, their performance is expected to be similar to that of the final year students on the staff image set, but like the first years on the final year set.

2.1. Method

2.1.1. Materials

Members of staff of the psychology department and final year

psychology students were asked to provide pictures of themselves (referred to as targets) and their first-degree relatives – parents, siblings or children. These were deliberately ‘ambient’ images, the better to reflect natural face processing abilities. We obtained 14 such pairs for the staff set and 18 for the final year set. If not already digital these were scanned. The images were cropped and presented at 280 × 400 pixels, about 8 × 11 cm on screen.

2.1.2. Participants

Thirty-five first year psychology students took part in return for course credit: 3 were dropped due to not following instructions, leaving 22 female, 10 male, mean age 18.7 years. Fifteen final year students, 11 female, mean age 22.1, and 16 research postgraduates, 15 female, mean age 31.1, from the psychology department took part voluntarily.

2.1.3. Design

Dependent variables are proportion correct for matching, with 0.5 being chance and 1.0 perfect; and similarity rating on a scale from 1 to 7 for related and unrelated individuals. There are three participant groups: first year, final year and postgraduate.

2.1.4. Procedure

The staff and undergraduate photo sets were shown separately, with order of presentation of the two sets alternated. Participants were shown a relative photograph at the top of the screen, and two potential targets at the bottom. They were told the top image would be the sibling, parent or child of one of the bottom two. Pairs of potential relatives were defined at set up, choosing foils from among the set that looked plausible, having a similar general appearance. Each target photograph therefore appeared twice, always in the same pair, once with a correct match and once without, though participants were not told this. Order of presentation of the triplets was randomized by Eprime. Participants were allowed as long as they wished to make a decision. When they had completed all triplets, they were then asked to rate the similarity of each pair of target and relative faces, on a scale from 1 to 7. Each target face was rated twice, once for resemblance to its true relative and once for the non-relative with which it had been presented in the first stage. The photographs to be rated were presented side by side, again with no time limit for response. Participants were then asked whether or not they knew each of the targets (i.e. the staff or undergraduate pictures) and finally whether they already knew the match between target and relative (e.g. because they had seen a photograph by a desk). When they had completed the first set, they were shown the second, with identical procedure.

2.1.5. Analysis

Linear mixed effect (LME) modelling for all experiments was carried out using R (R Core Team, 2021) and the packages lme4 (Bates, Mächler, Bolker, & Walker, 2015) and lmerTest (Kuznetsova, Brockhoff, & Christensen, 2017). For all models, both participants and images were entered as random factors, allowing variable intercepts and slopes with respect to the key variable of resemblance. For experiments 2 and 3, the full model is singular, indicating a lack of meaningful data for the random slopes. The equivalent models without random slopes are reported in supplementary analyses; the results are qualitatively identical.

2.2. Results

Where participants reported knowing that a pair was related, the responses for both that pair and the other triplet with the same target photograph were removed from the data. Thus if someone knew that John was Paul's son, then that response was removed, as was the response to the other occasion that Paul appeared, this time as a foil target. This was done in case participants realised that each target person appeared twice and therefore logically could not be related to the relative in the second case. This was probably unnecessarily

conservative: participants were not told that each target would appear twice and in any case responses to the case where the target appeared as a foil before the veridical pairing would be unlikely to be affected. In fact, the incidence of such knowledge was low, and only 12 cases had to be removed, all from postgraduate participants evaluating the staff target set.

Three of the first year participants were removed, since they claimed to know almost all of the targets, both staff and final year. It seems likely they were thinking that the identity check was a memory test and correctly asserting that they had seen the images in the first part of the study. Table 1 shows the average recognition of the targets for each of the participant groups. As expected, the first year students knew very few of the target identities in either image set, the final year sample knew about half of both, while the postgraduate sample knew most of the staff targets but very few final years.

Fig. 1 shows the proportion of correct matches of relative to target for the two different target sets. It is apparent that performance is better in the staff set, but the two sets were in no way equated for difficulty and are not directly comparable. The pattern between groups is as expected; postgraduates and final years do better than first years on the staff set, while final years do better than the other two groups on the final year set. Analysis by group using ANOVA, confirming this pattern, is reported in supplementary analyses.

To enter all the data in a single analysis, an LME model was used to test the effect of knowing the target on matching accuracy. The results, summarised in Table 2a, indicate a highly significant effect of knowing the target, with an increase in proportion correct of 0.13, Fig. 2a.

Fig. 2b shows the similarity ratings for the two target groups. It is apparent that the related pairs are rated as more similar than the unrelated ones but that there is an effect of knowing the target only when the pairs match (ie are actually related). The similarity ratings were also analysed by an LME with fixed factors of whether the target was known, whether the pair of faces was a match and the interaction between these factors. The results, in Table 2b, show the expected large effect of whether the pair are related (the similarity rating is 0.92 higher on average when they are). Overall, there is a small negative effect of knowing the target, qualified by a highly significant interaction such that the similarity rating is on average 0.81 higher when the target is known and the pair are a match.

2.3. Discussion

As predicted, familiarity with the target facilitates recognition of family resemblance. While the experiment was designed as a between groups study and is analysed as such in supplementary analyses, the use of an LME allows focus on the key variable, namely whether a particular participant knows a particular target. Such familiarity results in a large increase in matching accuracy and an increase in rated similarity for related pairs. That there is a small negative affect of familiarity on overall similarity score indicates that there is some tendency for a familiar target to look less like an unrelated face.

There is a possibility, however, that since the similarity rating followed the matching decision, the one affected the other. That is, when assessing similarity, a participant might remember their decision that the pair now presented either were or were not a match and change their rating, perhaps subconsciously. A further experiment with the two decisions made by different groups of participants would be required to rule this possibility out.

The LME estimated the intercept, i.e. the matching performance for

Table 1

Percentage of each target group reported to be known by each participant group.

Participant group	First year	Final Year	Postgraduate
Final year targets	0.9%	41.3%	5.4%
Staff targets	1.8%	47.3%	76.6%

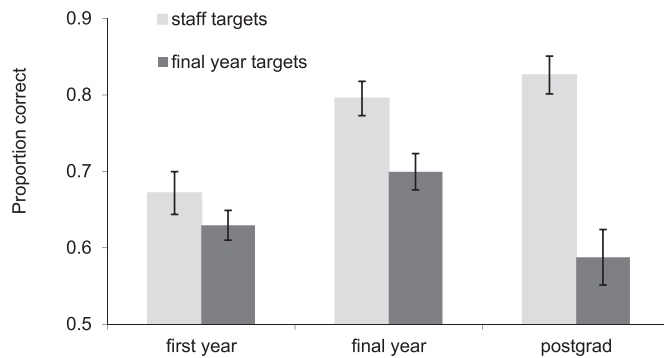


Fig. 1. Matching performance by each of the three participant groups on the two target sets. Error bars are standard errors.

Table 2

Summary of LME results for Experiment1.

	Estimate (Standard error)	t-value (Approx df)	p-value
a) Proportion correct			
Intercept	0.65 (0.034)	19.21 (37.98)	<0.001
Know target	0.13 (0.027)	4.95(29.06)	<0.001
b) Similarity rating			
Intercept	3.45 (0.12)	28.39 (79.91)	<0.001
Know target	-0.24 (0.12)	-2.04 (70.52)	0.045
Match	0.92 (0.049)	18.78 (4242)	<0.001
Match × Know target	0.81 (0.10)	7.89 (4193)	<0.001

unfamiliar faces, to be 0.65, better than chance, but not by much, consistent with previous studies using unfamiliar faces discussed above. Familiarity raised this to 0.78, significantly higher. Ratings of similarity followed the same pattern for truly related pairs, with increased rated similarity when familiar with the targets.

A possible explanation for the data is that the familiar participants knew more of the relatives than they reported. When asked, they might believe the answer to be no, while in fact having some memory of having met the relative, or of having seen their photograph. It would be very hard to control for this completely. One approach might be to do away with genuine familiarity at all, and do an experiment where participants are first familiarised with half of the targets in some way. It is an open question as to whether such rapid familiarisation would produce the effects reported here.

Nonetheless, these data are consistent with the islands of expertise proposal. While categorical processing of familiar faces (Beale & Keil, 1995) suggests that a relative's face might look more different from the target to someone familiar with them, the data show an increased ability to detect the underlying resemblance. This can be explained, at a surface level, by the difference in the task. Those unfamiliar with the faces have

to resort to face matching, while those familiar can first identify the targets and then use their existing knowledge of the faces to decide which the relative most resembles. This is despite the third 'relative' face being unfamiliar to all participants. Familiarity with the target allows the underlying similarities to be perceived.

The next two experiments test whether this advantage might also apply to a recognition task where the study items are previously unfamiliar faces. Where a particular study item resembles a known face, does that enhance subsequent recognition (Experiment 2) and does identification of the resemblance have to be conscious (Experiment 3)?

3. Experiment 2

In a typical test of unfamiliar face memory, participants are shown a set of faces and asked to make some decision about them, such as how distinctive they are. After some distractor task, they are shown a bigger set of faces, some of which show identities seen at study, and asked to decide which are new and which old. How successfully a given identity is recognised will depend on a variety of factors, such as how distinctive it is (Wickham, Morris, & Fritz, 2000). For familiar faces, the underlying process is very different. At study, a face will produce a recognition response, which may include a name or other semantic information. At test, assuming the new picture is also recognised, the task becomes one of whether person X was seen at study, rather than having to match the facial appearance to study items. So, what if an unfamiliar face happens to resemble someone who is already known, for example, your friend John? If the resemblance is strong enough to generate a name, then it seems plausible that the same semantic processing will operate. The prediction, then, is that unfamiliar study faces that a participant thinks look like someone they know will be better remembered at test. Since everyone knows a different set of faces, we would expect each participant to remember different faces. This might account for some of the otherwise unexplained variance between individuals in typical face recognition studies.

A secondary aspect of this experiment was to test possible differences caused by using the same or different images at study and test. If the same image is used at study and test, there is the possibility of doing the task by image recognition. Even matching two different images of the same face is surprisingly difficult (Bruce et al., 1999). Using a different image at study and test is thought to require more specific face processing. It is therefore possible that, if there is an effect of prior familiarity on the ability to remember a face, this will show more strongly when the image is different, since both processes require deep face processing. Consider the same image condition: at study, a participant may think, that image looks like Fred. At test, the same image appears, and they can see it is the image that looked like Fred so they may be more confident in their response. The different image condition should on average be harder, precisely because the image is different, therefore there is more room for improvement. So what might be the effects of resemblance to a known identity? Different pictures of an individual

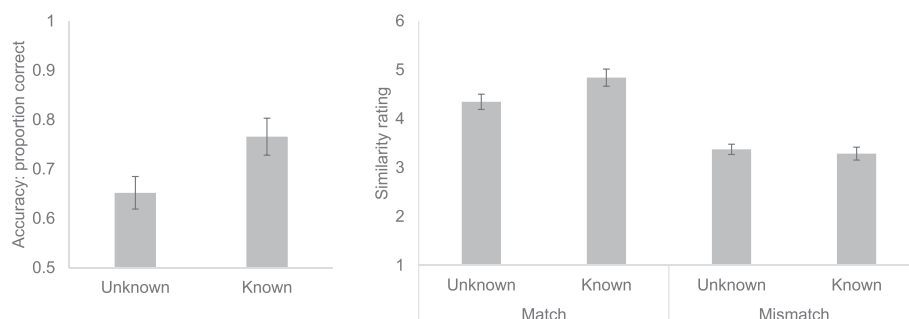


Fig. 2. a) Effect of knowing the target on proportion correct, b) Effect of knowing the target on similarity rating of match and mismatched pairs. Error bars are standard errors by items.

vary in how good a likeness they are thought to be (Mileva, Young, Kramer, & Burton, 2019; Ritchie, Kramer, & Burton, 2018). It follows that different images of an unfamiliar face will vary in how much they resemble some other, already familiar individual. The study picture might look like Fred; the test image may look less, or more, like Fred. This can be expected to add variance to any effect but it is unclear how the different factors may affect the effect size.

The experiment reported here was part of a larger study that involved eye-tracking participants while they undertook three different tasks; a memory study, a matching study and a composite task (Richler, Mack, Gauthier, & Palmeri, 2009). Only the behavioural results from the memory study are presented here.

3.1. Method

3.1.1. Participants

Thirty students, mean age 22 (SD = 6.2), 4 male, 24 female, 2 other, from the University of Stirling took part in return for course credit.

3.1.2. Materials

The face images used came from a set collected at the University of Surrey. Two different frontal images were chosen of each of 60 people, half male. The photographs were taken some weeks apart, enough for obvious differences in hair length or style, or the emergence of facial hair. The images were divided into 4 sets of 30, each containing half the identities, with different sets for the two photographs of each identity. Counterbalancing allowed both images of all the identities to be used in turn as study items. Images were loosely cropped within an oval to remove any clothing and presented in colour at 283×371 pixels on a Tobii 1750 eye tracker which has a 17 in. monitor at 1280×1024 resolution.

3.1.3. Procedure

Participants first completed the study phase of the memory experiment. They were shown 15 male and 15 female faces and asked to rate each on a resemblance scale of 1–7 how much the face reminded them of someone they knew, either personally or by being famous. Faces stayed on screen until response. The order of the faces was randomized for each participant.

Participants then completed 40 trials of a split face holistic processing task, which took about 3 min. They then completed 30 trials of an inverted version of the Glasgow Face Matching Task (Burton, White, & McNeill, 2010), which again took about 3 min. Both these tasks use monochrome images and are therefore unlikely to cause much interference with the memory study. They are not discussed further here.

The recognition stage of the memory study followed. Participants saw pictures of 30 previously unseen identities and the 30 seen at study. Of these, half showed the identical picture, while the other half showed a new, unseen image. Participants were asked to respond to each with a score from 1 to 7, where 1 was certain they had not seen that identity at study, 7 was certain they had and 4 was completely uncertain as to whether they had.

3.2. Results

Of interest here is whether participants better remembered faces they thought looked like someone they already knew. Since each participant knows different people, the analysis has to be by items, comparing the memorability of a face when it looked familiar with when it did not. One participant was omitted from analysis, since they only ever responded either 1 or 7 to the resemblance question. Since there were very few resemblance ratings of 6 and 7, they were combined into one group for analysis. The key dependent variable is the recognition confidence rating at test given to items that were in the study set, referred to as hit confidence.

Fig. 3 shows the relationship between the rated resemblance at study

and the confidence of recognition at test. It is apparent that the confidence ratings for the same images are far higher than for different images but that both show the same upward slope with rated resemblance. In fact the lines of best fit have identical slopes of 0.20 increase in hit confidence for each step in resemblance. Note that the bottom end of the different image line, for a resemblance of 1, is below 4, which indicates a tendency to think they were unseen at study. For comparison, the average recognition score for the truly unseen images is 2.11, significantly lower (paired *t*-test by subjects, $t(29) = 13.9$, $p < .001$). Participants are able to do the task, if not very confidently.

An LME with random factors for participant and image, and fixed factors of resemblance and image type at test (same/different), confirmed a large effect of image type, a highly significant effect of resemblance and no interaction between them (Table 3). The estimate of the effect of a unit change in resemblance on hit confidence from the more sophisticated analysis is slightly lower, at 0.18.

3.3. Discussion

The data showed the predicted effect of resemblance, with an almost monotonic increase in hit confidence with strength of resemblance. It seems that, if a face consciously reminds a participant of someone familiar, then they are better able to recognise the same identity later.

The surprising finding is that the slope is identical for the same and different image conditions. In truth, it is surprising that the relationship is so linear at all, given that both the resemblance and recognition variables are really ordinal. In the introduction to this experiment, various factors were identified that might affect the effect size in the different image condition. The recognition task is harder, allowing more room for improvement and requiring more face processing, as opposed to superficial image matching. Image variability is likely also to affect how strongly a given image resembles a known identity and in turn, the size of any familiarity benefit. It seems that any overall change in the effect size is too small to be detected here.

A limitation of Experiment 2 is that participants were prompted to consider whether each study face resembled anybody they know. This prompt is likely to cause participants to form associations between study faces and known faces that the participant themselves would not generate spontaneously. The idea of islands of expertise would be more significant if there is an effect of resemblance when not overtly conscious of it. This might be possible if the improved performance is not only due to overt semantic associations (that face looks like X) but also to improved resolution in the 'face space' near a known identity. It

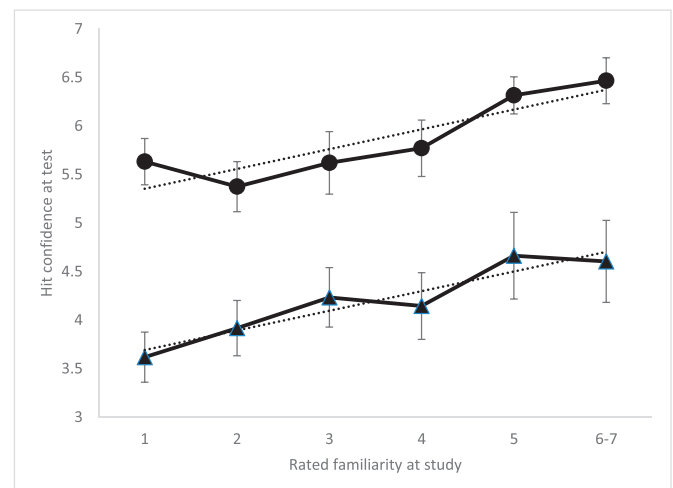


Fig. 3. Hit confidence at test against resemblance rating at study. Circles, same image at test, triangles, different image at test, each with a best linear fit line, dotted. Error bars are standard errors by items.

Table 3
Summary of LME results for Experiments 2 and 3.

Experiment 2	Estimate (standard error)	t-value (approximate df)	p-value
Intercept	3.49 (0.26)	13.57 (59.15)	<0.001
Resemblance	0.18 (0.067)	2.77 (56.61)	0.0076
Image type	1.35 (0.25)	5.35 (821.2)	<0.001
Resemblance x Image type	−0.028 (0.081)	−0.34 (782.9)	0.73
Experiment 3 full model			
Intercept	3.91 (0.14)	28.81 (110.5)	<0.001
Resemblance	0.15 (0.04)	3.43 (1482)	<0.001
Prompt	−0.40 (0.13)	−3.03 (2040)	0.002
Resemblance x Prompt	0.16 (0.06)	2.57 (1962)	0.01
Experiment 3 unprompted only			
Intercept	3.90 (0.13)	29.40 (50.96)	<0.001
Resemblance	0.16 (0.047)	3.36 (83.97)	0.0012

isn't possible to tell people not to think about resemblance at study, since doing so would have precisely the opposite effect. They can be asked about some other characteristic, without saying anything about resemblance beforehand, then asked after the test phase whether each face reminds them of anyone. That forms the basis of Experiment 3, where participants were asked to rate faces for trustworthiness at study, unprompted about resemblance. They then completed a second memory experiment with a different set of faces, where they were prompted to think about resemblance at study.

4. Experiment 3

4.1. Method

4.1.1. Participants

Forty-four students at the University of Stirling, mean age 20 (SD = 2.2), 16 male, 28 female, took part voluntarily. They were told only that it was a study of factors affecting the recognition of faces.

4.1.2. Materials

Ninety-six pairs of images were selected from the Glasgow unfamiliar faces database (Burton et al., 2010), half male. The C1 set were used for study items and the DV set for test. These images are taken in the same sitting, one using a camera, the other being a still taken from a digital video.

4.1.3. Procedure

The overall experiment consists of two memory tests, using different sets of face images. In the first study phase, participants were shown 24 face images, one at a time, and asked to rate them for perceived trustworthiness on a scale from 1 to 5, labelled as not, slightly, moderately, very and completely trustworthy. The faces were all male or all female, counterbalanced by participant number. They were then shown a distractor video of cute puppies for 30 s, before the test phase. They were presented with 48 face images, half showing identities from the study phase, and asked to indicate whether they had seen that person before, on a scale from 1 to 6. These response options were labelled as certain no, think no, guess no, guess yes, think yes and certain yes. This differs from the 7 point scale used in Experiment 2 in omitting the central 'don't know' response and forcing participants to make a decision. Following the test phase, participants were shown the study images again, one at a time, and asked whether the face reminded them of anyone they knew, either personally or as a celebrity, on a scale from 1 to 5. These response options were labelled as 1) does not remind me of anyone, 2) slight, 3) moderate, 4) distinct resemblance and 5) looks very like someone I know. The scale was shorter than that used in Experiment 2 as the 7 point scale there was used sparsely.

This was followed by another study phase, using the opposite sex set.

This time, participants were asked about resemblance at study, using the same prompt and response scale. Another distractor video was followed by the final test phase, differing from the first only in the sex of the image set.

4.2. Results

Fig. 4 shows the hit confidence at test for studied faces, averaged by items for both prompt conditions. It is apparent that there is a general increase in hit confidence with resemblance rating whether or not participants were prompted to think about resemblance. An LME with participants and images as random factors, and resemblance and prompt condition as fixed factors, reported in Table 3, confirms significant effects of both resemblance and prompt and the interaction between them.

It is noteworthy that the effect of prompt is negative (Table 3), estimated at −0.4, while in Fig. 4 the line for prompted lies above that for unprompted everywhere except for resemblance level 1. This is because the majority of responses are at resemblance level 1; relatively few of the images reminded participants of anyone. The model identifies that the average effect of the prompt is to reduce hit confidence, with the interaction indicating a higher effect of resemblance when prompted to think about it.

Since the aim of this experiment was to assess whether there would be an effect of resemblance if unprompted about it at study, Table 3 also reports an LME for only the first, unprompted part of the experiment. This confirms a highly significant effect of resemblance, estimated at 0.16 on the confidence scale per step of resemblance.

4.2.1. Trustworthiness

There have been reports that trustworthiness may affect the memorability of faces (Felisberti & Pavey, 2010). Here, no effects were found (LME in supplementary analysis).

4.2.2. Study time

A possible issue of concern is that differences between parts 1 and 2 are caused by the amount of time participants spent studying each face. If they spent longer studying faces when assessing resemblance, that might explain the bigger effect at test. A paired samples *t*-test showed that the time spent assessing trustworthiness in part 1 ($M = 3077$ ms) was significantly longer than the time to assess resemblance in part 2 ($M = 2319$ ms, $t(43) = -3.72$, $p = .001$). This cannot therefore explain the bigger effect of resemblance in part 2.

4.3. Discussion

Participants were better at recognising new pictures of someone that reminded them of an already familiar face. Prompting participants to

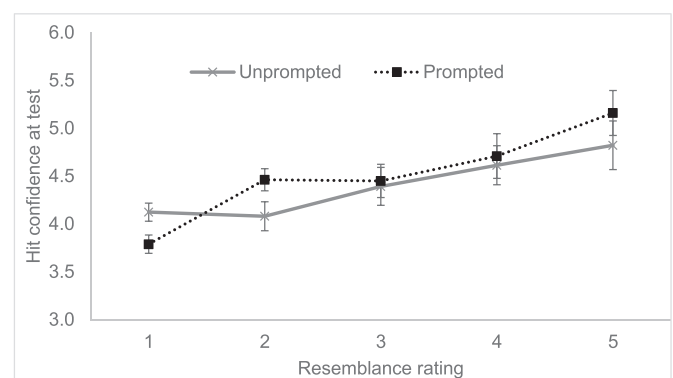


Fig. 4. Experiment 3 hit confidence scores for studied faces, from part 1, unprompted, and part 2, prompted about resemblance at study. Error bars are standard errors by items.

think about resemblance at study produced a larger effect, though there was a robust effect when unprompted. This suggests that there may be two components to the effect. One is an overt, conscious process where the resemblance is identified and the participant actively considers the likeness to the known face. The other may be unconscious, with the novel face merely 'looking familiar'.

This sense of general familiarity was termed 'context free familiarity' in early experiments on the memorability of faces. Vokey and Read (1988) also asked their participants about whether faces in a memory study reminded them of someone they knew. However, they averaged this familiarity across participants and concluded it had rather little effect on recognition. They comment that this is probably because people know different faces. Had they analysed their data by items, as here, they may have uncovered an effect of resemblance.

An unexpected result is the overall decrease in hit confidence when participants are prompted to think about resemblance at study. This effect is driven by the lower confidence when the faces did not resemble anyone. One possible explanation is related to task demands, whereby asking people to think about resemblance at test causes them to pay less attention to those faces that do not resemble anyone. Alternatively, they may be actively using resemblance at test and tending to reject faces that do not remind them of anyone as being unseen. If so, then previously unseen test faces (foils) that resemble a familiar face might be less confidently rejected, a prediction for future study, since resemblance information was collected here only for study items.

5. An outline computational explanation

How might these effects of resemblance manifest themselves computationally? This section illustrates one possibility with reference to the Lewis Face-space-R model (Lewis, 2004), which is a formalisation of some aspects of the more generic face space proposals of Valentine (1991). Valentine's model postulated that faces are represented as locations in a multi-dimensional space. The nature and number of the dimensions is clearly of interest but also remains unclear. Busey (1998) obtained similarity ratings on a set of faces of bald men and then used multi-dimensional scaling to generate six dimensions: age; race; facial adiposity (plumpness); facial hair; aspect ratio (short-fat/long-thin); and facial hair colour. Shepherd, Ellis, and Davies (1977) used a similar method to derive just three dimensions: hair, face shape and age. It is clear that while such methods may describe the type of a face, they would not serve to distinguish individuals. In fact the dimensions are typical of those used during unfamiliar face processing, which is what the participants in their studies were doing. Familiar processing must be more subtle, since people are able to distinguish very similar faces such as 'identical' twins. Lewis (2004) obtained an estimate of between 15 and 22 dimensions.

Central to Face-space-R is the notion of a learned exemplar unit per face identity, whose activation is given by a Gaussian radial basis function (Broomhead & Lowe, 1988). That is, the unit's activation will peak when the input face has the correct value on each dimension and falls off like the normal distribution curve as the location of the face moves away from the optimal. These are a possible instantiation of the Face Recognition Units (FRUs) in the box model of Bruce and Young (1986). All the exemplar units (one per known identity) respond to some (perhaps vanishingly small) extent when a face is input and a competition between them decides which is the closest.

Fig. 5 shows a simple, one dimensional illustration of the basic coding mechanism of Face-space-R. The x-axis represents some arbitrary dimension in face space. There are four known faces depicted, represented by the four solid lines. Each of these Gaussian curves represents the response sensitivity to an input in that location on the x-axis. The Gaussians have different widths, reflecting the variability of that face on that dimension. The dotted and dashed curves represent face image inputs: in Face-space-R, these also appear as a Gaussian shaped zone of activation on the dimension. The initial response of each of the four

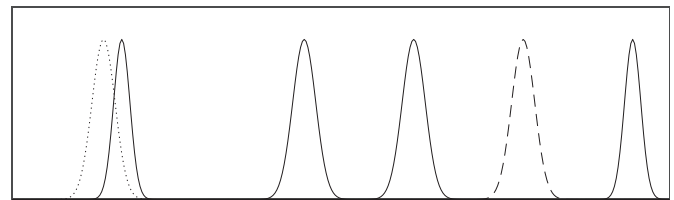


Fig. 5. A one dimensional illustration of Face-space-R. See text for explanation.

known face units depends on the overlap between their response curves and the input activation curve. The dashed line corresponds to an unfamiliar face that does not closely resemble any known face. The dotted line corresponds to a face that is similar to the left-most of the known faces and will therefore produce a response, albeit not as strong as an actual image of the known face.

An explanation for the Experiment 1 data in Face-space-R would be that a relative's face is similar enough, across enough dimensions, to cause sufficient activation in the relevant FRU to trigger a sense of familiarity. If the similarity is strong enough then familiarity might give way to false recognition; one sister photograph was sometimes mistaken for the person herself. The familiarity signal makes the difference, without it the participant is left trying to compare image similarities across the various dimensions.

For Experiments 2 and 3, there are two potential modes of operation. If the similarity is close enough to a known face to go above some threshold and activate the semantic information for the face (as in, that looks like Fred), then that semantic label becomes part of the stored memory for the face. If it does not, there will still be a signal from the FRU, and the proposal here is that this signal itself becomes part of the stored memory of the face. This additional signal makes the difference in the unprompted condition of Experiment 3. Provided the test exemplar image is similar enough it will generate the same familiarity signal, supplementing the stored representation of the appearance of the studied face and enhancing recognition accuracy.

Within the model, the spread (SD of the Gaussian) of the receptive field will have a big effect on activation. For simplicity the simulations in Lewis (2004) used a fixed SD, but in practice this would surely vary, both across the various dimensions for a given face and between different faces. As specified, Face-space-R is not a learning model but there would have to be a process that adjusts the centres and spread of the radial basis functions to match a new identity. Burton, Kramer, Ritchie, and Jenkins (2016) present evidence that faces have idiosyncratic modes of variation and that part of learning a face is learning how its appearance varies. Within the Face-space-R model, this would be accommodated by the receptive field having a spread for each dimension that is proportional to the variability that identity shows on that dimension. If genetic (or perhaps environmental) relatedness also influences the way in which faces vary in appearance (e.g. having similar smiles) it would increase the likelihood of a relative's face generating a familiarity signal.

6. General discussion

Familiar face processing is remarkably accurate; given enough experience, it is possible to distinguish 'identical' twins at a glance. By contrast, unfamiliar face processing is effortful and error-prone, even for matching different pictures taken at the same time (Bruce et al., 1999). Young and Burton (2018) therefore argue that we are truly expert only at processing familiar faces. The novel proposal here is that this makes familiar faces into 'islands of expertise' within the space of possible faces and that faces that resemble those we are familiar with should inherit some of that expertise. Experiment 1 showed that we are better able to identify relatives of familiar people. Experiments 2 and 3 showed that we are better able to remember unfamiliar faces that resemble a known one, even when not prompted to think about the resemblance at study.

These findings and the islands of expertise model are supported by

Strathie, Hughes-White, and Laurence (2020), work which was prompted by a conference presentation of results from Experiment 1. They use famous individuals and their less well known siblings to show that participants are better able to identify relatives of familiar faces. With their image set, which used unfamiliar Spanish celebrities as the control group, this manifested as being less able to identify the related image in the unfamiliar condition. Their second experiment reports that it is easier to match different pictures of an individual if they are related to someone who is already familiar (e.g. two pictures of Brad Pitt's brother Doug). Strathie et al. (2020) go on to show that these structural similarities can be captured in a Principal Components Analysis (PCA) model. Face images are better reconstructed from a PCA model derived from images of their sibling than are unrelated identities. The implication is that siblings share some common variability in appearance which both humans and a PCA model can identify.

One of the findings of the first experiment is that similarity ratings increase with familiarity for the correct pair but are little affected for the incorrect pairs. A possible explanation, in terms of the Face-space-R analysis presented above, is that similarity judgements are largely carried out in the featural domain of unfamiliar face processing. That is, given this task, participants estimate the resemblance of features and general characteristics, such as face shape. However, with the correct pair, the relative's face may provoke some response in the face recognition unit for the known target face, adding a sense of familiarity to boost the similarity estimate. For the incorrect pairing, the relative face will not cause any such activation and the more basic feature matching holds sway. Knowledge of the target face has little effect on the perceived dissimilarity of the relative's face. Lorusso, Brelstaff, Brodo, Lagorio, and Grosso (2011) suggest that judgement of similarity and dissimilarity may be different processes. This is consistent with the observation from face matching studies that performance on hits and false positives do not correlate (Megreya & Burton, 2006). Some people do well at telling faces apart, others at telling them together; the two abilities seem to be largely independent. It may be that the added familiarity signal that derives from an unfamiliar face resembling a known one helps with matching but not so much with rejecting mismatches. This is consistent with the results from Strathie et al. (2020) where familiarity effects are shown only in the matching conditions.

The island of expertise model offers a potential explanation for the abilities of 'super-recognizers' (Russell, Duchaine, & Nakayama, 2009): individuals with face recognition performance more than two standard deviations above average. Someone who naturally has markedly better face recognition than average will tend to learn and remember more faces. This will populate their face space, such that a novel face will more often resemble one that is already known. This will in turn make the novel face more memorable, bootstrapping the process so that their abilities continue to grow. When the space is sufficiently populated, the internal process might work like triangulating a radio source: a resemblance to more than one known face allowing a more precise fix on the appearance of the novel face and its likely variability. This would further improve the resolution and accuracy of processing of unfamiliar faces. In common with typical participants, super-recognizers show an 'other race' deficit (Bate et al., 2019; Robertson, Black, Chamberlain, Megreya, & Davis, 2020), albeit still performing better than controls. While their intrinsic face processing ability still helps, they lose the advantage afforded by already knowing many similar-looking people. They have fewer islands of expertise for less familiar facial types. At the other end of the scale, people with well below average recognition learn relatively few faces and would rarely benefit from prior resemblance.

A prediction from the results of Experiment 3 is that different participants should consistently remember different faces within a set. This would require retesting the same participants using the same identities (possibly using different photographs) some time apart. Quite how long apart would be hard to judge; individuals with good memory for faces might be able to recall the previous test if it is done too soon. This would confound the results since a face that was, for whatever reason,

memorable the first time would be likely to stick in the memory better.

7. Conclusion

A variety of experimental evidence suggests that we are truly expert only in the faces that we know well (Young & Burton, 2018). The three experiments reported here suggest that some of this enhanced expertise extends to faces that resemble known identities. That relatively few faces provoke a 'resemblance' response suggests that, for most of us, our internal 'face space' is sparsely filled, otherwise every new face would be well coded in terms of those we already know.

Ethics

All experimental studies were approved by local ethics committees and conducted in accordance with British Psychological Society guidance. All participants gave informed consent.

Data availability

Data files are available at <https://osf.io/rz24m/>.

Author statement

I am the sole author of this work.

Acknowledgements

Lauren Bulloch collected data for Experiment 1, Rebecca Riches for Experiment 2. Thanks to friends and colleagues for providing photographs for experiment 1. Dan Carragher helped remove some of the ambiguities in my writing. The work was supported by the UK Engineering and Physical Sciences Research Council [grant number EP/N007743/1, Face Matching for Automatic Identity Retrieval, Recognition, Verification and Management].

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cognition.2021.104765>.

References

- Alvergne, A., Oda, R., Faurie, C., Matsumoto-Oda, A., Durand, V., & Raymond, M. (2009). Cross-cultural perceptions of facial resemblance between kin. *Journal of Vision*, 9(6). ISI:000272270500023.
- Angeli, A., Davidoff, J., & Valentine, T. (2008). Face familiarity, distinctiveness, and categorical perception. *The Quarterly Journal of Experimental Psychology*, 61(5), 690–707.
- Bate, S., Bennetts, R., Hasshim, N., Portch, E., Murray, E., Burns, E., & Dudfield, G. (2019). The limits of super recognition: An other-ethnicity effect in individuals with extraordinary face recognition skills. *Journal of Experimental Psychology: Human Perception and Performance*, 45(3), 363–377. <https://doi.org/10.1037/xhp0000607>.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>.
- Beale, J. M., & Keil, F. C. (1995). Categorical effects in the perception of faces. *Cognition*, 57(3), 217–239.
- Bredart, S., & French, R. M. (1999). Do babies resemble their fathers more than their mothers? A failure to replicate Christenfeld and Hill (1995). *Evolution and Human Behavior*, 20(2), 129–135.
- Bressan, P., & Dal Martello, M. F. (2002). Talis pater, talis filius: Perceived resemblance and the belief in genetic relatedness. *Psychological Science*, 13(3), 213–218.
- Bressan, P., & Grassi, M. (2004). Parental resemblance in 1-year-olds and the Gaussian curve. *Evolution and Human Behavior*, 25(3), 133–141.
- Broomhead, D. S., & Lowe, D. (1988). Multivariable functional interpolation and adaptive networks. *Complex Systems*, 2, 321–355.
- Bruce, V., Henderson, Z., Greenwood, K., Hancock, P. J. B., Burton, A. M., & Miller, P. I. (1999). Verification of face identities from images captured on video. *Journal of Experimental Psychology: Applied*, 5(4), 339–360.
- Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, 77(3), 305–327.

- Burton, A. M., Kramer, R. S. S., Ritchie, K. L., & Jenkins, R. (2016). Identity from variation: Representations of faces derived from multiple instances. *Cognitive Science*, 40(1), 202–223. <https://doi.org/10.1111/cogs.12231>.
- Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow face matching test. *Behavior Research Methods*, 42(1), 286–291.
- Burton, A. M., Wilson, S., Cowan, M., & Bruce, V. (1999). Face recognition in poor-quality video: Evidence from security surveillance. *Psychological Science*, 10(3), 243–248.
- Busey, T. A. (1998). Physical and psychological representations of faces: Evidence from morphing. *Psychological Science*, 9(6), 476–483.
- Carey, S. (1992). Becoming a face expert. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 335(1273), 95–103. <https://doi.org/10.1098/rstb.1992.0012>.
- Christenfeld, N. J. S., & Hill, E. A. (1995). Whose baby are you. *Nature*, 378(6558), 669.
- Clutterbuck, R., & Johnston, R. A. (2004). Matching as an index of face familiarity. *Visual Cognition*, 11(7), 857–869.
- Curby, K. M., & Gauthier, I. (2009). The temporal advantage for individuating objects of expertise: Perceptual expertise is an early riser. *Journal of Vision*, 9(6), 7. <https://doi.org/10.1167/9.6.7>.
- Dal Martello, M. F., & Maloney, L. T. (2006). Where are kin recognition signals in the human face? *Journal of Vision*, 6(12), 2. <https://doi.org/10.1167/6.12.2>.
- Dwyer, D. M., & Vladeanu, M. (2009). Perceptual learning in face processing: Comparison facilitates face recognition. *Quarterly Journal of Experimental Psychology*, 62(10), 2055–2067.
- Felisberti, F. M., & Pavey, L. (2010). Contextual modulation of biases in face recognition. *PLoS One*, 5(9), Article e12939. <https://doi.org/10.1371/journal.pone.0012939>.
- Gauthier, I., Skudlarski, P., Gore, J., & Anderson, A. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience*, 3(2), 191–197.
- Hancock, P. J. B., Bruce, V., & Burton, A. M. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Sciences*, 4(9), 330–337.
- Hole, G. J. (1994). Configurational factors in the perception of unfamiliar faces. *Perception*, 23(1), 65–74. <https://doi.org/10.1068/p230065>.
- Johnston, R. A., & Edmonds, A. J. (2009). Familiar and unfamiliar face recognition: A review. *Memory*, 17(5), 577–596.
- Kikutani, M., Roberson, D., & Hanley, J. R. (2008). What's in the name? Categorical perception for unfamiliar faces can occur through labeling. *Psychonomic Bulletin & Review*, 15(4), 787–794.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(1), 1–26. <https://doi.org/10.18637/jss.v082.i13>.
- Levin, D. T., & Beale, J. M. (2000). Categorical perception occurs in newly learned faces, other-race faces, and inverted faces. *Perception and Psychophysics*, 62(2), 386–401.
- Lewis, M. B. (2004). Face-space-R: Towards a unified account of face recognition. *Visual Cognition*, 11(1), 29–69.
- Lorusso, L., Brelstaff, G., Brodo, L., Lagorio, A., & Grosso, E. (2011). Visual judgments of kinship: An alternative perspective. *Perception*, 40(11), 1282–1289. <https://doi.org/10.1068/p6916>.
- Mclain, D. K., Setters, D., Moulton, M. P., & Pratt, A. E. (2000). Ascription of resemblance of newborns by parents and nonrelatives. *Evolution and Human Behavior*, 21(1), 11–23.
- Megreya, A. M., & Burton, A. M. (2006). Unfamiliar faces are not faces: Evidence from a matching task. *Memory & Cognition*, 34(4), 865–876.
- Mileva, M., Young, A. W., Kramer, R. S. S., & Burton, A. M. (2019). Understanding facial impressions between and within identities. *Cognition*, 190, 184–198. <https://doi.org/10.1016/j.cognition.2019.04.027>.
- Nesse, R. M., Silverman, A., & Bortz, A. (1990). Sex-differences in ability to recognize family resemblance. *Ethology and Sociobiology*, 11(1), 11–21.
- Oda, R., Matsumoto-Oda, A., & Kurashima, O. (2005). Effects of belief in genetic relatedness on resemblance judgements by Japanese raters. *Evolution and Human Behavior*, 26(5), 441–450.
- R Core Team. (2021). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rhodes, G., Hayward, W. G., & Winkler, C. (2006). Expert face coding: Configural and component coding of own-race and other-race faces. *Psychonomic Bulletin & Review*, 13(3), 499–505.
- Richler, J. J., Mack, M. L., Gauthier, I., & Palmeri, T. J. (2009). Holistic processing of faces happens at a glance. *Vision Research*, 49(23), 2856–2861. <https://doi.org/10.1016/j.visres.2009.08.025>.
- Ritchie, K. L., Kramer, R. S. S., & Burton, A. M. (2018). What makes a face photo a “good likeness”? *Cognition*, 170, 1–8. <https://doi.org/10.1016/j.cognition.2017.09.001>.
- Robertson, D. J., Black, J., Chamberlain, B., Megreya, A. M., & Davis, J. P. (2020). Super-Recognisers show an advantage for other race face identification. *Applied Cognitive Psychology*, 34(1), 205–216. <https://doi.org/10.1002/acp.3608>.
- Russell, R., Duchaine, B. C., & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic Bulletin & Review*, 16(2), 252–257. <https://doi.org/10.3758/PBR.16.2.252>.
- Saether, L., & Laeng, B. (2008). On facial expertise: Processing strategies of twins' parents. *Perception*, 37(8), 1227–1240.
- Shepherd, J. W., Ellis, H. D., & Davies, G. M. (1977). *Perceiving and remembering faces*. Home Office.
- Strathie, A., Hughes-White, N., & Laurence, S. (2020). The sibling familiarity effect: Is within-person facial variability shared across siblings? *PsyArXiv*. <https://doi.org/10.31234/osf.io/6j8ek>.
- Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 43A(2), 161–204.
- Vokey, J. R., & Read, J. D. (1988). Typicality, familiarity and the recognition of male and female faces. *Canadian Journal of Psychology*, 42(4), 489–495.
- Wickham, L. H. V., Morris, P. E., & Fritz, C. O. (2000). Facial distinctiveness: Its measurement, distribution and influence on immediate and delayed recognition. *British Journal of Psychology*, 91(1), 99–123. <https://doi.org/10.1348/000712600161709>.
- Young, A. W., & Burton, A. M. (2018). Are we face experts? *Trends in Cognitive Sciences*, 22(2), 100–110. <https://doi.org/10.1016/j.tics.2017.11.007>.