Facing the facts: Naive participants have only moderate insight into their face recognition and face perception abilities.

Anna K. Bobak, Viktoria R. Mileva & Peter J.B. Hancock

Psychology, Faculty of Natural Sciences, University of Stirling, UK

Word count (main text): 4578

Address for correspondence

Anna K. Bobak

Psychology

Faculty of Natural Sciences

Cottrell Building

University of Stirling

Stirling

FK9 4LA

United Kingdom

Telephone: +44(0) 1786 466367

Email: a.k.bobak@stir.ac.uk

Abstract

A reliable self-report measure to assess the broad spectrum of face recognition ability (FRA) from developmental prosopagnosia (DP) to super-recognition (SR) would make a valuable contribution to initial screening of large populations. We examined performance of 96 naive participants and seven SRs, using a range of face and object processing tasks and a newly developed 20-item questionnaire, the Stirling Face Recognition Scale (SFRS). Overall, our findings suggest that young adults have only moderate insight into their FRA, but those who have been previously informed of their (exceptional) performance, the SRs, estimate their FRA accurately. Principal Component Analysis of SFRS yielded two components. One loads on questions about low ability and correlates with perceptual tasks and one loads on questions about high FRA and correlates with memory for faces. We recommend that self-report measures of FRA should be used in addition to behavioural testing, to allow for cross-study comparisons, until new, more reliable instruments of self-report are developed. However, self-report measures should not be solely relied upon to identify highly skilled individuals. Implications of these results for theory and applied practice are discussed.

Key words: face recognition; super-recognisers; self-report; face perception; individual differences

1. Introduction

There are large individual differences in human face recognition ability (FRA) with some people excelling at this task, so-called super–recognisers (SRs; Russell, Duchaine, & Nakayama, 2009) and others struggling with recognising people from their faces owing to a condition called developmental prosopagnosia (DP). People with DP are unable to recognise both familiar and unfamiliar faces despite having no other associated deficits in visual acuity, intelligence, or known brain injury (Jones & Tranel, 2001).

While many face recognition researchers have concentrated their efforts in the last decade on mapping the cognitive and perceptual underpinnings of DP, much recent work is concerned with the individuals on the high end of the face recognition spectrum, the SRs. With the seminal study by Russell and colleagues (Russell et al., 2009) identifying people at the high end of the FRA, recent attempts to examine this relatively new field have concentrated on cognitive and perceptual abilities (Bobak, Bennetts, Parris, Jansari, & Bate, 2016) , eye-movement strategies (Bobak, Parris, Gregory, Bennetts, & Bate, 2017) and the applied value of superior face recognition (Bobak, Hancock, & Bate, 2016c; Bobak, Dowsett, & Bate, 2016b; Robertson, Noyes, Dowsett, Jenkins, & Burton, 2016).

The theoretical and applied value of research into superior face recognition is of particular interest for a number of reasons. In terms of theory, if superior face recognition and DP represent the opposite ends of one distribution, it is plausible to assume that the processing deficits associated with prosopagnosia such as lack of global precedence (Bentin, DeGutis, D’Esposito, & Robertson, 2007), holistic processing (DeGutis, Cohan, Mercado, Wilmer, & Nakayama, 2012), or perception of facial identity (White, Rivolta, Burton, Al-Janabi, & Palermo, 2017) will have their inverse in SRs. If this hypothesis were true, this may pave the way for tailored rehabilitation programmes for people affected by DP. The applied value of super-recognition research should also not be overlooked. Specifically, the extraordinary ability to match faces (Bobak et al., 2016; Robertson et al., 2016) and recognition of faces from video footage or line-ups (Bobak et al., 2016) would be of particular value to national security agencies. Pertinently, the London Metropolitan Police has a dedicated Super-Recogniser Unit working towards solving criminal investigations where face identity information, however scant, is available. The Unit has helped in solving several high profile cases, such as the Alice Gross inquiry (The New Statesman, 2016) and a series of upmarket thefts in London (The Guardian, 2016). However, the screening and classification of SRs remains a challenging issue.

In the context of extraordinary FRA, it is also important to mention studies with forensic examiners (see White, Norell, Phillips, & O'Toole, 2017 for a review). Surprisingly, facial examiners tend to show more featural processing strategies and display a reduced inversion effect (White, Phillips, Hahn, Hill, & O'Toole, 2015). Their performance on FRA tasks is important to note for several reasons: Firstly, forensic examiners receive training and have considerable experience in facial image comparison (White, Kemp, Jenkins, Matheson, & Burton, 2014); Secondly, it can be assumed that their motivation to perform well is high; Finally, it is possible that individuals aware of their high FRA self-select for the occupations where facial image comparison and person identifications are an inherent part of the job.

To date, authors predominantly classify SRs based on self-reported above average ability to memorise and discriminate faces, with scores of two standard deviations above the mean on the Cambridge Face Memory Test Long Form (CFMT+, Russell et al., 2009) and a supplementary, albeit not binding, score on the Cambridge Face Perception Test indicating enhanced face perception (CFPT, Duchaine, Germine, & Nakayama, 2007). The CFMT+ is a standardised test of episodic memory requiring participants to memorise six male faces and recognise them later in gradually degrading images over 102 trials. The CFPT asks participants to sort a line-up of faces in order of similarity to a target presented above, i.e. it has no memory component. Both these tests have been primarily developed to classify developmental and acquired prosopagnosia (DP and AP; Duchaine et al., 2007; Duchaine & Nakayama, 2006) and have become the gold standard neuropsychological assessment tools used by several laboratories worldwide (e.g. Bate & Bennetts, 2014). Additionally, some SRs appear to show exceptional performance at face matching (Bobak et al., 2016b; Bobak et al., 2016c), but such paradigms have not been typically included in the initial screening of self-reported SRs.

Although laboratory screening using a standardised assessment battery is the most rigorous means of testing an individual's FRA, it is not always the most appropriate when participants are located remotely or a face-to-face meeting is not possible. While most behavioural tests can be administered online, this method does not allow researchers to control for testing conditions such as lighting, wearing of glasses, or other distractions in participants' homes, all of which are important when relying on accuracy and reaction times in FRA assessment. A reliable screening tool that could be completed in a timely fashion prior to in-depth behavioural testing would be useful for testing large numbers of participants (e.g. recruits of national security agencies) or those in remote locations prior to inviting the selected subset for further testing in laboratories.

A self-report face recognition scale is one candidate for such test. A recent study by Shah, Sowden, Gaule, Catmur, & Bird (2015) reported strong negative correlations between a new self-report scale (PI-20) aimed to measure prosopagnosic traits and behavioural measures of FRA (CFMT and Famous Faces Test), but not object recognition (CCMT; Dennett et al., 2012). The authors concluded that the PI-20 may be a useful tool for screening large populations, such as police or border control officers to identify those with prosopagnosia. However, the sample of participants tested in this study contained both naive participants and individuals with prosopagnosia who were perhaps aware of their face recognition deficits following initial classification to the "DP" group. Pertinently, individuals with DP have been reported to have poor insight into their face-specific visual imagery (Tree, 2011; Tree & Wilkie, 2010) indicating that, similarly to typical perceivers, their self-reported cognitive abilities should be interpreted with caution.

This issue was addressed in a recent study by Gray, Bird, and Cook (2017) who reported robust associations between the PI-20 and the CFMT in two large samples of young adults. However, while highly significant, the correlations between the CFMT and the PI-20 were moderate ($r = 0.394$ and $r = 0.390$) and explained 15% of the variance in the CFMT data. Furthermore, a recent replication with a Portuguese sample (Ventura, Livingston, & Shah, in press) yielded a .40 correlation between the PI-20 and the CFMT attesting this moderate effect size is robust and replicable. The PI-20 correlated more strongly with the Glasgow Face Matching Test (GFMT, Burton, White, & McNeill, 2010), a task where participants have to decide if two images presented side by side belong to the same person or two different people (Shah et al., 2015). The PI-20 explained 24% of the variance ($r = .49$, $p < .001$) in the abbreviated GFMT. It is thus possible that PI-20 taps more into face perception, rather than face memory ability.

Additionally, Livingston and Shah (2017) re-analysed data from the Shah et al. (2015) study and reported that for validation study 2, the correlations between the PI-20 and the CFMT are $r = -.62$ and $r = -.70$ (highly significant) for DPs and typically developing individuals, respectively. However, the analyses were conducted with small subject groups and a large scale replication of these results is needed before drawing firm conclusions about the reliability of PI-20 in detecting prosopagnosia.

Other authors reported small effect sizes associated with self-reported FRA and objective measures of facial recognition. Bindemann and colleagues (Bindemann, Attard, & Johnston, 2014) showed that people have no insight into their ability to recognise unfamiliar faces and that this effect may be driven by conflating familiar and unfamiliar face processing. Furthermore, a recent study by Bobak, Pampoulov, and Bate (2016d) yielded small correlations between single questions about self-perceived FRA and the face recognition and face perception tests. Similar results were presented by Palermo and colleagues (2017) who showed that across five tests, self-reported FRA and various objective measurements yield small to medium effect sizes.

One reason for the above studies reporting weak correlations may have been a small number of items (e.g. single question in Bobak et al., 2016d) used to examine self-reported FRA. Nonetheless, a recent Italian study yielded a highly significant, albeit also weak, correlation between the CFMT scores and a 25 –item theory-driven questionnaire of face cognition (Turano & Viggiano, 2016).

Given the recent resurrection of the debate into the feasibility of finding a reliable self-reported face recognition measure/scale, and its applicability to national security settings where face processing tasks are an everyday part of the job, in this study we modified the PI-20 questionnaire (see supplementary material) by adding several questions based on personal communications of the authors with the so-called SRs. In this way we hoped to create a scale that would encompass the entire spectrum of FRA – from DP to SR, whereas the PI-20 was specifically designed with prosopagnosia in mind. The main aim of this study was to examine whether this new comprehensive self-report questionnaire, which we called the Stirling Face Recognition Scale (SFRS), would be able to predict participants' performance on standardised measures of face processing.

2. Method

2.1. Participants

Ninety nine students and visitors at University of Stirling took part in the study. Three participants were removed due to coming from a non-White Caucasian ethnic background and living in the UK for less than three years. This strategy was adopted to ensure social contact sufficient for individuation of faces used in this study (Meissner & Brigham, 2001). One further participant was removed from analyses due to completing the entire study in approximately 15 minutes and thus suggesting a lack of engagement in the tasks. In comparison, the other participants took on average 30-45 minutes to complete the study. The remaining 96 participants (25 male, 1 undisclosed) were included in the final analyses. The mean age of the sample was 23, SD = 4, range 18 – 35 years. All bar four participants were White Caucasian. The non – Caucasian participants all reported to have been living in the UK for more than three years. In addition to the sample of naive observers, seven SRs (four female), who all previously received feedback on their face recognition performance, were asked to fill in the SFRS. The mean age was 32, SD = 10, range 21 – 49 years. All bar one SR were of White Caucasian ethnicity and all had been living in the UK for over three years.

Furthermore, to increase the reliability of the Principal Component Analysis, we included data for the SFRS for participants from two undergraduate studies. Twenty participants (four male) aged 26 years (*SD* = 13, range 18-52) took part in Study 1 and 49 participants (18 male) aged 24 (*SD* = 9.8, range 18-58) took part in Study 2. Please note, these individuals only completed the SFRS and are thus not included in correlational analyses.

2.2. Materials

Stirling Face Recognition Scale (Items 2, 3, 4, 5,10,14,18 adapted from PI-20; Shah, Sowden, Gaule, Catmur, & Bird, 2015). The scale comprised of 20 face-recognition specific questions about one's experience in relation to their face recognition ability (see supplementary material). The scale included ten positive and ten reversed items (predominantly adapted from the PI-20 questionnaire) with the aim to encompass questions applicable to both the low and high end of the face recognition spectrum. Possible responses ranged from strongly disagree (score = 1) to strongly agree (score = 7). The maximum possible score was 140, there was no time limit on the completion of the scale, and participants were encouraged to answer the questions as honestly as possible.

Cambridge Face Memory Test Long Form (CFMT+; Russell et al., 2009). The upright version of this task is an extension of the original CFMT (Duchaine & Nakayama, 2006) routinely used to "diagnose" DP. In the test, participants study six male greyscale identities that have been cropped of all external features (ears and hair) and later choose the target faces amongst two distractors. The CFMT+ consists of 102 trials split into four sections with gradually increasing difficulty. First, the "learning" phase comprises of 18 trials and the studied images are identical to those later presented at test. Participants then review all six identities for 20 seconds and proceed to the "novel" phase, where the target identities and distractors are presented under novel lighting and varying poses over 30 trials. The subsequent two phases (24 and 30 trials respectively), present images that become progressively more difficult due to the overlaying of noise, various facial expressions, and the presence of hair (not included at the study stage). The CFMT+ is currently mainly used to classify SRs in psychological research (e.g. Bobak et al., 2016; Russell, Chatterjee, & Nakayama, 2012; Russell et al., 2009).

Cambridge Car Memory Test (CCMT; Dennett et al., 2012). The CCMT was included in this study as a control test of object memory. The CCMT comprises of 72 trials and was devised as a counterpart to the original, short, CFMT. Similarly to the face memory task, the trials become progressively more difficult throughout the test and the only discrepancy between these assessments is that images of cars are used instead of face identities.

Models Face Matching Task (MFMT; Dowsett & Burton, 2015). The MFMT is a simultaneous face matching test where two images of male identities (with hair, ears, and clothing present) are displayed on the screen side by side and the participants' task is to decide whether the two images show the same person or two different people. The original set comprises of 90 pairs of photographs, 45 matched and 45 mismatched (77% overall accuracy), divided into three blocks of equal difficulty (15 matched and 15 mismatched trials each). For the purpose of this study we asked participants to complete one block (30 trials in total). There is no time limit for completion of this task.

2.3.Procedure

All naive participants took part in the following tasks in the same order[1] either voluntarily (as a part of a third year student project) or for a small monetary reimbursement. Data were collected on students' personal computers or machines available in the Psychology Department. Participants were instructed to respond in their own time in all tasks; their response times were not restricted or recorded. The study was approved by the Departmental

[1] The SRs, for whom the CFMT+ scores were known prior to the commencement of this study completed only the SFRS. This is because some, but not all, SR participants have previously taken part in the extended version of the MFMT with the same images used in this task. It was thus inappropriate to repeat the shortened assessment and the analyses including these seven individuals are only reported for the CFMT+.

Ethics Committee. All volunteers signed an informed consent and were fully debriefed following the end of the study. Data analysis was performed using the SPSS 23 software.

3. Results

Overall, the SFRS had very good internal reliability of Cronbach's α = .88, Split-half Spearman- Brown coefficient = .89 suggesting that all items were worthy of retention.

3.1. Descriptive Statistics

Table 1. Shows the number of participants and mean scores for all tasks employed in this study.

Table 1. Descriptive statistics for all tests employed in the study

Task (scale)	*N* (excluding SRs)	Mean *(SD)*	Range
SFRS (total items)	165	101.5 *(14.9)*	58 - 139
CFMT (total items)	96	66.4 *(13.3)*	38 - 96
CCMT (total items)	96	46.4 *(10.9)*	24 - 72
MFMT (% correct)	96	*74.42 (10.2)*	50- 93

3.2.Principal Component Analysis (PCA).

The SFRS responses for all naive participants (*N* = 165) were entered into a PCA model with two components (as indicated by parallel analysis). There were no theoretical reasons to assume that the resulting components would be entirely orthogonal, therefore we applied the oblimin rotation to the analysis. Our findings suggest that *Component 1* reflected primarily questions about being *poor* at face recognition, including both face memory and general face processing (social factors and contextual recognition), while *Component 2* encompassed mainly questions about being *good* at face recognition, especially face memory (see *Table 2*); in combination both factors explained 40.6% of the variance.

Table 2. Pattern matrix for factor analysis of the SFRS.

Item	Rotated component loadings (oblimin)	
	Face processing (1)	Face memory (2)
Q19. I often fail to recognise someone who knows me.	.71	
Q16. I sometimes find movies hard to follow because of difficulties recognizing characters.	.67	
Q10. When people change their hairstyle, or wear hats, I have problems recognizing them.	.67	
Q14. I have to try harder than other people to memorize faces.	.66	
Q12. I sometimes have to warn new people I meet that I am 'bad with faces'.	.63	
Q18. I struggle to recognize people without hearing their voice.	.59	
Q4. Anxiety about poor face recognition has led me to avoid certain social or professional situations.	.58	
Q6. I often mistake people I have met before for strangers.	.55	
Q2. When I was at school I struggled to recognize my classmates.	.33	
Q3. I find it very easy to picture individual faces in my mind with great detail.		.73
Q11. I am very good at spotting family resemblance in a group of unfamiliar people.		.61
Q13. I often recognise people I have met briefly before, but they have no idea who I am.		.60
Q15. I 'never forget a face'.	.38	.57
Q5. I am better than most people at putting a 'name to a face'.		.56
Q1. I know exactly where I first met someone (work meetings, parties, etc.).		.52
Q17. I find it easy to recognise my family and friends from their childhood photographs.		.52

Q9. Faces are enough for me to recognise people- I don't need to hear their voice or see their whole body.	.37	.48
Q20. I easily recognise someone even if I can only see part of their face.		.48
Q7. I tend to forget faces very quickly after seeing them.	.37	.46
Q8. I find it very easy to recognise familiar people when I meet them out of context (e.g. meeting a work colleague unexpectedly in the gym).		.39
Eigenvalues	6.41	1.71
% of variance	32.04	8.54

Both components were subsequently entered into correlational analyses with the three behavioural tests (CFMT+, MFMT, and the CCMT). *Component 1* correlated moderately, but significantly with both the CFMT+ (r =.28, p = .006, N = 96) and the MFMT (r =.34, p = .001, N = 96), and *Component 2* yielded a moderate association with CFMT+ (r =.32, p = .001, N = 96) only (Figure 1).

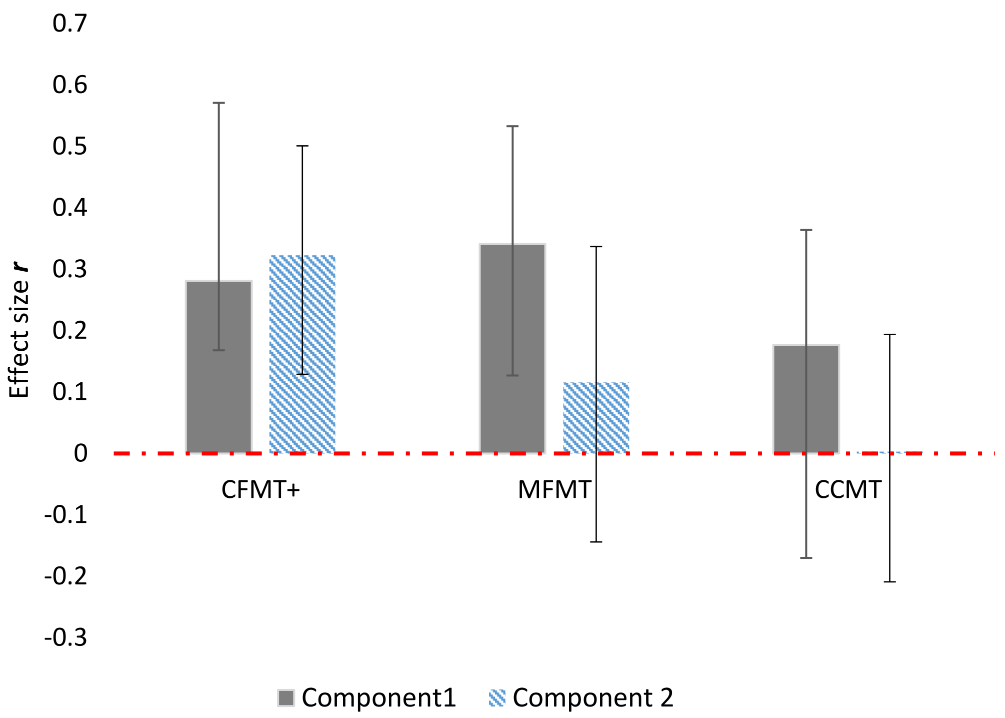

Figure 1. Effect sizes for correlations between PCA components and behavioural tasks. Error bars represent 95% CIs for coefficient *r*.

3.3.Correlational analyses[2]

Scores from three behavioural tasks were correlated with the SFRS using Pearson's correlations (Table 3). These analyses revealed moderate correlations between all variables. The SFRS explained 12.9% and 7.9% of variance in performance on the CFMT+ (face recognition memory) and the MFMT (face perception) respectively. Importantly, the scale did not correlate with the CCMT (object memory task) suggesting that the questionnaire did not tap into general object processing ability and that the questionnaire is, indeed, face-specific.

[2] Correlations between single items of the SFRS and the behavioural measures employed in this study are available to access in supplementary materials.

Notably, Bayesian analyses provided very strong support for the relationship between the SFRS and CFMT+, but only moderate evidence for the relationship between the SFRS and the MFTMT, a task without a memory component. Evidence for the dissociation of the SFRS and the non-face CCMT was also moderate (see Table 3).

Table 3. The correlations (parametric and Bayes Factors) between self-perceived FRA and objective tests of face and object processing ability; *N* = 96. 95% Confidence Intervals for *r* coefficients are in parentheses.

Variables			1	2	3	4
1.	SFRS	Pearson's r	–			
		BF_{10}	–			
2.	CFMT+	Pearson's r	.36** [.18, .51]	–		
		BF_{10}	75.4	–		
3.	MFMT	Pearson's r	.28* [.06, .50]	.46** [.30, .60]	–	
		BF_{10}	5.60	6905	–	
4.	CCMT	Pearson's r	.12 [-.08, .30]	.23* [.03, .41]	.36** [.16, .53]	–
		BF_{10}	0.24	1.47	68.2	–

*Correlation significant at .05 level
**Correlation significant at .01 level

For the sake of completeness, we carried out additional analyses partialling out any effects of age and gender in the associations between the SFRS and the three behavioural tasks. The correlations and 95% CIs are presented in Table 4. The SFRS remained moderately associated with the CFMT+ and the MFMT, but not the CCMT suggesting that it is tapping into face-specific cognitive process.

Table 4. Partial correlations between self-perceived FRA and objective tests of face and object processing ability; *N* = 93. 95% Confidence Intervals for *r* coefficients are in parentheses.

Variables	Control variable	1	2	3	4	
1. SFRS	Age	–				
	Gender	–				
2. CFMT+	Age	.41** [.23, .57]	–			Pe
	Gender	.36** [.18, .51]	–			BI
3. MFMT	Age	.31* [.06, .51]	.43** [.25, .58]	–		Pe
	Gender	.27* [.03, .49]	.46** [.30, 60]	–		BI
4. CCMT	Age	.15 [-.05, .33]	.16 [-.05, .36]	.32* [.12 .50]	–	Pe
	Gender	.09 [-.11, .28]	.22* [.03, .41]	.35** [.17, 52]	–	BI

*Correlation significant at .05 level
**Correlation significant at .01 level

To illustrate the point that inclusion of special populations may produce misleading results in correlational analyses, we included the SFRS scores from seven super-recognisers who had previously received feedback on their performance on the CFMT+. Thus, they were aware of their exceptional face processing abilities and some took part in previous studies on super-recognition. The previously classified SRs constituted 6.80% of the sample, a proportion over twice as large as that estimated in a recent paper by Bobak and colleagues (2016). Pertinently, the size of the effect was inflated by the inclusion of these individuals (see Figure 1, $N = 103$, $r = .50$, $p < .001$, 95% CI [.33, .63]. Although a Fisher's Z statistic for the difference between these two correlations was non-significant ($z = -1.17$, $p = .242$), the shared variance increased to 24.7% following the inclusion of participants informed about their FRA.

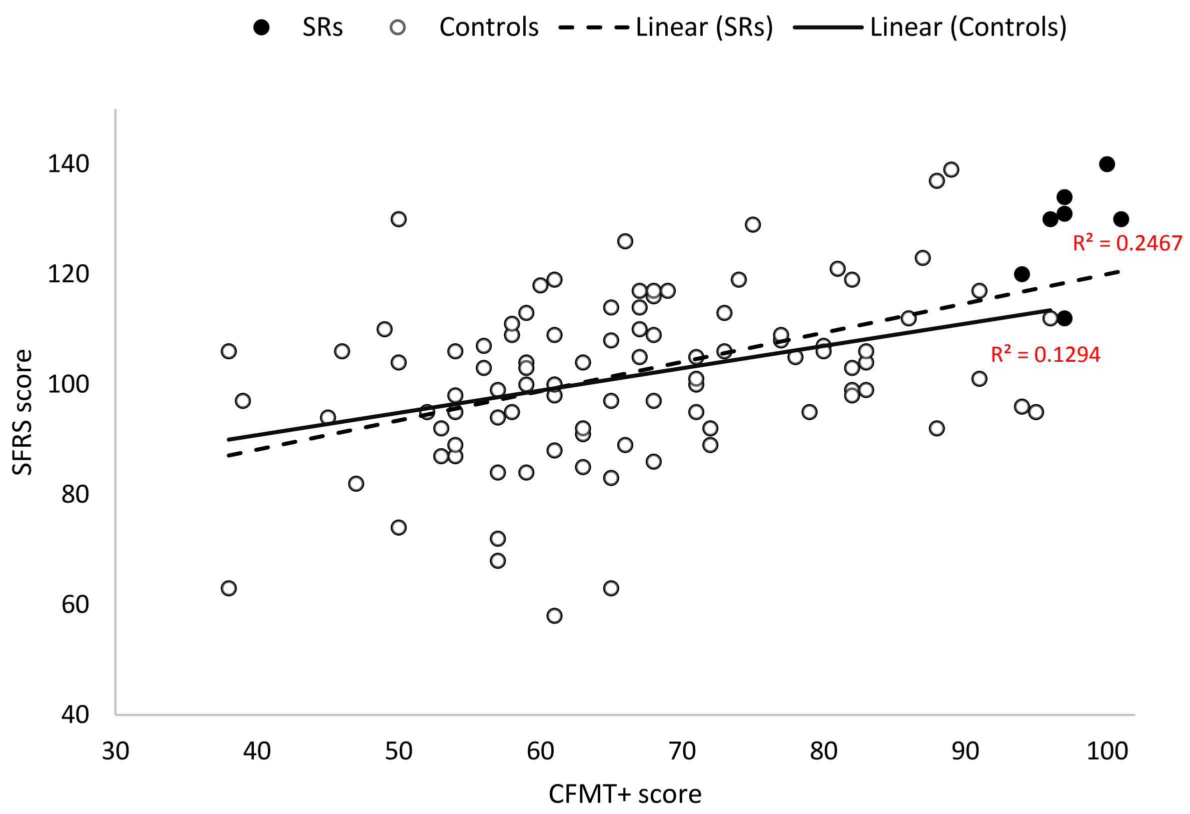

Figure 1. SFRS scores plotted against performance on the CFMT+ with and without the inclusion of six SRs

4. Discussion

To investigate whether people have insight into their own FRA we tested 96 participants using a new questionnaire with items sourced from personal communications with SRs and adapted from the recently published PI-20 scale (Shah et al., 2015). Specifically, statements such as “I know exactly where I first met someone” (1) or “I often recognise people I have met briefly before, but they have no idea who I am.” (13) were cited by SRs describing their experience in every-day life. Overall, our results suggest that young adults have only moderate insight into their FR and face perception abilities, but replicate the finding that multiple item scales can estimate it to some extent (Palermo et al., 2016) Specifically, Palermo and colleagues (2016) developed a new, 77-item scale to asses face processing ability and reported small to moderate correlations between the scale and several measures of face processing (e.g. r = .32 between the CFMT and the questionnaire). In a sample of undergraduate students not selected specifically for their FRA we noted a moderate, but strongly significant correlation with the CFMT+ and a small, but also significant correlation with the MFMT. There was no relationship between the SFRS and the CCMT, suggesting that the questionnaire taps well into FRA, rather than general visual memory (but see Gauthier et al., 2014; McGugin, Richler, Herzmann, Speegle, & Gauthier, 2012).

Pertinently, we report that individuals who have previously received feedback on their performance on laboratory tests of face processing, the SRs, are somewhat better at estimating their FRA using our self-report questionnaire. Scores on SFRS of five out of seven previously identified SRs were at least 1.9 SD above the unselected group's mean. It is plausible that they were using their prior knowledge from objective assessments to inform the answers in the questionnaire. These results are particularly important when validating any future self-report instruments of cognitive abilities and suggest that inclusion of special populations in otherwise opportunistically selected samples of participants may confound the outcomes of subsequent analyses. However, a recent re-analysis of the initial PI-20 data set (Shah et al., 2015) reported by Livingston and Shah (2017) and the moderate associations between the GFMT and PI-20 in a sample of naive participants (Shah et al., 2015) speak against this. In the present study, the correlations between the CFMT+ and the SFRS increased, albeit non-significantly when data from seven SRs (6.80% of the sample) were included in the analyses. This sampling discrepancy may be one underlying factor for the differences between studies reporting weak correlations between self-reported FRA (Bobak et al., 2016; Bowles et al., 2009; Palermo et al., 2016) and those asserting that individuals' insight is a reliable means of screening the general population and individuals in applied settings. An alternative possibility is that PI-20 taps more efficiently into individuals' face processing skills than other scales developed to measure self-reported FRA.

Perhaps the most striking finding here is the result of the PCA, which returned two components of the questionnaire data, explaining a cumulative variance of 40.6% (32.04% and 8.54%, respectively). The first factor loads mostly onto questions about being poor at face recognition, such as failing to recognise someone who knows me (Table 2). This factor correlates most strongly with performance on the MFMT but is also associated with the CFMT+. The second component loads mostly onto questions about being good at face recognition (e.g. I never forget a face) and correlates most strongly with performance on the CFMT. Our tentative interpretation is that there are two aspects to human face recognition ability, one that is primarily perceptual, which therefore correlates with the model matching task, but also underpins face memory, and one that is primarily about the ability to remember faces, and therefore correlates with the face memory task only. People who report being bad with faces may struggle perceptually, while those who report being good must also be able to remember them.

Megreya and Burton (2006) reported correlation in performance across a variety of face matching and recognition tasks. They found that ability to match unfamiliar faces correlated with other visual perception tasks, but not with ability to recognise familiar faces. We suggest that our first component, face matching ability, may equate with the ability identified by Megreya and Burton. More recently, Verhallen et al. (in press) also investigated associations between some standard measures of face recognition and a single-item self-report FRA scale. They identified only one underlying factor, that they term *f*, which correlated similarly strongly with both the CFMT and the Glasgow face matching task, as well as their single item FRA scale. It may be that a single-item scale is able to identify overall ability well enough but that a multiple question scale, such as our SFRS, is needed to identify the more subtle underlying components of human face perception

In general, the rather moderate pattern of correlations between the two behavioural face processing tasks and the SFRS mirrors the results from examinations of insight into other types of memory (Beaudoin & Desrichard, 2011), general intelligence (Paulhus, Lysy, & Yik, 1998), mathematical ability (Mason, 2003), or general cognitive ability in patients with frontal lobe epilepsy (Baños et al., 2004). Furthermore, a recent meta-analysis reported average correlation between ability self-evaluations and objective performance to be moderate (*M* = .29, *SD* = .11) and task dependent, i.e. self-report was more accurate when behavioural tasks were low in complexity and objective (Zell & Krizan, 2014). In an early report on DP, De Haan (1999) urged caution when interpreting self-reported FRA after noting that some individuals score within typical range on objective face processing tasks, despite reporting problems in everyday life. Conversely, other authors have recently shown that naive participants are frequently surprised when they find out their face recognition is poorer than what they had originally thought (Bowles et al., 2009; Grueter et al., 2007). This may be because despite profound deficit in FRA, they are still able to navigate successfully in everyday life by employing compensatory strategies (Palermo et al., 2016). Alternatively, it may be that objective tests such as the CFMT+ do not adequately reflect real life abilities.

The present study corroborates these results. Three naive participants who scored over two *SD*s above the control mean for the CFMT+ and between 0.54 and 1.51 *SD*s on MFMT subjectively rated their FRA between -0.43 and 0.71 *SD*s away from the control average on the SFRS. Conversely, out of three volunteers scoring poorly on the CFMT and the MFMT, two rated their FRA on par with the group's mean. One, however, reported self-perceived deficits in their ability to recognise faces and showed considerably low performance on both face processing tasks, but not the object task. In sum, this suggests that the self-report method is not an accurate way of identifying those exceptionally skilled at unfamiliar face processing and is of limited ability to identify deficits in it.

It is of note that our sample consisted of young adults, similar to the populations studied by most other researchers in attempting to examine self-reported FRA (Bobak et al., 2016; Bowles et al., 2009; Palermo et al., 2016). It is probable that older adults have longer experience with instances of failures and successes of their own face recognition, and would thus provide a more accurate estimate of their ability. An implicit measure of decisional bias (i.e. whether a participant is likely to under-estimate or over-estimate their own ability) would be helpful in clarifying the moderate correlations between the self-report and objective tests of face processing ability.

Furthermore, given that face recognition matures late, at approximately 30 years of age (Susilo, Germine, & Duchaine, 2013), and that (Bowles et al. (2009) reported no association between the self-rated FRA and the CFMT in participants aged over 55 years, a fruitful avenue for future research may be to investigate the insight into FRA reported by middle-aged adults. It is possible that having gained relatively accurate understanding of their own abilities at a later age, elderly individuals fail to adjust for their failing abilities in recognising faces and rate them based on how they were in most of their adult life. Given that children and older adults also show strong own-age bias in FRA (Anastasi & Rhodes, 2005), it is perhaps unsurprising that the scores on standardised assessment of face recognition show decline with age.

Should researchers completely abandon the self-report route in light of these findings? We argue that this is not the case. The pioneering studies with SRs (Russell et al., 2012, 2009) identified their participants based on anecdotal evidence, which were later corroborated by behavioural tests. Similarly, participants in later reports were predominantly identified by either a close person's referral or self-reported extraordinary instances of face memory. We believe it is important to develop this strand of research, but that it is pertinent to keep certain factors in mind. Firstly, researchers may wish to include a "free recall" of memorable instances of face recognition miscarriages or successes in the screenings of general populations and those suspected to be SRs or have DP. Systematic collation of such qualitative data could inform a more accurate self-report instrument for assessing individual FRA. Secondly, a standardised self-report instrument administered across laboratories would be a helpful addition to studies of special populations, such as individuals with DP or SRs. Additionally, self-report measures would be a useful tool in diagnosis of face recognition impairments in clinical settings both, in identifying stand-alone prosopagnosia (Barton & Corrow, 2016), and detecting concurrent face recognition deficits in ASD or dementia. Initial evidence for different types of questions tapping into separate components of the face recognition system (i.e. the uncovering of two components in our PCA), if replicated, may be helpful in mapping personal experience of different deficits in prosopagnosia, or specific strengths of SRs.

In summary, the evidence presented here suggests that young naive participants have moderate insight into their own FRA, but those who previously received feedback on their extraordinary performance (the SRs) are reasonably well informed about their face processing skills and can estimate it with higher degree of precision. This corroborates most reports suggesting that people have only moderate insight into their FRA (Bobak et al., 2016; Bowles et al., 2009; Grueter et al., 2007; Haan, 1999; Palermo et al., 2016). Correlations improve when the studied population includes those already known to be poor (Shah et al., 2015), or good (this study) are included, but this may be because these individuals already know their objective performance. Some differences between the SFRS and the PI-20 may be attributable to inherent differences between these two measures. PI-20 was designed to capture the low end of the FRA, whereas SFRS aimed to measure the broad spectrum from DR to superior recognition. Critically, we show that those particularly apt at face processing tasks who have not received any feedback on their performance in behavioural tests, are unaware of their high ability and report as average, or below average on the SFRS. The data reported here have important implications for applied settings where extraordinary face processing ability is an important part of every-day work, and for which SRs would be ideal personnel candidates (i.e. recognising individuals from photo stills/mug shots and CCTV). Pertinently, behavioural testing is currently the most reliable means of ascertaining individuals' FRA and while interesting for cross-study comparison and qualitative analyses, self-report is not a reliable measure of high aptitude in the domain of unfamiliar face recognition in its current form.

References

Anastasi, J. S., & Rhodes, M. G. (2005). An own-age bias in face recognition for children and older adults. *Psychonomic Bulletin & Review*, *12*(6), 1043–1047. https://doi.org/10.3758/BF03206441

Association, P. (2016, April 2). Serial thief jailed after being snared by ‘super recogniser’ police unit. *The Guardian*. Retrieved from https://www.theguardian.com/uk-news/2016/apr/02/serial-thief-jailed-after-being-snared-by-super-recogniser-police-unit

Baños, J. H., LaGory, J., Sawrie, S., Faught, E., Knowlton, R., Prasad, A., … Martin, R. C. (2004). Self-report of cognitive abilities in temporal lobe epilepsy: cognitive, psychosocial, and emotional factors. *Epilepsy & Behavior*, *5*(4), 575–579. https://doi.org/10.1016/j.yebeh.2004.04.010

Bate, S., & Bennetts, R. J. (2014). The rehabilitation of face recognition impairments: a critical review and future directions. *Frontiers in Human Neuroscience, 8*. https://doi.org/10.3389/fnhum.2014.00491

Beaudoin, M., & Desrichard, O. (2011). Are memory self-efficacy and memory performance related? A meta-analysis. *Psychological Bulletin*, *137*(2), 211–241. https://doi.org/10.1037/a0022106

Bentin, S., DeGutis, J. M., D’Esposito, M., & Robertson, L. C. (2007). Too Many Trees to See the Forest: Performance, Event-related Potential, and Functional Magnetic Resonance Imaging Manifestations of Integrative Congenital Prosopagnosia. *Journal of Cognitive Neuroscience*, *19*(1), 132–146. https://doi.org/10.1162/jocn.2007.19.1.132

Bindemann, M., Attard, J., & Johnston, R. A. (2014). Perceived ability and actual recognition accuracy for unfamiliar and famous faces. *Cogent Psychology*, *1*(1), 986903. https://doi.org/10.1080/23311908.2014.986903

Bobak, A. K., Bennetts, R. J., Parris, B. A., Jansari, A., & Bate, S. (2016). An in-depth cognitive examination of individuals with superior face recognition skills. *Cortex*, *82*, 48–62. https://doi.org/10.1016/j.cortex.2016.05.003

Bobak, A. K., Dowsett, A. J., & Bate, S. (2016). Solving the Border Control Problem: Evidence of Enhanced Face Matching in Individuals with Extraordinary Face Recognition Skills. *PLOS ONE*, *11*(2), e0148148. https://doi.org/10.1371/journal.pone.0148148

Bobak, A. K., Hancock, P. J. B., & Bate, S. (2015). Super-recognisers in Action: Evidence from Face-matching and Face Memory Tasks. *Applied Cognitive Psychology*, n/a-n/a. https://doi.org/10.1002/acp.3170

Bobak, A. K., Pampoulov, P., & Bate, S. (2016). Detecting Superior Face Recognition Skills in a Large Sample of Young British Adults. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01378

Bobak, A. K., Parris, B. A., Gregory, N. J., Bennetts, R. J., & Bate, S. (2017). Eye-movement strategies in developmental prosopagnosia and 'super' face recognition. *The Quarterly Journal of Experimental Psychology*, *70*(2), 201–217. https://doi.org/10.1080/17470218.2016.1161059

Bowles, D. C., McKone, E., Dawel, A., Duchaine, B., Palermo, R., Schmalzl, L., … Yovel, G. (2009). Diagnosing prosopagnosia: Effects of ageing, sex, and participant–stimulus ethnic match on the Cambridge Face Memory Test and Cambridge Face Perception Test. *Cognitive Neuropsychology*, *26*(5), 423–455. https://doi.org/10.1080/02643290903343149

Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow Face Matching Test. *Behavior Research Methods*, *42*(1), 286–291. https://doi.org/10.3758/BRM.42.1.286

DeGutis, J., Cohan, S., Mercado, R. J., Wilmer, J., & Nakayama, K. (2012). Holistic processing of the mouth but not the eyes in developmental prosopagnosia. *Cognitive Neuropsychology*, *29*(5–6), 419–446. https://doi.org/10.1080/02643294.2012.754745

Dennett, H. W., McKone, E., Tavashmi, R., Hall, A., Pidcock, M., Edwards, M., & Duchaine, B. (2012). The Cambridge Car Memory Test: a task matched in format to the Cambridge Face Memory Test, with norms, reliability, sex differences, dissociations from face memory, and expertise effects. *Behavior Research Methods*, *44*(2), 587–605. https://doi.org/10.3758/s13428-011-0160-2

Dowsett, A. J., & Burton, A. M. (2015). Unfamiliar face matching: Pairs out-perform individuals and provide a route to training. *British Journal of Psychology*, *106*(3), 433–445. https://doi.org/10.1111/bjop.12103

Duchaine, B., Germine, L., & Nakayama, K. (2007). Family resemblance: Ten family members with prosopagnosia and within-class object agnosia. *Cognitive Neuropsychology*, *24*(4), 419–430. https://doi.org/10.1080/02643290701380491

Duchaine, B., & Nakayama, K. (2006). The Cambridge Face Memory Test: results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia*, *44*(4), 576–585. https://doi.org/10.1016/j.neuropsychologia.2005.07.001

Gray, K. L. H., Bird, G., & Cook, R. (2017). Robust associations between the 20-item prosopagnosia index and the Cambridge Face Memory Test in the general population. *Royal Society Open Science*, *4*(3), 160923. https://doi.org/10.1098/rsos.160923

Grueter, M., Grueter, T., Bell, V., Horst, J., Laskowski, W., Sperling, K., … Kennerknecht, I. (2007). Hereditary Prosopagnosia: the First Case Series. *Cortex*, *43*(6), 734–749. https://doi.org/10.1016/S0010-9452(08)70502-1

Haan, E. H. F. D. (1999). A Familial Factor in the Development of Face Recognition Deficits. *Journal of Clinical and Experimental Neuropsychology*, *21*(3), 312–315. https://doi.org/10.1076/jcen.21.3.312.917

Jones, R. D., & Tranel, D. (2001). Severe Developmental Prosopagnosia in a Child With Superior Intellect. *Journal of Clinical and Experimental Neuropsychology*, *23*(3), 265–273. https://doi.org/10.1076/jcen.23.3.265.1183

Livingston, L. A., & Shah, P. (2017). People with and without prosopagnosia have insight into their face recognition ability. *The Quarterly Journal of Experimental Psychology*, *0*(0), 1–3. https://doi.org/10.1080/17470218.2017.1310911

Mason, L. (2003). High School Students' Beliefs About Maths, Mathematical Problem Solving, and Their Achievement in Maths: A cross-sectional study. *Educational Psychology*, *23*(1), 73–85. https://doi.org/10.1080/01443410303216

Palermo, R., Rossion, B., Rhodes, G., Laguesse, R., Tez, T., Hall, B., … McKone, E. (2016). Do people have insight into their face recognition abilities? *The Quarterly Journal of Experimental Psychology*, *0*(0), 1–16. https://doi.org/10.1080/17470218.2016.1161058

Robertson, D. J., Noyes, E., Dowsett, A. J., Jenkins, R., & Burton, A. M. (2016). Face Recognition by Metropolitan Police Super-Recognisers. *PLOS ONE*, *11*(2), e0150036. https://doi.org/10.1371/journal.pone.0150036

Russell, R., Chatterjee, G., & Nakayama, K. (2012). Developmental prosopagnosia and super-recognition: no special role for surface reflectance processing. *Neuropsychologia*, *50*(2). https://doi.org/10.1016/j.neuropsychologia.2011.12.004

Russell, R., Duchaine, B., & 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

Self-Report Measures of Intelligence: Are They Useful as Proxy IQ Tests? (1998). *Journal of Personality*, *66*(4), 525–554. https://doi.org/10.1111/1467-6494.00023

Shah, P., Sowden, S., Gaule, A., Catmur, C., & Bird, G. (2015). The 20 item prosopagnosia index (PI20): relationship with the Glasgow face-matching test. *Royal Society Open Science*, *2*(11). https://doi.org/10.1098/rsos.150305

Susilo, T., Germine, L., & Duchaine, B. (2013). Face recognition ability matures late: Evidence from individual differences in young adults. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(5), 1212–1217. https://doi.org/10.1037/a0033469

The super-recognisers of Scotland Yard. (n.d.). Retrieved 17 January 2017, from http://www.newstatesman.com/politics/uk/2016/08/super-recognisers-scotland-yard

Tree, J. J. (2011). Mental imagery in congenital prosopagnosia: A reply to Grüter et al. *Cortex*, *47*(4), 514–518. https://doi.org/10.1016/j.cortex.2010.11.005

Tree, J. J., & Wilkie, J. (2010). Face and object imagery in congenital prosopagnosia: A case series. *Cortex*, *46*(9), 1189–1198. https://doi.org/10.1016/j.cortex.2010.03.005

Turano, M. T., & Viggiano, M. P. (2016). The relationship between face recognition ability and socioemotional functioning throughout adulthood. *Aging, Neuropsychology, and Cognition*, *0*(0), 1–18. https://doi.org/10.1080/13825585.2016.1244247

Ventura, P., Livingston, L. A., & Shah, P. (in press). Adults have moderate-to-good insight into their face recognition ability: Further validation of the 20-item prosopagnosia index (PI20) in a Portuguese sample. *Quarterly Journal of Experimental Psychology*.

Verhallen, R. J., Bosten, J. M., Goodbourn, P. T., Lawrance-Owen, A. J., Bargary, G., & Mollon, J. D. (n.d.). General and specific factors in the processing of faces. *Vision Research*. https://doi.org/10.1016/j.visres.2016.12.014

White, D., Kemp, R. I., Jenkins, R., Matheson, M., & Burton, A. M. (2014). Passport Officers' Errors in Face Matching. *PLoS ONE*, *9*(8), e103510. https://doi.org/10.1371/journal.pone.0103510

White, D., Norell, K., Phillips, P. J., & O'Toole, A. J. (2017). Human Factors in Forensic Face Identification. In *Handbook of Biometrics for Forensic Science* (pp. 195–218). Springer, Cham. Retrieved from https://link.springer.com/chapter/10.1007/978-3-319-50673-9_9

White, D., Phillips, P. J., Hahn, C. A., Hill, M., & O'Toole, A. J. (2015). Perceptual expertise in forensic facial image comparison. *Proc. R. Soc. B*, *282*(1814), 20151292. https://doi.org/10.1098/rspb.2015.1292

White, D., Rivolta, D., Burton, A. M., Al-Janabi, S., & Palermo, R. (2017). Face matching impairment in developmental prosopagnosia. *The Quarterly Journal of Experimental Psychology*, *70*(2), 287–297. https://doi.org/10.1080/17470218.2016.1173076

Zell, E., & Krizan, Z. (2014). Do People Have Insight Into Their Abilities? A Metasynthesis. *Perspectives on Psychological Science*, *9*(2), 111–125. https://doi.org/10.1177/1745691613518075

This research was supported by Engineering and Physical Sciences Research Council as a part of Face Matching for Automatic Identity Retrieval, Recognition, Verification and Management project. Grant number: EP/N007743/1.

Dataset for this publication can be found here http://hdl.handle.net/11667/106