Running Head: THE LIMITS OF SUPER RECOGNITION

The Limits of Super Recognition:

An Other-Ethnicity Effect in Individuals with Extraordinary Face Recognition Skills

[1]Sarah Bate, [2]Rachel Bennetts, [1]Nabil Hasshim, [1]Emma Portch, [1]Ebony Murray, [3]Edwin Burns & [4]Gavin Dudfield

[1]Department of Psychology, Bournemouth University, UK

[2]College of Health and Life Sciences, Division of Psychology, Brunel University, UK;

[3]Department of Psychology, University of Richmond, USA; [4]Dorset Police, UK.

Word count: 8018

Address for correspondence:

Dr Sarah Bate

Department of Psychology

Faculty of Science and Technology

Bournemouth University

Poole

BH12 5BB

United Kingdom

Telephone: +44 (0) 1202 961918

Fax: +44 (0) 1202 965314

Email: sbate@bournemouth.ac.uk

Abstract

In the last decade there has been increasing interest in super-recognisers, who have an extraordinary ability to recognise faces. However, it has not yet been investigated whether these individuals are subject to the same biases in face recognition as typical perceivers. The most renowned constraint reported to date is the other-ethnicity effect, whereby people are better at recognizing faces from their own, compared to other, ethnicities. If super-recognisers also show this bias, it is possible that they are no better at other-ethnicity face recognition than typical native perceivers – a finding that would have important theoretical and practical implications. In the current study, eight Caucasian super-recognisers performed other-ethnicity tests of face memory and face matching. In Experiment 1, super-recognisers outperformed Caucasian but not Asian controls in their memory for Asian faces. In Experiment 2, a similar pattern emerged in some super-recognisers on a test of face matching. Finally, Experiment 3 examined the consistency of superior other-ethnicity face matching in relation to Caucasian controls, using Arab and Black faces. Only four super-recognisers consistently outperformed controls, and other-ethnicity matching performance was not related to Caucasian face-matching or own- or other-ethnicity face memory. These findings suggest that super-recognisers are subject to the same biases as typical perceivers, and are simply those at the top end of a common face recognition spectrum as opposed to a qualitatively different group of individuals.

Keywords: Face recognition, individual differences, super-recognisers, face matching.

Public Significance Statement

Existing work has identified people with extraordinary face recognition skills (super-recognisers), and some police forces use these people to assist with security tasks. However, super recognisers have only been tested with Caucasian faces, and it is unknown whether their superior skills extend to faces from other ethnicities. In the current study, some super-recognisers outperformed Caucasian typical perceivers across multiple other-ethnicity face recognition tests, but no individual outperformed typical native perceivers. Thus, there is no benefit in using a Caucasian super-recogniser over a native typical perceiver in policing settings.

The Limits of Super Recognition: An Other-Ethnicity Effect in Individuals with Extraordinary Face Recognition Skills

In recent years there has been growing interest in individual differences in the recognition of facial identity, and much work indicates that face recognition skills in the typical population reside on a continuum (e.g. Bate, Haslam, Parris & Kay, 2010; Richler, Cheung & Gauthier, 2011; Wang, Li, Fang, Tian & Liu, 2012; Wilmer, 2017). At one end of the spectrum are those with very poor face recognition skills (some of whom may have a condition known as developmental prosopagnosia: e.g. Bennetts, Murray, Boyce & Bate, 2017; Bowles et al., 2009; Dalrymple & Palermo, 2016; Duchaine & Nakayama, 2006), whereas those at the top end have an extraordinary ability to recognise faces and have been dubbed "super-recognisers" (Russell, Duchaine & Nakayama, 2009). The latter group of individuals have been used as a theoretical window into the typical face recognition system (e.g. Bennetts, Mole & Bate, 2017; Bobak, Bennetts, Parris, Jansari & Bate, 2016; Ramon et al., 2016; Russell, Chatterjee & Nakayama, 2012), and have spurred substantial real-world interest into the mobilisation of super-recognisers for forensic face recognition tasks. However, the limits of super recognition have not yet been established, and it is unknown whether the same biases and influences that act upon typical face recognition (for reviews see Herlitz & Lovén, 2013; Meissner & Brigham, 2001; Rhodes & Anastasi, 2012) also restrict performance at the top end.

One of the most robust and well-replicated findings in the human face recognition literature is the other-ethnicity effect (also known as the "own-race bias" or "other-race effect"), whereby recognition memory is reliably better for own- compared to other-ethnicity faces, and shows a substantial effect size in both laboratory and applied settings (Doty, 1998; Malpass & Kravitz, 1969; Meissner & Brigham, 2001). Pertinently, the other-ethnicity effect has also resulted in miscarriages of justice, with other-ethnicity eyewitness misidentifications historically accounting for a large proportion of wrongful convictions (Huff, Ratner, Sagarin & MacNamara, 1986). Modern policing frequently involves recognition of faces from a variety of ethnicities, and with the current movement towards the deployment of super-recognisers in policing settings, the limitations of super recognition urgently need to be understood. While it may be the case that super-recognisers' excellent face recognition performance extends to the recognition of other-ethnicity faces, perhaps even making super-recognisers immune to the other-ethnicity effect and allowing them to outperform native perceivers, an alternate possibility is that super-recognisers are still subject to the other-ethnicity effect. In the most severe scenario, the bias may reduce their other-ethnicity face recognition performance to that of their non-super-recogniser own-ethnicity counterparts.

Our existing theoretical understanding of both super recognition and the other-ethnicity effect prohibits a clear hypothesis from being posited on this issue. There has been much debate in the psychological literature about the underpinnings of extreme face recognition performance, with some authors suggesting that individuals with developmental prosopagnosia are simply those at the bottom end of a common face recognition spectrum, as opposed to being "qualitatively different" from the typical population (Barton & Corrow, 2016; Bate & Tree, 2017). The same argument can be directed to those with excellent face recognition skills, although only one study to date has addressed this question. Bobak, Parris, Bennetts, Gregory and Bate (2017) recorded the scanpaths elicited to facial stimuli in super-recognisers, individuals with developmental prosopagnosia, and control participants. The prosopagnosia group avoided the eye region and tended to look more at the mouth – a pattern that was distinctly different from the eye movement strategies that were observed in controls. Conversely, super-recognisers spent more time looking at the nose, and the same strategy was found to positively correlate with face recognition ability in the control group. This finding suggests that super-recognisers may simply be those at the top end of a typical face recognition spectrum – a conclusion that receives additional support from reports of heightened configural or holistic processing in these individuals (Bobak, Bennetts et al., 2016; Russell et al., 2009). Given that some studies suggest that holistic processing positively correlates with face recognition skills in the typical population (Richler et al., 2011; Wang et al., 2012; but for critiques of this claim see Richler, Floyd & Gauthier, 2015), existing evidence mostly supports the idea that super-recognisers are quantitatively but not qualitatively different to typical perceivers.

The latter finding is particularly pertinent given that variability in holistic processing is heavily implicated in one explanation of the other-ethnicity effect (but also see socio-cognitive and contact hypotheses: Hugenberg, Young, Bernstein, & Sacco, 2010; Levin & Banaji, 2006). Specifically, many studies suggest that typical perceivers show reduced holistic processing for other-ethnicity faces, and instead rely more heavily on a laboured featural or piecemeal processing strategy that is suboptimal for identity recognition (Hancock & Rhodes, 2008; Michel, Rossion, Han, Chung, & Caldara, 2006; Rhodes, Tan, Brake & Taylor, 1989). If this hypothesis is correct, and it is also true that super-recognisers are those at the top-end of a common face recognition spectrum that is to some extent underpinned by holistic processing skills, we might expect super-recognisers to be subject to the same other-ethnicity effect as typical participants. The size of this effect would have important theoretical implications: if super-recognisers' performance decreases for other-ethnicity faces at the same rate as controls, this would imply only a quantitative difference in the processing of own- compared to other-ethnicity faces. On the other hand, if the other-ethnicity effect is exaggerated in super-recognisers (i.e. typical recognition levels for other-ethnicity faces, but superior recognition for own-ethnicity faces), this would indicate that other-ethnicity faces are being processed in a qualitatively different manner, which is unaffected by the processes underpinning super recognition.

However, this issue is further complicated by indications that, similarly to developmental prosopagnosia (Bate & Bennetts, 2015; Minnebusch, Suchan, Ramon & Daum, 2007; Susilo & Duchaine, 2013), super recognition is heterogeneous in its cognitive presentation, with evidence of dissociations between face memory and face matching performance (Bate et al., 2018; Bobak, Bennetts, et al., 2016; Bobak, Dowsett & Bate, 2016; Bobak, Hancock & Bate, 2016). It is not inconceivable that the same patterns may extend to other-ethnicity faces, or that the other-ethnicity effect may differentially affect face memory and face matching in these individuals. Indeed, although the other-ethnicity effect has previously been observed in typical perceivers performing perceptual matching (e.g. Lindsay, Jack & Christian, 1991; Sangrigoli & de Schonen, 2004; Walker & Hewstone, 2006) or classification (e.g. Caldara, Rossion, Bovet & Hauert, 2004; O'Toole, Peterson & Deffenbacher, 1996) tasks, it is not always detected and effects do not tend to be as large as in memory paradigms (see Papesh & Goldinger, 2009). Further, it is plausible that performance on the same paradigm might vary across faces from different ethnicities: the perceptual mechanisms that underpin super recognition may be only suitably calibrated for the recognition of certain types of face, and differences in an individual's level of contact with different other-ethnicity faces might idiosyncratically influence performance.

The current investigation sought to address these issues in eight Caucasian super-recognisers performing a series of other-ethnicity face recognition tasks. In an initial experiment, we examined whether super-recognisers displayed an other-ethnicity effect in their memory for Caucasian versus Asian faces. Second, we examined the consistency of these effects when super-recognisers completed face matching tasks, again using Caucasian and Asian faces. In both experiments, we used same- and other-ethnicity control groups as we wished to examine (a) whether super-recognisers outperformed Caucasian controls in the other-ethnicity condition, and (b) if so, whether they also outperformed non-super-recogniser native perceivers. Finally, we examined the consistency of face matching performance across faces from two other ethnicities.

Experiment 1

Our first experiment sought to examine whether the other-ethnicity effect would affect the performance of super-recognisers in face memory. Because the other-ethnicity effect is most robust in memory paradigms (e.g. Papesh & Goldinger, 2009), this was deemed to be the most reliable means to initially investigate whether super recognition is also subject to the bias. Further, we used well-validated and reliable measures of memory that are known to be appropriately calibrated to detect individual differences in performance.

Method

Participants. Eight super-recognisers (seven male; age range = 22-49 years, M = 37.3, SD = 8.0) took part in this investigation. All were Caucasian and had lived in the UK for their entire lives. These individuals were police officers who had participated in a previous super-recogniser screening programme carried out by our laboratory, where they surpassed the performance of control participants by at least 1.96 SDs on two tests of face memory: the long form of the Cambridge Face Memory Test (CFMT+; Russell et al., 2009) and the Models Memory Test (MMT: Bate et al., 2018).

The CFMT+ is a dominant test used by most laboratories to detect super recognition (e.g. Bobak, Pampoulov & Bate, 2016; Ramon et al., 2016; Russell et al., 2009), whereas the MMT is a new test of face memory, developed within our laboratory, that is calibrated to differentiate between top-end performers. Detailed descriptions of the two tests can be found in the papers of Russell et al. (2009) and Bate et al. (2018). In brief, the CFMT+ is an extended version of the standard CFMT (Duchaine & Nakayama, 2006) – a test that is widely used to detect prosopagnosia (see Dalrymple & Palermo, 2016). Participants are required to encode six male faces and subsequently select target faces from 72 triads, under novel viewpoint or lighting conditions. The CFMT+ includes a further 30 more difficult trials in order to discriminate between high performers. The MMT employs a similar design to the CFMT+, but embraces the more real-world natural variability that occurs between different presentations of the same face (Young & Burton, 2017), and includes target-absent as well as target-present trials. Demographic background and inclusion test scores are presented for each super-recogniser participant in Table 1.

< Insert Table 1 >

Two groups of control participants took part in this study. One group contained 35 Caucasians (18 male; age range = 18-50 years, M = 33.2, SD = 10.0) who had lived in the UK for the majority of their lives; the second group contained 28 Asian control participants (16 female, age range = 20-50 years, M = 27.4, SD = 8.1), who had been raised in an Asian dominant country (primarily China or Singapore, with some Malay and Filipino participants), and had spent the majority of their lives there (M % of life spent within native country = 78.6, SD = 12.9). Sample size was based on previous reports (e.g. Malpass & Kravitz, 1969; Lindsay et al., 1991; McKone et al., 2012). We confirmed the appropriateness of the sample size with a power calculation using the average effect size of the other-ethnicity effect (d = 0.80) for Caucasian and Asian participants performing the original and Chinese versions of the CFMT in McKone et al.'s (2012) study. To obtain a power of 0.8 against this effect size, the required sample is 21 participants in each condition (the power calculation was performed within G*Power: Faul, Erdfelder, Lang & Buchner, 2007). We therefore stopped recruitment once this target had been met, but continued to test any additional participants who had already contacted us prior to this point. Ethical approval for all experiments was granted by the institutional Ethics Committee.

Materials. Two tests of face memory were used in this experiment. First, we assessed memory for Caucasian faces using the CFMT+ (see description above). Second, we used the Chinese version of the CFMT (McKone et al., 2012): a test that uses Chinese rather than Caucasian faces, but is otherwise identical in design to the original version of the CFMT (Duchaine & Nakayama, 2006; see above). This means that the CFMT+ has 30 more trials than the CFMT-Chinese. Although it is possible to calculate the 72-item score on the CFMT+ (thereby giving a score that matches the number of items in the CFMT-Chinese), we retained the full 102-item score in order to avoid ceiling effects in our super-recogniser participants. Existing norming data for Caucasian participants on the CFMT-Chinese (McKone et al., 2012) suggested that we were unlikely to encounter a ceiling effect on this test, despite its shorter length. However, it is possible that the unequal trial numbers may result in an under-estimation of the other-ethnicity effect. Previous work indicates that both tests have excellent psychometric properties. Cronbach's α for the standard (Wilmer et al., 2010) and Chinese (McKone et al., 2012) versions of the CFMT have been reported in the range of .89-.90.

Procedure. There were some differences in the manner that data were gathered between super-recogniser and control participants. Because super-recognisers had completed the CFMT+ (and the Caucasian matching test reported in Experiment 2) online within the initial screening procedure (the CFMT+ was always completed first), we used these scores in the current investigation to avoid practice effects. The super-recognisers then completed the CFMT-Chinese in a face-to-face testing session that occurred approximately six months later. The CFMT-Chinese was always completed first in this session, followed by the three matching tests that are reported in Experiments 2 and 3.

Due to their varied geographical locations, control participants completed both memory tests online within the same testing session. This session also included the four matching tests that are reported in Experiments 2 and 3. The CFMT+ and CFMT-Chinese were always performed first, in a counterbalanced order. Participants were instructed to take breaks as required between each test. Given previous work has observed no differences between online versus laboratory performance on the standard version of the CFMT (Germine et al., 2012), we were not concerned about differences in testing location between participants. Further, the testing platform optimized stimulus presentation to ensure consistency in the size that images were presented. All participants completed the tests on a laptop or desktop computer.

Statistical analyses: Initial analyses aimed to confirm that an other-ethnicity effect could be observed across the two tests in the two control groups. Given the smaller sample size of our super-recogniser group, and our expectation that there may be some heterogeneity in performance, we carried out case-by-case comparisons for each individual using Crawford and Garthwaite's (2002) modified *t*-tests. Based on the size of the control groups, previously reported reliability for the standard version of the CFMT (Bowles et al., 2009) and the CFMT-Chinese (McKone, Wan, Robbins, Crookes, & Liu, 2017), and the calculations presented by Crawford and Garthwaite (2006), we estimate that our statistical power to detect effects greater than 2 SDs from the mean was between 0.54 and 0.60.

Results

Controls. A 2 (test: CFMT+, CFMT-Chinese) x 2 (participant group: Caucasian, Asian) mixed factorial ANCOVA (controlling for age) on the performance of the two control groups resulted in a significant interaction, $F(1,60) = 42.31$, $p = .001$, $\eta\rho^2 = .41$. Follow-up analyses confirmed that Caucasian controls outperformed Asian controls on the CFMT+, whereas the reverse pattern emerged on the CFMT-Chinese, *F*(1,60) = 4.23, *p* = .04, $\eta\rho^2$ = .07 and *F*(1,60) = 20.99, *p* = .001, $\eta\rho^2$ = .26, respectively (see Figure 1A). The main effects of test and participant group were not significant: *F*(1,60) = 0.68, *p* = .41, $\eta\rho^2$ = .01, and *F*(1,60) = 2.05, *p* = .16, $\eta\rho^2$ = .03, respectively. There was no influence of participant age (*p*s > .74).

< *Insert Figure 1* >

Super-recognisers. All super-recognisers had previously demonstrated facilitated memory for Caucasian faces, given the CFMT+ was used as an index for initial super-recogniser inclusion criteria (for formal analyses and effect sizes, see Table 2). The eight super-recogniser participants performed in the range of 81.94-93.06% correct on the CFMT-Chinese (see Figure 1B). A related-samples *t*-test indicated that the super-recognisers as a group performed significantly poorer on the CFMT-Chinese (M = 85.76%, SD = 3.54) than the CFMT+ (91.79%, SD = 3.70), *t*(7) = 3.31, *p* = .01, *d* = 1.67. Percentage correct was higher on the CFMT+ compared to the CFMT-Chinese in all but one of the super-recognisers (see Figure 1B). There was no difference in the size of the other-ethnicity effect between the control (M = 4.18%, SD = 13.07) and super-recogniser (M = 6.03%, SD = 5.15) groups, *t*(41) = 0.39, *p* = .70, *d* = 0.19.

Individual performance on the CFMT-Chinese was examined in relation to the Caucasian and Asian control groups, using Crawford and Garthwaite's (2002) modified *t*-tests for single-case comparisons (see Table 2). Three super-recognisers outperformed the Caucasian control group, and the remaining five all performed above 1.5 SDs from the control mean. All but one of the super-recognisers performed less than one SD from the Asian control mean. The remaining individual scored 1.2 SDs above the mean – a score that was equalled or bettered by four of the Asian controls (see Figure 1B).

< *Insert Table 2* >

Discussion

Experiment 1 investigated the other-ethnicity effect for the memory of Caucasian and Asian faces in eight super-recognisers, compared to typical Caucasian and Asian control groups. Findings suggest that super-recognisers show some consistency in face memory for own- and other-ethnicity faces, in line with previous work that has found a correlation between performance in the CFMT-Chinese and CFMT-Australian in typical observers (Wan et al., 2017; although note that this study did not find a significant correlation in Caucasian control participants).

All of the super-recognisers had facilitated Caucasian face memory (this was the basic inclusion criteria for the current study), and their performance consistently clustered at the top-end of Caucasian control performance on the CFMT-Chinese. Although three individuals significantly surpassed Caucasian controls, this categorical approach to detecting superior performance is of course somewhat arbitrary. More importantly, the distribution of individual super-recogniser performance was at the top end of the Caucasian control distribution. However, it was more moderate when compared to Asian controls, overlapping with approximately the top half of the sample. This is despite the fact that some of our super-recognisers had achieved near ceiling own-ethnicity memory scores at screening: for example, one individual had scored 100% correct on the CFMT+, yet scored 87.5% on the CFMT-Chinese. Because the CFMT-Chinese does not contain the 30 more difficult items that are present in the CFMT+, it should (other-ethnicity issues aside) be a relatively easier test, meaning the other-ethnicity effect may actually have been underestimated. Thus, it appears that superior own-ethnicity face memory skills likely extend to above-average other-ethnicity memory skills, but do not surpass those of native perceivers.

Importantly, there was no difference in the size of the other-ethnicity effect between super-recogniser and control participants. While this indicates consistency in the magnitude of the effect across the face recognition spectrum, it should be noted that the other-ethnicity effect may have been further underestimated in the super-recognisers as they always completed the CFMT-Chinese after the CFMT+. Further, control participants completed the two tests within the same testing session, resulting in higher memory load and potential interference effects on the second test. While the order of the two tests was counterbalanced in these participants, mean performance on both tests may be slightly under-estimated. Nevertheless, findings from this study indicate a quantitatively-similar other-ethnicity effect across typical and superior perceivers.

Experiment 2

Having found evidence that super recognition is susceptible to the other-ethnicity effect in a memory paradigm, we proceeded to examine the consistency of this effect in a matching task, using faces from the same ethnicities as in Experiment 1. Given the heterogeneity that has previously been observed in the performance of super-recognisers across own-ethnicity memory and matching tasks (e.g. Bate et al., 2018; Bobak, Bennetts, et al., 2016), we expected that different patterns of performance might be observed in the eight individuals reported here.

Method

Participants. The same super-recogniser, Caucasian control and Asian control participants took part in this experiment.

Materials. Two face matching tests were used: one containing Caucasian faces and the other Asian faces. The Caucasian face matching test was previously developed in our laboratory and is suitably calibrated to detect superior performance (the Pairs Matching Test, PMT: Bate et al., 2018). This task assessed participants' ability to match simultaneously-presented pairs of Caucasian faces over 48 trials (24 male): half matched in identity and the remainder displayed two different individuals. All images were downloaded from Google image searches, and were cropped to display the entire face from the neck upwards. Mismatched faces were paired according to their perceived similarity to each other, and all images were adjusted to 10 cm in width and 14 cm in height. Trials were displayed in a random order until responses were made, and no time limit was imposed. Participants made key presses to elicit "same" or "different" responses. For the purposes of the current investigation, we used the exact same parameters to develop a new version of this test using Asian faces. As for the Caucasian test, the Asian test was also designed to be suitably challenging for top-end performers.

Procedure. As described in Experiment 1, super-recogniser participants had previously completed the Caucasian matching test online within the initial screening procedure (this always occurred immediately after the CFMT+). We used these scores in the current investigation to avoid practice effects. The super-recognisers completed the Asian matching test in the same face-to-face session as described for Experiment 1 (i.e. for the CFMT-Chinese), which occurred approximately six months after they completed the CFMT+ and Caucasian matching test. All participants completed the CFMT-Chinese first, followed by three matching tests (the Asian matching test reported in this experiment, and the two tests described in Experiment 3). The three matching tests were completed in a counterbalanced order.

Control participants completed the two matching tests online within the same testing session as the memory tests (reported in Experiment 1) and the two other matching tests (reported in Experiment 3). All participants completed the memory tests first, in a counterbalanced order. They then completed the four matching tests, also in a counterbalanced order. Participants were instructed to take breaks as required between each test.

Statistical analyses. Scores for all participants were calculated in terms of hits (the number of correct “same” responses) and correct rejections (the number of correct “different” responses), and summed for overall accuracy. This data was also used to calculate signal detection theory (SDT) measures of recognition. Due to the clearly non-normal distribution (negative skew) of the data, the analysis for this task used non-parametric measures of sensitivity (*A*) and bias (*b*) (Zhang & Mueller, 2005). The measure *A* ranges from 0 (chance performance) to 1 (perfect performance); the measure *b* is used as an indicator of response bias (i.e., whether the participant has a tendency to say that the target is present or absent; MacMillan & Creelman, 2005). A score of 0 indicates a neutral response criterion, whereas a positive score indicates conservative responding (a tendency to indicate that a target was not present) and a negative score indicates more liberal responding (a tendency to indicate that a target was present).

Similarly to Experiment 1, we carried out single case analyses for each of the super-recognisers, using Crawford and Garthwaite’s (2002) modified *t*-tests. Based on the size of the control groups, reliability of the tests (see below), and the calculations presented by Crawford and Garthwaite (2006), we estimated that our statistical power to detect effects greater than 2 SDs from the mean was between 0.50 and 0.55.

Results

Controls. The reliability of the two matching tests was initially calculated, using the performance of all control participants. Because previous work using the Caucasian matching test found low correlations between target-present and target-absent performance (Bate et al., 2018), we calculated Cronbach’s α separately for each type of trial. For the Caucasian test, α

was .74 for target-present trials and .79 for target-absent trials; values of .76 and .67 were returned, respectively, for the Asian test.

Overall patterns of performance in control participants were initially examined via a 2 (test: Caucasian, Asian) x 2 (trial: hits, correct rejections) x 2 (participant group: Caucasian, Asian) mixed factorial ANCOVA, controlling for age. The interaction between test and group was significant, $F(1,60) = 22.59$, $p = .001$, $\eta\rho^2 = .27$; follow-up analyses confirmed that Caucasian participants outperformed Asian participants on the Caucasian test, whereas the reverse pattern was observed on the Asian test, $F(1,60) = 8.19$, $p = .01$, $\eta\rho^2 = .12$ and $F(1,60) = 7.54$, $p = .01$, $\eta\rho^2 = .11$, respectively (see Figure 2A). A main effect of trial indicated a higher proportion of correct rejections (M = 73.48%, SE = 1.72) than hits (M = 66.38%, SE = 1.71) across the two tests, $F(1,60) = 5.33$, $p = .02$, $\eta\rho^2 = .08$. While participant age interacted with test, $F(1,60) = 5.27$, $p = .03$, $\eta\rho^2 = .08$ (older participants tended to perform slightly better on the Asian test, whereas there was virtually no benefit of age on the Caucasian test), there were no other significant effects of age (all *p*s > .10). The three-way interaction was also non-significant, $F(1,60) = 1.08$, $p = .30$, $\eta\rho^2 = .02$, as were the two-way interactions between trial and group and trial and test, $F(1,60) = 1.78$, $p = .19$, $\eta\rho^2 = .03$ and $F(1,60) = 1.27$, $p = .26$, $\eta\rho^2 = .02$, respectively. The main effects of test and participant group did not reach significance: $F(1,60) = 0.07$, $p = .80$, $\eta\rho^2 = .01$, and, $F(1,60) = .01$, $p = .94$, $\eta\rho^2 = .01$, respectively.

< Insert Figure 2 >

To explore whether this pattern of findings emerged without any influence of response bias, a 2 (test: Caucasian, Asian) x 2 (participant group: Caucasian, Asian) mixed factorial ANCOVA (controlling for age) was performed on *A*. A significant interaction between test and participant group was observed, $F(1,60) = 19.90$, $p = .001$, $\eta\rho^2 = .25$. Follow-up analyses confirmed the pattern from the previous ANCOVA: Caucasian controls outperformed Asian controls on the Caucasian test, whereas the reverse pattern emerged on the Asian test, $F(1,60) = 7.34$, $p = .01$, $\eta\rho^2 = .11$ and $F(1,60) = 5.98$, $p = .02$, $\eta\rho^2 = .09$, respectively (see Figure 2B). Again, age interacted with test, $F(1,60) = 5.71$, $p = .02$, $\eta\rho^2 = .09$, and there were no main effects of test or participant group: $F(1,60) = 0.11$, $p = .74$, $\eta\rho^2 = .01$, and $F(1,60) = 0.10$, $p = .75$, $\eta\rho^2 = .01$.

The same ANCOVA was performed on *b* (bias). The interaction between test and participant group was significant, $F(1,60) = 4.41$, $p = .04$, $\eta\rho^2 = .07$. Follow-up analyses indicated that Caucasian participants were more conservative (M = 1.50, SE = 0.11) than Asian participants (M = 1.16, SE = 0.13) in their responses on the Asian test, $F(1,60) = 4.0$, $p = .05$, $\eta\rho^2 = .06$., but *b* did not differ between Caucasian (M = 1.20, SE = .10) and Asian (M = 1.20, SE = .11) participants for the Caucasian matching test, $F(1,60) = 0.01$, $p = .99$, $\eta\rho^2 = .01$. The main effects of test and group were non-significant: $F(1,60) = 2.17$, $p = .15$, $\eta\rho^2 = .04$, and $F(1,60) = 1.48$, $p = .23$, $\eta\rho^2 = .02$, respectively; and there was no interaction between age and test for this measure, $F(1,60) = 1.06$, $p = .31$, $\eta\rho^2 = .02$. Finally, we examined the relationship between memory and matching performance for own-ethnicity and other-ethnicity faces, using *A*. Significant correlations emerged for both control groups for own- and other-ethnicity faces (see Table 3).

< Insert Table 3 >

Super-recognisers. Because analyses on the control data revealed no differences in the proportion of hits and correct rejections according to test or control group, we used overall accuracy scores to compare the performance of super-recognisers to controls[1]. As there was a difference in calibration between the two tests, these scores were converted to standardized scores. While a paired-samples *t*-test indicated a significant difference between performance on the Caucasian (M = 2.31, SD = 0.45) and Asian (M = 1.53, SD = 0.80) tests,

[1] As super-recognisers also displayed the same patterns of hits versus correct rejections as noted in controls (i.e. a greater number of correct rejections than hits), and this was consistent across the two tests, we proceeded to only look at overall accuracy scores.

t(7) = 2.31, *p* = .05, *d* = 1.20, modified *t*-tests indicated some heterogeneity in performance. Specifically, four super-recognisers exceeded the performance of Caucasian controls on the Caucasian matching test (see Table 4), and another three achieved scores that were at least 1.3 SDs above the control mean (see Figure 2C). One individual achieved a score that was very similar to the control mean. Five super-recognisers outperformed Caucasian controls on the Asian matching test, with the remaining three performing at least 1.2 SDs above the Caucasian control mean. No individual significantly outperformed the Asian control group on this test, although the highest performing super-recogniser equated the score of the top-performing Asian control (see Figure 2C).

Standardized scores were also used to compare the magnitude of the other-ethnicity effect in super-recogniser versus Caucasian controls. As observed for the memory tests, there was no difference in the size of the effect between control (M = 0.75, SD = 1.06) and super-recogniser (M = 0.79, SD = 0.97) groups, *t*(41) = 0.08, *p* = .94, *d* = 0.04. Finally, performance on this test was compared to that on the CFMT-Chinese (see Figure 2D). Only one super-recogniser outperformed Caucasian controls on both tasks, and one failed to significantly outperform controls on either task. The latter individual was the lowest performer on both tests, but there was little consistency in the rankings of the other super-recognisers across the two tests.

< *Insert Table 4* >

Discussion

Experiment 2 assessed Caucasian and Asian face-matching performance in the same eight super-recognisers and control groups as used in Experiment 1. However, a less clear pattern emerged for face matching compared to face memory. Previous work has examined the heterogeneity of super recognition, and reported super-recognisers (classified as such on the basis of their face memory skills) with and without facilitations in own-ethnicity versions of this task (e.g. Bate et al., 2018; Bobak, Bennetts, et al., 2016). The performance of the super-recognisers reported here sit well with this work, with four individuals significantly surpassing Caucasian controls on the own-ethnicity version of this task, three performing above 1.3 SDs of the mean, and one achieving a score that was very similar to the control mean. However, while this latter individual achieved a good score on the CFMT+, they were the bottom performer on our second screening test of face memory (the MMT), and achieved a moderate score on the Asian matching test. Thus, their overall profile suggests only a borderline case for super recognition inclusion.

A novel question addressed in the current paper is whether those individuals with superior matching skills for own-ethnicity faces also show superior performance when matching other-ethnicity faces. Unsurprisingly, the heterogeneity in matching performance prohibited group-based comparisons from reaching significance, but case-by-case comparison of individual performance in relation to own- and other-ethnicity control groups provides some insights. Five super-recognisers outperformed Caucasian controls on the Asian matching test, with three performing at least 1.2 SDs above the control mean. However, no individual surpassed the Asian control cut-off, with performance in the range of 0.49-1.64 SDs of the mean. Again, as for face memory, we found no evidence that Caucasian super-recognisers show superior performance on an other-ethnicity task compared to typical native perceivers. Given our Asian controls were recruited according to a fairly liberal contact criterion (most were currently living in the UK, albeit for a relatively short period of time), the effect would likely be larger when super-recognisers are compared to native perceivers who have lived in an Asian country for their entire life.

Similarly to Experiment 1, no difference in the size of the other-ethnicity effect was observed between super-recogniser and Caucasian control participants. However, as also noted for Experiment 1, the other-ethnicity effect may also have been underestimated in the super-recogniser group. Indeed, all super-recognisers had completed the Caucasian test as part of the initial screening process, whereas the Asian test was completed at a later date. Thus, their performance may be slightly inflated on the Asian test due to practice effects, potentially reducing the size of the other-ethnicity effect.

Experiment 3

Given the range of performance observed in the super-recognisers in Experiment 2, with five of the eight individuals outperforming Caucasian controls, we sought to examine the consistency of performance in two further matching tests using faces from different ethnicities (Arab and Black faces). In this Experiment we only compared performance to Caucasian controls, to establish whether super-recognisers can reliably outperform own-ethnicity typical perceivers when matching a range of other-ethnicity faces.

Method

Participants. The same super-recogniser and Caucasian control participants as described above took part in this experiment.

Materials and procedure. Two further versions of the matching task were prepared, one using Arab faces and the other using Black faces of various ethnic origins. The protocols of each test were otherwise identical to the matching tests used in Experiment 2. All participants took part in these two tests within the same testing session (please refer to the Procedure section of Experiment 2 for details of test order).

Statistical analyses. The proportion of hits and correct rejections were calculated for each participant, in addition to an overall accuracy score. The SDT measures A and b (bias) were calculated. Because we were interested in overall consistency in performance, data were also combined with that for the Caucasian and Asian matching tests from Experiment 2.

Results

Controls. Cronbach's α was calculated for the two new matching tests used in this Experiment. In the Arab test, values of .76 and .65 were calculated for target-present and target-absent trials, respectively; and .75 and .67 for the Black test.

A 4 (test: Caucasian, Asian, Arab, Black) x 2 (trial: hits, correct rejections) repeated measures ANCOVA (controlling for age) was initially performed on the control data. The interaction was not significant, $F(3,123) = 2.02$, $p = .11$, $\eta\rho^2 = .05$ (see Figure 3A); nor were there main effects of test or trial, $F(3,123) = 0.82$, $p = .48$, $\eta\rho^2 = .02$, and $F(1,41) = 1.11$, $p = .30$, $\eta\rho^2 = .03$. The influence of response bias was also evaluated by looking at performance on each test according to *A* and *b* (bias). One-way ANCOVAs controlling for age did not find a difference in performance across the four tests on either measure: $F(3,123) = 0.46$, $p = .71$, $\eta\rho^2 = .01$, and $F(3,123) = 1.54$, $p = .21$, $\eta\rho^2 = .04$, respectively. There was no influence of age in any analysis (all *p*s > .24). Descriptive statistics for each matching test are displayed in Table 5.

The correlation matrix for control performance on the four matching tests is presented in Table 6. While there were significant correlations between the three other-ethnicity tests, performance on the Caucasian test only correlated with the Asian and not the Arab or Black tests.

< Insert Figure 3 and Tables 5 & 6 >

Super-recognisers. Because analyses on the control data revealed no differences in the proportion of hits and correct rejections across the tests, for consistency we proceeded to analyse the data using overall percentage correct[2]. As observed in Experiment 2, there were differences in the relative difficulties between the four tests. We therefore calculated standardized scores in order to decipher whether an other-ethnicity effect emerged for the Arab and Black tests. Paired-samples *t*-tests did not detect significant differences at the group level for performance on the Caucasian (M = 2.31, SD = 0.45) compared to either the Arab (M = 2.21, SD = 0.63) or Black (M = 1.77, SD = 0.68) tests, $t(7) = 0.71$, $p = .50$, $d = 0.18$, and $t(7) = 2.26$, $p = .06$, $d = 0.94$, respectively. In addition, there was no difference between the magnitude of the effect between super-recognisers and controls for either the Arab (super-recognisers: M = 0.10, SD = 0.42; controls: M = 0.38, SD = 1.01) or Black (super-recognisers: M = 0.54, SD = 0.68; controls: M = 0.38, SD = 1.07) test: , $t(41) = 0.75$, $p = .46$, $d = 0.36$, and $t(41) = 0.41$, $p = .69$, $d = 0.18$, respectively.

In single-case analyses, four of the eight super-recognisers significantly outperformed controls on the Arab test (see Table 7), with the performance of two others approaching significance (*p*s = .06). One super-recogniser performed 1.72 SDs above the control mean, and the remaining individual achieved a score that was 1.2 SDs above the control mean. Three super-recognisers significantly outperformed controls on the Black matching test, with one missing the cut-off by one point. One performed at 0.76 SDs from the control mean, and three within 1.5 SDs of the control mean (see Figures 3B and 3C).

The consistency of the super-recognisers' performance across the four matching tests is displayed in Figures 3B and 3C, in standardized score units. Because there was evidence to suggest that the performance of controls was inter-related on the three other-ethnicity tasks, we also computed an overall other-ethnicity index score for face matching (by averaging the three standardized scores for these tests). Four of the eight super-recognisers exceeded the performance of controls by 1.96 SDs (see Figure 3B). While two of these had outperformed

[2] Again, the patterns of hits and correct rejections in the super-recognisers followed those observed in the controls. For brevity, we only report overall percentage correct.

controls on the Caucasian matching task, two had not. On the other hand, two of the four super-recognisers who did not exceed the cut-off on the index score for other-ethnicity matching did outperform controls on the Caucasian matching task (see Figure 3C). There was no consistency between the other-ethnicity index score for face matching and performance on the CFMT-Chinese or the CFMT+ (see Figure 4).

< Insert Figure 4 and Table 7 >

Discussion

Experiment 3 examined the consistency of super-recognisers' performance on two further other-ethnicity face matching tasks (using Arab and Black faces), with reference to the data collected in Experiment 2 (i.e. the Caucasian and Asian tests). Control data indicated a difference in difficulty levels between the tests, obscuring evidence of an other-ethnicity effect when data from all four tests were simultaneously entered into an ANCOVA. In the absence of native performers on the Arab and Black matching tests, we used standardized scores to control for the differences in test difficulty and examine the magnitude of the other-ethnicity effects for the super-recognisers. Given that control performance on the three other-ethnicity tasks was correlated, we deemed it appropriate to calculate an overall other-ethnicity matching index. This index provides a more straightforward measure of other-ethnicity matching than spotting consistency across multiple individual test scores. The lowest standardized scores on the index was 1.39, whereas four individuals outperformed controls by at least two SDs. While all index scores are in a respectable range, there is clearly some distance between the top and bottom performer.

However, the top-performing individuals were not consistently those who had achieved superior scores on the Caucasian face memory or face matching measures. This finding suggests that alternative influences, that are not necessarily specific to face recognition or matching ability, may underpin superior scores on other-ethnicity tasks. One possibility is that social contact or experience with the faces of particularly ethnicities assists the accuracy of top-performers. Unfortunately, we did not measure the levels of social contact experienced by each super-recogniser for the ethnicities used in this experiment. While all eight individuals had lived in the UK for their entire lives, it remains possible that the top-performers on the other-ethnicity face matching index are those who have had more experience with faces from the target ethnicities.

Finally, another point of difference between Experiments 2 and 3 was the effect of age. Specifically, an interaction between age and test ethnicity was observed in the ANCOVA reported in Experiment 2 but not in Experiment 3. This suggests a small benefit of increasing age was either only observable for the Asian test, or was too mild to emerge in the larger ANCOVA carried out in Experiment 3.

General Discussion

This investigation set out to examine whether super-recognisers are subject to the same other-ethnicity effect as typical perceivers in face memory and face matching tasks. In an initial experiment, an other-ethnicity effect was observed for the memory of Caucasian versus Chinese faces. While the super-recognisers all outperformed Caucasian controls on the CFMT-Chinese by at least 1.5 SDs, their performance was more moderate in comparison to Asian controls. Second, we investigated the same issue in a matching task. While five super-recognisers outperformed Caucasian controls on the Asian version of the test, no individual outperformed the Asian control group. In a third experiment, four of these individuals consistently displayed superior face matching skills for other-ethnicity faces. However, there was little association between super-recognisers' other-ethnicity face matching skills in relation to own-ethnicity matching or memory performance.

The latter finding is particularly striking, but its underpinnings are unclear. It is possible that the discrepancy results from idiosyncratic differences in levels of social contact with other-ethnicities. Alternatively, it may be that faces from different ethnicities tap different processing strategies. This hypothesis is supported by some of the correlational data from controls: performance on the Caucasian matching test only correlated with that on the Asian test, and not the Arab or Black tests (but see Kokje, Bindemann & Megreya, 2018, for a relationship between Caucasian and Arab face matching in typical Caucasian perceivers). A potential explanation for this effect comes from holistic processing accounts of the other-ethnicity effect (e.g. Hancock & Rhodes, 2008; Michel et al., 2006; Rhodes et al., 1989). If holistic processing is available for the matching of own-ethnicity faces but less available or absent for other-ethnicity faces, the initiation of a reduced or even different processing strategy for the latter might explain the lack of relationship across the two measures.

However, this does not explain why there was more consistency in own- and other-ethnicity performance for face memory compared to face matching. It may be that even own-ethnicity face matching – particularly unfamiliar face matching – relies heavily on featural processing mechanisms (Megreya & Burton, 2006), whereas the CFMT paradigm was designed to specifically tap holistic processing, via the use of visual noise in the third stage (Duchaine & Nakayama, 2006; McKone, Martini, & Nakayama, 2001). If holistic processing remains available (albeit to a lesser degree) in the CFMT-Chinese, this would explain the consistency of high performance for other-ethnicity face memory. However, if featural processing is more heavily relied upon for own- and other-ethnicity matching, this may be more easily disrupted and perhaps more influenced by variances in everyday contact with faces of different ethnicities.

Alternatively, it may be that face matching performance is generally less stable than face memory, as has been observed in previous work (e.g. Bindemann, Avetisyan & Rakow, 2012). This may simply result from the inherent differences between memory and matching paradigms: the CFMT paradigm allows repeated encoding of unfamiliar faces for recall, whereas the matching paradigm only displays each face on a single occasion. The richer experience with target faces in the memory paradigm may produce increased stability in performance as opposed to the single-exposure matching paradigms.

Whatever the explanation, it is clear from the current dataset that, while superior own-ethnicity memory performance may roughly lend itself to above-average other-ethnicity memory performance, neither own-ethnicity memory or matching performance provides a reliable indicator of other-ethnicity matching skills. It remains to be seen whether any super-recogniser is immune to the other-ethnicity effect, at least to the extent that can be detected by the calibration of face recognition tests, and whether other biases are also observed in these individuals (e.g. the own-age bias: Rhodes & Anastasi, 2012). Thus, future super-recogniser screening programmes should include other-ethnicity matching tests, and not rely on any own-ethnicity indicators to select the best performers for such tasks. Importantly, there appears to be little benefit in selecting a Caucasian super-recogniser above a typical native perceiver for an other-ethnicity face recognition task.

Finally, the findings reported here have important theoretical implications. In terms of super recognition, our data support existing hypotheses that super-recognisers are simply those who reside at the top end of a common face recognition spectrum (Barton & Corrow, 2016; Bate & Tree, 2017). Indeed, it seems that super-recognisers are subject to other-ethnicity effects in both face memory and face matching, with no individual outperforming native perceivers in either domain. Further, the size of the super-recognisers' other-ethnicity effect was similar to that of controls, for both face memory and the matching of Caucasian versus Asian faces. This argues against super recognition arising due to exaggerated effects of expertise (i.e. rapid acquisition of faces to which they have exposure) or abnormal processing strategies. Instead, these results suggest that super-recognition is likely a reflection of more general face-processing mechanisms such as holistic processing, or, potentially, differences in adaptive coding of faces.

Adaptive coding suggests that faces are coded in relation to one or many "norms" – average representations of all the faces a person encounters (with different norms for different social categories), and efficient recognition of faces requires norms that adapt to reflect our visual experience of faces (Rhodes, 2017). Notably, adaptive coding has been proposed as a way to account for effects of exposure and individual differences in face recognition performance: increased exposure to faces from a certain category (e.g. faces of an unfamiliar ethnicity) enhances the ability to discriminate between faces from that category (Rhodes, Watson, Jeffery, & Clifford, 2010); and differences in how well individuals adapt to faces has been associated with individual differences in face recognition for own, but not other-ethnicity faces (Rhodes, Jeffery, Taylor, Hayward, & Ewing, 2014). This raises the possibility that super-recognition arises as a result of enhanced adaptive coding, that is highly sensitive to the visual "diet" of faces that super-recognisers have been exposed to over the course of their lifetime (i.e. primarily own-ethnicity faces). To date, though, adaptive coding has not been examined in super-recognisers, so the precise mechanisms (or combinations of mechanisms) underlying super recognition remain to be determined.

Further, our data supports the idea that, in memory tasks, the other-ethnicity effect is likely due to a quantitative, rather than qualitative difference in the processing of own- and other-ethnicity faces, and that this quantitative difference is relatively consistent across individual differences in face recognition ability (although see Wan et al., 2017, for some exceptions). This is in line with the suggestion that other-ethnicity faces are processed in a less holistic manner than own-ethnicity faces (e.g., Hancock & Rhodes, 2008; Michel, Rossion, Han, Chung, & Caldara, 2006; Rhodes, Tan, Brake & Taylor, 1989). As discussed above, the matching tasks revealed a more mixed pattern of performance, with some super-recognisers showing evidence of a typical other-ethnicity effect, and others performing better on other-ethnicity than own-ethnicity matching. This pattern of performance might reflect the use of atypical or non-face-specific strategies during matching tasks (matching of unfamiliar faces of one's own ethnicity may elicit qualitatively different processing strategies than elicited by familiar faces: Megreya & Burton, 2006). If the other-ethnicity effect is related to variations in holistic processing across different faces, the idea that matching tasks generally elicit lower levels of holistic processing could explain why previous studies have found smaller and less consistent other-ethnicity effects in matching compared to memory tasks (see Papesh & Goldinger, 2009).

In sum, the findings reported here indicate that Caucasian super-recognisers are more useful in other-ethnicity face memory tasks than typical Caucasian perceivers, but there is less advantage in deploying these individuals when typical native perceivers are available. Only Caucasian super-recognisers have been reported in the literature to date. While no investigation to date has examined the phenomenon in individuals from other ethnicities, there is evidence of individual differences in typical non-Caucasian perceivers (McKone et al., 2017; Wan et al., 2017), in addition to case reports of non-Caucasians with developmental prosopagnosia (e.g., Zhao et al., 2016). The current work suggests that identification of super-recognisers from a variety of ethnicities would be useful to the police for other-ethnicity face recognition tasks, but, meanwhile, super-recogniser screening programmes should include other-ethnicity face matching tests to identify the most consistent top-performers at this task. As such, all matching tests used in this investigation are freely available and can be accessed on request.

Acknowledgements

SB is supported by a British Academy Mid-Career Fellowship (MD170004). The full data set accompanying this investigation is available for download (Bate et al., 2018). Materials can be accessed by contacting the corresponding author.

References

Barton, J.J.S., & Corrow, S.L. (2016). The problem of being bad with faces. *Neuropsychologia, 89*, 119–24.

Bate, S., Bennetts, R., Hasshim, N., Portch, E., Murray, E., Burns, E., & Dudfield, G. (2018). Super-recognisers and controls - own-race and other-ethnicity memory and matching dataset [Data set]. https://doi.org/10.17633/rd.brunel.6752825.v1.

Bate, S., Frowd, C., Bennetts, R., Hasshim, N., Murray, E., Bobak, A.K., Wills, H., & Richards, S. (2018). Applied sother-ethnicity effectening tests for the detection of superior face recognition. *Cognitive Research: Principals and Implications, 3,* 22.

Bate, S., Parris, B.A., Haslam, C., & Kay, J. (2010). Socio-emotional functioning and face recognition ability in the normal population. *Personality and Individual Differences, 48*, 239-242.

Bate, S., & Tree, J.J. (2017). The definition and diagnosis of developmental prosopagnosia. *Quarterly Journal of Experimental Psychology, 70*, 193-200.

Bennetts, R.J., Mole, J.A., & Bate, S. (2017). Super recognition in development: A case study of an adolescent with extraordinary face recognition skills. *Cognitive Neuropsychology, 34*, 357-376.

Bennetts, R.J., Murray, E., Boyce, T., & Bate, S. (2017). Prevalence of face recognition deficits in middle childhood. *Quarterly Journal of Experimental Psychology, 70,* 234-258.

Bindemann, M., Avetisyan, M., & Rakow, T. (2012). Who can recognize unfamiliar faces? Individual differences and observer consistency in person identification. *Journal of Experimental Psychology: Applied, 18*, 277-291.

Bobak, A.K., Dowsett, A., & Bate, S. (2016). Solving the border control problem: Evidence of enhanced face matching in individuals with extraordinary face recognition skills. *PLos One, 11*, e0148148.

Bobak, A.K., Hancock, P.J.B., & Bate, S. (2016). Super-recognizers in action: Evidence from face matching and face memory tasks. *Applied Cognitive Psychology, 30*, 81-91.

Bobak, A., Pampoulov, P., & Bate, S. (2016). Detecting superior face recognition skills in a large sample of young British adults. *Frontiers in Psychology, 7*, 1378.

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. *Quarterly Journal of Experimental Psychology, 70*, 201-217.

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.

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*, 423-455.

Caldara, R., Rossion, B., Bovet, P., & Hauert, C.-A. (2004). Event-related potentials and time course of the 'other-race' classification advantage. *NeuroReport*, *15,* 905-910.

Crawford, J.R., & Garthwaite, P.H. (2002). Investigation of the single case in neuropsychology: Confidence limits on the abnormality of test scores and test score differences. *Neuropsychologia, 40*, 1196-1208.

Crawford, J.R., & Garthwaite, P.H. (2006). Methods of testing for a deficit in single-case studies: Evaluation of statistical power by Monte Carlo simulation. *Cognitive Neuropsychology, 23*, 877-904.

Dalrymple, K.A., & Palermo, R. (2016). Guidelines for studying developmental prosopagnosia in adults and children. *WIREs Cognitive Science, 7*, 73-87.

Doty, N.D. (1998). The influence of nationality on the accuracy of face and voice recognition. *American Journal of Psychology, 111*, 191-214.

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 subjects. *Neuropsychologia, 44*, 576–585.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191.

Germine, L., Nakayama, K., Duchaine, B.C., Chabris, C.F., Chatterjee, G., & Wilmer, J. (2012). Is the Web as good as the lab? Comparable performance from Web and lab in cognitive/perceptual experiments. *Psychonomic Bulletin & Review, 19*, 847-857.

Hancock, K.J., & Rhodes, G. (2008). Contact, configural coding and the other-race effect in face recognition. *British Journal of Psychology, 99*, 45-56.

Herlitz, A., & Lovén, J. (2013). Sex differences and the own-gender bias in face recognition: A meta-analytic review. *Visual Cognition, 21*, 1306-1336.

Huff, C.R., Ratner, A., Sagarin, E., & MacNamara, D.E.J. (1986). Guilty until proved innocent: Wrongful conviction and public policy. *Crime & Delinquency, 32*, 518-544.

Hugenberg, K., Young, S.G., Bernstein, M.J., & Sacco, D.F. (2010). The Categorization-Individuation Model: An integrative account of the cross race recognition deficit. *Psychological Review, 117*, 1168-1187.

Kokje, E., Bindemann, M., & Megreya, A.M. (2018). Other-ethnicity correlations in the abilities to match unfamiliar faces. *Acta Psychologica, 185*, 13-21.

Levin, D.T., & Banaji, M.R. (2006). Distortions in the perceived lightness of faces: The role of race categories. *Journal of Experimental Psychology: General, 135*, 501-512.

Lindsay, D.S., Jack, P.C., & Christian, M.A. (1991). Other-race face perception. *Journal of Applied Psychology, 76*, 587-589.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection Theory: A User's Guide (2nd ed.).* Mahwah, N.J.: Lawrence Erlbaum Associates.

Malpass, R.S., & Kravitz, J. (1969). Recognition for faces of own and other race. *Journal of Personality and Social Psychology, 13*, 330-334.

McKone, E., Martini, P., & Nakayama, K. (2001). Categorical perception of face identity in noise isolates configural processing. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 573-599.

McKone, E., Stokes, S., Liu, J., Cohan, S., Fiorentini, C., Pidcock, M., Yovel, G., … Pelleg, M. (2012). A Robust Method of Measuring Other-Race and Other-Ethnicity Effects: The Cambridge Face Memory Test Format. *PLOS ONE, 7,* e47956.

McKone, E., Wan, L., Robbins, R., Crookes, K., & Liu, J. (2017). Diagnosing prosopagnosia in East Asian individuals: Norms for the Cambridge Face Memory Test–Chinese. *Cognitive Neuropsychology*, *34*, 253-268.

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

Meissner, C.A., & Brigham, J.C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law*, *7*, 3-35.

Michel, C., Rossion, B., Han, J., Chung, C.H., & Caldara, R. (2006). Holistic processing is finely tuned for faces of one's own race. *Psychological Science, 17*, 608-615.

O'Toole, A.J., Peterson, J., & Deffenbacher, K.A. (1996). An 'other-race effect' for categorizing faces by sex. *Perception, 25,* 669-676.

Papesh, M.H., & Goldinger, S.D. (2009). Deficits in other-race face recognition: No evidence for encoding-based effects. *Canadian Journal of Experimental Psychology, 63*, 253-262.

Ramon, M., Miellet, S., Dzieciol, A.M., Konrad, B.N., Dresler, M., & Caldara, R. (2016). Super-memorizers are not super-recognizers. *PLoS ONE, 11*, e0150972.

Rhodes, G. (2017). Adaptive coding and face recognition. *Current Directions in Psychological Science*, *26*(3), 218-224.

Rhodes, G., Jeffery, L., Taylor, L., Hayward, W. G., & Ewing, L. (2014). Individual differences in adaptive coding of face identity are linked to individual differences in face recognition ability. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(3), 897.

Rhodes, G., Tan, S., Brake, S., & Taylor, K. (1989). Expertise and configural coding in face recognition. *British Journal of Psychology, 80*, 313-331.

Rhodes, M.G., & Anastasi, J.S. (2012). The own-age bias in face recognition: A meta-analytic and theoretical review. *Psychological Bulletin, 138*, 146-174.

Rhodes, G., Watson, T. L., Jeffery, L., & Clifford, C. W. (2010). Perceptual adaptation helps us identify faces. *Vision Research*, *50*(10), 963-968.

Richler, J.J., Cheung, O.S., & Gauthier, I. (2011). Holistic processing predicts face recognition. *Psychological Science*, *22*, 464-471.

Richler, J.J., Floyd, J., & Gauthier, I. (2015). About-face on face recognition ability and holistic processing. *Journal of Vision*, *15*, 15.

Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic Bulletin & Review, 16*, 252–257.

Russell, R., Chatterjee, G., & Nakayama, K. (2012). Developmental prosopagnosia and super-recognition: no special role for surface reflectance processing. *Neuropsychologia*, *50*, 334-340.

Sangrigoli, S., & de Schonen, S. (2004). Effects of visual experience on face processing: A developmental study of inversion and non-native effects. *Developmental Science*, *7*, 74-87.

Walker, P.M., & Hewstone, M. (2006). A perceptual discrimination investigation of the own-race effect and intergroup experience. *Applied Cognitive Psychology*, *20*, 461-475.

Wan, L., Crookes, K., Dawel, A., Pidcock, M., Hall, A., & McKone, E. (2017). Face-blind for other-race faces: Individual differences in other-race recognition impairments. *Journal of Experimental Psychology: General*, *146*, 102-122.

Wang, R., Li, J., Fang, H., Tian, M., & Liu, J. (2012). Individual differences in holistic processing predict face recognition ability. *Psychological Science, 23*, 169-177.

Wilmer, J.B. (2017). Individual differences in face recognition: A decade of discovery. *Current Directions in Psychological Science*, *26*, 225-230.

Wilmer, J.B., Germine, L., Chabris, C.F., Chatterjee, G., Williams, M., Loken, E., … Duchaine, B. (2010). Human face recognition ability is specific and highly heritable. *Proceedings of the National Academy of Sciences, 107*, 5238-5241.

Young, A.W., & Burton, A.M. (2017). Recognizing faces. *Current Directions in Psychological Science*, *26*, 212-217.

Zhang, J., & Mueller, S. T. (2005). A note on ROC analysis and non-parametric estimate of sensitivity. *Psychometrika*, *70*(1), 203-212.

Zhao, Y., Li, J., Liu, X., Song, Y., Wang, R., Yang, Z., & Liu, J. (2016). Altered spontaneous neural activity in the occipital face area reflects behavioral deficits in developmental prosopagnosia. *Neuropsychologia*, *89*, 344-355.

Tables

Table 1: Control mean percentage accuracy (SD) for the CFMT+ and MMT, with demographic information and individual *z* scores for each super-recogniser. Chance is 33.33% on the CFMT+ and 25% on the MMT. Control norms (N = 40) are taken from Bate et al. (2018): cut-offs are set at 1.96 SDs from the control mean.

	Controls	SR1	SR2	SR3	SR4	SR5	SR6	SR7	SR8
Gender	20xM, 20xF	M	M	M	M	M	M	M	F
Age	33.6 (10.1)	38	37	22	44	49	36	40	32
CFMT+	67.77 (9.92)	2.36	2.26	2.06	2.46	2.46	3.25	2.46	2.06
MMT	53.25 (14.06)	2.06	2.53	2.61	2.69	3.01	2.38	2.77	2.69

Table 2: Results from modified t-tests for single-case comparisons (Crawford & Garthwaite, 2002) for the eight individual super-recognisers in relation to Caucasian and Asian controls performing the CFMT+ and CFMT-Chinese. Significant results are highlighted by an accompanying asterisk. Effect sizes (*z*-cc) are also reported for each calculation, alongside the percentage of the population achieving less extreme scores than each individual super-recogniser.

	Raw score (%)	Caucasian controls				Asian controls			
		t	*p*	*z*-cc	% population less extreme	*t*	*p*	*z*-cc	% population less extreme
CFMT+:									
SR1	91.18	2.13	.04*	2.16	97.99	2.14	.04*	2.17	97.91
SR2	90.20	2.04	.05*	2.07	97.55	2.06	.05*	2.10	97.56
SR3	88.24	1.86	.07	1.88	96.39	1.92	.07	1.95	96.71
SR4	92.16	2.23	.03*	2.26	98.36	2.21	.04*	2.25	98.21
SR5	92.16	2.23	.03*	2.26	98.36	2.21	.04*	2.25	98.21
SR6	100.00	2.97	.01*	3.01	99.73	2.79	.01*	2.84	99.53
SR7	92.16	2.23	.03*	2.26	98.36	2.21	.04*	2.25	98.21
SR8	88.24	1.86	.07	1.88	96.39	1.92	.07	1.95	96.71
CFMT-Chinese:									
SR1	83.33	1.63	.11	1.66	94.42	0.40	.69	.40	65.27
SR2	81.94	1.51	.14	1.54	93.03	0.28	.78	.29	60.91
SR3	84.72	1.75	.09	1.78	95.58	0.51	.61	.52	69.43
SR4	87.50	2.00	.05*	2.02	97.29	0.75	.46	.76	76.96
SR5	84.72	1.75	.09	1.78	95.58	0.51	.61	.52	69.43

SR6	87.50	2.00	.05*	2.02	97.29	0.75	.46	.76	76.96
SR7	83.33	1.63	.11	1.66	94.42	0.40	.69	.40	65.27
SR8	93.06	2.48	.02*	2.51	99.08	1.22	.23	1.24	88.29

Table 3: The correlation between *A* scores on same- and other-ethnicity memory and matching tests for Caucasian and Asian controls.

Control Group	Caucasian Tests	Asian Tests
Caucasian	.37*	.46*
Asian	.39*	.62**

*p < .05; **p < .001

Table 4: Modified *t*-tests for single-case comparisons (Crawford & Garthwaite, 2002) for the eight individual super-recognisers in relation to Caucasian and Asian controls performing the Caucasian and Asian matching tests. Significant results are highlighted by an accompanying asterisk. Effect sizes (*z*-cc) are also reported for each calculation, alongside the percentage of the population achieving less extreme scores than each individual super-recogniser.

	Raw score (%)	Caucasian controls				Asian controls			
		t	*p*	*z*-cc	% population less extreme	*t*	*p*	*z*-cc	% population less extreme
Caucasian test:									
SR1	70.83	0.32	.75	0.33	62.50	0.90	.37	0.92	81.27
SR2	87.50	2.41	.02*	2.45	98.93	2.95	.01*	3.00	99.67
SR3	81.25	1.63	.11	1.65	94.37	2.18	.04*	2.22	98.10
SR4	83.33	1.89	.07	1.92	96.63	2.44	.02*	2.48	98.91
SR5	85.42	2.15	.04*	2.18	98.07	2.69	.01*	2.74	99.40
SR6	89.58	2.67	.01*	2.71	99.43	3.20	.003*	3.26	99.83
SR7	89.58	2.67	.01*	2.71	99.43	3.20	.003*	3.26	99.83
SR8	79.17	1.37	.18	1.39	90.98	1.93	.06	1.96	96.77
Asian test:									
SR1	87.50	2.02	.05*	2.05	97.45	1.16	.26	1.18	87.12
SR2	81.25	1.25	.22	1.26	88.92	0.48	.63	0.49	68.30
SR3	87.50	2.02	.05*	2.05	97.45	1.16	.26	1.18	87.12
SR4	83.33	1.50	.14	1.53	92.91	0.71	.49	0.72	75.69
SR5	89.58	2.28	.03*	2.31	98.56	1.38	.18	1.41	91.07
SR6	83.33	1.50	.14	1.53	92.91	0.71	.49	0.72	75.69

SR7	91.67	2.54	.02*	2.58	99.21	1.61	.12	1.64	94.01
SR8	87.50	2.02	.05*	2.05	97.45	1.16	.26	1.18	87.12

Table 5: Mean percentage accuracy (SD) for Caucasian control participants on each of the matching tests.

	Hits	Correct rejections	Overall % accuracy	*A*	*b* (bias)
Caucasian	65.71 (16.45)	70.83 (16.01)	68.27 (7.86)	.75 (.09)	1.18 (.51)
Asian	63.93 (16.73)	78.57 (13.64)	71.25 (7.92)	.79 (.08)	1.46 (.77)
Arab	63.93 (16.79)	79.40 (13.17)	71.67 (7.99)	.79 (.09)	1.46 (.75)
Black	56.79 (17.16)	75.48 (14.25)	66.13 (8.97)	.73 (.11)	1.39 (.59)

Table 6: Correlation matrix for the performance of Caucasian control participants on the matching tests.

	Caucasian	Asian	Arab	Black
Caucasian	1	.53*	.12	.38
Asian		1	.47*	.45*
Arab			1	.57*
Black				1

*$p < .017$, sequential Bonferroni correction applied.

Table 7: Results from modified *t*-tests for single-case comparisons (Crawford & Garthwaite, 2002) for the eight individual super-recognisers in relation to Caucasian controls performing the Arab and Black matching tests. Significant results are highlighted by an accompanying asterisk. Effect sizes (*z*-cc) are also reported for each calculation, alongside the percentage of the population achieving less extreme scores than each individual super-recogniser.

	Raw score (%)	Caucasian controls			
		t	*p*	*z*-cc	% population less extreme
Arab test:					
SR1	87.50	1.95	.06	1.98	97.05
SR2	85.42	1.70	.10	1.72	95.06
SR3	91.67	2.47	.02*	2.50	99.06
SR4	81.25	1.18	.25	1.20	87.73
SR5	97.92	3.24	.003*	3.29	99.87
SR6	87.50	1.95	.06	1.98	97.05
SR7	91.67	2.47	.02*	2.50	99.06
SR8	91.67	2.47	.02*	2.50	99.06
Black test:					
SR1	77.08	1.20	.24	1.22	88.15
SR2	83.33	1.89	.07	1.92	96.64
SR3	85.42	2.12	.04*	2.15	97.93
SR4	79.17	1.43	.16	1.45	91.96
SR5	91.67	2.81	.01*	2.85	99.59
SR6	72.92	0.75	.46	0.76	76.97
SR7	79.17	1.43	.16	1.45	91.96

SR8	87.50	2.35	.02*	2.38	98.76

Figure Captions

Figure 1: (A) Mean percentage correct (SE) for Caucasian and Asian controls on the CFMT+ and CFMT-Chinese, and (B) performance of the super-recognisers in relation to the two control groups. The dotted line represents the 1.96 SD cut-off for the CFMT-Chinese according to the Caucasian control norming data. As the cut-off for the Asian control data hit ceiling on this test, it is not displayed on the graph.

Figure 2: Mean (A) percentage correct and (B) *A* score for Caucasian and Asian controls on the Caucasian and Asian face matching tests. Error bars represent standard error means. (C) represents the overall percentage correct scores for super-recognisers in relation to control participants, and (D) the relationship between super-recogniser performance (% correct) on the CFMT-Chinese and Asian matching test. Dotted lines represent the cut-offs from Caucasian norming data (no super-recogniser exceeded the cut-off from the Asian norming data on the Asian matching test).

Figure 3: (A) Mean percentage correct hits and correct rejections (CRs) for control participants on the four face matching tasks. Standardized scores on each test are displayed for super-recognisers who did (B) and did not (C) surpass the 1.96 cut-off (indicated by the grey line) on an overall index score of other-ethnicity face matching.

Figure 4: Overall other-ethnicity face matching index score for super-recognisers in comparison to (A) CFMT-Chinese and (B) CFMT+ performance.

Figure 1

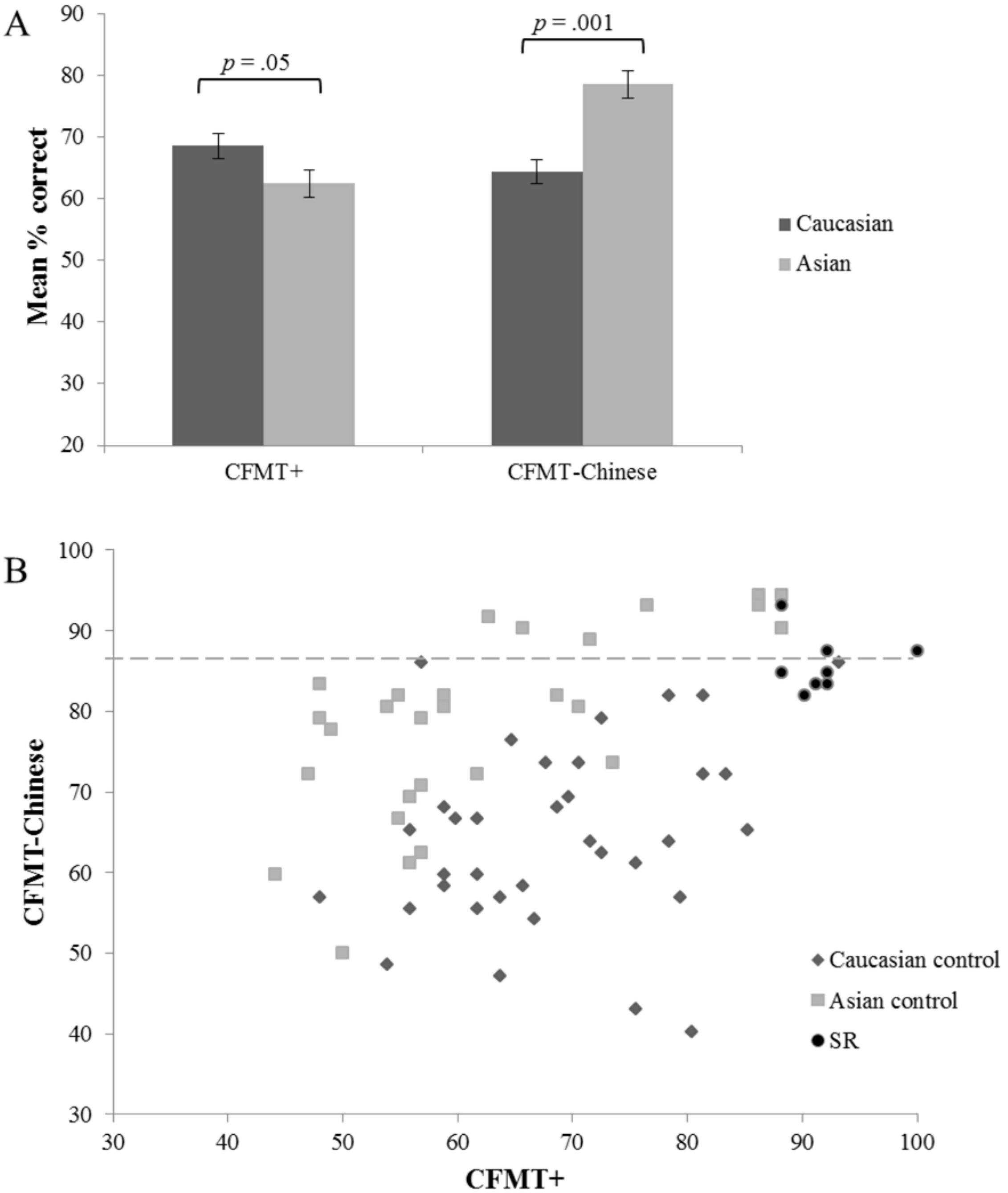

A
p = .05
p = .001
Mean % correct
90
80
70
60
50
40
30
20
CFMT+
CFMT-Chinese
Caucasian
Asian
B
CFMT-Chinese
100
90
80
70
60
50
40
30
40
50
60
70
80
90
100
CFMT+
Caucasian control
Asian control
SR

Figure 2

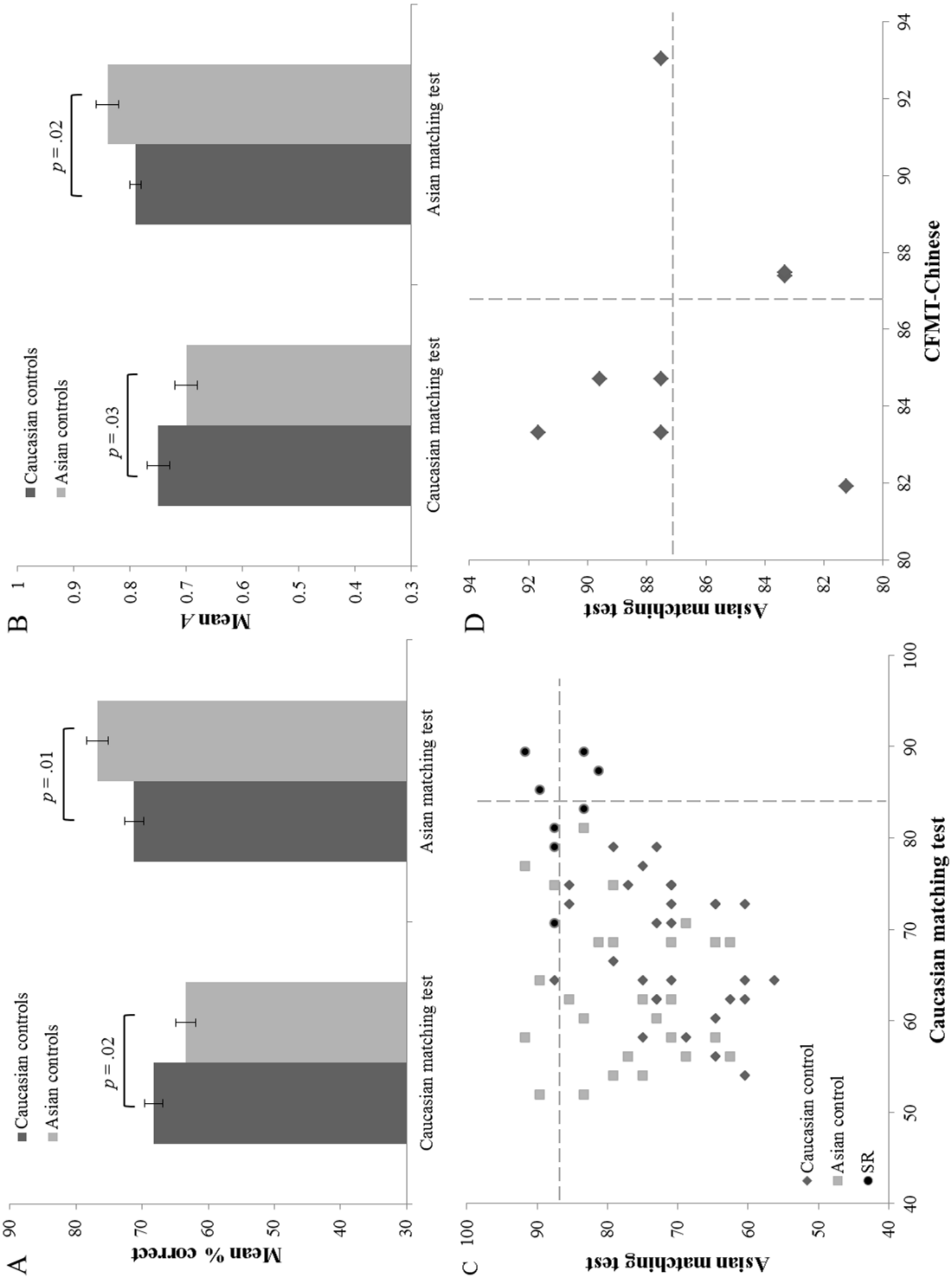

A
Mean % correct
Caucasian controls
Asian controls
p = .02
p = .01
Caucasian matching test
Asian matching test
B
Mean A
p = .03
p = .02
C
Asian matching test
Caucasian matching test
Caucasian control
Asian control
SR
D
CFMT-Chinese

Figure 3

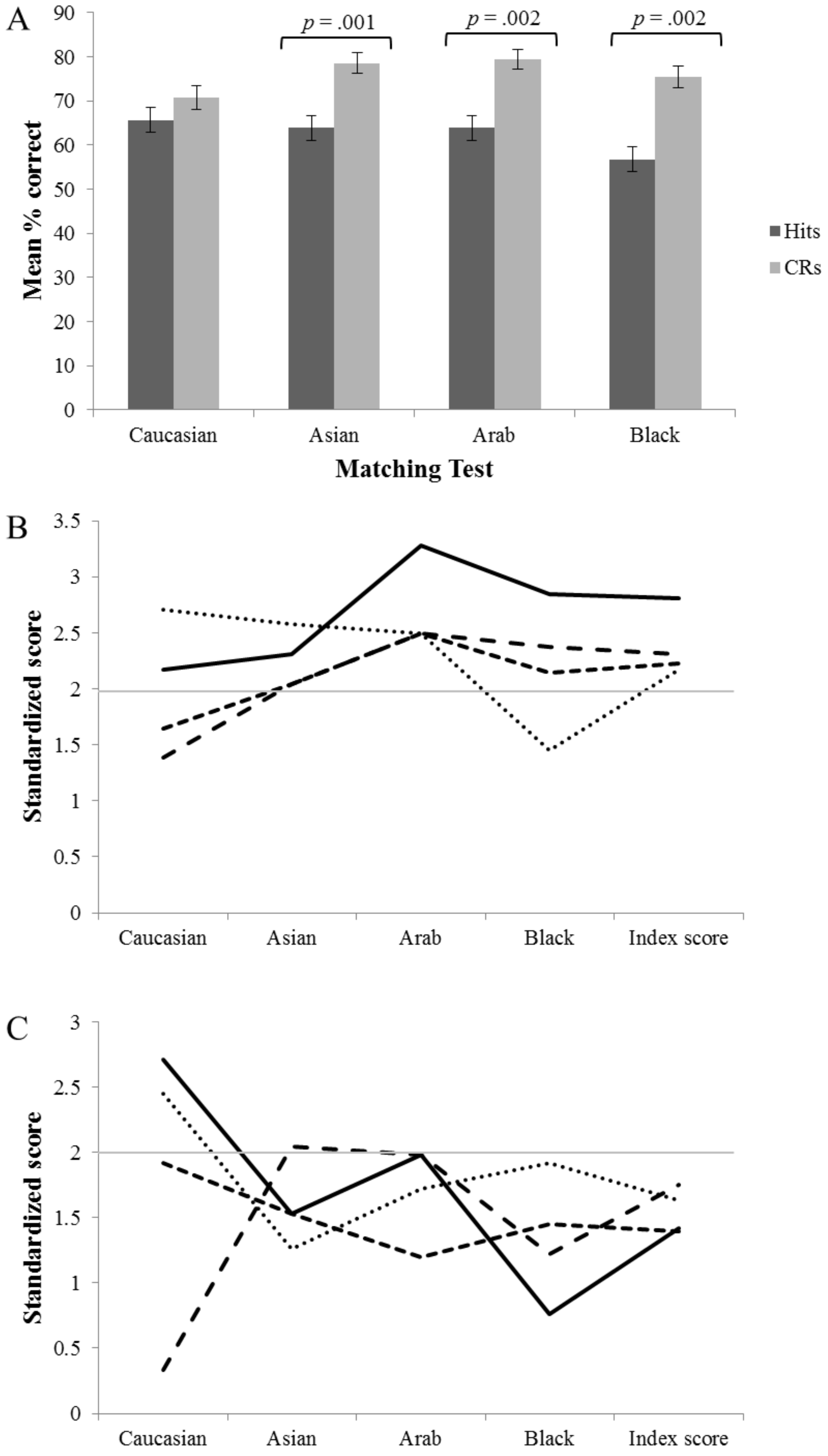
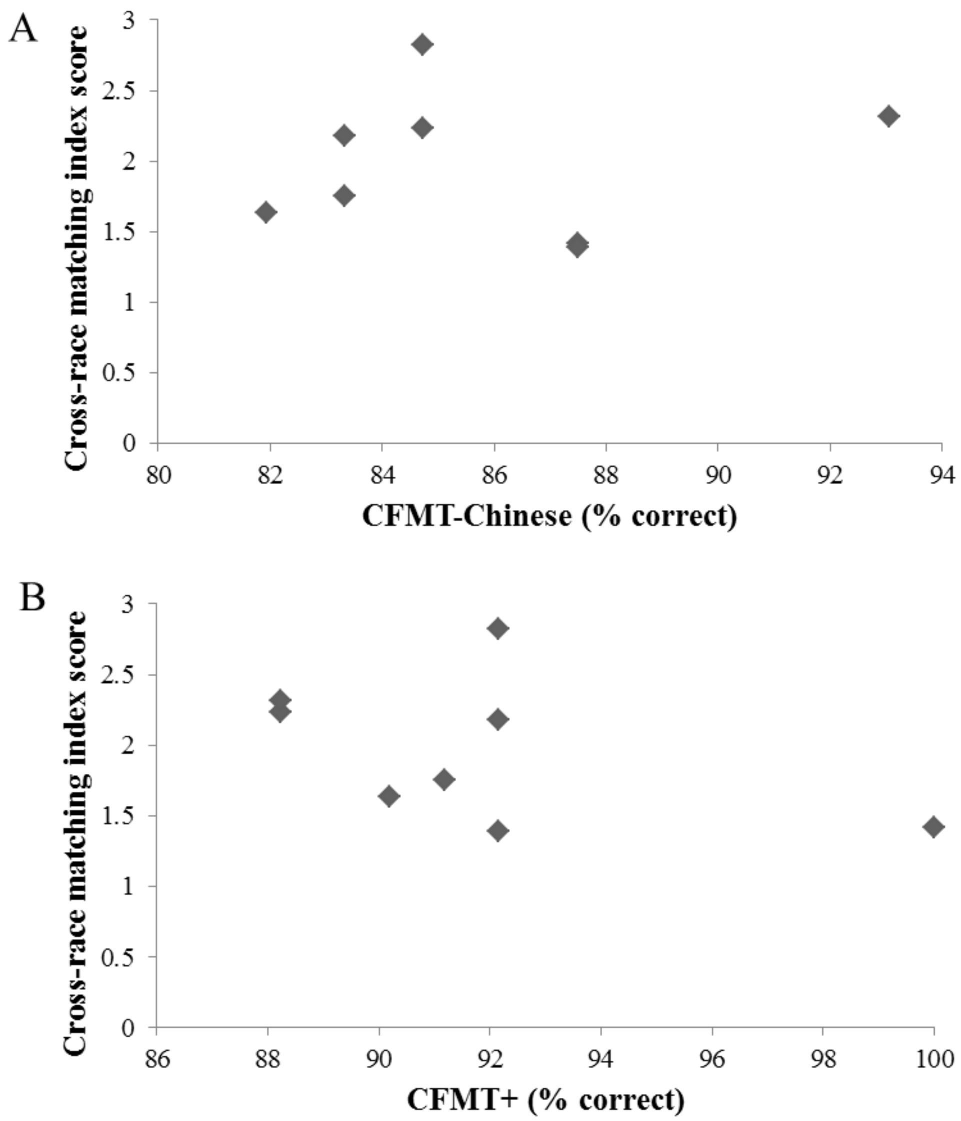

A
90
80
70
60
50
40
30
20
10
0
Mean % correct
p = .001
p = .002
p = .002
Hits
CRs
Caucasian
Asian
Arab
Black
Matching Test
B
3.5
3
2.5
2
1.5
1
0.5
0
Standardized score
Caucasian
Asian
Arab
Black
Index score
C
3
2.5
2
1.5
1
0.5
0
Standardized score
Caucasian
Asian
Arab
Black
Index score

Figure 4


A
Cross-race matching index score
3
2.5
2
1.5
1
0.5
0
80
82
84
86
88
90
92
94
CFMT-Chinese (% correct)
B
Cross-race matching index score
3
2.5
2
1.5
1
0.5
0
86
88
90
92
94
96
98
100
CFMT+ (% correct)