Received: 27 November 2018 | Revised: 1 August 2019 | Accepted: 1 October 2019
DOI: 10.1002/acp.3608

RESEARCH ARTICLE

WILEY

Super-Recognisers show an advantage for other race face identification

David J. Robertson[1] | **Jennifer Black**[1] | **Bethany Chamberlain**[1] | **Ahmed M. Megreya**[2] | **Josh P. Davis**[3]

[1]School of Psychological Sciences and Health, University of Strathclyde, Glasgow, UK

[2]Department of Psychological Sciences, College of Education, Qatar University, Doha, Qatar

[3]Department of Psychology, Social Work and Counselling, University of Greenwich, London, UK

Correspondence
David Robertson, School of Psychological Sciences and Health, University of Strathclyde, Glasgow, UK.
Email: david.j.robertson@strath.ac.uk

Summary

The accurate identification of an unfamiliar individual from a face photo is a critical factor in several applied situations (e.g., border control). Despite this, matching faces to photographic ID is highly prone to error. In lieu of effective training measures, which could reduce face matching errors, the selection of "super-recognisers" (SRs) provides the most promising route to combat misidentification or fraud. However, to date, super-recognition has been defined and tested using almost exclusively "own-race" face memory and matching tests. Here, across three studies, we test Caucasian participants' performance on own- and other-race face identification tasks (GFMT, MFMT, CFMT+, EFMT, CFMT-Chinese). Our findings show that compared to controls, high-performing typical recognisers (Studies 1 and 2) and SRs (Study 3) show superior performance on both the own- and other-race tests. These findings suggest that recruiting SRs in ethnically diverse applied settings could be advantageous.

KEYWORDS

face recognition, identity verification, individual differences, super-recogniser, unfamiliar face matching

1 | INTRODUCTION

The use of face photos for accurate identity verification is critical in maintaining border security and ensuring that correct convictions occur within the criminal justice system. At border control, passport officers are required to decide whether the face of a traveller matches their passport photo, and police officers are routinely required to match the face of a suspect to poor-quality closed-circuit television stills. In each of these cases, the target individuals are likely to be unfamiliar to the police officer or border control official. Despite this, it is now well established that matching pairs of unfamiliar faces is highly prone to error (Burton, 2013; Burton & Jenkins, 2011; Davis & Valentine, 2009; Hancock, Bruce, & Burton, 2000; Jenkins & Burton, 2011; Johnston & Edmonds, 2009; Robertson, 2018; Robertson & Burton, 2016). Notably, errors within this context may lead to travellers with fraudulent passports entering the country illegally or innocent suspects being convicted of a crime.

The Glasgow Face Matching Test (GFMT; Burton, White, & McNeill, 2010) is one of the most widely used tests of unfamiliar face matching (see also Benton, Sivan, Hamsher, Varney, & Spreen, 1983; Bruce et al., 1999; Fysh & Bindemann, 2018). The task uses Caucasian faces, and error rates from Caucasian viewers are typically around 20%. That is, on one in five occasions, individuals will incorrectly state that two faces show the same person when in

Data sharingThe data that support the findings of this study are available from the corresponding author upon request.

fact they are two different people (GFMT mismatch condition; analogous to a fraud attack at passport control). This non-trivial level of error can be exacerbated by a number of other factors such as greater within-person variability in the images (e.g., changes in pose, expression, and hairstyle; Bruce et al., 1999; Bindemann & Sandford, 2011; Megreya, Sandford, & Burton, 2013), the frequency of mismatch items (Papesh & Goldinger, 2014), time pressure (Bindemann, Fysh, Cross, & Watts, 2016; Fysh & Bindemann, 2017), matching fatigue (Alenezi, Bindemann, Fysh, & Johnston, 2015), poor sleep (Beattie, Walsh, McLaren, Biello, & White, 2016), and ageing (Megreya & Bindemann, 2015).

In addition, research has also shown that specialist recognisers (i.e., police officers and passport checkers) generally tend to perform no better on face-matching tasks than non-specialist controls (Burton, Wilson, Cowan, & Bruce, 1999; Kemp, Towell, & Pike, 1997; Tree, Horry, Riley, & Wilmer, 2017; White, Dunn, Schmid, & Kemp, 2015; White, Kemp, Jenkins, Matheson, & Burton, 2014; but see White, Phillips, Hahn, Hill, & O'Toole, 2015). In addition, a number of recent experiments have found it difficult to train people to be better at facial identification, with individual differences in performance often outweighing the magnitude of improvement (e.g., see Robertson et al., 2018; and White, Kemp, Jenkins, & Burton, 2014 for work on feedback training). The difficulty in trying to improve an individual's facial recognition ability is further supported by a recent paper by Towler et al. (2019), which showed that professional facial ID training courses, which are used by agencies across the world, appear to have little or no impact on an individual's person identification performance.

Therefore, focus has now shifted from improving the performance of typical recognisers to the selection of high performing individuals (see Balsdon, Summersby, Kemp, & White, 2018), known as super-recognisers (SRs), who naturally excel at face identification tasks as a result of an inherited (Wilmer et al., 2010), and face-specific ability (McCaffery, Robertson, Young, & Burton, 2018; Wilhelm et al., 2010; Yovel, Wilmer, & Duchaine, 2014). At present, a conservative definition of super-recognition is a minimum accuracy score of 93% (95/102 items correct; Bobak, Pampoulov, & Bate, 2016) on the Cambridge Face Memory Test: Long version (CFMT+; Russell, Duchaine, & Nakayama, 2009), a level of ability that should be present in around 2% of the general population. Recent work has started to assess the processes which may underpin super-recognition and the findings suggest that SRs may focus more on the inner features of unfamiliar faces (particularly the nose region; Bobak, Parris, Gregory, Bennetts, & Bate, 2017), as well as showing enhanced early stage encoding of incoming facial information, compared to typical recogniser controls (Belanova, Davis, & Thompson, 2018).

Despite the advances in establishing neurocognitive markers of super-recognition, the CFMT+ remains the gold standard test for SR categorisation. The CFMT+ is a Caucasian learned face memory test. Participants are asked to memorise the faces of six people, followed by a memory test (3AFC) which includes novel instances of the learned identities. However, as noted above, the critical task at border control and in criminal identification is *unfamiliar face matching*, which does not place any demands on memory and indeed, in the early phase of SR research it was not clear whether CFMT+ SRs would also excel in matching tasks. Research has now addressed that issue and has shown that the superior face memory ability found in CFMT+ SRs does generalise to the unfamiliar face matching domain. A series of recent studies have shown significantly greater accuracy rates for CFMT+ SRs on the GFMT and the more challenging Models Face Matching Test (MFMT) compared to typical recognisers (Bobak, Dowsett, & Bate, 2016; Davis, Lander, Evans, & Jansari, 2016; Robertson, Noyes, Dowsett, Jenkins, & Burton, 2016; see also Bobak, Hancock, & Bate, 2016; Davis, Treml, Forrest, & Jansari, 2018; Noyes, Hill, & O'Toole, 2018; Phillips et al., 2018 for similar findings with newly developed matching tests). In addition, recent individual difference studies have reported positive correlations of moderate strength, between scores on the CFMT+ and the GFMT (e.g., McCaffery et al., 2018; Verhallen et al., 2017; see Fysh, 2018; Fysh & Bindemann, 2018 for equivalent findings with the CFMT/Kent Face Matching Test). Such correlations across face matching and face memory tasks support the existence of Verhallen's *f* (Verhallen et al., 2017), as a common underlying mechanism for face processing akin to Spearman's *g* for intelligence (1927). In the applied context, these findings confirm that CFMT+ SRs can also excel on matching tasks and could therefore be deployed as passport checkers at border control or as officers in criminal identification units in policing.

The finding that CFMT+ SRs also excel at matching pairs of faces is important in terms of the general utility of SRs across different occupations. However, it must still be viewed with caution because the face tasks employed in these studies (CFMT+, GFMT, and MFMT) used only Caucasian faces (see Noyes & O'Toole, 2017), when in the real world, passport checkers and police officers regularly encounter faces from a wide range of ethnic groups. Data from the 2011 U.K. Census (ONS, 2011) showed that six distinct ethnic groups are represented by more than one million U.K. citizens (i.e., White British, all other White, mixed, Asian, Black, and with “other” category representing many additional ethnic groups), and an official may encounter many other non-UK ethnicities at an airport. Verifying an individual's identity from a face photo is challenging enough when the viewer and the target are from within the same ethnic group, however, due to a well-established psychological phenomenon known as the *other-race effect* (ORE), accurately identifying a person from a different ethnic group results in even poorer performance (see Meissner & Brigham, 2001 for a review).

The ORE emerges early in development with infants as young as 9 months of age showing preferential recognition for own-race faces, with initial exposure to predominantly own-race faces shaping adult perception and performance (Kelly et al., 2007; Meissner & Brigham, 2001; O'Toole, Deffenbacher, Valentin, & Abdi, 1994; Walker & Tanaka, 2003). The ORE is present both in the recognition of learned other-race faces (Marcon, Meissner, Frueh, Susa, & MacLin, 2010; McKone et al., 2012; Meissner & Brigham, 2001), and importantly, for the purposes of this study, in face matching tasks (Kokje, Bindemann, & Megreya, 2018; Megreya, White, & Burton, 2011; Meissner, Susa, & Ross, 2013). The presence of the effect in matching tasks suggests poorer encoding of other-race faces during early perceptual processing (Walker & Tanaka, 2003; Zhao & Bülthoff, 2013).

Meissner et al. (2013) demonstrated the ORE using a matching task which mirrored a passport control context, with image pairs showing a high-quality face photo of the "traveller" and a scanned photo-ID page from a passport. They reported the typical 20% error rate in the own-race condition (Mexican American observers/faces), which rose to 30% in the other-race condition (Mexican American observers/African American faces). In addition, findings from Megreya et al. (2011) displayed the ORE in a 1–10 matching task (UK/Egyptian faces/observers). Intriguingly, this study also reported moderate-to-strong correlations between accuracy rates on the own- and other-race tests for both groups (r = .60 UK Observers, r = .78 Egyptian observers), although the sample size here was small (N = 26 for both groups). This suggests that participants who excelled on the own-race task were also likely to excel on the other-race task (relative to a lower mean score). Recent work by Kokje et al. (2018) replicated both the ORE effect and the own-/other-race accuracy correlation with a larger sample (N = 74) using one–to–one matching tasks. However, they did not use the CFMT+, or the GFMT, when assessing individual differences in performance, limiting the generalisability of their findings to typical recognisers.

To date, only one paper by Bate et al. (2018a) has attempted to directly assess the performance of SRs on other-race face identification tests. Using a sample of eight Caucasian SRs, Bate et al. (2018a) presented participants with own and other-race face memory tests (Experiment 1) and own and other-race face matching tests (Experiments 2 and 3). They reported that their sample of SRs did not show a performance advantage over native typical recognisers (i.e., Asian observers/Asian face tests). However, the SRs did show an advantage over the Caucasian controls on the other-race face tests, although the accuracy cost for other-race faces remained, with no difference in magnitude compared to the control group. That is, the ORE was present in SRs, albeit from a higher baseline level of performance compared to controls. These are intriguing findings and they suggest that SRs may be performing at the top end of a face recognition continuum rather than displaying qualitatively different cognitive processes. However, the findings from Bate et al. (2018a) should be treated with caution, as both the size of the SR sample (N = 8) and its heterogeneity precluded statistical comparisons at the group level. Further work, with a larger SR sample is required to test the robustness of their findings.

Therefore, the present study sought to investigate individual differences in performance across a range of own- and other-race face tests in typical recognisers (Studies 1 and 2) and a large sample of Caucasian SRs (Study 3). In Study 1, we test a large sample of *typical recognisers* (Caucasian undergraduate students; N = 111) using a battery of facial identification tests that tap own-race face memory (CFMT+), own-race face matching (GFMT-short, MFMT-short), other-race face memory (CFMT-Chinese), and other-race face matching processes (Egyptian Face Matching Test [EFMT]-long), to assess whether *typical recognisers* who perform accurately on own-race tests also show similar levels of performance on the other-race tests. If that is the case, it would provide support for a common mechanism that underlies both own- and other-race identity perception. In Study 2, using a sample of typical recognisers (Caucasian undergraduate students; N = 43), we verify a shortened 40-item version of the other-race face matching test used in Study 1 (EFMT-short). Finally, in Study 3, in order to directly assess SRs' performance on own- and other-race face tasks, we test a large sample of Caucasian SRs (N = 35) using the CFMT+, Adult Face Recognition Test (AFRT), MFMT, and the other-race EFMT-short, relative to Caucasian typical recogniser controls (N = 420). Following the process reported by Bate et al. (2018a), we seek to assess whether Caucasian SRs outperform Caucasian controls on an other-race unfamiliar face matching test, to identify whether an other-race accuracy cost is evident in the SRs, and if so, to what extent.

2 | STUDY 1

In Study 1, we use four established face tests (CFMT-short, CFMT-Chinese [CFMT-C], GFMT-short, MFMT-short) and the 200-item EFMT-long (100 match/100 mismatch trials). Here, we seek to replicate previous work which has shown a robust correlation between the CFMT (learned face memory) and the GFMT (face matching). We also include the more challenging MFMT (face matching; highly variable male model images) as a direct correlation between this task and the CFMT has not been previously reported. Importantly, we also include an other-race face matching test (EFMT-long; Egyptian faces), and we assess whether this task produces an other-race accuracy cost and whether accuracy on the own-race GFMT generalises to the other-race EFMT-long. Although the focus of this paper is on other-race face matching, we also include the CFMT-Chinese version (McKone et al., 2012) to assess cross-domain (i.e., matching/memory) and cross-race correlations (Caucasian, Egyptian, Chinese). The short version of the CFMT is used in this study, rather than the CFMT+, and so we cannot determine if there are any SRs in the sample. Therefore, in Study 1, we test typical recognisers (undergraduate students) only.

2.1 | Method

2.1.1 | Ethical approval

Each study reported in this paper received ethical approval from the Ethics Committee of the University of Strathclyde School of Psychological Sciences and Health. Study 3 received concurrent approval from the University of Greenwich Research Ethics Committee.

2.1.2 | Participants

One hundred and eleven Caucasian participants with a mean age of 22 years (SD = 5, $Range$ = 18–53, 18 male) were recruited from the

University of Strathclyde School of Psychological Sciences and Health. All participants had normal or corrected-to-normal vision, each provided written informed consent, and upon completion of the study each received a course credit or an optional piece of confectionary.

2.1.3 | Stimuli and apparatus

Glasgow Face Matching Test

The GFMT (short version) consists of 40 pairs of unfamiliar Caucasian faces. The test contains an equal number of trials in which the face pairs show the same person (match condition) or two different but similar looking people (mismatch condition). See Figure 1 for an example image pair and Burton et al. (2010) for further details.

Models Face Matching Test

The MFMT (short version) consists of 30 pairs of unconstrained highly variable face photos of male models (15 match/15 mismatch). The MFMT is designed to be more difficult than the GFMT and, in line with the CFMT/CFMT+ distinction, is more likely to detect high-performers. See Figure 1 for an example image pair and Dowsett and Burton (2015) for further details.

Two hundred-item Egyptian Face Matching Test Long version

The EFMT-long consists of 200 pairs of unfamiliar male Egyptian faces (100 match/100 mismatch), an example is provided in Figure 1 (see Megreya et al., 2011 for further details).

Cambridge Face Memory Test

The CFMT (short version) is a well-established 72-item learned face recognition task, which increases in difficulty with the addition of within-person variability and visual noise to the image set. Figure 1 shows an example of the stimuli used in the CFMT, see Duchaine and Nakayama (2006) for further details.

Cambridge Face Memory Test—Chinese version

The CFMT-C follows an identical format to that described above for the CFMT with the exception that Chinese faces replace the

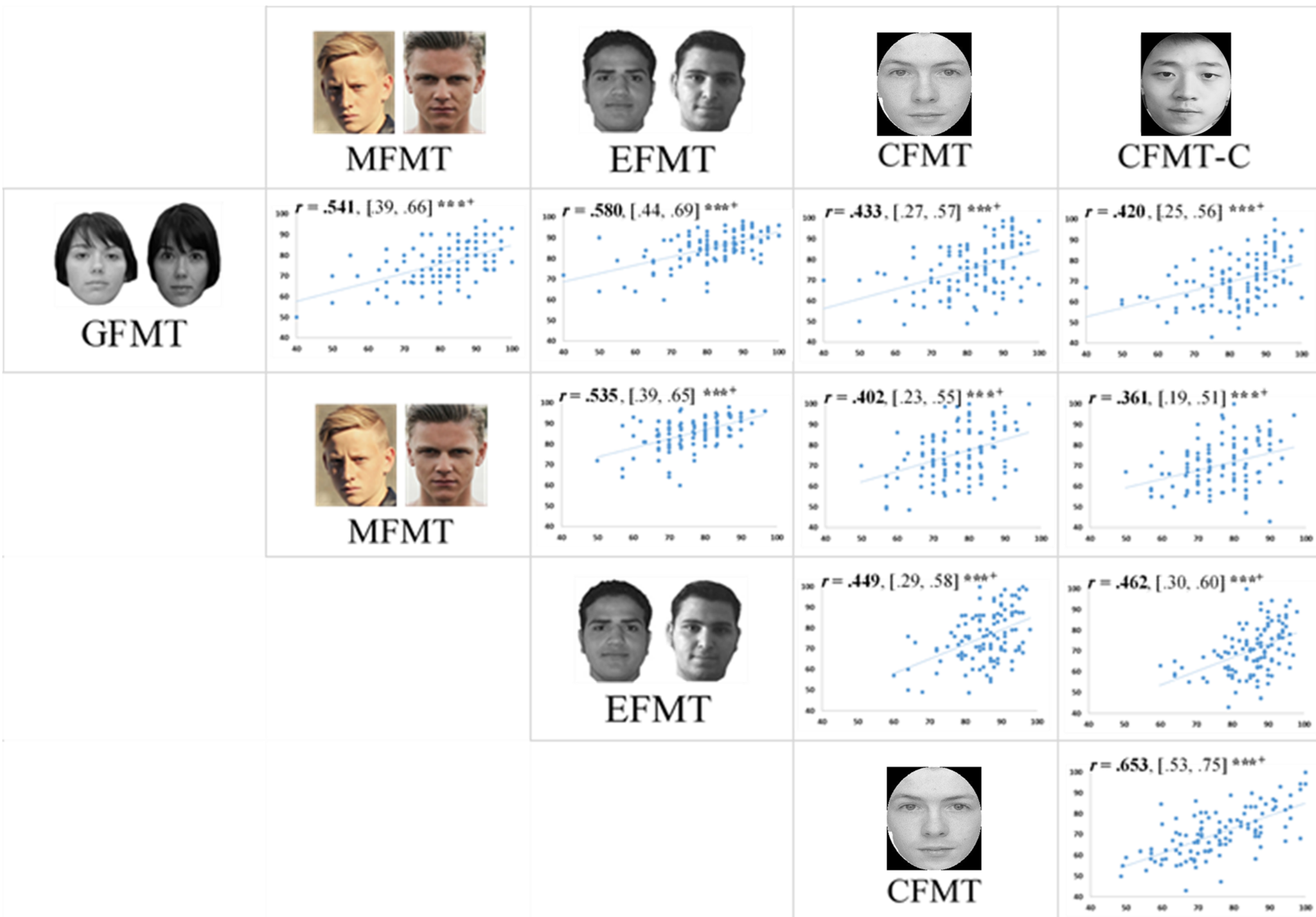

FIGURE 1 Correlation matrix for the five-face identification tests used in Study 1; Glasgow Face Matching Test (GFMT), Models Face Matching Test (MFMT), Egyptian Face Matching Test (EFMT), Cambridge Face Memory Test (CFMT), Cambridge Face Memory Test-Chinese Version (CFMT-C). *** $p < .001$, + significant after Benjamini–Hochberg and more conservative Bonferonni correction ($\alpha = .05/10 = .005$) for multiple comparisons. Note, due to copyright reasons, the faces we show for the MFMT and the CFMT/CMFT-C are not items from the tests but are a good approximation of the stimuli used (all images used are in the public domain and have CC0 licences). [Colour figure can be viewed at wileyonlinelibrary.com]

Caucasian faces used in the original test. See McKone et al. (2012) for further details.

2.1.4 | Procedure

Task order was randomised by domain (face matching tests, face memory tests) and then by test (GFMT/MFMT/EFMT-long, CFMT/CFMT-C). On each trial of the face matching tests, participants were required to decide whether the face pair showed the same person or two different people. Each trial remained on screen until response. For the face memory tests, participants were required to learn six target identities by viewing photos of them in three different orientations (left, forward facing, right) and to then detect photos of these identities in the presence of two foils in 3-AFC recognition trials. Recognition trials remained onscreen until response. All responses were made via keyboard key with the testing session lasting approximately 1 hr.

2.2 | Results

2.2.1 | Task accuracy

Unfamiliar face matching (GFMT, MFMT, EFMT-long)

For the own-race matching tasks, mean accuracy on the GFMT was 82% (*SD* = 11%, *Range* = 40%–100%), and 77% on the MFMT (*SD* = 10%, *Range* = 50%-97%), and these scores are in line with published norms (see Burton, White, & McNeil, 2010; Dowsett & Burton, 2015). As expected, mean accuracy on the MFMT was significantly lower than that found for the GFMT, $t(110) = 5.45$, $p < .001$, $d = .49$, supporting its use as an assessment tool for unfamiliar face matching ability at the top end of the performance distribution.

Although research shows that accuracy on other-race tasks is poorer than own-race face tasks, here we find that EFMT-long accuracy (*M* = 85%, *SD* = 8%, *Range* = 60%–98%) was significantly higher than both the GFMT and MFMT ($t(110) = 4.16$, $p < .001$, $d = .33$ for the GFMT; $t(110) = 11.06$ $p < .001$, $d = .90$ for the MFMT). This pattern is likely to be due to the fact that, as mentioned above, the GFMT and MFMT consist of the most difficult items from longer test sets. This is not the case for the EFMT-long, in which the full 200 trials were used, and so accuracy is likely to be inflated by the inclusion of a greater proportion of easy trials. Therefore, in Study 2 we develop a shortened version of this task based on trial accuracy data from the current data set.

Face recognition memory (CFMT, CFMT-C)

For the learned face memory tests, scores were again in line with published norms (see Duchaine & Nakayama, 2006; McKone et al., 2012), with a 76% mean accuracy rate for the CFMT (*SD* = 12%, *Range* = 49%–100%) and 71% for the CFMT-C (*SD* = 11%, *Range* = 43%–100%). The difference between the scores was significant, $t(110) = 5.69$, $p < .001$, $d = .52$, confirming that the other-race CFMT-C provided a more challenging face memory test for Caucasian observers, in comparison to the own-race CFMT.

2.2.2 | Individual differences

As our principal aim was to explore potential correlations between the different measures, we were more concerned with avoiding Type 2 than Type 1 errors and therefore report uncorrected statistics. However, as a reliability check, we also we also used the Benjamini–Hochberg procedure with a false discovery rate of 0.2 to correct for multiple comparisons, and we also report confidence intervals (see McCaffery et al., 2018).

Unfamiliar face matching (GFMT, MFMT, EFMT-long)

As seen in Figure 1, there was a significant positive correlation between the GFMT and the MFMT ($r(111) = .541$, uncorrected $p < .001$, 95% CI [.39, .66]) with individuals who perform highly on the GFMT also performing highly on the MFMT. This correlation replicates the effect reported by Bobak, Dowsett, and Bate (2016) and shows a level of stability in matching aptitude across the GFMT and the MFMT. It further supports the use of the MFMT as a more sensitive measure of face matching ability amongst high performers.

Importantly, participants' scores on the own-race GFMT and MFMT both correlated with the other-race EFMT-long ($r(111) = .580$, uncorrected $p < .001$, 95% CI [.44, .69] for the GFMT; $r(111) = .535$, uncorrected $p < .001$, 95% CI [.39, .65] for the MFMT). This finding extends previous research by Megreya et al. (2011) who reported a similar relationship using 1–10 face matching arrays. These findings suggest that individuals who perform highly in matching pairs of unfamiliar faces from their own race, are also likely to perform highly when exposed to other-race faces.

Face recognition memory (CFMT, CFMT-C)

Here, we replicate the strong positive correlation reported by McKone et al. (2012) between performance on the own-race Caucasian CFMT and the other-race CFMT-C, $r(111) = .653$, uncorrected $p < .001$, 95% CI [.53, .75]. This finding shows that individuals with a high aptitude for the recognition of new instances of a recently learned own-race face, are also like to perform well when the target identity is from a different ethnic group.

Cross-domain and cross-race correlations

As shown in Figure 1, all of the cross-domain (matching, memory) tests correlated with each other, suggesting shared underlying mechanisms for identity verification in both matching and memory contexts. Although it has previously been established that scores on the CFMT and the GFMT correlate (McCaffery et al., 2018; Verhallen et al., 2017), this is the first study to show such relationships between these tests and the other-race face tasks included in the battery. Importantly, we show a significant positive correlation between the CFMT and both the own-race GFMT ($r(111) = .433$, uncorrected $p < .001$, 95% CI [.27, .57]) and the other-race EFMT-long ($r(111) = .449$, uncorrected $p < .001$, 95% CI [.29, .58]). That is, aptitude on a face memory test generalises to both own- and other-race unfamiliar face matching accuracy. Taken together, the findings from Study 1 provide support for the view that a general face processing factor *f* (Verhallen et al., 2017) exists, which supports face processing across matching and memory domains for both own- and other-race faces.

3 | STUDY 2

As reported in Study 1, mean accuracy on the 200-item EFMT-long was higher than the 40-item GFMT-short (which consists of the 40 most challenging items from the GFMT-long). In Study 2, we follow the same procedure as Burton et al. (2010) by selecting the 40 most difficult items (i.e., least accurate responses) from the EFMT set used in Study 1, to create a shorter version of the task.

3.1 | Method

3.1.1 | Participants

Forty-three Caucasian participants were recruited from the University of Strathclyde School of Psychological Sciences and Health, with a mean age of 23 years (*SD* = 5, *Range* = 18–44, 11 male). All participants had normal or corrected-to-normal vision, each provided written informed consent, and upon completion of the study they received a course credit.

3.1.2 | Stimuli, apparatus, and procedure

In this study, only the GFMT and our shortened version of the EFMT were used. In line with the GFMT, the EFMT-short used in the present study consisted of 40 trials (20 match/20 mismatch). The tasks were presented on a Dell PC, task order was counterbalanced, and trial order was randomised across participants.

3.2 | Results

3.2.1 | Task accuracy

Mean accuracy on the shortened version of the other-race EFMT was 74%, significantly lower than the own-race GFMT (81%), $t(42) = 4.43$, $p < .001$, $d = .64$. As seen in Figure 2, accuracy on the EFMT-short was lower in both the match and mismatch conditions ($t(42) = 2.24$, $p = .030$, $d = .32$ for match; $t(42) = 2.97$, $p = .005$, $d = .44$ for mismatch), and in line with the GFMT, accuracy rates did not differ between EFMT-short match and mismatch conditions, $t < 1$. We note here that although the EFMT-short produced lower accuracy rates than the GFMT, without the inclusion of an Egyptian sample of participants we cannot say conclusively that our EFMT-short produces an ORE. It could be the case that the EFMT-short items are simply more difficult than the GFMT items, we thank Reviewer 3 for bringing this to our attention. However, 75% of the items used in our EFMT-short were also included in a longer test by Kokje et al. (2018), and an analysis of those items from that data set revealed that mean accuracy rates for the Egyptian observers was 78%, that is 4% more accurate than our Caucasian observers. This suggests that should Study 2 be replicated with the inclusion of an Egyptian sample, that it would be likely that the EFMT-short would generate an ORE on accuracy rates. Even were it to be the case that this data was not available from Kokje et al. (2018), the EFMT-short would still provide a valid measure with which to assess between group differences in identification accuracy using own- and other-race faces, which is the aim of Study 3.

3.2.2 | Individual differences

We replicate the findings from Study 1 with a significant positive correlation between overall scores on the GFMT and the EFMT-short ($r(43) = .454$, uncorrected $p = .002$, 95% CI [.18, .66]), again showing consistency in performance across own-race and other-race unfamiliar face matching tests. In addition, significant correlations were found

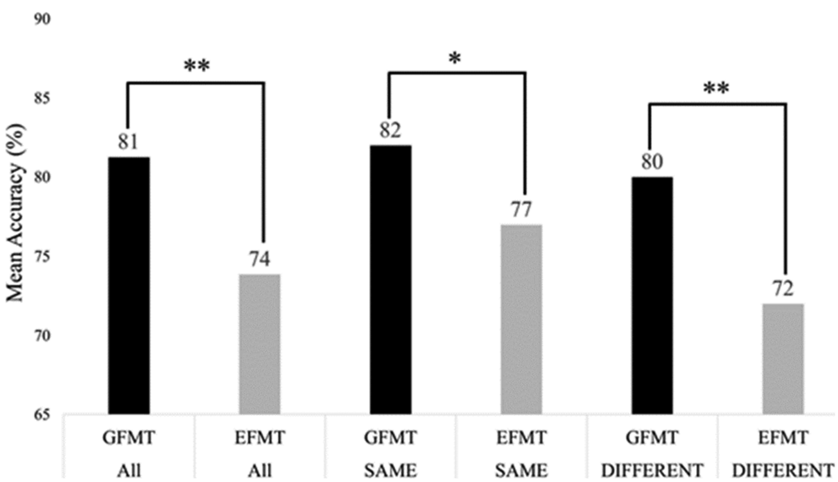

FIGURE 2 Mean accuracy for the Glasgow Face Matching Test (GFMT) and the shortened version of the Egyptian Face Matching Test EFMT (40 Trials), and separately, their match and mismatch conditions, *$p < .05$, **$p \le .005$

across the tests when the match and mismatch trials were analysed separately ($r(43) = .532$, uncorrected $p < .001$, 95% CI [.27, .71] for match trials; $r(43) = .390$, uncorrected $p = .010$, 95% CI [.10, .61] for mismatch trials). These correlations remained significant after applying both the Bonferroni and Benjamini–Hochberg corrections.

4 | STUDY 3

Having developed our short version of the EFMT, we now use this test to assess performance in a group of Caucasian SRs. In addition to the EFMT-short, we also test participants using the CFMT+, the old/new AFRT, the MFMT, and GFMT-short. In doing so, we seek to replicate previous work which has shown that CFMT+ SRs outperform typical recognisers on the GFMT and to extend this work to other-race face matching

4.1 | Method

4.1.1 | Participants

Seven hundred and forty-four participants were recruited via an existing University of Greenwich face recognition participant database. One hundred and sixty-five participants were removed from the data set as they were not Caucasian, or as a result of failing to input a valid participant code, or for not providing consent for their previous CFMT+ and GFMT test scores to be included in the current study. The final sample consisted of 420 Caucasian participants of mean age 36 years (*SD* = 12, *Range* = 16-75, 57% female). From this sample, we identified 60 individuals who met or exceeded the score required for categorisation as a SR (i.e., a score ≥ 95/102 on the CFMT+). Although meeting the CFMT+ cut-off score is the current standard practice for SR categorisation, we sought to increase the validity of our sample by only including those who had also shown a level of superior performance on the AFRT. Therefore, in order to "verify" individuals within our SR group, we excluded any participants who scored below the SR mean on the AFRT (a score of 83%). In doing so, we follow a similar approach to Belanova et al. (2018), but we use a more conservative AFRT cut-off score. Following that SR verification criteria, 25 participants were removed from the analysis (i.e., their SR status had not been verified).

The final groups consisted of 35 SRs with a mean age of 36 years (*SD* = 9, *Range* = 20-57, 57% female), and 360 typical recogniser controls of mean age 36 years (*SD* = 12, *Range* = 16–75, 58% female). The percentage of SRs in the sample was 8.3%, this is higher than the 2.4% we would expect in the typical population based on data from Bobak, Pampoulov, and Bate (2016), and is likely to be a consequence of the sample being recruited from an existing database which actively sought SRs. There was no significant difference in age between the groups, $t < 1$, and there was a similar proportion of male to female participants.

4.1.2 | Stimuli and apparatus

Participants had previously completed the CFMT+, the GFMT, and the AFRT. The AFRT is a White Caucasian face learning and recognition memory test, and in line with Belanova, Davis and Thompson (2018; see paper for full AFRT details), we used scores on this test as an additional criteria for verification of SR status. In this study, the participants completed the MFMT, the EFMT-short, and a morph detection task (the findings from this task are not described here), each task was presented online using Qualtrics.

4.1.3 | Procedure

Three thousand participants from the University of Greenwich Face Recognition Database were invited to take part in this study via an email advert. Each of these participants had previously completed the CFMT+, AFRT, and the GFMT and consented to having their scores on these measures retained for potential use in future studies. Participants completed the 40-item EFMT-short, the MFMT, and a morph detection task. All trials were self-paced, task order and trial order were randomised across participants, and feedback scores were provided at the end of the study.

4.2 | Results

4.2.1 | Task accuracy

Group comparisons

For the typical recogniser control group, mean accuracy rates on the tasks were: 80% for the CFMT+ (*SD* = 11%, *Range* = 46%–92%), 75% for the AFRT (*SD* = 9%, *Range* = 40%–95%), 91% for the GFMT (*SD* = 7%, *Range* = 58%–100%), 83% for the MFMT (*SD* = 9%, *Range* = 53%–100%), and 86% for the EFMT-short, the other-race face matching task (*SD* = 8%, *Range* = 55%–100%). Mean performance on each of these tests is around 8–10% higher than previously published norms, which is likely to be due to a recruitment bias in which those likely to take part in this study have an interest in superior face recognition abilities. Importantly, these results replicate our findings from Study 2, with poorer performance on our newly established short version of the other-race EFMT in comparison with the own-race GFMT, $t(359) = 10.87$, $p < .001$, $d = .64$.

For the SR group, mean accuracy on each of the tests was significantly higher than that found for the control group. Mean accuracy rates for the SR group were: 95% for the CFMT+ (*SD* = 2%, *Range* = 93%–100%; $t(393) = 8.36$, $p < .001$, $d = 1.43$ for the SR/control group comparison), 88% for the AFRT (*SD* = 5%, *Range* = 83%–100%; $t(393) = 8.19$, $p < .001$, $d = 1.49$), 97% for the GFMT (*SD* = 4%, *Range* = 88%–100%; $t(393) = 4.65$, $p < .001$, $d = .88$), 89% for the MFMT (*SD* = 7%, *Range* = 73%–100%; $t(393) = 4.24$, $< .001$, $d = .68$), and 94% for the other-race EFMT-short (*SD* = 6%, *Range* = 80%–100%; $t(393) =$

5.59, p < .001, d = 1.02; these comparisons remained significant after the application of the Bonferroni correction for multiple comparisons).

These findings replicate previous work which has shown that SRs who have been classified on the basis of CFMT+ scores (i.e., face memory) also outperform typical recognisers on unfamiliar face matching tests (i.e., GFMT/MFMT; see Bobak, Dowsett, & Bate, 2016; Davis et al., 2016; Robertson et al., 2016). However, the important finding is that SRs' face matching ability generalises to other-race faces. That is, a Caucasian SR working as a border control officer may be more likely to detect fraud attacks than a typical recogniser, even when the travellers involved are from outside their ethnic group.

It is important to note that although SRs display enhanced accuracy on the EFMT-short in comparison to controls, the SRs still performed less accurately on the other-race EFMT-short (94%) compared with the own-race GFMT (97%; t(34) = 2.67, p = .012, d = .49 for the difference). For the SRs, the mean difference in accuracy between the EFMT-short and GFMT was 3%, which was not significantly smaller than the 5% effect reported between the tests for the typical recogniser controls, t(393) = 1.53, p = .128, d = .32. However, again, this could be due to the recruitment bias in the control group outlined above, and when the size of the SR difference in accuracy between the own- and other-race tests (3%) was compared with the typical recognisers recruited for Study 2 (7%; students), the magnitude of the SR cost was found to be significantly smaller, t(76) = 2.33, p = .022, d = .44. We note again, that our claim that the EFMT-short produces an other-race task cost should be replicated in a fully crossed design, which includes native Egyptian observers.

4.2.2 | Individual differences

Typical recogniser control group

In line with the findings from Studies 1 and 2, we find significant positive correlations between the face memory tests used to categorise SRs (CFMT+, AFRT) and the own-race unfamiliar face matching tests (GFMT, MFMT; all r's > .27, all p's <. 001, 95% CI's [.14, .59]; all correlations remained significant after applying the Benjamini–Hochberg and Bonferroni corrections). Importantly, in these results, we replicate our previous finding which showed a significant positive correlation between the CFMT+ (own-race SR categorisation task) and the EFMT-short (other-race matching task; r(360) = .361, p < .001, 95% CI [.27, .45]) and between the own-race matching tasks (GFMT, MFMT), and the other-race EFMT-short (r(360) = .455, p < .001, 95% CI [.37, .53] for GFMT, r(360) = .436, p < .001, 95% CI [.35, .52] for MFMT).

Superrecogniser group

In contrast to the typical recogniser group, and as expected, there were no correlations between the CFMT+ and any of the other tests (all p's > .076), a consequence of selecting SRs on the basis of the CFMT+ scores, thus removing most of the variance from that set, which would allow for an individual differences analysis.

However, this reduction in variance had less of an effect on the remaining tests scores, and we again report a positive correlation between the EFMT-short and the MFMT, r(35) = .363, p = .032, 95% CI [.12, .56], and a correlation between the EFMT-short and GFMT, which approaches significance, r(35) = .319, p = .062, 95% CI [.07, .53].

Superior performance across all tests

Although the majority of SRs did produce scores above mean control performance across tasks, it is important to note that three SRs scored below the control mean on the GFMT, four SRs scored below the control mean on the MFMT, and two SRs scored below the control mean on the EFMT-short. Therefore, it is not the case that all SRs, as categorised by the CFMT+ and the AFRT, will always show superior performance on other facial identification tasks. Moreover, if we apply the conservative CFMT+ criteria for SR (i.e., ≥2 SDs above the control mean) to the other tests, then, as seen in Figure 3, 16/35 SRs achieved this for the GFMT, 3/35 for the MFMT, and 9/35 for the EFMT-short. Out of the sample of 35 SRs, only one participant achieved scores of 100% across each of the three face matching tests.

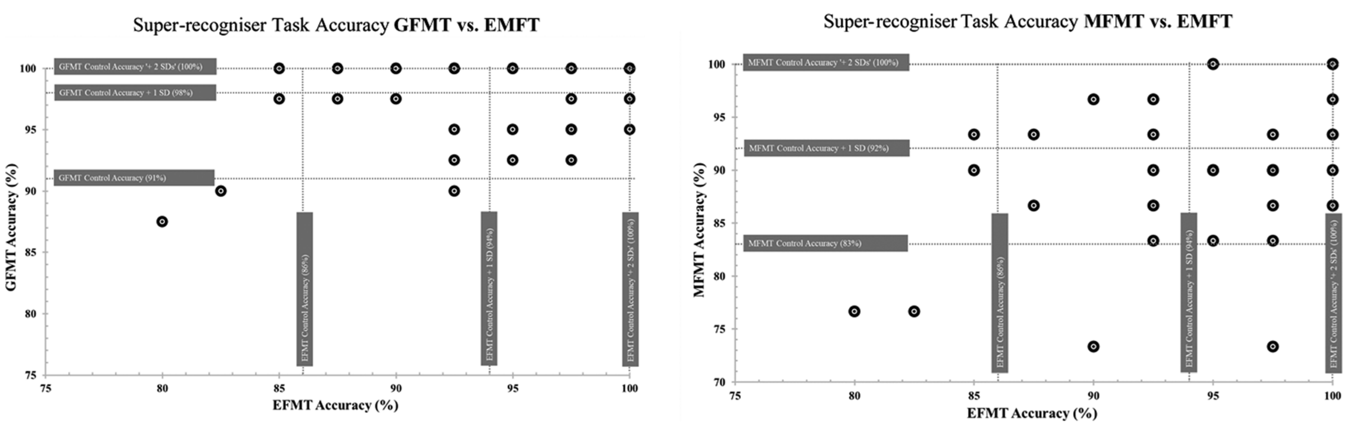

FIGURE 3 Left scatterplot shows the super-recogniser group correlation between the Glasgow Face Matching Test (GFMT; own-race matching) and the Egyptian Face Matching Test (EFMT; other-race matching). Right scatterplot shows the superrecogniser group correlation between the Models Face Matching Test (MFMT; own-race matching) and the EFMT (other-race matching)

This has implications in terms of the types of tests that should be used to categorise SRs for specific occupations, as outlined in the general discussion.

5 | GENERAL DISCUSSION

Across three studies, we demonstrate a consistent performance cost for other-race face identification, both in the context of recognition memory (Study 1; CFMT/CFMT-C) and unfamiliar face matching (Studies 2-3; GFMT, MFMT, EFMT), we show that Caucasian SRs do outperform Caucasian controls on an other-race face matching test but that an other-race accuracy cost remains evident in that group.

Study 1 is, to our knowledge, the first to assess cross-domain matching/memory performance in own-/other-race tasks using this battery of tests (CFMT, CFMT-C, GFMT, MFMT, EFMT-long) in a single well powered sample. The findings from Study 1 replicate previous work showing consistency in performance on the CFMT and GFMT (McCaffery et al., 2018; Verhallen et al., 2017), and we extend this to the more challenging MFMT. The latter effect supports the idea that individuals who excel on the CFMT and GFMT are also likely to perform well in more ecologically valid tasks that contain highly variable face photos. Most importantly, we show that performance on the CFMT and GFMT correlate with scores on the EFMT-long. This suggests that performing well on own-race face memory/matching tasks is likely to result in superior performance when an individual encounters faces from outside their own ethnic group (Kokje et al., 2018; McKone et al., 2012; Megreya et al., 2011; Meissner et al., 2013). This finding along with the other cross-domain correlations (e.g., CFMT-C vs. GFMT) adds further support to the idea that both face matching/memory and own-/other-race face processing may tap the same underlying cognitive and perceptual processes, which Verhallen et al. (2017) has termed *f*, a general face perception factor (analogous to Spearman's *g* in the study of intelligence; Spearman, 1927), which may be distinct from nonface cognitive abilities (McCaffery et al., 2018; Wilhelm et al., 2010). However, although Verhallen et al. (2017) used a variety of face tests to assess the potential for a general face factor, *f*, further work, including a variety of object based and other nonface tasks is required to test whether this factor is indeed specifically indicative of individual differences in face processing.

Having assessed cross-domain performance in typical recognisers in Study 1 and verified our 40-item EMFT-short in Study 2, in Study 3, we used a battery of tests to assess own- and other-race face identification in a set of Caucasian SRs in comparison to Caucasian controls. The findings showed that although there was an SR advantage for accurately matching pairs of other-race faces, with an 8% increase in mean performance over controls, SR accuracy on the EFMT-short was still lower than scores on the own-race GFMT. These findings support the recent work by Bate et al. (2018a), which also showed that SRs outperformed typical recognisers on other-race face tests but that an accuracy cost or ORE remained evident in the SR group. Both the study by Bate et al. (2018a) and the present findings provide support for the view that the SRs are displaying performance at the top end of a face recognition continuum, rather than engaging qualitatively different cognitive and perceptual processes. One limitation of the present study is that it did not include native Chinese (Study 1) or Egyptian (Studies 1 and 3) control groups and therefore we were not able to test whether SRs would outperform native observers. However, Bate et al. (2018a) did include native control groups and they found that although Caucasian SRs outperformed Caucasian controls on other-race tests, the native observers (e.g., Asian observers/Asian face test) outperformed both of these groups. This suggests that although employing a Caucasian SR at border control may lead to greater detection of fraud attacks by other-race travellers, a native observer who shares the fraudsters ethnic group would outperform that SR. Again, the sample size used in the study by Bate et al. (2018a) was small, so future work should seek to further assess this native observer versus SR advantage.

The persistence of an ORE in SRs in the study by Bate et al. (2018a) and the other-race accuracy cost reported in this paper, is consistent with the idea that SRs represent the top end of a face recognition continuum, rather than a qualitatively distinct ability. Bobak, Parris, Gregory, Bennetts, and Bate (2017) used eye-tracking to assess face processing in SRs, typical recognisers, and individuals with congenital prosopagnosia, and found that SRs spent a greater proportion of their time on the inner features of a face, particularly the nose region, when viewing social scenes. It could be this change in the time spent on the internal features of a face that is driving the SR advantage for other-race faces. A series of studies have shown that the ORE may result from failing to direct attention to those features, such as the nose region, of an other-race face that are likely to provide the most diagnostic information for accurate identity perception (Hills, Cooper, & Pake, 2013; Hills & Lewis, 2006; Hills & Pake, 2013). Therefore, it could be the case that SRs are naturally attuned to deploy their attention more efficiently and for longer to central regions of the face, leading to greater accuracy for both own- and other-race faces. This could explain the greater accuracy on the EFMT-short in the SR group relative to controls; and the smaller magnitude of the SR EFMT-short cost (3%) compared with typical recognisers (7% in Study 2; but n.s. 5% in Study 3).

An important consideration in terms of the applied potential of our findings relates to the fact that within SR research, group-level analyses (i.e., SRs vs. typical recognisers) can mask the fact that not all SRs, as categorised by scores on the CFMT+, always outperform typical individuals on other tests of face processing (Davis et al., 2016; Noyes et al., 2018). In both Study 1 and Study 2, we replicate the correlation between the CFMT+ and the GFMT reported by previous studies. This correlation suggests that CFMT+ SRs are also likely to perform above average in occupations where unfamiliar face matching is the critical task (i.e., passport control officer). However, these correlations are in the moderate range, and it is therefore the case that not all CFMT+ SRs are likely to be "super-face-matchers," and therefore, tests would need to be performed in conjunction with the CFMT+ before an individual could be considered as a suitable SR candidate for roles in which face matching is the critical task. Similarly, as outlined in Study 3, and as seen in Figure 3, not all CFMT+ SRs, or indeed higher performers on the GFMT, showed outstanding performance on the EFMT. Therefore, in the applied context, professions that are seeking to recruit SRs should employ a battery of tests to assess their suitability for the specific role (see Bate et al., 2018b; Ramon, Bobak, & White, 2019). It is not the case that selecting SRs on the basis of CFMT+ scores will ensure that each of these individuals will excel at unfamiliar face matching or indeed other-race unfamiliar face matching.

In conclusion, our findings of consistent associations in accuracy across face processing domains and ethnicity add weight to the notion that these processes may be served by the same underlying mechanism or *f*, a general face perception factor. SRs as a group, and to a large extent at the individual level, outperform typical recognisers on a test of other-race face matching but with an other-race accuracy cost remaining evident in this group. This SR advantage for other-race faces may be driven by more efficient attentional allocation to central regions of the face, particularly the nose, which are likely to provide greater diagnostic information for identity perception (Bobak, Parris, Gregory, Bennetts, & Bate, 2016; Hills et al., 2013; Hills & Lewis, 2006; Hills & Pake, 2013). Finally, police forces, border control agencies, and private organisations who seek to select and employ SRs must include other-race face tasks in their assessment battery to ensure, at the individual level, that the people they select also excel in verifying the identities of individuals from outside their own ethnic group. In doing so, this would provide an effective addition to counter measures that are designed to reduce fraud attacks at passport control and wrongful criminal convictions.

ACKNOWLEDGEMENTS

We would like to thank Dr. Stephen Butler, Ario Pellegrini Frugone, Lauren Gilmour, and Daisy Thomas for their help with data collection in Studies 2 and 3 and to the participants of the University of Greenwich Super Recogniser database for their ongoing support for this research.

CONFLICT OF INTEREST STATEMENT

The authors declare no competing interests.

ORCID

David J. Robertson https://orcid.org/0000-0002-8393-951X
Josh P. Davis https://orcid.org/0000-0003-0017-7159

REFERENCES

Alenezi, H. M., Bindemann, M., Fysh, M. C., & Johnston, R. A. (2015). Face matching in a long task: Enforced rest and desk-switching cannot maintain identification accuracy. *PeerJ*, *3*, e1184.

Balsdon, T., Summersby, S., Kemp, R. I., & White, D. (2018). Improving face identification with specialist teams. *Cognitive Research: Principles and Implications*, *3*(1), 25.

Bate, S., Bennetts, R., Hasshim, N., Portch, E., Murray, E., Burns, E., & Dudfield, G. (2018a). The limits of super recognition: An other-ethnicity effect in individuals with extraordinary face recognition skills. *Journal of Experimental Psychology: Human Perception and Performance*.

Bate, S., Frowd, C., Bennetts, R., Hasshim, N., Murray, E., Bobak, A. K., … Richards, S. (2018b). Applied screening tests for the detection of superior face recognition. *Cognitive Research: Principles and Implications*, *3*(1), 22.

Beattie, L., Walsh, D., McLaren, J., Biello, S. M., & White, D. (2016). Perceptual impairment in face identification with poor sleep. *Royal Society Open Science*, *3*(10), 160321.

Belanova, E., Davis, J. P., & Thompson, T. (2018). Cognitive and neural markers of super-recognisers' face processing superiority and enhanced cross-age effect. *Cortex*, *108*, 92–111. https://doi.org/10.1016/j.cortex.2018.07.008

Benton, A. L., Sivan, A. B., Hamsher, K., Varney, N. R., & Spreen, O. (1983). *Contribution to neuropsychological assessment*. NY: Oxford University Press.

Bindemann, M., Fysh, M., Cross, K., & Watts, R. (2016). Matching faces against the clock. *Perception*, *7*(5). 2041669516672219

Bindemann, M., & Sandford, A. (2011). Me, myself, and I: Different recognition rates for three photo-IDs of the same person. *Perception*, *40*(5), 625–627.

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.

Bobak, A. K., Hancock, P. J., & Bate, S. (2016). Super-recognisers in action: Evidence from face-matching and face memory tasks. *Applied Cognitive Psychology*, *30*(1), 81–91.

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*, 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. *The Quarterly Journal of Experimental Psychology*, *70*(2), 201–217.

Bruce, V., Henderson, Z., Greenwood, K., Hancock, P. J., Burton, A. M., & Miller, P. (1999). Verification of face identities from images captured on video. *Journal of Experimental Psychology: Applied*, *5*(4), 339.

Burton, A., Wilson, S., Cowan, M., & Bruce, V. (1999). Face recognition in poor-quality video: Evidence from security surveillance. *Psychological Science*, *10*, 243–248.

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

Burton, A. M., & Jenkins, R. (2011). *Unfamiliar face perception*. The Oxford handbook of face perception, 287-306.

Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow face matching test. *Behavior Research Methods*, *42*(1), 286–291.

Davis, J. P., Lander, K., Evans, R., & Jansari, A. (2016). Investigating predictors of superior face recognition ability in police super-recognisers. *Applied Cognitive Psychology*, *30*(6), 827–840.

Davis, J. P., Treml, F., Forrest, C., & Jansari, A. (2018). Identification from CCTV: Assessing police super-recognisers ability to spot faces in a crowd and susceptibility to change blindness. *Applied Cognitive Psychology*, *32*(3), 337–353.

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

Duchaine, B., & Nakayama, K. (2006). The Cambridge Face Memory Test: Results for neurologically intactindividuals and an investigation of its validity using inverted face stimuli and prosopagnosicparticipants. *Neuropsychologia*, *44*(4), 576–585.

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.

Fysh, M. C. (2018). Individual differences in the detection, matching and memory of faces. *Cognitive Research: Principles and Implications*, *3*(1), 20.

Fysh, M. C., & Bindemann, M. (2017). Effects of time pressure and time passage on face-matching accuracy. *Royal Society Open Science*, *4*(6), 170249. https://doi.org/10.1098/rsos.170249

Fysh, M. C., & Bindemann, M. (2018). The Kent face matching test. *British Journal of Psychology*, *109*(2), 219–231. https://doi.org/10.1111/bjop.12260

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

Hills, P. J., Cooper, R. E., & Pake, J. M. (2013). Removing the own-race bias in face recognition by attentional shift using fixation crosses to diagnostic features: An eye-tracking study. *Visual Cognition*, *21*(7), 876–898.

Hills, P. J., & Lewis, M. B. (2006). Reducing the own-race bias in face recognition by shifting attention. *The Quarterly Journal of Experimental Psychology*, *59*(6), 996–1002. https://doi.org/10.1080/17470210600654750

Ramon, M., Bobak, A. K., & White, D. (2019). Super-recognizers: From the lab to the world and backagain. *British Journal of Psychology*.

Hills, P. J., & Pake, J. M. (2013). Eye-tracking the own-race bias in face recognition: Revealing the perceptual and socio-cognitive mechanisms. *Cognition*, *129*(3), 586–597. https://doi.org/10.1016/j.cognition.2013.08.012

Jenkins, R., & Burton, A. M. (2011). Stable face representations. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *366*(1571), 1671–1683. https://doi.org/10.1098/rstb.2010.0379

Johnston, R. A., & Edmonds, A. J. (2009). Familiar and unfamiliar face recognition: A review. *Memory*, *17*(5), 577–596. https://doi.org/10.1080/09658210902976969

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Ge, L., & Pascalis, O. (2007). The other-race effect develops during infancy: Evidence of perceptual narrowing. *Psychological Science*, *18*(12), 1084–1089. https://doi.org/10.1111/j.1467-9280.2007.02029.x

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

Kokje, E., Bindemann, M., & Megreya, A. M. (2018). Cross-race correlations in the abilities to match unfamiliar faces. *Acta Psychologica*, *185*, 13–21. https://doi.org/10.1016/j.actpsy.2018.01.006

Marcon, J. L., Meissner, C. A., Frueh, M., Susa, K. J., & MacLin, O. H. (2010). Perceptual identification and the cross-race effect. *Visual Cognition*, *18*(5), 767–779.

McCaffery, J. M., Robertson, D. J., Young, A. W., & Burton, A. M. (2018). Individual differences in face identity processing. *Cognitive Research: Principles and Implications*, *3*(1), 21.

McKone, E., Stokes, S., Liu, J., Cohan, S., Fiorentini, C., Pidcock, M., … Pelleg, M. (2012). A robust method of measuring other-race and other-ethnicity effects: The Cambridge Face Memory Test format. *PLoS ONE*, *7*(10), e47956. https://doi.org/10.1371/journal.pone.0047956

Megreya, A. M., & Bindemann, M. (2015). Developmental improvement and age-related decline in unfamiliar face matching. *Perception*, *44*(1), 5–22. https://doi.org/10.1068/p7825

Megreya, A. M., Sandford, A., & Burton, A. M. (2013). Matching face images taken on the same day or months apart: The limitations of photo ID. *Applied Cognitive Psychology*, *27*(6), 700–706.

Megreya, A. M., White, D., & Burton, A. M. (2011). The other-race effect does not rely on memory: Evidence from a matching task. *The Quarterly Journal of Experimental Psychology*, *64*(8), 1473–1483. https://doi.org/10.1080/17470218.2011.575228

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*(1), 3.

Meissner, C. A., Susa, K. J., & Ross, A. B. (2013). Can I see your passport please? Perceptual discrimination of own-and other-race faces. *Visual Cognition*, *21*(9-10), 1287–1305.

Noyes, E., Hill, M. Q., & O'Toole, A. J. (2018). Face recognition ability does not predict person identification performance: Using individual data in the interpretation of group results. *Cognitive Research: Principles and Implications*, *3*(1), 23.

Noyes, E., & O'Toole, A. J. (2017). Face recognition assessments used in the study of super-recognisers. arXiv preprint arXiv:1705.04739.

O'Toole, A. J., Deffenbacher, K. A., Valentin, D., & Abdi, H. (1994). Structural aspects of face recognition and the other-race effect. *Memory & Cognition*, *22*(2), 208–224. https://doi.org/10.3758/bf03208892

Papesh, M. H., & Goldinger, S. D. (2014). Infrequent identity mismatches are frequently undetected. *Attention, Perception, & Psychophysics*, *76*(5), 1335–1349.

Phillips, P. J., Yates, A. N., Hu, Y., Hahn, C. A., Noyes, E., Jackson, K., … & Chen, J. C. (2018). Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms. *Proceedings of the National Academy of Sciences*, *115*(24), 6171–6176.

Robertson, D. J. (2018). Face recognition: Security contexts, super-recognizers, and sophisticated fraud. *The Journal of The United States Homeland Defence and Security Information Analysis Center (HDIAC)*, *5*(1), 6–10. Retrieved from. https://www.hdiac.org/system/files/HDIAC%20Journal_Volume%205_Issue_1_WEBSITE_01.pdf

Robertson, D. J., & Burton, M. (2016). Unfamiliar face recognition: Security, surveillance and smartphones. *The Journal of The United States Homeland Defence and Security Information Analysis Center (HDIAC)*, *3*(1), 14–21. Retrieved from. https://www.hdiac.org/wp-content/uploads/2018/04/Journal-Volume-3-Issue-1_B_Unfamiliar-Face-Recognition-1.pdf

Robertson, D. J., Mungall, A., Watson, D. G., Wade, K. A., Nightingale, S. J., & Butler, S. (2018). Detecting morphed passport photos: A training and individual differences approach. *Cognitive research: principles and implications*, *3*(1), 27.

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., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: People with extraordinary facerecognition ability. *Psychonomic bulletin & review*, *16*(2), 252–257.

Spearman, C. (1927). *The abilities of man*. Oxford: Macmillan.

Towler, A., Kemp, R. I., Burton, A. M., Dunn, J. D., Wayne, T., Moreton, R., & White, D. (2019). Do professional facial image comparison training courses work? *PLoS ONE*, *14*(2), e0211037.

Tree, J. J., Horry, R., Riley, H., & Wilmer, J. B. (2017). Are portrait artists superior face recognizers? Limited impact of adult experience on face recognition ability. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(4), 667–676. https://doi.org/10.1037/xhp0000328

UK Census (2011). Office For National Statistics, Retrieved from https://www.ons.gov.uk/census/2011census

Verhallen, R. J., Bosten, J. M., Goodbourn, P. T., Lawrance-Owen, A. J., Bargary, G., & Mollon, J. D. (2017). General and specific factors in the processing of faces. *Vision Research*, *141*, 217–227. https://doi.org/10.1016/j.visres.2016.12.014

Walker, P. M., & Tanaka, J. W. (2003). An encoding advantage for own-race versus other-race faces. *Perception*, *32*(9), 1117–1125. https://doi.org/10.1068/p5098

White, D., Dunn, J. D., Schmid, A. C., & Kemp, R. I. (2015). Error rates in users of automatic face recognition software. *PLoS ONE*, *10*(10), e0139827. https://doi.org/10.1371/journal.pone.0139827

White, D., Kemp, R. I., Jenkins, R., & Burton, A. M. (2014). Feedback training for facial image comparison. *Psychonomic Bulletin & Review*, *21*(1), 100–106.

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., Phillips, P. J., Hahn, C. A., Hill, M., & O'Toole, A. J. (2015). Perceptual expertise in forensic facial image comparison. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1814), 20151292.

Wilhelm, O., Herzmann, G., Kunina, O., Danthiir, V., Schacht, A., & Sommer, W. (2010). Individual differences in perceiving and recognizing faces—One element of social cognition. *Journal of Personality and Social Psychology*, *99*(3), 530–548. https://doi.org/10.1037/a0019972

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*(11), 5238–5241.

Yovel, G., Wilmer, J. B., & Duchaine, B. (2014). What can individual differences reveal about face processing? *Frontiers in Human Neuroscience*, *8*, 562.

Zhao, M., & Bülthoff, I. (2013). The other-race effect in face recognition is sensitive to face format at encoding. *Visual Cognition*, *21*(6), 722–725.

How to cite this article: Robertson DJ, Black J, Chamberlain B, Megreya AM, Davis JP. Super-Recognisers show an advantage for other race face identification. *Appl Cognit Psychol*. 2019; 1–12. https://doi.org/10.1002/acp.3608