Investigating predictors of superior face recognition ability in police super-recognisers

Josh P Davis [1], Karen Lander [2], Ray Evans [2], and Ashok Jansari [3]

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

[2] School of Psychological Sciences, University of Manchester, UK

[3] Department of Psychology, Goldsmiths, University of London, UK

Correspondence to:

Dr Josh P Davis

Department of Psychology, Social Work and Counselling

University of Greenwich

Avery Hill

London, SE9 2UG, UK

j.p.davis@greenwich.ac.uk

Word count (excluding titles, abstract, tables, figures, and references): 7,705

Acknowledgments: We would like to thank Mick Neville and Paul Smith from the Metropolitan Police Service for assistance with police recruitment and information as to police procedures; as well as Kelty Battenti, Jennie Bishop, Charlotte Byfield-Wells, Nezire Cornish, Catherine Culbert, Ima Fagerbakke, Beckie Hogan, Charlotte Ide, Desislava Ignatova, Ionela Jurj, Andreea Maigut, Jade Murray, Christopher Pluta, Sarah Poland, Lauren Smith, Have Sokoli, Jay Tamplin-Wilson, Nadine Wanke, and Sarah Wilson for help with recruitment and testing. We would also like to thank two anonymous reviewers for their very helpful comments on an earlier version of this manuscript.

Parts of this research were presented in *The Psychologist* (Davis, Lander, and Jansari, 2013), the *IEEE International Conference on Advanced Video and Signal-based Surveillance (AVSS) Karlsruhe Institute of Technology, Germany*, 25-28 August, 2015; the *Pre-SARMAC Face Day Workshop, University of Victoria*, *British Columbia, Canada*; and the 5*th* *International Conference on Imaging for Crime Prevention and Detection (ICDP 2013), Kingston University, UK*, December 2013.

Abstract

There are large individual differences in the ability to recognise faces. *Super-recognisers* are exceptionally good at face memory tasks. In London, a small specialist pool of police officers (also labelled ‘super-recognisers’ by the Metropolitan Police Service) annually makes 1,000’s of suspect identifications from CCTV footage. Some suspects are disguised, have not been encountered recently, or are depicted in poor quality images. Across tests measuring familiar face recognition, unfamiliar face memory and unfamiliar face matching, the accuracy of members of this specialist police pool was approximately equal to a group of non-police super-recognisers. Both groups were more accurate than matched control members of the public. No reliable relationships were found between the face processing tests and object recognition. Within each group however, there were large performance variations across tests, and this research has implications for the deployment of police worldwide in operations requiring officers with superior face processing ability.

There are large, individual differences in face recognition ability. These mainly inherited differences (Wilmer et al., 2010; Shakeshaft & Plomin, 2015), correlate with eyewitness identification accuracy (e.g., Bindemann, Brown, Koyas, & Russ, 2012), simultaneous face matching ability (e.g., Megreya & Burton, 2006), personality (Lander & Poyarekar, 2015; Li, Tian, Fang, Xu, & Liu, 2010), and propensity to process faces holistically (e.g., DeGutis, Wilmer, Mercado, & Cohan, 2013; Wang, Li, Fang, Tian, & Liu, 2012; although see Richler, Floyd, & Gauthier, 2015 for contrasting findings). Research examining extreme ability has mainly focussed on face blindness (*prosopagnosia),* particularly when coexisting with normal-range visual acuity and object recognition ability. *Acquired prosopagnosia* is a consequence of brain damage (e.g., Rossion et al., 2003; Jansari et al., 2015), whereas *developmental prosopagnosia,* often identified in childhood, is linked to no known damage (e.g., Duchaine, Germine, & Nakayama, 2007; Wilmer et al., 2010).

Some people however possess exceptionally good face processing ability (Bobak, Bennetts, Parris, Jansari, & Bate, 2016a; Bobak, Hancock, & Bate, 2016b; Bobak, Parris, Gregory, Bennetts, & Bate, 2016c; Robertson, Noyes, Dowsett, Jenkins, & Burton, 2016; Russell, Chatterjee, & Nakayama, 2012; Russell, Duchaine, & Nakayama, 2009; White, Dunn, Schmid, & Kemp, 2015a; White, Phillips, Hahn, Hill, & O’Toole, 2015b). Russell et al. (2009) found that four self-identifying *super-recognisers* performed far better than controls on the enhanced *Cambridge Face Memory Test (CFMT),* the *Cambridge Face Perception Test,* and a *Before They Were Famous* face test. The authors suggest that super-recognisers “are about as good (at face recognition) as many developmental prosopagnosics are bad” (p. 256), and as there appears to be a continuous spectrum of ability within the population, super-recognition and developmental prosopagnosia are convenient labels for individuals represented in the two tails of this spectrum.

Excellent face recognition ability has law enforcement implications. A majority of suspect identifications (‘*idents*’)[1] from CCTV images in London are made by a small pool of Metropolitan Police Service (MPS) officers and staff (*police identifiers;* [2] see Davis, Lander, & Jansari, 2013 for a description). By 2015, approximately 140 out of 48,000 MPS officers and civilian staff (e.g., cell detention officers) were members. Most were identified after making multiple idents from the MPS *Caught on Camera* ‘wanted’ website, [3] and their viewing of images of highly serious London-wide crimes, as well as less serious local crimes is now prioritised. Idents are the first step in a police investigation, and although most CCTV-identified suspects confess in interview when confronted with images (> 70%), not all cases proceed to court – often from lack of alternative evidence, meaning that guilt cannot always be established. Nevertheless, between April 2013 and December 2015, the total idents made in the MPS jurisdiction was approximately 13,000 – police identifiers made 9,000, substantially increasing sentencing rates in cases involving CCTV evidence in London.

Most MPS police identifiers are community-based front line officers, and their idents are mainly driven by knowledge of local familiar suspects, although some are disguised, have not been encountered recently, or are depicted in poor quality images. Functional theories postulate qualitatively different processing pathways for familiar and unfamiliar faces (e.g., Bruce & Young, 1986; for reviews see Burton, 2013; Johnston & Edmonds, 2009). With familiar faces, viewpoint- and expression-independent stored representations govern recognition - accuracy is high even with poor-quality images (e.g., Bruce, Henderson, Newman, & Burton, 2001; Burton, Wilson, Cowan, & Bruce, 1999). To reduce ceiling effects, familiar face recognition research normally employs impoverished images. Recognition performance is higher to moving images in these circumstances (e.g. Knight & Johnston, 1997), and is of applied interest as CCTV images are often low quality. This may simply be a consequence of additional information (there are more available frames), although some authors have suggested that movement may allow for individuating ‘motion signature’ extraction (e.g., Lander & Bruce, 2000; Lander, Bruce, & Hill, 2001; Lander & Chuang, 2005).

Idents are also sometimes of suspects the police identifier has never encountered in person, but recognises from previously viewed crime scene imagery. In contrast to familiar face recognition which is dominated by internal feature processing (e.g., eyes, mouth), unfamiliar face processing is driven by the external features (e.g., hairstyle, face shape; Bruce et al., 1999; Ellis, Shepherd, & Davies, 1979), as well as expression- and viewpoint-specific *pictorial* codes, making it far more prone to error (Burton, 2013; Jenkins, White, van Montfort, & Burton, 2011; Johnston & Edmonds, 2009). Hairstyle can be an unreliable identification cue, and environmental (e.g., lighting, viewpoint), appearance (e.g., expression, hairstyle change) or camera (e.g. lens type) variations can be interpreted as differences in facial structure. These variations may make images of two different people appear highly similar, or two images of the same person appear very different. Indeed, simultaneous unfamiliar face matching performance can be unreliable even with unlimited viewing time, high quality images, and targets present in person (e.g., Bruce et al., 1999; Davis & Valentine, 2009; Megreya & Burton, 2006; see Davis & Valentine, 2015 for a review).

Despite the problems associated with unfamiliar face processing, some police identifiers have been assigned to operations requiring the type of excellent unfamiliar face processing skills associated with super-recognisers. These include memorising photographs to locate suspects at crowded events; matching images of suspects across footage taken of different crimes possessing similar characteristics; and reviewing footage to locate persons of interest. A few police identifiers have been attached to a *Proactive Super-Recogniser Unit* in order to perform these tasks full-time. Robertson et al. (2016) describe four members of this unit as possessing unfamiliar and familiar face processing skills “which far exceed the general population” (p. 5). However, this unit forms a minority of the police identifier pool.

The current research therefore employed four face processing tests to examine whether the performance of the mainly community-based front line pool of MPS police identifiers, matched a group of super-recognisers, meeting the inclusion criteria for this ability employed in previous research (e.g., Bobak et al., 2016a; Russell et al., 2009). Demographically-matched controls provided performance baselines. The primary aim was to determine whether the police identifier’s high ident rates were indicative of super-recognition ability. A further aim was to develop a greater understanding of the skill sets associated with both police identifiers and super-recognisers, in order to determine whether any characteristics in common could explain the police identifier’s successes. For this reason, the tests in the current research were based on factors that might influence ident accuracy from sometimes impoverished CCTV images. These included the recognition of familiar faces, some not seen for many years, from degraded moving and static images; distinguishing briefly learnt unfamiliar faces from arrays of physically similar distracters; extrapolating identity from one facial viewpoint to a second; an inclination to focus on the more reliable and stable internal facial features when learning new faces (as opposed to peripheral details such as hairstyle); confidence; and simultaneous unfamiliar face matching.

In addition, theories based mainly on prosopagnosia research suggest that faces may be ‘special’ in that either due to adaptation (e.g., Duchaine & Nakayama, 2006) or expertise (e.g., Gauthier, Skudlarski, Gore, & Anderson, 2000); they are processed by dedicated domain-specific cortical pathways. Recent evidence also suggests that super-recognisers’ superior skills may also be face-specific (Bobak *et al*., 2016a), and an *Object Memory Test* examined whether super-recognition ability extended to an alternative visual memory task.

The super-recognisers by definition were as a group hypothesised to be more accurate at the four face processing tests than the controls. Staff in roles in which face memory ability is important, often perform no better at face processing tests than members of the public (e.g., passport officers: White, Kemp, Jenkins, Matheson, & Burton, 2014; police officers: Burton et al., 1999). However, recent research has consistently shown that there are individual differences in ability within these groups (White et al., 2014; White et al., 2015a; 2015b; Wilkinson & Evans, 2009), including the MPS (Robertson et al., 2016). Therefore due to having displayed exceptional performance in an operational context, the police identifiers as a group were also expected to be more accurate at the four face processing tests than the controls. Nevertheless, as some super-recognisers' performances vary across different face processing tests (Bobak et al., 2016a; 2016b), and their high rates of idents may be acquired in diverse circumstances, in advance it was unclear whether this advantage would be found with *all* police identifiers. Indeed, although simultaneous face matching and face memory performance normally correlates (e.g., Lander & Poyarekar, 2015; Megreya & Burton, 2006), some prosopagnosics are able to match faces within the normal range (Dalrymple**,** Garrido, & Duchaine, 2014), and not all super-recognisers are excellent at unfamiliar face matching (Bobak et al., 2016b). This suggests that face memory and matching may, in some cases, draw on different mechanisms (see also White et al., 2015b).

For this reason, as well as group level analyses, individual analyses were conducted in the manner of neuropsychological research by comparing the performance of each police identifier and super-recogniser on each test against the controls. This allowed us to measure test performance consistency and to generate an estimate of the proportion of the general population each would be expected to exceed (e.g., see Bobak et al., 2016a; Crawford, Garthwaite, & Porter, 2010). Finally, a correlational component examined the relationships between the test performances of participants, with an expectation that outcomes on the four face processing tests, but not necessarily the object memory task, would positively correlate.

Method

Design

This study received University of Greenwich Research Ethics Committee approval. It primarily employed an independent-measures design comparing the performance of super-recognisers, police identifiers, and controls on five tests, conducted in the following order - *Unfamiliar Face Memory Array Test* (Bruce et al., 1999); *Famous Face Recognition Test* (Lander *et al.*, 2001); *Object (Flowers) Memory Test*; *Old/New Unfamiliar Face Memory Test*; *Glasgow Face Matching Test* (Burton, White, & McNeill, 2010). A correlational design examined the relationships between test performances.

Participants

Super-recognisers (*n* = 10; 40% female; aged 24-44 years, M = 34.4 (SD = 7.3); 20% left-handed (LH); 50% white-Caucasian, 20% Indian, 30% other ethnicity) had previously achieved scores in the top 2% on the extended CFMT (Russell et al., 2009; Range: 93.1%-99.0%; M = 94.3%, SD = 1.9), based on results from more than 700 visitors (M = 68.7%, SD = 13.8) to a public engagement with science initiative held at London's Science Museum. [4] Their scores exceed the criteria (88.2%) for super-recognition employed in previous research (e.g., Bobak et al., 2016a; Russell et al., 2009).

Police identifiers (*n* = 36; 19.4% female; aged 24-58 years, M = 38.1 (SD = 9.1); 8.6% LH; 87.5% white-Caucasian, 9.4% black, 3.1% other ethnicity) were invited to participate by senior MPS officers, and relieved from normal duties. They were members of an MPS pool of volunteer ('super-recogniser') officers and staff informally established in 2011-2012 (see Davis et al., 2013). Many of those tested were 'founder' members of the pool, although the pool expanded from nearly 30 to over 100 during the data collection period. The inclusion criteria for the current research was a minimum of 15 idents within a 12 month period from 2011 to 2014 (ident rates varied – the most successful police identifier made more than 180 idents in a single year). Five additional police identifiers meeting inclusion criteria declined to participate. The remaining members were either not given time out of their duties, or had not achieved the minimum inclusion criteria at the time.

Controls (*n* = 143; 24.5% female; aged 19-61 years, M = 34.4 (SD = 10.2); 9.7% LH; 92.3% white-Caucasian, 4.9% black, 3.8% other ethnicity), were non-student members of the public recruited by research assistants, via posters, and adverts on social media. These adverts described the study as measuring face recognition ability and that it would take up to two hours. No compensation was paid to controls or super-recognisers.

There were no between-group gender, $\chi^2(2, 189) = 1.89$, $p > .1$; age, $F(2, 177) = 1.72$, $p > .1$; or handedness differences, $\chi^2(2, 179) = 1.20$, $p > .1$. However, ethnicity differed, $\chi^2(2, 185) = 17.47$, $p < .001$, the proportion of white-Caucasian controls and police identifiers was approximately equal ($p > .05$), but was higher than that of the super-recognisers ($p < .05$).

Materials and Procedure

Famous Face Recognition Test (Lander *et al*., 2001): This test consisted of two counterbalanced sets of 15 male and 15 female celebrity faces, and 5 male and 5 female unknown faces taken approximately 12-years previously. The more recent media profile of the celebrities varied substantially, although no data of this were collected. Images were degraded by thresholding. Each face was shown for 5-sec, half *moving* (20); half *static* (20). If moving in Set A, a face was static in Set B and vice versa (80 trials). Participants provided a name or semantic information or stated they did not recognise the face. Famous face responses were categorised as *hits*: participants provided correct names or individuating information (e.g., for Angela Lansbury, - “a writer in the TV drama – *Murder She Wrote”* was accredited with a hit); *misidentifications*: incorrect names/information; or *misses:* failures to recognise the face, or non-individuating responses (e.g. “actress”). With unknown faces, responses were *correct rejections*: correctly identified as unfamiliar; or *false alarms:* incorrect names/identities. Participants were subsequently presented with a list of the celebrity names and asked whether they should have recognised them. *Conditionalised Naming Rates* (CNR) were calculated by excluding response data to a celebrity from analyses if the participant claimed they would not have recognised that face.

Unfamiliar Face Memory Array Test: The test stimuli were originally designed for a face matching study (Bruce et al., 1999). In this *memory* design, across four counterbalanced versions, participants completed 40 trials in which a single colour white-Caucasian male image was displayed for 5-sec from a frontal perspective (20 faces) or a 30 degree angle (20 faces). Each was almost immediately followed by an array of 10 randomly arranged colour frontal same-day different-camera faces each marked with a number (1-10). Participants, warned in advance that half the trials were target-absent, attempted to identify the target by supplying an array number, or if not present, to reject the array. Target-present outcomes were either *hits* (correct array number), *misidentifications* (incorrect array number), or *misses* (‘not present’ response). Target-absent outcomes were *correct rejections* (‘not present’ response) or *false alarms* (incorrect array number). Decision confidence ratings were collected immediately after each trial (1: low – 5: high). There were no test phase time limits. Based on 240 participants, mean target-present hit rates, and mean target-absent correct rejection rates were both 70% in Bruce et al.'s (1999) first face matching experiment.

Old/New Unfamiliar Face Memory Test: This test was designed to measure the propensity of participants to focus on the internal regions when learning new faces, as opposed to hairstyle and other peripheral information. In the learning phases of two counterbalanced versions, 20 randomly ordered sequentially presented colour photos depicted unfamiliar white-Caucasian males (10 faces) and females (10 faces) from the waist up for 5-sec each. Context was provided (e.g., room, clothing), although participants were instructed to remember the faces. Almost immediately, participants viewed 40 sequentially presented faces with external facial features and background cues obscured and judged whether each was old (*hit*: 20 faces) or new (*correct rejection:* 20 faces). Participants were not forewarned that only the internal features would be shown. Different same-day photos of the same person were used in learning and test phases. There were no test phase time limits.

Glasgow Face Matching Test (short version: Burton, White, & McNeill, 2010): This self-paced standardised test consists of 40 pairs of simultaneously presented white-Caucasian faces in greyscale. Participants respond 'same' (*hit*: 20 faces) or 'different' (*correct rejection*: 20 faces). There were no response time limits. Based on 194 participants, Burton et al. found the normalised mean hit rates on this test to be 79.8%, correct rejections rates - 82.5%.

Object Memory Test: Participants viewed a series of 20 colour flower photographs for 5-sec each in the learning phase. In the test phase, they viewed 40 photographs, and judged whether each was 'old' (*hit*: 20 flowers), or 'new' (*correct rejection*: 20 flowers). The same 'old' photographs were used in both phases. There were no test phase time limits.

Participants could take rests between tests so that total time varied from 75 to 120 minutes. No performance feedback was provided.

Data of hits and correct rejections allowed calculation of signal detection theory (see Green & Swets, 1966; Macmillan & Creelman, 1991) measures of sensitivity (d'), and bias (C).[5] In memory, or matching tests, high positive values of d' indicate good discrimination of 'old' and 'new,' or 'same' and 'different' stimuli respectively. Independent of sensitivity, criterion (C) or response bias measures the tendency to respond 'old' or 'new' under conditions of uncertainty. Negative values of C are indicative of a conservative response bias or a tendency to respond 'new' in memory tests - 'different' in matching tests; positive values indicate liberal response biases or a tendency to respond 'old' or 'same' respectively. To calculate sensitivity and bias for the *Famous Face Recognition Test* and the *Unfamiliar Face Memory Array Test*, misidentification responses were pooled with misses (see Bobak et al., 2016b; Mickes, Moreland, Clark, & Wixted, 2014).

Results

The mean proportions of hits, and correct rejections, as well as mean sensitivity (d') and response bias (C) for each test were calculated, and independent-measures ANOVAs compared between-group outcomes. Games-Howell post hoc tests were employed throughout ($\alpha = .05$). Modified t-tests for single cases (Crawford et al., 2010), compared the d' scores of police identifiers and super-recognisers against the control mean.

Famous Face Recognition Test

After completing the face naming phase of this test (Sets A and B), participants read a list of the 30 celebrity names and reported which celebrity faces they would never have recognised. These data were used to calculate *Conditionalised Naming Rates* (CNR) by excluding data the unrecognised names. A between-groups ANOVA on these data was significant, $F(2, 186)$

= 7.53, $p = .001$, $\eta^2 = .075$. Super-recognisers ($M = 2.2$) and police identifiers ($M = 3.5$) did not differ ($p > .05$), but both groups claimed they would never have recognised fewer faces than controls ($M = 6.5$, $p < .05$). This may reflect genuine differences in celebrity knowledge. However, controls and police identifiers were drawn from a similar background; and retrospective responses of this type may be biased by previous face naming performances. Therefore a series of 3 (group) x 2 (presentation: moving, static) ANOVAs were conducted on both the conditionalised naming rate $_{\text{CNR}}$ data *and* the *unconditionalised* $_{\text{UN}}$ data that were not adjusted for name recognition. Figure 1 displays the CNR data outcomes.

Anticipating the between-groups results below, super-recognisers and police identifiers were more accurate than controls at recognising degraded famous faces, and correctly rejecting unfamiliar faces. All groups were better at recognising moving faces than static. However, an interaction on the unconditionalised $_{\text{UN}}$ hit rate results only, revealed that super-recognisers and police identifiers derived a greater advantage from movement than controls in the recognition of famous faces.

Figure 1 about here

Hits: The group main effect was significant, F_{CNR} (2, 186) = 9.37, $p < .001$, $\eta^2 = .092$; F_{UN} (2, 186) = 12.21, $p < .001$, $\eta^2 = .116$. Super-recogniser ($M_{\text{UN}} = 0.68$) and police identifier hit rates ($M_{\text{UN}} = 0.64$) did not differ ($p > .05$), but were higher than controls ($M_{\text{UN}} = 0.49$, $p <$ .05). The presentation main effect was significant, F_{CNR}(1, 186) = 83.18, $p < .001$, $\eta^2 = .309$; F_{UN}(1, 186) = 82.07, $p < .001$, $\eta^2 = .306$. Moving image hit rates were higher ($M_{\text{UN}} = 0.64$) than static ($M_{UN} = 0.56$). The interaction was significant for the unconditionalised analyses only, F_{CNR}(2, 186) = 2.54, $p = .081$, $\eta^2 = .027$; F_{UN}(2, 186) = 4.52, $p = .012$, $\eta^2 = .046$. The movement-advantage was greater for super-recognisers ($M_{\text{UN}\ Moving} = 0.73$, $_{\text{UN}\ Static} = 0.62$, $_{\text{UN}}$

difference = 0.11) and police identifiers ($M_{UN\ Moving}$ = 0.68, $_{UN\ Still}$ = 0.59, $_{UN\ diff}$ = 0.09) than controls ($M_{UN\ Moving}$ = 0.52, $_{UN\ Static}$ = 0.46, $_{UN\ diff}$ = 0.06; $p < .05$).[6]

Correct rejections: The group main effect was significant, $F(2, 186) = 3.09$, $p = .048$, $\eta^2 = .032$. Super-recognisers ($M = 0.83$) and police identifiers did not differ ($M = 0.78$, $p > .05$). Both made more correct rejections than controls ($M = 0.68$, $p < .05$). [7]

Sensitivity: The group main effect was significant, $F_{CNR}(2, 186) = 10.89$, $p < .001$, $\eta^2 = .105$; $F_{UN}(2, 186) = 12.68$, $p < .001$, $\eta^2 = .120$; super-recogniser d/ (M_{UN} = 1.60) did not differ from police identifiers (M_{UN} =1.25, $p > .05$), but control d/ was lower (M_{UN} = 0.52, $p < .05$). The presentation main effect was significant, $F_{CNR}(1, 186) = 25.17$, $p < .001$, $\eta^2 = .119$; $F_{UN}(1, 186) = 19.64$, $p < .001$, $\eta^2 = .095$; d/ was higher to moving (M_{UN} =1.28) than static images (M_{UN} = 0.97).

Response bias: The presentation main effect was significant, $F_{CNR}(1, 186) = 13.85$, $p = .001$, $\eta^2 = .069$; $F_{UN}(1, 186) = 9.99$, $p = .002$, $\eta^2 = .051$, static image responses were more liberal (M_{UN} = 0.34) than moving (M_{UN} = 0.23).

Figure 2 about here

Individual level analyses: Modified t-tests for single cases (Crawford et al., 2010), compared the d'_{CNR} scores of police identifiers and super-recognisers against the control mean. Two super-recognisers [20.0%; d'_{CNR} range = 2.45-2.74; $t(143) \geq 1.72$, $p < .05$, one-tailed, z = 1.72-2.03] and three police identifiers [8.3%; d'_{CNR} range = 2.44-2.91; $t(143) \geq 1.71$, $p < .05$, one-tailed, z = 1.71-2.22] scored significantly higher. Figure 2 depicts the 95% confidence intervals of the proportion of the general population each police identifier and super-recogniser would be expected to exceed on this test.

Unfamiliar Face Memory Array Test

Five 3 (group) x 2 (viewpoint: full face (FF), three-quarters (3/4)) mixed ANOVAs compared performance outcomes (see Figure 2) and confidence. To anticipate the results below, in comparison to controls, and regardless of facial viewpoint, super-recognisers and police identifiers were more accurate and confident at recognising briefly seen faces from subsequent arrays of 10, and at rejecting arrays not containing that face. For all participants, accuracy was positively related to confidence.

Figure 3 about here

Hits: The group main effect was significant, $F(2, 186) = 9.11$, $p < .001$, $\eta^2 = .089$. Super-recogniser ($M = 0.78$) and police identifier hit rates ($M = 0.72$) did not differ ($p > .05$), but higher than controls ($M = 0.63$, $p < .01$). The viewpoint effect was significant, $F(1, 186) = 11.29$, $p < .001$, $\eta^2 = .057$. FF hit rates were higher ($M = 0.76$) than 3/4 face ($M = 0.66$).

Correct rejections: The group main effect was significant, $F(2, 186) = 8.18$, $p < .001$, $\eta^2 = .081$. Super-recogniser ($M = 0.76$) and police identifier correct rejection rates ($M = 0.73$) did not differ ($p > .05$), but were higher than controls ($M = 0.60$, $p < .05$). The viewpoint effect was marginally significant, $F(1, 186) = 3.87$, $p = .051$, $\eta^2 = .020$. FF correct rejection rates were slightly higher ($M = 0.72$) than 3/4 faces ($M = 0.67$).

Sensitivity: The group main effect was significant, $F(2, 186) = 13.26$, $p < .001$, $\eta^2 = .125$; super-recogniser ($M = 1.52$) and police identifier $d^{/}$ ($M = 1.45$) did not differ ($p > .05$), but was higher than controls ($M = .72$, $p < .05$). The viewpoint effect was significant, $F(1, 186) = 17.88$, $p < .001$, $\eta^2 = .088$; FF $d^{/}$ was higher ($M = 1.53$) than 3/4 face ($M = 1.00$).

Response bias: No effects were significant ($p > .2$).

Confidence: The group main effect was significant, $F(2, 186) = 9.39$, $p < .001$, $\eta^2 = .092$. Super-recogniser ($M = 3.88$) and police identifier confidence ($M = 3.64$) did not differ ($p > .05$), but both were more confident than controls ($M = 3.29$, $p < .05$). The viewpoint effect was significant, $F(1, 186) = 23.53$, $p < .001$, $\eta^2 = .112$. Confidence was higher to FF ($M = 3.73$) than 3/4 faces ($M = 3.48$). A Pearson's correlation test also revealed a significant positive relationship, $r(189) = 0.43$, $p < .001$, between the mean sensitivity ($d^/$) scores of all participants and their mean decision confidence ratings across all trials.

Figure 4 about here

Individual level analyses: Modified t-tests comparing $d^/$ scores revealed that three super-recognisers [30.0%; $d^/$ range = 2.07-2.35; $t(143) \geq 1.67$, $p < .05$, one-tailed, z = 1.68-2.02] and seven police identifiers [19.4%; $d^/$ range = 2.12-3.24; $t(143) \geq 1.74$, $p < .05$, one-tailed, z = 1.74-3.09] scored significantly higher than the control mean ($n = 143$, $M = 0.68$, $SD = 0.83$) (see Figure 4 for the proportion of population expected to perform below each super-recogniser and police identifier (95% CI)).

Old/New Unfamiliar Face Memory Test

The mean outcomes are displayed in Figure 5. Anticipating the results, there were no group differences between super-recognisers and police identifiers, and although effect sizes were small in the comparison between these groups and controls on hits and correct rejections, when measuring sensitivity, controls were less accurate.

Figure 5 about here

Hits and correct rejections: The ANOVA on hits was significant, $F(2, 186) = 4.89$, $p = .009$, $\eta^2 = .050$. Super-recogniser hit rates were marginally higher than controls ($p = .073$). The correct rejection ANOVA was significant, $F(2, 186) = 7.10$, $p = .001$, $\eta^2 = .071$. Police identifier correct rejection rates were higher than controls ($p < .001$).

Sensitivity and response bias: The d' ANOVA was significant, $F(2, 186) = 15.25$, $p < .001$, $\eta^2 = .141$. There were no differences between super-recognisers and police identifiers ($p > .05$). Police identifier d' was significantly higher ($p < .05$); super-recogniser d' marginally higher ($p = .053$) than controls. There were no response bias effects, $F(2, 186) < 1$, $\eta^2 < .01$.

Figure 6 about here

Individual level analyses: Modified t-tests on d' scores revealed that five super-recognisers [50.0%; d' range = 2.43-2.87; $t(143) \geq 2.01$, $p < .05$, one-tailed, $z = 2.02$-2.71] and ten police identifiers [27.8%; d' range = 2.24-2.87; $t(143) \geq 1.71$, $p < .05$, one-tailed, $z = 1.72$-2.71] scored significantly above the control mean ($n = 143$, $M = 1.16$, $SD = 0.63$) (see Figure 6 for the proportion of the population (95% CI) expected to perform below each super-recogniser and police identifier).

Glasgow Face Matching Test

The mean outcomes are displayed in Figure 7. To summarise the results below, even though the controls were more accurate than the normative published *Glasgow Face Matching Test* data, super-recogniser and police identifier sensitivity was reliably higher.

Figure 7 about here

Three one sample t-tests comparing revealed that the hit rates of super-recognisers (*M* = 0.94), $t(9) = 3.68$, $p = .005$; police identifiers (M = 0.96), $t(35) = 14.03$, $p < .001$; *and* controls (M = 0.89), $t(142) = 8.58$, $p < .001$, were significantly higher than the published normative test data (M = 0.80; Burton *et al*., 2010). Similar effects were found with correct rejections (normative data M = 0.83), super-recognisers (M = 0.96), $t(9) = 7.47$, $p < .001$; police identifiers (*M* = 0.93), $t(35) = 10.88$, $p < .001$; controls (*M* = 0.89), $t(142) = 6.34$, $p < .001$, indicating all groups were more accurate than that expected by the general population.

The ANOVAs on hits, $F(2, 186) = 4.78$, $p = .009$, $\eta^2 = .049$; correct rejections, $F(2, 186) = 3.36$, $p = .037$, $\eta^2 = .035$, and d', $F(2, 184) = 8.64$, $p < .001$, $\eta^2 = .085$ were significant. Super-recognisers and police identifiers did not differ but were more accurate than controls on all outcomes ($p < .05$). There were no response bias effects, $F(2, 186) < 1$, $\eta^2 < .01$.

Figure 8 about here

Individual level analyses: Three super-recognisers and four police identifiers scored 100%. However, modified t-tests revealed that their d'scores did not significantly differ [d' = 3.92; $t(143) = 1.45$, $p > .05$ one-tailed, $z = 1.46$] from the control mean ($n = 143$, $M = 2.72$, $SD = 0.82$). (See Figure 8 for the proportion of population (95% CI) expected to perform below each super-recogniser and police identifier).

Object Memory Test

The mean outcomes are displayed in Figure 9. In anticipation, super-recogniser hit rates were higher than police identifiers and controls; partly due to a liberal response bias by the super-recognisers. However, there were no group differences in correct rejections or sensitivity.

Figure 9 about here

Hits: This ANOVA was significant, $F(2, 186) = 4.76$, $p = .010$, $\eta^2 = .049$. Super-recogniser hit rates were higher than police identifiers and controls ($p < .05$), who did not differ ($p > .05$).

Correct rejections and sensitivity: The ANOVAs on correct rejections, $F(2, 186) < 1$, and d' were not significant $F(2, 184) = 2.67$, $p = .072$, $\eta^2 = .028$.

Response bias: This ANOVA was significant, $F(2, 186) = 3.68$, $p = .027$, $\eta^2 = .038$; Super-recogniser responses were more liberal than police identifiers and controls.

Figure 10 about here

Individual level analyses: Modified t-tests comparing d' scores revealed that two super-recognisers [20.0%; $d' = 3.21$; $t(143) = 1.75$, $p < .05$, one-tailed, $z = 1.76$] and one police identifier [2.8%; $d' = 3.24$; $t(143) = 1.79$, $p < .05$, one-tailed, $z = 1.80$] scored significantly above the control mean ($n = 143$, $M = 1.88$, $SD = 0.75$) (see Figure 10 for the proportion of population (95% CI) expected to perform below each group member).

Overall mean performance on the four face processing tests

The mean scores for each participant on each outcome (hits, correct rejections, sensitivity ($d^{/}$), response bias (C)) across the four face processing tests were calculated ((*Famous Face Recognition Test* + *Unfamiliar Face Memory Array Test* + *Old/New Unfamiliar Face Memory Test* + *Glasgow Face Matching Test*)/4), and a final series of one-way ANOVAs compared the between-group mean scores (see Figure 11).[8]

As previously, super-recognisers and police identifiers did not differ on any measure ($p > .05$), but their hit rates, F_{CNR} (2, 186) = 21.33, $p < .001$, $\eta^2 = .187$; F_{UN} (2, 186) = 23.68, $p < .001$, $\eta^2 = .203$, correct rejection rates, $F(2, 186) = 14.91$, $p < .001$, $\eta^2 = .138$ and sensitivity, F_{CNR} (2, 186) = 32.91, $p < .001$, $\eta^2 = .261$; F_{UN} (2, 186) = 33.58, $p < .001$, $\eta^2 =$ .265 were higher than controls ($p < .05$). No response bias effects were significant ($p > .2$).

Figure 11 about here

Figure 12 about here

Individual level analyses: Modified t-tests comparing $d^{/}_{CNR}$ scores revealed that nine super-recognisers (90.0%) and 32 police identifiers (88.9%) scored above the control mean (n = 143, M = 1.35, SD = 0.48). Five super-recognisers [50.0%; $d^{/}_{CNR}$ = 2.32-2.58; $t(143) \geq 2.03$, $p < .05$, one-tailed, z = 2.04-2.58] and eleven police identifiers [30.6%; $d^{/}_{CNR}$ = 2.16-3.10; $t(143) \geq 1.69$, $p < .05$, one-tailed, z = 1.70-3.67] scored significantly higher [9] (see Figure 12 for the population proportion (95% CI) expected to perform below each group member).

Relationships between performances on each test

Data for static and moving images from the *Famous Face Recognition Test* $_{CNR}$; and full-face and 3/4 view faces on the *Unfamiliar Face Array Memory Test* were pooled. Pearson's correlation tests examined the relationship between mean outcomes on the tests (Table 1).

Table 1 about here

On the four face processing tests, only the correlations between hits, correct rejections, and $d^{/}$ on the *Famous Face Recognition Test* and the *Unfamiliar Face Memory Array Test*, and separately the *Glasgow Face Matching Test*, and the *Old/New Face Memory Test* were significant suggesting that these tests were partly measuring different constructs. However, there were no significant correlations between the *Object Memory Test*, and any of the face processing tests except the *Old-New Unfamiliar Face Memory Test*, which may be due to identical test designs.

Discussion

This research found that London MPS police identifiers who have made multiple identifications (idents) of suspects from CCTV were, as a group, more accurate than controls drawn from the general public on tests of familiar and unfamiliar face memory, and simultaneous unfamiliar face matching. The scores of the police identifiers did not reliably differ from a group of super-recognisers, meeting diagnostic criteria for this outstanding face recognition ability (e.g., Bobak et al., 2016a; 2016b; Russell et al., 2009). The 36 police identifiers were some of the first to join a pool of approximately 140 (by 2015) designated MPS officers regularly tasked with viewing crime scene images. This pool comprises less than 0.1% of the MPS workforce of 48,000. A growing body of research has demonstrated that some individuals in the general population possess exceptional ability, and it is not surprising therefore that some police are also outstanding (see also Robertson et al., 2016).

These between-group differences are particularly striking as there was evidence of a recruitment bias. Even though super-recognisers and police identifiers scored higher than controls on the *Glasgow Face Matching Test*, the controls in turn scored higher than the original published data of 194 participants on this test (Burton *et al.*, 2010). This suggests that the controls as a group may possess a higher ability than the ‘average’ member of the general public. An explanation is that lower ability controls may have been deterred, and higher ability controls attracted by the challenge of the advertised two-hour face recognition tests.

Individual analyses, in which the sensitivity scores of each super-recogniser and police identifier were compared to the control mean on each test, demonstrated however, that only a few performed exceptionally across all four face processing tests. Comparable inconsistencies across tests have been found in previous super-recognition research (Bobak et al., 2016a; 2016b). Based on research finding similar variance in the performances of developmental prosopagnosics, Bobak and colleagues suggest that super-recognition may be cognitively heterogeneous, underpinned by different neurological processes. The current research was not designed to answer this, although the results are consistent with this proposal. Nonetheless, the differences found here may simply be due to design anomalies (the surprise removal of external features from learning-to-test in the *Old/New Unfamiliar Face Test*); brief exposure times (unfamiliar face memory test learning phases were only 5-sec); ceiling effects (*Glasgow Face Matching Test*), interest in contemporary culture (*Famous Face Recognition Test*), anxiety and aversion to taking tests (anecdotal post-test feedback) and fatigue (the tests took up to two hours).

Nevertheless, on the combined mean scores across the four tests (Figure 12), almost all super-recognisers and police identifiers scored above the control mean (50% grid line) suggesting better than average ability, and 50% of the super-recognisers and 30.6% of the police identifiers scored *significantly* higher. Figure 12 also shows that the performance profile of the highest scoring super-recognisers and police identifiers was remarkably similar, with their scores expected to exceed high proportions of the population, suggesting they effectively belong in the same super-recognition category. In addition, both groups also achieved higher hit *and* correct rejection rates than controls on most tests. A correct rejection requires *rejection recollection* (e.g., Brainerd, Reyna, Wright, & Mojardin, 2003), or *recall to reject* (Rotello & Heit, 2000). In a police context, high correct rejection rates would generate few false leads. The ability to reliably recognise suspects may partly be based on the ability to know that you have not seen a face before.

The four face processing test results contrast with those of the *Object Memory Test*, in which there were no reliable between-group differences, and many super-recognisers and police identifiers scored below the control mean (Figure 10). Except for significant relationships between the *Object Memory Test* and the *Old-New Unfamiliar Face Memory Test*, probably due to identical test designs accessing similar memorial processes, there were also no relationships between face and object recognition. These results support previous research suggesting super-recognition ability is face-specific (Bobak et al., 2016a), as well as theoretical proposals suggesting that faces are *special* in they are processed by *domain-specific* cortical networks (e.g., Duchaine & Nakayama, 2006), and prosopagnosia is associated with a breakdown of those networks. An alternative *domain-general* theoretical viewpoint based on face expertise is that prosopagnosia is associated with a failure of the processes involved in the recognition of any object for which the sufferer possesses expertise, in this case faces (e.g., Gauthier et al., 2000). However, conclusions are limited in that only one non-face test was included here. Super-recognisers may prove to be superior at many alternative visual processing tests not involving faces.

The *Famous Face Recognition Test* results also support the suggestion that super-recognisers may possess a larger-than-normal long term facial representation capacity, as they recognised a greater number of 12-year-old degraded celebrity faces – some no longer appearing regularly in the media. Nevertheless, consistent with previous research (e.g., Lander *et al*., 2001), performance was higher with moving images, with stronger effect sizes greater for super-recognisers and police identifiers, but only on the unconditionalised hit rate data. One explanation could be that these groups more effectively extract the increased information available from the additional frames. However, effect sizes were small and further research is required to isolate processes. Regardless, the display of moving images on police wanted websites should enhance ident rates by all witnesses.

Superior performance on the extended version of the CFMT (Russell et al., 2009) has been the primary *objective* method of diagnosing super-recognition ability (e.g., Bobak et al., 2016a). A score of 90 out of 102 (88.2%) being the minimum standard. On a continuum of ability, such a label will always be arbitrary, cutting off those fractionally below the selected mark. Nevertheless, the individual analyses conducted here demonstrate that two out of ten (20%) super-recognisers (SR9, SR10) who easily exceeded this standard (CFMT scores = 96.1%; 93.1%) performed relatively poorly at the four face processing tests. Indeed, believing that SR10's exceptionally low performance on the *Glasgow Face Matching Test* was potentially due to experimenter error, they were asked to take the test a second time a few months later. Their second performance virtually equalled their first. As noted above, similar inconsistencies have been found previously (Bobak et al., 2016a), and we agree with the authors that future research should examine the utility of the CFMT, or indeed of any other test used to diagnose super-recognition ability.

For similar reasons, based on individual performances on the tests reported here, it is not possible to suggest a minimum threshold for police identifier pool membership. Even though the scores of the lower performing police identifiers may have been adversely influenced by the design anomalies described above, an arbitrary cut off point would still exclude some who would otherwise make substantial numbers of idents. Poor test performances can also be explained simply by some police identifier's high levels of familiarity with suspects who regularly get captured on camera – which in most cases would not require extraordinary ability. Indeed, in anecdotes, some claimed their idents were often based on cues such as distinctive clothing, hairstyle, body shape, gait, tattoos or scars, and that the face was sometimes a less important cue. Although the current research did not test claim veracity, to rule out this factor, none of the tests included targets with distinctive marks or were depicted wearing the same clothing in learning and test phases.

The results of many police identifiers who do show a superior pattern of performance on the tests reported here; together with the findings of Robertson et al. (2016), nevertheless suggest that it would be a worthwhile policy measuring the face processing abilities of all police officers. As has proved so successful for the London MPS, giving super-recogniser police time out of their normal roles to view crime scene images should increase suspect identification and sentencing rates. However, the variability in test performances, also supports the proposal that a range of tests might be required for selection to specific operations that require different types of superior face processing ability.

There are a number of factors that limit conclusions. Super-recogniser numbers were low – although more than for any previous published research. Moreover, the police identifiers tested comprise a minority of the MPS pool, and the abilities of the remainder cannot be assumed. However, ident rates have continued to rise as pool membership has increased, suggesting the group tested here is unlikely to be unrepresentative. In addition, most participants, and all stimuli were white-Caucasian, and as a result these results do not inform as to whether super-recognisers display the typical cross-ethnicity (for a meta-analysis see Meissner & Brigham, 2001) or cross-age effects (e.g., Perfect & Moon, 2005) in which performance is normally worse to other-group faces. In addition, learning phase face exposure was only 5-sec in both unfamiliar face memory tests. Longer exposure would likely improve performance (Bornstein, Deffenbacher, Penrod, & McGorty, 2012), although longer delays between learning and test would likely reduce accuracy (for a meta-analysis see Deffenbacher, Bornstein, McGorty, & Penrod, 2008). However, conclusions are limited due to the short term nature of these tasks.

The issue of motivation should also be considered when examining between-group differences on any cognitive test (see Duckworth, Quinn, Lynam, Loeber, & Stouthamer-Loeber, 2011). Offering rewards can enhance face matching test performance (Moore & Johnston, 2013), and it might be expected that due to their ident successes, the police identifiers might gain intrinsic motivation from face processing competence tasks (See Deci, 1971, White, 1959). Indeed, their (slightly) lower performance on the *Object Memory Test*, which did not include faces, might suggest lower motivation on that task alone. However, motivational influences may be limited. Previous findings have found no differences between students and either non-specialist police (e.g., Burton et al., 1999) or passport officers (White et al., 2015) at face matching. Furthermore, the high control performance on the *Glasgow Face Matching Test*, after more than an hour of testing, suggests no lack of motivation.

In summary, this research demonstrated that some specialist police officers regularly identifying suspects from crime scene images possess superior face processing ability. It is probable that alternative organisations might benefit by employing individuals with this ability (e.g., passport control, security). However, any work that requires face-to-face but irregular contact with clients may be enhanced by selecting employees with outstanding face recognition ability. Nevertheless, it would be a risky policy if a suspect ident by a *super-recogniser police identifier* was given inappropriately high evidential weight by the courts.

Super-recognisers are not super-human. No participant in the current research achieved a perfect score on all tests, and police records indicate that police identifiers are occasionally mistaken with their idents - sometimes those from low quality imagery are candidly tentative. Furthermore, all police and forensic experts may be susceptible to confirmation biases by unduly interpreting evidence confirming pre-existing beliefs as to a suspect's guilt, while disregarding exculpatory evidence (see Edmund, Davis, & Valentine, 2015; Kassin, Dror, & Kukucka, 2013). Placing undue weight on a super-recogniser police identifier ident might increase these risks. Some legal protections to insulate suspects from these effects are in place in England and Wales, due to the enactment of the Police and Criminal Evidence Act (1984; Codes of Practice, Code D). These codes apply to crime scene image viewing. At present, few police forces worldwide appear to have identified super-recognisers in their ranks. This may partly be because the UK, and in particular London led the world in large scale CCTV implementation. If other jurisdictions do assimilate these processes; similar protections to protect the rights of suspects are required. Nevertheless, super-recogniser police identifiers should be viewed as an important investigative tool in the armoury of law enforcement. A tentative super-recogniser identification may provide the first vital lead pointing towards alternative stronger evidence of guilt.

Footnotes

[1] For MPS records, when a police officer identifies one suspect from one crime scene they are accredited with one ident. If three suspects are identified from one crime scene, they are accredited with three idents. If they recognise two suspects in images of two different crimes they are accredited with four idents. Finally, if more than one officer independently recognises the same suspect, each is accredited with one ident.

[2] Without any previous empirical evaluation of their abilities, these officers and civilian staff have also been described by the MPS as super-recognisers. For clarity, here, they are referred to as ‘police identifiers’. ‘Super-recogniser’ is used for individuals who have scored in the top 2% on the enhanced-CFMT (Russell et al., 2009).

[3] *http://content.met.police.uk/Site/caughtoncamerametcu*

[4] Volunteer participants at the museum viewed museum organised posters or online adverts inviting participation in a study measuring face recognition ability. The aim was to replicate the Russell et al. (2009) study with a larger sample. The participants completed the CFMT, an adapted version of Russell et al.'s *Before They Were Famous Test* and a *Navon Letters Test*. The results will be reported elsewhere. However, the highest scorers on the CFMT who left contact details were invited to take part in the research reported here.

[5] For the calculation of $d^{/}$ and C, hit rates of 1.0 were converted using the formula, $Padj = 1 - 1 / (2N)$, whereas false alarm rates of 0.0 were converted using the formula, $Padj = 1 / (2N)$ [where N = number of items].

[6] Further analyses examined whether the hit rate movement-advantage for super-recognisers and police identifiers was related to other aspects of face processing ability. Two Pearson’s correlation tests found however no relationship between movement-advantage scores (calculated by subtracting static image hit rates from moving image hit rates) and face-specific recognition (calculated by subtracting hit rates on the *Object Memory Test*, from those on the *Old-New Unfamiliar Face Memory Test* (see Wang et al., 2012), $r(189) < .01$, p

> .5, or with mean hit rates on the three unfamiliar face processing tests (*p* > .1 all tests). These results suggest no reliable relationship between the familiar face movement advantage and unfamiliar face memory and matching.

[7] CNR analyses are not conducted on correct rejections as all stimuli depict unknown individuals.

[8] Calculated using *Famous Face Recognition Test* $_{CNR}$ *and* $_{UN}$ data for hits, d', and C.

[9] Using *Famous Face Recognition Test* $_{UN}$ data instead, five super-recognisers, and ten police identifiers scored significantly higher than the control mean.

References

Bindemann, M., Brown, C., Koyas, T., & Russa, A. (2012). Individual differences in face identification postdict eyewitness accuracy. *Journal of Applied Research in Memory and Cognition, 1*, 96-103. DOI: 10.1016/j.jarmac.2012.02.001

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

Bobak, A. K., Hancock, P. J. B. and Bate, S. (2016b). Super-Recognizers in action: Evidence from face matching and face memory tasks. *Applied Cognitive Psychology, 30*(1), 81-91. DOI: 10.1002/acp.3170

Bobak, A. K., Parris, B. A., Gregory, N. J., Bennetts, R. J., & Bate, S. (2016c). Eye-movement strategies in developmental prosopagnosia and “super” face recognition. *Quarterly Journal of Experimental Psychology*, DOI: 10.1080/17470218.2016.1161059

Bornstein, B. H., Deffenbacher, K. A., Penrod, S. D., & McGorty, E. K. (2012). Effects of exposure time and cognitive operations on facial identification accuracy: A meta-analysis of two variables associated with initial memory strength. *Psychology, Crime and Law, 18*, 472–490. doi:10.1080/1068316X.2010.508458

Brainerd, C. J., Reyna, V. F., Wright, R., & Mojardin, A. H. (2003). Recollection rejection: False-memory editing in children and adults. *Psychological Review, 110*, 762–784. http://dx.doi.org/10.1037/0033-295X.110.4.762

Bruce, V., Henderson, Z., Greenwood, K., Hancock, P. J. B., Burton, A. M., & Miller, P. (1999). Verification of face identities from images captured on video. *Journal of Experimental Psychology: Applied, 5*, 339-360. http://dx.doi.org/10.1037/1076-898X.5.4.339

Bruce V., Henderson, Z., Newman, C., & Burton, A. M. (2001). Matching identities of familiar and unfamiliar faces caught on CCTV images. *Journal of Experimental Psychology: Applied, 73*, 207-218. http://dx.doi.org/10.1037/1076-898X.7.3.207

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology, 77*, 305-327. DOI: 10.1111/j.2044-8295.1986.tb02199.x

Burton, A. M. (2013). Why has research in face recognition progressed so slowly? The importance of variability. *The Quarterly Journal of Experimental Psychology, 66*, 1467-1485. doi:10.1080/ 17470218.2013.800125

Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow face matching test. *Behavior Research Methods, 42*, 286–291. DOI: 10.3758/BRM.42.1.286

Burton, A. M., Wilson, S., Cowan, M., & Bruce, V. (1999). Face recognition in poor quality video: Evidence from security surveillance. *Psychological Science, 10*, 243–248. doi: 10.1111/1467-9280.00144

Crawford, J. R., Garthwaite, P. H., & Porter, S. (2010). Point and interval estimates of effect sizes for the case-controls design in neuropsychology: Rationale, methods, implementations, and proposed reporting standards. *Cognitive Neuropsychology, 27*, 245-260. DOI:10.1080/02643294.2010.513967

Dalrymple, K. A., Garrido, L., & Duchaine, B. (2014). Dissociation between face perception and face memory in adults, but not children, with developmental prosopagnosia. *Developmental Cognitive Neuroscience, 10*, 10-20. http://dx.doi.org/10.1016/j.dcn.2014.07.003 1878-9293

Davis, J. P., Lander, K., & Jansari, A. (2013). I never forget a face. *The Psychologist, 26(10),* 726-729.

Davis, J. P. & Valentine, T. (2009). CCTV on trial: Matching video images with the defendant in the dock. *Applied Cognitive Psychology, 23,* 482–505. DOI: 10.1002/acp.1490

Davis, J. P., & Valentine, T. (2015). Human verification of identity from photographic images. In T. Valentine and J. P. Davis (Eds.), *Forensic Facial Identification: Theory and Practice of Identification from Eyewitnesses, Composites and CCTV* (pp. 211-238). Chichester: Wiley-Blackwell.

Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology, 18,* 105-115. doi.org/10.1037/h0030644

DeGutis, J., Wilmer, J., Mercado, R. J., & Cohan, S. (2013). Using regression to measure holistic face processing reveals a strong link with face recognition ability. *Cognition, 126*, 87-100. doi:10.1016/j.cognition.2012.09.004

Deffenbacher, K. A., Bornstein, G. H., McGorty, E. K., & Penrod, S. D. (2008). Forgetting the once-seen face: Estimating the strength of an eyewitness's memory representation. *Journal of Experimental Psychology: Applied, 14,* 139–150. doi:10.1037/1076-898X.14.2.139

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

Duchaine, B., & Nakayama, K. (2006). Developmental prosopagnosia: A window to content-specific face processing. *Current Opinions in Neurobiology*, *16*, 166-173. doi:10.1016/j.conb.2006.03.003

Duckworth, A. L., Quinn, P. D., Lynam, D. R., Loeber, R., & Stouthamer-Loeber, M. (2011). Role of test motivation in intelligence testing. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(19), 7716–7720. http://doi.org/10.1073/pnas.1018601108

Edmund, G., Davis, J. P., & Valentine, T. (2015). Expert analysis: Facial image comparison. In T. Valentine and J. P. Davis (Eds.), *Forensic Facial Identification: Theory and Practice of Identification from Eyewitnesses, Composites and CCTV* (pp. 239-262). Chichester: Wiley-Blackwell.

Ellis, H. D., Shepherd, J. W. & Davies, G. M. (1979). Identification of familiar and unfamiliar faces from internal and external features: some implications for theories of face recognition. *Perception, 8*, 431-439

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience, 3*, 191-197. doi:10.1162/jocn.1997.9.5.605

Green, D. M., & Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. New York: Wiley.

Jansari, A., Miller, S., Pearce, L., Cobb, S., Sagiv, N., Williams, A. L., Tree, J. J., & Hanley, J. R. (2015). The man who mistook his neuropsychologist for a popstar: when configural processing fails in acquired prosopagnosia. *Frontiers in Human Neuroscience, 9,* 390. doi: 10.3389/fnhum.2015.00390

Jenkins, R., White, D., van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition, 121,* 313-323. DOI: 10.1016/j.cognition.2011.08.001

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

Kassin, S. M., Dror, I. E., & Kukucka, J. (2013). The forensic confirmation bias: Problems, perspectives and proposed solutions. *Journal of Applied Research in Memory and Cognition, 2*, 42–52. doi:10.1016/j.jarmac.2013.01.001

Knight, B., & Johnston A. (1997). The role of movement in face recognition. *Visual Cognition, 4*, 265-273. DOI:10.1080/713756764

Lander, K., & Bruce, V. (2000). Recognizing famous faces: Exploring the benefits of facial motion. *Ecological Psychology*, *12*, 259-272. DOI:10.1207/S15326969ECO1204_01

Lander, K., Bruce, V., & Hill, H. (2001). Evaluating the effectiveness of pixilation and blurring on masking the identity of familiar faces. *Applied Cognitive Psychology, 15*, 101-116. DOI: 10.1002/1099-0720(200101/02)15:1<101:AID-ACP697>3.0.CO;2-7

Lander, K., & Chuang, L. (2005). Why are moving faces easier to recognise? *Visual Cognition, 12,* 429-442. DOI:10.1080/13506280444000382

Lander, K., & Poyarekar, S. (2015). Famous face recognition, face matching and extraversion. *Quarterly Journal of Experimental Psychology*, *68*(9), 1769-1776. doi: 10.1080/17470218.2014.988737

Li, J., Tian, M., Fang, H., Xu, M., & Liu, J. (2010). Extraversion predicts individual differences in face recognition. *Communicative and Integrative Biology, 3-4*, 295-298. DOI:10.4161/cib.3.4.12093

Macmillan, N., & Creelman, C. D. (1991). *Detection theory: A user's guide*. New York: Cambridge University Press.

Megreya, A. M., & Burton, A. M. (2006). Unfamiliar faces are not faces: Evidence from a matching task. *Memory & Cognition, 34,* 865-876. DOI: 10.3758/BF03193433

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*, 1-35. http://dx.doi.org/10.1037/1076-8971.7.1.3

Mickes, L., Moreland, M., Clark, S., & Wixted, J. (2014). Missing the information needed to perform ROC analysis? Then compute d', not the diagnosticity ratio. *Journal of Applied Research in Memory & Cognition, 3,* 58–62. doi:10.1016/j.jarmac.2014.04.007

Moore, R. M., & Johnston, R. A. (2013). Motivational incentives improve unfamiliar face matching accuracy. *Applied Cognitive Psychology*, *27*, 754–760. doi:10.1002/acp.2964

Perfect, T. J., & Moon, H. (2005). The own age effect in face recognition. In J. Duncan, and P. McLeod (Eds.), *Measuring the Mind: Speed, Control, and Age* (pp. 317–340). Oxford: Oxford University Press.

Police and Criminal Evidence Act (PACE) (1984). Codes of Practice (Code D). (2011). https://www.gov.uk/government/publications/pace-code-d-2011

Richler, J. J., Floyd, R. J., & Gauthier, I. (2015). About-face on face recognition ability and holistic processing. *Journal of Vision*, *15*(9), 15–12. http://doi.org/10.1167/15.9.15

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), e0150036–8. http://doi.org/10.1371/journal.pone.0150036

Rossion, B., Caldara, R., Seghier, M., Schuller, A. M., Lazeyras, F. & Mayer, E. (2003). A network of occipito-temporal face-sensitive areas besides the right middle fusiform gyrus is necessary for normal face processing. *Brain, 126*, 2381-2395. DOI: http://dx.doi.org/10.1093/brain/awg241

Rotello, C. M., & Heit, E. (2000). Associative recognition: A case of recall-to-reject processing. *Memory & Cognition, 28,* 907–922. http://dx.doi.org/10.3758/BF03209339

Russell, R., Chatterjee, G., & Nakayama, K. (2012). Developmental prosopagnosia and super-recognition: no special role for surface reflectance processing. *Neuropsychologia, 50,* 334-340. doi:10.1016/j.neuropsychologia.2011.12.004

Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic Bulletin & Review, 16*, 252-257. doi:10.3758/PBR.16.2.252

Shakeshaft, N. G., & Plomin, R. (2015). Genetic specificity of face recognition. *Proceedings of the National Academy of Sciences of the United States of America, 112*(41), 12887–12892. doi: 10.1073/pnas.1421881112

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. doi: 10.1177/0956797611420575

White, R. W. (1959). Motivation reconsidered: The concept of competence. *Psychological Review*, *66*, 297-333. doi.org/10.1037/h0040934

White, D., Dunn, J. D., Schmid, A. C., & Kemp, R. I. (2015a). Error rates in users of automatic face recognition software. *PloS One, 10*(10), e0139827–14. doi.org/10.1371/journal.pone.0139827

White. D., Kemp, R. I., Jenkins, R., Matheson, M., & Burton, A. M. (2014). Passport officers' errors in face matching. *PLoS One*, *9*, e103510. doi:10.1371/journal.pone.0103510

White, D., Phillips, P. J., Hahn, C. A., Hill, M., & O'Toole, A. J. (2015b). Perceptual expertise in forensic facial image comparison. *Proceedings of the Royal Society B: Biological Sciences, 282*(1814), 1292–1298. doi.org/10.1098/rspb.2015.1292

Wilkinson, C., & Evans, R. (2011). Are facial image analysis experts any better than the general public at identifying individuals from CCTV images? *Science and Justice, 49*, 191–196. doi:10.1016/j.scijus.2008.10.011

Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Williams, M., Loken, E., Nakayama, K., & Duchaine, B. (2010). Human face recognition ability is specific and highly heritable. *Proceedings of the National Academy of Sciences of the United States of America, 107*, 5238-5241. doi:10.1073/pnas.0913053107

Figure 1: a) Mean hits (proportions); b) correct rejections (proportions); c) sensitivity (d'); and d) response bias (C) on the Famous Face Recognition Test separately for static (black bars) and moving (grey bars) images (error bars = standard error of the mean)

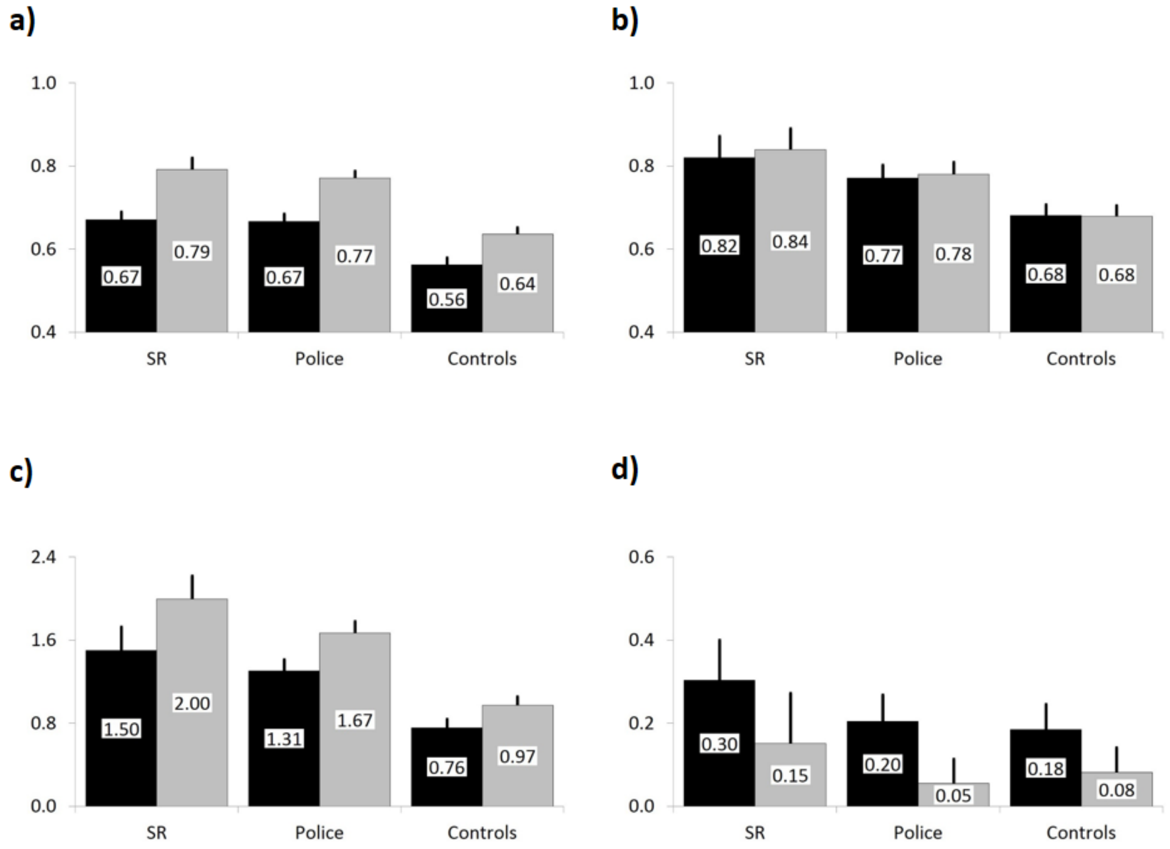

Figure 2: Upper and lower bound confidence intervals (95%) of the estimated proportion of the general population expected to fall below each super-recogniser (SR, n = 10) and police identifier (PI, n = 36) based on their d′ scores from the Famous Face Test $_{CNR.}$ *To enhance interpretability, the SRs and PIs are batched in groups of (mainly) ten; and ordered and numbered based on their overall mean* d'_{CNR} *score across the four face-based tests (see Figure 12). The 50% grid line represents the control mean, so that 50% of the population would be expected to achieve above this level, 50% below. For interpretation purposes, the lower bound confidence intervals from Figure 2 demonstrate that 9 super-recognisers (90%), and 29 police identifiers (80.6%) scored above the control mean, and thus would likely score higher than 50% of the population. Similarly, 5 super-recognisers (50%) and 10 police identifiers (27.8%) would likely score higher than 75% of the population.*

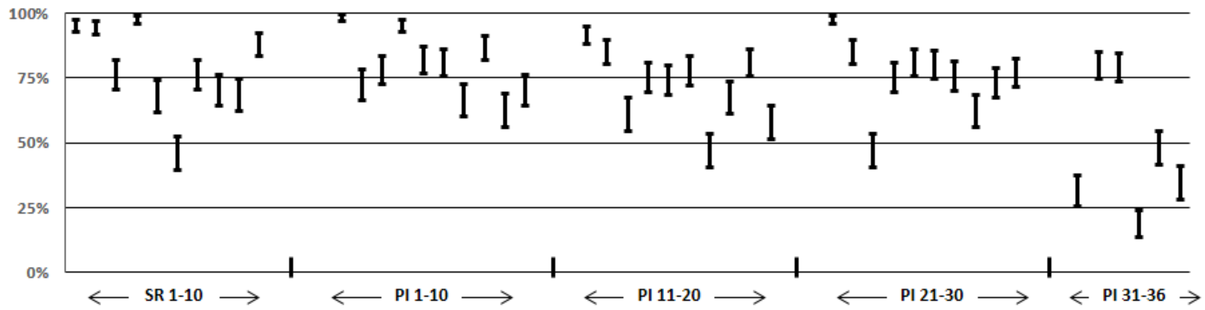

Figure 3: a) Mean hits (proportions); b) correct rejections (proportions); c) sensitivity (d/); and d) response bias (C) on the Unfamiliar Face Memory Array Test separately for first phase full face (grey bars) and three-quarter (black bars) view images (error bars = standard error of the mean)

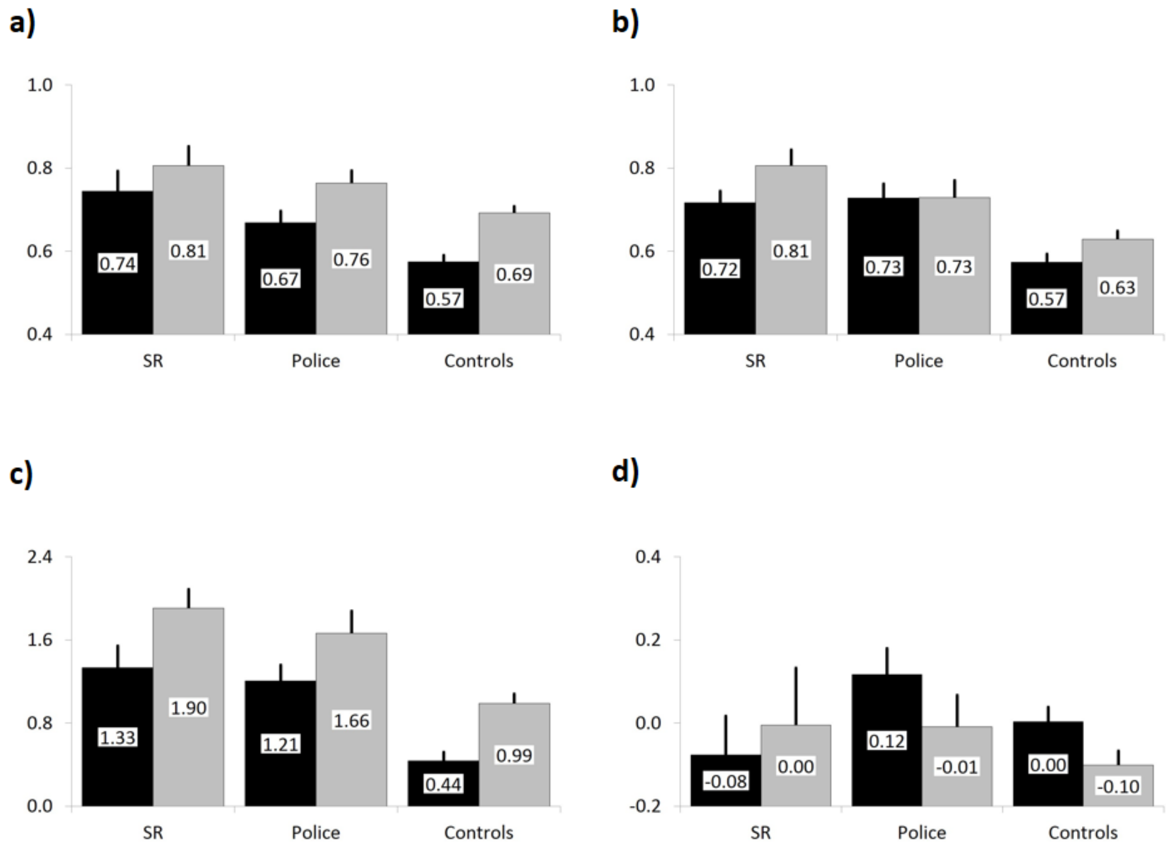

Figure 4: Upper and lower bound confidence intervals (95%) of the estimated proportion of the general population expected to fall below each super-recogniser and police identifier based on their d′ scores from the Unfamiliar Face Memory Array Test (see Figure 2 for interpretation)

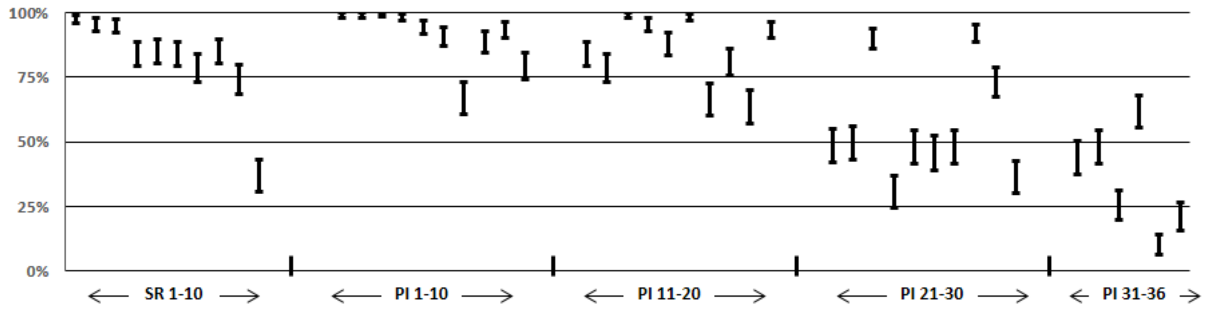

Figure 5: a) Mean hits (proportions); b) correct rejections (proportions); c) sensitivity (d'); and d) response bias (C) on the Unfamiliar Old/New Face Memory Test (error bars = standard error of the mean)

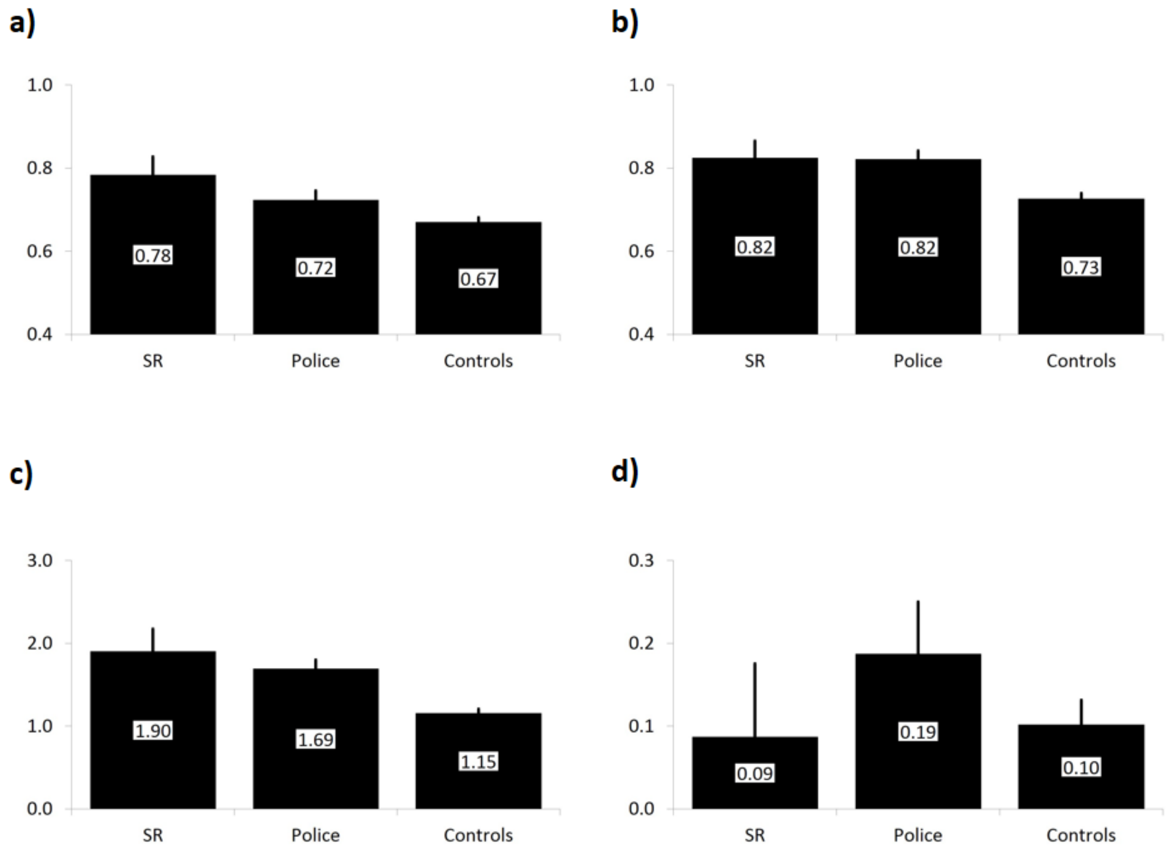

Figure 6: Upper and lower bound confidence intervals (95%) of the estimated proportion of the general population expected to fall below each super-recogniser and police identifier based on their d′ scores from the Old/New Unfamiliar Face Memory Test (see Figure 2 for interpretation)

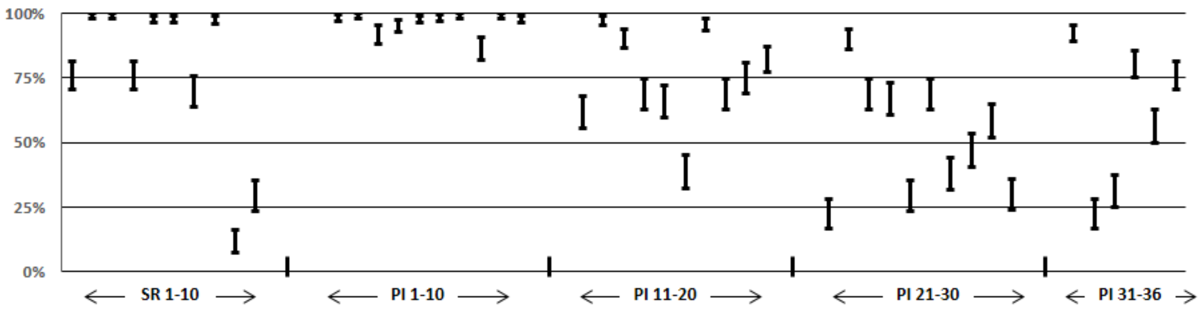

Figure 7: a) Mean hits (proportions); b) correct rejections (proportions); c) sensitivity (d/); and d) response bias (C) on the Glasgow Face Matching Test (error bars = standard error of the mean)

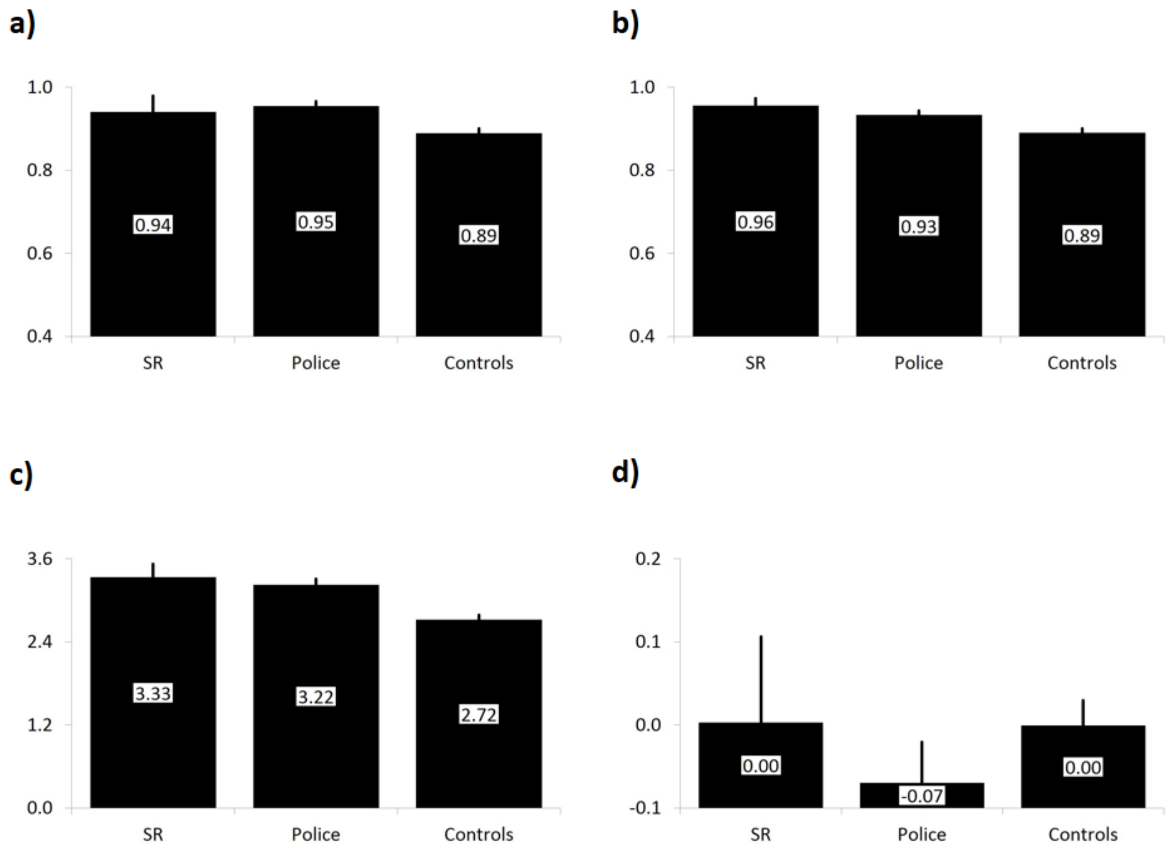

Figure 8: Upper and lower bound confidence intervals (95%) of the estimated proportion of the general population expected to fall below each super-recogniser and police identifier based on their d' *scores on the Glasgow Face Matching Test (see Figure 2 for interpretation)*

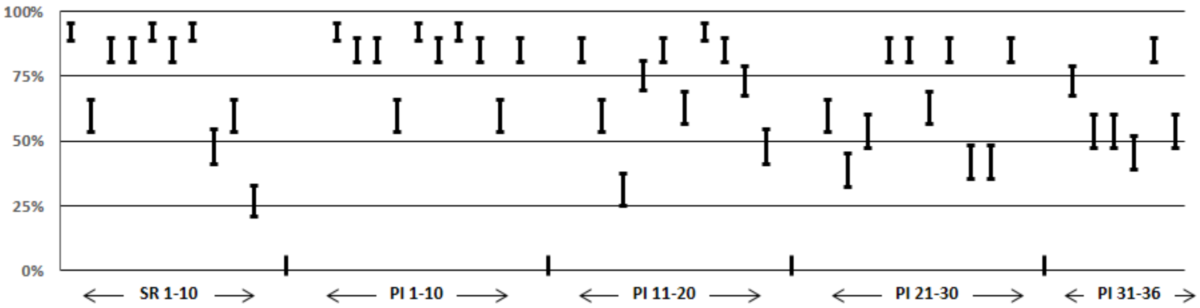

Figure 9: a) Mean hits (proportions); b) correct rejections (proportions); c) sensitivity (d/); and d) response bias (C) on the Object Memory Test (error bars = standard error of the mean)

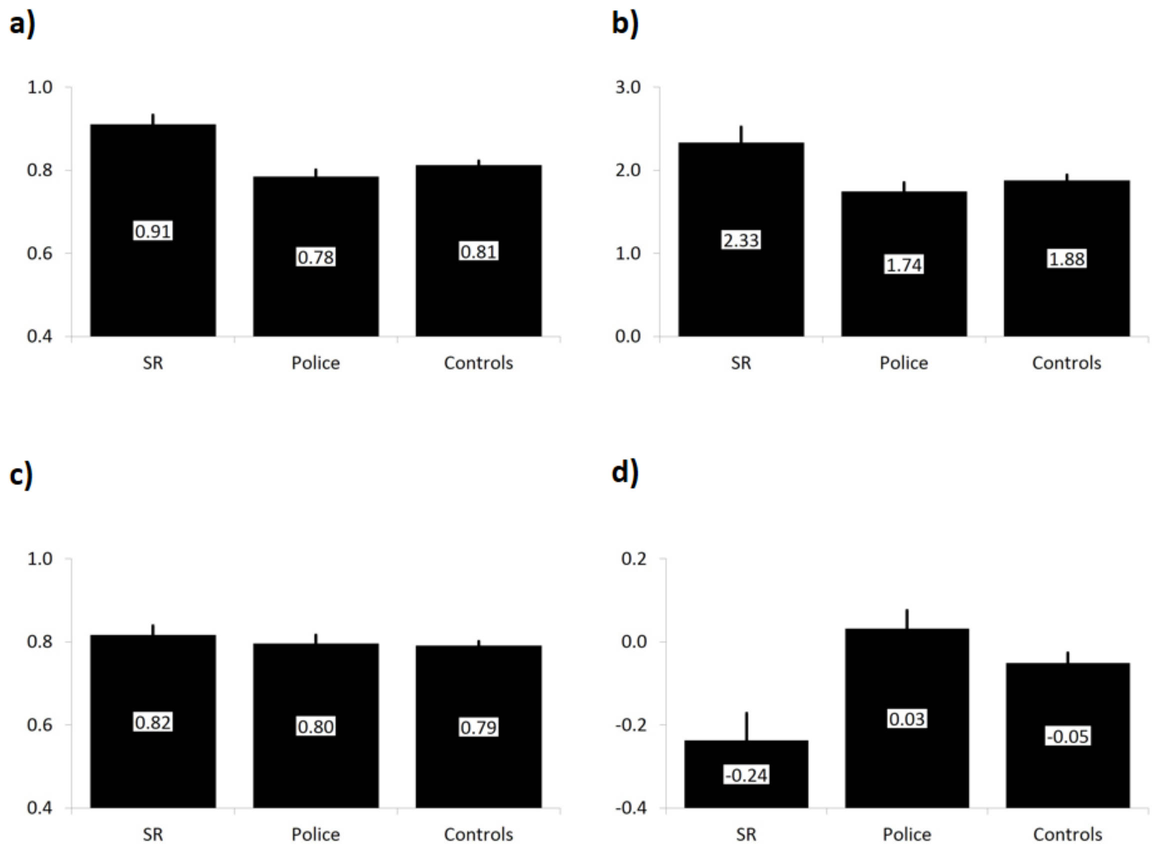

Figure 10: Upper and lower bound confidence intervals (95%) of the estimated proportion of the general population expected to fall below each super-recogniser and police identifier based on their d' *scores on the Object Memory Test (see Figure 2 for figure interpretation)*

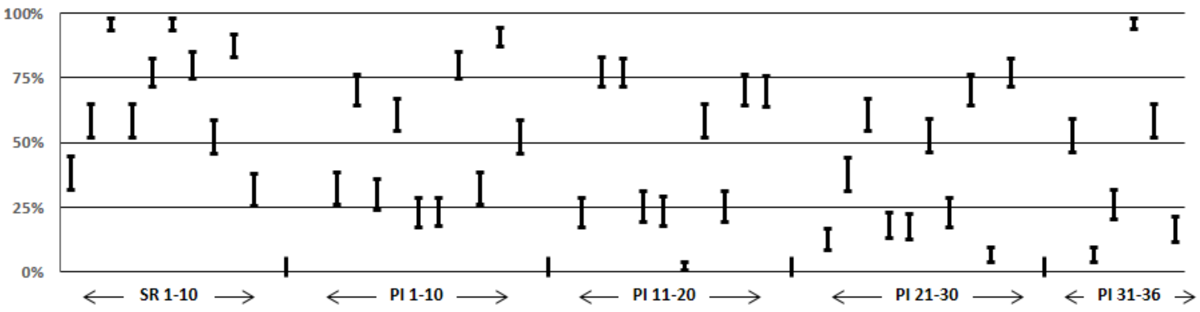

Figure 11: a) Mean hits (proportions); b) correct rejections (proportions); c) sensitivity (d/); and d) response bias (C) across the four face processing tests (error bars = standard error of the mean)

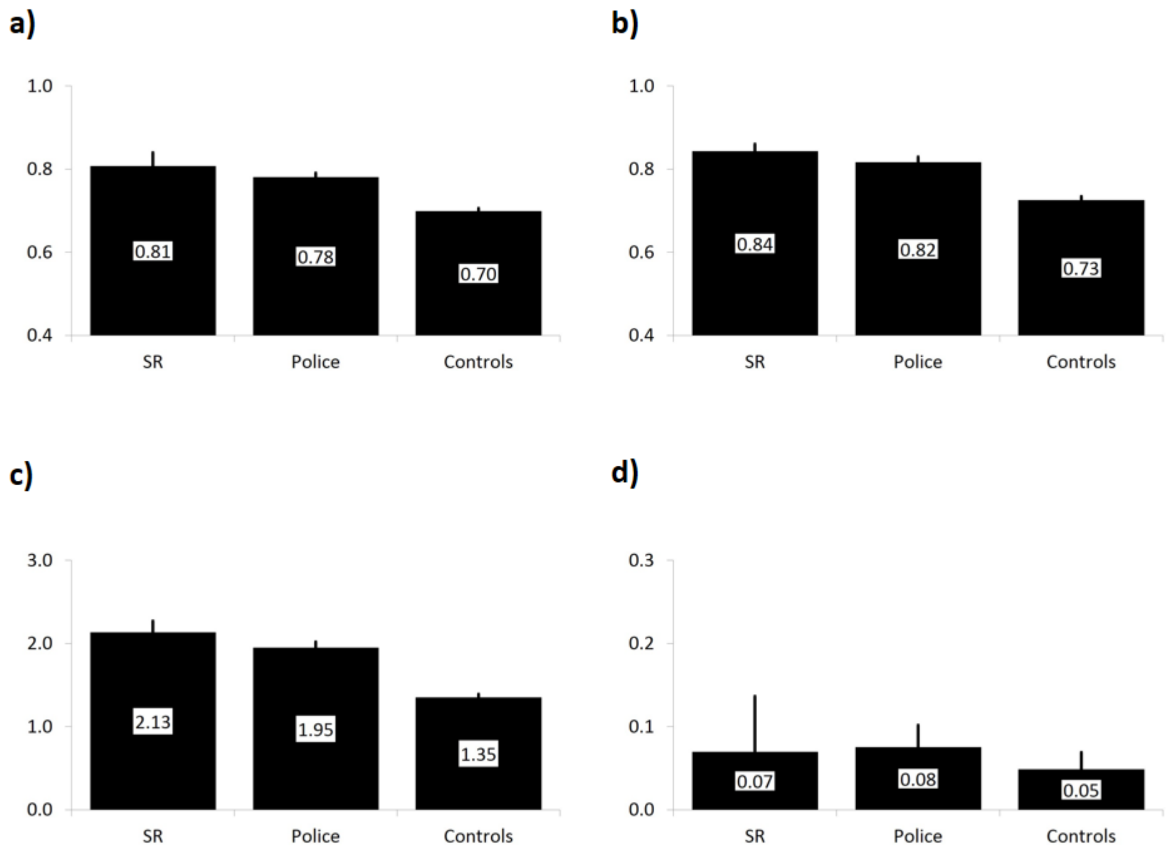

Figure 12: Upper and lower bound confidence intervals (95%) of the estimated proportion of the general population expected to fall below each super-recogniser and police identifier based on their d/ scores across the four face processing tests (see Figure 2 for figure interpretation)

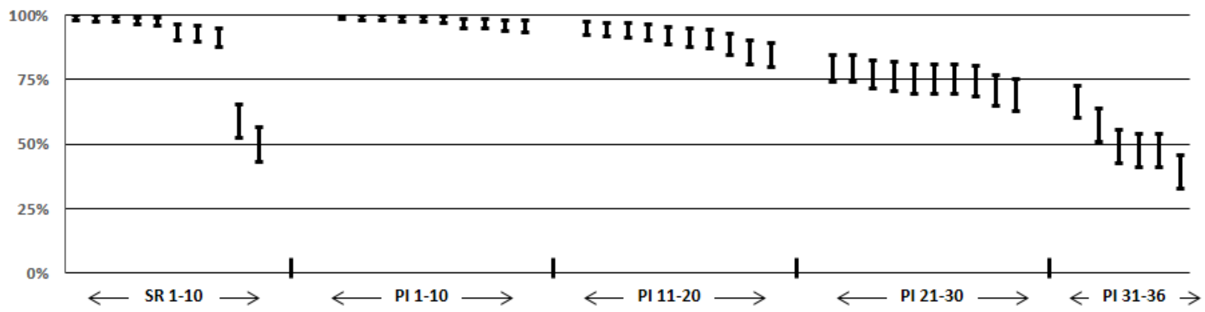

Table 1: Correlation coefficients between a) hits b) correct rejections c) sensitivity (d/) and response bias (C) on the Famous Face Recognition Test $_{CNR}$ (FFT); the Unfamiliar Face Memory Array Test (UFAT); the Old-New Unfamiliar Face Memory Test (O/NUFT), the Glasgow Face Matching Test (GFMT) and the Object Memory Test (OMT) (n = 189)

a) Hits

	T1: FFT	T2: UFAT	T3: O/NUFT	T4: GFMT
T2: UFAT	.19 **			
T3: O/NUFT	.12	.08		
T4: GFMT	.13	.30 **	.09	
T5: OMT	.10	.10	.07	-.08

b) Correct rejections

	T1: FFT	T2: UFAT	T3: O/NUFT	T4: GFMT
T2: UFAT	.16 *			
T3: O/NUFT	.10	.20 **		
T4: GFMT	-.01	.17 *	.21 **	
T5: OMT	-.07	.11	.29 **	.05

c) Sensitivity (d/)

	T1: FFT	T2: UFAT	T3: O/NUFT	T4: GFMT
T2: UFAT	.27 **			
T3: O/NUFT	.14	.29 **		
T4: GFMT	.18 *	.28 **	.31 **	
T5: OMT	-.06	.12	.15 *	<.01

d) Response bias (C)

	T1: FFT	T2: UFAT	T3: O/NUFT	T4: GFMT
T2: UFAT	.10			
T3: O/NUFT	.11	.04		
T4: GFMT	-.10	.12	<.01	
T5: OMT	.06	.10	.20 **	-<.01

$^*p < .05$ $^{**}p < .01$