The worldwide public impact of identifying super-recognisers for police and business Dr Josh P Davis

Reader in Applied Psychology, University of Greenwich, London

Home Page: http://www.gre.ac.uk/people/rep/faculty-of-education-and-health/josh-davis

(note: this is the pre-proof version of an article published in and © Cognitive Bulletin, 2019)

Davis, J. P. (2019). The worldwide public impact of identifying super-recognisers for police and business. *The Cognitive Psychology Bulletin, 4, Spring 2019*, 17-21.

Abstract

Super-recognisers occupy the extreme top end of a wide spectrum of human face recognition ability. Although test scores provide evidence of super-recognisers' quantitative superiority, their abilities may be driven by qualitatively different cognitive or neurological mechanisms. Some super-recognisers scoring exceptionally highly on multiple short-term face memory tests do not achieve superior performances on measures of simultaneous face matching, long term face memory and/or spotting faces in a crowd. Heterogeneous performance patterns have implications for police, security or business aiming to utilise super-recognisers' superior skills. Drawing on a global participant base ($n \approx 6{,}000{,}000$), as well as theory and empirical research, this paper describes the background, development, and employment of tests designed to measure four components of superior face processing to assist in recruitment and deployment decisions.

Background.

My PhD mainly conducted at London's Science Museum investigated the *Forensic Identification of CCTV images of Unfamiliar Faces* (2007). As with previous research (e.g. Bruce, Henderson, Greenwood, Hancock, Burton, & Miller, 1999), the results revealed poor human unfamiliar face identification performance across a series of experiments (e.g. Davis & Valentine, 2009). Completed by over 3,000 participants, tests included single-trial high-quality close-up photo-to-photo, video-to-photo, and live target-to-video/photo simultaneous unfamiliar face matching designs none of which drew on memory. A photo-anthropometric technique, now discredited, but commonly employed in court then by expert witnesses was also shown to be unreliable (Davis, Valentine, & Davis, 2010).

Accordingly, I was sceptical when I first heard about 20 London Metropolitan Police Service (MET) police officers (out of a 30,000+ workforce) making a disproportionately large number of suspect identifications from a newly established *Caught on Camera* wanted CCTV image website (Davis, Lander, & Jansari, 2013). I had not heard of *super-recognisers*. The first paper describing participants with extraordinarily good face recognition abilities had not long been published (Russell, Duchaine, & Nakayama, 2009). Nevertheless, in April 2011, some of these 20 police outperformed controls and matched the excellent score patterns of super-recognisers, on long-term familiar face recognition, short-term unfamiliar face recognition and face matching tests (Davis, Lander, Evans, & Jansari, 2016); although some of the super-recognisers might not now meet the more rigorous inclusion criteria described in this paper. In unpublished interviews, police scoring relatively poorly claimed to mainly make multiple identifications of familiar repeat offenders. A large body of empirical evidence suggests familiar and unfamiliar faces are processed using different mechanisms (e.g., Johnston, & Edmonds, 2009), with familiar recognition being far more reliable (e.g. Bruce, Henderson, Newman, & Burton, 2001). However, most CCTV images needing identification will depict suspects unfamiliar to most police, and it is superior unfamiliar face processing that is the hallmark of super-recognition.

Superior face recognisers in the Metropolitan Police Service (MET).

Following the London riots of August 2011, the MET instigated *Operation Withern*. Of the 5,000 rioters charged, 4,000 were initially identified from CCTV and other images. Over 600 were identified by the police who had contributed to our research (Davis et al., 2016). One officer identified 180 rioters. In contrast, a computerised face recognition system identified one rioter only, a consequence of suspect disguise, poor image quality from the typical high placement of CCTV cameras, and the mainly night-time conditions. Most police again claimed pre-existing familiarity with identified rioters, although some were only very slightly familiar, having been viewed in previous footage, or in incidental fleeting, past encounters. Following the riots, and the establishment of improved CCTV collection and distribution systems, the MPS created a super-recogniser pool of front line officers who were the first to view crime specialism or local images. Many had not been empirically tested, although, again based mainly on familiarity, in total they made thousands of suspect identifications (Keefe, 2016, Venkataramanan, 2015).

In 2013, our online bespoke pilot tests of short-term and long-term multi-ethnic unfamiliar face memory using old/new two-phase designs, and simultaneous face matching were uploaded to the Qualtrics system (BSc Psychology projects: Kamińska, 2014; Maigut, 2014). These were advertised on the MET intranet and taken by over 1,400 police. The highest scorers were reported to the MET. Some had never made a CCTV identification. In August 2014 they were deployed to high-profile police investigations drawing on their superior skills (e.g. see Keefe, 2016). These required searching through extremely large volumes of footage to match unfamiliar targets captured by geographically and temporally separated cameras varying in image quality and field of view. They often used clothing, body and gait cues as well as faces, although thousands of bystanders also needed excluding.

In May 2015, some of the same police joined the newly-established *Proactive Super-Recogniser Unit* at New Scotland Yard to review current and cold cases. Employing techniques such as *snapping* unfamiliar suspects from images of different crimes, identification rates substantially increased, and some police identified 100s of previously unfamiliar suspects per annum. Concern was directed at high identification volume reliability, the risk of error from confirmation bias, and the potential prejudicial influence of super-recogniser testimony in court (i.e.

super-powers) (see Davis et al., 2016; Davis, Treml, Forrest, & Jansari 2018). Risks remain, particularly if jurisdictions inappropriately weight super-recogniser evidence. However, the MET publicised super-recogniser unit suspect identification error rates, and instigated blind review procedures, indicative of risk awareness (Keefe, 2016). Moreover, most prosecuted cases needed no super-recogniser to appear in court, as the identifications led investigation teams to stronger evidence of guilt.

Although the MET no longer refers to a super-recogniser unit the work continues (e.g. the annual Notting Hill Carnival; the recent Salisbury Novichok poisoning case; Medeiros, 2018). Investigations like these informed the design of University of Greenwich tests developed as part of a European Commission funded project (LASIE, 2014). These tests have been used by international public-sector bodies to recruit super-recognisers from within their organisations, and in addition over 30 permanent full-time roles have been created by businesses. Most of this work has only been possible because of worldwide media and public interest.

Public engagement with super-recognition research.

In June 2015, a BBC Future online article describing the MET super-recognisers (Venkataramanan, 2015) linked to a 14-trial fun test: *Could you be a Super-Recogniser?* At the time of writing nearly six million participants have taken this test. Figure 1 displays score frequencies out of 14 for the first few months of data collection (M = 9.64, SD = 1.90). Perhaps you contributed? Score patterns remain consistent, but no test with this small number of trials could reliably diagnose ability – it was never meant to. Continued media interest has ensured a constant flow of participants (see Figure 2), and on completion, requests for volunteers to contribute to collaborative research devised in five continents have been successful (n > 250,000). There is however a bias to attract participants who are on average far better at face recognition than is typical – limiting representativeness (e.g. Belanova, Davis, & Thompson, 2018). Fortunately, this means that the hundreds of volunteers who have additionally piloted the tests described below have achieved scores on multiple tests surpassing any published super-recogniser threshold. My assistants are nevertheless mindful of those who repeat tests inflating performances. Their work often involves instigating quality assurance measures.

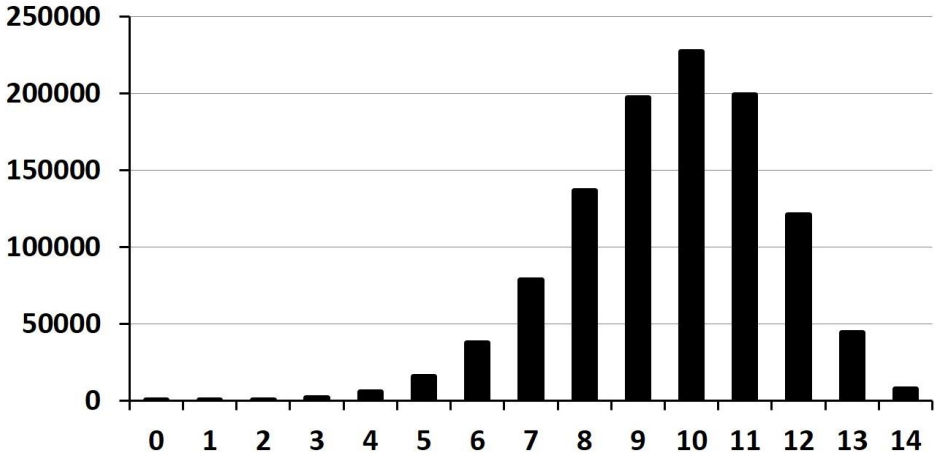

Figure 1: Frequencies of total scores (out of 14) of the first participants on the 14-trial test: Could you be a Super-Recogniser? (n = 1,070,845)

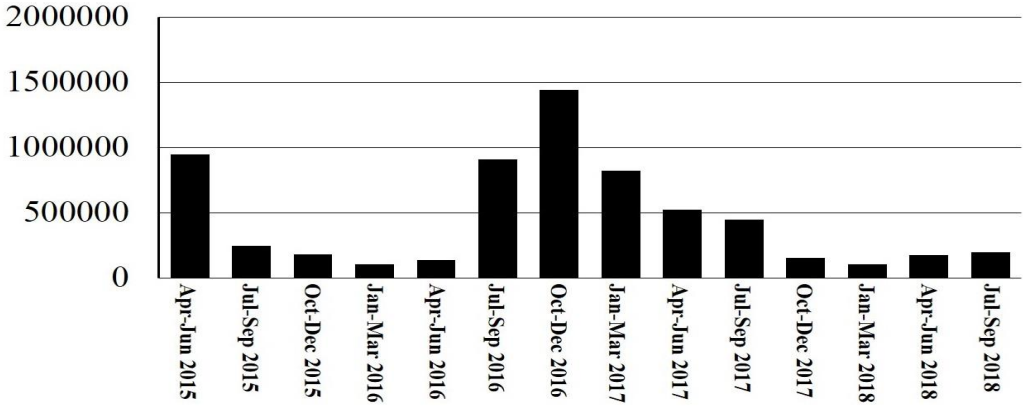

Figure 2: Since its online inception in April 2015, quarterly frequencies of participants taking University of Greenwich face processing tests including, Could you be a Super-Recogniser?

Defining minimum super-recognition test thresholds

There has been a move to standardise super-recognition definitions (Noyes et al., 2017). Replicating the first super-recognition research by Russell et al. (2009), primary super-recognition group inclusion criteria in most research have been anecdotal claims of exceptional face recognition abilities and scores 2 SDs above control means on the standardised short-term *Cambridge Face Memory Test: Extended* (CFMT+), representing the top 2% of scorers (e.g., Belanova et al., 2018; Bobak, Bennetts, Parris, Jansari, & Bate, 2016), sometimes verified by

additional tests. Developmental prosopagnosics who cannot recognise even highly familiar faces tend to score about 2 SDs below the mean on the CFMT+, and assuming a normal distribution, Russell et al. suggested that 1-2% of the population might be super-recognisers. Nevertheless, self-reports of outstanding abilities may be unreliable (e.g. Bate et al., 2018), while control means and SDs on the CFMT+ have slightly varied, impacting thresholds. Super-recognition may also be driven by qualitatively different cognitive and neurological processes (e.g., Belanova et al., 2018; Bobak, Bennetts et al., 2016), and it may not be appropriate to assume super-recognisers are simply located at the extreme high end of a quantitative spectrum of ability. Indeed, some who score exceptionally highly on a series of short-term face memory tests do not always achieve superior performances on alternative tests (e.g., simultaneous face matching). This suggests skill heterogeneity, although, "some individuals may perform in the superior range on that occasion simply by chance, whereas others may fall short of the cut-off value due to extraneous variables such as fatigue, illness or simply having a bad day" (Bate et al., 2018, p 3.). Nevertheless, in practical terms for policing and security, super-recognisers need to possess a superior range of reliable face processing skills for deployment to different tasks.

To be defined as a super-recogniser (as per the University of Greenwich testing procedures), a participant is normally required to consistently score within the typical super-recogniser range on a battery of 10 or more multi-ethnic face tests measuring four components.

a) *Short-term face learning and recognition.* Four tests mainly employing two-phase old-new designs are included. In these, participants view a sequence of faces and are required to identify them from different images later. Learning times and trial numbers vary.

b) *Simultaneous face matching:* Four tests employ different designs with varying image quality including high-quality close-up photos, CCTV-type images, and passport-type images taken up to 10-years apart.

c) *Long-Term Face Memory*: In this test adapted from Bretfelean & Davis (2017; see also Davis, Bretfelean, & Thompson, under review), participants view 10 videos of different target-actors for one-minute each, and with minimum delays of seven-days identify targets from line-ups.

d) *Spotting a Face in a Crowd Test*. In this test adapted from Davis et al. (2018) replicating police CCTV review procedures, participants are provided with four photos of eight different actors and asked to locate them, almost always clothed differently, in videos depicting crowded London tourist spots.

In total, the tests take about 4 hours, and although many are online, some are completed in examination conditions. Different target-distractor proportions are employed to reduce guessing in the short-term memory and simultaneous face matching tests, and the worst test outcome can be discounted as an anomaly, particularly if problematic (distraction, internet crash etc.). Long-term memory and spotting faces in a crowd are only assessed by one test each. Although similar issues may arise, participants have far longer to learn faces in videos on the *Long-Term Face Memory Test,* and there are no time pressures on the *Spot the Face in a Crowd Test*.

As employment requirements differ dependent on organisation, the tests also differ (e.g. face ethnicity range, skill-specific measures). As with previous research, participants may excel on one, two, three or four components; while some high-scorers on one or more components display anomalous below-control mean performances on others. Indeed, on the *Long-Term Face Memory Test*, about 20% of super-recognisers scoring in the top 2% on at least two short-term face recognition tests, scored below the mean of controls selected as possessing estimated population-average ability on the same two tests (Davis et al., under review). Well under 1% of the population would be likely to achieve super-recognition criteria on all four components.

Piloting tests with superior-face-recogniser samples.

To establish super-recogniser thresholds on the test battery, super-recognisers are invited to pilot tests if they have scored a) a minimum of 95 out of 102 on the CFMT+ (Russell et al., 2009), a test used in the vast majority of super-recognition research; while this score, the highest published research minimum threshold for super-recognition was achieved by approximately 2% of participants in a representative UK sample (Bobak Pampoulov, & Bate, 2016); b) a maximum score of 40 out of 40 on the simultaneous *Glasgow Face Matching Test* (GFMT) (Burton, White, & McNeill, 2010); and c) a score above the mean of super-recognisers achieving criteria a and b on a second ability-verifying short-term old-new face recognition test (different tests have been employed here).

As noted, some super-recognisers are poor at simultaneous face matching, suggesting dissociations between the perceptual and memorial mechanisms of superior face processing. The requirement to achieve maximum scores on the GFMT ensures only super-recognisers who also possess good face matching abilities are invited to pilot research. All tests have been piloted by over 100 super-recognisers.

With police and business projects, clients are advised to select individuals whose performances on the four categories are consistently within the typical super-recognition range (i.e. those whose scores are above or close to the pilot super-recogniser mean). Such consistency is rare. In one recent police project, well over 4,000 participants started an 11-test process; fewer than 20 achieved the four-component threshold.

Conclusions

Most of the University of Greenwich super-recogniser tests have been inspired by the growing research literature that followed Russell et al. (2009). Nevertheless, there will always be a risk of generating circular criteria for any skill, when thresholds are defined by participants who have achieved similar criteria on roughly similar tests. Test validity can only be assessed by evaluating workplace performances. Even then, a genuine police super-recogniser may rarely identify a suspect from CCTV, owing to never viewing appropriate images. It is also never possible to predict how any individual might perform at real-world tasks. Nevertheless, the pilot tests used with MET police possessed good predictive power given the later crime detection successes; and we have extended business contracts due to satisfaction with super-recogniser staff. The tests are regularly updated, partly driven by super-recogniser feedback. Currently, the decision making of the best computerised face recognition systems are enhanced when combined with those of super-recognisers (Philips et al., 2018). With the rapid improvement of face recognition and video search algorithms, occupational roles for super-recognisers in future may change. Research will therefore be required to generate new tests to select for any additional skills required.

References

Bate, S., Frowd, C., Bennetts, R., Hasshim, N., Murray, E., Bobak, A. K., Wills, H., & Richards, S. (2018). Applied screening tests for the detection of superior face recognition. *Cognitive Research: Principles and Implications, 3*(1), 22. https://doi.org/10.1186/s41235-018-0116-5

Belanova, E., Davis, J. P., & Thompson, T. (2018). Cognitive and neural markers of super-recognisers' face processing superiority and enhanced cross-age effect. *Cortex, 98*, 91-101. https://doi.org/10.1016/j.cortex.2018.07.008

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

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). doi:10.3389/fpsyg.2016.01378

Bretfelean, L. D., & Davis, J. P. (2017). Super-recognition: Long term reliable memory of unfamiliar faces following a 'fleeting glimpse'. *Undergraduate Prize Winner, British Psychological Society: Cognitive Section Annual Conference*, 30 August 2017.

Bruce, V., Henderson, Z., Greenwood, K., Hancock, P., Burton, A. M., & Miller, P. (1999). Verification of face identities from images captured on video. *Journal of Experimental Psychology: Applied, 5*, 339-360. doi: 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, 7*, 207-218. doi:10.1037/1076-898X.7.3.207

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

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. DOI: 10.1002/acp.3260

Davis, J. P., Lander, K., & Jansari, A. (2013). I never forget a face. *The Psychologist, 26*(10), 726-729. http://thepsychologist.bps.org.uk/volume-26/edition-10/i-never-forget-face

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. DOI: 10.1002/acp.3405

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., & Davis, R. E. (2010). Computer assisted facial comparison of photographic images taken from full-face and profile viewpoints. *Forensic Science International, 200*, 165-176. DOI: 10.1016/j.forsciint.2010.04.012

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

LArge Scale Information Exploitation of Forensic Data (LASIE) (2014). European Commission 7th Framework Programme. SEC-2013.1.6-1: 607480, May 2014. http://www.lasie-project.eu/

Medeiros, J. (2018). How police super-recognisers cracked the Russian novichok case. *Wired*, 8 September 2018, https://www.wired.co.uk/article/salisbury-novichok-poisoning-russia-suspects

Noyes, E., Phillips, P. J., & O'Toole, A. J. (2017). What is a super-recogniser? In M. Bindemann & A. M. Megreya (eds.), *Face processing: Systems, Disorders, and Cultural Differences*. New York, NY: Nova.

O'Keefe, P. (2016). The detectives who never forget a face. *The New Yorker*, 22 August 2016, http://www.newyorker.com/magazine/2016/08/22/londons-super-recognizer-police-force

Phillips, P. J., Yates, A. N., Hu, Y., Hahn, C. A., Noyes, E., Jackson, K. et al. (2018). Face recognition accuracy of forensic examiners, super-recognizers, and face recognition algorithms. *Proceedings of the National Academy of Science (PNAS).* http://www.pnas.org/content/early/2018/05/22/1721355115

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

Venkataramanan, M. (2015). The super power police now use to tackle crime. *BBC Future,* http://www.bbc.com/future/story/20150611-the-superpower-police-now-use-to-tackle-crime, 11 June 2015