Masters of disguise: Super-recognisers' superior memory for concealed faces

Josh P Davis, Donata Tamonytė
Department of Psychology, Social Work and Counselling,
University of Greenwich
London, UK
Email: j.p.davis@gre.ac.uk

***Abstract* —The deployment of police super-recognisers (SRs) with exceptional face recognition ability, has transformed the manner in which some forces manage CCTV evidence. In London, SRs make high numbers of suspect identifications, sometimes of suspects in disguise. In two experiments measuring immediate and one-week memory of faces in disguise, SRs were more accurate and confident than controls at correctly identifying targets, and ruling out faces not seen before. Accuracy and confidence were highest when targets wore no disguise, followed by hat and plaster, sunglasses, and balaclavas respectively. Even in the balaclava condition, SR performance was more accurate than chance. These findings join an accumulating body of empirical evidence demonstrating that SRs possess wide-ranging enhanced face processing abilities, and their deployment should complement ever improving computerised face recognition systems.**

1. Introduction

The study of *super-recognisers* (SR) who possess exceptional face recognition ability [1-8], has enhanced knowledge of the wide ability spectrum in the population. Interest in SRs, in the top 1-2% of ability, is driven by policing and security implications. Following the 2011 London Riots, SR police identified a third of the 5,000 convicted rioters from CCTV [9]. A full time London police SR unit was also established, substantially enhancing suspect identification rates [4, 5, 8]. Superior face processers also work in passport offices, complementing face recognition algorithms in identifying fraudulent applications [e.g. 10]. A large body of research aims to improve face recognition algorithms and develop security systems that best interact with human operators. Parallel research on SRs is essential, as future identity decisions may be made by machine-human combinations. Understanding the limits of SR abilities will therefore assist face recognition algorithm developers.

Not surprisingly, disguise reduces human and computer face recognition accuracy [e.g., 11], and yet London rioters were often disguised with only the eyes visible. Other cues (build, clothing, gait) may facilitate identification, and some rioters, tracked through different CCTV camera feeds, were videoed removing disguises. Sometimes, the disguise was never removed on camera.

Identified suspects are mainly familiar to police SRs, and not surprisingly, familiar faces are better recognised than unfamiliar faces, which theories suggest are processed using different cognitive mechanisms [12, 13]. Unfamiliar face identification is in contrast unreliable [14, 15], yet most London SR Unit identifications are of *unfamiliar* suspects, matched across different crime footage.

Familiar face recognition mainly draws on internal features (eyes, mouth), whereas external feature extraction (face shape, hairstyle) drives unfamiliar face recognition [16], explaining why recognition of a friend after hairstyle change is easy, in contrast to changes in someone encountered less frequently. SRs however, appear to 'learn' novel faces more effectively, extrapolating facial identity across different viewpoints after brief exposures [4]. As such, disguises covering different facial features may differentially effect identification by individuals differing in face recognition ability. This has policing implications, as beliefs that disguised face image identification is impossible may be made at early investigative stages by someone with 'average ability', leading to case closure, when in fact identification by SRs *might* be possible.

To investigate this, in Experiment 1, SRs and 'average-ability' controls, completed a 40-trial *Disguised Face Memory Test*, in which a single target face in no disguise, sunglasses, hat and plaster, or balaclava, was followed by an array of 10 undisguised faces. The disguises were suggested by police as exemplars of recent cases. In Experiment 2, SRs and controls viewed a 1 min target video in similar disguise conditions, and at least one week later viewed a video line-up. In both experiments, in half the trials the target was present (TP), half were target-absent (TA). Correct target identifications in TP trials were categorised as *hits*. Correct TA trial responses in were categorised as *correct rejections (CR)*. Here, participants correctly identified the target as absent from the array/line-up.

To hypothesise, SRs were predicted to outperform controls on all outcomes. Disguises, particularly the balaclava as it covered most of the face, were expected to reduce accuracy, and in Experiment 2, longer delays were expected to additionally impact accuracy [see 17]. However, the sunglasses, and the hat and plaster only partly obscured internal, or external facial features; and therefore no specific predictions were made as to the relative impact of these disguises on SR and control TP or TA accuracy.

2. Experiment 1

2.1 Method

2.1.1 Design

A 2 (independent measures: *Group*: SR, control) x 4 (repeated measures: *Disguise*: no disguise, sunglasses, hat and plaster, balaclava) x 2 (repeated measures: target presence: target present: TP, target absent: TA) mixed design was employed. The dependent variables were TP trial hits (rates of correct target identifications), and correct rejections (CR; rates of correctly responding the target was not present) of TA trials.

2.1.2 Participants

Participants had contributed to the first author's previous unpublished research, and had volunteered for more. Those meeting SR and control criteria were invited by e-mail to take the online Qualtrics platform tests. [1] Most previous SR research has employed the *Cambridge Face Memory Test*: *Extended* (CFMT+) [7], with a minimum SR threshold of 90/102 (2 SD above the mean) (= top 2%) [e.g., 1-3, 6]. However, a recently recommended higher criterion was used here (min = 95/102) [see 18] for SRs (n = 106, 66.0% female; 82.1% white; aged 18 - 63; M = 35.7; SD = 10.1). Controls (n = 101, 51.5% female; 89.2% white; aged 18 - 72; M = 43.4; SD = 13.4) had scored within 1 SD (83.0 – 58.4) of the estimated population mean (CFMT+: M = 70.7, SD = 12.3 [see 18]).

As expected, SR's CFMT+ scores (M = 96.4, SD = 1.3) were higher than controls (M = 73.6, SD = 4.8), $t(205)$ = 47.04, $p < .001$, Cohen's d = 6.48. Unexpectedly, SRs were also younger, $t(204)$ = 4.66, p = .003; and mainly female, $\chi^2(1, 201)$ = 6.31, p = .012.

2.1.3 Materials and procedure

Disguised Face Memory Test [DFMT; adapted from 12]: Before starting, participants provided consent, and were correctly informed that some faces would be in disguise, and that half the trials would be TA. Participants attempted to familiarise themselves to a series of 40 single video stills of a white male target face for 8-sec in one of four disguise conditions (see *Fig. 1*). They almost immediately attempted to identify the target from an array of 10 white male undisguised faces. Participants clicked on a number associated with each face (1-10) or responded 'not present', and gave a decision confidence rating (0: guessing-100: highly confident). All images were high-quality, taken on the same day and half the arrays contained a frontal facial photograph of the target within an array of nine foils (TP), or with the target replaced by an extra foil (TA). Each array had been constructed by selecting target-similar faces from a database of 200 trainee police officers. Disguises were added using GIMP software. [2] Each disguise was shown to each participant in Stage 1 an equal number of times: TP = 5 trials in each disguise condition; TA = 5 trials in each disguise condition). After completing two practice trials, this test consisted of 40 trials with stimuli randomly ordered and fully counterbalanced across 8 versions.

Hit rates were defined as the proportion of correct TP identifications in each disguise condition (out of 5). CR rates were the proportion of CRs in each disguise condition in TA trials (out of 5). Mean decision confidence rates in each condition were also calculated.

Fig. 1: Examples of artificially edited disguises in Experiment 1, clockwise from top left: no disguise; sunglasses; balaclava; hat and plaster (for copyright reasons original images from [13] are not depicted)

2.2 Results

A series of 2 (group: SR, control) x 4 (disguise: no disguise; sunglasses; hat and plaster; balaclava) mixed ANOVAs were conducted on each outcome (hits, CRs, confidence). Post-hoc comparisons were performed using the Bonferroni correction ($p < .05$). Where appropriate the Greenhouse-Geisser adjustment was applied to violations of sphericity. Occasional missing data (confidence) were treated as such on specific analyses only.

Accuracy (hits and CRs) (see *Figs. 2,a-b*)*:* There were significant effects of *group* on hits, $F(1, 205)$ = 16.76, $p < .001$, η^2 = .076; and CRs, $F(1, 205)$ = 32.43, $p < .001$, η^2 = .137. SRs outperformed controls on both accuracy measures, although CR rate effect sizes were stronger.

[1] www.qualtrics.com

[2] https://www.gimp.org (Version 2.8)

There were significant effects of *disguise* on hits, $F(2.87, 588.91) = 90.67$, $p < .001$, $\eta^2 = .307$; and CRs, $F(2.87, 587.50) = 51.54$, $p < .001$, $\eta^2 = .200$. Post-hoc paired comparisons found that hits and CRs were highest in the no disguise condition, followed successively and significantly by the hat and plaster, sunglasses, and balaclava condition respectively (all *p's* < .05).

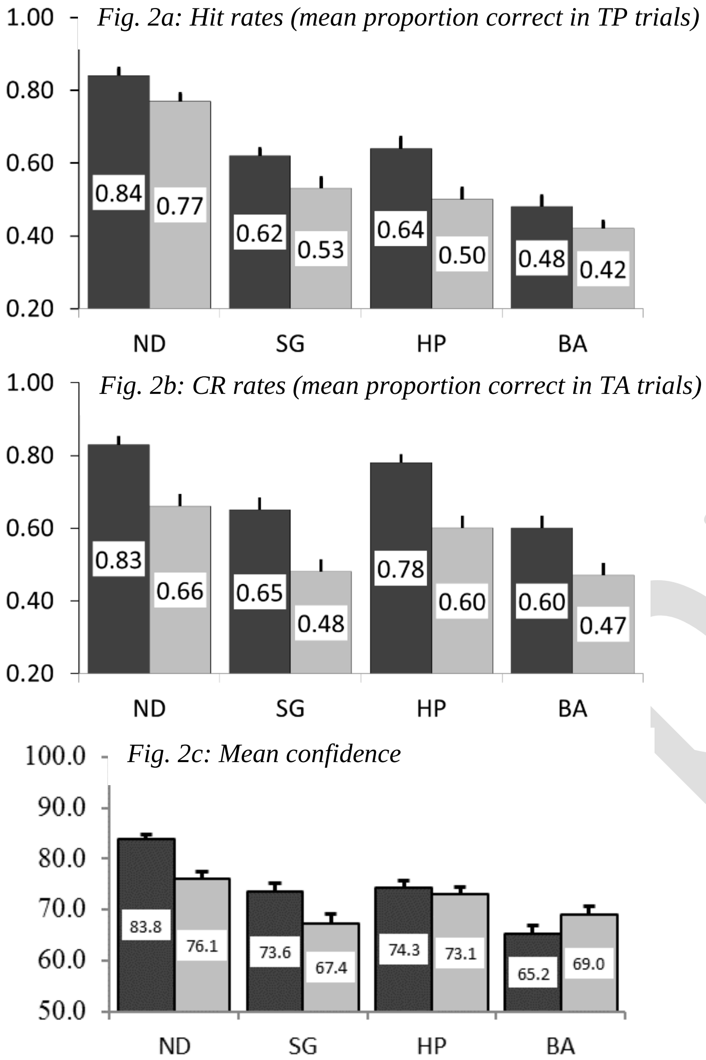

Fig 2. a) Mean hits (proportions); b) CRs (proportions); c) confidence on the DFRT (ND = no disguise; SG = sunglasses; HP = hat and plaster; BA = balaclava) separately for SRs (black bars) and controls (grey bars) in Experiment 1 (error bars = standard error of the mean: SEM)

Mean confidence (*Fig. 2c*): A 2 (group) x 4 (disguise) x 2 (condition: TP, TA) ANOVA revealed significant main effects of *group*, $F(1, 142) = 24.93$, $p < .001$, $\eta^2 = .149$; SR's were more confident than controls; *disguise*, $F(2.77, 392.62) = 57.49$, $p < .001$, $\eta^2 = .288$; confidence was significantly highest in the no disguise condition, followed by the hat and plaster, sunglasses, and balaclava condition respectively (all *p's* < .05); and *presence*, $F(1, 142) = 8.08$, $p = .005$, $\eta^2 = .054$; confidence was higher in TP trials.

There was a disguise x presence interaction, $F(2.72, 386.73) = 12.77$, $p < .001$, $\eta^2 = .083$. In TP trials, confidence was highest in the no disguise condition and lowest in the balaclava condition ($p < .001$); there were no significant differences between the sunglasses, and the hat and plaster conditions ($p > .2$). With TA trials, confidence was highest in the no disguise condition, followed by the hat and plaster condition ($p < .05$); there were no significant differences between the sunglasses and the balaclava conditions ($p > .05$). The other interactions were not significant ($p > .05$).

2.3 Experiment 1 Discussion

On all outcomes, and regardless of disguise or target presence in Experiment 1, SRs outperformed controls and were also more confident. As first phase exposure time was only 8-sec, these findings are consistent with research demonstrating that SRs are more accurate at learning and identifying novel faces, and at ruling out faces not been seen before – skills that may draw on different memorial processes. Accuracy and confidence in all participants was adversely affected by disguise, with those covering the external features or the eyes impacting most (balaclava followed by sunglasses), supporting research suggesting that external features drive unfamiliar face recognition, and the eyes and internal features are more important as faces become familiarised [16]. Nevertheless, even in the hardest balaclava condition, hit rates in TP trials were well above chance levels (chance = 1/10 = 0.10) by both SRs (0.48) and controls (0.42). The implications are discussed below. However, the stimuli used in Experiment 1 had all been taken on the same day, were artificially disguised, and the delay between Stages 1 and 2 was brief. Most police investigations involve images taken at different times and recognition is normally required after longer delays. Experiment 2 addressed these issues.

3. Experiment 2

3.1 Method

3.1.1 Design

Participants viewed a target video in a disguise condition in Stage 1. At least one week later in Stage 2, they viewed a video line-up in a 2 (*Group*: SR, control) x 4 (*Disguise*: no disguise, sunglasses, hat and plaster, balaclava) x 2 (*Presence*: TP, TA) mixed design. The dependent variables were hits in TP conditions, and CRs in TA conditions. Confidence and delay were also analysed.

3.1.2 Participants

Participants were recruited in the same manner as Experiment 1. Of the participants who completed Stage 1, more controls than SRs failed to contribute to Stage 2 (drop outs: SRs = 30 out of 99 Stage 1 starters: 30.3%; controls = 78 out of 143: 54.5%), $\chi^2(1, 242) = 13.91$, $p < .001$, $\phi = .240$. However, Stage 1 disguise condition had no impact on whether participants completed Stage 2 or not ($p > .2$).

Participants who completed Stage 2 were SRs ($n = 69$, 53.6% female; 78.3% white; aged 19-61; $M = 33.7$ years; $SD = 9.7$) and controls ($n = 65$, 50.8% female; 83.1% white; aged 18 - 65; $M = 36.4$ years; $SD = 14.6$).

SRs scored significantly higher on the CFMT+ ($M = 96.3$, $SD = 1.3$) than controls ($M = 74.7$, $SD = 5.6$), $t(132) = 31.21$, $p < .001$, Cohen's d = 5.34. Unlike Experiment 1, there were no differences in age, $t(124) = 1.23$, $p > .2$; or gender proportions, $\chi^2(1, 132) < 1$.

3.1.3 Materials

Video stimuli (Stage 1): The single target was filmed in four 10-sec colour video clips wearing no disguise, sunglasses, hat and plaster, or balaclava (*see Fig. 3*). In each clip, he faces the camera, turns right, forward, left, then faces the camera again. The video was looped so that duration of each final video was one minute.

Video line-ups (Stage 2): Four PROMAT™ [3] video line-ups were created (two TP and two TA to vary target and foil order – this had no effects, $p > .2$) by a police officer, following legal codes of practice in England and Wales [19]. High target-similarity foils were visually selected from a short list after entering keywords matching the target's description (e.g., age, gender) into a database. Each line-up consists of nine 15-sec sequentially presented colour head-and-shoulders video clips. Members face the camera; look left, right and forward. TP videos contained the target and eight foils; TA clips nine foils. The sequence repeats twice, with line-up numbers displayed (see [20] for a video of the procedure). There was a delay of 12 months between filming the target videos for Stage 1 and 2.

Fig. 3. Stills from the Experiment 2 Stage 1 videos depicting the target in each disguise condition (from left: no disguise, sunglasses, hat and plaster, balaclava)

3.1.4 Procedure

In Stage 1, participants viewed one of the randomly assigned 1-min target videos and recorded their *confidence* in being likely to later recognise the suspect (0: highly unlikely-100: very likely). An e-mail invite was sent one week later for Stage 2, during which participants were warned that the suspect 'may or may not be present', and were randomly assigned to view a password protected TP or TA video line-up. They then selected a number associated with each line-up member (1-9) which corresponded to the target identity in the line-up if present or rejected the line-up if absent, and provided confidence (0: guessing-100: absolutely certain). Hit, CR and confidence measures were calculated in the same manner as in Experiment 1, except participants made a single identification decision only.

3.2 Results

The mean delay between Stage 1 and 2 was 10.3 days ($SD = 12.3$). Delay for SRs and controls did not differ, $t(132) = 1.43$, $p = .155$. For all participants, regardless of disguise or target presence, there was no correlation between CFMT+ scores and Stage 1 confidence, $r(131) = 0.14$, $p = .116$. However CFMT+ scores and Stage 1 confidence correlated with Stage 2 accuracy, $r_{CFMT+}(134) = 0.33$, $p < .001$; $r_{Stage\ 1\ conf}(129) = 0.29$, $p = .001$; and confidence, $r_{CFMT+}(129) = 0.27$, $p = .002$; $r_{Stage\ 1\ conf}(127) = 0.34$, $p < .001$.

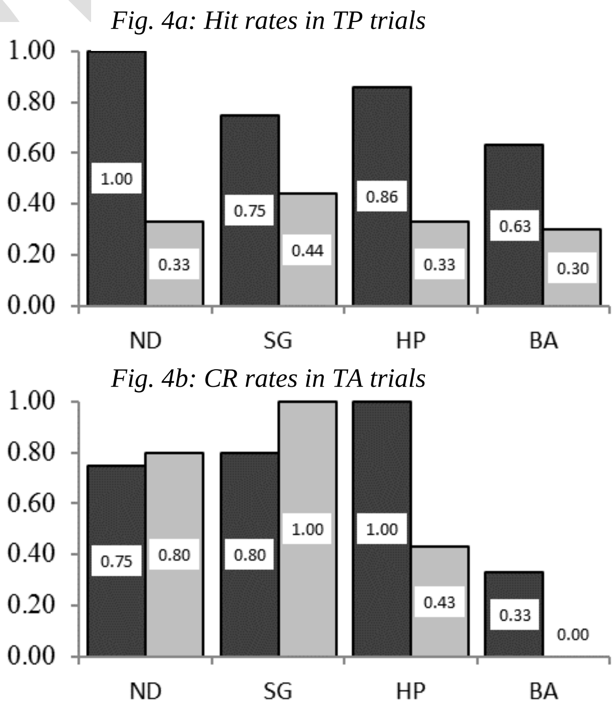

Fig 4. a) Mean hits (proportions); b) CRs (proportions) in Experiment 2 (ND = no disguise; SG = sunglasses; HP = hat and plaster; BA = balaclava) separately for SRs (black bars) and controls (grey bars) (error bars = SEM)

[3] Promat Envision International, Nelson, Lancashire, UK

Fig. 4a-b displays the mean hit and CR rates in each disguise condition for SRs and controls.

Accuracy: Due to low expected counts in some conditions violating statistical assumptions of a three-way loglinear analyses, it was not possible to combine group and disguise conditions into single analyses in TP and TA conditions. Therefore two separate sets of analyses were conducted to increase statistical power.

In the first, TP and TA trials were combined, and a series of 2 (group) x 2 (accuracy: correct, incorrect) chi-squared tests found that SRs were more accurate than controls in the no disguise, $\chi^2(1, 29) = 4.55$, $p = .033$, $\phi = .396$; hat and plasters, $\chi^2(1, 26) = 8.33$, $p = .004$, $\phi = .566$; and balaclava conditions, $\chi^2(1, 24) = 4.20$, $p = .040$, $\phi = .418$; but not the sunglasses condition, $\chi^2(1, 55) = 1.21$, $p = .271$, $\phi = .149$.

For the second analysis, the SR and control data were pooled, and a 4 (disguise) x 2 (accuracy) chi-squared test was also significant, $\chi^2(3, 134) = 11.00$, $p = .012$, $\phi = .287$. The only significant post doc test was that accuracy in the balaclava condition was lower than the other three conditions ($p < .05$), which did not differ ($p > .05$).

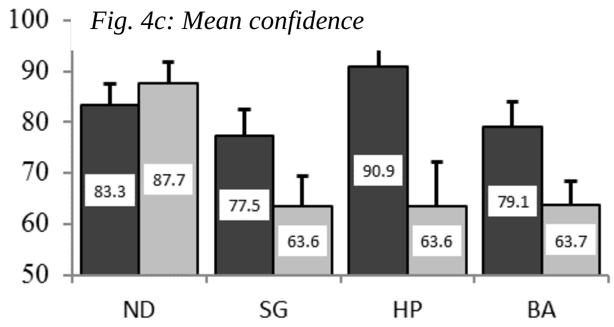

Fig 4. c) Mean confidence in Experiment 2 (ND = no disguise; SG = sunglasses; HP = hat and plaster; BA = balaclava) separately for SRs (black bars) and controls (grey bars) (error bars = SEM)

Confidence: A 2 (group) x 4 (disguise) x 2 (target presence) ANOVA conducted on confidence found a significant group effect, $F(1, 113) = 8.75$, $p = .004$, $\eta^2 = .072$; SRs were more confident than controls. The disguise main effect was significant, $F(3, 113) = 2.84$, $p = .041$, $\eta^2 = .070$; The only significant post-hoc comparison was that confidence in the no disguise condition was higher than the sunglasses condition ($p < .05$). The target presence main effect and all interactions were not significant ($p > .15$).

Delay: There was no significant correlation between delay and accuracy although the effects were in the expected negative direction, $r(134) = -.14$, $p = .101$.

3.3 Experiment 2 Discussion

As with Experiment 1 and expectations, SRs outperformed controls in three of the four disguise conditions (no disguise, hat and plaster, balaclava). The only non-significant comparison was in the sunglasses condition, mainly because the controls unexpectedly had the highest rates of accuracy in that condition. However, even when combined with the SR data, the accuracy rates in the sunglasses condition did not significantly differ from accuracy in the no disguise or hat and plasters conditions. Nevertheless, consistent with Experiment 1, accuracy in the balaclava condition was significantly worse than in the other three conditions, although rates of correct target identifications by SRs (0.63) and controls (0.30) in TP conditions were again above chance despite a delay of at least one week (chance = 1/9 = 0.11).

SRs were also more confident than controls in all disguise conditions, although unexpectedly, the only significant disguise effect was that confidence was highest in the no disguise condition, and lowest in the sunglasses condition.

It is noteworthy however, that rates of correct line-up rejections in the TA no disguise and sunglasses conditions were at, or close to ceiling by both groups, and SR's CR rates were additionally at ceiling in the hat and plasters condition, suggesting a bias to (correctly) respond not present under conditions of uncertainty. However, it is important to note that the conclusions from this experiment may be limited as all participants were exposed to the same target individual in Stage 1, and each encountered only one trial. Further research is required to examine whether effects would generalise to other actors (and gender, ethnicity etc.).

4. General Discussion and Conclusions

Previous research has shown that SRs possess superior abilities at short-term unfamiliar face recognition [e.g. 1, 7], simultaneous unfamiliar face matching [2, 8], recognising distorted 15-year-old famous face images [4] and spotting a face in a crowd [5]. The current research was the first to demonstrate that this advantage transfers to immediate and delayed (one-week) recognition of unfamiliar disguised faces. In both experiments, using a robust threshold for SR group membership, regardless of disguise condition, SRs outperformed 'average-ability' controls, and expressed higher confidence in their decisions. Although not all effects were significant in Experiment 2, mainly due to low statistical power, the pattern of disguise results was similar to Experiment 1, supporting theories suggesting that disguise covering external features is most detrimental to unfamiliar face recognition [16] regardless of ability.

Identification accuracy was slightly reduced by wearing a hat and plaster (a possible tactic used to avoid detection by face recognition algorithms). However, internal features, particularly eyes, are the most important for face familiarisation [21] and may negatively impact face learning as sunglasses covering the eyes reduced performance more. Nevertheless, the strongest reductions in identification accuracy and confidence were when targets wore balaclavas covering almost all of the face but the eyes (*Figs. 1, 3*), and yet even in this condition, SRs and controls were far more accurate than chance levels alone. After a week, SRs were twice as likely as controls to be correct in this condition.

There are implications here for computer scientists testing face recognition algorithms, or systems to assist police review CCTV footage [e.g. 22]. Most research testing algorithms employs 'average' ability humans as controls [23]. These algorithms surpass average-ability humans with high and medium quality footage, although humans outperform the top systems with lower quality footage - which is common with much CCTV evidence. However, it is not clear whether the same conclusions would be made when comparing computer systems with SRs.

There are some limitations to this research. Most suspect identifications from CCTV are made by those familiar with suspects, and images may be of lower quality than those used here. Even though SRs may be more accurate than controls at unfamiliar face recognition of heavily disguised faces, performance was not at 100%, and if giving identification evidence in court, the risks of error should be acknowledged. Nevertheless, it is very clear that the effective deployment of SRs in police forces worldwide should have a positive impact on crime detection and homeland security.

Acknowledgments

The project was funded by the LArge Scale Information Exploitation of Forensic Data (LASIE) project (*European Commission 7th Framework Programme. SEC-2013.1.6-1: 607480).* Thanks go to Paul Smith (Metropolitan Police Service), Monika Durova, Luiza-Diandra Bretfelean, and Lauren Jensen for participant recruitment assistance.

References

[1] A. K. Bobak, R. J. Bennetts, B. A. Parris, A. Jansari, and S. Bate, An in-depth cognitive examination of individuals with superior face recognition skills, *Cortex*, vol. 82, 2016, pp. 48-62.

[2] A. K. Bobak, P. J. B. Hancock, and S. Bate, Super-Recognizers in action: evidence from face matching and face memory tasks, *Applied Cognitive Psychology,* vol. 30, 2016, pp. 81-91.

[3] A. K. Bobak, B. A. Parris, N. J. Gregory, R. J. Bennetts, and S. Bate, Eye-movement strategies in developmental prosopagnosia and "super" face recognition, *Quarterly Journal of Experimental Psychology,* vol. 70, 2017, pp. 201-217.

[4] J. P. Davis, K. Lander, R. Evans, and A. Jansari, Investigating predictors of superior face recognition ability in police super-recognisers, *Applied Cognitive Psychology,* vol. 30, 2016, pp. 827–840.

[5] J. P. Davis, F. Treml, C. Forrest, and A. Jansari (unpublished). Identification from CCTV: Identification from CCTV: Assessing police super-recognisers ability to spot faces in a crowd and susceptibility to change blindness. *Applied Cognitive Psychology*

[6] R. Russell, G. Chatterjee, and K. Nakayama, Developmental prosopagnosia and super-recognition: no special role for surface reflectance processing, *Neuropsychologia,* vol. 50, 2012, pp. 334-340.

[7] R. Russell, B. Duchaine, and K. Nakayama, Super-recognizers: people with extraordinary face recognition ability, *Psychonomic Bulletin & Review,* vol. 16, 2009, pp. 252–257.

[8] D. J. Robertson, E. Noyes, A. J. Dowsett, R. Jenkins, and A. M. Burton, Face recognition by Metropolitan Police super-recognisers, *PloS One,* vol. 11, 2016, e0150036–8.

[9] J. P. Davis, K. Lander, and A. Jansari, I never forget a face, *The Psychologist,* vol. 26, 2013, pp.726-729.

[10] D. White, J. D. Dunn, A. C. Schmid, and R. I. Kemp, Error rates in users of automatic face recognition software, *PloS One,* vol. 10, 2015, e0139827–14.

[11] K. E. Patterson, and A. D. Baddeley, When face recognition fails, *Journal of Experimental Psychology: Human Learning and Memory*, vol. 3, 1977, pp. 406-417.

[12] V. Bruce, and A. Young, Understanding face recognition, *British Journal of Psychology*, vol. 77, 1986, pp. 305-327.

[13] R. A. Johnston, and A. J. Edmonds, Familiar and unfamiliar face recognition: a review, *Memory,* vol. 17, 2009, pp. 577-596.

[14] V. Bruce, Z. Henderson, K. Greenwood, P. J. B. Hancock, A. M. Burton, and P. Miller, Verification of face identities from images captured on video, *Journal of Experimental Psychology: Applied,* vol. 5, 1999, pp. 339-360.

[15] J. P. Davis, and T. Valentine (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,* 2015, Chichester: Wiley-Blackwell, pp. 211-238.

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

[17] K. A. Deffenbacher, G. H. Bornstein, E. K. McGorty, and S. D. Penrod, Forgetting the once-seen face: Estimating the strength of an eyewitness's memory representation, *Journal of Experimental Psychology: Applied*, vol. 14, 2008, pp. 139–150.

[18] A. K Bobak, P. Pampoulov, and S. Bate, Detecting superior face recognition skills in a large sample of young British adults, *Frontiers in Psychology*, vol. 7, 2016, p. 1378.

[19] Police and Criminal Evidence Act (PACE) (1984). Codes of Practice (Code D). (2017). Home Office.

[20] J. P. Davis, A. C. Maigut, D. Jolliffe, S. Gibson, and C. Solomon, Holistic facial composite creation and subsequent video line-up eyewitness identification paradigm. *Journal of Visualized Experiments,* vol. 106, 2015, e53298.

[21] C. O'Donnell, and V. Bruce, Familiarisation with faces selectively enhances sensitivity to changes made to the eyes, *Perception,* vol. 30, 2001, 755-764.

[22] LArge Scale Information Exploitation of Forensic Data (LASIE) project (*European Commission 7th Framework Programme. SEC-2013.1.6-1: 607480)*

[23] A. J. O'Toole, X. An, J. Dunlop, V. Natu, and P. J. Phillips, Comparing face recognition algorithms to humans on challenging tasks, *ACM Transactions on Applied Perception*, vol. 9, 2012, p. 16.