TooManyEyes: Super-recogniser directed identification of target individuals on CCTV

ICDP 2017: 8th IET International Conference of Imaging for Crime Detection and Prevention

M.L. Durova [1], A. Dimou [2], G. Litos [3], P. Daras [4], & J.P. Davis [5]

[1, 5] Department of Psychology, Social Work, and Counselling, University of Greenwich, UK, j.p.davis@gre.ac.uk
[2-4] Information Technologies Institute, Centre for Research and Technology Hellas, Thessaloniki 57001, Greece
Contact details: [1] M.Durova@greenwich.ac.uk, [2] dimou@iti.gr, [3] daras@iti.gr, [4] litos@iti.gr, [5] j.p.davis@gre.ac.uk
For correspondence: j.p.davis@gre.ac.uk

Keywords: super-recognisers; face recognition; CCTV; face matching

Abstract

For the current research, a 'Spot the Face in a Crowd Test' (SFCT) comprising six video clips depicting target-actors and multiple bystanders was loaded on TooManyEyes, a bespoke multi-media platform adapted here for the human-directed identification of individuals in CCTV footage. To test the utility of TooManyEyes, police 'super-recognisers' (SRs) who may possess exceptional face recognition ability, and police controls attempted to identify the target-actors from the SFCT. As expected, SRs correctly identified more target-actors; with higher confidence than controls. As such, the TooManyEyes system provides a useful platform for uploading tests for selecting police or security staff for CCTV review deployment.

1 Introduction

A growing body of recent literature has been devoted to understanding individual differences in face recognition, and the core attributes of people with exceptional face memory skills [e.g. 1-5]. At one end of the ability spectrum are *Developmental Prosopagnosics* (DPs) who may be adversely affected by a neurodevelopmental impairment in processing facial stimuli but display no apparent neurological damage [6, 7]. The antitheses are *Super-Recognisers* (SRs), who compared with the typical population, score higher on tests assessing face perception, simultaneous face matching, and familiar and unfamiliar face recognition [e.g., 1, 4, 5, 8], while performing at about the same level as controls at object recognition [3, 4], which suggests that SR, like DP is primarily face-specific.

Forensic interest in SRs is a consequence of police procedures being improved by deploying individuals who possess superior face recognition abilities. Closed Circuit Television (CCTV) surveillance is prevalent worldwide. Estimates suggest 4,285,000 cameras in the UK [9], 30,000,000 in the USA [10], and continued widespread international installation, producing ever higher quality images, is expected. CCTV footage of a crime scene provides a permanent record of events and of suspects involved and it can have a powerful impact in court. It is clear that to make the best evidential use of images; for instance, when conducting reviews of large quantities of footage, police could more effectively deploy SRs who are exceptionally good at identifying suspects from such evidence.

The 2011 London Riots first placed SR police in the public eye. From approximately 5,000 images, one officer identified 180 rioters, 20 officers identified 609, while in stark contrast, state-of-the-market face recognition software only made a single suspect identification, mainly due to the typically poor CCTV imagery taken at night, from above head height, with rioters in disguise [11, 12]. Following these successes, the London Metropolitan Police Service (MPS) created a team of full-time SRs who together annually identify well over 1,000 suspects from CCTV evidence. Based on additional evidence, many are subsequently convicted. Informative media reports of case successes have been published [13, 14], and some SRs have been deployed in operations in which, after memorising multiple 'wanted' suspect photos, they observe crowds in real time, often successfully identifying those suspects (e.g. Notting Hill Carnival with crowds of over 1 million).

To account for the variety of complex factors that could contribute to efficiency at CCTV review and the identification of criminals, in the MPS, SRs have been selected to operational roles drawing on their abilities based on, a) a track record of making multiple identifications of suspects from the MPS *Caught on Camera* wanted suspect website; or b) from exceptionally high performances across a battery of tests assessing their face recognition and simultaneous face matching abilities [e.g. 14]. However, CCTV review operations require alternative skills such as vigilance and close attention to detail that not all SRs may possess. Therefore, SR police have been tested on a *Spot the Face in the Crowd Test* (SFCT), where they have to search for unfamiliar target actors playing 'missing persons' in footage depicting crowds of bystanders [15]. The SFCT replicates a CCTV review task and draws on facial memory and simultaneous unfamiliar face matching, as well as concentration, vigilance, and attention on the task in hand – skills not assessed in typical face recognition and matching tests.

Recent research [15], testing full-time SRs from the MPS SR Unit in London (n = 7), other police (n = 92) and controls drawn from the public (n = 152) found positive relationships between performances on an early version of the SFCT, the

extended version of the *Cambridge Face Memory Test* (CFMT+) [5], used in most previous research to allocate SRs to SR groups [e.g., 1-5], and a Change Blindness test [16]. As a group, members of the MPS SR Unit outperformed the other groups on the three tests, although not all achieved SR criteria used in previous research. Nevertheless, on the 18 min SFCT, SR Unit police made more identifications of targets, with significantly higher confidence, while their rates of false positives of bystanders were lower, and were also made with significantly lower confidence, possibly reflecting experience of the difficulty in attempting to identify persons of interest across different CCTV feeds of sometimes indifferent quality footage. However, within each group there were large individual differences in performances on the different tests which may partly reflect that they draw on other skills outside face recognition and face matching, while overall accuracy also positively correlated with time to complete the SFCT, which for some participants took over 2 hours.

The aim of the current study was to examine whether *TooManyEyes*, a bespoke automated online remote system for human-directed identification of individuals in CCTV footage could offer an effective platform for the SFCT, which currently requires an administrator to provide face-to-face instructions to participants. Following extensive piloting, an abridged version of the SFCT was loaded on the *TooManyEyes* platform, and the performance of MPS SR Unit and other police exceeding SR criteria [5] was compared with controls, who were MPS non-SR police officers and staff, but who also regularly view CCTV footage. Based on previous research [13], it was hypothesised that SR police would outperform controls, as operationalised by higher rates of correct target identifications (hits), and fewer false positives of bystanders (i.e. incorrect identification of non-targets). Confidence in identification decisions was also expected to positively correlate with accuracy.

2 Method

2.1 Participants

Participants (n = 26 police officers and civilian staff) were selected by the Metropolitan Police Service (MPS). *Super-recognisers* were current or past members of the MPS SR unit, highly experienced at reviewing CCTV evidence and identifying suspects (n = 6; males = 6). Controls mainly worked within the Central Forensic Image Team, or the Forensics Department at the MPS (n = 20; males = 18).

2.2 Design

The abridged version of the SFCT consisted of six video clips (the original version [15] had 11 clips) in which four target-actors walking through the environment were depicted in four of the clips. Two clips were empty of targets, and one target appeared in two clips. The remaining three targets appeared in one clip each. Participants attempted to identify the targets in clips which included multiple continually varying numbers of bystanders. No crimes were depicted. The design of the study was between-participants in that SRs and controls were compared on rates of correct identifications of targets (hits), incorrect identifications of non-targets (false positives), correct rejections of 'empty' clips, as well as mean confidence in hits and false positives.

2.3 Materials

2.3.1 *Spot the Face in the Crowd Test (SFCT) (abridged)*

An abridged version of the SFCT was designed (see [13] for full description). This consisted of six separate video clips (labelled A-F) (total time = 8 min 16 sec). Each clip depicted two (n = 1 clip), one (n = 3 clips), or zero (n = 2 clips) actors. One actor appeared in two clips. Beforehand, participants were able to familiarise themselves to the four target-actors – depicted in four photos each, and could also simultaneously compare the photos to the footage during the test itself. Table 1 lists the play time of each clip, the number and 'letter' of each actor/actress depicted, the approximate number of bystanders (the estimated value was based on the actual number depicted at the mid-point of each clip), and the time on screen of each actor (sec). It also lists the mean performance on each clip (see below for explanation).

The abridged SFCT was administered through the *TooManyEyes* online platform. Participants had to individually register to the system and provide consent before proceeding further. After registration, participants were directed to a 'HOME' page consisting of, (1) an instructions video (explaining the system features); (2) three practice trials (two clips contained target-actors not included in the main test, while one clip was empty of target-actors), and, (3) the final test itself consisting of the six video clips of the abridged version of the SFCT.

2.3.2 *TooManyEyes*

TooManyEyes[1] is a multimedia focused micro-tasking framework that has been developed for crowd-sourcing applications. It is a web platform that allows the creation of applications, to which volunteers can participate from their web browsers. It is designed to cover all stages of the application management, including the creation, distribution and assessment of the results from a large-scale user base, through an intuitive interface. The *TooManyEyes* platform has been adapted to accommodate the needs of measuring the ability of a person to spot a face in the crowd by providing face recognition tests. The aim of *TooManyEyes* is to provide a toolbox that significantly simplifies the creation of the tests, their e-management and the assessment of the results.

Task Creation: With *TooManyEyes*, the creation of a test (a.k.a. *task*, as defined by crowd-sourcing terminology) is performed in a series of easy steps. A new task is created entering some basic details (Figure 1. *a*) and building it up from elementary testing blocks (micro-tasks). Each micro-task is built in three steps, namely selecting the people to

[1] http://toomanyeyes.net/

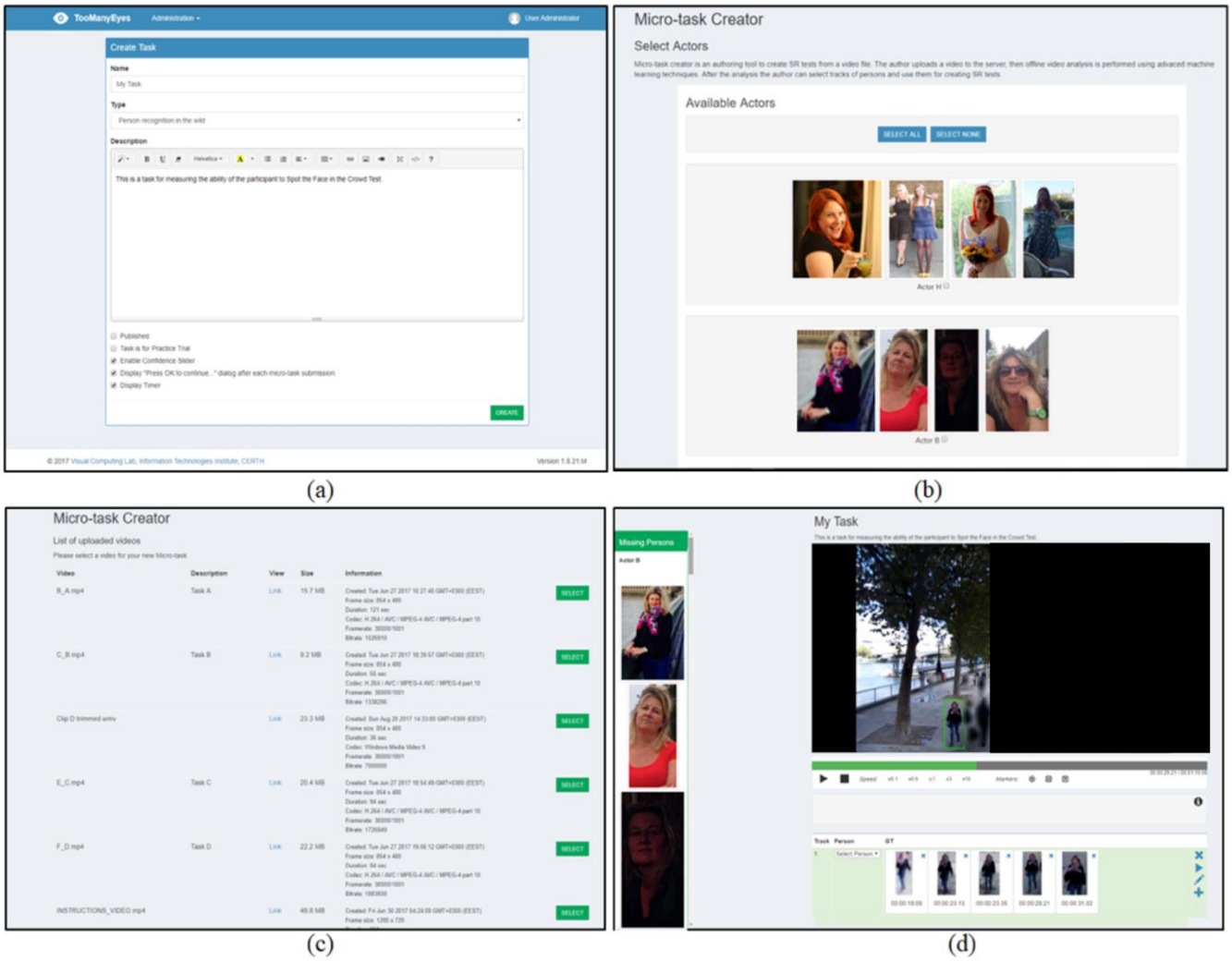

Figure 1: Required steps for creating a new task using the TooManyEyes platform.

recognize from a gallery of photos (Figure 1. *b*), selecting the relevant videos from a repository (Figure 1. *c*), and annotating the video with a simple and user-intuitive tool (Figure 1. *d*). The task creator supports all possible variations of the test, including target absent tasks (e.g. the video depicts no targets), or tasks where the targets are to be spotted multiple times in the same or different videos. A completed example can be seen in
Figure 2, in which the six tasks of the SFCT have been imported into the TooManyEyes system. Each clip has been listed as a separate task. It should be noted, however, that although the images loaded on the system depicted only Caucasian females, future use of the system can benefit from the implementation of racial and gender diversity in order to investigate cross-rate effects and error proneness in eyewitness identification [17].

Task management: Once created, the tasks have to be distributed to the participants. While this may seem a trivial task, it is very challenging when it comes to scaling the test to a large audience. For this purpose, the *TooManyEyes* platform integrates a management framework to simplify the procedure. It features two distinct roles, namely the administrator and the users (i.e. participants) of the platform. The administrator is able to create applications (a.k.a. *projects*) that may contain several tasks that will measure the ability of the participant to identify people in the crowd.

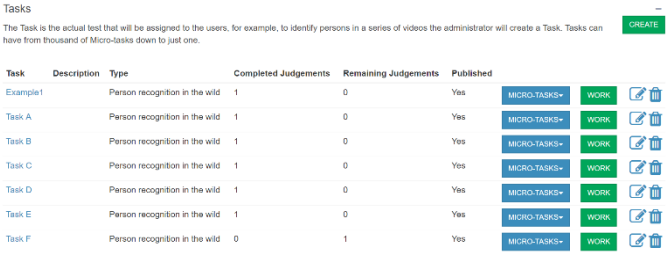

Figure 2: Screenshot showing task creation of the 6-video clip abridged version of the SFCT (See text for details)

The participants can be allocated to different *organizations*, while each organization can be assigned to one or more projects. Its members are notified in order to participate to the related tasks. The *TooManyEyes* platform provides the

administrator with considerable flexibility in the management of the organizations and the projects offering customizable settings.

In the current instance, the project was the abridged version of the SFCT and it was assigned to an organization created specifically for the test.

Task participation

The participant can sign up to the *TooManyEyes* platform providing some information for his/her identification. The amount of information provided depends on the needs of their organization and is customizable. It may range from personal information (name, organization, gender, mail, address etc.) to completely anonymous participation using a pre-distributed login code. Participants are required to provide informed consent, however, the terms of use may vary according to the organisation. Once a participant logs in, a list of projects eligible for participation is provided.

Selecting a test, participants are provided with written and visual instructions for the task they are invited to complete. Moreover, they are requested to complete practice tests to verify that they understand the task and the functionality of the platform. When completed, a task is de-activated to avoid any confusion.

Task assessment

The answers (a.k.a. *judgements*, as defined by crowdsourcing terminology) provided by the participants are collected centrally for further processing. Each judgement comprises the actual input given, the time required for the completion of the task, a confidence level for each answer, and possible comments pinpointing evidence leading to the provided answer (see Figure 4). All provided judgements are automatically assessed based on the ground truth that has been provided during the creation of the task. However, the administrator has the ability to review and correct (when and if needed) the assessments.

The *TooManyEyes* platform extracts the following statistics for each clip and in total: *Duration, Total Judgements, Hits, Misses, Correct Rejections, False Alarms* and *Confidence*. These statistics are also extracted at multiple levels, namely participant-wise, project-wise and organization-wise. The administrator can either view or extract the results from the platform for further analyses.

Participants can be provided with an overview of their final results in accordance with the policy of their organization. The policy can restrain them either from viewing the results, or viewing them only after the completion of all tasks.

2.3 Procedure

The study received ethical approval from the University of Greenwich Research Ethics Committee. Permissions to film videos at the tourist sites for the SFCT were provided by land owners, with the proviso that most locations and any bystanders were not to be depicted or identifiable in published articles. Signs warned bystanders of filming.

Participants were invited to complete the SFCT through the automated online remote system *TooManyEyes* across two different testing sessions with identical conditions/settings. The testing sessions, conducted on MPS laptops took place in an MPS police office. The first author was present during both sessions in case of any technical or other issues.

For the current research, participants first practiced on three clips. Entering the actual SFCT (abridged) test, they were then provided with a list of four photos of the four different targets to be identified in the SFCT video clips (see Figure 1b for examples of photos of some of the targets). Subsequently, they were provided with the SFCT video clips to examine. For their convenience, they could watch the video in slow/fast motion, skip parts, create loops or freeze the video. *TooManyEyes* supports multiple screens and full screen functionalities, although for the current research, target photos and the SFCT video clips were shown on a single laptop screen. If a participant believed they identified a target-actor, they marked it by drawing a bounding box around that person (see Figure 4 for a screen shot depicting some of the tools offered by the *TooManyEyes* system and a correctly identified target-actor boundary box). The platform saves the selection offering additional fields for target identification, confidence level and additional comments (see Figure 4). Participants can review and change their answers in multiple stages before finalizing their participation. In this case the participant has commented on the similarity of the scarf worn by the target in Figure 1b and the target-actor in Figure 4, demonstrating that identification decisions on the SFCT are not necessarily made on the basis of physical appearance alone.

Please note that the image shown in Figure 4 has been blurred for the purpose of making bystanders unidentifiable.

Analyses

Data were amalgamated and analysed using IBM SPSS for Windows of each participants' correct identifications of target-actors (hit rates), incorrect identifications of bystanders (False Positives: FPs), and correct rejections of clips empty of target-actors (Correct Rejections: CRs). Time taken to complete the abridged SFCT and mean confidence in hits and FPS were also analysed.

Table 1: Individual video clip (Task A-F), timing (min-sec), approximate bystander numbers (n), target-actors (a-d), time of each target-actor on screen in video clips (sec), mean hit rates (proportions), mean false positive rates (FPs) and CR rates to empty clips (proportions) on the revised SFCT (see text for explanation).

Video clip	A	B	C	D	E	F	
Time (min)	1.58	0.54	1.34	1.32	1.34	2.04	
Bystanders	12	14	28	19	12	17	
Actor	*b*	-	*d*	-	*c*	*a*	*b*
Screen Time (sec)	7	-	4	-	4	11	25
Hits	0.74	-	0.33	-	0.74	0.85	0.52
FPs	0.56	0.33	0.41	0.33	0.52	0.52	0.07
CRs	-	0.67	-	0.56	-	-	

3 Results

Table 1 displays individual clip (A-F) time (min), the number of bystanders at the mid-point of the clip, the actors shown in each clip (a-d), and the proportion of hits, correct rejections and false positives as a function of clip. Performances ranged from 93% of participants identifying Actor b in Clip A, to 33% of participants identifying Actor d in Clip C. The highest rates of FPs were also in Clip A. Table 2 depicts outcomes as a function of group. Scores of police SRs were higher than controls on hits, CRs, and FPs, and they also took longer to complete the test. A series of independent-measures t-tests compared the SRs and controls on these data. SRs' hit rates, $t(24) = 2.40$, $p = .025$, Cohen's $d = 1.14$, and confidence in hits, $t(24) = 2.30$, $p = .031$, Cohen's $d = 1.32$ were significantly higher than controls. No other effects were significant (all: $t < 1.30$, $p > .2$, Cohen's $d < 0.60$).

Table 2: Performance outcomes on the SFCT

	SRs (n = 6)		Controls (n = 20)	
	Mean	SD	M	SD
Completion Time (min)*	46.54	18.31	37.08	12.10
Proportion Hits	0.73	0.16	0.53	0.19
Hits Confidence	9.09	0.54	7.01	2.16
Proportion CRs	0.75	0.42	0.50	0.43
Number of FPs	2.17	1.94	3.35	1.95
FP Confidence	5.63	3.15	4.64	2.34

**Completion time data were not collected from all participants(n = 13)*

4 Discussion

This research demonstrated that the *TooManyEyes* system provides a suitable platform for the *Spot the Face in the Crowd Test* (SFCT), a different version of which has been employed to test police involved in CCTV review and other similar duties. As expected, consistent with previous research [15], police who have experience of working in a full time Metropolitan Police Service (MPS) CCTV review Super-Recogniser (SR) Unit, some of whom may possess exceptional face recognition ability, significantly outperformed police controls in terms of higher rates of correct identifications of target-actors depicted in the SFCT. SRs' confidence in correct identifications was also significantly higher than controls, again supporting predictions and in line with previous research [15]. However, probably due to low statistical power, from low participant numbers, no differences were found in rates of false bystander identifications (false positives) or correct rejections of empty clips, although the scores were in the expected direction.

Figure 4: Example screenshot from the SFCT in which a bounding box has been selected around the target-actor depicted in Figure 1b using the TooManyEyes system (with enlarged image below). The functionality of some of the tools provided with the TooManyEyes system is also depicted. Often moving footage can provide more identifiation information than a single frame as depicted here.

Feedback, obtained from the police participants following the testing sessions was highly positive, and it is clear that this methodology maps onto the types of task of police who are deployed to these activities. A more demanding SFCT could however be loaded for more detailed examination of abilities – for instance involving the recognition of faces of different

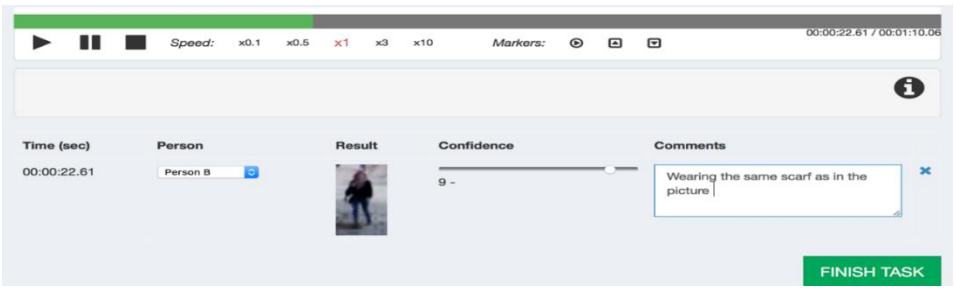

Figure 3: An example of the participants working bar where they have to select the letter of the identified person, their confidence, and any comment they might have.

ethnicities and genders [17] or faces in disguise, commonly found with CCTV footage of crime scenes, and also a task that SRs have also been shown to exceed average-ability controls at [18]. Indeed, the current test would not be sufficient for measuring individual differences in performance, as there were too few measurable outcomes to create a standardised version. In addition, a test measuring actor behavioural expectancies and/or threat detection analyses could be devised for loading on the system. Nevertheless, it is clear that the *TooManyEyes* system would provide a suitable platform on which to load a remote battery of different CCTV review tests for selecting international SR police to CCTV review operations and evidence gathering.

Acknowledgements

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 Diandra Bretfelean for assistance with testing participants.

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, A. Dowsett and S. Bate, "Solving the Border Control Problem: Evidence of Enhanced Face Matching in Individuals with Extraordinary Face Recognition Skills", PLOS ONE, vol. 11, no. 2, p. e0148148, 2016.

[3] 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.

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

[6] S. Bate and J. Tree, "The definition and diagnosis of developmental prosopagnosia", The Quarterly Journal of Experimental Psychology, vol. 70, no. 2, pp. 193-200, 2016.

[7] A. Damasio, H. Damasio and G. Van Hoesen, "Prosopagnosia: Anatomic basis and behavioral mechanisms", Neurology, vol. 32, no. 4, pp. 331-331, 1982.

[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] C. Norris, M. McCahill and D. Wood, "Editorial. The Growth of CCTV: a global perspective on the international diffusion of video surveillance in publicly accessible space.", Surveillance & Society, vol. 2, pp. 110-135, 2004.

[10] "Surveillance society: New high-tech cameras are watching you", Popular Mechanics (PM), 2017. [Online]. Available: http://www.popularmechanics.com/military/a2398/4236865/ [Accessed: 03-Sep-2017].

[11] "London riots: Most wanted suspect CCTV images released - BBC News", BBC News, 2011. [Online]. Available: http://www.bbc.co.uk/news/uk-england-london-16171972. [Accessed: 29-Aug-2017].

[12] "The police 'super-recognisers' putting names to faces - BBC News", BBC News, 2015. [Online]. Available: http://www.bbc.co.uk/news/uk-england-34544199. [Accessed: 29-Aug-2017].

[13] M. Venkataramanan, "The superpower police now use to tackle crime", BBC, 2015. [Online]. Available: http://www.bbc.com/future/story/20150611-the-superpower-police-now-use-to-tackle-crime. [Accessed: 29- Aug- 2017].

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

[15] J. P. Davis, C. Forrest, F. Treml, and A. Jansari (in press). 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

[16] S. Smart, M. Berry and D. Rodriguez, "Skilled Observation and change blindness: A comparison of law enforcement and student samples", Applied Cognitive Psychology, vol. 28, no. 4, pp. 590-596, 2014.

[17] W. K. Hugenberg and M. Bernstein, "The Cross-Race Effect and Eyewitness Identification: How to Improve Recognition and Reduce Decision Errors in Eyewitness Situations", Social Issues and Policy Review, vol. 7, no. 1, pp. 83-113, 2013.

[18] J. P. Davis and D. Tamonytė, "Masters of disguise: Super-recognisers' superior memory for concealed unfamiliar faces, Proceedings of the 2017 Seventh International Conference on Emerging Security Technologies, pp. 44-49, 2017.