ETHI Brief: The use and impact of facial recognition technology

Prof. Rob Jenkins, Department of Psychology, University of York, UK
4 April 2022

---

1. Summary of evidence

Over five decades of research has led to a substantial scientific literature on face recognition by humans and technology. This research has established several key facts that are critical to this enquiry.

1.1. Face recognition by humans

1.1.1. Unfamiliar and familiar face recognition are fundamentally different. Recognition of *familiar* faces (faces that we know) is highly accurate [1]. In contrast, Recognition of *unfamiliar* faces (faces that we don't know) is surprisingly inaccurate [2]. Virtually all of applied face recognition concerns unfamiliar face. FRT mimics principles of unfamiliar face recognition by humans.

1.1.2. Intuition is a poor predictor of face recognition accuracy. We are bad judges of which unfamiliar faces are easy or difficult to identify [3]. Most people are poor at this task and yet believe the task is more straightforward than it is [4].

1.1.3. Human decision making is prone to biases in which errors are influenced by contextual information [5]. For example, biases can be induced by 'match score' information displayed to operators by FRT [6], which can potentially lead to confident misidentification errors.

1.1.4. There are large individual differences in face recognition ability. This is a stable cognitive trait that is largely determined by genetic factors [7].

1.1.5. Own-race faces are better recognised than other-race faces. This is the finding around the world for all known race combinations. The effect of demographic factors on accuracy is small. However, people expect it to be large [8]. Demographic group disparities in face recognition accuracy reflect the viewer's social diet of faces. This is important for at

least two reasons. First, we should expect performance disparities even in the absence of group prejudice. Second, 'social diet of faces' has a clear analogue in FRT—specifically, the composition of image sets that are used in training.

1.2. Face Recognition Technology (FRT)

1.2.1. Benchmark tests of algorithm accuracy are typically conducted in idealised conditions. Image quality in these tests is unrealistically good due to controlled image capture conditions, which are not possible in criminal investigations and surveillance.

1.2.2. We do not know the accuracy of FRT in realistic applied settings. There is little evidence of FRT accuracy in tests that capture the range of real-world operational deployments. Proper end-to-end evaluation must include measurement of accuracy of the total system which includes both FRT and the people that use FRT.

1.2.3. We do not know how police currently use FRT. There is a lack of information and understanding relating to how FRT is used in police investigations. Lack of clarity can obscure potential for improvement, and prohibits public scrutiny of the costs and benefits of FRT. In the US, it recently became apparent that police forces do not keep records or audit the use of FRT, despite its widespread adoption [9, 10].

1.2.4. Even in idealised testing conditions, errors are more likely for some demographic groups than for others [11]. Disparities in error rate across demographic groups reflect the composition of image sets used to train FRT.

1.3. Human factors in FRT

1.3.1. FRT requires human oversight in the form of intervention, interpretation and monitoring. This is because in legal and forensic applications, FRT does not 'recognise' faces. Instead, automated database search delivers a 'candidate list' of potential matches. Final face identity decisions are made by human operators who select faces from the candidate list and compare them to the search target [12].

1.3.2. Human operators make frequent errors when reviewing candidate lists. For lists of just eight faces, operators made errors about 50% of the time [13]. This level of performance is consistent with research on eyewitness identification, which is known to be unreliable, with well-meaning witnesses often mistakenly identifying innocent suspects [14].

1.3.3. Human oversight is built into regulation. The *UK Surveillance Camera Code of Practice* [15] states that use of FRT "should always involve human intervention before decisions are taken that affect an individual adversely". A similar principle of human oversight has been publicly adopted by the Australian Federal Government that serve to identify a person will never be made by technology alone".

1.4. Approaches to optimising human performance

1.4.1. Selection of high aptitude individuals. The probability of error can be substantially reduced by (i) selecting people with high face recognition ability to become operators of FRT, and (ii) providing them with appropriate training [13]. Face recognition ability is independent of IQ and general visual ability [17], but can be reliably measured using tests that target face recognition specifically [18].

1.4.2. Wisdom of crowds analysis. Recent research shows that aggregating face identity judgments made independently by multiple humans or algorithms greatly improves facial recognition accuracy. Together with recent advances in understanding individual differences in face recognition ability, this 'wisdom of crowds' approach is a promising path to minimising errors associated with the use of FRT [19].

1.4.3. Effective training. Lab-based tests suggest that some training methods provide benefits to face identification, but these are small in comparison to selection and recruitment [20]. Commercial training courses are often ineffective despite positive reviews from trainees [21], and so it is important that FRT training courses are formally evaluated for their effectiveness.

---

2. Principles for use of FRT in forensic settings

2.1. Appropriate human oversight and attention to human operators in the design and implementation of facial recognition systems. Humans are a critical component to ensuring the accuracy of FRT. Deliberate efforts must be made to ensure that the humans involved in face recognition decisions are highly skilled, either by targeted recruitment or evidence-based training. Scientifically validated tests of face identification ability can be used to assess suitability of FRT operators and monitor performance over time. In addition, the design of face recognition systems needs to incorporate appropriate checks and balances to minimise the risk of consequential errors.

2.2. Transparency. Use of FRT in the legal system should be accompanied by transparent disclosure of information relating to the accuracy, strengths, limitations, and operation of this technology. It is also necessary to understand how the technology is being used and to design regulation that ensures appropriate use.

2.3. Development of an expert workforce in facial recognition. If FRT is to be adopted in forensic practice, then new types of expert practitioners and researchers are required to design, evaluate, oversee, and explain the resultant face identification systems. Because these systems incorporate human and AI decision making, people with broad expertise in related disciplines are required.

---

3. Further supporting information

3.1. Accuracy and errors

3.1.1. Benchmarking. Standard benchmarking tests—of the type quoted by algorithm vendors when asked about the accuracy of their FRT—underestimate the error that would be found in the vast majority of forensic applications. In standard tests, image quality is often much higher, and databases that are being searched are often much smaller, than those in criminal investigations. Test images are also of 'compliant' subjects who are not trying to avoid identification, and are looking straight at the camera in good lighting. These conditions are not representative of conditions that are typically encountered in investigations.

3.1.2. Candidate lists. In benchmarking tests, accuracy is often measured as the probability that a matching face from a given database is returned as the highest ranked possible match. However, because this probability is reduced for low quality imagery used in criminal investigations, it means that human operators must review large lists of images – sometimes over 100 – to ensure they do not miss a match. Recent studies have tested the face matching accuracy of staff using FRT in their daily work and found errors in 50% of their candidate list review decisions [13], despite displaying lists of just eight images. Errors indicate that the probability of human operators selecting an innocent person from an FRT candidate list face – a 'false positive' identification – is alarmingly high.

3.1.3. Cascading errors. It may be tempting to conclude that because FRT is used to generate investigative leads, and not to make definitive judgements of identity, the problem

of 'false positives' in this context is not serious. However, broader understanding of the influences of cognitive bias in forensic decision making – and the compounding effects that an error at one stage of an investigation can have on interpretation of subsequent sources of information [5] – suggest that even erroneous investigative leads can lead to serious outcomes such as wrongful arrest [22].

3.2. Implementation

3.2.1. System implementation. Proper implementation of facial recognition systems is more complex than simply purchasing the latest algorithm. While algorithm vendors should do more than simply install the software, they cannot be expected to provide the necessary oversight. This requires 'in-house' testing using images, procedures, and tools that are representative of the organisation's casework, rather than relying on vendor performance rates or standard benchmark tests. Critically, this testing should consider both FRT and human processing accuracy. It requires specialist staff to be accountable for the whole system, including algorithms, workflow design, and delegation of tasks to humans. Staff must have the expertise to measure accuracy and to report risks associated with deployment of algorithms for specific uses.

3.2.2. System support. The monetary cost of implementing a facial recognition system far exceeds the cost of purchasing and installing FRT in IT systems and incorporating it into workflow. Budget allocations for FRT must include funds for ensuring appropriate use and oversight of the system, including regular testing of facial recognition system accuracy in operational deployment, selection of human operators, staff training, career development of FRT specialist teams, routine monitoring and auditing of FRT use, and user experience design.

References

[1] Burton, A. M., Jenkins, R., & Schweinberger, S. R. (2011). Mental representations of familiar faces. British Journal of Psychology, 102(4), 943-958.

[2] Burton, A. M., & Jenkins, R. (2011). Unfamiliar face perception. The Oxford handbook of face perception, 28, 287-306.

[3] Zhou, X., & Jenkins, R. (2020). Dunning–Kruger effects in face perception. Cognition,

203, 104345.

[4] Ritchie, K. L., Smith, F. G., Jenkins, R., Bindemann, M., White, D., & Burton, A. M. (2015). Viewers base estimates of face matching accuracy on their own familiarity: Explaining the photo-ID paradox. Cognition, 141, 161-169.

[5] Dror, I. E., Morgan, R. M., Rando, C., & Nakhaeizadeh, S. (2017). The bias snowball and the bias cascade effects: Two distinct biases that may impact forensic decision making. Journal of Forensic Sciences, 62(3), 832-833.

[6] Howard, J. J., Rabbitt, L. R., & Sirotin, Y. B. (2020). Human-algorithm teaming in face recognition: How algorithm outcomes cognitively bias human decision-making. PLOS One, 15(8), e0237855.

[7] 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, 107(11), 5238-5241.

[8] Zhou, X., Burton, A. M., & Jenkins, R. (2021). Two Factors in Face Recognition: Whether you know the person's face and whether you share the person's race. Perception, 50(6), 524-539.

[9] Facial Recognition Technology: Federal Law Enforcement Agencies Should Better Assess Privacy and Other Risks. US Government Accountability Office. Via: https://www.gao.gov/products/gao-21-518; see also https://www.perpetuallineup.org/, https://humanrights.gov.au/our-work/rights-and-freedoms/publications/human-rights-and-technology-final-report-2021 (Chapter 9)

[10] https://www.flawedfacedata.com

[11] Grother, P. et al. (2019). Face Recognition Vendor Test (FVRT): Part 3, Demographic Effects. https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8280.pdf

[12] Footnote 1, Pages 7-9; Davies, B. et al (2018). An Evaluation of South Wales Police's Use of Automated Facial Recognition. https://crimeandsecurity.org/feed/afr

[13] White, D., Dunn, J. D., Schmid, A. C., & Kemp, R. I. (2015). Error rates in users of

automatic face recognition software. PLOS One, 10(10), 1-14.

[14] Steblay, N., Dysart, J., Fulero, S., & Lindsay, R. C. (2001). Eyewitness accuracy rates in sequential and simultaneous lineup presentations: A meta-analytic comparison. Law and Human Behavior, 25(5), 459-473.

[15] https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1010815/Surveillance_Camera_Code_of_Practice_update.pdf

[16] https://www.theguardian.com/technology/2019/sep/29/plan-for-massive-facial-recognition-database-sparks-privacy-concerns

[17] Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Gerbasi, M., & Nakayama, K. (2012). Capturing specific abilities as a window into human individuality: The example of face recognition. Cognitive Neuropsychology, 29(5-6), 360-392.

[18] White, D., Guilbert, D., Varela, V. P., Jenkins, R., & Burton, A. M. (2022). GFMT2: A psychometric measure of face matching ability. Behavior research Methods, 54(1), 252-260.

[19] Phillips, P. J., Yates, A. N., Hu, Y., Hahn, C. A., Noyes, E., Jackson, K., Cavazos, J. G., Jeckeln, G., Ranjan, R., Sankaranarayanan, S., Chen, J., Castillo, C. D., Chellappa, R., White, D., & O'Toole, A. J. (2018). Face recognition accuracy of forensic examiners, super-recognizers, and face recognition algorithms. Proceedings of the National Academy of Sciences, 115(24), 6171-6176.

[20] Towler, A., Kemp, R. I., & White, D. (2021). Can face identification ability be trained? Evidence for two routes to expertise. In M. Bindemann (Ed.), Forensic face matching: Research and practice: Oxford University Press.

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

[22] https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html