Article:
BURTON, MIKE, Dimmock, Izzy and Salton, Becca (2026) Polling clerks' errors in face matching. Royal Society Open Science. 251989. ISSN: 2054-5703

https://doi.org/10.1098/rsos.251989

ROYAL SOCIETY OPEN SCIENCE

royalsocietypublishing.org/journal/rsos

Research

Cite this article: Burton AM, Dimmock I, Salton B. 2026 Polling clerks' errors in face matching. *R. Soc. Open Sci.* **13**: 251989.
https://doi.org/10.1098/rsos.251989

Received: 16 October 2025
Accepted: 1 December 2025

Subject Category:
Psychology and cognitive neuroscience

Subject Areas:
psychology

Keywords:
face matching, face perception, photo ID, polling clerks, elections

Author for correspondence:
A. Mike Burton
e-mail: mike.burton@york.ac.uk

Polling clerks' errors in face matching

A. Mike Burton[1,2], Izzy Dimmock[1] and Becca Salton[1]

[1]Department of Psychology, University of York, York, UK
[2]Faculty of Society and Design, Bond University, City of Gold Coast, Queensland, Australia

AMB, 0000-0002-2035-2084

The UK 2024 general election was the first for which voters were required to prove their identity using a photo Identity Document (ID). Opposition to this change focused on the way it might affect participation, rather than the accuracy of photo ID itself. However, research demonstrates that people are generally poor at matching unfamiliar faces, including checking ID. We tested polling clerks who were responsible for ID checks in the 2024 general election. Using standard tests of face matching, polling clerks made multiple errors in simple match/mismatch decisions (Glasgow Face Matching Test: 16% errors; Kent Face Matching Test: 36% errors), levels of performance that are typical across the population. Individual differences were large, with some clerks performing accurately, and others much less so. There was a significant decline in accuracy over age, with older clerks (on average) scoring poorly. The same tests were administered to a volunteer group of university students, who had taken no part in the election. These students made very similar levels of errors. As with the general population and other professional groups, polling clerks are prone to error in face matching. For this reason, photo ID does not provide a reliable proof of identity at the point of voting.

1. Introduction

Matching unfamiliar faces is highly error prone. For over 20 years, laboratory-based studies have demonstrated that viewers typically find it difficult to decide whether two simultaneously presented photos show the same person or not [1–4]. Average error rates of 20–30% are common in these studies, depending on the particular tests used, even when photos are taken in good visual conditions and close together in time. This contrasts strongly with *familiar* face matching, which most viewers find very easy, even when images are distorted or shown in poor quality [2,5].

Photo ID continues to be used as a means of proving identity in many countries. This seems odd, given the evidence

THE ROYAL SOCIETY PUBLISHING

that unfamiliar face matching is difficult. We have previously suggested that human viewers' excellent recognition of *familiar* faces misleads us into thinking that this will generalize to all faces, and so most people do not realize how difficult it is to identify unfamiliar people [6,7]. This idea is supported by studies showing that people generally have poor levels of insight into their own face recognition abilities [8,9].

Does evidence from laboratory tasks really address the problem of photo ID in practice? Most laboratory-based studies use photo-to-photo tasks, whereas in common ID usage, the viewer has the real person present, adding three-dimensional movement and social cues to the information available. In fact, many studies have demonstrated that viewers are generally poor at person-to-photo matching, and performance is typically no better than photo-to-photo levels [10–13]. Furthermore, while some laboratory studies use photos taken specifically for the experiment at hand, there are several studies using real ID, passports and driving licences that elicit very similar levels of (relatively poor) matching performance [14–16].

Critically for this study, there have been many demonstrations that professional employees, trained in face matching, are no better than untrained, inexperienced controls. For example, studies have shown very similar performance, on average, for untrained student viewers, serving passport officers [13], police officers [5,17], bank tellers and notaries [18] and supermarket check-out staff, bar staff and bouncers [11,19]. These results seem to compromise legal standards, and in some cases, public security. One group of professionals has sometimes been reported to outperform other professionals and the general public on face-matching tests [20,21]. These are *forensic examiners,* working for police and national security organizations, and who have been specifically trained in facial comparison techniques, typically emphasising part-by-part comparisons. However, even those classed as forensic examiners do not always outperform novice face matchers [22], and their performance remains well below the optimal levels [21].

While we have emphasized generally poor levels of unfamiliar face-matching performance in the studies discussed above, these results reflect *average* performance. In fact, research consistently shows large and stable individual differences in face processing [23–25], with different types of face tasks (memory, matching, line-ups, eye movements) correlating strongly [26–28]. These results combine to suggest that the generally poor average levels of unfamiliar face matching, observed across multiple settings, imply that checking of photo ID is highly likely to be error prone, *on average*, in any real-world setting. One solution to this problem is to recruit employees who have naturally high levels of face-matching ability, and this approach has been validated in the literature [29–31]. However, this is not routinely done across the range of jobs that include a requirement to check photo ID, whether for security-critical tasks (e.g. passport officers) or in more routine settings (e.g. check-out staff).

This study reports face-matching performance in a group so far unstudied: polling clerks. In the UK general election of 2024, voters were required, for the first time, to prove their identity at polling stations using government-approved photo ID (passports, driving licences, etc.). This represented a significant change in the law, introduced in the UK Elections Acts 2022. A government policy paper (2022) stated that 'Showing photographic identification is a reasonable and proportionate way to confirm that someone is who they say they are when voting, thus stamping out the potential for voter fraud …' [32]. The change was politically controversial, with debates focusing on the likelihood that this new stipulation would encourage or discourage different sections of society to exercise their vote. However, there was very little debate about the efficacy of the measure—it was assumed that photo IDs could be checked accurately, and this responsibility fell to *polling clerks*, specially recruited for the single day of the election in order to run polling station operations.

In this study, we directly test polling clerks' face-matching ability, recruiting participants who played this role in the 2024 UK general election. We measured their abilities on two standard face-matching tests, which we also applied to a group of student participants. While we hypothesized that polling clerks were likely to show matching ability similar to that previously reported in the general population, it is important to test this group specifically, in order to establish whether government policy to require photo ID is likely to address issues of concern in the security of elections.

2. Method

2.1. Participants

Polling clerks were recruited through a local UK city council. The council sent an email to the list of people whom it had hired in this role for the UK general election in July 2024. Testing took place in December 2024 and January 2025. Volunteers responding to the email were further asked whether they had acted as polling clerks in the general election, and 58 confirmed that they had. One was subsequently excluded on the grounds of outlying performance (see §3), leaving a total of 57 (22 male, 35 female; mean age = 50, s.d. = 17.3). A group of 50 undergraduate students volunteered to take part; none of them had ever worked as polling clerks (12 male, 37 female, 1 non-binary; mean age = 20.9, s.d. = 1.2). All participants were provided with clear explanations of the procedure before beginning the study, and all provided written consent.

2.2. Materials

Two standard face-matching tests were used, the Glasgow Face Matching Test (GFMT) [3], and the Kent Face Matching Test (KFMT) [4]. These tests have a similar format: both comprise 40 items, each showing a face pair. Half of the items show the same person and half different people. Participants are shown one pair at a time and asked to respond 'same' or 'different' to each pair.

The GFMT is the easier test and shows two images taken on the same day (matching items), at the same size and displayed in grey scale. The KFMT shows items taken several months apart, at different sizes and displayed in colour. Models for both tests were students. In the case of the GFMT, both images in each pair were studio photos (though taken with different cameras), while for the KFMT, one of the images was a studio photo, and the other was their student ID card. Examples are shown in figure 1. Norms for the two tests, taken from their original reporting, give a mean accuracy of 81.3% for the GFMT [3] and 66% for the KFMT [4].

2.3. Procedure

Testing was run online, using the *Gorilla* experimental platform. Participants provided consent and completed a short demographic questionnaire before starting the experiment proper. They recorded their age and gender and were also asked, 'have you ever received any training in face recognition?'. If the answer to the question was 'yes', they were invited to provide a short description of this training. The GFMT was then administered, followed by the KFMT, and within each test, the order of items was randomized independently for each participant. For each item on both tests, participants were instructed: 'If you think the faces are the same person, please press the "same" button. If you think the faces are different people, please press the "different" button'. Both tests were self-paced, and face pairs remained on screen until a decision was recorded by button press.

3. Results

One polling clerk was excluded from the analysis, scoring more than 3 s.d. below the mean for both the GFMT and the KFMT. (This participant actually scored below chance in both tests, at 48% and 43%, respectively.) This left a final sample of 57 polling clerks and 50 students, analysed below.

3.1. Accuracy levels

Table 1 shows mean performance levels for the two tests. It is clear that both groups score close to the previously published norms for these tests (see §2.2). 2 × 2 ANOVA shows a main effect of test (GFMT scores being higher than KFMT, $F_{1,105} = 508.6$, $p < 0.001$,$\eta_p^2 = 0.83$) but no main effect of group (polling clerks versus students, $F_{1,105} = 0.013$, $p = 0.911$, $\eta_p^2 < 0.001$) and no interaction ($F_{1,105} = 1.32$, $p = 0.253$, $\eta_p^2 = 0.012$).

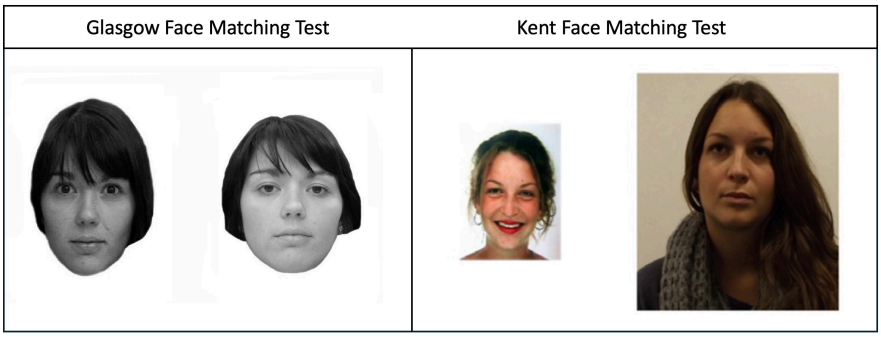

Figure 1. Examples from the two matching tests. (Both items show 'same' matches.)

Table 1. Mean accuracy (%) for the two tests. (Standard deviations are given in brackets.)

	GFMT	KFMT
polling clerks	83.7 (9.9)	64 (8.6)
students	85 (8.4)	63.1 (8.3)

3.2. Demographics and training

There was no evidence of gender differences for accuracy on either test. One participant identified as neither male nor female, for those remaining (34 male, 72 female): GFMT means were 83% versus 85% (male/female), $t_{104} = 0.87$, $p = 0.386$, $B_{10} = 0.31$; KFMT means were 62% versus 64% (male/female), $t_{104} = 1.1$, $p = 0.273$, $B_{10} = 0.37$).

The age range for students was highly restricted and much greater for polling clerks (see §2.1). Figure 2 shows the association between age and face-matching performance for the polling clerks only. There was a significant negative correlation for both matching tests (GFMT: Pearson's $r = -0.31$, $p = 0.017$; KFMT: Pearson's $r = -0.27$, $p = 0.042$).

Only a minority of the polling clerks reported having received training in face recognition: 11 out of 57. Their brief descriptions of this training suggested an informal level of analysis; for example, responses included: 'online training—if the person's features look the same regardless of colour or style of hair, then you can assume they are the same person'; 'as part of election training—look for nose/eyes/face shape'. A comparison of polling clerks who reported receiving training and those who did not revealed no differences: GFMT means were 85 versus 84% (trained/untrained), $t_{55} = 0.39$, $p = 0.7$, $B_{10} = 0.34$; KFMT means were 64 versus 64% (trained/untrained), $t_{55} = 0.07$, $p = 0.94$, $B_{10} = 0.32$).

4. Discussion

The results clearly show that a group of polling clerks, operating during the UK general election of 2024, scored relatively poorly on a standard test of face matching. On the GFMT, in which viewers compare pictures taken just a few seconds apart, they make about 16% errors. The KFMT requires comparison of a photo with a genuine ID, and so perhaps more fairly reflects the demands of their role. The polling clerks made 36% errors. This level of error is extraordinarily high in any practical setting—all the more so when one considers that the test takes place in optimal conditions, with no time pressure to make a decision, showing faces in good lighting and very similar pose.

As described above, people are generally not aware of how difficult unfamiliar face matching is. The poor rates of performance observed in polling clerks are actually very similar to those reported elsewhere, in both the general population and specialist groups. As confirmation that our testing circumstances were not unusual, we observed very similar performance in a group of students, quite unlike the polling clerks demographically, but achieving very similar levels of accuracy on face matching. We know from previous research that performance on these tests is highly correlated with performance in real settings [13,31]. It therefore seems safe to conclude that UK polling clerks cannot reliably judge a person's identity from their photo ID.

This result is barely surprising, given previous research. However, it is also worth pointing out, in common with previous studies, that *some* polling clerks are performing extremely well. For example,

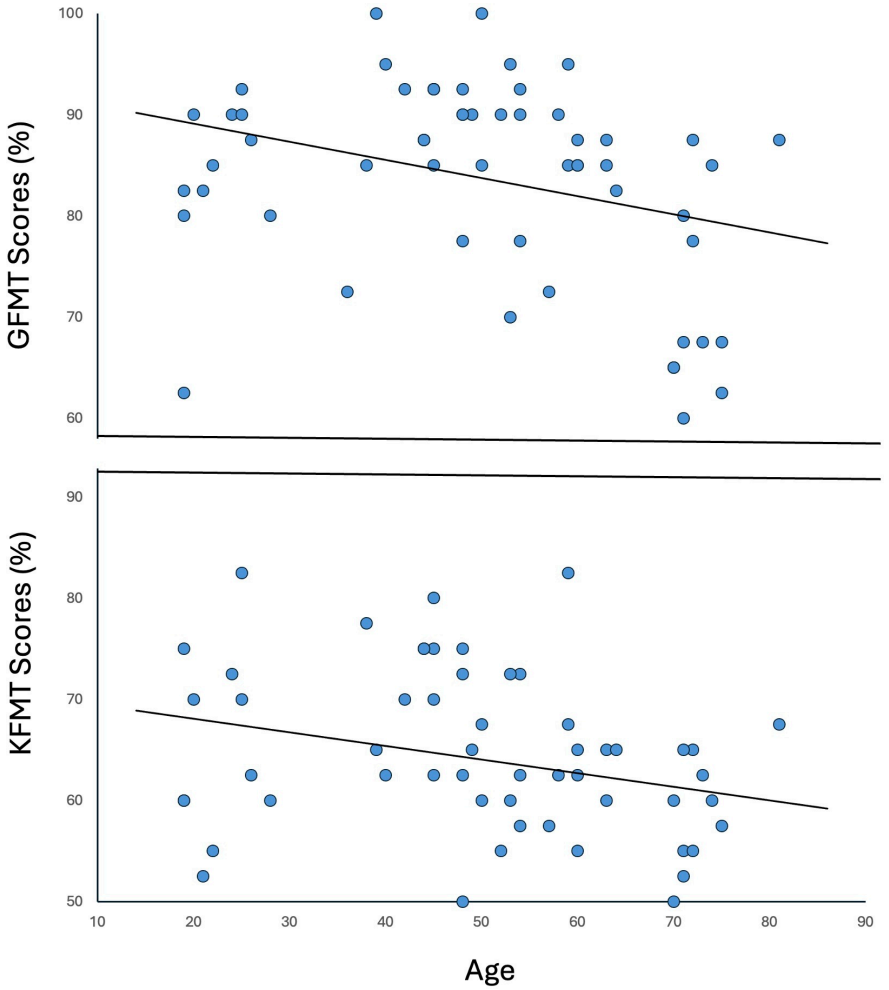

Figure 2. Associations between polling clerks' age and their performance on two tests of face matching. Trend lines represent linear model fit: GFMT, $R^2 = 0.099$; KFMT, $R^2 = 0.0723$.

figure 2 shows that two participants scored very highly on the GFMT—achieving 100% on that test. Some participants also scored well on the KFMT. However, the same figure shows some participants scoring very poorly—at or close to chance levels (50%).

Could this problem be solved by introducing more extensive training? We were surprised that most polling clerks did not report having received any training in face matching, though general training in administering the polling operation was provided. However, established research suggests that training is unlikely to help. In one extensive study [17], Towler and colleagues tested the face-matching ability of nearly 400 participants, before and after taking courses provided to international security agencies, and adhering to the standards outlined by the Facial Identification Scientific Working Group. Most courses produced no effect of training on participants' face-matching ability, though the longest courses produced a very small improvement on some tests. No courses substantially reduced the number of errors made by the trainees. Despite this, 93% of those attending believed that their performance had improved.

An alternative strategy to improve systems based on human face matching is to recruit people who are naturally high performers in this task [29,33]. This approach has been taken by a number of government-level security organizations. However, extensive pre-employment testing may be impractical for everyday tasks such as the sales of age-restricted goods. Nevertheless, for security-critical roles, such as polling clerks, it may be possible to exclude some of the most poorly performing applicants, using available tests with published norms.

There is one further aspect of the data we have collected which should raise concern for the security of the polling system. Both the tests reported in figure 2 show a significant decline in performance as participants get older. Polling clerks represent a very wide age range, and the decline in *average* performance reported here is consistent with studies of face recognition in the literature [34,35]. This

effect is similar across the two tests, but note that it is relatively small by comparison to the individual differences in performance overall. In short, there are high and low performers in face matching for all ages, though the average shows a decline in later years.

Despite the fact that this particular occupational group—polling clerks—performed as predicted on face matching, it is important to establish this empirically, if we are to advise policy-makers on security procedures. We have no doubt that these clerks behaved in a highly professional manner throughout and are highly motivated to perform well. However, it seems to be a constraint on human performance that unfamiliar face matching is variable among people and surprisingly poor on average. We continue to argue that unfamiliar face matching is a poor way to establish one's identity.

Ethics. This research was conducted in accordance with the Declaration of Helsinki. It was approved by the Psychology Departmental Ethics Committee, University of York, UK; Approval Number 2425307. All participants provided written informed consent.

Data accessibility. All data are available at the OSF [36].

Declaration of AI use. We have not used AI-assisted technologies in creating this article.

Authors' contributions. A.M.B.: conceptualization, data curation, formal analysis, methodology, project administration, supervision, writing—original draft, writing—review and editing; I.D.: conceptualization, data curation, formal analysis, investigation, methodology, writing—original draft, writing—review and editing; B.S.: conceptualization, data curation, formal analysis, investigation, methodology, writing—original draft, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Acknowledgements. We would like to thank Nicole Collingwood, who identified polling clerks as an interesting group of participants to study.

References

1. Bruce V, Henderson Z, Greenwood K, Hancock PJB, Burton AM, Miller P. 1999 Verification of face identities from images captured on video. *J. Exp. Psychol. Appl.* **5**, 339–360. (doi:10.1037//1076-898x.5.4.339)
2. Bruce V, Henderson Z, Newman C, Burton AM. 2001 Matching identities of familiar and unfamiliar faces caught on CCTV images. *J. Exp. Psychol. Appl*. **7**, 207–218.
3. Burton AM, White D, McNeill A. 2010 The Glasgow face matching test. *Behav. Res. Methods* **42**, 286–291. (doi:10.3758/BRM.42.1.286)
4. Fysh MC, Bindemann M. 2018 The Kent face matching test. *Br. J. Psychol*. **109**, 219–231. (doi:10.1111/bjop.12260)
5. Burton AM, Wilson S, Cowan M, Bruce V. 1999 Face recognition in poor-quality video: evidence from security surveillance. *Psychol. Sci*. **10**, 243–248. (doi:10.1111/1467-9280.00144)
6. Ritchie KL, Smith FG, Jenkins R, Bindemann M, White D, Burton AM. 2015 Viewers base estimates of face matching accuracy on their own familiarity: explaining the photo-ID paradox. *Cognition* **141**, 161–169. (doi:10.1016/j.cognition.2015.05.002)
7. Young AW, Burton AM. 2017 Recognizing faces. *Curr. Dir. Psychol. Sci*. **26**, 212–217. (doi:10.1177/0963721416688114)
8. Zhou X, Jenkins R. 2020 Dunning-Kruger effects in face perception. *Cognition* **203**, 104345. (doi:10.1016/j.cognition.2020.104345)
9. Bobak AK, Mileva VR, Hancock PJ. 2019 Facing the facts: naive participants have only moderate insight into their face recognition and face perception abilities. *Q. J. Exp. Psychol*. **72**, 872–881. (doi:10.1177/1747021818776145)
10. Davis JP, Valentine T. 2009 CCTV on trial: matching video images with the defendant in the dock. *Appl. Cogn. Psychol*. **23**, 482–505. (doi:10.1002/acp.1490)
11. Kemp R, Towell N, Pike G. 1997 When seeing should not be believing: photographs, credit cards and fraud. *Appl. Cogn. Psychol*. **11**, 211–222. (doi:10.1002/(sici)1099-0720(199706)11:33.3.co;2-f)
12. Megreya AM, Burton AM. 2008 Matching faces to photographs: poor performance in eyewitness memory (without the memory). *J. Exp. Psychol. Appl*. **14**, 364–372. (doi:10.1037/a0013464)
13. White D, Kemp RI, Jenkins R, Matheson M, Burton AM. 2014 Passport officers' errors in face matching. *PLoS ONE* **9**, e103510. (doi:10.1371/journal.pone.0103510)
14. Kramer RSS, Mohamed S, Hardy SC. 2019 Unfamiliar face matching with driving licence and passport photographs. *Perception* **48**, 175–184. (doi:10.1177/0301006619826495)
15. McCaffery JM, Burton AM. 2016 Passport checks: interactions between matching faces and biographical details. *Appl. Cogn. Psychol*. **30**, 925–933. (doi:10.1002/acp.3281)
16. Feng X, Burton AM. 2021 Understanding the document bias in face matching. *Q. J. Exp. Psychol. (2006)* **74**, 2019–2029. (doi:10.1177/17470218211017902)
17. Towler A, Kemp RI, Burton AM, Dunn JD, Wayne T, Moreton R, White D. 2019 Do professional facial image comparison training courses work? *PLoS ONE* **14**, 1–17. (doi:10.1371/journal.pone.0211037)

18. Papesh MH. 2018 Photo ID verification remains challenging despite years of practice. *Cogn. Res.* **3**, 19. (doi:10.1186/s41235-018-0110-y)
19. Robertson DJ, Burton AM. 2021 Checking ID-cards for the sale of restricted goods: age decisions bias face decisions. *Appl. Cogn. Psychol.* **35**, 71–81. (doi:10.1002/acp.3739)
20. Phillips PJ *et al*. 2018 Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms. *Proc. Natl Acad. Sci. USA* **115**, 6171–6176. (doi:10.1073/pnas.1721355115)
21. White D, Phillips PJ, Hahn CA, Hill M, O'Toole AJ. 2015 Perceptual expertise in forensic facial image comparison. *Proc. R. Soc. B* **282**, 20151292. (doi:10.1098/rspb.2015.1292)
22. Claydon JR, Fysh MC, Prunty JE, Cristino F, Moreton R, Bindemann M. 2023 Facial comparison behaviour of forensic facial examiners. *Appl. Cogn. Psychol.* **37**, 6–25. (doi:10.1002/acp.4027)
23. White D, Burton AM. 2022 Individual differences and the multidimensional nature of face perception. *Nat. Rev. Psychol.* **1**, 287–300. (doi:10.1038/s44159-022-00041-3)
24. Wilmer JB. 2017 Individual differences in face recognition: a decade of discovery. *Curr. Dir. Psychol. Sci.* **26**, 225–230. (doi:10.1177/0963721417710693)
25. Baker KA, Stabile VJ, Mondloch CJ. 2022 Stable individual differences in unfamiliar face identification: evidence from simultaneous and sequential matching tasks. *Cognition* **232**, 105333. (doi:10.1016/j.cognition.2022.105333)
26. Peterson MF, Lin J, Zaun I, Kanwisher N. 2016 Individual differences in face-looking behavior generalize from the lab to the world. *J. Vis.* **16**, 1–18. (doi:10.1167/16.7.12)
27. Bobak AK, Hancock PJB, Bate S. 2016 Super-recognisers in action: evidence from face-matching and face memory tasks. *Appl. Cogn. Psychol.* **30**, 81–91. (doi:10.1002/acp.3170)
28. McCaffery JM, Robertson DJ, Young AW, Burton AM. 2018 Individual differences in face identity processing. *Cogn. Res.* **3**, 1–15. (doi:10.1186/s41235-018-0112-9)
29. Bate S, Frowd C, Bennetts R, Hasshim N, Murray E, Bobak AK, Wills H, Richards S. 2018 Applied screening tests for the detection of superior face recognition. *Cogn. Res.* **3**. (doi:10.1186/s41235-018-0116-5)
30. Bobak AK, Dowsett AJ, Bate S. 2016 Solving the border control problem: evidence of enhanced face matching in individuals with extraordinary face recognition skills. *PLoS ONE* **11**, e0148148. (doi:10.1371/journal.pone.0148148)
31. Robertson DJ, Noyes E, Dowsett AJ, Jenkins R, Burton AM. 2016 Face recognition by metropolitan police super-recognisers. *PLoS ONE* **11**, 1–8. (doi:10.1371/journal.pone.0150036)
32. Department for Levelling Up, Housing & Communities. 2022 *Protecting the integrity ofour elections: voter identification at polling stations and the new voter card*. See https://www.gov.uk/government/publications/voter-identification-at-polling-stations-and-the-new-voter-card.
33. Balsdon T, Summersby S, Kemp RI, White D. 2018 Improving face identification with specialist teams. *Cogn. Res.* **3**, 25. (doi:10.1186/s41235-018-0114-7)
34. Germine LT, Duchaine B, Nakayama K. 2011 Where cognitive development and aging meet: face learning ability peaks after age 30. *Cognition* **118**, 201–210. (doi:10.1016/j.cognition.2010.11.002)
35. Megreya AM, Bindemann M. 2015 Developmental improvement and age-related decline in unfamiliar face matching. *Perception* **44**, 5–22. (doi:10.1068/p7825)
36. Burton M. 2025 Polling clerks errors in face matching. Retrieved from https://osf.io/cu529