Routledge
Taylor & Francis Group

Volume 31 – Issue 7 – August 2023
Visual Cognition

Visual Cognition

Co-Editors: Benjamin Balas & Jonathan Flombaum

Routledge

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/pvis20

Recognizing newly learned faces across changes in age

Laura Sexton, Mila Mileva, Graham Hole, Ailsa Strathie & Sarah Laurence

To cite this article: Laura Sexton, Mila Mileva, Graham Hole, Ailsa Strathie & Sarah Laurence (07 Mar 2024): Recognizing newly learned faces across changes in age, Visual Cognition, DOI: 10.1080/13506285.2024.2315813

To link to this article: https://doi.org/10.1080/13506285.2024.2315813

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

View supplementary material

Published online: 07 Mar 2024.

Submit your article to this journal

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=pvis20

VISUAL COGNITION
https://doi.org/10.1080/13506285.2024.2315813

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Recognizing newly learned faces across changes in age

Laura Sexton[a,b], Mila Mileva[c], Graham Hole[d], Ailsa Strathie[a] and Sarah Laurence [a]

[a]School of Psychology and Counselling, The Open University, Milton Keynes, UK; [b]School of Psychology, University of Sunderland, Sunderland, UK; [c]School of Psychology, University of Plymouth, Plymouth, UK; [d]School of Psychology, University of Sussex, Falmer, UK

ABSTRACT

We examined how well faces can be recognized despite substantial age-related changes, using three behavioural experiments plus Mileva et al.'s (2020, Facial identity across the lifespan. *Cognitive Psychology*, 116, 101260) PCA + LDA computational model of face recognition. Participants and the model were trained on a set of faces at one age (with each facial identity depicted in multiple images) and tested on their ability to recognize those individuals in images taken at a different age. The younger images were aged 20–30 years and the older images were either 20 or 40 years older. The computational model showed high accuracy, but it performed better if faces were learnt in their younger versions and testing was with the older images than vice versa. The humans did not show this age-direction effect. Although their recognition of faces across either a 20- or 40- year age gap was poor, it was significantly above chance, suggesting that we can extract identity diagnostic information despite substantial changes in outward appearance.

ARTICLE HISTORY
Received 24 February 2023
Accepted 22 January 2024

KEYWORDS
Face recognition; facial ageing; face perception

Introduction

Recognizing faces across changes in age poses a significant challenge in a range of applied settings, such as identifying a person from their old passport/ID card or recognizing a person who has been missing for several years. One reason for looking at how people cope with age-related change is that it acts as a "test-bed" for ideas about how we cope with change in unfamiliar faces more generally. While there is a lot of research looking at how well we can recognize unfamiliar faces despite changes in viewpoint, context and lighting etc. (see Hancock et al., 2000; Johnston & Edmonds, 2009), age-related changes produce more profound physical alterations to a face (Friedman, 2005). Yet there is something that remains in the face that permits recognition to occur (Mileva et al., 2020). So, investigating how we cope with age-related changes has the potential to tell us something about how we cope with change in general. (For one thing, it questions claims that face recognition is based on fine grain configural processing (Freire et al., 2000; Maurer et al., 2002; also see Burton et al., 2015; Hole et al., 2002), because the configuration of a person's face at 20 and 70 years is typically very different).

From a theoretical point of view, it is interesting that people can cope with age-related changes at all, given that even in adulthood, the craniofacial and textural changes are often quite profound. The ageing process results in substantial changes to skin texture and facial volume (Coleman & Grover, 2006), nose size (Edelstein, 1996); hair, eyebrow, and eyelash colour (Panhard et al., 2012), and jaw size and angle (Ohm & Silness, 1999). Appearance can also be affected by environmental and health-related factors, such as sun damage, stress, or cigarette smoking (De Jager et al., 2018). These factors will differ from person to person, therefore, facial ageing results from both generic and idiosyncratic changes to facial appearance.

To recognize the same person in age-separated images, one must extract information diagnostic of identity, despite substantial changes in outward appearance. Previous research by Longmore et al. (2017) suggests that recognition of newly learned faces becomes less accurate as the age gap

CONTACT Sarah Laurence Sarah.Laurence@open.ac.uk School of Psychology and Counselling, The Open University, Milton Keynes, MK7 6AA, UK
Supplemental data for this article can be accessed online at https://doi.org/10.1080/13506285.2024.2315813.

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

between the learned and test faces increases. Performance was below chance when learning a face at one age and subsequently trying to recognize the same person across a 40-year age gap. This suggests that whatever identity-diagnostic information people were extracting from a face image, was not shared with an image of the same person 40 years later. In Longmore's study, however, participants learned faces at each age (e.g., 20 years/60 years) from a single image.

A growing body of work has found that exposure to within-person variability in appearance plays a key role in face learning (Andrews et al., 2015; Baker et al., 2017; Kramer et al., 2017b; Matthews & Mondloch, 2018; Murphy et al., 2015; Ritchie & Burton, 2017). Learning a face from multiple images showing within-person variability in appearance (i.e., variability across different encounters with a face) results in better face learning than from multiple images taken during a single encounter (Ritchie & Burton, 2017). The approach taken by many researchers examining face learning from variable images is to source multiple images of each identity from a Google search (e.g., using the name of a celebrity, such as "Barack Obama" as a search term; Andrews et al., 2015; Ritchie & Burton, 2017; Kramer et al., 2017b; Matthews & Mondloch, 2018; Honig et al., 2022) or use images provided by a model as experimental stimuli (e.g., Baker et al., 2017; Murphy et al., 2015, Matthews & Mondloch, 2022; Menon et al., 2015; Zhou et al., 2018). The number of images per identity has varied across studies (e.g., 10 images per identity in Ritchie & Burton, 20 images per identity in Andrews et al., 2015), with a benefit of high variability over low variability observed with as few as two images per identity (Menon et al.). While this approach will result in age differences across the experimental stimuli, very little research has explicitly studied the ability to generalize across age (i.e., ageing has not been the focus). Learning a face at one age and subsequently recognizing the same face 20, 30 or even 40 years later is a more challenging task. There are more ways in which the face could potentially vary across substantial age changes than when recognizing two face images showing a similar age. General knowledge about how faces age might be helpful, to some extent, but this cannot compensate for the influence of unknown environmental factors (e.g., whether someone's face has aged faster due to an unhealthy lifestyle) or genetic factors (Djordjevic et al., 2016; Mayes et al., 2010).

Can the variability we learn at one age generalize to a very different age? Evidence from Mileva et al. (2020) suggests that computer models of face recognition can generalize to other ages surprisingly well. Mileva et al. modelled bottom-up and top-down contributions to face recognition using Principal Component Analysis (PCA) alone or combined with Linear Discriminant Analysis (LDA), respectively (see Kramer et al., 2018b). In one set of simulations, they trained their model on identities from a single time period and tested how well novel images of those identities could be classified at three different time periods (no time difference, 20-year difference, 40-year difference). Classification accuracy was greatest when there was no time difference and worsened as the time difference increased. However, performance was well above chance even when there was a 40-year age difference between the training and test images, at least with the additional top-down information, necessary for LDA, that codes for the identity each picture in the training set belonged to.

Few behavioural studies with human participants have investigated our ability to generalize across substantial age differences. Bruck et al. (1991) examined how well participants could recognize their old school classmates at their 25-year reunion. Participants were asked to match old high school yearbook photographs with recent images taken of each classmate. Compared to an age-matched control group who were unfamiliar with the persons depicted in the photographs, former classmates displayed superior performance when matching the old and new photographs (control 33% vs. classmates 49%, chance performance was 10%). However, when a face was rated as very familiar in high school, and the face could also be named, the correct matching rate was as high as 84%.

The Before they were Famous Task (BTWF; Russell et al., 2009) also measures the ability to generalize across age, but unlike Bruck et al., participants recognize younger instances of celebrities, taken at a time before they were famous (usually in childhood). Participants are tasked with identifying the individual in each image by name or by some other non-ambiguous semantic information. Studies which have employed the BTWF task have shown that it is possible to recognize at least some of the identities in

this task despite the massive changes in appearance and the participants' lack of experience with these celebrities during their childhood. For example, Bindemann et al. (2014) found that 28% of past photographs could be recognized. Comparing these results to Bruck et al. (1991) suggests that if we are familiar with someone when they are young, we can more easily use this experience to recognize that person when they are older than the other way around. Indeed, recent evidence from Matthews et al. (2022) has shown that children are less sensitive to images showing a highly personally familiar face (their parent) taken before they were born compared to images taken after they were born.

The faces used by Bruck et al. (1991) and Bindemann et al. (2014) were pre-experimentally familiar faces. Less is known about how newly learned, previously unfamiliar faces can be recognized across age changes. In the present research, we familiarized participants with five face images of new identities. We then measured how well those same faces could be recognized 40 years later/earlier. This is considerably more challenging because there is limited variability available to learn at one age. Crucially, we aimed to determine whether the variability in the images was enough to generalize across age.

An additional aim of the proposed research was to determine whether the direction matters when generalizing across age. Facial ageing goes in one direction: from young to old. However, due to photography and access to people's past photographs on social media, ageing can also be reversed from old to young. For example, Angela Merkel was in the public eye in her role as Chancellor of Germany in her 50s and 60s and Karyn Parsons, who played Hilary Banks in the Fresh Prince of Bel-Air, was in the public eye in her 20s. How well could someone recognize Angela Merkel in her 20s and Karyn Parsons in her 50s (see Figure 1)?

Figure 1. Karyn Parsons in her 50s [Markus Wissmann / Alamy Stock Photo].

There is some evidence that the direction of ageing might play a role in the ability to generalize across age. Schneider and Carbon trained participants on a set of identities from eight younger (5–30 years of age) or eight recent images (40–60 years old). These participants were then asked to rate how typical instances of the trained identities were relative to their mental representation. Crucially, the images they were asked to rate were from across the entire age range (i.e., the trained ages and the untrained ages). These typicality ratings were subjected to a cluster analysis, revealing additional and more defined clusters in the data from participants who learned each identity with younger images compared to those who learned them from more recent images. This suggests that learning the target identities from younger, rather than older and more recent, instances results in participants being able to more readily pick up more finely-grained differences in recent instances. George and Hole (1998) examined how well participants could recognize faces across changes in age. Participants learned a series of facial images featuring children and teenagers (one image per identity). At test, participants were presented with either an identical image, a novel image taken at the same age, or an older or younger image. The researchers observed that participants were equally as good at recognizing the learned identities from same-age or older-age images, but they performed worse when the test face was younger. There are two possible explanations for this pattern of results. First, it could be that the direction of ageing matters, with the recognition of younger faces a more difficult task than the recognition of older faces. If we consider an evolutionary perspective proposed by Schneider and Carbon (2021), there is little benefit to recognizing individuals from past images. Photographs are a relatively recent invention; therefore, nearer-to-present experiences are more likely than long-past experiences to provide salient information for identity recognition (Schneider & Carbon).

Second, the younger faces in George and Hole (1998) might have been more difficult to recognize because they were less similar in appearance than the older versions. Faces undergo age-related changes, with greater structural changes occurring in childhood (e.g., craniofacial changes) than during adolescence/early adulthood (Enlow, 1982). For example, the age-related changes between 5 and 15 are greater than between 15 and 25. Therefore, the result may not have anything to do with ageing per se but may reflect the facial similarity between images, with younger child test faces in George and Hole (1998) being less similar to the learned instances than older test faces. In addition, young children's faces are also particularly difficult to recognize (Michalski et al., 2019), relative to adult faces (Kramer et al., 2018a; White et al., 2015), and the participants in George and Hole's study were adults. It remains to be seen whether an ageing asymmetry effect is present for face images depicting ageing during adulthood.

Here, we aimed to determine whether the ageing asymmetry effect could be observed using the computational model of face recognition developed by Mileva et al. (2020; Simulation 3). Then, we collected behavioural data to substantiate the modelling results. Participants/models were trained on a set of young adult faces (e.g., 20–30 years old), with each facial identity depicted in multiple ambient images. The ability to recognize instances of the trained identities at an older age (e.g., 60–70 years old) was then tested (young-to-old). We also ran a condition in which the age of the training/test faces was reversed (i.e., training with older instances and testing with younger instances of the same identities; old-to-young). Sequential matching was used in Experiment 2. This methodology has been found to produce better face learning than simultaneous matching, perhaps because it contains a memory component which forces participants to abstract identity information from the images contained in the learning set (see Sandford & Ritchie, 2021 who used arrays of up to three images per identity in the training set). In Experiments 3a and 3b we used a different memory paradigm in which the training stimuli were all presented during an initial learning phase before recognition was subsequently tested (similar to George & Hole, 1998). We were interested in whether performance across substantial changes in age would be above chance and whether the direction of age change mattered: is it easier to recognize faces that change from young to old (which happens in nature) than to recognize faces that change from old to young (which never happened before the invention of photography)?

Experiment 1: modelling data

Mileva et al. (2020) describe how two multidimensional models of face recognition deal with challenging differences across time – a PCA-based approach that captures physical image variability only (representing unfamiliar face processing) and a more guided approach where PCA is combined with a clustering technique based on additional top-down information about the identity each image belongs to (PCA + LDA, representing familiar face processing). These models were trained on images of 10 male identities taken between the 1960s and 2010s (depicting these identities from their early 20s to their 70s or 80s). When the two models were trained on images spanning the entire range of available images, both models performed well above chance, with a clear advantage from the additional top-down information in the PCA + LDA model. However, when the models were trained on images spanning a limited time period only (e.g., training on images from the 1960s only and testing on images from the 1980s), the PCA only model could not generalize across time, whereas the additional top-down information in the PCA + LDA model was sufficient to keep performance substantially above chance, even when there was a 40-year gap between the training and test images. Here, we take a similar approach to explore the effect of generalization direction on the overall recognition accuracy. That is, we ask whether it is easier to recognize someone in their 70s when we know how they used to look in their 20s (young-to-old condition) or to recognize them in their 20s when we know how they look in their 70s (old-to-young condition).

Method

Our general method follows the procedure for the PCA + LDA model, representing familiar face recognition, developed for Simulation 3 in Mileva et al. (2020). For consistency, the same set of 2100 images

was used to test recognition across time – it consisted of 210 images of each of 10 male identities (70 images for three pre-determined time periods: 1960s-70s, 1980-1990s, and 2000s-10s). Images were downloaded from the online archive Getty Images (https://www.gettyimages.co.uk/) – all were ambient images, represented in grayscale for consistency. In order to approximate our pre-existing experience with and knowledge of face variability, these 2100 images were represented in a multidimensional face space created by applying PCA to a much larger set of 6100 naturally-varying images (see Simulation 2: Background set images in Mileva et al., 2020). This way, each image was represented in this larger face space and described with a vector of 285 numbers that captured the location of this specific image within the first 285 face texture PCs in the space. These first PCs captured 95% of the variance in face texture. LDA was then applied to a training set of 600 images all from the same time period (60 images per identity) and a test set of 100 images from a different time period was projected in this newly-created LDA space. Finally, the Euclidean distance between each test and training image within the LDA space was calculated. Classification was considered accurate when the closest image to the test image belonged to the same identity. As the image set included images of 10 identities, chance classification accuracy is 10%.

This procedure was repeated for 49 iterations with seven different training image sets (all containing 600 images, 60 per identity, with each image appearing an equal number of times across the seven sets) and seven different test image sets (100 images, 10 per identity). We used these simulations to compare classification accuracy with young-to-old generalization, where images in the training set were taken earlier than images in the test set (e.g., training on images taken during the 60s and testing on images taken in the 80s) and old-to-young generalization, where images in the training set were taken after images in the test set (e.g., training on images taken in the 80s and testing on images taken in the 60s). Using the three pre-determined time periods, we did this for images taken approximately 20 years apart (1960s vs 1980s and 1980s vs 2000s) and for images taken approximately 40 years apart (1960s vs 2000).

Results and discussion

Figure 2 shows the average classification accuracy across the 49 iterations of the procedure, separately for each age gap and generalization direction

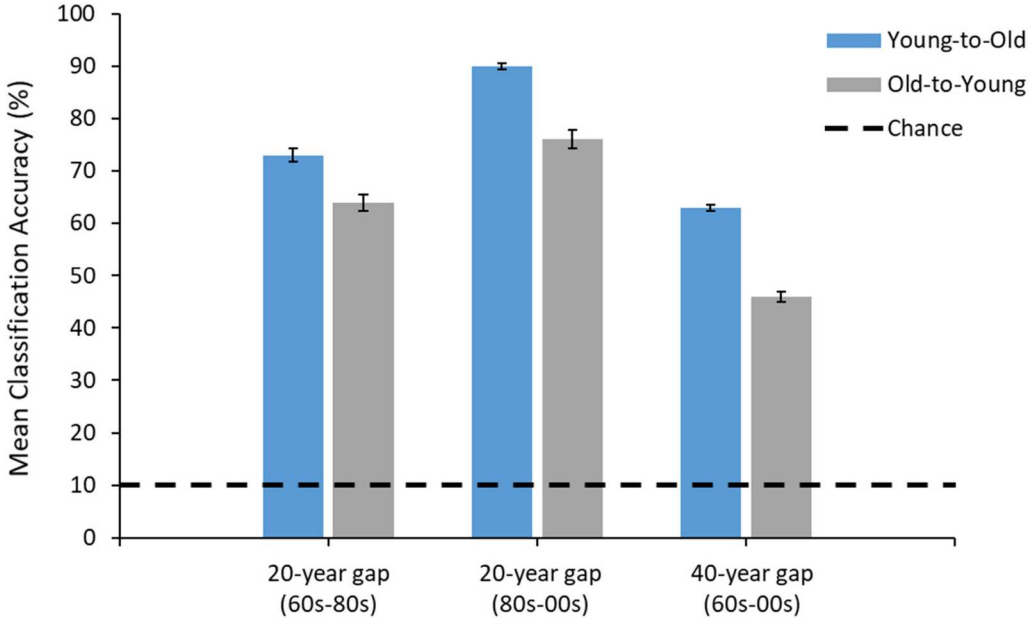

Figure 2. Mean PCA + LDA classification accuracy for young-to-old and old-to-young generalization at three different age gaps between the training and test images. Dashed line represents chance classification performance and error bars show the standard error across the 49 iterations of the classification procedure.

condition. A 2 × 3 between subjects ANOVA with factors: generalization direction (young-to-old vs old-to-young) and age gap (20-year gap, comparing generalization between images taken in the 1960s and 1980s, 20-year gap, comparing generalization between images taken in the 1980s and 2000s, and 40-year gap) showed significant main effects of generalization direction ($F(1, 288) = 190.83$, $p < .001$, $\eta_p^2 = .40$) and age gap ($F(2, 288) = 283.64$, $p < .001$, $\eta_p^2 = .66$), as well as a significant interaction between them ($F(2, 288) = 5.41$, $p < .005$, $\eta_p^2 = .04$). Simple main effects showed that classification was always significantly more accurate when generalizing from younger to older images. However, this effect was strongest with a 40-year gap between the training and test images ($F(1, 288) = 102.07$, $p < .001$, $\eta_p^2 = .26$) than with a 20-year gap ($F(1, 288) = 30.15$, $p < .001$, $\eta_p^2 = .09$ for the 60s-80s generalization and $F(1, 288) = 69.44$, $p < .001$, $\eta_p^2 = .19$ for the 80s-00s generalization).

These results suggest an effect of ageing direction. Training the model with younger faces and testing with older faces resulted in higher classification accuracy than training with old faces and testing with young faces. Next, we aim to replicate this effect with behavioural data.

Experiment 2: Sequentially matching faces across a 40-Year age gap

In Experiment 2, we used an independent-measures experimental design to examine (a) whether participants can recognize the same person across a 40-year age gap and (b) the effect of ageing direction on participants' ability to sequentially match faces. Participants were randomly allocated to one of four conditions. In the first condition, participants were trained on a set of young adult faces and were subsequently asked to match instances of the trained identities at an older age (young to old). In the second condition, participants were trained on a set of older adult faces and subsequently asked to match instances of the trained identities at a younger age (old to young). In the third condition, participants were trained on and tested with young adult faces (young to young), and in the fourth, older adult faces were used for training and testing (old to old). Response sensitivity and response bias were determined using signal detection measures d-prime (d') and criterion (c) respectively.

Method

Participants

Participants were 134 individuals (69 female, 65 male) aged between 25–35 ($M = 29.01$, $SD = 2.89$), recruited through Prolific (https://www.prolific.co/). Participants had normal or corrected to normal vision, were fluent in English, and had no known cognitive or face recognition impairments. All participants completed an electronic consent form prior to the commencement of the experimental task. Ethical approval for the three behavioural experiments in this paper was obtained from the Human Research Ethics Committee at the Open University.

Thirty-three participants completed the old-to-young (OTY) version of the task, 30 completed the young-to-old (YTO), 35 completed the old-to-old (OTO) and 36 completed the young-to-young (YTY). After completion of the matching task, participants' familiarity with the target faces was assessed. No participant could name more than one of the stimulus faces; thus, all participants were deemed to be unfamiliar with the stimulus set.

Materials

For the target stimuli, we collected 12 images each of 20 male European celebrities (actors, musicians, TV presenters etc.) who were active during the period 1940-2020. White male faces were selected in order to match the stimuli set used to train the PCA + LDA model in Experiment 1. For each individual, six photos were selected of them at a younger age (20-30 years old) and six images were selected of them at an older age (60-70 years old). For each identity, the gap between the younger and older faces was approximately 40 years.

An additional 40 images of male European celebrities (20 taken at a younger age, 20 taken at an older age) were selected for use as foils in the mismatch trials. Each of these foils was matched to a target face based on age and visual appearance.

Target and foil stimuli were selected from the online archive Getty Images which provides specific information about when exactly each image was taken. The images varied greatly in lighting, hairstyle, emotional expression and facial hair. Only images featuring a clear, unobstructed view of the face that could be dated were selected. The internal features were fully visible in each image and were not

Figure 3. Example images that represent the stimuli used in this experiment (and Experiment 3a). The first two (younger) images depict the subject at 25–28 years old. The second two (older) images depict the subject at 65–68 years old. (Due to restrictions on the reuse and distribution of the experiment images, it is not possible to reproduce them in this paper).

obscured by clothing or accessories. All images were converted to greyscale and resized to 200 × 300 pixels using GIMP (https://www.gimp.org/). Figure 3 provides an example of the types of image used in this experiment (and Experiments 3a and 3b).

Procedure

The experimental task was administered using the online experiment generator platform, Gorilla (www.gorilla.sc). Participants completed 20 experimental trials (10 matches, 10 mismatches), with each of the 20 target identities shown in either a matching or mismatching pair. Trial type was randomized between participants so that half of the participants saw a given target face in a match trial and half in a mismatch.

In each trial, participants were shown a 5-image array featuring a single target identity at either a younger or older age (e.g., five images of the French actor Alain Delon aged 30). These images were presented simultaneously on the screen, in a single horizontal line for 8000ms. After a 500ms interstimulus interval, participants were shown a test face. The test face was a single image of a similar age to the trained faces (in the OTO and YTY conditions) or a face at a different age (in the OTY and YTO conditions). In half of the trials (match trials), the test face was the same identity as in the array, whereas, in the other half (mismatch trials), the image was of a different identity. Participants were tasked with identifying whether the test face belonged to the same person as the previously presented array or a different person, novel to the experiment. Participants made their response by clicking on one of two buttons ("same" or "different") presented on the screen alongside the target face. Trial presentation order was randomized between participants.

After completing the main task, participants' familiarity with the target stimuli was assessed. Participants were shown the same twenty arrays again, one by one, and were asked to indicate how familiar they were with each face before completing the experiment, making their response on a sliding scale from 0 (unknown) to 9 (very familiar).

Results

To examine the impact of ageing direction (old-to-old, old-to-young, young-to-old and young-to-young) on face recognition ability, we performed separate one-way independent measures ANOVA on accuracy rates (percentage of correct trials), d-prime (d') and criterion (c). Table 1 presents the means and standard deviations for these performance metrics.

The results of our analysis revealed significant differences in accuracy, $F(3, 130) = 44.22$, $p < .001$, $\eta_p^2 = .51$, and response sensitivity, $F(3, 130) = 45.45$, $p < .001$, $\eta_p^2 = .52$, between the four age conditions. However, there was no observable effect of age

Table 1. Mean accuracy, sensitivity (d') and response bias (c) in the four age conditions.

	Accuracy (%)		D-prime (d')		Criterion (c)	
Condition	*M*	SD	*M*	SD	*M*	SD
OTO	90.4	10.9	2.49	0.86	0.02	0.28
OTY	65.8	11.5	0.82	0.67	0.07	0.4
YTO	67	8.96	0.87	0.5	0.02	0.35
YTY	82.5	10.2	1.87	0.68	0.03	0.43

direction on response bias, $F(3, 130) = 0.61$, $p = .61$, $\eta_p^2 = .02$. A further examination of the accuracy rates revealed that participants displayed higher accuracy in the OTO condition compared to the other three conditions (all $ps < .001$). Additionally, accuracy was higher in the YTY condition than in the OTY ($p < .001$) and YTO conditions ($p < .001$). However, there was no difference in accuracy between the YTO and OTY conditions ($p > .05$). Similar significant pairwise comparisons were observed in the analysis of the d' data.

To determine whether there was support for the alternative hypothesis (i.e., higher accuracy for the YTO condition than the OTY condition), we followed up the non-significant effects with Bayesian t-tests in JASP (Love et al., 2019). We used the default prior in JASP and found weak evidence for the alternative hypothesis with respect to task accuracy ($BF_{10} = 0.28$), response sensitivity ($BF_{10} = 0.27$) and response bias ($BF_{10} = 0.4$).

One sample *t*-tests revealed that d' was above chance in all four conditions: old-to-old, $t(34) = 17.2$, $p < .001$; old-to-young, $t(32) = 7.11$, $p < .001$; young-to-old, $t(29) = 9.55$, $p < .001$; and young-to-young, $t(35) = 16.4$, $p < .001$.

Experiment 3a & 3b: Old/new recognition memory test

The findings from Experiment 2 revealed that participants were better able to match faces that were similar in age than when they were significantly different in age. Importantly, however, performance was above chance even when age differed (both the YTO and OTY conditions were above chance). This suggests that the variability participants had learned from five images in the array was sufficient to generalize across a 40-year age gap, at least when averaging the data across all 20 identities.

The sequential matching task did not show any evidence of an age direction effect. Performance in the YTO and OTY conditions was equivalent. This contrasts with the modelling data presented in Experiment 1 and the findings from George and Hole (1998). The purpose of Experiment 3a and 3b was, therefore, to replicate the findings of Experiment 2, but this time with a learning paradigm similar to that used by George and Hole. Participants first completed a learning phase in which they were trained with five images of each of 10 new identities. In a subsequent test phase, participants' ability to recognize the faces over a 40-year age gap (Experiment 3a) and a 20-year age gap (Experiment 3b) was measured. Half of the participants in each experiment were trained on a set of young adult faces and were subsequently asked to recognize instances of the trained identities at an older age (YTO). The other half were trained on a set of older adult faces and were subsequently asked to recognize instances of the trained identities at a younger age (OTY).

Experiment 3a: 40-year Age Gap

Method

Participants. Participants were 65 individuals (32 female, 33 male) aged between 25–35 ($M = 30.28$, $SD = 3.17$), recruited through Prolific. Inclusion criteria were the same as for Experiment 2. Individuals who participated in Experiment 2 were excluded from this study. Thirty-two participants completed the OTY version of the task, and 33 completed the YTO. None of the participants could name/identify more than one of the target identities, as assessed via a post-task familiarity check. Therefore, all participants were deemed to be unfamiliar with the stimulus set.

Materials. This experiment used a subset of 80 images, selected from the stimuli used in Experiment 2. Ten target identities and 20 matching foils were selected from the original stimuli set. For each target identity, six photos showing them at a younger age (20-30 years old) were selected and six images were selected of them at an older age (60-70 years old). For each matching foil, a single image of them at an older or younger age was chosen.

Procedure. The experimental task was administered using Gorilla. In the first phase of the task (the learning phase), participants learned ten target identities at either a younger or older age. For each identity, participants were first asked to view a five-image array for 8000ms. Then, each of the five images was presented individually and sequentially for 2000ms each.

An initial pilot test ($N = 30$) suggested that the above procedure was insufficient for participants to accurately learn the target identities. Therefore, participants were also provided with the celebrity's first

name to encourage the processing of identities as both percepts and concepts (Schwartz & Yovel, 2016; Yovel & Abudarham, 2021). Previous research conducted by Juncu et al. (2020) has demonstrated that associating facial stimuli with corresponding names enhances participants' ability to differentiate between different individuals and integrate diverse images of the same individual into a unified mental representation. This suggests that including names not only contributes to the perceptual representation of facial features but also aids in the construction of a conceptual representation of an individual. Furthermore, the incorporation of names may capture the influence of top-down processes, mirroring the design of our modelling study, which integrated both bottom-up and top-down processing. Consequently, we might anticipate that the findings of this study will align more closely with those of Experiment 1 than Experiment 2. Names were shown onscreen underneath each array and individual image.

In the second phase of the task (the recognition phase), participants viewed the same ten identities from phase one at a different age: if they had seen younger faces in phase one, they saw older faces in phase two, and vice versa. The previously seen identities were randomly intermixed with ten novel faces of the same age. Participants were advised that they would be shown a series of facial images, and that some of the images would feature the individuals from the learning phase at a different age. Their task was to identify whether they had seen each face before (during the learning phase of the task) or whether it was completely novel. Participants responded by clicking on one of two buttons presented on the screen alongside the target face. If they had seen the face during the learning phase, they were asked to click on a box labelled "previously seen". If they had not seen the face before, they selected a button marked "not seen before."

After completing the main task, participants' familiarity with the target stimuli was assessed. Participants were shown each of the ten arrays again and were asked to identify the given celebrity, either by their full name or by some other unique semantic information (e.g., a film or tv show they had starred in). As participants were provided with the celebrity's first name during the learning phase of the task, they were only considered to be familiar with the celebrity if they knew the full name (i.e., first name and surname). Participants typed their responses into open-ended text boxes that appeared on the screen underneath the array.

Results. An independent samples *t*-test revealed no significant difference in accuracy between the two age conditions, $t(63) = 0.94$, $p = .35$, $d = 0.23$, $BF_{10} = 0.37$, suggesting that participants were equally accurate when recognizing faces from old to young ($M = 51.9$, $SD = 9.31$) than from young to old ($M = 54.1$, $SD = 9.72$). Furthermore, there was no effect of age direction on d-prime, $t(63) = 0.72$, $p = .48$, $d = 0.18$, $BF_{10} = 1.45$, or criterion, $t(63) = 0.85$, $p = .4$, $d = 0.21$, $BF_{10} = 0.35$.

One sample *t*-tests revealed that d-prime was significantly above chance in the young-to-old condition, $t(32) = 2.55$, $p = .02$, $d = 0.44$. However, recognition performance did not differ significantly from chance in the old-to-young condition, $t(31) = -0.05$, $p = .13$, $d = 0.28$. In Experiment 3b, we therefore aimed to make the task easier by reducing the age gap between the learning and test faces.

Experiment 3b: 20-year Age Gap

Method

The method used in Experiment 3b was the same as in 3a, apart from the details listed below.

Participants

Participants were 60 individuals (43 female, 17 male) aged between 22–34 ($M = 29.18$, $SD = 3.12$). Thirty participants completed the OTY version of the task, and 30 completed the YTO. None of the participants could name/identify more than one of the target identities, as assessed via a post-task familiarity check. Therefore, all participants were deemed to be unfamiliar with the stimulus set.

Materials

This experiment used the same older images as Experiment 3a. For each individual, a new set of six younger photos (age range = 40-50) was sourced using the same criteria described previously. For each identity, the gap between the younger and older faces was approximately 20 years.

Results

An independent samples *t*-test revealed no significant difference in accuracy between the two age conditions, $t(63) = 0.77$, $p = 0.45$, $d = 0.19$, $BF_{10} = 0.34$, suggesting that participants were equally accurate when recognizing faces from old to young ($M = 65.3$, $SD = 8.9$) than from young to old ($M = 63.2$, $SD = 12.7$). Moreover, there was no significant effect of age direction on either d-prime, $t(52.8) = 1.13$, $p = .26$, $d = 0.30$, $BF_{10} = 0.45$, or criterion, $t(63) = 1.49$, $p = .14$, $d = 0.39$, $BF_{10} = 0.66$.

One sample t-tests revealed that sensitivity rates were significantly above chance in both the young-to-old, $t(29) = 5.57$, $p < .001$, $d = 1.02$, and old-to-young conditions, $t(29) = 9.63$, $p < .001$, $d = 1.76$ (see Figure 4).

Taken together, Experiments 3a and 3b show no evidence of an effect of ageing direction.

Measuring image similarity in old and young arrays

The modelling data in Experiment 1 showed an effect of ageing direction that was not present in the behavioural data in Experiments 2, 3a & b. Previous research suggests that exposure to increased variability typically facilitates face learning (e.g., Andrews et al., 2015; Murphy et al., 2015; Ritchie & Burton, 2017). One possible explanation for our observed findings is that younger facial images may inherently possess greater variability, with the PCA + LDA model able to utilize this variability more effectively for identity recognition compared to human perceivers. To explore this hypothesis further, we examined whether the image variability in the arrays was comparable for the young and older face sets used in Experiments 2, 3a & b by applying PCA to these images.

Pre-processing & principal components analysis (PCA)

All images were first resized to 380×570 pixels and represented in a.bmp format. PCA was performed using InterFace software (Kramer et al., 2017a) and as part of this process face shape and face texture are separated. Face shape was determined by aligning 82 fiducial points to each face using a semi-automatic algorithm and texture is determined by warping each image to a standard shape (see Burton et al., 2016 for a more detailed description of

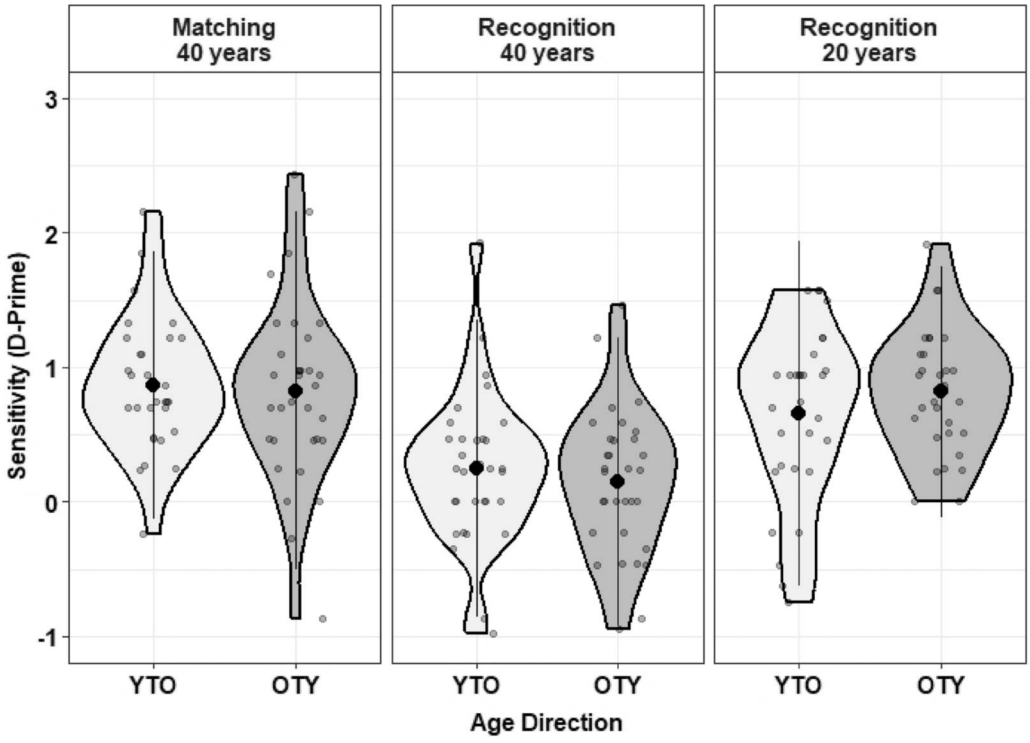

Figure 4. Plots showing sensitivity (d') rates from Experiments 2 (Matching 40 years), 3a (Recognition 40 years) and 3b (Recognition 20 years).

the process). In order to approximate people's existing knowledge and experience with faces, a PCA was used on a large face set containing 6100 images (see the description of the "background set" in Mileva et al., 2020). This was done in order to more accurately represent the variability in pose, lighting, emotional expressions, age, image quality, etc. we experience in everyday life. The PCA generated a 100-dimensional face shape and face texture spaces, with these 100 Principal Components (PCs) explaining 99.9% of shape variance and 91.6% of texture variance. The images included in the old and young arrays used in Experiment 2 and 3b, were then represented in these shape and texture spaces, producing a shape and texture vector for each image that represented where this image was located in our large face space. Image similarity was measured by calculating the Euclidean distance between pairs of images included in the old and young arrays. This way, images that are located closer together should be more similar (in shape or in texture) than images that are located further apart from one another.

Each array consisted of five images, therefore, a total of 10 pairwise distances were calculated, separately for old and young arrays showing each of the 10 identities. Figure 5 shows the average Euclidean distance within the face shape and face texture spaces for the old and young image arrays used in Experiment 2 and 3a. While old and young arrays seem to show comparable levels of variability in face shape ($t(9) = 0.74$, $p = .477$), young arrays contained significantly higher levels of variability in face texture ($t(9) = 4.51$, $p = .001$).

Figure 6 shows the average Euclidean distance within the face shape and face texture spaces for old and young image arrays used in Experiment 3b. Old and young arrays seem to show comparable levels of variability in both face shape ($t(9) = 1.18$, $p = .269$) and face texture ($t(9) = 2.10$, $p = .065$), though the latter comparison was approaching significance, with more texture variability in the young, rather than in the old, arrays.

The similarity analysis suggests there were higher levels of variability in the younger arrays than the older arrays. This is relevant to the findings presented in Experiments 2 and 3. Previous research using PCA + LDA has only used texture components as there is evidence that identity is more strongly carried by the texture components (Andrews et al., 2015; Rogers et al., 2022). In Experiment 1, we found that when using PCA + LDA, generalizing from young to

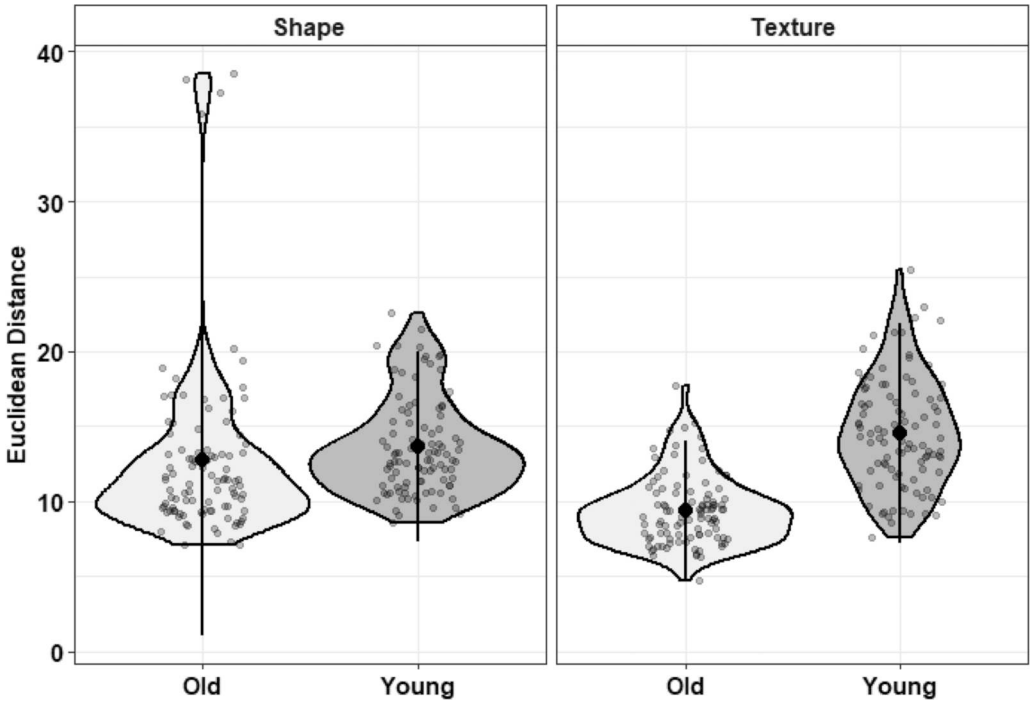

Figure 5. Plots showing the image similarity data from Experiments 2 and 3a for face shape and texture.

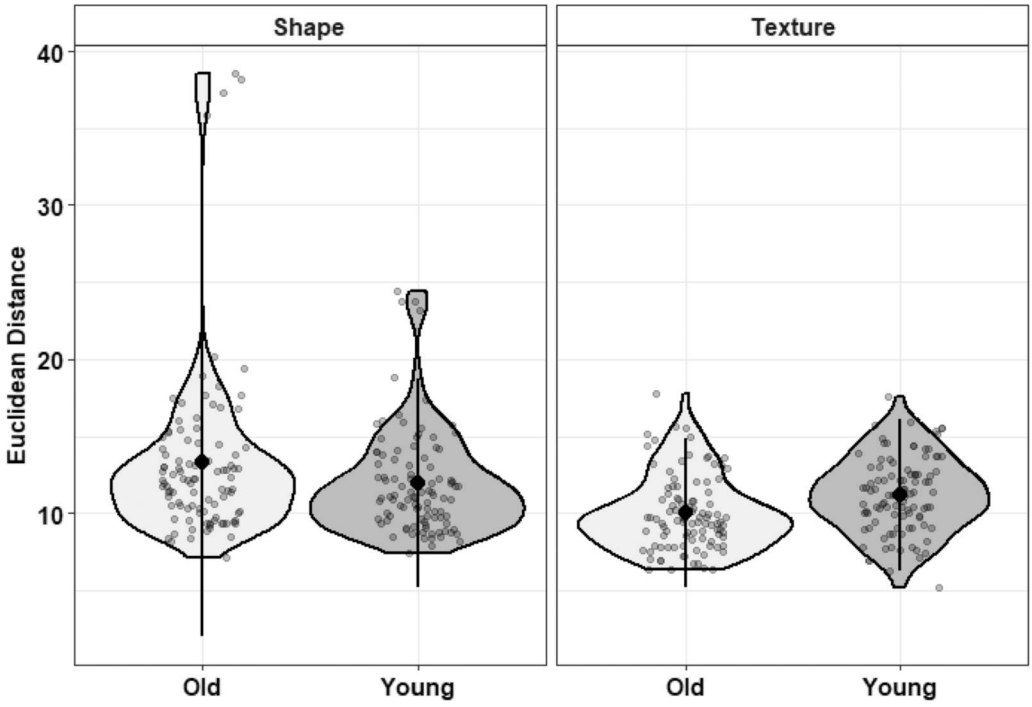

Figure 6. Plots showing the image similarity data from Experiment 3b for face shape and texture.

old is easier (i.e., classification is more accurate) than generalizing from old to young. One possibility is that the higher levels of texture variability enhance the algorithm accuracy (more variability in the young set means better learning because the modelling approach can make good use of this extra variability), but human performance may not benefit from the same additional textural information because we cannot make such good use of this increased variability.

General discussion

Our results show that recognizing faces across changes in age is a challenging task. We found impaired performance when matching faces across a 40-year age gap relative to no age gap (Experiment 2). Adding in a longer-term memory component made the task more challenging (Experiment 3a) and reducing the age gap to 20 years improved performance (Experiment 3b). In addition, while we found a clear effect of direction of ageing in the modelling data (Experiment 1), we found no evidence for this across three behavioural experiments (Experiment 2, 3a and 3b). A measure of image similarity revealed that the younger arrays contained more variability than the older arrays, raising the possibility that the computational model used in Experiment 1 was better able to exploit variability than human participants. The theoretical implications of the findings are discussed below.

Idiosyncratic variability generalizes across age

Like previous research using face matching (Davis & Valentine, 2009; Megreya et al., 2013; Meissner et al., 2013; Mileva et al., 2020), in Experiment 2 we found that adding a substantial age gap between the array and test face reduced accuracy on a sequential matching task. Nevertheless, participants were able to recognize facial identity across substantial changes in age across almost all conditions in all three experiments (OTY – Experiment 3a was the exception). Performance was poor, but recognition accuracy was still at above-chance levels. This suggests the variability contained in five images depicting a person at one age was enough to enable generalization to a very different age for at least some of the identities.

General knowledge about ageing might be a mechanism by which we are able to predict how someone might look at a very different age. However, this is limited by the extent to which ageing reflects generic vs. idiosyncratic changes. Recent theories of face recognition suggest that

learning a face involves learning the idiosyncratic ways in which the face varies (Burton et al., 2016). For example, the actor Harrison Ford has an idiosyncratic smile that can be observed in both younger and older photographs of him. It is, therefore, likely that despite substantial age-related changes in facial structure and texture, idiosyncratic features remain. Learning a face from multiple images may help the viewer to determine the nature of these diagnostic idiosyncratic features.

Computational modelling and human behaviour

Experiment 1 revealed an ageing asymmetry effect in a computer model of face recognition that was not present in three behavioural experiments. There are, however, important differences between the model and human participants. Most notably, the model was trained on considerably more images (60 per identity) than the participants in Experiments 2 and 3 (five per identity). The PCA + LDA procedure required at least 27 images per identity, making it impossible to fully simulate the behavioural experiments. Any comparison between the two is therefore limited, and we can only speculate about why the model produced different results to human participants. For example, based on the results of the similarity analysis we propose that increased variability might enhance the performance of the computational model, but humans may not be able to take advantage to the same extent.

Although we found above-chance performance in Experiments 2 and 3b, perhaps our experimental tasks were not sensitive enough to detect an ageing direction effect because participants were at the limits of capability. The tasks were challenging, and it was remarkable that participants could generalize across 40 years at all, especially given that they had only learned each of the faces from five images. Future research should increase the number of images in the learning set (or use real-world faces that are familiar at one age only) to determine if an effect of ageing direction is present when faces are more familiar to participants.

Variability in older and younger faces

Participants were young adults; therefore, you might expect an own-age bias (OAB) with younger faces being easier to recognize than older faces (e.g., Anastasi & Rhodes, 2005; Rhodes & Anastasi, 2012; Wiese et al., 2013). This is not what we found in practice. Experiment 2 found that performance was better in the OTO condition than the YTY condition. There are a few possible explanations for this. Firstly, previous research from Proietti et al. (2019) has found the OAB is not present when participants are shown images containing variability as this encourages participants to attend to individuating information. Second, variations in image quality and resolution between the younger images (captured during the 1940s-60s) and the older images (captured during the 1980s-2000s) may have played a role in the improved learning and recognition of the older faces in Experiment 2. Interestingly, despite the enhanced quality of the older images, there seems to be no discernible effect on performance when it comes to recognizing faces across different age transitions (e.g., in the YTO and OTY conditions). It's plausible that the increased variability inherent in the younger images might have offset any potential drawbacks arising from their lower image quality. Alternatively, differences in image properties between the images taken in different time periods (e.g., the 1950s and the 1990s) might be less important than identity-specific information contained in those images. Evidence to support this comes from the simulations in Mileva et al. (2020), who tested classification in two ways. For the main simulations, classification was tested as a 1 in 10. This was because there were 10 identities and for every test image they found the closest training image. If it belonged to the same identity, classification was correct, if not, it was incorrect. Mileva et al. also used a slightly different approach where they trained the model in such a way that each identity in each of the three time periods (60s-70s, 80s-90s and 00s-10s) was represented as a separate person. Classification in this simulation was therefore 1 in 30. This approach allowed Mileva et al to further examine what mistakes the model would make – would it pick an image showing a different identity from the same time period (because of similarities in the properties of images taken during this time) or whether it would pick an image showing the same identity but from a different time period? In all cases, most of the mistakes were when picking an image showing the same person but from a different time

period. This shows that the models could easily overcome the clear differences in the properties of these images (i.e., they were not clustering images taken during the same period together, but images belonging to the same person). An interesting direction for future research will be to determine whether humans make a similar pattern of errors.

Finally, it is possible that because the older adult faces did not vary as much across instances, this resulted in the recognition task being easier as the test face was more similar to the face images in the training set. However, this should also be detrimental to face learning. Exposure to increased variability typically facilitates learning (e.g., Andrews et al., 2015; Murphy et al., 2015; Ritchie & Burton, 2017), so in theory the more variable younger faces should be easier to learn than the less variable older faces. The nature of within-person variability as we age is an interesting topic that has received relatively little attention. If our finding that older faces are less variable can be replicated across different training sets then this has practical implications, for example, photo identification should be easier to recognize/match for older faces than for younger ones.

However, it should be noted that PCA primarily emphasizes shape and texture characteristics and while it can capture some elements of hair (e.g., hair colour and presence of hair through differences in light and dark, the presence of a fringe or facial hair), it cannot account for exact differences in hairstyle. These external features have been shown to hold significance, particularly for human observers, in the recognition of unfamiliar identities (Ellis et al., 1979; Latif & Moulson, 2022). Therefore, PCA's inherent limitations in considering these external features may have led to an underestimation of the overall variability in the facial images. Additionally, the faces used in the present research were celebrities and all males. Therefore, future work should examine this in non-celebrity faces that are more representative of the population (e.g., women, people of different ethnicities).

Conclusion

The focus of the present research was to examine face recognition despite substantial age-related changes. We showed an effect of ageing direction for a computer model of face recognition, but the effect was not present in three behavioural experiments. Above-chance recognition of faces across a 20- and 40-year age gap suggests that people can extract identity diagnostic information despite substantial changes in outward appearance. An interesting direction for future research would be to establish exactly what identity-diagnostic information is required to successfully recognize the same person across a substantial change in age.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was funded by an Economic and Social Research Council New Investigator grant awarded to Sarah Laurence (ES/R005788/2).

ORCID

Sarah Laurence http://orcid.org/0000-0002-3771-5910

References

Anastasi, J. S., & Rhodes, M. G. (2005). An own-age bias in face recognition for children and older adults. *Psychonomic Bulletin & Review*, *12*(6), 1043–1047. https://doi.org/10.3758/BF03206441

Andrews, S., Jenkins, R., Cursiter, H., & Burton, A. M. (2015). Telling faces together: Learning new faces through exposure to multiple instances. *Quarterly Journal of Experimental Psychology*, *68*(10), 2041–2050. https://doi.org/10.1080/17470218.2014.1003949

Baker, K. A., Laurence, S., & Mondloch, C. J. (2017). How does a newly encountered face become familiar? The effect of within-person variability on adults' and children's perception of identity. *Cognition*, *161*, 19–30. https://doi.org/10.1016/j.cognition.2016.12.012

Bindemann, M., Attard, J., & Johnston, R. A. (2014). Perceived ability and actual recognition accuracy for unfamiliar and famous faces. *Cogent Psychology*, *1*(1), 986903. https://doi.org/10.1080/23311908.2014.986903

Bruck, M., Cavanagh, P., & Ceci, S. J. (1991). Fortysomething: Recognizing faces at one's 25th reunion. *Memory & Cognition*, *19*(3), 221–228. https://doi.org/10.3758/BF03211146

Burton, A. M., Kramer, R. S., Ritchie, K. L., & Jenkins, R. (2016). Identity from variation: Representations of faces derived from multiple instances. *Cognitive Science*, *40*(1), 202–223. https://doi.org/10.1111/cogs.12231

Burton, A. M., Schweinberger, S. R., Jenkins, R., & Kaufmann, J. M. (2015). Arguments against a configural processing account of familiar face recognition. *Perspectives on*

Psychological Science, *10*(4), 482–496. https://doi.org/10.1177/1745691615583129

Coleman, S. R., & Grover, R. (2006). The anatomy of the aging face: Volume loss and changes in 3-dimensional topography. *Aesthetic Surgery Journal*, *26*(Suppl 1), S4–S9. https://doi.org/10.1016/j.asj.2005.09.012

Davis, J. P., & Valentine, T. (2009). CCTV on trial: Matching video images with the defendant in the dock. *Applied Cognitive Psychology*, *23*(4), 482–505. https://doi.org/10.1002/acp.1490

De Jager, S., Coetzee, N., & Coetzee, V. (2018). Facial adiposity, attractiveness, and health: A review. *Frontiers in Psychology*, *9*, 2562. https://doi.org/10.3389/fpsyg.2018.02562

Djordjevic, J., Zhurov, A. I., Richmond, S., & Visigen Consortium. (2016). Genetic and environmental contributions to facial morphological variation: A 3D population-based twin study. *PLoS One*, *11*(9), e0162250. https://doi.org/10.1371/journal.pone.0162250

Edelstein, D. R. (1996). Aging of the normal nose in adults. *The Laryngoscope*, *106*(S81), 1–25. https://doi.org/10.1097/00005537-199609001-00001

Ellis, H. D., Shepherd, J. W., & Davies, G. M. (1979). Identification of familiar and unfamiliar faces from internal and external features: Some implications for theories of face recognition. *Perception, 8*(4), 431–439. https://doi.org/10.1068/p080431

Enlow, D. H. (1982). *Handbook of facial growth* (2nd ed.). W. B. Saunders Company.

Freire, A., Lee, K., & Symons, L. A. (2000). The face-inversion effect as a deficit in the encoding of configural information: Direct evidence. *Perception*, *29*(2), 159–170. https://doi.org/10.1068/p3012

Friedman, O. (2005). Changes associated with the aging face. *Facial Plastic Surgery Clinics of North America*, *13*(3), 371–380. https://doi.org/10.1016/j.fsc.2005.04.004

George, P. A., & Hole, G. J. (1998). Recognising the ageing face: The role of age in face processing. *Perception*, *27*(9), 1123–1134. https://doi.org/10.1068/p271123

Hancock, P. J., Bruce, V., & Burton, A. M. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Sciences*, *4*(9), 330–337. https://doi.org/10.1016/S1364-6613(00)01519-9

Hole, G. J., George, P. A., Eaves, K., & Rasek, A. (2002). Effects of geometric distortions on face-recognition performance. *Perception*, *31*(10), 1221–1240. https://doi.org/10.1068/p3252

Honig, T., Shoham, A., & Yovel, G. (2022). Perceptual similarity modulates effects of learning from variability on face recognition. *Vision Research*, *201*, 108128. https://doi.org/10.1016/j.visres.2022.108128

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

Juncu, S., Blank, H., Fitzgerald, R. J., & Hope, L. (2020). Do image variability and names in missing person appeals improve prospective person memory? *Journal of Applied Research in Memory and Cognition*, 9(3), 410–418. https://doi.org/10.1037/h0101854

Kramer, R. S., Jenkins, R., & Burton, A. M. (2017a). InterFace: A software package for face image warping, averaging, and principal components analysis. *Behavior Research Methods*, *49*(6), 2002–2011. https://doi.org/10.3758/s13428-016-0837-7

Kramer, R. S., Jenkins, R., Young, A. W., & Burton, A. M. (2017b). Natural variability is essential to learning new faces. *Visual Cognition*, *25*(4-6), 470–476. https://doi.org/10.1080/13506285.2016.1242522

Kramer, R. S., Mulgrew, J., & Reynolds, M. G. (2018a). Unfamiliar face matching with photographs of infants and children. *PeerJ*, *6*, e5010. https://doi.org/10.7717/peerj.5010

Kramer, R. S., Young, A. W., & Burton, A. M. (2018b). Understanding face familiarity. *Cognition*, *172*, 46–58. https://doi.org/10.1016/j.cognition.2017.12.005

Latif, M., & Moulson, M. C. (2022). The importance of internal and external features in recognizing faces that vary in familiarity and race. *Perception*, *51*(11), 820–840. https://doi.org/10.1177/03010066221122299

Longmore, C. A., Santos, I. M., Silva, C. F., Hall, A., Faloyin, D., & Little, E. (2017). Image dependency in the recognition of newly learnt faces. *Quarterly Journal of Experimental Psychology*, *70*(5), 863–873. https://doi.org/10.1080/17470218.2016.1236825

Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, J., Ly, A., Gronau, Q. F., Smíra, M., Epskamp, S., Matzke, D., Wild, A., Knight, P., Rouder, J. N., Morey, R. D., & Wagenmakers, E. J. (2019). JASP: Graphical statistical software for common statistical designs. *Journal of Statistical Software*, *88*(2), 1–17. https://doi.org/10.18637/jss.v088.i02

Matthews, C. M., & Mondloch, C. J. (2018). Improving identity matching of newly encountered faces: Effects of multi-image training. *Journal of Applied Research in Memory and Cognition*, *7*(2), 280–290. https://doi.org/10.1016/j.jarmac.2017.10.005

Matthews, C M, & Mondloch, C J. (2022). Learning faces from variability: Four-and five-year-olds differ from older children and adults. *Journal of Experimental Child Psychology*, *213*, 105259. https://doi.org/10.1016/j.jecp.2021.105259

Matthews, C. M., Mondloch, C. J., Lewis-Dennis, F., & Laurence, S. (2022). Children's ability to recognize their parent's face improves with age. *Journal of Experimental Child Psychology, 223*, 105480. https://doi.org/10.1016/j.jecp.2022.105480

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences*, *6*(6), 255–260. https://doi.org/10.1016/S1364-6613(02)01903-4

Mayes, A. E., Murray, P. G., Gunn, D. A., Tomlin, C. C., Catt, S. D., Wen, Y. B., Zhou, L. P., Wang, H. Q., Catt, M. & Granger, S. P. (2010). Environmental and lifestyle factors associated with perceived facial age in Chinese women. *PLoS One*, *5*(12), e15270. https://doi.org/10.1371/journal.pone.0015270

Megreya, A. M., Sandford, A., & Burton, A. M. (2013). Matching face images taken on the same day or months apart: The limitations of photo ID. *Applied Cognitive Psychology*, *27*(6), 700–706. https://doi.org/10.1002/acp.2965

Meissner, C. A., Susa, K. J., & Ross, A. B. (2013). Can I see your passport please? Perceptual discrimination of own- and other-race faces. *Visual Cognition*, *21*(9-10), 1287-1305. https://doi.org/10.1080/13506285.2013.832451

Menon, N., White, D., & Kemp, R. I. (2015). Variation in photos of the same face drives improvements in identity verification. *Perception*, *44*(11), 1332–1341. https://doi.org/10.1177/0301006615599902

Michalski, D., Heyer, R., & Semmler, C. (2019). The performance of practitioners conducting facial comparisons on images of children across age. *PLoS One*, *14*(11), e0225298. https://doi.org/10.1371/journal.pone.0225298

Mileva, M., Young, A. W., Jenkins, R., & Burton, A. M. (2020). Facial identity across the lifespan. *Cognitive Psychology*, *116*, 101260. https://doi.org/10.1016/j.cogpsych.2019.101260

Murphy, J., Ipser, A., Gaigg, S. B., & Cook, R. (2015). Exemplar variance supports robust learning of facial identity. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(3), 577. https://doi.org/10.1037/xhp0000049

Ohm, E., & Silness, J. (1999). Size of the mandibular jaw angle related to age, tooth retention and gender. *Journal of Oral Rehabilitation*, *26*(11), 883–891. https://doi.org/10.1046/j.1365-2842.1999.00464.x

Panhard, S., Lozano, I., & Loussouarn, G. (2012). Greying of the human hair: A worldwide survey, revisiting the '50' rule of thumb. *British Journal of Dermatology*, *167*(4), 865–873. https://doi.org/10.1111/j.1365-2133.2012.11095.x

Proietti, V., Laurence, S., Matthews, C. M., Zhou, X., & Mondloch, C. J. (2019). Attending to identity cues reduces the own-age but not the own-race recognition advantage. *Vision Research*, *157*, 184–191. https://doi.org/10.1016/j.visres.2017.11.010

Rhodes, M. G., & Anastasi, J. S. (2012). The own-age bias in face recognition: A meta-analytic and theoretical review. *Psychological Bulletin*, *138*(1), 146. https://doi.org/10.1037/a0025750

Ritchie, K. L., & Burton, A. M. (2017). Learning faces from variability. *Quarterly Journal of Experimental Psychology*, *70*(5), 897–905. https://doi.org/10.1080/17470218.2015.1136656

Rogers, D., Baseler, H., Young, A. W., Jenkins, R., & Andrews, T. J. (2022). The roles of shape and texture in the recognition of familiar faces. *Vision Research*, *194*, 108013. https://doi.org/10.1016/j.visres.2022.108013

Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic Bulletin & Review*, *16*(2), 252–257. https://doi.org/10.3758/PBR.16.2.252

Sandford, A., & Ritchie, K. L. (2021). Unfamiliar face matching, within-person variability, and multiple-image arrays. *Visual Cognition*, *29*(3), 143–157. https://doi.org/10.1080/13506285.2021.1883170

Schneider, T. M., & Carbon, C. C. (2021). The Episodic Prototypes Model (EPM): On the nature and genesis of facial representations. *i-Perception*, *12*(5), 204166952110541. https://doi.org/10.1177/20416695211054105

Schwartz, L., & Yovel, G. (2016). The roles of perceptual and conceptual information in face recognition. *Journal of Experimental Psychology: General*, *145*(11), 1493. https://doi.org/10.1037/xge0000220

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), e0139827.

Wiese, H., Komes, J., & Schweinberger, S. R. (2013). Ageing faces in ageing minds: A review on the own-age bias in face recognition. *Visual Cognition*, *21*(9-10), 1337–1363. https://doi.org/10.1080/13506285.2013.823139

Yovel, G., & Abudarham, N. (2021). From concepts to percepts in human and machine face recognition: A reply to Blauch, Behrmann & Plaut. *Cognition*, *208*, 104424. https://doi.org/10.1016/j.cognition.2020.104424

Zhou, X., Matthews, C. M., Baker, K. A., & Mondloch, C. J. (2018). Becoming familiar with a newly encountered face: Evidence of an own-race advantage. *Perception*, *47*(8), 807–820. https://doi.org/10.1177/0301006618783915