Article:

Andrews, Timothy J orcid.org/0000-0001-8255-9120, Rogers, Daniel, Mileva, Mila orcid.org/0000-0003-0537-9702 et al. (3 more authors) (2023) A narrow band of image dimensions is critical for face recognition. Vision Research. 108297. ISSN 0042-6989

https://doi.org/10.1016/j.visres.2023.108297

Vision Research 212 (2023) 108297

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

A narrow band of image dimensions is critical for face recognition

Timothy J. Andrews *, Daniel Rogers, Mila Mileva, David M. Watson, Ao Wang, A. Mike Burton

Department of Psychology, University of York, York YO10 5DD, UK

ARTICLE INFO

Keywords:
Shape
Texture
Recognition
Face
Dimensions
Identity

ABSTRACT

A key challenge in human and computer face recognition is to differentiate information that is diagnostic for identity from other sources of image variation. Here, we used a combined computational and behavioural approach to reveal critical image dimensions for face recognition. Behavioural data were collected using a sorting and matching task with unfamiliar faces and a recognition task with familiar faces. Principal components analysis was used to reveal the dimensions across which the shape and texture of faces in these tasks varied. We then asked which image dimensions were able to predict behavioural performance across these tasks. We found that the ability to predict behavioural responses in the unfamiliar face tasks increased when the early PCA dimensions (i.e. those accounting for most variance) of shape and texture were removed from the analysis. Image similarity also predicted the output of a computer model of face recognition, but again only when the early image dimensions were removed from the analysis. Finally, we found that recognition of familiar faces increased when the early image dimensions were removed, decreased when intermediate dimensions were removed, but then returned to baseline recognition when only later dimensions were removed. Together, these findings suggest that early image dimensions reflect ambient changes, such as changes in viewpoint or lighting, that do not contribute to face recognition. However, there is a narrow band of image dimensions for shape and texture that are critical for the recognition of identity in humans and computer models of face recognition.

1. Introduction

The ability to recognise a person from their face is fundamental to the way we interact with them. Models of face processing propose that faces are first represented in a pictorial code that contains detailed information about the image, but is then transformed into a more abstract structural code that can be used for perception (Bruce & Young, 1986, 2012). This transformation from a pictorial to a structural representation is important because, as we interact with faces in a natural environment, the shape and texture of a face can vary dramatically due to movement of the head and changes in lighting. To be useful, the cognitive processes involved in recognition must be able to ignore these ambient image changes to reveal an invariant, structural representation that can be useful for recognition (Burton, 2013).

The distinction between familiar and unfamiliar faces demonstrates the transformation from a pictorial to a structural code. While photographs of unfamiliar faces can be remembered and later recognised remarkably well, recognition performance with unfamiliar faces degrades as soon as any changes are made between learnt and test images (Bruce, 1982; Longmore, Liu, & Young, 2008; Hancock, Bruce, & Burton, 2000; Kemp, Towell, & Pike, 1997). In contrast, the behavioural hallmark of familiar face recognition is that it is remarkably successful across substantial changes in expression, viewing angle, and lighting conditions (Bruce, 1994; Bruce & Young, 2012; Burton, 2013). Models of face recognition propose key structural representations for familiar faces that are known as Face Recognition Units (FRUs), which selectively respond to faces from a particular identity (Bruce & Young, 1986; Burton et al., 1990). Although most theories of face recognition recognise the importance of FRUs, or some similarly abstractive representation, the nature of the visual properties that are used in this structural code remains unresolved.

A distinction between shape and texture is often used to investigate which visual properties are important for recognition. For example, all face images consist of a set of edges created by abrupt changes in reflectance due to the shapes and positions of facial features and a broader pattern of reflectance based on the surface properties of the face known as texture (Andrews et al., 2016). A range of evidence suggests that the texture of the face is more important than shape for the recognition of identity (Hole et al., 2002; Burton, Jenkins, Hancock & White, 2005; Russell et al., 2007; Russell & Sinha, 2007). For example,

* Corresponding author.
E-mail address: timothy.andrews@york.ac.uk (T.J. Andrews).

https://doi.org/10.1016/j.visres.2023.108297
Received 14 December 2022; Received in revised form 7 July 2023; Accepted 12 July 2023
Available online 30 July 2023

familiar face recognition is not substantially affected if the texture is presented on a standardized shape (Burton et al., 2005), or when face shape is distorted by stretching the image (Hole et al., 2002). In contrast, line drawings of faces, which contain shape information, but lack any texture, are not usually sufficient for recognition (Davies et al., 1978; Leder, 1999). Similarly, recognition of facial identity is disrupted by contrast-reversal, which causes large changes in texture but does not affect the shape of the face (Bruce & Langton, 1994; Russell et al., 2006; Harris, Young & Andrews, 2014). Nevertheless, there are some changes in the texture of the face, such as changes in the direction of lighting, that are unlikely to be diagnostic of identity.

Certain aspects of shape or 'second order configural properties' of a face (Maurer et al, 2002) have also been suggested to play an important role in face recognition (McKone & Yovel, 2009; Tanaka and Gordon, 2011; Piepers & Robbins, 2012). Empirical support for this perspective comes from studies that show shape information can be used to discriminate unfamiliar face images (O'Toole et al., 1999; Jiang, Blanz & O'Toole, 2006; Russell et al., 2007; Russell & Sinha, 2007, Caharel et al., 2009; Jiang, Blanz and Rossion, 2011; Lai, Oruc & Barton, 2013; Itz et al., 2016). However, it is not clear whether shape information is a useful cue for the recognition of identity across naturally varying images of familiar identities (Burton et al., 2015a). One challenge for a configural account of face recognition is that shape cues (particularly from the internal features of the face) can vary quite dramatically as a result of rigid and non-rigid movements, such that the spatial distances between features can often vary as much within a person as between people (Burton et al., 2015b).

The aim of this study was to investigate how shape and texture contribute to the recognition of identity in humans and computer models of face perception. Principal components analysis (PCA) is one of a number of related techniques used to derive statistical descriptions of image sets (Moon & Phillips, 2001; Gong et al., 2000) and has been used to explain the variance in face images that can be related to perception (Turk & Pentland, 1991; O'Toole et al., 1993; Hancock, Burton & Bruce, 1996; Calder et al., 2001; Jozwik et al., 2022). When PCA is applied to faces, it delivers novel dimensions (eigenfaces), which together form a face-space, within which to characterise any face image (Scheuchenpflug, 1999; Tredoux et al., 2002; Nestor et al., 2013). The concept that individual faces are represented by their values across different dimensions is central to multidimensional models of face perception (Valentine, 1991; Valentine et al., 2016), and PCA provides one operationalisation of such a space. Within this framework, 'early' dimensions capture most variance within a learning set, and tend to be associated with coarse-scale image variation (for example coding changes in head orientation or whether an image is brighter on one side or the other). Later components, capturing progressively smaller variance, tend to capture finer-scale information. While PCA is applied to a particular set of faces, it is important to note that, given a sufficiently large sample, the resulting space generalises well. So, components derived from one set of faces, tend to capture the variance of novel sets well – particularly if training sets incorporate a range of variation in face images. This means that it is appropriate to ask how information about people's identity is carried in the components derived from PCA – might some dimensions be more useful than others in capturing identity?

Here, we investigated which image dimensions from a PCA of shape and texture could be used to predict recognition performance in humans and a computer model of faces. The relative importance of different dimensions on the recognition of identity was determined by measuring the effect of removing them in a variety of combinations. Our findings show that there is a relatively narrow band of image dimensions that are most important in forming a structural representation that is used for the recognition of faces.

2. Methods

2.1. Overview

We collected data on human face perception using three different tests involving identification: (i) a card sorting task in which participants were asked to group previously unseen faces together by identity; (ii) a matching task in which viewers indicated whether pairs of unfamiliar faces showed the same or different people; (iii) two familiar face recognition tasks in which participants identified famous faces. For each of the face tasks, we carried out PCA on task stimuli along with a database of face images (N > 6000) representing a large range of naturally occurring ('ambient') photographs. We then asked whether the physical representation of task stimuli, within PCA-space, could account for human performance. We sampled over different ranges of components to do this (for example, discounting early dimensions). Finally, for the unfamiliar face tasks (i and ii, above) we asked whether an automated face recognition system, implemented as a Deep Convolutional Neural Network (DCNN) and trained on a very large data set of face images, could predict human performance on the face stimuli used in the behavioural tasks. Once again, we also examined the capacity of PCA, sampled over different ranges, to predict the performance of the DCNN. We now describe the details of these tests and the procedures employed.

2.2. Participants

Opportunity sampling was used to recruit participants who had grown up in the UK. For each experiment we computed an a-priori power analysis ($\alpha = 0.05$, power level $= 0.8$) using G*Power (3.1.9.7, Faul et al., 2007) to determine the minimum sample size required to find an effect (if one was present) for each experiment. We recruited 60 participants (31 female, mean age: 23.3 years) for the card sorting experiment (Pearson's r correlation p H1 = 0.3, based on a moderate correlation, Cohen, 2013). We recruited 70 participants (58 female, mean age: 20.3 years) for the face matching experiment (experiment (p H1 = 0.3 based on a moderate correlation, Cohen, 2013). We recruited 99 participants (61 female, mean age: 25.4) for familiar face recognition Experiment 1 (repeated measures ANOVA-within factors, 4 measurements, $\eta_p^2 = 0.02$ indicating a small expected effect size, Cohen, 2013) and 103 participants (66 female, mean age: 26.4) for familiar face recognition Experiment 2 (repeated measures ANOVA-within factors, 5 measurements, $\eta_p^2 = 0.015$ indicating a small expected effect size, Cohen, 2013). All participants gave their written informed consent. The study was approved by the Psychology Ethics Committee at the University of York.

2.3. Card sorting task

The card sorting task was taken from Jenkins et al. (2011). Twenty images of each of two Dutch celebrities, Chantel Janzen (CH) and Bridget Maasland (BM), were used (40 images in total). These individuals were unfamiliar to our participants. The criteria for image selection were that they showed the face in roughly frontal aspect, exceeded 150 pixels in height and were free from occlusions. Other than these restrictions, the images were free to vary in a way that reflects the variability found in natural viewing. The images were printed in grey scale to a size of 35 × 50 mm and laminated. Participants were given a shuffled stack of the 40 face images and asked to sort the faces into piles according to identity. We then calculated the proportion of participants that sorted each pair of faces into the same pile. This task was self-paced.

2.4. Face matching task

The matching task was taken from the Models Face Matching Test (Dowsett & Burton, 2015). There were 90 trials. In each trial, a pair of face images was presented together. In half of the trials, the faces were from the same identity and in the remaining half of the trials the faces were from a different identity. Participants viewed images that were presented at a distance of approximately 57 cm, such that each image subtended approximately 7.8 × 10.2 degrees of visual angle. The two images were separated laterally by approximately 4.5 degrees. Participants were asked to indicate whether each pair of faces was from the same identity or a different identity. The task was self-paced.

2.5. Familiar face recognition task

The familiar face recognition task comprised two experiments involving naming familiar faces. Participants completed this experiment online using the Pavlovia platform (PSYCHOJS, Version 2020.2). Each trial began with a white fixation cross superimposed on a grey background for 0.5 s. This was followed by a centrally positioned face. Participants pressed one of two buttons to indicate if the face was familiar or unfamiliar. Participants were instructed to respond as quickly and as accurately as possible. If participants indicated that the face was familiar, a new screen would appear containing a response box for participants to type the name or biographical information of the person. When this was complete, a new trial began. If participants indicated that the face was unfamiliar a new trial began immediately. The order in which the faces were presented was randomised for each participant. After each experiment, participants completed a familiarity check to test their ability to recognise the familiar faces, in which a novel high-resolution image from each identity was presented to participants and their task was to name the identity depicted in each image.

The responses for the familiarity check were cross referenced with the responses given for the main experiment. For each participant, the identities that were not recognised in the familiarity check were automatically removed from the main analysis. Participants entered biographical information about the person (instead of their name) 1.5% of the time for Experiment 1 and 2% of the time for Experiment 2. Biographical information was judged to be a match if it was deemed specific enough to the target identity, for example a description of "actor", "musician" or "politician" would result in a non-match (even if these labels were true) but a description of "actor who played Harry Potter" was deemed specific enough for a match. 86.6% of the faces in the first experiment and 92.3% of the faces in the second experiment were recognized during the familiarity check. Accuracy and response time were calculated from these trials.

The familiar face images used for this task depicted A-list celebrities, most of who are well-known Hollywood actors/actresses. Images were collected using a Google Image Search by entering the name of a celebrity and downloading images classified as "large" by the search engine (size of 900 × 900 pixels and above) where the face was broadly front-facing and no part of it being obstructed (e.g. by other parts of the body, clothing or accessories). Apart from these criteria, the images varied naturally across lighting, emotional expressions, hairstyle, facial hair, etc.

In the first familiar face experiment (Exp. 1), we used 24 familiar face images (8 female). Fig. 1 shows the 4 conditions that were created by the selective removal of shape and texture principal components from the images: 0 (no PCs removed), 4 (PCs 1–4 removed), 8 (PCs 1–8 removed), 12 (PCs 1–12 removed). This gave a total of 144 (24 * 4) images. From these images, we created 4 stimulus sets in which there were 6 images from each of the 4 conditions giving a total of 24 images. In each stimulus set, there was only one image from each identity. Participants were allocated randomly to each image set.

In the second familiar face experiment (Exp. 2), we used 20 familiar face images. There were 5 conditions that were created by the selective removal of shape and texture principal components from the images: 0 (no PCs removed), 1–4 (PCs 1–4 removed), 5–8 (PCs 5–8 removed),

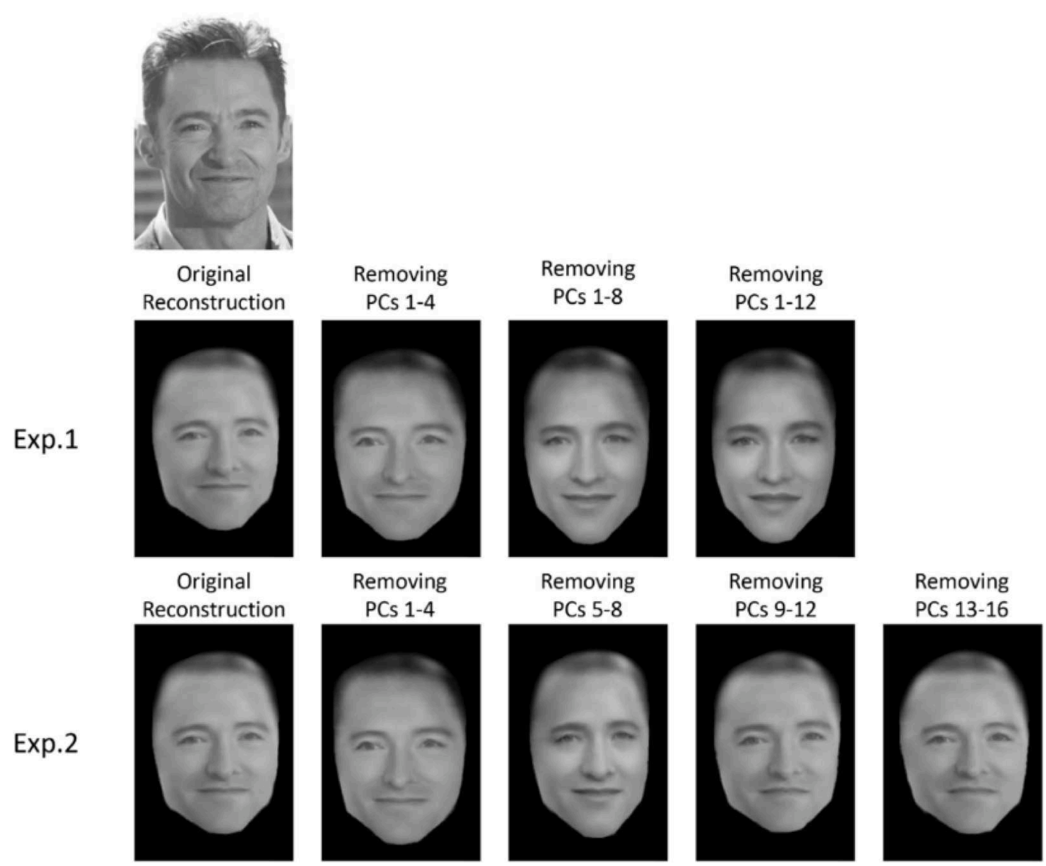

Fig. 1. The effect of removing different bands of principal components from an example familiar face image from the two familiar face recognition experiments.

9–12 (PCs 9–12 removed) and 13–16 (PCs 13–16 removed). This gave rise to at total of 100 (20 * 5) images. From these images, we created 5 stimulus sets in which there were 4 images from each of the 5 conditions giving a total of 20 images. In each stimulus set, there was only one image from each identity. Participants were allocated randomly to each image set.

2.6. Principal components analysis (PCA)

To approximate natural variation across faces and represent our pre-existing experience with faces in daily life, PCA was performed on a large image set containing 6100 images (see the 'background set' described in Mileva et al., 2020). The set contained a varying number of images for each identity (between 1 and 170 images) in order to simulate different levels of familiarity and all images were ambient, capturing variability across age, pose, lighting conditions, emotional expressions, image quality, and ethnicity. Images were rescaled to 380 × 570 pixels. To be consistent across all image sets, we converted all images to greyscale. The shape of each image was determined by aligning 82 fiducial points to each face using the Interface software package (Kramer, Jenkins and Burton, 2016). The x, y coordinates from each image were then entered into the principal components analysis for shape. The texture of each face was generated by warping each image to a standard shape. The intensity values of each pixel within the standard shape were then entered into a principal components analysis of texture. This procedure generated principal components that captured the ways in which images in the set varied, both in terms of shape and texture. We used the first 100 PCs which explained 99.9% of the shape variance and 91.6% of the texture variance.

The images used in the card sorting task (N = 40) and in the face matching task (N = 180) were then projected into this 100-dimensional face space, producing unique a vector with shape and texture reconstruction coefficients for each image that described its location within the face space. These reconstruction coefficients were used in the analyses described below. The images used in the familiar face recognition task (N = 24) were also projected into the same large face space. To determine the similarity between pairs of faces, a correlation was performed on the PC loadings of one image with the PC loadings of another image. These values could then be correlated with corresponding behavioural measures on the card sorting or face matching tasks. In the familiar face recognition task, InterFace software package (Kramer et al., 2016) was used to perform different manipulations to each image in order to neutralise the effect of a small number of shape and texture PCs. This was done by assigning a value of 0 to each shape and texture component within the specified range. Three different manipulations were applied to create images for the first experiment, neutralising the effect of both shape and texture PCs 1–4, 1–8, and 1–12. The second experiment used narrower ranges of PCs (1–4, 5–8, 9–12, and 13–16) to more precisely determine the PCs related to face identity. All other PCs were left intact. Table 1 shows the variance explained by different bands of PCs in the Familiar Face Recognition Task.

Table 1
Variance explained by different bands of principal components in the Familiar Face Recognition Task.

	Shape (%)	Texture (%)
Exp. 1		
PCs 1–4	96.55	63.67
PCs 1–8	98.48	73.05
PCs 1–12	99.15	77.81
Exp. 2		
PCs 1–4	96.55	63.67
PCs 5–8	1.93	9.38
PCs 9–12	0.67	4.76
PCs 13–16	0.31	2.62

2.7. Deep convolutional neural network (DCNN)

The VGG-Face DCNN (Parkhi, Vedaldi & Zisserman, 2015) was used to compare the similarity of face images. This DCNN consists of 13 convolutional layers and 3 fully connected (Fc) layers. The input to the network is an image of size 224 × 224 pixels; images are cropped to a square bounding box centred on the face and rescaled to this resolution. Each convolutional layer is followed by one or more non-linear layers, such as rectified linear units or max pooling. The first two FC layers have 4096 dimensions and the final FC layer has 2622 dimensions. The DCNN was trained on over 2.6 M face images from over 2.6 K identities. Face recognition on the Labeled Faces in Wild dataset (Huang et al., 2007) and YouTube Faces (Wolf et al., 2011) for VGG-Face is 99.9% and 97.4%, respectively.

We tested the ability to decode face identity from the PCA and DCNN outputs using a signal detection approach. Our prediction was higher correlations for the same- than different-identity pairs. We measured decoding sensitivity to identity by calculating the area under the receiver operating characteristic (ROC) curve. We converted this AUC to a value of d' according to the formula $d' = \sqrt{2} \times \Phi^{-1}(\mathrm{AUC})$, where Φ^{-1} is the inverse of the standard normal cumulative distribution function.

2.8. Transparency and openness

These studies were conducted in compliance with the Transparency and Openness Promotion Guidelines. The experiments in this study were not preregistered. All publicly available material has been cited. Anonymised data will be made freely available upon publication.

3. Results

3.1. Card sorting task

To determine the role of shape and texture on the perception of identity, participants performed a card sorting task with the same 40 face images (Jenkins et al., 2011). Participants were simply asked to group the photographs according to identity, so that different photos of the same person are grouped together. Participants were not told how many identities to expect and were free to group the images as they wished. Although the correct number of piles was 2, participants made on average about 8 piles (mean ± SEM: 7.8 ± 0.4; range: 2–15). Most piles contained images from the same identity. However, there were occasional errors in which a pile contained more than one identity (mean ± SEM: 1.0 ± 0.2; range: 0–4).

From the card sort, a probability matrix was generated (Fig. 2A). Each cell in the matrix represents the probability (averaged across participants) that two faces were perceived to have the same identity (i.e. were sorted in the same pile). The higher values in the top left and bottom right of the similarity matrix show that participants typically sorted images into piles from one identity. On average, the probability that two images with the same identity were placed in the same pile was significantly higher than for two images with different identities (within-person: 0.41 + 0.01; between-person: 0.02 + 0.001; $t(758) = 35.0$, $p < 0.0001$).

Next, we asked whether the probability that two images were sorted into the same pile could be predicted by the shape or texture of the images. To do this, we performed a representational similarity analysis comparing the behavioural similarity matrix with a corresponding similarity matrix based on the PCA data of the physical characteristics of the same images. The PCA matrix was generated by correlating (Pearson's r) the PC loadings from one image with the PC loadings from a different image. Thus, there were corresponding values for image similarity and perceptual similarity across all image pairs. These values were Fisher transformed prior to further statistical analysis.

A correlation between the behavioural and PCA data was then

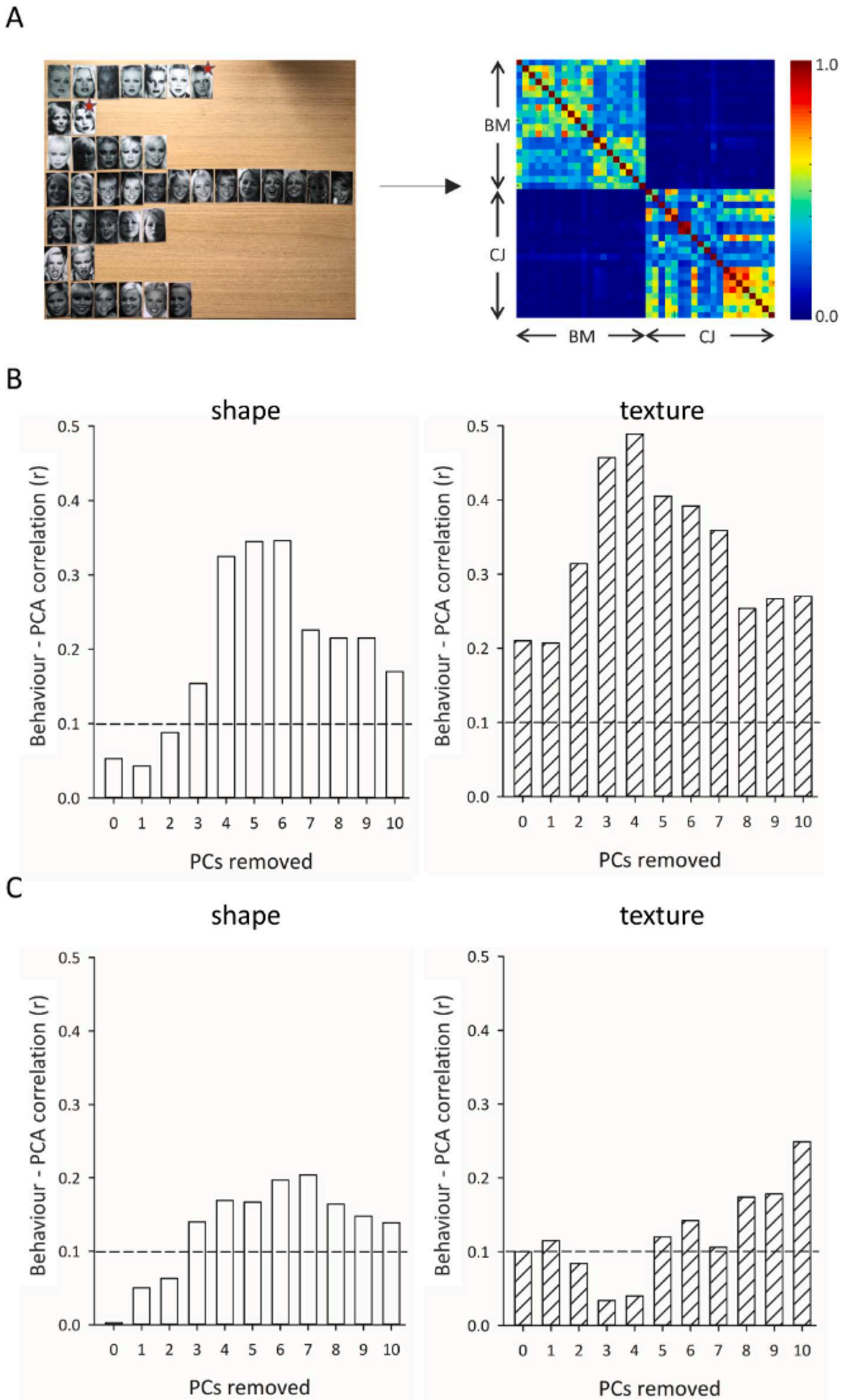

Fig. 2. Results from the card sorting task. (A) An example of a card sort from a participant showing individual piles of faces that the participant determined had the same identity (left). The probability (0.0 – 1.0) that individual pairs of images were sorted into the same pile was calculated across all participants, such that each cell in the similarity matrix (right) represents the probability that two of the images were placed in the same pile. There were 20 images of 2 identities (BM, CJ) giving a perceptual similarity matrix of 40*40. To determine whether the probability that two images were placed in the same pile could be predicted by the shape or texture of the images, a correlation was performed between the perceptual similarity matrix and corresponding shape and texture matrices from the PCA data. This analysis was performed separately for within-person (B) and between-person (C) comparisons. The results show that the correlation between shape or texture and perceptual similarity increased when early principal components (PC) were removed from the analysis. The dashed line shows the critical r values at $p < 0.05$.

performed independently for the within-person (Fig. 2B) and between-person (Fig. 2C) comparisons to avoid inflating comparisons. For the within-person comparisons, the ability of both shape and texture to explain patterns of behavioural responses in the card sorting task increased when the first principal components were removed from the analysis. This reached a maximum after 5 PCs were removed for shape and 4 PCs were removed for texture. Interestingly, if more PCs were removed the correlation began to decrease. For example, the correlation decreased if 7 or more PCs were removed for shape or when 5 or more PCs were removed for texture. A broadly similar pattern was evident for between-person comparisons, although the correlations were generally lower, reflecting the lower incidence (variance) of between-person grouping in this task.

To determine whether shape or texture information from the PCA data alone could be used to discriminate identity, we compared the within-person PCA correlations (an image of BM with a different image of BM or an image of CJ with a different image of CJ) to between-person PCA correlations (an image of BM with an image of CJ) using d'. Fig. 3

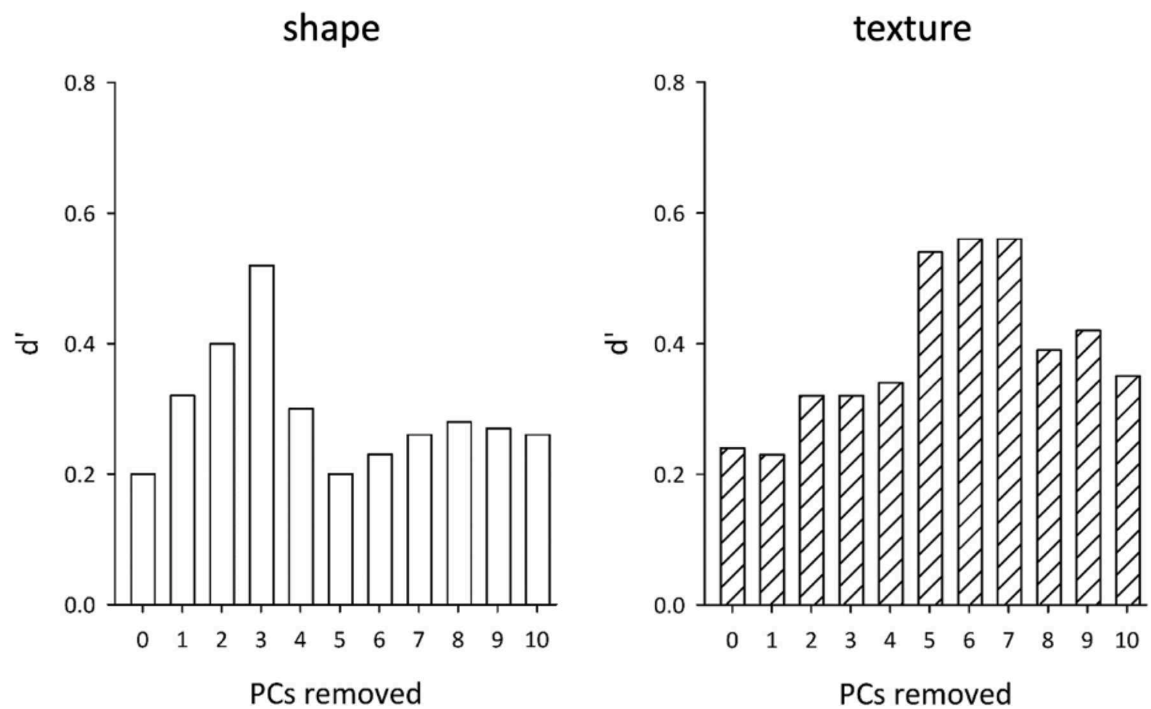

Fig. 3. Sensitivity (d') to facial identity using shape and texture information. A similarity matrix was calculated from the PCA by correlating the PC loadings for each pair of faces. The within-person correlations were then compared with the between-person correlations. Removing PCs increased the difference between within-person and between-person values for both shape and texture.

shows that within-person correlations were higher than between-person correlations for both shape and texture when all PCs were included in the analysis. However, as PCs were removed, the difference between the within-person and between-person values increased for both shape and texture. This difference was maximal for shape after the first 3 PCs were removed, whereas the difference was maximal for texture when the first 6 PCs were removed. However, removing more PCs resulted in a reduction in the within-person and between-person difference.

Next, we asked whether performance on the card sorting task could be predicted by performance of a DCNN (VGG-Face) that has been trained discriminate face identity (Parkhi et al, 2015). The DCNN compared all face pairs from the 40 images in the card sorting task. Fig. 4A shows the similarity matrices from the convolutional (Conv) and fully-connected (Fc) layers. Higher values on the top-left or bottom-right quadrants (within-person) compared to the bottom-left and top-right quadrants (between-person) of the matrices indicate greater discrimination of identity. Fig. 4B shows a statistical analysis (d') of the within-person vs between-person values. This shows that the ability of the DCNN to discriminate between two identities increased from the convolutional layers (1–13) to the fully-connected (14–16) layers.

We then asked whether differences in similarity of faces from the DCNN could be predicted by the perceptual probability that two faces were placed in the same pile in the card sorting task. The correlation between behavioural and DCNN is shown for within-person (Fig. 4C) and between-person (Fig. 4D) comparisons. These results again show there is an increase in the correspondence between the human observers and the DCNN from the convolutional to the fully connected layers. This provides support for the validity of the DCNN as a model of human behaviour, particularly in the later fully connected layers.

Finally, we asked whether the similarity of images in the DCNN could be predicted by variation in the shape or texture across images from the PCA data. Because the highest performance is shown for the fully-connected layers, we focussed on fully-connected layer 7. We used this layer, because the final fully-connected layer (Fc8) is the distribution of activations over trained identities. Fc7, on the other hand, is typically interpreted as a set of visual features that are useful for generic face recognition. The similarity (correlation) between image pairs in shape and texture was compared with the similarity (correlation) in the output of the DCNN for within-person (Fig. 5A) and between-person (Fig. 5B) comparisons. The within-person analysis shows that the correlation between the DCNN and PCA data for both shape and texture increased when the initial PCs were removed from the analysis, but then decreased when more PCs were removed from the analysis. This is similar to the pattern for the behavioural to PCA correlation shown in Fig. 2 following the removal of PCs. A similar pattern was evident for shape in the between-person comparisons, but there was a less clear pattern for texture. Together, these findings suggest a similarity in the way in which human observers and the DCNN represent faces and how performance is related to image properties.

3.2. Face matching task

In this task, participants made same or different identity judgements on pairs of unfamiliar face images from the Models Face Matching Task (Dowsett & Burton, 2015). There were 90 trials in the matching task and overall accuracy was 76.2 ± 8.8 % (mean ± SD). To compare behavioural data on the matching task, we calculated the average proportion of same responses for each image pair across all participants. We then asked whether the tendency of participants to report that an image pair as the same identity was predicted by the similarity in either the shape or texture of the images from the PCA data.

Similarity in shape and texture across images was measured by correlating the PC loadings of one image with the PC loadings of another image. Fig. 6 shows the correlation between the behavioural judgements (proportion same) and the similarity of the images in shape and texture across trials. The ability of shape or texture to predict perceptual responses was not significant when all PCs were included in the analysis. However, significant correlations between behavioural responses and shape or texture became apparent when the initial PCs for shape and texture were removed from the analysis.

Next, we asked if the proportion of same responses could be predicted by the similarity of images from the DCNN. A correlation between

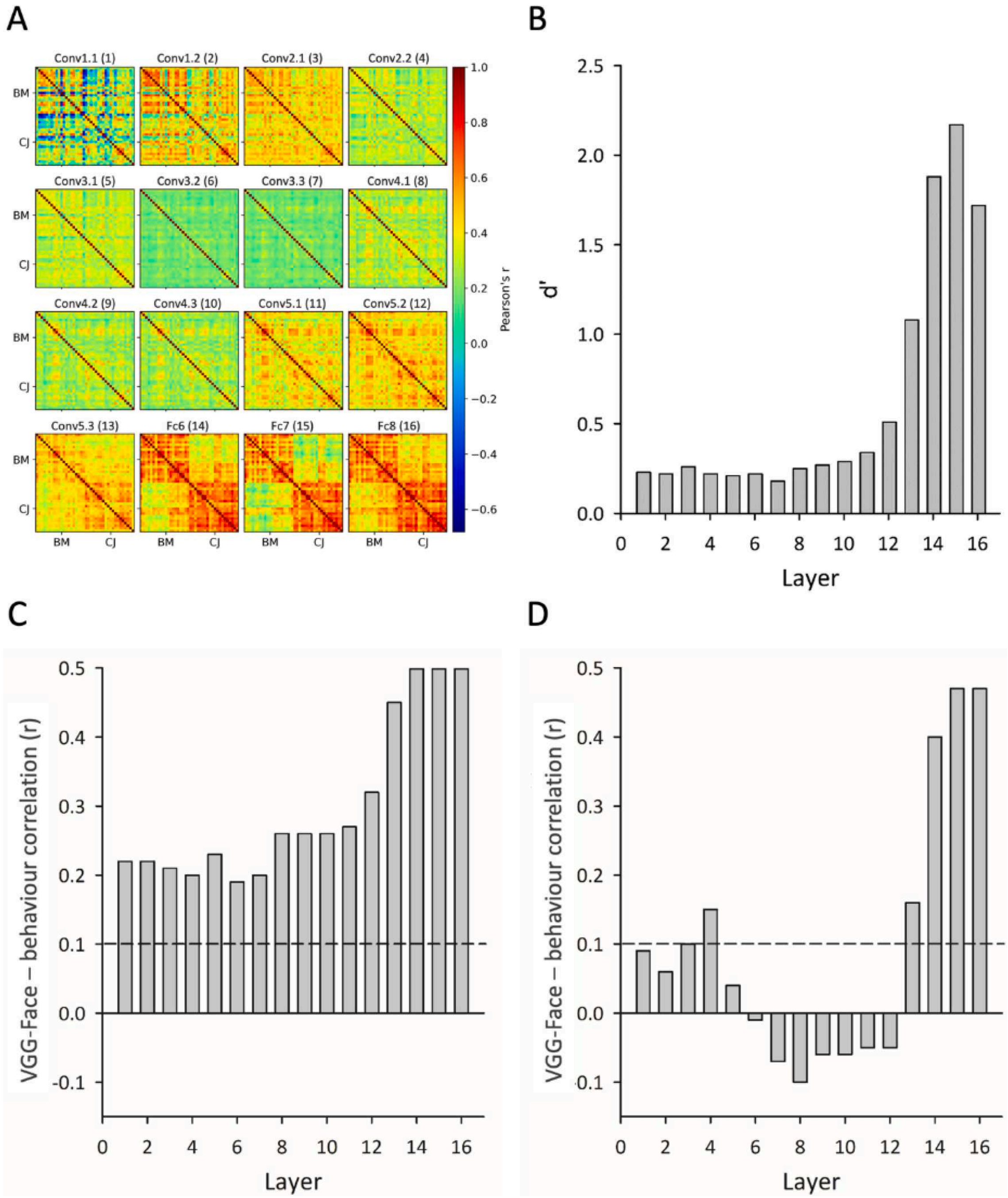

Fig. 4. (A) Similarity matrices from the images in the card sorting task calculated separately in the 16 layers of a DCNN (VGG-Face). (B) A comparison between within-person and between-person values from each layer of the DCNN (convolutional: 1–13; fully-connected 14–16). Higher within-person compared to between-person sensitivity (d') were found in the fully connected layers. Correlations between the DCNN and behavioural responses for (C) within-person and (D) between-person comparisons show that the fully connected layers were most closely linked to perception. Dashed lines show the critical values at $p < 0.05$.

behavioural judgements (proportion same) and similarity in DCNN is shown in Fig. 7A. This shows that the early layers of the DCNN do not predict behavioural judgements on the matching task. However, higher layers, particularly the fully-connected layers (14–16) show a significant correlation with the behavioural judgements of similarity. Consequently, we restricted subsequent analyses to the final fully-connected layer (layer 16 – Fc8) and asked whether the shape and texture of the face images could predict the performance of the DCNN. Fig. 7B shows the correlation between the DCNN and shape or texture was not significant when all PCs were included in the analysis. However, when the initial PCs were removed from the analysis, there was a significant correlation with both shape and texture.

3.3. Familiar face recognition

In the final experiments of this study, we explored how the removal of PCs would affect familiar face recognition. Given that we found similar patterns for removing shape and texture PCs in the unfamiliar face tasks, we decided to remove both shape and texture PCs simultaneously in the familiar face recognition task. In Exp. 1, we measured the recognition of familiar faces in which different numbers of PCs were removed from the image. There were 4 conditions in which either the

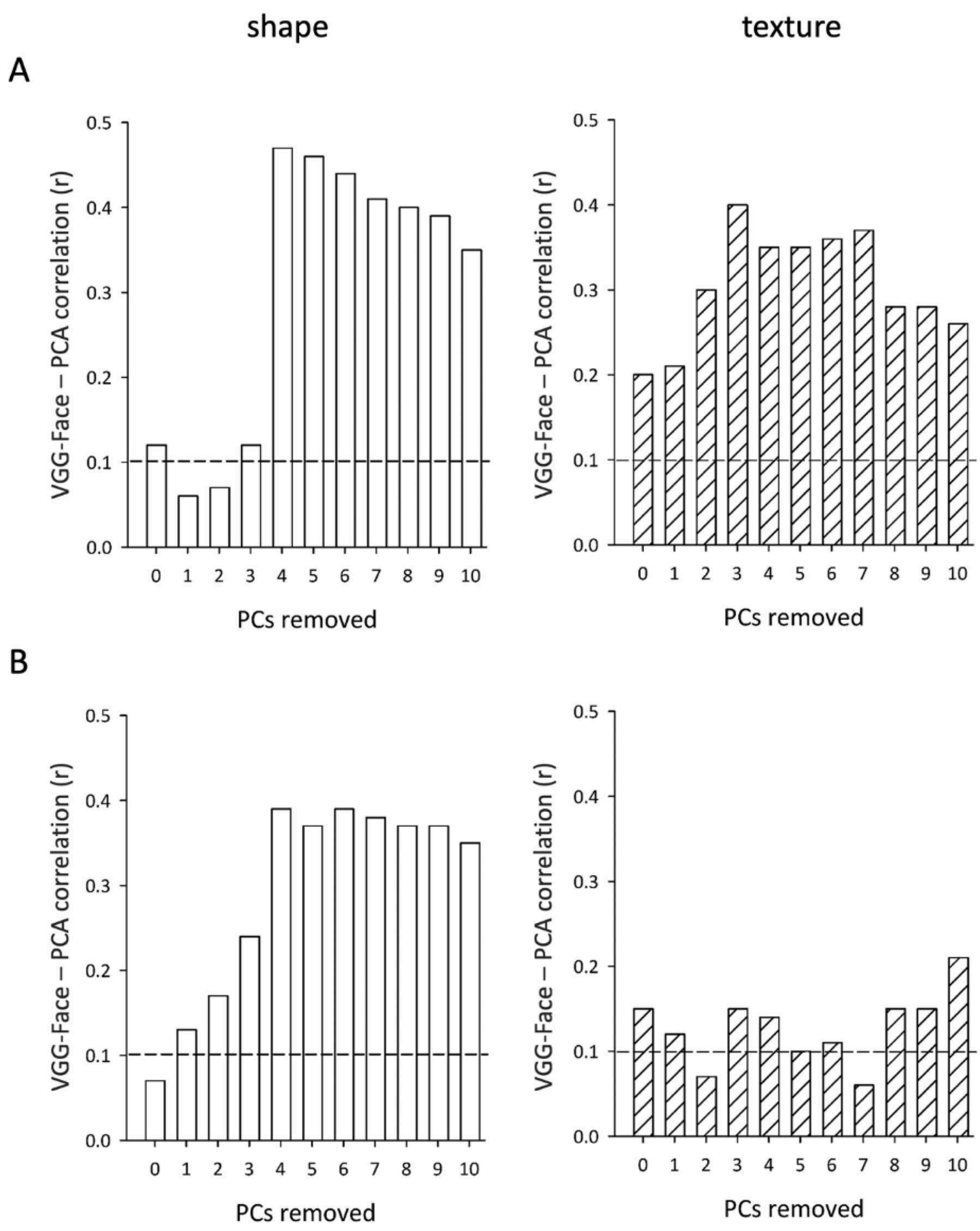

Fig. 5. The role of shape and texture in the representation of faces in VGG-Face. The output from layer Fc7 was used to generate a similarity matrix. A correlation was then performed between the VGG-Face similarity matrix and the shape or texture similarity matrices from the PCA. This analysis was performed separately for within-person (B) and between-person (C) comparisons. The results show that the correlation between shape or texture and DCNN increased when early principal components (PC) were removed from the analysis. The dashed line shows the critical r value at $p < 0.05$.

first 0, 4, 8 or 12 PCs from both shape and texture were removed from the image. Fig. 8A shows the accuracy and response time for each condition. A repeated-measures ANOVA revealed a significant effect of condition for both accuracy ($F(2.44, 239.1) = 225.7$, $p < .001$, $\eta^2_p = .70$) and response time ($F(2.41, 236.6) = 619.2$, $p < .001$, $\eta^2_p = .86$). Planned comparisons showed that there was a significant difference between the 0 and 4 PCs, which was due to an increased recognition accuracy ($t(98) = 2.72$, $p = .008$, $d = 0.27$) and a decreased response time ($t(98) = 5.94$, $p < .001$, $d = 0.60$) for the 4 PCs condition. There was also a significant difference between the 0 and 8 PC conditions, and between the 0 and 12 PC conditions for accuracy and response time. However, these differences were due to a decrease in accuracy (0:8, $t(98) = 11.43$, $p < .001$, $d = 1.15$; 0:12, $t(98) = 21.04$, $p < .001$, $d = 2.11$) and an increase in response time (0:8, $t(98) = 26.44$, $p < .001$, $d = 2.66$; 0:12, $t(98) = 28.47$, $p < .001$, $d = 2.86$). These data show that removal of the initial PCs improves accuracy and reduces response time, whereas removal of later PCs reduces accuracy and increases response time.

In Exp. 2, we investigated the effect of removing bands of PCs.

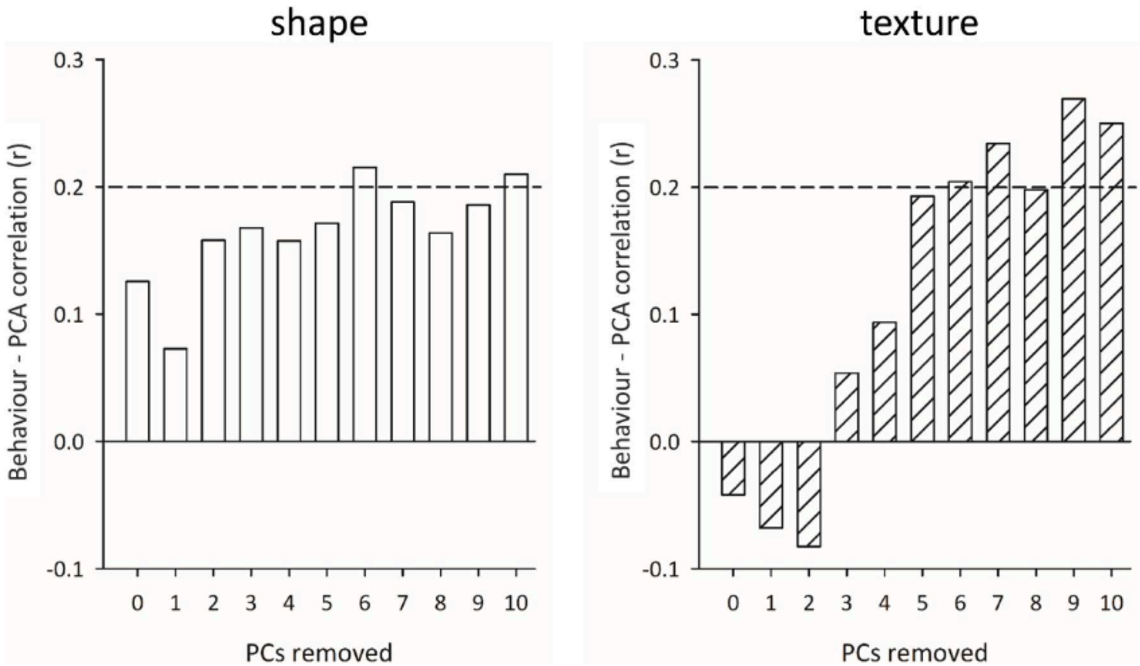

Fig. 6. The role of shape and texture in the prediction of behavioural responses in a matching task. To determine whether the probability that two images were reported as having the same identity could be predicted by the shape or texture of the images, a correlation was performed between the proportion of same responses across the 90 trials and their similarity in shape or texture. The results show that the correlation between behavioural judgements and shape or texture increased when early principal components (PC) were removed from the analysis. Dashed lines show the critical r value at $p < 0.05$.

Participants viewed images in which 0, 1–4, 5–8, 9–12 or 13–16 PCs of shape and texture were removed from the image. Fig. 8B shows the accuracy and response time for each condition. A repeated-measures ANOVA revealed a significant effect of condition for both accuracy (F (3.22, 327.9) = 125.9, p <.001, η^2_p =.55) and response time (F(3.56, 362.6) = 422.9, p <.001, η^2_p =.81). Planned comparisons showed that there was a significant difference between the 0 and 1–4 PC conditions, which was again due to an increased recognition (t(102) = -1.840, p =.034, d = 0.18) and a decreased response time (t(102) = 2.40, p =.009, d = 0.24) for the 1–4 PC condition. There was also a significant difference between the 0 and 5–8 PCs conditions, and also between the 0 and 9–12 PCs conditions for accuracy and response time. These differences were due to a decrease in accuracy (0:5–8, t(102) = 13.08, p <.001, d = 1.29); 0:9–12, t(102) = 12.88, p <.001, d = 1.27) and an increase in response time (0:5–8, t(102) = 23.26, p <.001, d = 2.30; 0:9–12, t(102) = 26.51, p <.001, d = 2.61)). Finally, there was no significant difference in accuracy between the 0 and 13–16 PCs removed conditions (t(102) = -0.21, p =.418, d = 0.02). However, there was a slight increase in response time (t(102) = -4.04, p <.001, d = 0.40). Nevertheless, the accuracy (t(102) = -13.27, p <.001, d = 1.31) and response time (t(102) = 25.00, p <.001, d = 2.46) were significantly different between the 9–12 and 13–16 conditions. These data show that removal of the initial PCs improves accuracy and reduces response time, whereas the selective removal of intermediate bands of PCs reduces accuracy and increases response time. Finally, removal of later bands of PCs has a minimal effect on recognition.

4. Discussion

The aim of this study was to determine what information is necessary for recognition of identity from faces. To address this issue, a principal components analysis was used to reveal the underlying dimensions of shape and texture in naturally varying face images from different identities. This allowed us to compare the importance of these image dimensions on a range of tasks involving judgements of identity. Our key finding is that the perception and recognition of identity from faces is critically dependent on a narrow band of statistical variation, expressible in terms of a simple linear decomposition of image sets.

As faces have a similar structure, the ability to discriminate identity must be based on encoding subtle differences between images. A further challenge for successful face recognition is that, as a result of changes in viewing conditions, each face can generate an almost infinite number of images. So, it is necessary for the recognition system to differentiate between information in the image that provides cues about identity from other image variation that does not. Models of face processing address this issue by proposing that information about faces is first represented in an image-based or pictorial code, which is then transformed into a structural code that can be used for recognition (Bruce & Young, 1986, 2012; Burton et al., 1990). However, the precise image properties that are used in this structural code have not been fully resolved. The face-space model provides a framework for explaining how variance across faces might be represented in a structural code that is used for recognition (Valentine, 1991; Valentine et al., 2016). In this model, different properties of the face are represented in a multidimensional face space. Each face is represented by a location in this multidimensional space, such that faces that are close together are perceived to be more similar and those that are separated by larger distances are perceived to be more different. Nevertheless, it has not been clear how many dimensions are important or what they might represent.

PCA allows an objective, data-driven approach to understand image variation across faces (Turk & Pentland, 1991; O'Toole et al., 1993; Calder et al., 2001; Jozwik et al., 2022). This technique can be used to reveal a number of principal components that account for variance in face images. In this study, we asked whether these principal components from the PCA correspond to the dimensions that are used for a structural code that could underpin the recognition of faces. To do this, we compared the similarity of the values across different dimensions for pairs of faces with the same identity or with different identities. We then progressively removed principal components or dimensions from the analysis to determine the effect on judgements of identity in unfamiliar faces. Removing the initial image dimensions of shape and texture increased the correlation between image similarity and patterns of

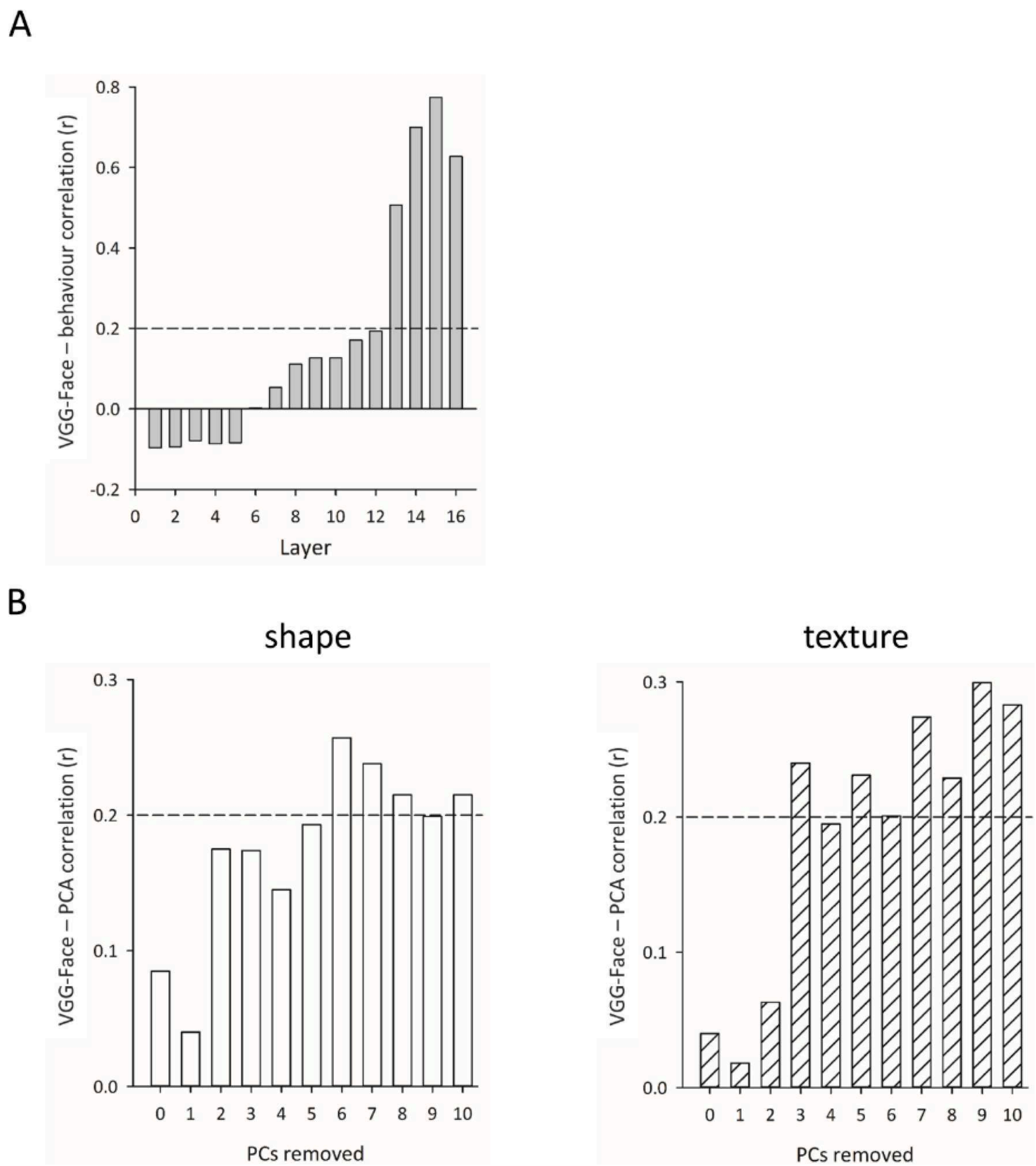

Fig. 7. (A) Similarity of each pair of the 90 images in the matching calculated from different layers of the DCNN. (B) Correlation between similarity values from each layer of the DCNN and behavioural judgements (proportion same). Higher correlations were found in the fully connected layers. (C) Correlation between similarity of images from the final fully-connected layer (Fc8) of the DCNN and the similarity in shape and texture. The results show that the correlation between the DCNN and shape or texture increased when early principal components (PC) were removed from the analysis. Dashed lines show the critical r value at $p < 0.05$.

perceptual performance on sorting (Jenkins et al., 2011) and matching (Dowsett & Burton, 2015) tasks with unfamiliar faces. However, removing more shape and texture dimensions led to a reduction in the difference between same and different identity images and reduced the correlation between perceptual performance and image similarity. Our findings that the distance in PC space predicts perceptual performance when early PCs are removed from the analysis fundamentally extends previous work showing a link between the perception of face images and their position in PCA space (Turk & Pentland, 1991; O'Toole et al., 1993; Hancock, Burton & Bruce, 1996; Calder et al., 2001; Burton et al., 2016; Jozwik et al., 2022).

The importance of texture in the perception of identity is consistent with previous studies showing that the manipulation of texture information in the face through contrast reversal has a significant effect on recognition (Bruce & Langton, 1994; Russell et al., 2006; Harris et al. 2014). Moreover, the recognition of faces also becomes much more difficult when texture is removed from the image (Davies et al., 1978; Leder, 1999; Burton et al., 2005). Finally, the recognition of hybrid faces in which the texture from one identity is combined with the shape from another familiar identity (Burton et al., 2005, 2015) is strongly biased toward the texture of the image (Andrews et al., 2016; Rogers et al., 2022). However, the increased correspondence between the perception of identity and texture when the initial image dimensions were removed suggests that not all texture information contributes to recognition. Presumably, these early image dimensions reflect ambient changes in the texture (e.g., illumination) that are not diagnostic of an identity. For practical reasons we used greyscale images in this study. Although faces can be reliably recognized from greyscale images, it would be interesting

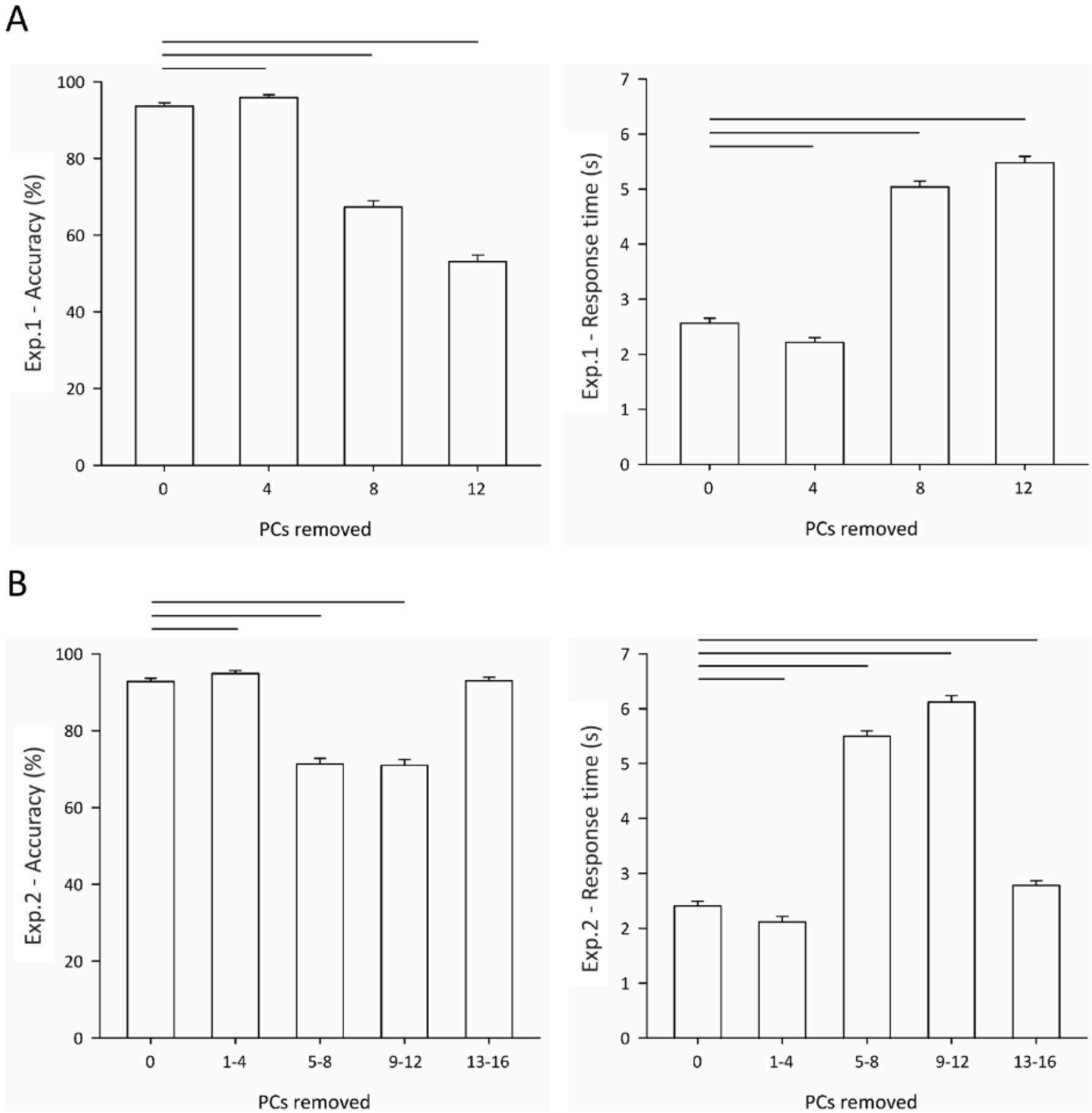

Fig. 8. Familiar face recognition experiment. (A) In Exp.1, removal of 4 PCs resulted in increased accuracy and a reduction in response time to familiar faces. However, removal of 8 or 12 PCs resulted in decreased accuracy and increased response time. (B) In Exp.2, removal of the PCs 1–4 again resulted in increased accuracy and a reduction in response time to familiar faces. Removal of 5–8 or 9–12 PCs resulted in a significant decrease in accuracy and an increased response time. However, removal of PCs 13–16 had no effect on accuracy and a limited effect on response time. Horizontal lines indicate significant differences ($p < 0.05$) relative to the 0 PCs condition (original reconstruction). Error bars indicate standard error of the mean.

to see how colour also contributes to the texture information that is important for face recognition in future studies.

The shape or configuration of the face has also been suggested to be important for face recognition and is often referred to as configural processing (McKone & Yovel, 2009; Tanaka & Gordon, 2011; Piepers & Robbins, 2012; Rogers et al., 2022). However, a challenge for configural accounts of face recognition is that they do not specify which aspects of shape are important (Burton et al., 2015). Natural variation in face images caused by rigid changes in viewpoint or non-rigid changes (such as in expression or during speech) can often lead to large changes in the configuration or shape of the face. So, it has not been clear whether shape could be a useful cue for face recognition (Burton et al., 2015). Our analysis shows that when the early PCs are removed from the analysis, the shape or configuration of the face does contain information that allows the discrimination of identity. That is, face images with the same identity had a more similar shape than images from different identities. Similarity in shape also predicted whether images were perceived to have the same identity. This shows that PCA could provide a useful way of operationalising how the shape or configuration of the face can be used for the perception of identity.

We next asked whether a computer model of face recognition using a deep convolutional neural network (DCNN) was sensitive to shape and texture information in a similar way to human observers. DCNNs have made significant progress in solving the complex problem of recognizing faces across naturally occurring changes in the image (O'Toole et al., 2018). Indeed, the development of DCNNs has been influenced by the organisation of the primate visual system (Kriegeskorte, 2015; Yamins and DiCarlo, 2016). However, it remains unclear whether these provide useful models of the human visual system (Cichy & Kaiser, 2019). In this study, we used a DCNN trained to discriminate faces (VGG-Face; Parkhi, Vedaldi & Zisserman, 2015) and found that performance was correlated with perceptual judgements of identity, particularly in the later convolutional and fully connected layers. We also found that the shape and texture of the face images predicted performance on the DCNN in a similar way to human perception. These findings concur with a recent study that showed that DCNNs trained on faces use shape and texture information in a similar way to humans (Daube et al., 2021). Our results extend these findings showing that the correspondence between performance on the DCNN and variance in shape and texture increases when the early image dimensions were removed from the analysis. Together, these findings provide evidence in support of DCNNs being useful models of human face perception (O'Toole et al., 2018; Abudarham et al., 2019).

Finally, we explored which dimensions of the image are important for familiar faces. To tackle this question, we removed PCs or image dimensions from familiar face images and measured the effect on recognition. In the first experiment, we found that removing the early PCs of shape and texture increased recognition and decreased response time. In contrast, removal of later PCs reduced recognition and increased response time. In the second experiment, we asked whether there are select bands of image dimensions that are important for recognition. Again, we found that removing the early PCs improved recognition compared to when no PCs were removed. Conversely, we found that removing intermediate bands of PCs resulted in a significant decrease in the recognition accuracy and increased response time. However, removing later PCs had a minimal effect on recognition compared to no PCs being removed. These findings suggest that the initial PCs reflect ambient image information that is not used for recognition. However, there is an intermediate band of PCs that plays a key role in recognition. These could reflect the multidimensional structural code that is used for recognition.

A key feature of our study was the use of ambient face images that reflect the image variation that occurs in natural viewing. Although the early image dimensions for both shape and texture explain most of the image variance, they do not appear to contain information that is important for the recognition of identity. In contrast, intermediate image components which represent more subtle changes in the shape and texture of the image appear to be important for recognition. These findings are relevant to the debate surrounding the difference between unfamiliar and familiar face perception (Young & Burton, 2018a, 2018b, 2021; Rossion, 2018; Sunday & Gauthier, 2018; Blauch, Behrmann & Plaut, 2021a; 2021b; Yovel & Abudarham, 2021). Our proposal is that an important aspect of the change from a pictorial representation that is used for unfamiliar face perception to a structural representation that is used for familiar face recognition involves the removal of ambient information in the image. The ability to recognise familiar faces would appear to depend on the ability to ignore this irrelevant information and focus on the image properties that are important for recognition. On the other hand, the difficulty in the recognition of unfamiliar faces may reflect the inability to ignore this information. It is interesting to speculate whether the ability to ignore irrelevant information may explain individual differences on tasks of face recognition (White & Burton, 2022).

Finally, it is important to note that there are many other signals available in face images, in addition to identity - for example, the sex, age, expression or pose of a face. The degree to which these different sources of information are processed independently by human viewers remains a topic of debate (Young, 2018; Duchaine & Yovel, 2015; Connolly et al, 2019). Our study opens the interesting possibility that the information space derived from statistical techniques, like PCA, contains discrete bands capturing information about different aspects of the face. In previous work, it has been shown that early components can capture not only superficial image noise, but also gross characteristics of faces that are independent of identity, such as expression (Calder & Young, 2005), gender (O'Toole et al 1993), race (O'Toole et al 1994), age (Burt & Perrett, 1995) and pose (Burton et al, 2016). It is now possible to interrogate the information space derived from face sets in order to establish whether there are specific ranges carrying more subtle personal information, such as expressions or facial speech. Focussing on the statistical information space, rather than the frequency analysis of individual images, opens up a potentially useful route for future research in face perception more generally, beyond recognition of identity. It is interesting to note that this possibility arises from a comparatively simple linear analysis such as PCA. Of course, many more complex, non-linear, decompositions of statistical face-space are possible, but nevertheless, the approach described here offers a (perhaps surprising) degree of interpretability.

In conclusion, our results suggest that an intermediate band of image dimensions contains the structural code that is used to discriminate identity. We show that variation across these image dimensions predicts performance of humans and computer models of face recognition. Recent studies in face recognition have shown that the discrimination of identity from a PCA is improved by the addition of a classifier (Kramer et al., 2017; Kramer et al., 2018). These results suggest that these classifiers may improve recognition by increasing the weight of these critical band of PCs or image dimensions. These findings provide a new perspective for understanding of the structural code that underpins the recognition of faces.

CRediT authorship contribution statement

Timothy J. Andrews: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing – original draft, Visualization, Supervision, Supervision. **Daniel Rogers:** Methodology, Investigation, Formal analysis. **Mila Mileva:** Software, Methodology, Formal analysis, Resources. **David M. Watson:** Software, Formal analysis, Investigation. **Ao Wang:** Investigation, Formal analysis. **A. Mike Burton:** Conceptualization, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

We would like to thank Helena von Werthern, Lucy Hammond, and Jessica Barrett for their help on data collection for the sorting task. We would also like to thank Rob Jenkins, Andy Young and members of the York face group for helpful comments during the course of this study.

References

Abudarham, N., Shkiller, L., & Yovel, G. (2019). Critical features for face recognition. *Cognition, 182*, 73–83.

Andrews, T. J., Baseler, H. A., Jenkins, R., Burton, A. M., & Young, A. W. (2016). Contributions of feature shapes and surface cues to the recognition and neural representation of facial identity. *Cortex, 80*, 280–291.

Blauch, N. M., Behrmann, M., & Plaut, D. C. (2021a). Computational insights into human perceptual expertise for familiar and unfamiliar face recognition. *Cognition, 208*, Article 104341.

Blauch, N. M., Behrmann, M., & Plaut, D. C. (2021b). Deep learning of shared perceptual representations for familiar and unfamiliar faces: Reply to commentaries. *Cognition, 208*, Article 104484.

Bruce, V. (1982). Changing faces: Visual and non-visual coding processes in face recognition. *British journal of psychology, 73*(1), 105–116.

Bruce, V., & Langton, S. (1994). The use of pigmentation and shading information in recognising the sex and identities of faces. *Perception, 23*, 803–822.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British journal of psychology, 77*(3), 305–327.

Bruce, V., & Young, A. (2012). *Face perception.* Hove, East Sussex: Psychology Press.

Burt, D. M., & Perrett, D. I. (1995). Perception of age in adult Caucasian male faces: Computer graphic manipulation of shape and colour information. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 259*(1355), 137–143.

Burton, A. M., Bruce, V., & Johnston, R. A. (1990). Understanding face recognition with an interactive activation model. *British Journal of Psychology, 81*(3), 361–380.

Burton, A. M. (2013). Why has research in face recognition progressed so slowly? The importance of variability. *Quarterly Journal of Experimental Psychology, 66*, 1467–1485.

Burton, A. M., Jenkins, R., Hancock, P. J., & White, D. (2005). Robust representations for face recognition: The power of averages. *Cognitive Psychology, 51*, 256–284.

Burton, A. M., Kramer, R. S. S., Ritchie, K. L., & Jenkins. (2015). Identity from variation: Representations of faces derived from multiple instances. *Cognitive Science, 40*, 202–223.

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.

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

Caharel, S., Jiang, F., Blanz, V., & Rossion, B. (2009). Recognizing an individual face: 3D shape contributes earlier than 2D surface reflectance information. *Neuroimage, 47*, 1809–1818.

Calder, A. J., Burton, A. M., Miller, P., Young, A. W., & Akamatsu, S. (2001). A principal component analysis of facial expressions. *Vision research, 41*(9), 1179–1208.

Calder, A. J., & Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Reviews Neuroscience, 6*(8), 641–651.

Cichy, R. M., & Kaiser, D. (2019). Deep neural networks as scientific models. *Trends Cogn Sci, 23*, 305–317.

Connolly, H. L., Young, A. W., & Lewis, G. J. (2019). Recognition of facial expression and identity in part reflects a common ability, independent of general intelligence and visual short-term memory. *Cognition and Emotion, 33*(6), 1119–1128.

Daube, C., Xu, T., Zhan, J., Webb, A., Ince, R. A., Garrod, O. G., & Schyns, P. G. (2021). Grounding deep neural network predictions of human categorization behavior in understandable functional features: The case of face identity. *Patterns, 2*(10), Article 100348.

Davies, G., Ellis, H. D., & Shepherd, J. (1978). Face recognition accuracy as a function of mode of representation. *Journal of Applied Psychology, 63*, 180–187.

Dowsett, A. J., & Burton, A. M. (2015). Unfamiliar face matching: Pairs out-perform individuals and provide a route to training. *British journal of psychology, 106*(3), 433–445.

Duchaine, B., & Yovel, G. (2015). A revised neural framework for face processing. *Annual review of vision science, 1*, 393–416.

Sunday, M. A., & Gauthier, I. (2018). Face expertise for unfamiliar faces: A commentary on Young and Burton's "Are we face experts?". *Journal of Expertise, 1*(1), 35–41.

Gong, S., McKenna, S. J., & Psarrou, A. (2000). *Dynamic vision: From images to face recognition*. London, UK: Imperial College Press.

Hancock, P. J., Burton, A. M., & Bruce, V. (1996). Face processing: Human perception and principal components analysis. *Memory & cognition, 24*(1), 26–40.

Hancock, P. J., Bruce, V., & Burton, A. M. (2000). Recognition of unfamiliar faces. *Trends in cognitive sciences, 4*(9), 330–337.

Harris, R. J., Young, A. W., & Andrews, T. J. (2014). Brain regions involved in processing facial identity and expression are differentially selective for surface and edge information. *NeuroImage, 97*, 217–223.

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

Itz, M. L., Schweinberger, S. R., & Kaufmann, J. M. (2016). Effects of caricaturing in shape or color on familiarity decisions for familiar and unfamiliar faces. *PLoS ONE, 11*(2), e0149796.

Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition, 121*(3), 313–323.

Jiang, F., Blanz, V., & Rossion, B. (2011). Holistic processing of shape cues in face identification: Evidence from face inversion, composite faces, and acquired prosopagnosia. *Visual Cognition, 19*, 1003–1034.

Jozwik, K. M., O'Keeffe, J., Storrs, K. R., Guo, W., Golan, T., & Kriegeskorte, N. (2022). Face dissimilarity judgments are predicted by representational distance in morphable and image-computable models. *Proceedings of the National Academy of Sciences, 119*(27). e2115047119.

Kemp, R., Towell, N., & Pike, G. (1997). When seeing should not be believing: Photographs, credit cards and fraud. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 11*(3), 211–222.

Kramer, R. S. S., Jenkins, R., & Burton, A. M. (2016). InterFace: A software package for face image warping, averaging, and principal components analysis. *Behavior Research Methods, 6*, 2002–2011.

Kramer, R. S. S., Young, A. W., & Burton, A. M. (2018). Understanding face familiarity. *Cognition, 172*, 46–58.

Kramer, R. S. S., Young, A. W., Day, M., & Burton, A. M. (2017). Robust social categorization emerges from learning the identities of very few faces. *Psychological Review, 124*(2), 115–129.

Kriegeskorte, N. (2015). Deep neural networks: A new framework for modeling biological vision and brain information processing. *Annu. Rev. Vis. Sci., 1*, 417–446.

Lai, M., Oruç, I., & Barton, J. J. (2013). The role of skin texture and facial shape in representations of age and identity. *Cortex, 49*(1), 252–265.

Leder, H. (1999). Matching person identity from facial line drawings. *Perception, 28*, 1171–1175.

Longmore, C. A., Liu, C. H., & Young, A. W. (2008). Learning faces from photographs. *Journal of Experimental Psychology: Human Perception and Performance, 34*(1), 77.

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences, 6*, 255–260.

McKone, E., & Yovel, G. (2009). Why does picture-plane inversion sometimes dissociate perception of features and spacing in faces, and sometimes not? Toward a new theory of holistic processing. *Psychonomic Bulletin & Review, 16*, 778–797.

Mileva, M., Young, A. W., Jenkins, R., & Burton, A. M. (2020). Facial identity across the lifespan. *Cognitive psychology, 116*, Article 101260.

Moon, H., & Phillips, P. J. (2001). Computational and performance aspects of PCA-based face-recognition algorithms. *Perception, 30*, 303–321.

Nestor, A., Plaut, D. C., & Behrmann, M. (2013). Face-space architectures: Evidence for the use of independent color-based features. *Psychological Science, 24*, 1294–1300.

O'Toole, A. J., Abdi, H., Deffenbacher, K. A., & Valentin, D. (1993). Low-dimensional representation of faces in higher dimensions of the face space. *JOSA A, 10*(3), 405–411.

O'Toole, A. J., Deffenbacher, K. A., Valentin, D., & Abdi, H. (1994). Structural aspects of face recognition and the other-race effect. *Memory & Cognition, 22*(2), 208–224.

O'Toole, A. J., Vetter, T., & Blanz, V. (1999). Three-dimensional shape and two-dimensional surface reflectance contributions to face recognition: An application of three-dimensional morphing. *Vision Research, 39*, 3145–3155.

O'Toole, A. J., Castillo, C. D., Parde, C., Hill, M. Q., & Chellappa, R. (2018). Face space representations in deep convolutional neural networks. *Trends in Cognitive Science, 22*(9), 794–809.

Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). Deep face recognition.

Piepers, D. W., & Robbins, R. A. (2012). A review and clarification of the terms "holistic", "configural", and "relational" in the face perception literature. *Frontiers in Psychology, 3*. Article 559.

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*, Article 108013.

Rossion, B. (2018). Humans are visual experts at unfamiliar face recognition. *Trends in cognitive sciences, 22*(6), 471–472.

Russell, R., Sinha, P., Biederman, I., & Nederhouser, M. (2006). Is pigmentation important forface recognition? Evidence from contrast negation. *Perception, 35*, 749–759.

Russell, R., & Sinha, P. (2007). Real-world face recognition: The importance of surface reflectance properties. *Perception, 36*(9), 1368–1374.

Russell, R., Biederman, I., Nederhouser, M., & Sinha, P. (2007). The utility of surface reflectance for the recognition of upright and inverted faces. *Vision Research, 47*, 157–165.

Scheuchenpflug, R. (1999). Predicting face similarity judgements with a computational model of face space. *Acta Psychologica, 100*, 229–242.

Tanaka, J. W., & Gordon, I. (2011). Features, configuration and holistic face processing. In A. J. Calder, G. Rhodes, M. H. Johnson, & J. V. Haxby (Eds.), *The Oxford handbook of face perception* (pp. 15–30). Oxford, United Kingdom: Oxford University Press.

Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal of cognitive neuroscience, 3*(1), 71–86.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology Section A, 43*(2), 161–204.

Valentine, T., Lewis, M. B., & Hills, P. J. (2016). Face-space: A unifying concept in face recognition research. *The Quarterly Journal of Experimental Psychology, 69*(10), 1996–2019.

Yamins, D. L. K., & DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nat. Neurosci., 19*, 356–365.

Young, A. W. (2018). Faces, people and the brain: The 45th Sir Frederic Bartlett Lecture. *Quarterly Journal of Experimental Psychology, 71*(3), 569–594.

Young, A. W., & Burton, A. M. (2018a). Are we face experts? *Trends in cognitive sciences, 22*(2), 100–110.

Young, A. W., & Burton, A. M. (2018b). What we see in unfamiliar faces: A response to Rossion. *Trends in cognitive sciences*, 472–473.

Young, A. W., & Burton, A. M. (2021). Insights from computational models of face recognition: A reply to Blauch, Behrmann and Plaut. *Cognition, 208*, Article 104422.

Yovel, G., & Abudarham, N. (2021). From concepts to percepts in human and machine face recognition: A reply to Blauch. *Behrmann & Plaut. Cognition, 208*, Article 104424.

White, D., & Burton, A. M. (2022). Individual differences and the multidimensional nature of face perception. *Nature Reviews Psychology, 1*(5), 287–300.

Wolf, L., Hassner, T., & Maoz, I. (2011, June). Face recognition in unconstrained videos with matched background similarity. In *CVPR 2011* (pp. 529-534). IEEE.