Cognition 182 (2019) 73–83

Contents lists available at ScienceDirect

Cognition

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

Original Articles

Critical features for face recognition

Naphtali Abudarham[a], Lior Shkiller[a], Galit Yovel[a,b,*]

[a] School of Psychological Sciences, Tel Aviv University, Tel Aviv, Israel
[b] Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel

ARTICLE INFO

Keywords:
Face recognition
Familiar faces
Face space
Deep neural network
Feature space

ABSTRACT

Face recognition is a computationally challenging task that humans perform effortlessly. Nonetheless, this remarkable ability is better for familiar faces than unfamiliar faces. To account for humans' superior ability to recognize familiar faces, current theories suggest that different features are used for the representation of familiar and unfamiliar faces. In the current study, we applied a reverse engineering approach to reveal which facial features are critical for familiar face recognition. In contrast to current views, we discovered that the same subset of features that are used for matching unfamiliar faces, are also used for matching as well as recognition of familiar faces. We further show that these features are also used by a deep neural network face recognition algorithm. We therefore propose a new framework that assumes similar perceptual representation for all faces and integrates cognition and perception to account for humans' superior recognition of familiar faces.

1. Introduction

Face recognition is a computationally challenging task that requires fine discrimination between similarly looking images of different identities, as well as generalization across different images of the same individual. Although humans are considered experts in face recognition, studies have shown that our face recognition abilities are superior to faces we are familiar with, whereas our ability to match unfamiliar faces is error-prone (Young & Burton, 2017b, 2017a). These findings led to the suggestion that familiar face recognition depends on a different set of facial features, based on the extensive experience that we have with them than those used for unfamiliar faces. For example, it has been suggested that familiar face recognition is primarily based on internal facial features, whereas unfamiliar face matching is primarily based on external facial features (Ellis, Shepherd, & Davies, 1979; Kramer, Towler, Reynolds, & Burton, 2017; O'Donnell & Bruce, 2001; Young, Hay, McWeeny, Flude, & Ellis, 1985). According to another view, the representation of familiar faces is based on the average of their different appearances, which excludes superficial image-based information that may dominate the representation of unfamiliar faces (Jenkins & Burton, 2011). This view further posits that throughout our experience with variable images of familiar faces, we learn the idiosyncratic features that remain invariant across their different appearances and are unique for each identity. This view therefore suggests that a different set of features is used to recognize different familiar identities (Burton, Kramer, Ritchie, & Jenkins, 2016).

In a recent study, we used a novel *reverse engineering* approach to reveal which facial features are critical for face identity. We found a subset of features for which humans have high perceptual sensitivity to detect differences between different identities (high-PS features) (Abudarham & Yovel, 2016) (see Fig. S2). We then showed that systematically changing high-PS features changes the identity of faces, whereas changing features for which humans have low perceptual sensitivity (low-PS features) did not change the identity of faces (see Fig. 1). Importantly, these high-PS features remain invariant across different head views (Abudarham & Yovel, 2016), making them useful not only for discrimination between identities but also for generalizing across different appearances of the same identity.

Nevertheless, this subset of features was shown to be critical for unfamiliar faces and may not generalize to familiar faces, with which we have much greater experience. Thus, the goal of the current study was to use the same reverse engineering approach to reveal which features are critical for familiar face recognition. This allowed us to test the common view that different facial features are used for the identification of familiar and unfamiliar faces.

To that end, in Experiment 1 we first examined the role of high-PS vs. low-PS features in a familiar face matching task, using the same matching task that was used for unfamiliar faces in our previous study (Abudarham & Yovel, 2016). An important difference between familiar and unfamiliar faces is that familiar faces are represented in memory. Features that are used for matching two faces presented simultaneously, may not be used for matching a familiar face to its representation in

* Corresponding author at: School of Psychological Sciences & Sagol School of Neuroscience, Tel Aviv University, Tel Aviv 69978, Israel.
E-mail address: gality@post.tau.ac.il (G. Yovel).

https://doi.org/10.1016/j.cognition.2018.09.002
Received 2 April 2018; Received in revised form 3 September 2018; Accepted 3 September 2018

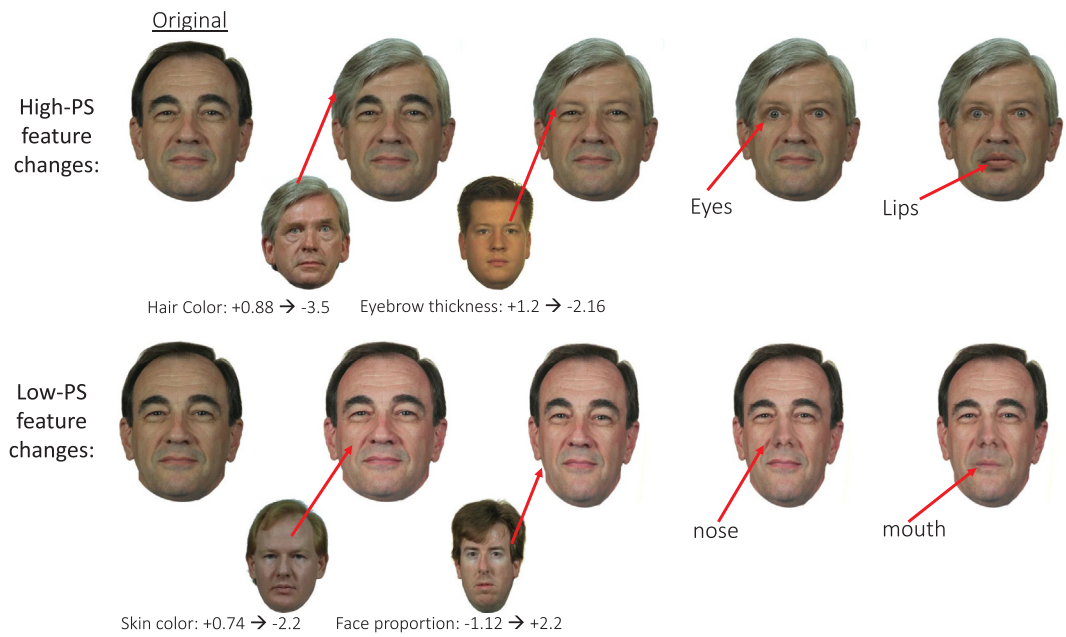

Fig. 1. To reveal which features are critical for the identity of the face we replaced 5 features for which we have high perceptual sensitivity (high-PS features) (top) or 5 features for which we have low perceptual sensitivity (low-PS features) (bottom) (Abudarham & Yovel, 2016). Features were taken from faces of different identities (see methods for more information about the feature changing procedure). Similarity matching between the original face and the changed face on the far right showed that changing high-PS features changed the identity of the face whereas changes in low-PS features did not change the identity of the face (see Fig. 3 for similarity matching results).

memory. Therefore, in Experiments 2 and 3 we examined whether these features are also used for face recognition. Finally, the features that we found correspond to semantic descriptions of facial features (e.g., eyes, mouth), and may therefore overlook visual information that cannot be described by these labels. We therefore examined whether these features are also used by a face recognition algorithm, that is not bound to these semantic meanings. Recently, Deep Neural Network (DNN) algorithms have reached human level performance on unconstrained ("wild") facial images, in which faces appear in various poses, expressions and illuminations (Schroff & Philbin, 2015). These advances are the result of the capability of DNNs to extract the invariant information through supervised learning with many different images of the same identity (O'Toole, Castillo, Parde, Hill, & Chellappa, 2018). We therefore hypothesized that a DNN may be tuned to the same invariant, high-PS features that humans use for face recognition (Experiment 4).

2. Experiment 1 – Critical features for matching familiar faces

To determine whether changing high-PS features, but not low-PS features, changes the identity of a familiar face, we used a matching task similar to the one we used in a previous study with unfamiliar faces (Abudarham & Yovel, 2016). Familiar faces were modified by either changing five high-PS features or five low-PS features (Fig. 2, Fig. S3). We presented participants with pairs of celebrity faces, before and after feature changes, and asked them to rate whether the two pictures belong to the same person or to different people (Fig. 3A, top). Pairs of same identity and different identity faces were also presented to obtain baseline performance to which matching abilities for low-PS and high-PS pairs can be compared.

2.1. Method

2.1.1. Participants

All participants were Amazon-Mechanical-Turk workers, participating in the experiment for payment (approximately 1$ per 15 min of work). A total of 38 participants (American residence, 18 females, 28 Caucasians, 6 East-Asians, 2 African-American, 1 Hispanic/Latin and 1 Middle-Eastern, ages 23–66 (mean 39.4, standard deviation (SD) 13.6) performed the experiment.

2.1.2. Stimuli

Ten American celebrities – all adult Caucasian males – were selected for the experiment. For each identity, we downloaded from the internet two frontal neutral expression images, with no glasses, hat or facial hair, and with adequate lighting and quality. All pictures were cropped from the background, and cut below the chin, leaving just the face, including the hair and ears. One of these images was selected as a "base" picture, a picture that was later modified, and the other designated as a "reference" picture, which was left unchanged. Additional 100 frontal pictures of Caucasian male faces, with no glasses or facial hair, were taken from the Color FERET database (Phillips, Wechsler, Huang, & Rauss, 1998) and cropped in the same way.

2.1.3. Face tagging: converting faces into feature vectors, and measuring face-space distances

In our previous study we described faces as feature-vectors embedded in a multidimensional feature space. We showed that by perceptually assigning values to a set of 20 features, we can measure distances between faces, and these distances were correlated with perceptual face similarity scores. In this study, we repeated this procedure with familiar faces and converted each one of the faces in our database into a feature-vector representation (see Fig. S1 for an example of feature-vectors of two celebrity faces). For the 100 faces from the color-FERET dataset we used the feature-vector representations obtained in our previous study (Abudarham & Yovel, 2016). For tagging the ten celebrity faces we ran a face-tagging procedure. To provide participants with a large enough dataset for tagging, allowing them to judge facial features with respect to a variance of feature sizes and shapes, we created a dataset of 60 face images. These 60 images included the selected 10 celebrity faces, 20 pictures of other celebrities of similar characteristics as the original 10, and 30 randomly selected pictures from the 100 color-FERET dataset. In the tagging procedure, participants were asked to rank each of the 20 features for each of the sixty faces on a scale of −5 and +5 (for example: how thick are the

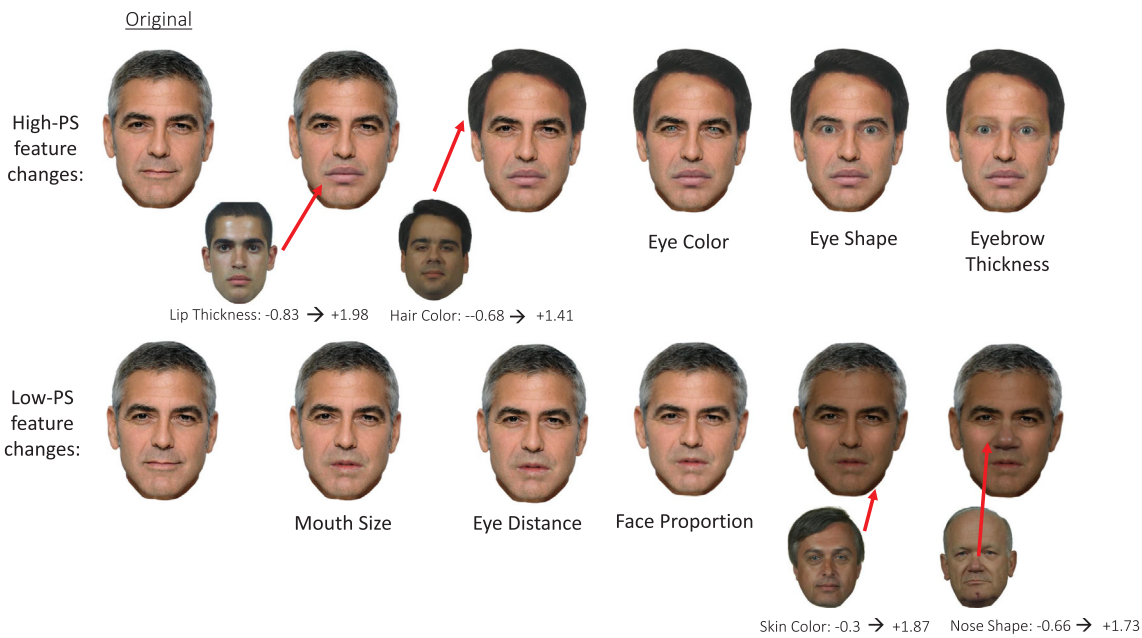

Fig. 2. An example of the effect of changing 5 high-PS features and 5 low-PS features on a familiar face (George Clooney) (see methods for more information about the feature changing procedure).

lips? how large are the eyes?). We then normalized these values to get a relative rating score (Fig. S1). A total of sixty-eight participants participated in the tagging experiment, each participant rated 7 features (to avoid fatigue), such that an average of 10 participants tagged each feature in each face.

2.1.4. Facial feature substitution

Based on the feature vectors that we obtained for each face, we were able to select features from other faces, to replace the original features (see Fig. 2). The features were taken from faces in our original database of 100 tagged faces used in the previous study (Abudarham & Yovel, 2016). Feature replacement was done by copying and pasting a feature with an opposite and as far away value as possible. For example, to replace thin lips, we took one of the three thickest lips available in the database (we did not always take the top thickest, to avoid repeating the same feature in too many faces). This face modification process was performed by a professional graphic designer using Adobe Photoshop ©. We created two sets of changed pictures: in one set we changed high-PS features, namely the lip-thickness, the hair, eye-color, eye-shape and eyebrow-thickness. In the second set, we changed low-PS features, namely mouth-size, eye-distance, face-proportion, skin-color and nose. (See Fig. 2 for an example of George Clooney) (see Fig. S3 for 5 high and low-PS feature changes for the 10 celebrity faces). Following feature substitution, we ran again the face tagging procedure, with different participants, and computed feature vectors for the changed faces. This allowed us to calculate the face-space distances (the sum of absolute differences between feature values, or L1-norm) between faces before and after change.

2.1.5. Face matching task

Following the feature substitution, we had 4 images for each celebrity face: two original images ("base" and "reference"), one image with high-PS feature changes, and one with low-PS feature changes. Of the two original images of each identity, the "base" image and the "reference" image, only the "base" image of each identity was modified. The "reference" image was left unchanged and was paired with either the same identity "base" image, the high-PS, low-PS or the different identity image (see Fig. 3A, top). This ensured that during matching, the low-level image information was different between the pair of matched images, making sure that matching was not based on low-level image-based similarity.

Thus, for each celebrity we created four pairs: *Same pair* – "reference" and "base" images of the same identity, *Different* pair– "reference" image and the "reference" image of another celebrity, *High-PS* pair– the "reference" image and the image with high-PS changes, and *Low-PS* pair – the "reference" image and the image with low-PS changes. We used these 40 pairs to create two versions of a matching task, each version with 30 face pairs: Version 1 – the 10 Same pairs, the 10 Different pairs, 5 High-PS pairs and 5 Low-PS pairs, and Version 2 – the 10 Same pairs, the 10 Different pairs, the 5 High-PS pairs not included in Version 1, and the 5 Low-PS pairs not included in Version 1. Each subject completed one of two versions of the experiment, in which half of the faces were changed in their high-PS features, and the other half in their low-PS features.

To determine perceptual similarity between face pairs, the two face images were presented simultaneously, side by side on the computer screen, and participants were asked to judge whether they belonged to the same person or to different people. The participants' response was on a scale of 1 to 6: '1' indicating "definitely the same person"; '2' indicating "same person"; '3' indicating "possibly the same person"; '4' indicating "possibly different people"; '5' indicating "different people" and '6' indicating "definitely different people". The two pictures were presented on the screen until response, after which the next two faces were presented. The order of the pairs, as well as the right-left positions of the images within each pair were randomized across participants.

2.2. Results & discussion

Fig. 3A shows a histogram of the dissimilarity ratings for the four types of pairs (Fig. 3A, middle row) and an average dissimilarity rating score for each of the four conditions (Fig. 3A bottom row). A repeated measures ANOVA on the averaged rating revealed a significant difference between the 4 conditions ($F(3,111) = 257.75$, $p \ll 0.001$, $\eta^2 = 0.87$). Post-hoc comparisons of the four different conditions with p-value corrected for six comparisons ($p = .008$) revealed a significant difference between all four conditions ($t(37) > 8.43$, $p < .0001$), except the difference between high-PS and different faces ($t(37) = 2.2$, $p = .03$). These findings indicate that changing high-PS features

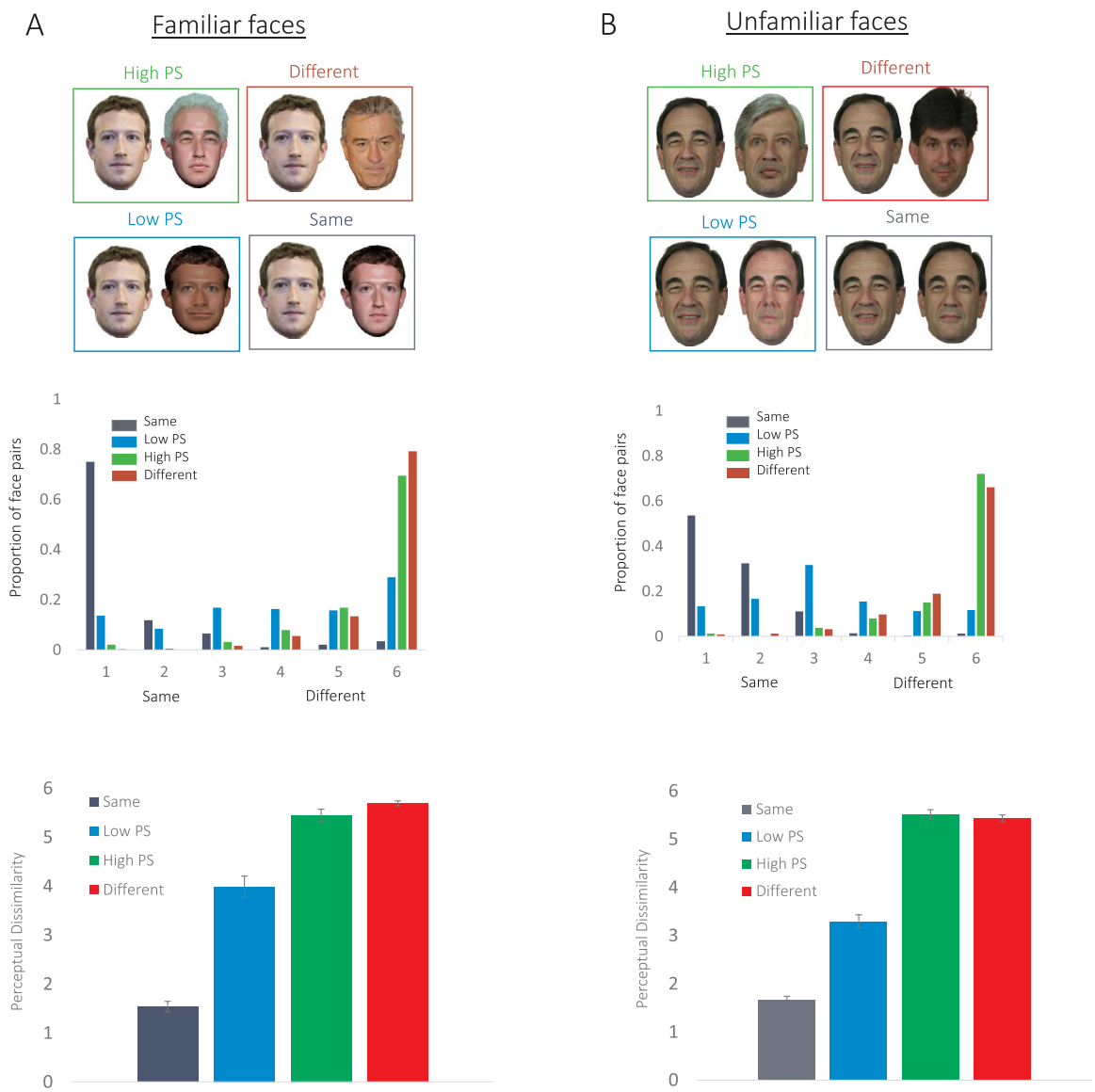

Fig. 3. Results of the face matching task (Experiment 1). A: Familiar faces; B: Unfamiliar faces. Top: The four types of face pairs presented in the experiment. Middle: histogram of the distribution of same-different identity rating from 1 – definitely same identity to 6 – definitely different identity. Bottom: The average similarity rating scores for each pair type shows that matching the original and a changed high-PS face yielded scores that are similar to different pairs.

changed the identity of the face. Low-PS changes were not perceived as the same person but also did not make faces look like different people. Thus, low-PS features are noticeable changes but do not change the identity of a face. These results are similar to results we obtained in our previous study with unfamiliar faces (see Fig. 3B based on data taken from Abudarham & Yovel, 2016).

We also compared the feature vector distance (L1-norm: the sum of absolute differences between 20 feature values) for the original faces and faces that underwent high and low-PS changes, based on the feature tagging data. Overall, the distances between the original and changed faces were not significantly different for High-PS changes (M = 24.05, STD = 1.4) and Low-PS changes (M = 22.05, STD = 4.3) ($t(9) = 1.37$, $p = 0.2$). Yet, as reported above, the perceptual distance (i.e. how different the changed faces were perceived) was larger for High-PS changes. These results indicate that although the face-space distance based on all 20 features was the same for both types of changes, different features contributed differently to the perceptual distance between faces. In other words, the direction of change in space is more critical than its distance.

Finally, we computed image-based differences by calculating the pixel-based Euclidean distance between the original face and its high and low-PS changed images. This analysis revealed slightly larger distance between the original and low-PS changes (M = 1817.17, STD = 653.62) than the original and high-PS changes (M = 1298.49, STD = 702.41), $t(9) = 2.35p = .04$. Thus, low-level imaged based differences cannot account for the perceptual effects. Additional analysis based on Earth-Movers-Distance (EMD) revealed no difference between high and low-PS changes (for details on EMD analysis see results of Experiment 2).

In summary, our results suggest that the same features that were critical for unfamiliar face matching are also critical for familiar face matching. These findings are therefore inconsistent with the suggestion that different features are used to determine the identity of familiar and unfamiliar faces (Ellis et al., 1979; Kramer et al., 2017; O'Donnell & Bruce, 2001; Young et al., 1985). Instead, our data show that the same subset of features is used to determine the identity of all faces regardless of familiarity or their specific identity.

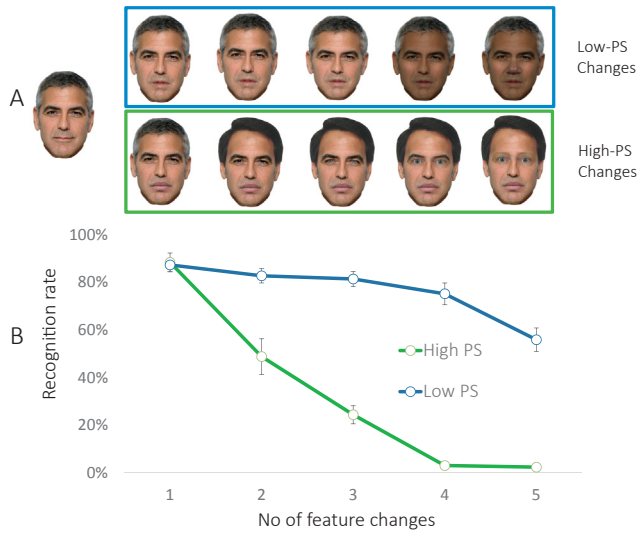

Fig. 4. Results of the face recognition task (Experiment 2). A: The order and type of feature changes that were presented during the face recognition task. Participants were first presented with the 5 changes of all faces until they were presented with the original faces. B: Results show proportion of recognition as a function of the number of changes. Changing 3 or more high PS features resulted in misidentification of the celebrity faces. Error bars indicate the standard error of each condition.

3. Experiment 2 – Critical features for familiar face recognition

In Experiment 1 we found that high-PS features are critical for familiar face *matching*. However, matching two images presented simultaneously may not necessarily depend on the same features used to match an incoming face to its representation in memory. Furthermore, in real life we hardly ever need to match different images of unfamiliar or familiar faces, but often need to recognize familiar faces by matching them to their representation in memory. In Experiment 2, we therefore examined whether high-PS, but not low-PS features are also critical for familiar face recognition. In addition, we asked how many feature changes are needed to change the identity of a face. To that end, we presented faces in a face recognition task, and measured recognition performance as a function of the number of high-PS or low-PS feature changes (see Fig. 4A for an example of the face of George Clooney).

3.1. Methods

3.1.1. Participants

Participants were 40 Amazon-Mechanical-Turk workers (American residents, 22 females, 34 Caucasians, 3 Hispanic/Latins, 1 African American, 1 South-Asian and 1 Native American, Ages 21–51, mean age 32.7, STD = 7.9), participating in the experiment for payment (approximately 1$ per 15 min of work). They were randomly assigned to two groups: 20 participants were presented with faces that were modified by changing high-PS features, and the other 20 participants were presented with faces that were modified by changing low-PS features (see Fig. 4A). Two out of the 20 participants that were presented with high-PS changes were excluded because their data were corrupted.

3.1.2. Stimuli

We used the same original pictures, of ten celebrities, as in Experiment 1. For each identity, we generated 10 different modified faces in which we gradually changed either 5 high-PS features or 5 low-PS features. The order of changes corresponded to the descending order of the 5 top PS scores of features (see Fig. S2). For high-PS changes, faces were gradually changed in the following order: lip-thickness, hair, eye-color, eye-shape, and eyebrow-thickness. For low-PS changes we had the following order: mouth-size, eye-distance, face-proportion, skin-color and nose, matching the ascending order of the low-PS features. Thus, a total of 10 identities with 10 modifications for each (5 High-PS and 5 Low-PS feature changes) were used in the experiment. Fig. 4A shows the low and high-PS modifications for one identity.

3.1.3. Procedure

Participants were presented with faces that were changed in high-PS or in low-PS features. The images were presented one at a time, in the following order: first the 10 faces with 5 feature changes (the order within these 10 faces was randomized across participants), then the 10 faces with 4 feature changes, and so on until finally the 10 original celebrity images were presented. Thus, each participant was presented with a total of 60 faces. For each image, participants were asked to write the name of the person in the image. If they did not recall the name they were asked to describe him as best as they could. If they could not recognize the face they marked "I cannot recognize this face". After they typed the response they pressed a key, which initiated the presentation of the next face. Faces were presented on the screen until participants initiated the presentation of the next face

3.1.4. Data analysis

All responses were analyzed manually to determine whether the participant correctly identified the face in the image. For example, responses such as "Mark Zuckerberg", "Zuckerberg" or "The Facebook guy" were all accepted as correct recognition of the picture of Mark Zuckerberg. For each participant, each reply was scored "1" for correct recognition, and "0" for an incorrect recognition or an "I cannot recognize this face" response. Some subjects could not recognize the original, unchanged celebrity faces, so we removed the data of these unrecognized faces from the final analysis. The faces that were not recognized in their original form, and were excluded from the analysis, were also not recognized when one or more feature changes were made and therefore yielded an overall 0% recognition rate. Recognition rates were then calculated for each celebrity face, and for each number of feature changes, as the average correct recognition across participants (see Fig. 4B). The original faces were not included in the analysis because after removing the identities that were not recognized, they all had the same score of 100% recognition for all faces and participants.

3.2. Results & discussion

Fig. 4B shows recognition rates for high and low-PS changes as a function of the number of changes. A mixed two-way ANOVA with PS type (high/low) as a between participants factor, and number of changes (1–5) as a within subject factor, revealed a significant effect of PS type indicating better recognition for low-PS than high-PS changed faces ($F(1,36) = 106.2$, $p < .00001$, $\eta^2 = .75$). A significant effect of number of features indicates that recognition decreases with the increase in the number of feature changes ($F(4,144) = 115.68$, $p < .00001$, $\eta_p^2 = .76$). An interaction between the two variables indicates that recognition dropped much more steeply for high-PS than for low-PS features ($F(4,144) = 3.5$, $p < .00001$, $\eta_p = .52$). Independent sample t-tests corrected for 6 multiple comparisons ($p = .008$) revealed a significant difference between high and low-PS changes for 5, 4, 3 and 2 changes ($t(36) > 33$, $p < .0001$, $d > 1.4$) ($t(36) > 33$, $p < .001$, $d > 1.4$) but not for 1 change ($t(36) < .04$, $p > .96$). Thus, by changing only 2–3 high-PS changes we caused a significant drop in recognition of familiar faces.

The results of this experiment clearly show that high-PS features are critical for face recognition. Changing more than three high-PS features completely modified the identity of a familiar face such that - less than 5% of the participants were able to recognize it. In contrast, about 50% of the celebrity faces in which five low-PS features were replaced could be recognized. These results indicate that high-PS features are not only used for matching unfamiliar or familiar faces presented simultaneously

but are also central to the representation of the identity of a familiar face in memory.

Whereas our findings show that changing 5 high-PS features changed the identity of the face, the relative contribution of each high-PS feature may be different for different faces. For example, Fig. S4 shows that for some celebrity faces (e.g. Justin Bieber) changing the hair had a larger effect than others (e.g., Matt Damon). Nevertheless, changing all 5 high-PS features reduced recognition to less than13% (%5 on average across all 10 faces) for all 10 celebrity faces.

To verify that the effect found on face recognition was not merely the result of changes at the image level, we used the Earth Mover's Distance (EMD) metric as a measure of image-based changes (Rubner, Tomasi, & Guibas, 2000), to compare the modified images to the original images. Fig. S5 shows the mean EMD between modified images and original images, as the number of changes increases, and recognition level indicated by proportion of errors. Results indicate that the pattern of results of EMD scores was different from the pattern of results of recognition scores. The delta in EMD as we change features, is a function of the area of the image that was changed. We see that for high-PS changes, the large increase in EMD is caused by changing the hair, which indeed takes up a large area in the image. Changing eye-color or eye-shape contribute very little to EMD, but have a significant effect on recognition. As for low-PS changes, changing skin color results in a large EMD change, but is not reflected in recognition scores. Finally, the EMD for final high-PS and low-PS changes are equal, but recognition scores are much worse for five high-PS than five low-PS changes. In fact, low-PS changes involved changing *all* the pixels in the face (excluding the hair), but generated a very small effect on recognition, while in high-PS changes large areas of the face remain untouched (the cheeks, chin, jaw, forehead, nose), but recognition dropped to zero. These results indicate that face processing is not affected by low-level image changes, but is a result of processing high-level features, in particular features for which we have high perceptual sensitivity and are invariant across different appearances of the same identity (Abudarham & Yovel, 2016).

In the next set of experiments, we examined the relative contribution of different high-PS features by changing the order in which they were modified.

4. Experiment 3: The relative contribution of different High-PS features for face recognition

To assess the relative contribution of individual feature changes, we altered the order in which features were changed. Since there are 120 different orders of 5 feature changes, we had to select a few of interest. We started with changes in the reverse order to the one that we used in Experiment 2.

4.1. Experiment 3A

4.1.1. Methods

4.1.1.1. Participants. Twenty Amazon-Mechanical-Turk workers (American residents, 10 females, 12 Caucasians, 3 African-Americans, 3 Hispanic/Latins, 1 East-Asian and 1 Native American, ages 25–53, mean age 33.75, STD = 7.7), participated in the experiment for payment (approximately 1$ per 15 min of work).

4.1.1.2. Stimuli. In the previous experiment the order of feature changes started with features with the highest PS score (lip thickness, hair, eye-shape, eye-color, eyebrow-thickness) (see Fig. S2). Here we reversed the order of feature changes (i.e. the order was: eyebrow-thickness, eye-shape, eye-color, hair and lip-thickness).

4.1.2. Results & discussion

Fig. 5 shows recognition rates for the new order of the high-PS changes relative to the low and high-PS changes used in Experiment 2.

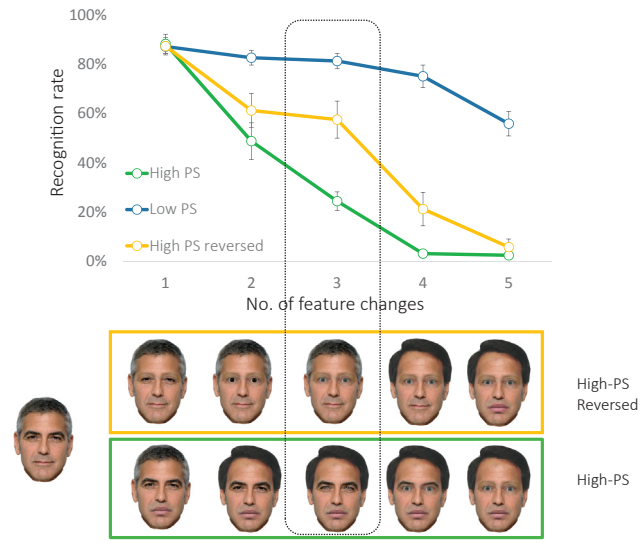

Fig. 5. Results of the face recognition task (Experiment 3A) in which the order of high-PS changes was reversed (orange line) shows overall similar findings across the two orders of changes, with an important role for the hair in face recognition. Error bars indicate the standard error of each condition.

A mixed ANOVA with Type (high-PS original order, high-PS reversed order) as a between subject factor, and number of changes as a within subject factor, revealed a main effect of no. of changes ($F(4, 144) = 139.75$, $p < 0001$, $\eta^2 = .77$), a marginally significant effect of type of change ($F(1,36) = 4.13$, $p = .05$, $\eta_p^2 = .10$) and an interaction of no. of changes and order of changes ($F(4, 144) = 5.37$, $p < .001$, $\eta_p^2 = .13$). The interaction reflects the difference in recognition for 3 feature changes in which the hair changed in the original order but not in the reversed order. As a result, performance level was higher when the original hair was present than when it was changed ($t(36) = 3.44$, $p < .001$).

Results therefore suggest that hair has important contribution for face recognition. We therefore conducted another experiment in which we changed the hair first or last and examined recognition rate for the same faces.

4.2. Experiment 3B

4.2.1. Method

4.2.1.1. Participants. Forty Amazon-Mechanical-Turk workers (American residents, 20 females, 31 Caucasians, 4 East-Asians, 2 Hispanic/Latins, 1 South-Asian and 1 African-American, ages 18–56, mean age 35, STD = 9.1) participated in the experiment for payment (approximately 1$ per 15 min of work). 20 of them were randomly allocated to the hair-first version of the experiment and the rest to the hair-last version. The data of one participant in the hair-first version was corrupted, and was excluded from analysis.

4.2.1.2. Stimuli. We used the same stimuli but changed the features by either changing the hair first or last (see Fig. 6). The order of the other feature changes was similar to the order used in Experiment 2.

4.2.2. Results & discussion

A mixed ANOVA with Type (Hair First, Hair Last) as a between subject factor, and number of changes as a within subject factor, revealed a main effect of no. of changes ($F(4,148) = 115.68$, $p < 0001$, $\eta_p^2 = .76$), a main effect of type of feature ($F(1,36) = 8.93$, $p < .001$, $\eta_p^2 = .74$). and an interaction of no. of changes and order of changes ($F(4,148) = 39.50$, $p < .001$, $\eta_p^2 = .52$). Independent sample t-tests corrected for 6 multiple comparisons ($p = .008$) revealed better performance for hair-last than hair-first changes for 2 and 4 changes (t

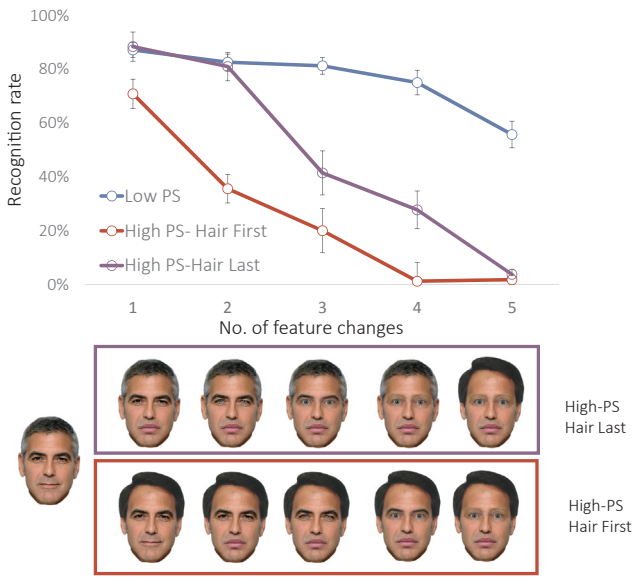

Fig. 6. Results of the face recognition (Experiment 3B) in which the hair was changed first or last shows that changing the hair generates a larger drop in face recognition even for celebrity faces. Error bars indicate the standard error of each condition.

(37) > 3.84p < .0001) and a marginally significant difference for 1 and 3 changes (t(37) > 2.6, p < .01). There was no significant difference between five changes, which were based on identical images in the two versions of the experiment (p = .35).

These findings show a dramatic effect of hair on familiar face recognition. When the hair is changed last, recognition level for high-PS features drops below low-PS features changes after 3–4 feature changes. Whereas when the hair is changed first, recognition level drops below low-PS changes already at 2 feature changes.

Hair is often cropped in face recognition experiments that focus on the role of the internal facial features for face recognition. Here we show that face recognition is highly influenced by the hair even in celebrities that tend to modify their hair styles. These findings suggest that despite the fact that the hair is a relatively variable facial feature, it is an integral and critical part of the representation of the identity of a person.

5. Experiment 4 – Critical features for a deep neural network face recognition algorithm

Our results so far reveal a subset of features that are used to determine the identity of familiar and unfamiliar faces in matching and recognition tasks. However, these features correspond to semantic descriptions of facial features (eyes, mouth). It was therefore important to assess whether these features are also used by a face recognition algorithm that is not bound to these semantic meanings. To this end we tested the similarity scores for the feature vectors of the original and changed pictures, obtained using a face-recognition DNN. As mentioned above, DNNs are trained on an unconstrained set of face images that vary greatly within an individual and have recently reached human face recognition level (O'Toole et al., 2018).

5.1. Methods

5.1.1. Stimuli

We used the 10 celebrity face images that were used in Experiment 2 and 15 unfamiliar faces that were manipulated in the same way (Abudarham & Yovel, 2016). All images were pre-processed using the DLIB image processing library: the faces were aligned to match eyes and nose locations across all faces, and then were cropped to a size of 96 by 96 pixels (cropping out the hair).

5.1.2. Face recognition algorithm

We used the OpenFace DNN (https://cmusatyalab.github.io/openface/) for face recognition, and the pretrained model nn4.small2.v1.t7 available for download from the *OpenFace* site. OpenFace is trained with ~500,000 images from labeled face recognition datasets, CASIA-WebFace (Yi, Lei, Liao, & Li, 2014) and FaceScrub (Ng & Winkler, 2014). The algorithm learns a mapping from face images to an embedding space where distances in the embedding space directly correspond to a measure of face similarity (faces of the same person have small distances and faces of different people have large distances). OpenFace trains its output to be a feature vector of length 128 using a triplet-based loss function. The triplets consist of two matching face images and one non-matching face image and the loss aims to separate the positive pair from the negative by a distance margin.

5.1.3. Image similarity measurements

Image similarity scores are Euclidian distances between feature vectors. We computed similarity scores between original pictures and changed pictures for each type of change (high-PS or low-PS changes). We also computed similarity scores for Same pairs (the "base" and "reference" pictures), and for Different pairs (all pairs of "base" pictures). To compare human and machine distance scores we normalized each of the ratings by dividing it with its largest score (6 for human dissimilarity scores and 2 for DNN dissimilarity scores)

5.2. Results & discussion

Fig. 7 shows the mean similarity scores for the face images used in the human matching task described in Experiment 1. The results for machine face matching are very similar to the results of human face matching. A repeated measure ANOVA on machine similarity scores for the four face conditions with face stimuli as a random variable revealed a main effect of condition ($F(3,72) = 117.63$, $p < .0001$, $\eta^2 = .36$). Post hoc t-tests corrected for 6 comparisons (p = .008) revealed a significant difference among all conditions ($t(24) > 3.5$, $p < .002$) in the same direction as was found for Humans.

To directly compare the representations of humans and machines, a mixed ANOVA with DNN/Human as a between-groups factor and Condition (Same, Low-PS, High-PS, Different) as repeated measures reveal a significant main effect of Condition ($F(3,144) = 388.83$, $p < .0001$, $\eta^2 = .89$) and an interaction between the two factors ($F(3,144) = 3.68$, $p < .02$, $\eta^2 = .07$), which reflects the larger difference between high-PS and different faces for the algorithm than humans ($F(1,48) = 10.33$, $p < .005$, $\eta^2 = .18$). This may result from the fact that the algorithm training is performed on faces with cropped hair, which has a major effect on human face recognition (see Fig. 6). We therefore further examined the effect of each feature on the representation of faces by the DNN by comparing the feature vector distance of faces that differed in only one feature. This analysis revealed that out of the high-PS features, hair and eye color are not used as much by the algorithm, and out of the low-PS features, the nose is used by the algorithm (see Fig. S6).

In addition, we computed similarity scores between original pictures and changed pictures for each type of change (high-PS or low-PS changes) across the different layers of the DNN. Results show larger distance scores for low-PS than high-PS changes in low-level layers of the DNN, indicating that low-PS changes results in larger low-level image changes. In higher layers of the DNNs, results reversed, showing higher distance scores for high-PS than low-PS changes, consistent with human performance (see Fig. S7).

In summary, we found that the same high-PS features that are critical for human face matching, are also used by DNNs trained with unconstrained faces. The fact that the same features are critical also for

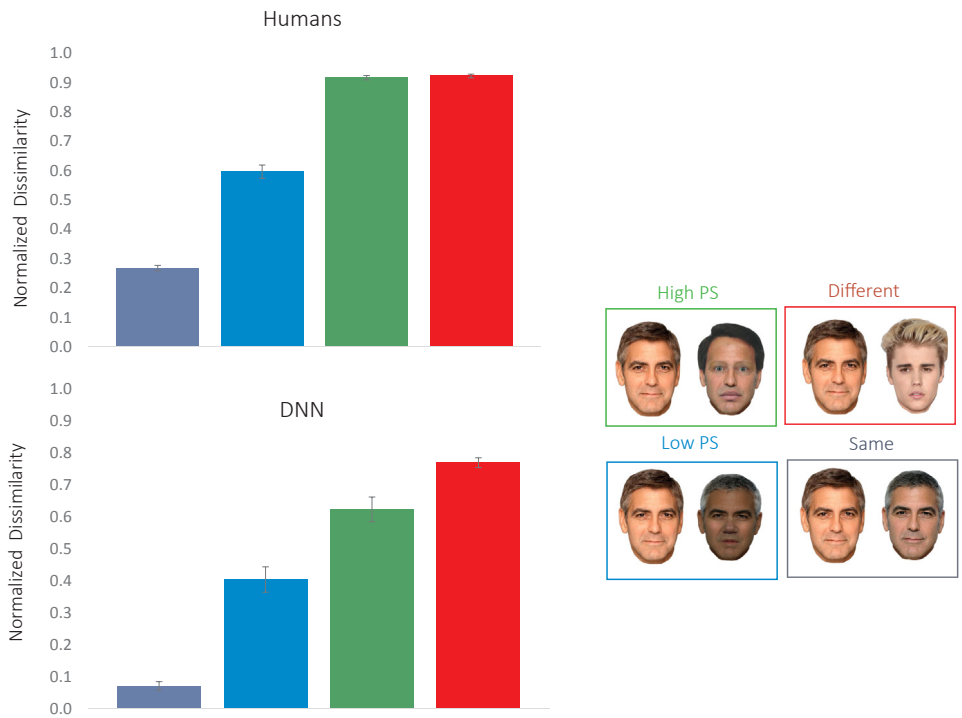

Fig. 7. Similarity scores between the four types of face pairs in humans and a DNN face recognition algorithm (Experiment 4) shows that like humans, DNNs show higher sensitivity to high-PS than low-PS features. Error bars indicate the standard error of each condition.

machine face recognition, indicates that these features have visual importance beyond their semantic meaning. We do not argue that the feature-vector representation of the DNN that we used specifically has eyebrow thickness or lip thickness encoded in it, but that high-PS features are highly correlated with the information represented in the network. We suggest that these features are important because they are invariant across different appearances of the same identity, whereas low-PS features, such as skin color and eye distance, vary across different appearances, such as lighting and head pose.

6. General discussion

We discovered a subset of facial features that are critical for human familiar and unfamiliar face matching, human familiar face recognition and DNN face recognition. To reveal these features, we used a novel *reverse-engineering* approach, in which we replaced different facial features until faces were perceived as different identities. We found that faces that differed in 5 or even 3 high-PS features but not those that differed in low-PS features, were perceived as different identities. An important difference between high and low-PS features is that the latter tend to vary under variations in appearance caused by changes in pose (Abudarham & Yovel, 2016), in illumination or expression, thus making them less useful not only for discrimination between identities but also for generalization across different appearances of the same individual.

Our findings have important implications for several major areas in the study of face recognition. First, they suggest a novel categorization of facial features instead of the commonly used external-internal and part-configuration categorization; second, in contrast to current theories that suggest that different facial features are used for familiar and unfamiliar faces, our findings suggest that the same features are used for familiar and unfamiliar face processing. Third, our findings go beyond current comparisons between human and machine face recognition, which primarily focus on performance levels, and examine the similarity of the internal face representations in humans and machines. We will discuss each of these topics in turn.

The face recognition literature has long been interested in discovering which facial features determine the identity of the face. Two prominent categorizations of facial features have been extensively examined in previous studies: external vs. internal features, and configural vs. part-based categorization. With respect to the external-internal dichotomy, it has been suggested that internal features are more critical for the identification of familiar faces but not for unfamiliar faces (Ellis et al., 1979; Young et al., 1985). These studies typically crop external facial features and present either the external part, which includes the ears, face outline and the hair or only the internal features. Our method, which provided us with a much more detailed examination of the role of 20 different facial features, each was measured independently, and relies on human perceptual sensitivity (Abudarham & Yovel, 2016), revealed a new categorization that is inconsistent with the external-internal distinction. Out of the 5 high-PS features that we changed, 4 are internal features (lip-thickness, eye-color, eye-shape and eyebrow-thickness), and the fifth is the hair, which had a large effect on both familiar and unfamiliar face recognition (Fig. 5, see also (Sinha & Poggio, 1996)). The importance of eyebrows and the eyes for face recognition has been also shown by Sadr et al. (2003) who found impaired recognition for familiar faces presented without eyebrows or eyes. Furthermore, low-PS changes such as eye-distance and mouth-size are also internal features but were found to be non-critical for face identity. Importantly, our fine distinction allowed us to separately test features such as lip thickness vs. mouth size or eyebrow thickness vs. eyebrow shape, showing that the former but not the latter in these two examples, are important for the identity of the face. These findings suggest that a gross categorization to internal and external features is inept to reveal which features are used for face recognition.

A second frequently used categorization of facial features is the part vs. configural (i.e., spacing) distinction. Whereas earlier studies indicated that configural processing is more important than part-based processing for face recognition (Le Grand, Mondloch, Maurer, & Brent, 2001) later studies have shown that both spacing and part-based information are used for face recognition (Yovel & Duchaine, 2006; Yovel

& Kanwisher, 2004, for review see McKone & Yovel, 2009). The special role of configural processing in face recognition has been also challenged by studies showing that compressed familiar faces can be easily recognized despite the fact that their eye distance and face proportions are largely distorted (Gilad-Gutnick, Harmatz, Tsourides, Yovel, & Sinha, 2018; Hole, George, Eaves, & Rasek, 2002; Sandford & Burton, 2014). Our findings that eye distance and face proportion are not used for face recognition are also consistent with these results. On the other hand, eye shape and eyebrow- and lip-thickness, which are considered local, part-based features, were found to be important for face recognition. These data are therefore inconsistent with the common notion that face recognition does not depend on face parts. Instead, we suggest an alternative categorization, according to which facial features that are critical for face recognition are those that remain invariant across different appearances of the same individual, and can be easily discriminated across different identities.

A second important finding our study reveals is that the same critical features are used for familiar and unfamiliar faces. Recent studies have convincingly shown that our ability to match different images of the same individual is relatively poor for unfamiliar faces, but is superior for familiar faces (Jenkins, White, Van Montfort, & Mike Burton, 2011; Ritchie et al., 2015; Young & Burton, 2017b). These findings have led to the suggestion that the perceptual representation that we acquire during our experience with familiar faces is robust to these different variations. In particular, in contrast to the face space theory and our findings that the same set of features is used for all faces, Burton and colleagues (2015) have suggested that different dimensions are used for the representation of different familiar faces. This representation is based on the experience that we have with different images of that same person from which we extract the individual's invariant features that are different from the invariant features of other individuals. However, if a different set of dimensions is used to represent different familiar faces in memory, what is the process by which an incoming face is recognized as familiar? What feature-extraction is applied to a face *before* we recognize it, and before we know whether it is familiar or not? Face recognition is a process by which the representation of a face input is matched to its representation in memory. Such matching cannot be performed if faces in memory are coded along different dimensions than the dimensions used for initial measurements of the face images. Our findings that face identity of familiar and unfamiliar faces is based on the same set of features enable direct matching of the representation of an incoming face stimulus to a representation in memory, to determine if a face is familiar and if so who that person is. We therefore suggest that the perceptual representation of unfamiliar faces is based on our life-long experience with familiar faces. The face processing system learns which features are suitable for both discrimination and generalization of familiar faces, and then applies the same features for matching unfamiliar faces.

What may therefore underlie the better generalization across images of the same identity for familiar than unfamiliar faces in a face-matching task? Fig. 8 (left) shows an example of a pair of face images of the same person presented in a matching task. The face-space theory, which is based on more general theories of pattern-recognition, assumes that image matching involves feature-extraction and conversion into feature-vectors that are classified as the same or different person based on the distance between them (Fig. 8, center). For unfamiliar faces, all the information that we have is the images, and therefore very dissimilar images of a person may be wrongly categorized as different people. For familiar faces, we also have a stored representation in memory that can be matched to the feature vectors created for the face images presented in the matching task. The representation of faces in memory is based on the different encounters that we had with a familiar person. The larger variability among the images of the same individual that we encounter, the richer is their representation in memory (Burton et al., 2016; Kramer et al., 2017). We can therefore find a match to each of the two different images in the matching task in memory, if we have a similar representation of each image from previous encounters. In some other cases we may not have a similar representation of the familiar person in memory. For example, if we see a younger-age picture of a person we had met at an older age, we may not be able to match the younger image to the representation in memory of its older version (see Fig. 8). Once we learn that it is the same individual, we can update its representation in memory and now match its younger and older face images to the updated representation in memory.

But if the two faces are different perceptually, how can we link them to the same individual in memory? Here we suggest two possible models (Fig. 8, right): *A perceptual single-prototype model* and a *conceptual multiplesub-prototypes model* (Vanpaemel & Storms, 2008). The perceptual single-prototype model is consistent with the average face model suggested by Burton and colleagues (Jenkins & Burton, 2011) in which the representation of familiar faces in memory is the averaged representation of all the different images that we encountered of a given individual (right panel, top). An incoming image will be correctly matched as long as it is perceptually similar to the prototype/averaged representation. As also noted by Burton and colleagues (2011), the average face may not be a good representation of the large variability across different images of the same person. Thus, a conceptual multi sub-prototype model suggests that if two images of the same individual are different beyond a certain threshold, such that averaging them would generate an image that is too far from each of them, we may create an additional sub-prototype and keep them separately. In other words, we may cluster similar images of a person, and represent each cluster by the average feature vector of all the images in the cluster, thus creating multiple sub-prototypes, similar to the Varying Abstraction Model (VAM) suggested by Vanpaemel and Storms (2008) (see also Love, Medin, Gureckis, 2004). Because faces in memory are linked to the same *conceptual* information, matching such different images of the same individual must be based on conceptual similarity rather than perceptual similarity (right panel, bottom). An example of such conceptual representation are the *concept neurons* revealed in the medial temporal lobe that respond to very different images of the same individual that cannot be matched perceptually (e.g., Halle Barry with a cat-woman mask and a line drawing of her face) as well as their name (Quiroga, Reddy, Kreiman, Koch, & Fried, 2005; Quiroga, 2012). Future experiments are needed to directly test these two different representations of familiar faces in memory and better characterize the conceptual multiple sub-prototypes model. Importantly, these representations do not assume different critical features for familiar and unfamiliar faces but still account for the better performance that is typically found for familiar than unfamiliar faces.

The face stimuli that we used to examine the representation of familiar faces were celebrity faces, with which we have limited experience relative to personally familiar faces. It is therefore possible that our results may not generalize to personally familiar faces for which changing high-PS features may have a smaller effect. Whereas this is an empirical question that will need to be examined in future studies, it is noteworthy, that many of the previous studies that posited that familiar face recognition relies on a different set of features than unfamiliar faces used celebrity faces (e.g. Ellis et al., 1979; Kramer et al., 2017; Young et al., 1985). Furthermore, the superb generalization across different images of the same individual was found for celebrity faces (Jenkins et al., 2011; Ritchie & Burton, 2016). Thus, our findings are relevant to many previous studies that used celebrity faces to study the differences between the representations of familiar and unfamiliar faces. The challenge of testing our predictions for personally familiar faces would be the need to use different face sets for different individuals. Nevertheless, this study is feasible and should be conducted in the future.

A third finding reported in our study is that the same features that were found to be important for human face recognition are also used by a DNN trained on unconstrained faces. Comparing human and machine face recognition adds further support for the validity of our suggested

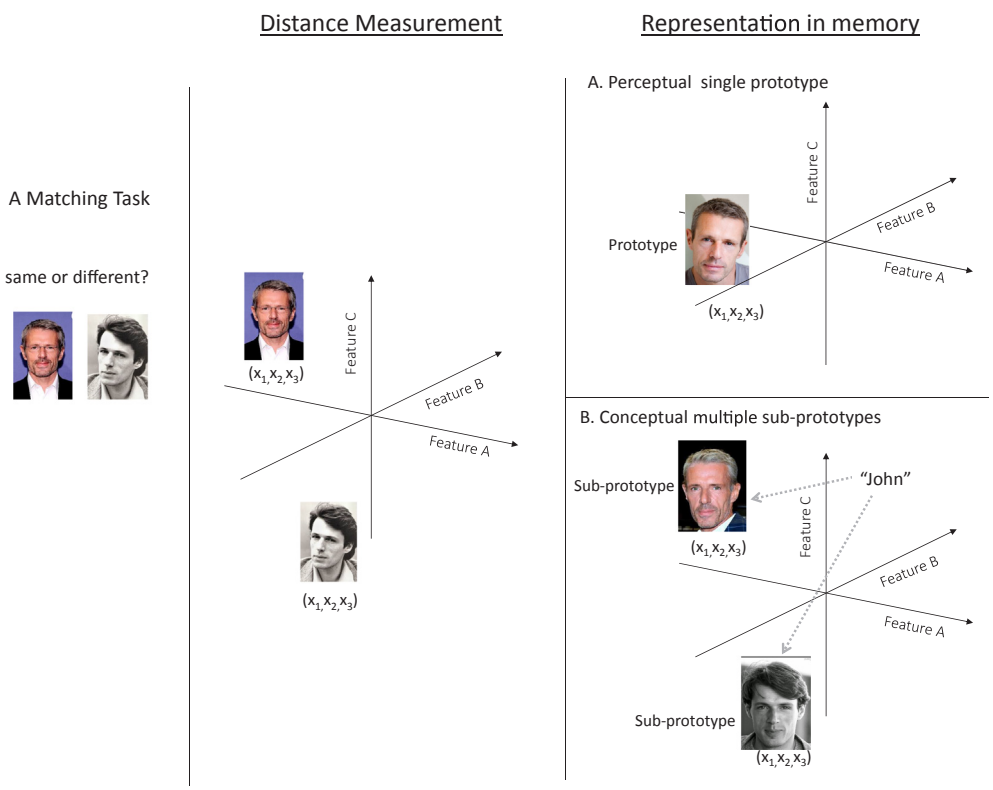

Fig. 8. In a face matching task two different images of the same individual are presented side by side (left panel shows for example the French actor and singer Lambert Wilson). The two faces are measured based on their critical features and the distance between the feature vectors is used for classification (center panel). If faces are familiar they are represented in memory based on the same critical features such that an incoming image can be matched to these representations (right panel). We propose two different types of representations that may result in better matching of familiar than unfamiliar faces. According to a perceptual single-prototype representation, the representation of face identity in memory is an averaged image/the most representative image (prototype) of previous encounters with that person (Jenkins & Burton, 2011). If this representation is perceptually similar to the two different images then these different images can be linked perceptually. According to a conceptual multiple sub-prototype model, if two representations of the same individual are perceptually different, one prototype may not be a good representation of both and therefore each identity is represented by multiple sub-prototypes that can be linked based on their conceptual similarity.

feature categorization. Even though the features that we used to tag faces are "nameable" features, i.e. features with semantic meaning, the fact that machine face recognition is also sensitive to the same features indicates that high-PS features capture the essence of the identity of a face.

Previous studies that compared human and machine face recognition have focused on performance level, aiming to develop an algorithm that reaches the level of humans or beyond (Phillips et al., 2018; Phillips, Hill, Swindle, & O'Toole, 2015). Our novel method allowed us to go beyond the performance level and compare the nature of the representation of humans and machines. In particular, we showed that humans and DNNs are tuned to the same facial information and overlook other facial information that may not be useful for face recognition. We believe that this similarity in representations is because both humans and DNNs are trained on high variability of labeled images of the same and different individuals and need to learn which features remain invariant across different images of the same individual and at the same time vary across different identities.

Finally, it is noteworthy that the features that we discovered are useful for the identity of Caucasian male faces. Features such as eye color and hair color that are critical for Caucasian faces are unlikely to be used for Asian or African faces, which show smaller variations in these features. Similarly, hair and eye-brow thickness that are also used for Caucasian faces are unlikely to be used to discriminate infant faces that hardly have hair or eyebrows, and are all look very similar to one another (Yovel et al., 2012). Our findings may therefore account also for the well-established other-race effect, indicating that lower performance for other-race faces may be due to the usage of set of facial features that are not diagnostic for the race we had no experience with. Our methods can be therefore used to determine the critical features of any category of faces based on the same principle of perceptual sensitivity and the reverse engineering approach (Abudarham & Yovel, 2016). These features are typically learned through our real-life experiences with different categories of faces. As mentioned above, we suggest that the training needed to extract these features is based on our experience with familiar faces, rather than passive exposure to unfamiliar faces. This claim is based on previous studies that show that it is the association of faces with person-related information that improves face recognition rather than the pure perceptual exposure to a large number of faces (Schwartz & Yovel, 2016; Tanaka & Pierce, 2009; Yovel et al., 2012).

In summary, we discovered a subset of features that determine the identity of familiar and unfamiliar faces in humans and a face recognition DNN. We suggest that these features are learned through our rich experience with familiar faces to obtain optimal generalization and discrimination. These features are then applied to determine the similarity between any incoming face stimuli and are therefore similar to both familiar and unfamiliar faces. Future studies are needed to further explore the mechanisms that account for the different abilities that humans show for familiar than unfamiliar faces, and the extent to which they are based on conceptual rather than perceptual matching as was suggested here.

Acknowledgements

Portions of the research in this manuscript use the FERET database of facial images collected under the FERET program, sponsored by the DOD Counterdrug Technology Development Program Office. The said images were processed by the author for this specific experiment. You may not use any of the images in this experiment without written permission from NIST and from the author.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.cognition.2018.09.002.

References

Abudarham, N., & Yovel, G. (2016). Reverse engineering the face space: Discovering the critical features for face identification. *Journal of Vision, 16*(3), 40. https://doi.org/10.1167/16.3.40.
Burton, A. M., Jenkins, R., & Schweinberger, S. R. (2011). Mental representations of

familiar faces. *British Journal of Psychology, 102*(4), 943–958.
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. https://doi.org/10.1111/cogs.12231.
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.
Gilad-Gutnick, S., Harmatz, E. S., Tsourides, K., Yovel, G., & Sinha, P. (2018). Recognizing Facial Slivers. *Journal of Cognitive Neuroscience, 30*(7), 951–962.
Hole, G. J., George, P. A., Eaves, K., & Rasek, A. (2002). Effects of geometric distortions on face-recognition performance. *PERCEPTION-LONDON-, 31*(10), 1221–1240.
Jenkins, R., & Burton, A. M. (2011). Stable face representations. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 366*(1571), 1671–1683. https://doi.org/10.1098/rstb.2010.0379.
Jenkins, R., White, D., Van Montfort, X., & Mike Burton, A. (2011). Variability in photos of the same face. *Cognition, 121*(3), 313–323. https://doi.org/10.1016/j.cognition.2011.08.001.
Kramer, R. S. S., Towler, A., Reynolds, M. G., & Burton, A. M. (2017). Familiarity and Within-Person facial variability: The importance of the internal and external features. *Perception*. https://doi.org/10.1177/0301006617725242.
Le Grand, R., Mondloch, C. J., Maurer, D., & Brent, H. P. (2001). Neuroperception: Early visual experience and face processing. *Nature, 410*(6831), 890.
Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review, 111*(2), 309.
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*(5), 778–797. https://doi.org/10.3758/PBR.16.5.778.
Ng, H. W., & Winkler, S. (2014). A data-driven approach to cleaning large face datasets. *IEEE International Conference on Image Processing (ICIP), 265*(265), 530.
O'Donnell, C., & Bruce, V. (2001). Familiarisation with faces selectively enhances sensitivity to changes made to the eyes. *Perception, 30*(6), 755–764. https://doi.org/10.1068/p3027.
O'Toole, A. J., Castillo, C. D., Parde, C. J., Hill, M. Q., & Chellappa, R. (2018). Face space representations in deep convolutional neural networks. *Trends in Cognitive Sciences, 22*(9), 794–809. https://doi.org/10.1016/j.tics.2018.06.006.
Phillips, P. J., Hill, M. Q., Swindle, J. A., & O'Toole, A. J. (2015). Human and algorithm performance on the PaSC face Recognition Challenge. In 2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems, BTAS 2015. http://doi.org/10.1109/BTAS.2015.7358765.
Phillips, P. J., Wechsler, H., Huang, J., & Rauss, P. J. (1998). The FERET database and evaluation procedure for face-recognition algorithms. *Image and Vision Computing, 16*(5), 295–306.
Phillips, P. J., Yates, A. N., Hu, Y., Hahn, C. A., Noyes, E., Jackson, K., ... O'Toole, A. J. (2018). Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms. *Proceedings of the National Academy of Sciences, 201721355*. https://doi.org/10.1073/pnas.1721355115.
Quiroga, R. Q. (2012). Concept cells: The building blocks of declarative memory functions. *Nature Reviews Neuroscience, 13*(8), 587–597. https://doi.org/10.1038/nrn3251.
Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature, 435*(7045), 1102–1107. https://doi.org/10.1038/nature03687.
Ritchie, K. L., & Burton, M. A. (2016). Learning faces from variability. *The Quarterly Journal of Experimental Psychology, 0218*(May), 1–9. https://doi.org/10.1080/17470218.2015.1136656.
Ritchie, K. L., Smith, F. G., Jenkins, R., Bindemann, M., White, D., & Burton, A. M. (2015). Viewers base estimates of face matching accuracy on their own familiarity: Explaining the photo-ID paradox. *Cognition, 141*, 161–169. https://doi.org/10.1016/j.cognition.2015.05.002.
Rubner, Y., Tomasi, C., & Guibas, L. J. (2000). The Earth Mover' s distance as a metric for image retrieval. *Work*. https://doi.org/10.1023/A:1026543900054.
Sandford, A., & Burton, A. M. (2014). Tolerance for distorted faces: Challenges to a configural processing account of familiar face recognition. *Cognition, 132*(3), 262–268. https://doi.org/10.1016/j.cognition.2014.04.005.
Sadr, J., Jarudi, I., & Sinha, P. (2003). The role of eyebrows in face recognition. *Perception, 32*(3), 285–293.
Schroff, F., & Philbin, J. (2015). A Unified Embedding for Face Recognition and Clustering. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*(pp. 815–823).
Schwartz, L., & Yovel, G. (2016). The roles of perceptual and conceptual information in face recognition. *Journal of Experimental Psychology: General, 145*(11), 1493–1511. https://doi.org/10.1037/xge0000220.
Sinha, P., & Poggio, T. (1996). I think I know that face.
Tanaka, J. W., & Pierce, L. J. (2009). The neural plasticity of other-race face recognition. *Cognitive, Affective, & Behavioral Neuroscience, 9*(1), 122–131. https://doi.org/10.3758/CABN.9.1.122.
Vanpaemel, W., & Storms, G. (2008). In search of abstraction: The varying abstraction model of categorization. *Psychonomic Bulletin and Review*. https://doi.org/10.3758/PBR.15.4.732.
Yi, D., Lei, Z., Liao, S., & Li, S. Z. (2014). Learning face representation from scratch. *ArXiv, 1411*, 7923.
Young, A. W., & Burton, A. M. (2017b). Recognizing faces. *Current Directions in Psychological Science, 26*(3), 212–217. https://doi.org/10.1177/0963721416688114.
Young, A. W., & Burton, A. M. (2017a). Are we face experts ? *Trends in Cognitive Sciences, xx, 1–11*. https://doi.org/10.1016/j.tics.2017.11.007.
Young, A. W., Hay, D. C., McWeeny, K. H., Flude, B. M., & Ellis, A. W. (1985). Matching familiar and unfamiliar faces on internal and external features. *Perception, 14*(6), 737–746.
Yovel, G., & Duchaine, B. (2006). Specialized face perception mechanisms extract both part and spacing information: Evidence from developmental Prosopagnosia. *Journal of Cognitive Neuroscience, 18*(4), 580–593. https://doi.org/10.1162/Jocn.2006.18.4.580.
Yovel, G., Halsband, K., Pelleg, M., Farkash, N., Gal, B., & Goshen-Gottstein, Y. (2012). Can massive but passive exposure to faces contribute to face recognition abilities? *Journal of Experimental Psychology-Human Perception and Performance, 38*(2), 285–289. https://doi.org/10.1037/a0027077.
Yovel, G., & Kanwisher, N. (2004). Face perception: Domain specific, not process specific. *Neuron, 44*(5), 889–898. https://doi.org/10.1016/J.Neuron.2004.11.018.