

advances.sciencemag.org/cgi/content/full/6/22/eaba3274/DC1

Supplementary Materials for

Same father, same face: Deep learning reveals selection for signaling kinship in a wild primate

M. J. E. Charpentier*, M. Harté, C. Poirotte, J. Meric de Bellefon, B. Laubi, P. M. Kappeler, J. P. Renoult

*Corresponding author. Email: marie.charpentier@umontpellier.fr

Published 27 May 2020, *Sci. Adv.* **6**, eaba3274 (2020)

DOI: [10.1126/sciadv.aba3274](https://doi.org/10.1126/sciadv.aba3274)

This PDF file includes:

Supplementary Methods
Supplementary Results
Fig. S1
Tables S1 to S7
References

SUPPLEMENTARY MATERIALS

Supplementary Methods

Image database and pre-processing

The mandrill face database includes ~16k images representing 276 different mandrills originating from the wild studied population (Gabon; 12.9k images), the semi-captive population of the Centre International de Recherche Medicale de Franceville (CIRMF, Gabon; 2.7k images) and other sources (internet, the Wildlife Reserves of Singapore, Zoo of Grandy; 0.4k images). Images from Gabon (wild and CIRMF populations) were taken between 2012 and 2018, using different camera models. Pictures represent individuals that are awake and passive, awake and active (i.e. feeding, grooming, vocalizing) or anesthetized during one of the bi-annual captures (representing 1.1k images). We frequently photographed active individuals using the slow burst mode of cameras, which allowed to capture variation in face position and expression while avoiding identical frames. The multiple frames obtained while using the slow burst mode are hereafter referred to as “a burst-mode series”. Images were then manually oriented to align pupils horizontally, and cropped to generate square portraits centered on the nose and excluding the ears. Each portrait was manually labelled based on face position and image quality. Face position included two levels:

- P0: the face is in profile view (approx. $>30^\circ$), either from below or above, or in frontal view but significantly occluded ($>50\%$).
- P1: the face is in frontal view (approx. $<30^\circ$) and occlusion covers less than 50% of the face.

Image quality included five levels:

- Q0: very bad quality; it impossible to recognize the individual without contextual information, even for experienced field assistants.
- Q1: poor quality; individual recognition is possible but challenging for experienced field assistants.
- Q2: medium quality; individual recognition is easy for field assistants but the portrait does not meet Q3 criteria.
- Q3: high, “passport” quality image; individual recognition is easy for field assistants, the face is in frontal view, it has a neutral expression and is not partially occluded, the image is sharp, has no shadow or lighting spot. This quality level excludes images that meet Q4 criteria.
- Q4: a single image of a burst-mode series, which meets Q3 criteria.

We excluded P0 and Q0 images from all analyses. P1Q1234 dataset (simplified as Q1234 hereafter) contained 14,8k images.

Dataset partitions

The Q1234 dataset represents 276 individuals belonging to one or several of the following age classes: infant, juvenile, adolescent (only for males), subadult (only for males) and adult. Because the face of an individual varies considerably between its different age classes (table S2), we used ind-age classes rather than individuals for the identification task, that is, we treated two ind-age classes representing the same individual as distinct and independent classes. The Q1234 dataset contains 343 ind-age classes (mean number of images per class: 50.1; range: [1,419]).

Q1234 was split into a learning set and a test set, which were different for the adult female and juvenile analyses. For the adult female analysis, the learning set included pictures of semi-captive and captive males and females of all age classes, as well as wild individuals from the studied wild population but adult females. For the juvenile analysis, the learning set was the same

as above but it included wild adult females and excluded wild juveniles. Each learning set was itself split into a training set and a validation set. The validation set was used to parameterize the model for the face identification task. It contained two images of each class. For a correct evaluation of training performances, we ensured that none of the validation image was from a burst-mode series that also contained images present in the training set. The training set contained all other images of the learning set. Because we were able to reach high performances (see results) despite a large imbalance between classes in the training sets, we did not attempt to correct for this imbalance. Last, the test set included either adult females or juveniles from the wild population. To maximize the quality of resemblance measurements, we selected only Q4 images for the adult female test set. For the juvenile test set, we selected Q34 because we had insufficient Q4 images. The characteristics of the different image datasets are given in table S3.

Face identification

We trained a deep convolutional neural network (DNN) to identify individual (ind-age) mandrills as a goal to learn a deep representation of mandrill faces. We applied a transfer learning procedure (45) by initializing the training with VGG-Face, a network that has learned to recognize 2,6k different humans from 2,6M portrait pictures. VGG-Face uses a VGG-16 architecture, which consists of 5 blocks of convolution (that filters images), nonlinear ReLU activation (that sets to zero all negative values) and max pooling (that selects a maximum locally) layers (*CB*), a flatten layer (*Fl*) that converts a 2D matrix of feature activations into a vector, two fully connected layers (*FC*) and a softmax classification layer (*SM*). VGG-16 can thus be written as *CB1–CB2–CB3–CB4–CB5–Fl–FC1–FC2–SM*. For transfer learning, we replaced *SM* by a new layer of dimension fitting the number of classes in the new mandrill identification task, which varied with datasets. We included two dropout layers (with 50% dropout probability), one after each *FC* layer, to limit the risk of overfitting. We trained the network using a stochastic gradient descent with momentum optimizer with initial learning rate of 10^{-5} for *CB* layers and 10^{-5} for *FC* and *SM* layers. The learning rate decreased by a factor 10 every 5 epochs. Learning continued until the validation loss did not decrease further after three consecutive epochs, which required approximately 15 epochs (Fig. S1). In order to match the input size of VGG-Face, mandrill portraits were downsized to $224 \times 224 \times 3$ prior to analyses. We set the batch size (the number of images used for optimization during one iteration) to 32. This is small value compared to standard practice; however, as previously demonstrated (46), we found that a small batch size reduced overfitting significantly compared to larger sizes (64 or 128; results not shown). We limited overfitting further by using “data augmentation” (47). The transformations used in data augmentation were chosen to reproduce the spurious variation that could occur during the manual processing of images (i.e. cropping and alignment). Each iteration, images were shifted horizontally and vertically (by a number of pixels random selected within the range [-40; 40]), rotated (range of degree: [-20; 20]) and scaled (range of factor: [-0.7; 1.2]). We did not flip images horizontally, assuming that some bilaterally asymmetrical features could be important for individual recognition. All deep learning analyses were ran on a single NVIDIA GeForce GTX 1080 GPU with MATLAB.

Face verification

After evaluating performance in face identification with the validation sets, we retrained the DNN using the full learning set to maximize the number of images. VGG-Mandrill, the newly trained DNN, was then used to extract deep feature activation vectors, a compact and informative representation of a mandrill face. The distance between feature activation vectors predicts the

resemblance between images (48). In this study, we followed previously published procedures (21) and used a χ^2 distance calculated with normalized features. Normalization was achieved by first subtracting to each activation of a vector the minimal value found for this feature across the entire learning set. This set to zero the lower boundary of the feature space (this step of the normalization is necessary *only* for testing the effect of activations before ReLU transformation; see below). Next, we divided each activation by the maximum value found for this feature across the entire learning set (or by 0.05 if the maximum value was under this threshold, to avoid division by a small number). This set to one the upper boundary of the feature space. Last, we normalized each feature vector by its L2-norm (Euclidean normalization).

Studies on face verification, the task of identifying whether two faces represent the same individual or not, performed best when learning a distance metric in the deep feature space (39). Because the various features in the deep feature space contribute differently to predicting facial resemblance, a learned distance metric aims at finding the feature weights that optimize face verification. Following a previous study (21), we used a linear support vector machine (SVM) to learn a distance metric. We randomly selected 15k pairs of images representing different individuals and 15k pairs representing same individuals, and for each pair we calculated the χ^2 difference $(f_1[i]-f_2[i])^2/(f_1[i]+f_2[i])$, where f_1 and f_2 are the normalized feature vectors of the two images in a pair and i the index of a feature. Then, we ran the SVM with the χ^2 differences as explanatory variables and 0 (different-individual pairs) or 1 (same-individual pairs) as a response variable. The SVM output the accuracy of the face verification task as well as the weight of each feature. Weights were eventually used to calculate a weighted χ^2 distance as $\chi^2(f_1, f_2) = \sum_i \omega_i (f_1[i] - f_2[i])^2 / (f_1[i] + f_2[i])$, where ω is the vector of feature weights.

Face verification with test sets

We evaluated the SVM trained for face verification on test tests and compared different feature spaces. Feature activations were extracted from *FC1* (as a vector of length 4,096), *FC2* (vector of length 4,096) or both layers (*FC12*: vector of length 8,192). We further analyzed the importance of the nonlinear transformation of the deep feature space by extracting activations either before or after the ReLU activation function. For this evaluation, we built a balanced dataset with an equal number of same-individual and different-individual pairs of portrait images, all from test sets. The number of different-individual pairs was set to match the number of same-individual pairs; different-individual pairs were randomly selected.

Supplementary Results

Reproductive skew and number of kin

For each female involved in the study, we calculated an average number of maternal half-sisters (MHS) and paternal half-sisters (PHS) present in the study group at the end of the study, as follows (table S1). MHSmin represents the minimum number of MHS. MHS are accurately known from non-ambiguous behavioral observations or genetic analyses. MHSmax represents the probable maximum number of MHS. MHSmax was obtained by considering that each adult female gave birth to a living offspring (of either sex) every 18 months from 4yrs-old to the end of the study. We removed known deaths from MHSmax and divided this figure by two to consider only female MHS. The number of MHS for each study female was the mean between MHSmin and MHSmax. For PHS, and based on patterns of reproductive skew and on known or estimated dates of birth in the study group, we estimated that the alpha males sired on average 60% of offspring each year. We considered that non-dominant males sired, on average, 15% of offspring

each year. We calculated the number of PHS based on the total number of offspring born each year, on the alpha males' tenure and male presence in the group, and on whether female's father was the alpha male during her conception or not. We obtained the figures presented in table S1. Note that we did not use these figures for our statistical analyses -based on true MHS and PHS-; these figures only served to estimate the number of MHS and PHS available, on average, for a female mandrill (see the Introduction in the main text).

Face identification

VGG-Mandrill was able to identify mandrill faces with high accuracy and generalization capacity (i.e. limited overfitting; Fig. S1). The highest accuracy was 91.1% for the adult female learning set and 91.9% for the juvenile learning set (table S4). The highest accuracies were reached with the sets including the largest number of images. More precisely, a wider image set (i.e. with more classes) was better than a deeper image set (i.e. with more images per class), and maximizing the number of images was more important than maximizing their quality (adding poor quality images *increased* performance).

Face verification

The accuracy of face verification was evaluated on test sets directly, using either Q1234-12-Tr+V (which yielded the highest accuracy in face identification) or Q234-25-TR+V (which is more stringent regarding both image quality and the number of images per class) for training the DNN to identify faces and for learning the distance metric. The highest accuracy was 83% for the adult female test set (table S5), and 90% for the juvenile test set (table S6). The higher accuracy with the juvenile set compared to the adult female set is expected given that same-individual pictures in this set are on average more similar than same-individual pictures in the adult female set, which contains only a single image per burst-mode series.

We found a difference in the best parameter settings between the adult female and juvenile sets only for the type of learning set. With the adult female set, the highest accuracy was reached using the fullest learning set (Q1234-12-Tr+V), as in the face identification task. With the juvenile set, the highest accuracy was reached with the more stringent learning set (Q234-25-Tr+V), as opposed to the face identification task. Note, however, that the influence of the type of learning set is less (maximum $\pm 3\%$) than that of other parameters.

The choice of the DNN layer at which feature activations are extracted influenced the accuracy of face verification by up to $\pm 5\%$. For either test set, *FC1* activations yielded higher performance compared to *FC2* and *FC12* activations. A more influencing factor is the use of a new metric learned specifically for face verification. A learned distance metric raised the accuracy by up to 15 % (with the adult female set) and 17 % (with the juvenile set). Last, the non-linearization of activations was even more influencing. Extracting activations after a ReLU transformation boosted performance by up to 17 % (with the adult female set) and 25 % (with the juvenile set).

Based on these results, for both the adult female and juvenile analyses we estimated resemblance between pairs of mandrill faces by:

- training the DNN to perform face identification on Q1234-12-Tr+V learning set,
- extracting the *FC1* feature activation vectors (after ReLU transformation) of Q1234-12-Tr+V images,
- training a SVM classifier with these vectors to verify face identity and compute feature weights,
- extracting the *FC1* feature activation vectors (after ReLU transformation) of test sets,
- calculating the weighted distance between every pairs of images in each test set,

- averaging pairwise distances for every different pairs.

Supplementary Figure

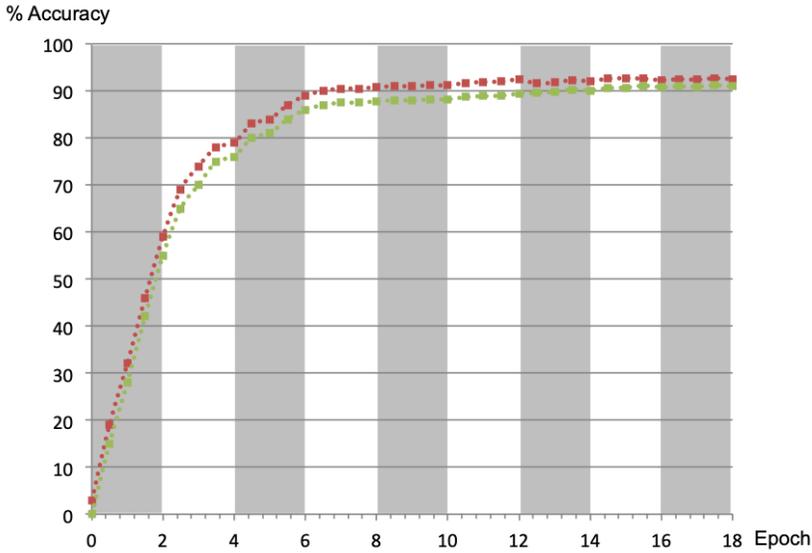


Fig. S1. Learning curve in the face identification task. Evolution of accuracy for the training set (f-Q1234-12-Tr; in red) and validation set (f-Q1234-12-V; in green) during a typical run. The accuracy stabilized after 15 epochs at around 91% for the validation set; the run thus automatically stopped after 18 epochs (based on a stabilization of the loss). The small difference between the training accuracy and the validation accuracy indicates limited or no effect of overfitting.

Supplementary Tables

Table S1. Estimated number of MHS and PHS. For each female involved in the study, an average number of maternal half-sisters (MHS) and paternal half-sisters (PHS) available were calculated (see above).

	Mean number (SD)	Variance
MHS	4.7 (2.5)	6.1
PHS	9.9 (8.8)	77.9

Table S2. Kinship and social behavior among pairs aged less than 2 yrs apart. Statistics obtained from Generalized Linear Mixed Models (proc GENMOD, SAS Studio) with a negative binomial distribution performed to study the relationships between grooming and spatial association recorded across 38 adult females aged less than 2 yrs apart (N=64 with 8 MHS, 34 PHS and 22 NK) and a set of explanatory variables, including kinship. Note that the statistical model on aggression did not converge.

	<i>Explanatory variables</i>	χ^2	<i>P</i>
Grooming	Kinship	0.67	0.71
	Rank difference	1.11	0.57
	Age difference	7.19	<0.01
	Kinship*Age difference	6.49	0.04
Association	Kinship	11.30	<0.01
	Rank difference	6.19	0.05
	Age difference	0	1
	Kinship*Age difference	7.04	0.03

Table S3. Characteristics of image sets. The name of image sets is coded as followed: whether the dataset is for the adult female (f) or juvenile (j) analysis – the level of image quality (e.g., Q1234 = all qualities retained) – the minimal number of images per class before the training/validation partition (either 12 or 25) – the use of the image sets (Tr: training, V: validation, Te: test). Wild: wild population, ad.: adult, fem.: female, #: number of.

Set name	Sex & age categories	# images	# classes	Mean image/class
f-Q1234-12-Tr	all but ad. fem. from Wild	8429	158	53.3
f-Q1234-25-Tr	all but ad. fem. from Wild	8218	141	58.3
f-Q234-12-Tr	all but ad. fem. from Wild	7914	132	60.0
f-Q234-25-Tr	all but ad. fem. from Wild	7386	108	68.4
f-Q1234-12-V	all but ad. fem. from Wild	316	158	2
f-Q1234-25-V	all but ad. fem. from Wild	182	141	2
f-Q234-12-V	all but ad. fem. from Wild	264	132	2
f-Q234-25-V	all but ad. fem. from Wild	216	108	2
f-Q4-Te	ad. fem. from Wild	421	55	7.6
j-Q1234-12-Tr	all but juv. from Wild	13772	202	68.2
j-Q1234-25-Tr	all but juv. from Wild	11901	155	76.8
j-Q234-12-Tr	all but juv. from Wild	13155	200	65.7
j-Q234-25-Tr	all but juv. from Wild	11548	155	74.5
j-Q1234-12-V	all but juv. from Wild	404	202	2
j-Q1234-25-V	all but juv. from Wild	310	155	2
j-Q234-12-V	all but juv. from Wild	400	200	2
j-Q234-25-V	all but juv. from Wild	310	155	2
j-Q1234-Te	all but juv. from Wild	472	50	9.4
j-Q34-Te	juv. from Wild	383	50	7.6

Table S4. Performance of VGG-Mandrill in face identification. The accuracy (\pm sem) is averaged over 5 runs. Highest accuracy in bold.

Training set	Validation set	Accuracy \pm sem
Adult females		
f-Q1234-12-Tr	f-Q1234-12-V	0.911 \pm 0.118
f-Q1234-25-Tr	f-Q1234-25-V	0.898 \pm 0.163
f-Q234-12-Tr	f-Q234-12-V	0.883 \pm 0.297
f-Q234-25-Tr	f-Q234-25-V	0.867 \pm 0.304
Juveniles		
j-Q1234-12-Tr	j-Q1234-12-V	0.919 \pm 0.321
j-Q1234-25-Tr	j-Q1234-25-V	0.906 \pm 0.478
j-Q234-12-Tr	j-Q234-12-V	0.872 \pm 0.677
j-Q234-25-Tr	j-Q234-25-V	0.851 \pm 0.694

Table S5. Performance of face verification on the female adult test sets. The accuracy (\pm sem) is averaged over 5 runs. Highest accuracy in bold.

Learning set	Layer	ReLU	Metric Learning	Accuracy \pm sem
Q1234-12-Tr+V	<i>FC1</i>	No	No	0.639 \pm 0.016
Q1234-12-Tr+V	<i>FC2</i>	No	No	0.636 \pm 0.014
Q1234-12-Tr+V	<i>FC12</i>	No	No	0.655 \pm 0.012
Q1234-12-Tr+V	<i>FC1</i>	Yes	No	0.705 \pm 0.018
Q1234-12-Tr+V	<i>FC2</i>	Yes	No	0.652 \pm 0.016
Q1234-12-Tr+V	<i>FC12</i>	Yes	No	0.646 \pm 0.010
Q1234-12-Tr+V	<i>FC1</i>	No	Yes	0.679 \pm 0.019
Q1234-12-Tr+V	<i>FC2</i>	No	Yes	0.652 \pm 0.017
Q1234-12-Tr+V	<i>FC12</i>	No	Yes	0.699 \pm 0.013
Q1234-12-Tr+V	<i>FC1</i>	Yes	Yes	0.834 \pm 0.015
Q1234-12-Tr+V	<i>FC2</i>	Yes	Yes	0.785 \pm 0.016
Q1234-12-Tr+V	<i>FC12</i>	Yes	Yes	0.796 \pm 0.010
Q234-25-Tr+V	<i>FC1</i>	No	No	0.601 \pm 0.011
Q234-25-Tr+V	<i>FC2</i>	No	No	0.611 \pm 0.018
Q234-25-Tr+V	<i>FC12</i>	No	No	0.612 \pm 0.017
Q234-25-Tr+V	<i>FC1</i>	Yes	No	0.680 \pm 0.009
Q234-25-Tr+V	<i>FC2</i>	Yes	No	0.662 \pm 0.009
Q234-25-Tr+V	<i>FC12</i>	Yes	No	0.659 \pm 0.014
Q234-25-Tr+V	<i>FC1</i>	No	Yes	0.648 \pm 0.013
Q234-25-Tr+V	<i>FC2</i>	No	Yes	0.655 \pm 0.016
Q234-25-Tr+V	<i>FC12</i>	No	Yes	0.656 \pm 0.014
Q234-25-Tr+V	<i>FC1</i>	Yes	Yes	0.820 \pm 0.015
Q234-25-Tr+V	<i>FC2</i>	Yes	Yes	0.776 \pm 0.010
Q234-25-Tr+V	<i>FC12</i>	Yes	Yes	0.785 \pm 0.010

Table S6. Performance of face verification on the female juvenile test sets. The accuracy (\pm sem) is averaged over 5 runs. Highest accuracy in bold.

Learning set	Layer	ReLU	Metric Learning	Accuracy \pm sem
Q1234-12-Tr+V	<i>FC1</i>	No	No	0.616 \pm 0.008
Q1234-12-Tr+V	<i>FC2</i>	No	No	0.603 \pm 0.012
Q1234-12-Tr+V	<i>FC12</i>	No	No	0.603 \pm 0.011
Q1234-12-Tr+V	<i>FC1</i>	Yes	No	0.740 \pm 0.007
Q1234-12-Tr+V	<i>FC2</i>	Yes	No	0.692 \pm 0.018
Q1234-12-Tr+V	<i>FC12</i>	Yes	No	0.699 \pm 0.002
Q1234-12-Tr+V	<i>FC1</i>	No	Yes	0.613 \pm 0.009
Q1234-12-Tr+V	<i>FC2</i>	No	Yes	0.634 \pm 0.010
Q1234-12-Tr+V	<i>FC12</i>	No	Yes	0.634 \pm 0.011
Q1234-12-Tr+V	<i>FC1</i>	Yes	Yes	0.868 \pm 0.010
Q1234-12-Tr+V	<i>FC2</i>	Yes	Yes	0.862 \pm 0.009
Q1234-12-Tr+V	<i>FC12</i>	Yes	Yes	0.856 \pm 0.014
Q234-25-Tr+V	<i>FC1</i>	No	No	0.645 \pm 0.012
Q234-25-Tr+V	<i>FC2</i>	No	No	0.627 \pm 0.018
Q234-25-Tr+V	<i>FC12</i>	No	No	0.619 \pm 0.013
Q234-25-Tr+V	<i>FC1</i>	Yes	No	0.769 \pm 0.010
Q234-25-Tr+V	<i>FC2</i>	Yes	No	0.704 \pm 0.008
Q234-25-Tr+V	<i>FC12</i>	Yes	No	0.715 \pm 0.005
Q234-25-Tr+V	<i>FC1</i>	No	Yes	0.641 \pm 0.009
Q234-25-Tr+V	<i>FC2</i>	No	Yes	0.672 \pm 0.012
Q234-25-Tr+V	<i>FC12</i>	No	Yes	0.679 \pm 0.009
Q234-25-Tr+V	<i>FC1</i>	Yes	Yes	0.897 \pm 0.010
Q234-25-Tr+V	<i>FC2</i>	Yes	Yes	0.855 \pm 0.013
Q234-25-Tr+V	<i>FC12</i>	Yes	Yes	0.866 \pm 0.009

Table S7. Sample sizes and face distance (mean and SD) across pairs of females. For comparison, we included all pairs of photographs of the same female, at different ages (¹Maximum difference in age: 6.9 yrs, ²Maximum difference in age: 2.0 yrs).

		Number of pairs of females (pairs of pictures)	Mean face distance (SD)
Juvenile females	PHS	16 (111)	12.89 (2.13)
	MHS	5 (181)	15.80 (1.0)
	NK	27 (1297)	15.07 (2.27)
	Identical females ¹	53 (2087)	10.34 (2.66)
Adult females	PHS	50 (2219)	12.83 (2.17)
	MHS	30 (1564)	12.94 (2.08)
	NK	79 (3209)	13.83 (1.89)
	Identical females ²	18 (2071)	9.64 (3.37)

REFERENCES AND NOTES

1. W. D. Hamilton, The genetical evolution of social behaviour I. *J. Theor. Biol.* **7**, 1–52 (1964).
2. A. S. Griffin, S. A. West, Kin selection: Fact and fiction. *Trends Ecol. Evol.* **17**, 15–21 (2002).
3. T. Clutton-Brock, Breeding together: Kin selection and mutualism in cooperative vertebrates. *Science* **296**, 69–72 (2002).
4. M. J. Sheehan, E. A. Tibbetts, Specialized face learning is associated with individual recognition in paper wasps. *Science* **334**, 1272–1275 (2011).
5. M. J. Sheehan, M. W. Nachman, Morphological and population genomic evidence that human faces have evolved to signal individual identity. *Nat. Comm.* **5**, 4800 (2014).
6. J. B. Cole, M. Manyama, J. R. Larson, D. K. Liberton, T. M. Ferrara, S. L. Riccardi, M. Li, W. Mio, O. D. Klein, S. A. Santorico, B. Hallgrímsson, R. A. Spritz, Human facial shape and size heritability and genetic correlations. *Genetics* **205**, 967–978 (2017).
7. D. Tsagkrasoulis, P. Hysi, T. Spector, G. Montana, Heritability maps of human face morphology through large-scale automated three-dimensional phenotyping. *Sci. Rep.* **7**, 45885 (2017).
8. M. J. E. Charpentier, M. Harté, B. Ngoubangoye, A. Herbert, P. M. Kappeler, Visual discrimination of kin in mandrills. *Ethology* **123**, 251–259 (2017).
9. D. Pfefferle, A. J. N. Kazem, R. R. Brockhausen, A. V. Ruiz-Lambides, A. Widdig, Monkeys spontaneously discriminate their unfamiliar paternal kin under natural conditions using facial cues. *Curr. Biol.* **24**, 1806–1810 (2014).
10. G. Kaminski, S. Dridi, C. Graff, E. Gentaz, Human ability to detect kinship in strangers' faces: Effects of the degree of relatedness. *Proc. Biol. Sci.* **276**, 3193–3200 (2009).
11. A. Alvergne, E. Huchard, D. Caillaud, M. J. E. Charpentier, J. M. Setchell, C. Ruppli, D. Féjan, L. Martinez, G. Cowlshaw, M. Raymond, Human ability to visually recognize kin within primates. *Int. J. Primatol.* **30**, 199–210 (2009).
12. A. J. N. Kazem, A. Widdig, A. Visual phenotype matching: Cues to paternity are present in rhesus macaque faces. *PLOS ONE* **8**, e55846 (2013).
13. A. Widdig, The impact of male reproductive skew on kin structure and sociality in multi-male groups. *Evol. Anthropol.* **22**, 239–250 (2013).
14. M. Charpentier, P. Peignot, M. Hossaert-McKey, O. Gimenez, J. M. Setchell, E. J. Wickings, Constraints on control: Factors influencing reproductive success in male mandrills (*Mandrillus sphinx*). *Behav. Ecol.* **16**, 614–623 (2005).

15. M. J. E. Charpentier, P. Peignot, M. Hossaert-McKey, J. E. Wickings, Kin discrimination in juvenile mandrills *Mandrillus sphinx*. *Anim. Behav.* **73**, 37–45 (2007).
16. S. C. Alberts, Paternal kin discrimination in wild baboons. *Proc. Biol. Sci.* **266**, 1501–1506 (1999).
17. A. Widdig, P. Nürnberg, M. Krawczak, W. J. Streich, F. B. Bercovitch, Paternal relatedness and age-proximity regulate social relationships among adult female rhesus macaques. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 13769–13773 (2001).
18. A. Widdig, Paternal kin discrimination: The evidence and likely mechanisms. *Biol. Rev. Camb. Philos. Soc.* **82**, 319–334 (2007).
19. M. Cords, T. Minich, S.-J. Roberts, C. Sleator, Evidence for paternal kin bias in the social affiliation of adult female blue monkey. *Am. J. Primatol.* **80**, e22761 (2018).
20. O. M. Parkhi, A. Vedaldi, A. Zisserman, Deep face recognition, in *Proceedings of the British Machine Vision Conference (BMVC)*, X. Xie, M. W. Jones, G. K. L. Tam, Eds. (BMVA Press, 2015), pp. 41.1–41.12.
21. Y. Taigman, M. Yang, M. A. Ranzato, L. Wolf, DeepFace: Closing the gap to human-level performance in face verification, in *2014 IEEE Conference on Computer Vision and Pattern Recognition (IEEE, 2014)*, pp. 1701–1708.
22. J. B. Silk, Kin selection in primate groups. *Int. J. Primatol.* **23**, 849–875 (2002).
23. D. Haig, *Genomic Imprinting and Kinship* (Rutgers Univ. Press, 2002).
24. T. Moore, D. Haig, Genomic imprinting in mammalian development: A parental tug-of-war. *Trends Genet.* **7**, 45–49 (1991).
25. D. K. McLain, D. Setters, M. P. Moulton, A. E. Pratt, Ascription of resemblance of newborns by parents and nonrelatives. *Evol. Hum. Behav.* **21**, 11–23 (2000).
26. F. Levréro, G. Carrete-Vega, A. Herbert, I. Lawabi, A. Courtiol, E. Willaume, P. M. Kappeler, M. J. E. Charpentier, Social shaping of voices does not impair phenotype matching of kinship in mandrills. *Nat. Comm.* **6**, 7609 (2015).
27. A. Alvergne, C. Faurie, M. Raymond, Father-offspring resemblance predicts paternal investment in humans. *Anim. Behav.* **78**, 61–69 (2009).
28. M. Heistermann, T. Ziegler, C. P. van Schaik, K. Launhardt, P. Winkler, J. K. Hodges, Loss of oestrus, concealed ovulation and paternity confusion in free-ranging Hanuman langurs. *Proc. Biol. Sci.* **268**, 2445–2451 (2001).

29. K. A. Abernethy, L. J. T. White, E. J. Wickings, Hordes of mandrills (*Mandrillus sphinx*): Extreme group size and seasonal male presence. *J. Zool.* **258**, 131–137 (2002).
30. J. Holt-Lunstad, T. B. Smith, J. B. Layton, Social relationships and mortality risk: A meta-analytic review. *PLOS Med.* **7**, e1000316 (2010).
31. J. B. Silk, S. C. Alberts, J. Altmann, Social bonds of female baboons enhance infant survival. *Science* **302**, 1231–1234 (2003).
32. M. Busson, M. Authier, C. Barbraud, P. Tixier, R. R. Reisinger, A. Janc, C. Guinet, Role of sociality in the response of killer whales to an additive mortality event. *Proc. Natl Acad. Sci. U.S.A.* **116**, 11812–11817 (2019).
33. D. L. Cheney, J. B. Silk, R. M. Seyfarth, Network connections, dyadic bonds and fitness in wild female baboons. *R. Soc. Open Sci.* **3**, 160255 (2016).
34. E. A. Archie, J. Tung, M. Clark, J. Altmann, S. C. Alberts, Social affiliation matters: Both same-sex and opposite-sex relationships predict survival in wild female baboons. *Proc. Biol. Sci.* **281**, 20141261 (2014).
35. M. Y. Akinyi, J. Tung, M. Jeneby, N. B. Patel, J. Altmann, S. C. Alberts, Role of grooming in reducing tick load in wild baboons (*Papio cynocephalus*). *Anim. Behav.* **85**, 559–568 (2013).
36. K. Shutt, A. MacLarnon, M. Heistermann, S. Semple, Grooming in Barbary macaques: Better to give than to receive? *Biol. Lett.* **3**, 231–233 (2007).
37. M. E. Hauber, P. W. Sherman, Self-referent phenotype matching: Theoretical considerations and empirical evidence. *Trends Neurosc.* **24**, 609–616 (2001).
38. M. D. Hauser, J. Kralik, C. Botto-Mahan, M. Garrett, J. Oser, Self-recognition in primates: Phylogeny and the salience of species-typical features. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 10811–10814 (1995).
39. M. Wang, W. Deng, Deep face recognition: A survey. arXiv:1804.06655 [cs.CV] (2018).
40. J. F. H. Cuthill, N. Guttenberg, S. Ledger, R. Crowther, B. Huertas, Deep learning on butterfly phenotypes tests evolution's oldest mathematical model. *Sci. Adv.* **5**, eaaw4967 (2019).
41. P. Peignot, M. J. E. Charpentier, N. Bout, O. Bourry, U. Massima, O. Dosimont, R. Terramorsi, E. J. Wickings, Learning from the first release project of captive-bred mandrills *Mandrillus sphinx* in Gabon. *Oryx* **42**, 122–131 (2008).

42. C. Poirotte, F. Massol, A. Herbert, E. Willaume, P. M. Bomo, P. M. Kappeler, M. J. E. Charpentier, Mandrills use olfaction to socially avoid parasitized conspecifics. *Sci. Adv.* **3**, e1601721 (2017).
43. L. Benoit, S. Mboumba, E. Willaume, P. M. Kappeler, M. J. E. Charpentier, Using next-generation sequencing methods to isolate and characterize 24 simple sequence repeat loci in mandrills (*Mandrillus sphinx*). *Conserv. Genet. Resour.* **6**, 903–905 (2014).
44. J. P. Gutiérrez, F. Goyache, A note on ENDOG: A computer program for analysing pedigree information. *J. Anim. Breed. Genet.* **122**, 172–176 (2005).
45. J. Yosinski, J. Clune, Y. Bengio, H. Lipson. How transferable are features in deep neural networks? in *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, K. Q. Weinberger, Eds. (Curran Associates, Inc., 2014), pp. 3320–3328.
46. D. Masters, C. Luschi, Revisiting small batch training for deep neural networks. arXiv:1804.07612 [cs.LG] (2018).
47. L. Perez, J. Wang, The effectiveness of data augmentation in image classification using deep learning. arXiv:1712.04621 [cs.CV] (2017).
48. R. Zhang, P. Isola, A. A. Efros, E. Shechtman, O. Wang, The unreasonable effectiveness of deep features as a perceptual metric, in *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (IEEE, 2018), pp. 586–595.