Behavior Research Methods
https://doi.org/10.3758/s13428-022-01830-7

Check for updates

Face templates for the Chicago Face Database

Balbir Singh[1] · **Ashleigh Gambrell**[1] · **Joshua Correll**[1]

Accepted: 9 March 2022
© The Psychonomic Society, Inc. 2022

Abstract

Researchers often need to manipulate faces, such as developing a continuum between two faces or averaging a set of faces. In order to do so, researchers use morphing software, but they first need to fit a template to the idiosyncratic landmarks in each face. In this paper, we present a set of landmark templates for the Chicago Face Database (CFD; Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago Face Database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*(4), 1122–1135). The CFD is a free online face database containing images of faces of people from various races and genders. We provide templates for each of 597 neutral (non-expressive) faces in version two of the CFD. Our templates are unique because the facial landmarks were hand placed by researchers. Hand placing facial landmarks allows for more accurate placement of landmarks than a computer-generated template. Historically, hand-placed templates were created by individual labs and not shared. In this paper, we describe how our templates were created, and some possible uses for the templates. We hope that our templates ease the burden for other researchers to manipulate faces.

Keywords Face perception · Social perception · Face manipulation · Morph

The Chicago Face Database (CFD; Lakshmi et al., 2021; Ma et al., 2015, 2020) comprises over 800 faces and has been downloaded thousands of times by researchers around the world. The first two versions of the database contained American faces of different races including Black, White, Latinx, and Asian. The database was recently expanded to include multiracial faces as well as Indian faces photographed in Delhi, India.

Researchers use the CFD for a variety of experiments, which can range from person perception (e.g., Wilson et al., 2017), to interpersonal interactions (e.g., Olivera-La Rosa et al., 2020) to stereotyping and prejudice (e.g., Rees et al., 2019). Depending on the research design and methodology, researchers may want to manipulate the faces they present to participants. Often, manipulation involves averaging faces together (e.g., Ji & Hayward, 2021; Ritchie et al., 2020; Valentine et al., 2004), and morphing/transforming faces (e.g., Beale & Keil, 1995; Walker & Tanaka, 2003).

In order to manipulate a face using morphing software, a set of facial landmarks typically needs to be placed on the image. Combined, these facial landmarks are referred to as a template and allow software to differentiate details of the face. As the number of landmarks placed on the face increases, the specificity of the face approximation increases. Figure 1 shows a face with an applied template. In the template shown, 17 landmarks are used to approximate each eye. Some landmarks outline the edges of the eyelids, while others approximate the shape of the iris and the pupil. These landmarks are numbered so that the same landmark can be placed at the same relative position across faces. For example, suppose landmark 100 indicates the left pupil of a specific face. Landmark 100 will also indicate the left pupil for every other face that is included in the set.

There are two ways a face can be templated. Either the template can be automatically placed by software or the template can be hand-placed. Computer-generated templates are often used by computer scientists (Baddar et al., 2016; Juhong & Pintavirooj, 2017) and have been used in

✉ Balbir Singh
balbir.singh@colorado.edu

1 Department of Psychology and Neuroscience, University of Colorado Boulder, Muenzinger D343D, 345 UCB, Boulder, CO 80309, USA

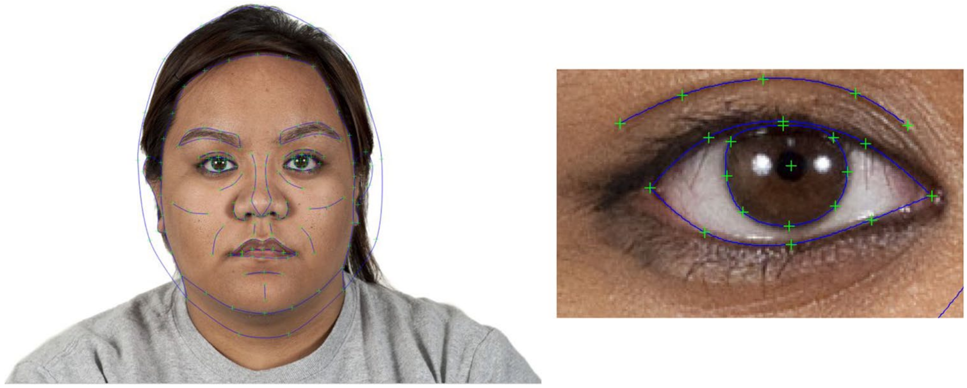

Fig. 1 Example template superimposed onto a face. On the left is an Asian Female face from the CFD with a template superimposed on the face. On the right is a zoomed in view of the model's left eye

psychological research as well (Jones et al., 2021). Dlib is one such machine learning toolkit (King, 2009) that can be used in conjunction with other Python packages to template faces. Unfortunately, computer-generated templates do not always detect facial structures and place landmarks correctly (Jones et al., 2021). For computer scientists who may be looking at patterns that emerge across hundreds or thousands of faces, this measurement error may not pose much of a problem (Baddar et al., 2016). However, errors can create artifacts in averages or morph continua that are problematic for laboratory studies (Scherhag et al., 2019; Venkatesh et al., 2020). Even more troubling, these errors seem to depend on race (and other demographic variables) such that faces of some races are more likely to have errors in their facial landmarks compared to other races (Jones et al., 2021; see Abdurrahim et al., 2018 for a review of errors in automatic face recognition). If researchers are looking to manipulate a specific face, computer-generated templates are less accurate than hand-placed templates.

Hand placing templates is tedious and time-consuming. Templating has often been done by individual labs around the world in order to complete specific projects. These templates are not usually shared (see DeBruine & Jones, 2017 for an exception), and therefore the manipulated faces are not reproducible. It is important to note that there can be ethical and or legal constraints on researchers that can prevent them from being able to share templates. Since templates outline the shape and features of the face, the template itself contains identifying information. Researchers should only share templates if they have permission to share and distribute the models' face.

The Chicago Face Database (CFD) is a freely available and downloadable resource for academics and researchers. To aid researchers who would like to manipulate faces easily and reproducibly, we have built a set of hand-placed templates for all of the neutral faces in version two of the CFD. Each of these templates has been placed by one researcher and checked by a second researcher to ensure consistency within and between faces. In consultation with the authors of the CFD, we are able to make our set of facial landmarks available on the CFD website: https://www.chicagofaces.org/resources/. Researchers can visit the Chicago Faces website, download the CFD, and then navigate to the resources tab to download these templates. Each template is written as a .tem file, which can be used with Psychomorph (Tiddeman et al., 2005) and WebMorph (DeBruine, 2018). These programs natively read .tem files and are free and open access to researchers.

Psychomorph and its online counterpart WebMorph are designed to read in images and template files. These programs can then be used to manipulate faces. To manipulate faces, both Psychomorph and WebMorph calculate a vector of change between two images. For example, let's consider the left pupil landmark. For the first face, the pupil may be 2 mm from the top of the eyelid. On the second face, the pupil may be 4 mm from the top of the eyelid. Psychomorph and WebMorph use this difference along with every other difference from all of the landmarks and the relationships between each landmark to calculate a vector of change from one individual face to the other individual face. This vector can then be applied to either of the faces, which results in a morphed face that sits between each of the two faces (at the percentage of the user's choosing).

Below we describe how the templates were created. We then discuss several potential uses of the templates. Although there may be many more, we believe these examples include common and interesting uses of the templates.

Fig. 2 Hairline examples. Three images illustrating how hairline landmarks were placed on three use cases. (Left), When the top of the forehead and hairline are visible, landmarks were placed at the edge of the hairline. (Middle), When the edge of the hairline is not visible, but hair does not occlude the face, landmarks were placed at the edge of hair. (Right), When the edge of the hairline is not visible and the hair occludes the face, landmarks were placed at the presumed edge of the hairline

Method

We have built templates for all of the neutral faces that are included in version two of the CFD. Version two consists of 597 images of separate models. Each of the templates was created using WebMorph. WebMorph defaults to a 189-landmark template referred to as the "FRL-face" template, which we used for all of the templates[1]. WebMorph first asks the user to place an initial landmark on the left pupil, a second landmark on the right pupil, and a third landmark on the center of the bottom of the upper lip. After the first three landmarks are placed on the face, an initial rough template is automatically mapped onto the face. The generated template is a rough estimation of the facial features and is generally a relatively poor representation of the face. The researcher working on the template must move each landmark to its correct place on the face.

The template outlines the following features of the faces: Fifteen landmarks estimate the outer hairline as well the throat/larynx. Thirty landmarks estimate the inner hairline, jaw, and overall shape of the face. Five landmarks estimate each ear lobe. Inside the face, each structure of the face is outlined by landmarks: right and left eyebrows ($k = 8$ each), right and left eye crease ($k = 5$ each), right and left eye outline ($k = 8$ each), right and left iris ($k = 8$ each), right and left pupil ($k = 1$ each), right and left circle under eye ($k = 3$ each), right and left cheekbone ($k = 3$ each), nose ($k = 11$), right and left nostril ($k = 7$ each), right and left smile line ($k = 3$ each), right and left philtrum ($k = 2$ each), outer lips ($k = 12$), upper and bottom lip ($k = 5$ each), crease between the bottom lip and chin ($k = 3$), and the chin dimple ($k = 2$).

[1] If a user would like to use the 189-point FRL-face template, it can be accessed on WebMorph.org. This template is a creation of Lisa Debruine, PhD.

Procedure

Training researchers

Four researchers templated faces and each of them was trained by the first author. The trainer would first template a face with the trainee present and then would ask the trainee to template a face on their own. They would then discuss any changes and continue this process until the templating style matched the trainer's. Researchers templated one race and one gender at a time. All landmark templates were initially placed by the trained researcher. Once the landmark template was completed, either the first or second author reviewed the completed template to ensure consistency. Following the initial review of this manuscript, templates were again reviewed by the first author to ensure consistency with the templating style and the FRL-face template.

Template procedure

Each image from the CFD was uploaded into WebMorph. Images were not manipulated or edited in any way prior to upload. Once the image was uploaded, a researcher applied the 189-landmark default FRL-face template and then moved the landmarks to match the face structure.

It is important to note that some of the landmark placements require subjective judgment. For example, the smile line represents the nasolabial fold. This fold can range from the outer nostrils to the lip. With only three landmarks to estimate this fold, researchers need to decide where to place each landmark. However, researchers made every effort to be consistent both within faces (placing the left and right smile line landmarks in the same relative places) and between faces.

The inner hairline deserves special mention. The hairline of faces fell into three categories, in some faces the hairline was distinct and there was little to no hair obstructing the hairline. In other faces, hair partially obstructed the hairline but most of the forehead was observable. Lastly, some faces

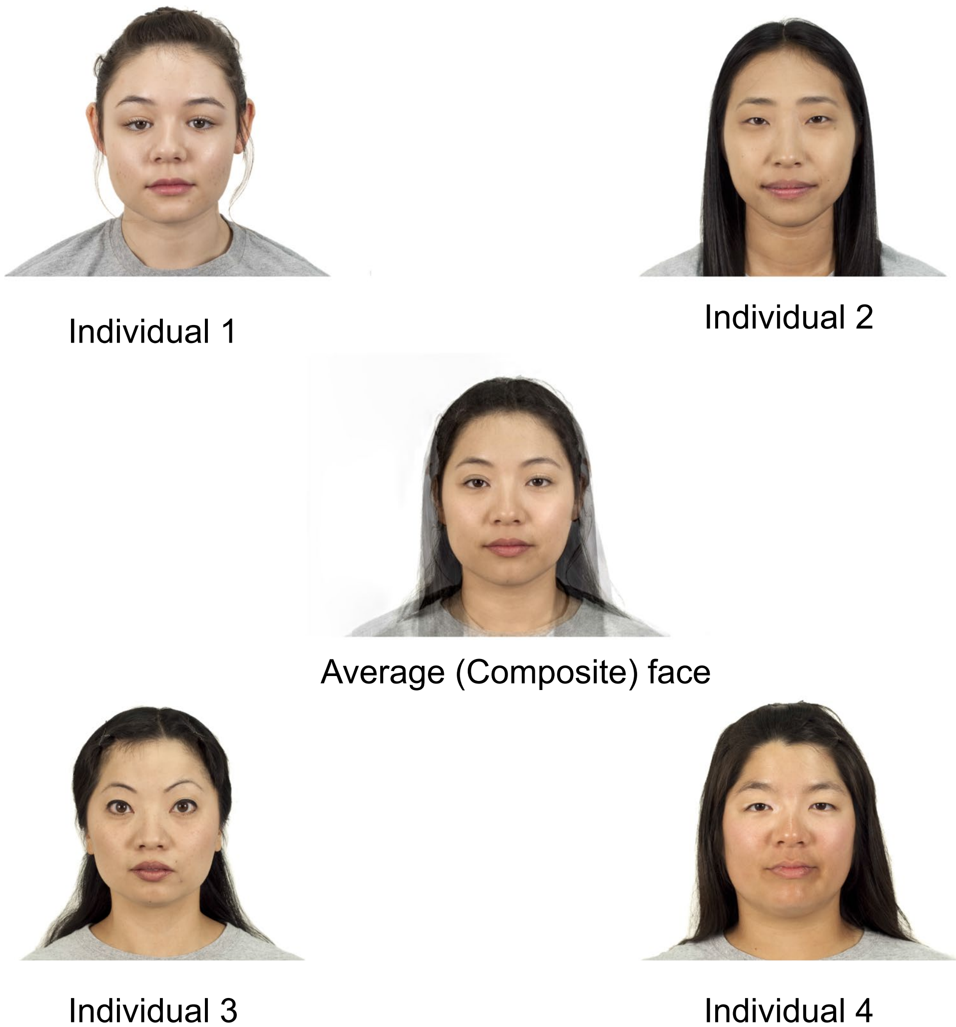

Fig. 3 Example average face. The center face is the average (determined by WebMorph) of the surrounding four faces. The shape, color, and texture have all been modified to the same degree

either did not have a distinct hairline (bald heads) or had a lot of hair that swept across the forehead so that the hairline was not visible. The researchers used the following guidelines. When the hairline was easily discernible, the inner hairline landmarks were mapped to the edge of the hairline. When there was a small amount of hair on the forehead, the researcher mapped the landmarks to the edge between the hair and the visible forehead. Lastly, when the hairline was indiscernible, the inner hairline landmarks were placed onto the occluding hair, based on the researcher's estimate of where the hairline would be, see Fig. 2[2].

[2] This hairline decision is optimal for some image transformations but not others. If two faces are being morphed and the researcher would like to produce a realistic face, then these templates should work for the researcher. If a researcher plans to use the templates in other ways (examining sexual dimorphism) than modifications to the templates may be required.

Possible template uses

In this section, we will describe a few possible uses of the templates. This is not an exhaustive list. Researchers may find novel and creative ways to use these templates (see Sutherland et al., 2017 for a review of face manipulation).

Averaging faces

One of the most common uses of morphing software involves averaging a set of faces (see Fig. 3 for an example). Once a user has a set of images that are templated, it is trivial to create an average in Psychomorph and WebMorph. Users simply denote which faces they want included in the average, and the software generates the image. The details on how the averaging function works are available online on WebMorph and Psychomorph (DeBruine, 2021; Sutherland, 2015).

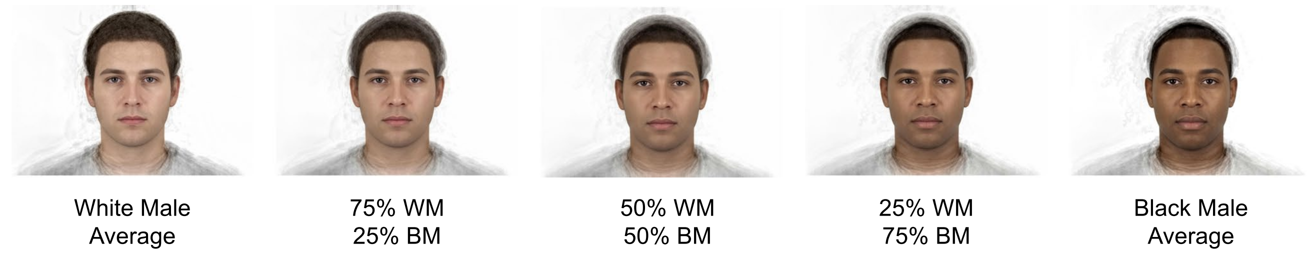

Fig. 4 Example morph continua. The face on the left is an average White Male face created by a set of CFD faces. The face on the left is an average Black Male face created by a set of CFD faces. These two average faces were then morphed. The percentages refer to how much the shape, color and texture was adjusted from one end of the continua to the other. The resulting continua is shown

Averaged faces can be used for a multitude of experiments. For example, face averages can be used to assess attractiveness (e.g., Valentine et al., 2004), category representations (e.g., Lloyd et al., 2020; Mangini & Biederman, 2004), the presentation of depression in the face (Scott et al., 2013), as well as how individual identities or groups are encoded (Davis et al., 2020; Ji & Hayward, 2021, respectively).

Morph continua

Another common manipulation involves generating a continuum of images that gradually transition between two faces. This is done by morphing two individual faces to create a new face that blends the two individual faces to different degrees (see Fig. 4 for an example, which uses average faces as the endpoints of the continuum). WebMorph and Psychomorph allow the researcher to individually control the shape, texture, and color of the morph. In short, these three variables allow the user to fine tune the morph to increase believability (see Tiddeman et al., 2001, 2005). This type of morphing has been used many times in the literature. From investigating the detection and recognition of emotion (Corneille et al., 2007), race (Levin & Angelone, 2002; Walker & Hewstone, 2006; Walker & Tanaka, 2003), and identity (Beale & Keil, 1995; Campanella et al., 2003; Levin & Beale, 2000), the possibilities are extensive.

Transforming (morphing a third face)

One of the reasons we chose to build these templates using Psychomorph and WebMorph, rather than another software package, was the unique way in which these two programs handle face manipulation. Psychomorph and WebMorph give the user the flexibility to employ three source faces in the manipulating process, rather than just two individuals. The first two individual faces are used to define the vector of change. The user can then apply that vector to transform a third face (e.g., DeBruine, 2005). This process is "an extension of morphing" referred to as *transforming* (Sutherland et al., 2017, p. 539). For example, a researcher might first create an average Black male face and an average White male face and define the vector of change between those two faces. Since that vector describes the difference between the averages, it should approximate the difference between a typical Black male face and a typical White male face. The researcher can then adjust a third face, such as a biracial (Black & White) face, adjusting it either in the direction of the Black-average pole of the vector or the White-average pole. This process effectively transforms the biracial face to appear more prototypically Black or prototypically White (see Fig. 5).

Using templates to measure faces

Lastly, templates can be used to measure facial features (e.g., Cai et al., 2019; Holzleitner et al., 2014). As previously discussed, the templates consist of landmarks that correspond to critical points, like the pupils of the right and left eyes or the tip of the nose, in a two-dimensional representation of the face. Each template consists of a list of (x, y) coordinates, and researchers can use those coordinates to measure the distances between landmarks.

Without the template, researchers could certainly manually measure the distance between any two features of the face. In fact, the CFD includes roughly 30 metrics of facial structure for each neutral face, including the distance between the pupils, the width of the face, the length of the nose, etc., which were measured in Adobe Photoshop. This process is tedious and time-consuming. On the other hand, using the templates we have developed, researchers can easily compute the distance between any pair of landmarks. With 189 individual landmarks, these templates easily allow a researcher to derive more than 17,000 possible measures of facial structure. Measurement could, of course, be automated using a program like R or Python or SPSS, to find the pupil-to-pupil distance (or any other distance) for all of

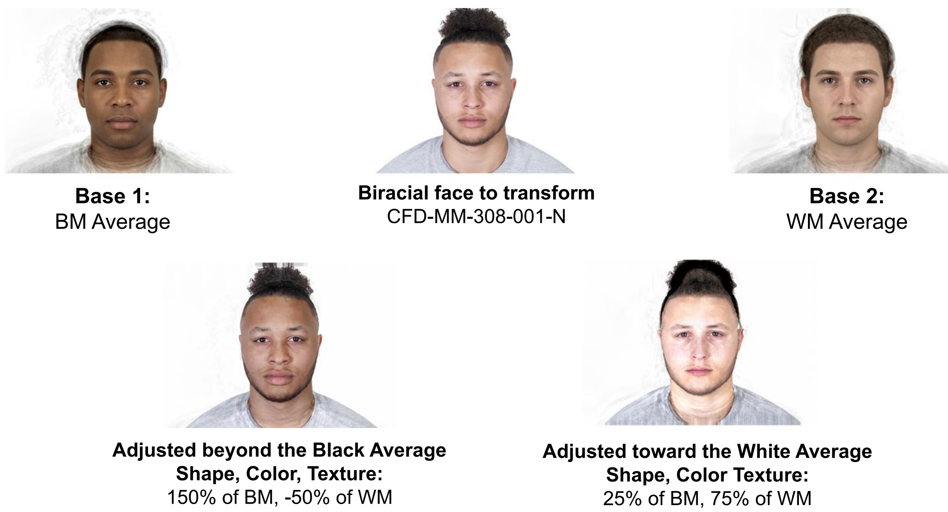

Fig. 5 Applying a vector of change to a distinct third face (transforming). The above example shows a race transformation. Base 1 is a composite of the Black Male faces in the CFD. Base 2 is a composite of the White Male faces in the CFD. The face to be transformed is a Biracial model with Black and White ancestry. We calculated the vector of change between each of the averages and applied it to the Biracial face. This allows us to transform the biracial model to a "Black" version of his face and a "White" version of his face. The shape, color and texture have all been modified to the same degree

the faces in the CFD. This method of measuring features of the face can be extended to configurations within the face, or global face measurements such as width to height ratios, shapes of the nose, and many others. To aid in this process, we created a single CSV file that includes the x,y coordinates of all 189 landmarks for all of the faces in CFD.

Conclusions

We believe these templates are a valuable resource to the community of researchers who use faces in their research. These templates should aid in manipulating CFD faces and may stimulate novel research questions and potentially novel methodologies. We hope that by making templates available for a commonly used face database, there will be a reduction of duplicative work. It is important to note that this reduction does come with a drawback. These templates all conform to the particular templating style adopted by our laboratory, which may result in certain idiosyncrasies. Nonetheless, we hope that the provided templates ease some of the barriers to face manipulation and face research more broadly.

Acknowledgements We would like to thank the entire Correll Lab at the University of Colorado Boulder for their assistance on this project.

Availability of data and materials The templates are available at https://www.chicagofaces.org/resources/

Code availability Not applicable.

Funding This project has not received any financial support.

Declarations

Conflicts of interest There are no conflicts of interest to report.

Ethics approval Not applicable.

Consent to participate Not applicable.

References

Abdurrahim, S. H., Samad, S. A., & Huddin, A. B. (2018). Review on the effects of age, gender, and race demographics on automatic face recognition. *The Visual Computer, 34*(11), 1617–1630. https://doi.org/10.1007/s00371-017-1428-z

Baddar, W. J., Son, J., Kim, D. H., Kim, S. T., & Ro, Y. M. (2016). A deep facial landmarks detection with facial contour and facial components constraint. *IEEE International Conference on Image Processing (ICIP), 2016*, 3209–3213. https://doi.org/10.1109/ICIP.2016.7532952

Beale, J. M., & Keil, F. C. (1995). Categorical effects in the perception of faces. *Cognition, 57*(3), 217–239. https://doi.org/10.1016/0010-0277(95)00669-X

Cai, Z., Hahn, A. C., Zhang, W., Holzleitner, I. J., Lee, A. J., DeBruine, L. M., & Jones, B. C. (2019). No evidence that facial attractiveness, femininity, averageness, or coloration are cues to susceptibility to infectious illnesses in a university sample of young adult women. *Evolution and Human Behavior, 40*(2), 156–159. https://doi.org/10.1016/j.evolhumbehav.2018.10.002

Campanella, S., Hanoteau, C., Seron, X., Joassin, F., & Bruyer, R. (2003). Categorical perception of unfamiliar facial identities, the face-space metaphor, and the morphing technique. *Visual Cognition, 10*(2), 129–156. https://doi.org/10.1080/713756676

Corneille, O., Hugenberg, K., & Potter, T. (2007). Applying the attractor field model to social cognition: Perceptual discrimination is facilitated, but memory is impaired for faces displaying evaluatively congruent expressions. *Journal of Personality and Social Psychology, 93*(3), 335–352. https://doi.org/10.1037/0022-3514.93.3.335

Davis, E. E., Matthews, C. M., & Mondloch, C. J. (2020). Ensemble coding of facial identity is not refined by experience: Evidence from other-race and inverted faces. *British Journal of Psychology, bjop.12457*. https://doi.org/10.1111/bjop.12457

DeBruine, L. M. (2005). Trustworthy but not lust-worthy: Context-specific effects of facial resemblance. *Proceedings of the Royal Society B: Biological Sciences, 272*(1566), 919–922. https://doi.org/10.1098/rspb.2004.3003

DeBruine, L. (2018). *Debruine/WebMorph: Beta Release 2*. Zenodo. https://doi.org/10.5281/ZENODO.1162670

DeBruine, L. (2021). *WebMorph manual*. https://debruine.github.io/webmorph/

DeBruine, L., & Jones, B. (2017). *Face research lab London set* (p. 281699312 Bytes) [Data set]. figshare. https://doi.org/10.6084/M9.FIGSHARE.5047666.V3

Holzleitner, I. J., Hunter, D. W., Tiddeman, B. P., Seck, A., Re, D. E., & Perrett, D. I. (2014). Men's facial masculinity: When (body) size matters. *Perception, 43*(11), 1191–1202. https://doi.org/10.1068/p7673

Ji, L., & Hayward, W. G. (2021). Metacognition of average face perception. *Attention, Perception, & Psychophysics, 83*(3), 1036–1048. https://doi.org/10.3758/s13414-020-02189-7

Jones, A. L., Schild, C., & Jones, B. C. (2021). Facial metrics generated from manually and automatically placed image landmarks are highly correlated. *Evolution and Human Behavior, 42*(3), 186–193. https://doi.org/10.1016/j.evolhumbehav.2020.09.002

Juhong, A., & Pintavirooj, C. (2017). Face recognition based on facial landmark detection. *2017 10th Biomedical Engineering International Conference (BMEiCON)*, 1–4. https://doi.org/10.1109/BMEiCON.2017.8229173

King, D. E. (2009). Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research, 10*, 1755–1758.

Lakshmi, A., Wittenbrink, B., Correll, J., & Ma, D. S. (2021). The India face set: International and cultural boundaries impact face impressions and perceptions of category, membership. *Frontiers in Psychology, 12*, 627678. https://doi.org/10.3389/fpsyg.2021.627678

Levin, D. T., & Angelone, B. L. (2002). Categorical perception of race. *Perception, 31*(5), 567–578. https://doi.org/10.1068/p3315

Levin, D. T., & Beale, J. M. (2000). Categorical perception occurs in newly learned faces, other-race faces, and inverted faces. *Perception & Psychophysics, 62*(2), 386–401. https://doi.org/10.3758/BF03205558

Lloyd, E. P., Sim, M., Smalley, E., Bernstein, M. J., & Hugenberg, K. (2020). Good cop, bad cop: Race-based differences in mental representations of police. *Personality and Social Psychology Bulletin, 46*(8), 1205–1218. https://doi.org/10.1177/0146167219898562

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago Face Database: A free stimulus set of faces and norming data. *Behavior Research Methods, 47*(4), 1122–1135. https://doi.org/10.3758/s13428-014-0532-5

Ma, D. S., Kantner, J., & Wittenbrink, B. (2020). Chicago Face Database: Multiracial expansion. *Behavior Research Methods*. https://doi.org/10.3758/s13428-020-01482-5

Mangini, M. C., & Biederman, I. (2004). Making the ineffable explicit: Estimating the information employed for face classifications. *Cognitive Science, 28*(2), 209–226. https://doi.org/10.1207/s15516709cog2802_4

Olivera-La Rosa, A., Chuquichambi, E. G., & Ingram, G. P. D. (2020). Keep your (social) distance: Pathogen concerns and social perception in the time of COVID-19. *Personality and Individual Differences, 166*, 110200. https://doi.org/10.1016/j.paid.2020.110200

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rees, H. R., Rivers, A. M., & Sherman, J. W. (2019). Implementation intentions reduce implicit stereotype activation and application. *Personality and Social Psychology Bulletin, 45*(1), 37–53. https://doi.org/10.1177/0146167218775695

Ritchie, K. L., Mireku, M. O., & Kramer, R. S. S. (2020). Face averages and multiple images in a live matching task. *British Journal of Psychology, 111*(1), 92–102. https://doi.org/10.1111/bjop.12388

Scherhag, U., Rathgeb, C., Merkle, J., Breithaupt, R., & Busch, C. (2019). Face recognition systems under morphing attacks: A survey. *IEEE Access, 7*, 23012–23026. https://doi.org/10.1109/ACCESS.2019.2899367

Scott, N. J., Kramer, R. S. S., Jones, A. L., & Ward, R. (2013). Facial cues to depressive symptoms and their associated personality attributions. *Psychiatry Research, 208*(1), 47–53. https://doi.org/10.1016/j.psychres.2013.02.027

Sutherland, C. (2015). *A basic guide to Psychomorph*. University of York.

Sutherland, C., Rhodes, G., & Young, A. W. (2017). Facial image manipulation: A tool for investigating social perception. *Social Psychological and Personality Science, 8*(5), 538–551. https://doi.org/10.1177/1948550617697176

Tiddeman, B. P., Burt, D. M., & Perrett, D. I. (2001). Computer graphics in facial perception research. *IEEE Computer Graphics and Applications, 21*(5), 42–50.

Tiddeman, B. P., Stirrat, M. R., & Perrett, D. I. (2005). Towards realism in facial image transformation: Results of a wavelet MRF method. *Computer Graphics Forum, 24*(3), 449–456. https://doi.org/10.1111/j.1467-8659.2005.00870.x

Valentine, T., Darling, S., & Donnelly, M. (2004). Why are average faces attractive? The effect of view and averageness on the attractiveness of female faces. *Psychonomic Bulletin & Review, 11*(3), 482–487. https://doi.org/10.3758/BF03196599

Venkatesh, S., Zhang, H., Ramachandra, R., Raja, K., Damer, N., & Busch, C. (2020). Can GAN generated morphs threaten face recognition systems equally as landmark based morphs? - Vulnerability and detection. *2020 8th International Workshop on Biometrics and Forensics (IWBF)*, 1–6. https://doi.org/10.1109/IWBF49977.2020.9107970

Walker, P. M., & Hewstone, M. (2006). A perceptual discrimination investigation of the own-race effect and intergroup experience. *Applied Cognitive Psychology, 20*(4), 461–475. https://doi.org/10.1002/acp.1191

Walker, P. M., & Tanaka, J. W. (2003). An encoding advantage for own-race versus other-race faces. *Perception, 32*(9), 1117–1125. https://doi.org/10.1068/p5098

Wilson, J. P., Hugenberg, K., & Rule, N. O. (2017). Racial bias in judgments of physical size and formidability: From size to threat. *Journal of Personality and Social Psychology, 113*(1), 59–80. https://doi.org/10.1037/pspi0000092

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.