THIS IS A PREPRINT OF THE PEER REVIEWED ARTICLE TO APPEAR IN JOURNAL OF PERSONALITY AND SOCIAL PSYCHOLOGY. THE MOST RECENT VERSION IS AVAILABLE AT https://osf.io/zn79k/

AUTHOR NOTES ARE AVAILABLE AT: https://goo.gl/9b2aR2

Deep neural networks are more accurate than humans at detecting sexual orientation from facial images

Yilun Wang, Michal Kosinski

Graduate School of Business, Stanford University, Stanford, CA94305, USA

michalk@stanford.edu

Author Note:

YW and MK collected the data and conducted the analysis; MK wrote the paper.

Abstract

We show that faces contain much more information about sexual orientation than can be perceived and interpreted by the human brain. We used deep neural networks to extract features from 35,326 facial images. These features were entered into a logistic regression aimed at classifying sexual orientation. Given a single facial image, a classifier could correctly distinguish between gay and heterosexual men in 81% of cases, and in 71% of cases for women. Human judges achieved much lower accuracy: 61% for men and 54% for women. The accuracy of the algorithm increased to 91% and 83%, respectively, given five facial images per person. Facial features employed by the classifier included both fixed (e.g., nose shape) and transient facial features (e.g., grooming style). Consistent with the prenatal hormone theory of sexual orientation, gay men and women tended to have gender-atypical facial morphology, expression, and grooming styles. Prediction models aimed at gender alone allowed for detecting gay males with 57% accuracy and gay females with 58% accuracy. Those findings advance our understanding of the origins of sexual orientation and the limits of human perception. Additionally, given that companies and governments are increasingly using computer vision algorithms to detect people's intimate traits, our findings expose a threat to the privacy and safety of gay men and women.

Keywords: sexual orientation, face, facial morphology, prenatal hormone theory, computational social science, big data, privacy, artificial intelligence

Deep neural networks are more accurate than humans at detecting sexual orientation from facial images

The science of judging one's character from their facial characteristics, or physiognomy, dates back to ancient China and Greece (Jenkinson, 1997). Aristotle and Pythagoras were among its disciples, and the latter used to select his students based on their facial features (Riedweg, 2005). Such beliefs have persisted and grown in popularity over the centuries. Robert FitzRoy, the captain of the Beagle, believed that Darwin's nose revealed a lack of energy and determination, and was close to barring him from the historic voyage (Glaser, 2002). Cesare Lombroso, the founder of criminal anthropology, believed that criminals could be identified by their facial features. He claimed, for example, that arsonists have a "softness of skin, an almost childlike appearance, and an abundance of thick straight hair that is almost feminine" (Lombroso, 1911, p. 51). By the eighteenth century, physiognomy "was not merely a popular fad but also the subject of intense academic debate about the promises it held for future progress" (Porter, 2003, p. 497).

Physiognomy is now universally, and rightly, rejected as a mix of superstition and racism disguised as science (Jenkinson, 1997). Due to its legacy, studying or even discussing the links between facial features and character became taboo, leading to a widespread presumption that no such links exist. However, there are many demonstrated mechanisms that imply the opposite. Such mechanisms can be arranged into three groups. First, there is much evidence that character can influence one's facial appearance (e.g., Lõhmus, Sundström, & Björklund, 2009; Zebrowitz & Collins, 1997). For example, women that scored high on extroversion early in life tend to become more attractive with age (Zebrowitz, Collins, & Dutta, 1998). Second, facial appearance can alter one's character. Facial appearance drives first impressions of others, influencing our expectations and behavior toward them, which, in turn, shapes their character (Berry, 1991; Berry & Brownlow, 1989; Penton-Voak, Pound, Little, & Perrett, 2006; Todorov, Said, Engell, & Oosterhof, 2008; Zebrowitz & Collins, 1997; Zebrowitz et al., 1998). Good-looking people, for example, receive more positive social feedback, and thus tend to become even more extroverted (Lukaszewski & Roney, 2011). Finally, there is a broad range of factors affecting both facial appearance and one's traits. Those include pre- and post-natal hormonal levels (Jones et al., 2015; Lefevre, Lewis, Perrett, & Penke, 2013; Whitehouse et al., 2015), developmental history (Astley, Stachowiak, Clarren, & Clausen, 2002), environmental factors, and gene expression (Ferry et al., 2014). Testosterone levels, for instance, significantly affect both: behavior (e.g., dominance) and facial appearance (e.g., facial-width-to-height-ratio; Lefevre et al., 2014).

The existence of such links between facial appearance and character is supported by the fact that people can accurately judge others' character, psychological states, and demographic traits from their faces (Zebrowitz, 1997). For example, we can easily and accurately identify others' gender, age, race, or emotional state—even from a glimpse of their faces (Brown & Perrett, 1993; Macrae & Bodenhausen, 2000; Roberts & Bruce, 1988). People also judge, with some minimal accuracy, others' political views (e.g., Rule & Ambady, 2010; Samochowiec, Wänke, & Fiedler, 2010), honesty (e.g., Bond, Berry, & Omar, 1994), personality (e.g., Borkenau, Brecke, Möttig, & Paelecke, 2009), sexual orientation (e.g., Rule & Ambady, 2008), or even the likelihood of winning an election (e.g., Ballew & Todorov, 2007; Little, Burriss, Jones, & Roberts, 2007; Todorov, Mandisodza, Goren, & Hall, 2005). Such judgments are not very accurate, but are common and spontaneous. Importantly, the low accuracy of humans when judging character from others' faces does not necessarily mean that relevant cues are not prominently displayed. Instead, people may lack the ability to detect or interpret them. It is possible that some of our intimate traits are prominently displayed on the face, even if others cannot perceive them. Here, we test this hypothesis using modern computer vision algorithms.

Recent progress in AI and computer vision has been largely driven by the widespread adoption of deep neural networks (DNN), or neural networks composed of a large number of hidden layers (LeCun, Bengio, & Hinton, 2015). DNNs mimic the neocortex by simulating large, multi-level networks of interconnected neurons. DNNs excel at recognizing patterns in large, unstructured data such as digital images, sound, or text, and analyzing such patterns to make predictions. DNNs are increasingly outperforming humans in visual tasks such as image classification, facial recognition, or diagnosing skin cancer (Esteva et al., 2017; LeCun et al., 2015; Lu & Tang, 2014). The superior performance of DNNs offers an opportunity to identify links between characteristics and facial features that might be missed or misinterpreted by the human brain.

We tested our hypothesis on a specific intimate trait: sexual orientation. We chose this trait for three main reasons. First, it is an intimate psycho–demographic trait that cannot be easily detected by others. While people can detect others' sexual orientation from both neutral and expressive faces (Rule & Ambady, 2008; Tskhay & Rule, 2015), or even from a single facial feature such as the mouth, eyes, or hair (Lyons, Lynch, Brewer, & Bruno, 2014; Rule, MacRae, & Ambady, 2009), the accuracy of such judgments is very limited, ranging from 55 to 65% (Ambady, Hallahan, & Conner, 1999; Lyons et al., 2014; Rule et al., 2009). The links between facial features and sexual orientation, however, may be stronger than what meets the human eye.

Recent evidence shows that gay men and lesbians,[1] who arguably have more experience and motivation to detect the sexual orientation of others, are marginally more accurate than heterosexuals (Brambilla, Riva, & Rule, 2013).

Second, the widely accepted prenatal hormone theory (PHT) of sexual orientation predicts the existence of links between facial appearance and sexual orientation. According to the PHT, same-gender sexual orientation stems from the underexposure of male fetuses or overexposure of female fetuses to androgens that are responsible for sexual differentiation (Allen & Gorski, 1992; Jannini, Blanchard, Camperio-Ciani, & Bancroft, 2010; Udry & Chantala, 2006). As the same androgens are responsible for the sexual dimorphism of the face, the PHT predicts that gay people will tend to have gender-atypical facial morphology (Bulygina, Mitteroecker, & Aiello, 2006; Rhodes, 2006; Whitehouse et al., 2015). According to the PHT, gay men should tend to have more feminine facial features than heterosexual men, while lesbians should tend to have more masculine features than heterosexual women. Thus, gay men are predicted to have smaller jaws and chins, slimmer eyebrows, longer noses, and larger foreheads; the opposite should be true for lesbians. Furthermore, as prenatal androgen levels also drive the sexual differentiation of behaviors and preferences during adulthood (Meyer-Bahlburg, 1984; Udry, 2000), the PHT predicts that gay people may tend to adopt gender-atypical facial adornments, expressions, and grooming styles. Such gender-atypical behaviors and preferences

[1] Following the APA's recommendation, the term "gay" is used to refer to same-gender sexual orientation.

might also be encoded in gay culture, further amplifying the effect of the prenatal androgen levels.

Previous empirical evidence provides mixed support for the gender typicality of facial features of gay men and women. Huges and Bremme (2011) studied a sample of 60 images of gay men and concluded that gay men had, on average, more feminine facial features. Lyons et al. (2014) asked 120 human judges to rate the masculinity and femininity of 80 faces of men and women. They found that on average, heterosexual women and gay men were rated as more feminine and less masculine than lesbians and heterosexual men. However, Skorska, Geniole, Vrysen, McCormick, and Bogaert (2015) used a sample of 390 photographs of men and women, and found that both lesbians and gay men had more masculine faces than heterosexual women and men, respectively. Valentova, Kleisner, Havlíček, and Neustupa (2014, p. 353) used a sample of facial images of 40 gay and 40 heterosexual men, and found that on average, gay men had relatively wider and shorter faces, smaller and shorter noses, and larger and more rounded jaws, or “a mosaic of both feminine and masculine features.” Such mixed findings might be attributed to the difficulty of precisely defining and measuring facial femininity. They might also be attributed to the fact that the difference between gay and heterosexual faces may be too subtle to be reliably detected in the small samples employed in these studies. This study aims to address those limitations by using a much larger sample size and data-driven methods, including an algorithm-based measure of facial femininity.

Finally, the predictability of sexual orientation could have serious and even life-threatening implications to gay men and women and the society as a whole. In some cultures, gay men and women still suffer physical and psychological abuse at the hands of governments, neighbors, and even their own families. Perhaps due to discrimination and stigmatization, gay people are also at a higher risk of depression, suicide, self-harm, and substance abuse (King et al., 2008). Consequently, their well-being and safety may depend on their ability to control when and to whom to reveal their sexual orientation. Press reports suggest that governments and corporations are developing and deploying face-based prediction tools aimed at intimate psycho–demographic traits, such as the likelihood of committing a crime, or being a terrorist or pedophile (Chin & Lin, 2017; Lubin, 2016). The laws in many countries criminalize same-gender sexual behavior, and in eight countries—including Iran, Mauritania, Saudi Arabia, and Yemen—it is punishable by death (UN Human Rights Council, 2015). It is thus critical to inform policymakers, technology companies and, most importantly, the gay community, of how accurate face-based predictions might be.

This work examines whether an intimate psycho–demographic trait, sexual orientation, is displayed on human faces beyond what can be perceived by humans. We address this question using a data-driven approach. A DNN was used to extract features from the facial images of 35,326 gay and heterosexual men and women. These features were entered (separately for each gender) as independent variables into a cross-validated logistic regression model aimed at predicting self-reported sexual orientation. The resulting classification accuracy offers a proxy for the amount of information relevant to the sexual orientation displayed on human faces. We also explore the features employed by the classifier and examine whether, as predicted by the PHT, the faces of gay men and women tend to be gender atypical. Furthermore, we compare the accuracy of the computer algorithm with that of human judges. Human accuracy does not only provide a baseline for interpreting the algorithm's accuracy, but it also helps to examine whether the nonstandardized facial images used here are not more revealing of sexual orientation than standardized facial images taken in a controlled environment. Finally, using an independent sample of gay men's facial images, we test the external predictive validity of the classifier developed here.

Study 1a: Using Deep Neural Network to Detect Sexual Orientation

In Study 1a, we show that a DNN can be used to identify sexual orientation from facial images. Previous studies linking facial features with sexual orientation used either images of neutral[2] faces taken in a laboratory (e.g., Skorska et al., 2015; Valentova et al., 2014) or self-taken images obtained from online dating websites (e.g., Hughes & Bremme, 2011; Lyons et al., 2014; Rule & Ambady, 2008; Rule, Ambady, Adams, & Macrae, 2008). We employed the latter approach, as such images can be collected in large numbers, from more representative samples, and at a lower cost (from the perspective of both the participants and researchers). Larger and more representative samples, in turn, enable the discovery of phenomena that might not have been apparent in the smaller, lab-based samples. Additionally, using self-taken, easily accessible digital facial images increases the ecological validity of our results, which is particularly important given their critical privacy implications.

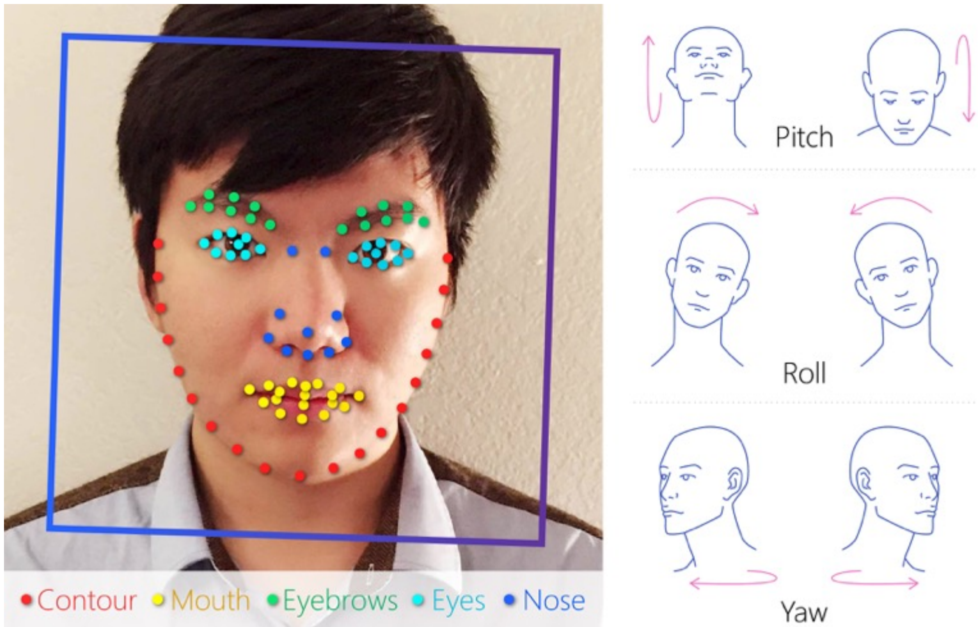
Images taken and uploaded by the participants have a number of potential drawbacks. They may vary in quality, facial expression, head orientation, and background. Furthermore,

[2] We believe that no face can be truly "neutral." People may systematically differ in the expression that they adopt when instructed to "adopt a neutral expression." Furthermore, even an image of a perfectly neutral face (e.g., taken under anesthesia) would still contain traces of typically adopted expressions (e.g., laugh lines), grooming style (e.g., skin health), and one's environment (e.g., tan).

given that they were originally posted on a dating website, they might be especially revealing of sexual orientation. We take several steps to mitigate the influence of such factors. First, the facial features are extracted using a DNN that was specifically developed to focus on non-transient facial features, disregarding the head's orientation and the background. Second, Study 1b investigates the areas of the face employed by the classifier and shows that the classifier focuses on the face and does not rely on the background. Third, Studies 1c and 2 explore the facial features used by the classifier and shows that they are consistent with the theory (PHT). Fourth, Studies 3 and 4 show that the images used here were not substantially more revealing of sexual orientation than images of neutral faces taken in a controlled setting or images obtained from Facebook.

Methods

Facial images. We obtained facial images from public profiles posted on a U.S. dating website. We recorded 130,741 images of 36,630 men and 170,360 images of 38,593 women between the ages of 18 and 40, who reported their location as the U.S. Gay and heterosexual people were represented in equal numbers. Their sexual orientation was established based on the gender of the partners that they were looking for (according to their profiles).



A B

Figure 1. Graphical illustration of the outcome produced by Face++. Panel A illustrates facial landmarks (colored dots, n=83) and facial frame (blue box). Panel B illustrates pitch, roll, and yaw parameters that describe the head's orientation in space.

The location of the face in the image, outlines of its elements, and the head's orientation were extracted using a widely used face-detection software: Face++.[3] Figure 1 shows the output of Face++ in a graphical format. The colored dots (Panel A) indicate the location of the facial landmarks outlining the contour and elements of the face. Additionally, Face++ provided the estimates of the head's yaw, pitch, and roll (Panel B).

Based on the Face++ results, we removed images containing multiple faces, partially hidden faces (i.e., with one or more landmarks missing), and overly small faces (i.e., where the

---

[3] Face++ can be accessed at http://www.faceplusplus.com.

distance between the eyes was below 40 pixels). We also removed faces that were not facing the camera directly (i.e., with a yaw greater than 15 degrees and a pitch greater than 10 degrees).

Table 1

Frequencies of Users and Facial Images, and the Age Distribution in the Final Sample Used in Study 1

	Men		Women	
	Gay	Heterosexual	Lesbian	Heterosexual
Unique users	3,947	3,947	3,441	3,441
Median age (IQR)	33 (30–36)	33 (30–36)	29 (25–34)	29 (25–34)
Total images	8,996	8,645	7,457	10,228
Users with at least:				
1 image	3,947	3,947	3,441	3,441
2 images	2,438	2,439	2,878	2,037
3 images	1,363	1,367	1,951	1,058
4 images	562	731	1,114	494
5 images	219	327	491	223

Note. IQR stands for interquartile range.

Next, we employed Amazon Mechanical Turk (AMT) workers to verify that the faces were adult, Caucasian, fully visible, and of a gender that matched the one reported on the user's profile. We limited the task to the workers from the U.S., who had previously completed at least 1,000 tasks and obtained an approval rate of at least 98%. Only faces approved by four out of six workers were retained. See Figure S1 for the instructions presented to the workers.

Finally, we randomly removed some users to balance the age distribution of the sexual orientation subsamples and their size—separately for each gender. The final sample contained 35,326 facial images of 14,776 gay and heterosexual (50/50%) men and women (53/47%; see

Table 1 for details). Facial images were cropped using the facial frame provided by Face++ (the blue box in Figure 1), and resized to 224 x 224 pixels.

Extracting facial features using a deep neural network. Facial features were extracted from the images using a widely employed DNN, called VGG-Face (Parkhi, Vedaldi, & Zisserman, 2015). VGG-Face was originally developed (or *trained*) using a sample of 2.6 million images for the purpose of facial recognition (i.e., recognizing a given person across different images). VGG-Face is similar to traditional scoring keys accompanying psychometric tests. A traditional scoring key can be used to convert responses to test questions into one or more psychometric scores, such as a single IQ score, or a set of five Big Five personality scores. VGG-Face translate a facial image into 4,096 scores subsuming its core features. Unfortunately, unlike psychometric scores, VGG-Face scores are not easily interpretable. A single score might subsume differences in multiple facial features typically considered to be distinct by humans (e.g., nose shape, skin tone, or eye color).

VGG-Face offers two main advantages in the context of this study. First, successful facial recognition depends on the DNN's ability to detect facial features that are unlikely to vary across images. Thus, VGG-Face aims at representing a given face as a vector of scores that are as unaffected as possible by facial expression, background, lighting, head orientation, image properties such as brightness or contrast, and other factors that can vary across different images of the same person. Consequently, employing VGG-Face scores enabled us to minimize the role of such transient features when distinguishing between gay and heterosexual faces. Second, employing a DNN trained on a different sample and for a different purpose, reduces the risk of overfitting (i.e., discovering differences between gay and heterosexual faces that are specific to our sample rather than universal). We also tried training a custom DNN directly on the images in our sample; its accuracy was somewhat higher, but it exposed us to the risk of overfitting.

Training classifiers. We used a simple prediction model, logistic regression, combined with a standard dimensionality-reduction approach: singular value decomposition (SVD). SVD is similar to principal component analysis (PCA), a dimensionality-reduction approach widely used by social scientists. The models were trained separately for each gender.

Self-reported sexual orientation (gay/heterosexual) was used as a dependent variable; 4,096 scores, extracted using VGG-Face, were used as independent variables. To prevent overfitting, we used a 20-fold cross-validation when estimating the predictions. The sample was split into 20 subsamples; one of the subsamples (test set) was put aside, while the remaining 19 subsamples (training sets) were used to train the prediction model. As the number of independent variables was relatively large (4,096) when compared with the number of number of cases (7,083 in the smallest training set), we used SVD to extract n=500 dimensions[4] from the independent variables. This helped to reduce the number of independent variables and eliminate redundant information.

A logistic regression model was trained to classify sexual orientation (a dependent variable) using 500 singular values extracted from VGG-Face scores (independent variables). Least absolute shrinkage and selection operator (LASSO; Hastie, Tibshirani, & Friedman, 2009) was used for variable selection and regularization when training the regression model. The

[4] Dimensions extracted by SVD are referred to as singular values; they are an equivalent of principal components in the context of PCA.

LASSO penalty parameter α was set to 1; the regularization parameter λ was automatically estimated using 10-fold cross-validation.

Finally, the model built on the training set, combining the SVD dimensionality reduction and logistic regression, was used to predict the sexual orientation of the participants in the test set. This procedure was repeated 20 times to assign a probability (ranging from 0 to 1) of being gay to all images in the sample.

For many users, more than one facial image was available. This enabled us to examine how the accuracy changes with the number of facial images available. To produce an aggregate probability of being gay based on *n* images, the probabilities associated with a randomly selected set of *n* images (ranging from 1 to 5) of a given participant were averaged.[5] Thus, a participant with three facial images was described by three probabilities of being gay: one based on a single randomly selected image, one based on two randomly selected images, and one based on all three images.

Results

The accuracy of predicting sexual orientation from facial images is presented in Figure 2. Across this paper, the accuracy is expressed using the area under receiver operating characteristic curve (AUC) coefficient. AUC represents the likelihood of a classifier being correct when presented with the faces of two randomly selected participants—one gay and one heterosexual.

[5] Logit transformation is used whenever the probabilities are averaged in this work. This means that the probabilities are logit transformed and averaged, and the resulting values are converted back into probabilities using an inverse-logit transformation.

The AUC = .50 (or 50%) indicates that the classifier is correct only half of the time, which is no better than a random draw. The AUC = 1.00 (or 100%) indicates that the classifier is always correct. AUC is an equivalent of the Wilcoxon signed-rank test coefficient, used more widely in social sciences.

Among men, the classification accuracy equaled AUC = .81 when provided with one image per person. This means that in 81% of randomly selected pairs—composed of one gay and one heterosexual man—gay men were correctly ranked as more likely to be gay. The accuracy grew significantly with the number of images available per person, reaching 91% for five images. The accuracy was somewhat lower for women, ranging from 71% (one image) to 83% (five images per person).

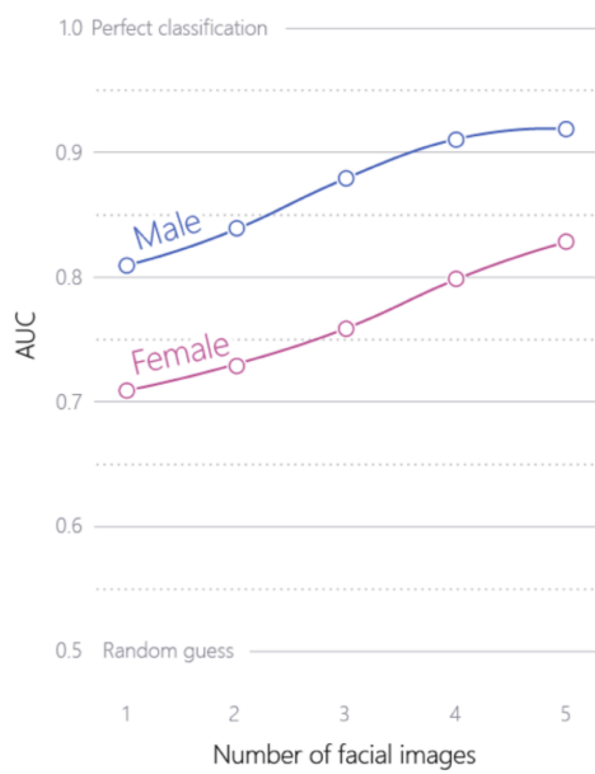


Figure 2. The accuracy of the DNN-based sexual orientation classifier against the number of images used in the classification.

Study 1b: Elements of the Facial Image Employed by the Classifier

The high accuracy of the classifier developed in Study 1a indicates that facial images contained much information related to sexual orientation, and that much of this was captured by the facial features extracted using the VGG-Face. This section examines which parts of the facial image enabled the classification. We address this question by masking parts of a facial image and measuring the degree to which the prediction has changed. If a given area of the image is important to the classifier, masking it is likely to significantly alter the prediction (and vice versa).

Methods

Facial images. The results were produced separately for each gender. Facial images of 100 male and 100 female users were randomly drawn from the sample used in Study 1a. The faces were adjusted to ascertain that a given facial feature (e.g., the mouth) was in exactly the same place in all of the images. This was achieved by warping images (using piecewise linear 2D transformation) to align them along nine landmarks (the left and right eye corners, left and right mouth corners, nose tip, and left and right nose corners).

Sexual orientation classifier. We used the remaining images from Study 1a to train the sexual orientation classifiers (separately for men and women) following the procedure described in Study 1a.

Analysis. We used the sexual orientation classifiers to estimate the probability of being gay for the faces in the samples used here. Next, an area of 7 x 7 pixels in the top-left corner was masked in all 100 images and the probability of being gay was estimated again. The procedure was repeated 1,024 times while sliding the mask across the grid covering the entire image, composed of 32 x 32 squares (each sized at 7 x 7 pixels). The average absolute change in the probability of being gay, resulting from masking a given area of the image, was used as a proxy for the importance of a given area to the prediction of sexual orientation.

Results

The results are presented in Figure 3 as heat maps showing the degree to which masking a given part of an image changes the classification outcome. The color scale ranges from blue (no change) to red (substantial change). Heat maps reveal that, for both genders, classification mainly relied on the facial area and ignored the background. The most informative facial areas among men included the nose, eyes, eyebrows, cheeks, hairline, and chin; informative areas among women included the nose, mouth corners, hair, and neckline. The heat maps are not symmetrical because duplicated facial features, such as eyes, may prompt the classifier to focus on only one of them and ignore the other as redundant.

The results presented here confirm that the VGG-Face scores extracted here focus on the facial features rather than on other parts of the image.

Figure 3. Heat maps showing the degree to which masking a given part of an image changes the (absolute) classification outcome, which is a proxy for the importance of that region in

classification. The color scale ranges from blue (no change) to red (substantial change). The color-coded squares were smoothed using 2D Gaussian filtering.

Study 1c: Facial Features Predictive of Sexual Orientation

Having established that the classification is based on facial features (as opposed to other elements of the image), we turn our attention to the differences between gay and heterosexual faces that enabled the classification. We examine this question by aggregating images classified as most and least likely to be gay in Study 1a.

Methods

Facial images. The results were produced separately for each gender. We used facial images and accompanying probabilities of being gay from Study 1a and retained those containing faces facing the camera directly (the head's pitch and yaw, as estimated by Face++, was lower than two degrees). Next, we selected a subset of images classified as most likely to be gay and a subset of images classified as least likely to be gay. We used subsets of 500 images per set to generate average landmarks' locations and 100 images per set to generate composite faces.

Average landmarks' location. The distances between facial landmarks, extracted using Face++ (see Figure 1), were normalized by setting the distance between the pupils to 1. The faces were centered and rotated to align the eyes horizontally, and the landmark coordinates were averaged.

Composite face. To obtain clearer composite faces, the images were warped using a piecewise linear 2D transformation along the average location of Face++ landmarks (the pixels of each image were transformed using bi-cubic interpolation). The values of corresponding pixels were averaged across images to produce composite faces.

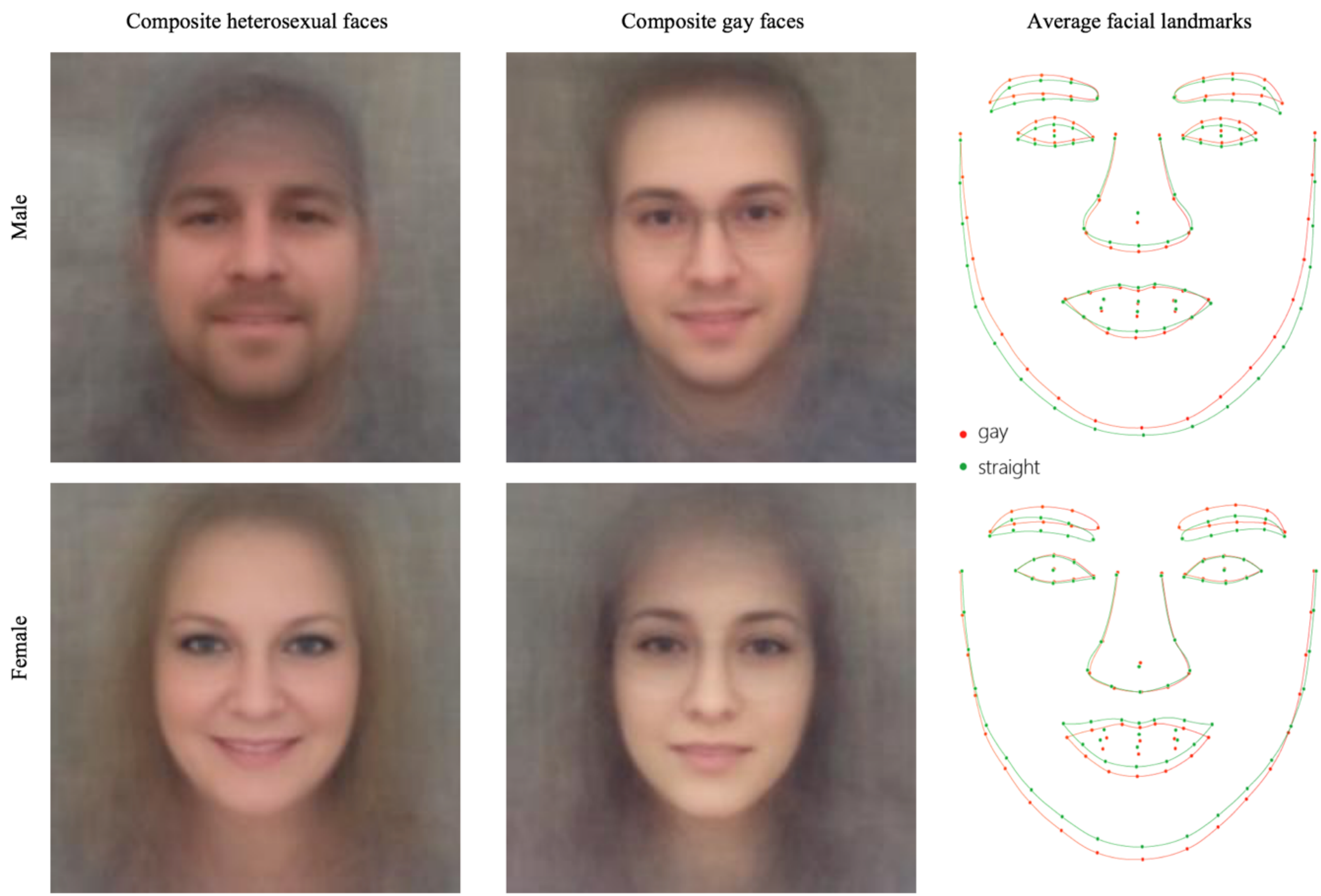
Results

Figure 4 shows the average landmark locations and aggregate appearance of the faces classified as most and least likely to be gay. Consistent with the PHT, gay faces tended to be gender atypical. Average landmark locations revealed that gay men had narrower jaws and longer noses, while lesbians had larger jaws. Composite faces suggest that gay men had larger foreheads than heterosexual men, while lesbians had smaller foreheads than heterosexual women. The differences between the outlines of faces and facial features of gay and heterosexual individuals are further explored in Study 3.

The gender atypicality of gay faces extended beyond morphology. Gay men had less facial hair, suggesting differences in androgenic hair growth, grooming style, or both. They also had lighter skin, suggesting potential differences in grooming, sun exposure, and/or testosterone levels.[6] Lesbians tended to use less eye makeup, had darker hair, and wore less revealing clothes (note the higher neckline), indicating less gender-typical grooming and style. Furthermore, although women tend to smile more in general (Halberstadt, Hayes, & Pike, 1988), lesbians smiled less than their heterosexual counterparts. Additionally, consistent with the association between baseball caps and masculinity in American culture (Skerski, 2011), heterosexual men

[6] Male facial image brightness correlates 0.19 with the probability of being gay, as estimated by the DNN-based classifier. While the brightness of the facial image might be driven by many factors, previous research found that testosterone stimulates melanocyte structure and function leading to a darker skin. (This is also why males tend to have darker skin than females in a given population; Glimcher, Garcia, & Szabó, 1978; Jablonski & Chaplin, 2000).

and lesbians tended to wear baseball caps (see the shadow on their foreheads; this was also confirmed by a manual inspection of individual images). The gender atypicality of the faces of gay men and lesbians is further explored in Study 2.



394

395 *Figure 4*. Composite faces and the average facial landmarks built by averaging faces classified as most and least likely to be gay.

Study 2: Gender Atypicality of Gay People's Faces

The qualitative analysis of the composite faces and average landmarks' locations for gay and heterosexual faces presented in Study 1c suggest that the faces of gay men and lesbians tend to be gender atypical. We test this hypothesis by using a data-driven measure of facial femininity: the DNN-based gender classifier.

Methods

Facial images. We used facial images and accompanying probabilities of being gay estimated in Study 1a.

Facial femininity. We measured facial femininity by using a gender classifier that assigns a probability of being female to each facial image. This gender classifier was developed on an independent sample of 2,891,355 facial images of Facebook users obtained from the myPersonality.org project (Kosinski, Matz, Gosling, Popov, & Stillwell, 2015). We used the same approach to preprocess facial images and train the classifier, as described in Study 1a. This time, however, we used gender as the dependent variable. This gender classifier was applied to all facial images in the sample used in Study 1a. The accuracy of this classifier, when predicting gender, equaled AUC = .98.

Results

The results show that the faces of gay men were more feminine and the faces of lesbians were more masculine than those of their respective heterosexual counterparts. Among men, the data-driven measure of facial femininity positively correlated with the probability of being gay (r

= 0.20; $p<.001$; 95% CI [0.19, 0.21]).[7] The opposite was true for women ($r = -0.21$; $p<.001$; 95% CI [-0.21, -0.20]).

Facial femininity alone allowed for classifying gay and heterosexual faces with some accuracy: AUC = .57 for men and AUC = .58 for women (based on one facial image).

Study 3: Morphology-Based Classifier

Study 1c shows the differences between the outlines of faces and facial features of gay and heterosexual individuals. The current study shows that such basic non-transient morphological features, such as the outline of the nose or facial contour, provide enough information to accurately classify sexual orientation.

Methods

Facial images. We used the same sample as in Study 1a.

Extracting morphological features. We extracted morphological features from the coordinates of the 83 landmarks outlining important facial features provided by Face++ (see Figure 1). To subsume the shape of a given facial feature, such as the nose, we computed Euclidean distances between the landmarks belonging to that feature. For example, as there are 10 landmarks outlining the nose (see Figure 1), its morphology was subsumed by a vector of 10 x 9 = 90 Euclidean distances. To account for the differing sizes of the faces in facial images, the distances were normalized by dividing them by the distance between the pupils.

[7] Pearson product-moment correlation was used. Probabilities were logit transformed.

This approach was applied to the following facial elements: nose, eyes, eyebrows, mouth, contour of the face, and entire face (see Figure 1 for the mapping between landmarks and facial elements).

Training classifiers. The classifiers were trained, separately for each facial element and for all facial landmarks combined, following a procedure similar to the one used in Study 1a. Here, however, we used Euclidean distances instead of the VGG-Face scores as independent variables. If the number of distances describing a given facial element was higher than 500, we used SVD to reduce their number to 500 (in the same way as SVD was used to reduce the number of VGG-Face scores in Study 1a).

Results

The accuracies of the landmark-based classifiers based on five images per person are presented in Figure 5. The results show that the shape of individual facial elements enabled high classification accuracy for both genders. A notably high accuracy was provided by facial contour alone (red landmarks in Figure 1): 75% for men and 63% for women. This provides additional support for the link between jaw shape between gay and heterosexual faces observed in Study 1c (see Figure 4). While the outline of the eyes, eyebrows, and mouth is—to some extent—affected by facial expression and grooming, facial contour is relatively inflexible, emphasizing the predictive power of fixed morphological traits.

The high performance of the contour-based classifiers, and fair performance of the nose-based ones, suggest that the shape of these (relatively fixed) facial elements is sufficient to detect sexual orientation. Overall, the performance of the landmark-based classifiers is remarkable given how little information from the original image is retained in the landmarks' locations.

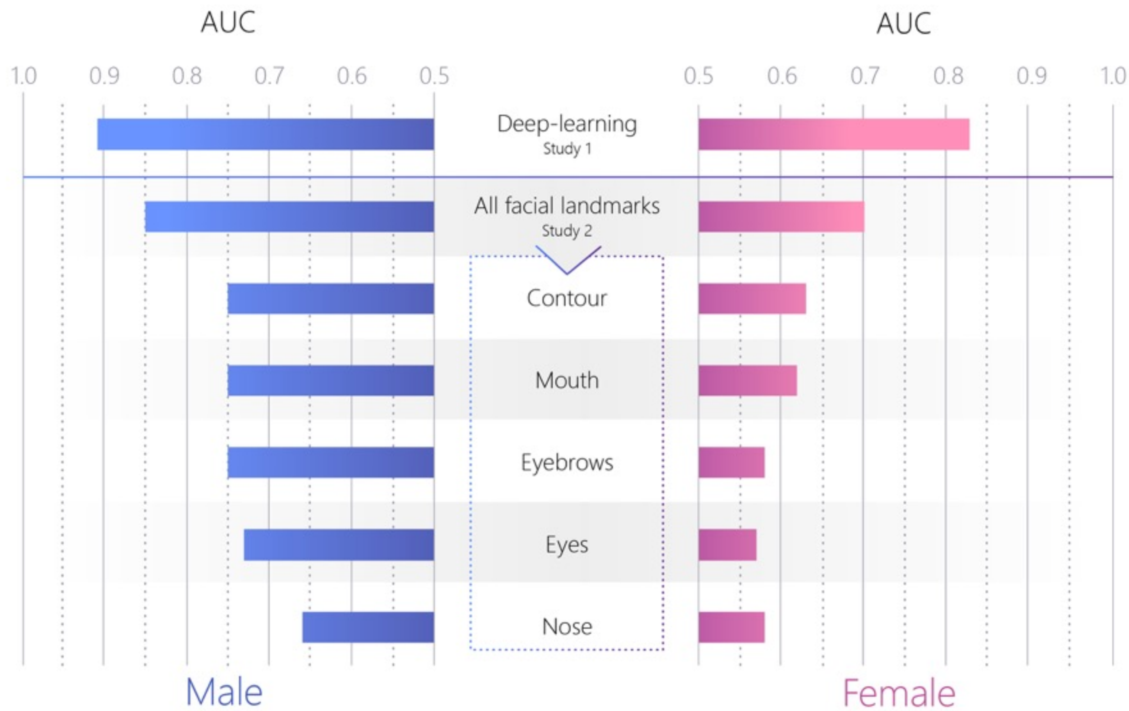


Figure 5. The accuracy of the landmark-based classifiers, when provided with five images per person. The accuracy of the DNN-based classifier trained in Study 1a is displayed on top of the figure for comparison.

Study 4: Human Judges

Study 1a shows that sexual orientation can be accurately determined from non-standardized facial images using a DNN. Study 3 shows that even the most basic non-transient morphological features, such as the shape of the contour of the face, provide sufficient information to predict sexual orientation. It is possible, however, that facial images posted on a dating website are particularly revealing of sexual orientation. Perhaps the users selected the photos that their desired partners might find the most appealing.

We tested this hypothesis by employing a sexual orientation classifier of known accuracy: human judges.[8] We show that the accuracy of the human judges, who were presented with the facial images employed in Study 1a, does not differ from the human judges' accuracy reported in the previous studies employing both: standardized images taken in the lab and dating website profile pictures.

Methods

Facial images. The 35,326 faces from Study 1a were randomly paired, resulting in 50,000 pairs for each gender (each face could be assigned to multiple pairs).

Human judges. We employed AMT workers from the U.S., who had previously completed at least 1,000 tasks and obtained an approval rate of at least 98%. They were asked to select the facial image more likely to represent a gay (or, in half of the cases, heterosexual) person from two, randomly ordered, facial images (one belonging to a gay and one to a heterosexual individual). Note that the accuracy of human judges on a task designed in this way is an equivalent of the AUC coefficient used to express the algorithms' accuracy. The instructions presented to the workers are shown in Figure S2.

Results

Human judges achieved an accuracy of AUC=.61 for male images and AUC=.54 for female images. This is comparable with the accuracy obtained in the previous studies, which ranged from approximately 55 to 65% (Ambady et al., 1999; Lyons et al., 2014; Rule et al.,

---

[8] We also considered applying the DNN-based classifier to the samples used in previous studies. We could not, however, convince their authors to share their samples with us.

2009). It is also compatible with the findings of Study 1a, which show that female faces are less revealing of sexual orientation. Finally, it demonstrates that the facial images used in our study were not unusually revealing of sexual orientation (at least to humans).

Study 5: Beyond Dating Website Facial Images

This study shows that the accuracy of the DNN-based classifier trained in Study 1a is not limited to facial images collected on a dating website, but could also correctly classify facial images recorded in a different environment: Facebook.

Methods

Facial images. We obtained a sample of 14,438 facial images of 6,075 openly gay men from the myPersonality database (Kosinski et al., 2015). Gay males were identified using two variables. First, we used the Facebook Audience Insights platform[9] to identify 50 Facebook Pages most popular among gay men, including Pages such as: “I love being Gay,” “Manhunt,” “Gay and Fabulous,” and “Gay Times Magazine.” Second, we used the “interested in” field of users’ Facebook profiles, which reveals the gender of the people that a given user is interested in. Males that indicated an interest in other males, and that liked at least two out of the predominantly gay Facebook Pages, were labeled as gay. Among the gay men identified in this way, and for whom relationship data was available, 96% reported that their significant other was male. Unfortunately, we were not able to reliably identify heterosexual Facebook users.

Those images were preprocessed and their VGG-Face scores extracted using the procedure described in Study 1a. The final sample contained n=918 facial images of unique users, characterized by an average age of 30 and interquartile range of [27–34]. This sample was

[9] https://www.facebook.com/ads/audience-insights

matched with two subsamples (of gay and heterosexual males) of facial images used in Study 1a. Those subsamples matched the Facebook sample in both size and age distribution.

Results

We applied the classifier trained in Study 1a (employing the VGG-Face scores as an independent variable) to distinguish between the faces of male gay Facebook users, male heterosexual dating-website users, and male gay dating-website users. The classifier could accurately distinguish between gay Facebook users and heterosexual dating-website users in 74% of cases, but was virtually unable to distinguish between gay Facebook users and gay dating-website users (53%). This demonstrates that the classifier trained in Study 1a can correctly identify facial images of gay men obtained in a different environment. It also shows that this classifier is largely insensitive to the origin of the image, as it was unable to distinguish between gay Facebook users and gay dating website users.

General Discussion

The findings reported in this work show that our faces contain more information about sexual orientation than can be perceived or interpreted by the human brain. Study 1a showed that facial features extracted by a DNN can be used to accurately identify the sexual orientation of both men and women. Study 1b showed that the predictions are based on the facial area and not the background. Study 1c revealed that the faces of gay men and lesbians had gender-atypical features, as predicted by the PHT. This was corroborated by the results of Study 2 showing that the probability of being gay was positively correlated with facial femininity among males and negatively correlated with female facial femininity. The high accuracy of the classifier based on the shape of facial elements, presented in Study 3, confirmed that much of the information about sexual orientation is retained in fixed facial features, such as the facial contour or shape of the nose. Study 4 revealed that the non-standardized facial images used in Study 1a were not especially revealing of sexual orientation—at least to human judges, whose accuracy was the same as in previous studies, some of which employed images of neutral faces taken in a carefully controlled environment. Study 5 further corroborated these results by showing that the DNN-based classifier developed in Study 1a performs similarly when presented with facial images of gay men collected in a different environment.

Our results provide strong support for the PHT, which argues that same-gender sexual orientation stems from the underexposure of male fetuses and overexposure of female fetuses to prenatal androgens responsible for the sexual differentiation of faces, preferences, and behavior (Allen & Gorski, 1992; Jannini et al., 2010; Udry & Chantala, 2006). Consistent with the predictions of the PHT, gay men's and gay women's faces were gender atypical—in terms of both fixed (e.g., nose shape) and transient facial features (e.g., grooming style). Some of the differences between gay and heterosexual individuals, such as the shape of the nose or jaw, are most likely driven by developmental factors. In other cases, nature and nurture are likely to be as intertwined as in many other contexts. For example, it is unclear whether gay men were less likely to wear a beard because of nature (sparser facial hair) or nurture (fashion). If it is, in fact, fashion (nurture), to what extent is such a norm driven by the tendency of gay men to have sparser facial hair (nature)? Alternatively, could sparser facial hair (nature) stem from potential differences in diet, lifestyle, or environment (nurture)? Interestingly, female faces seem to be less revealing of sexual orientation, suggesting a weaker link between sexual orientation and prenatal androgen levels among females, or larger fluidity of their sexual orientation.

Identifying links between facial features and psychological traits by employing methodology similar to the one used here could boost our understanding of the origins and nature of a broad range of psychological traits, preferences, and psychological processes. Many of the factors that can be approximated from human faces, such as pre- and post-natal hormonal levels (Jones et al., 2015; Lefevre et al., 2013; Whitehouse et al., 2015), developmental history (Astley et al., 2002), environmental factors, and genes (Ferry et al., 2014), are otherwise difficult to measure. Identifying links between facial features with known links to such factors and psychological traits or behaviors could provide a convenient avenue to generate hypotheses that could be later verified in experimental studies. We hope that future research will explore the links between facial features and other phenomena, such as personality, political views, or psychological conditions.

Importantly, we would like to warn our readers against misinterpreting or overinterpreting this study's findings. First, the fact that the faces of gay men and lesbians are, on average gender atypical, does not imply that all gay men are more feminine than all heterosexual men, or that there are no gay men with extremely masculine facial features (and vice versa in the case of lesbians). The differences in femininity observed in this study were subtle, spread across many facial features, and apparent only when examining averaged images of many faces. Second, our results in no way indicate that sexual orientation can be determined from faces by humans. In fact, Study 4 confirms that humans are rather inaccurate when distinguishing between facial images of gay and homosexual individuals. Finally, interpreting classification accuracy is not trivial and is often counterintuitive. The AUC = .91 does not imply that 91% of gay men in a given population can be identified, or that the classification results are correct 91% of the time. The performance of the classifier depends on the desired trade-off between precision (e.g., the fraction of gay people among those classified as gay) and recall (e.g., the fraction of gay people in the population correctly identified as gay). Aiming for high precision reduces recall, and vice versa.

Let us illustrate this trade-off in a simulated scenario based on the results presented in this work. We simulated a sample of 1,000 men by randomly drawing participants, and their respective probabilities of being gay, from the sample used in Study 1a. As the prevalence of same-gender sexual orientation among men in the U.S. is about 6–7% (Sell, Wells, & Wypij, 1995), we drew 70 probabilities from the gay participants, and 930 from the heterosexual participants. We only considered participants for whom at least 5 facial images were available; note that the accuracy of the classifier in their case reached an AUC = .91.

Setting the threshold above which a given case should be labeled as being gay depends on a desired trade-off between precision and recall. To maximize precision (while sacrificing recall), one should select a high threshold or select only a few cases with the highest probability of being gay. Among 1% (i.e., 10) of individuals with the highest probability of being gay in our simulated sample, 9 were indeed gay and 1 was heterosexual, leading to the precision of 90% (9/10 = 90%). This means, however, that only 9 out of 70 gay men were identified, leading to a low recall of 13% (9/70 = 13%). To boost recall, one needs to sacrifice some of the precision. Among 30 individuals with the highest probability of being gay, 23 were gay and 7 were heterosexual (precision = 23/30 = 77%; recall = 23/70 = 33%). Among the top 100 males most likely to be gay, 47 were gay (precision = 47%; recall = 68%).

This study has a number of limitations. We used nonstandardized images characterized by varying quality, head orientation, or facial expression. This provides for higher ecological validity and a larger, more representative sample, but also introduces confounders (as discussed in Study 1a). Additionally, as the images were obtained from a dating website, they might have been especially revealing of sexual orientation. We believe that we sufficiently addressed this problem by employing a model specifically trained to focus on non-transient facial features (Study 1a), by showing that facial features enabling the prediction were consistent with the theory (PHT; Studies 1c and 2), and by making sure that the images used here were not substantially more revealing of sexual orientation than images of neutral faces taken in a controlled setting (Study 4) or images obtained from Facebook (Study 5). Another issue pertains to the quality of the ground truth: it is possible that some of the users categorized as heterosexual were, in fact, gay or bisexual (or vice versa). However, we believe that people voluntarily seeking partners on the dating website have little incentive to misrepresent their sexual orientation. Furthermore, if some of the users were, in fact, wrongly labelled, correcting such errors would likely *boost* the accuracy of the classifiers examined here. Additionally, despite our attempts to obtain a more diverse sample, we were limited to studying white participants from the U.S. As the prejudice against gay people and the adoption of online dating websites is unevenly distributed across groups characterized by different ethnicities, we could not find sufficient numbers of non-white gay participants. We believe, however, that our results will likely generalize beyond the population studied here. They are consistent with the PHT of sexual orientation, which was supported by variety of studies of humans and other mammals (Hines, 2010). As the exposure to gender-atypical androgen levels is likely to affect the faces of people of different races to a similar degree, it is likely that their facial features are equally revealing of sexual orientation. Finally, it is possible that individuals with more discernibly gay faces are more likely to “come out.” If true, a classifier trained on the faces of openly gay users would be less accurate when detecting non-openly gay individuals. While we do not have data to test this hypothesis, it must be noted that coming out depends on many social, cultural, and legal factors.

Users who came out in our sample may wish or need to maintain their privacy in many contexts and places. Thus, while some faces might be less revealing, many others may prevent their owners from controlling their privacy of sexual orientation.

This brings us to perhaps the most critical nontheoretical ramification of these findings: privacy. Previous studies found that sexual orientation can be detected from an individual's digital footprints, such as social network structure (Jernigan & Mistree, 2009) or Facebook Likes (Kosinski, Stillwell, & Graepel, 2013). Such digital footprints, however, can be hidden, anonymized, or distorted. One's face, on the other hand, cannot be easily concealed. A facial image can be easily taken and analyzed (e.g., with a smartphone or through CCTV). Facial images of billions of people are also stockpiled in digital and traditional archives, including dating platforms, photo-sharing websites, and government databases. Such pictures are often easily accessible; Facebook, LinkedIn, and Google Plus profile pictures, for instance, are public by default and can be accessed by anyone on the Internet. Our findings suggest that such publicly available data and conventional machine learning tools could be employed to build accurate sexual orientation classifiers. As much of the signal seems to be provided by fixed morphological features, such methods could be deployed to detect sexual orientation without a person's consent or knowledge. Moreover, the accuracies reported here are unlikely to constitute an upper limit of what is possible. Employing images of a higher resolution, larger numbers of images per person, larger training set, and more powerful DNN algorithms (e.g., He, Zhang, Ren, & Sun, 2015) could further boost accuracy.

Some people may wonder if such findings should be made public lest they inspire the very application that we are warning against. We share this concern. However, as the governments and companies seem to be already deploying face-based classifiers aimed at detecting intimate traits (Chin & Lin, 2017; Lubin, 2016), there is an urgent need for making policymakers, the general public, and gay communities aware of the risks that they might be facing already. Delaying or abandoning the publication of these findings could deprive individuals of the chance to take preventive measures and policymakers the ability to introduce legislation to protect people. Moreover, this work does not offer any advantage to those who may be developing or deploying classification algorithms, apart from emphasizing the ethical implications of their work. We used widely available off-the-shelf tools, publicly available data, and methods well known to computer vision practitioners. We did not create a privacy-invading tool, but rather showed that basic and widely used methods pose serious privacy threats. We hope that our findings will inform the public and policymakers, and inspire them to design technologies and write policies that reduce the risks faced by homosexual communities across the world.[10]

The growing digitalization of our lives and rapid progress in AI continues to erode the privacy of sexual orientation and other intimate traits. Policymakers and technology companies seem to believe that legislation and new technologies offering individuals more control over their digital footprints can reverse this trend. However, the digital environment is very difficult to police. Data can be easily moved across borders, stolen, or recorded without users' consent. Furthermore, even if users were given full control over their data, it is hard to imagine that they would not share anything publicly. Most people want some of their social media posts, blogs, or

[10] The results reported in this paper were shared, in advance, with several leading international LGBTQ organizations.

profiles to be public. Few would be willing to cover their faces while in the public. As this and other studies show (e.g., Kosinski et al., 2013), such willingly shared digital footprints can be used to reveal intimate traits. Consequently, we believe that further erosion of privacy is inevitable, and the safety of gay and other minorities who may be ostracized in some cultures hinges on the tolerance of societies and governments. The postprivacy world will be a much safer and hospitable place if inhabited by well-educated, tolerant people who are dedicated to equal rights.

Acknowledgments: The authors thank Klaus Fiedler and other reviewers for their great comments on the earlier version of this manuscript. We would also like to thank Samuel Gosling, Robert Sternberg, Raphael Silberzahn, Martie Haselton, Amir Goldberg, Poruz Khambatta, Anonymous Gabriella, Jason Rentfrow, Kai Ruggeri, Pierre Dechant, Brent Roberts, David Mack, and Nicole Paulk for their critical reading of the earlier version of this manuscript. Also, we thank Isabelle Abraham for proofreading and Mariia Vorobiova for graphical design. Finally, we would like to thank the creators of Face++ for allowing us to use their software free of charge.

References

Allen, L. S., & Gorski, R. A. (1992). Sexual orientation and the size of the anterior commissure in the human brain. *Proceedings of the National Academy of Sciences*, *89*(15), 7199–7202. http://doi.org/10.1073/pnas.89.15.7199

Ambady, N., Hallahan, M., & Conner, B. (1999). Accuracy of judgments of sexual orientation from thin slices of behavior. *Journal of Personality and Social Psychology*, *77*(3), 538–547. http://doi.org/10.1037/0022-3514.77.3.538

Astley, S. J., Stachowiak, J., Clarren, S. K., & Clausen, C. (2002). Application of the fetal alcohol syndrome facial photographic screening tool in a foster care population. *The Journal of Pediatrics*, *141*(5), 712–717. http://doi.org/10.1067/mpd.2002.129030

Ballew, C. C., & Todorov, A. (2007). Predicting political elections from rapid and unreflective face judgments. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(46), 17948–17953. http://doi.org/10.1073/pnas.0705435104

Berry, D. S. (1991). Accuracy in social perception: Contributions of facial and vocal information. *Journal of Personality and Social Psychology*, *61*(2), 298–307. http://doi.org/10.1037/0022-3514.61.2.298

Berry, D. S., & Brownlow, S. (1989). Were the physiognomists right? Personality correlates of facial babyishness. *Personality and Social Psychology Bulletin*, *15*(2), 266–279. http://doi.org/10.1177/0146167289152013

Bond, C. F., Berry, D. S., & Omar, A. (1994). The kernel of truth in judgments of deceptiveness. *Basic and Applied Social Psychology*, *15*(4), 523–534. http://doi.org/10.1207/s15324834basp1504_8

Borkenau, P., Brecke, S., Möttig, C., & Paelecke, M. (2009). Extraversion is accurately perceived after a 50-ms exposure to a face. *Journal of Research in Personality*, *43*(4), 703–706. http://doi.org/10.1016/j.jrp.2009.03.007

Brambilla, M., Riva, P., & Rule, N. O. (2013). Familiarity increases the accuracy of categorizing male sexual orientation. *Personality and Individual Differences*, *55*(2), 193–195. http://doi.org/10.1016/j.paid.2013.02.023

Brown, E., & Perrett, D. I. (1993). What gives a face its gender? *Perception*, *22*(7), 829–840. http://doi.org/10.1068/p220829

Bulygina, E., Mitteroecker, P., & Aiello, L. (2006). Ontogeny of facial dimorphism and patterns of individual development within one human population. *American Journal of Physical Anthropology*, *131*(3), 432–443. http://doi.org/10.1002/ajpa.20317

Chin, J., & Lin, L. (2017, June 26). China's all-seeing surveillance state is reading its citizens' faces. *The Wall Street Journal*. Retrieved from https://www.wsj.com

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, *542*(7639), 115–118. http://doi.org/10.1038/nature21056

Ferry, Q., Steinberg, J., Webber, C., FitzPatrick, D. R., Ponting, C. P., Zisserman, A., & Nellåker, C. (2014). Diagnostically relevant facial gestalt information from ordinary photos. *eLife*, *2014*(3). http://doi.org/10.7554/eLife.02020.001

Glaser, G. (2002). *The Nose: A Profile of Sex, Beauty, and Survival*. Simon and Schuster.

Glimcher, M. E., Garcia, R. I., & Szabó, G. (1978). Organ culture of mammalian skin and the effects of ultraviolet light and testosterone on melanocyte morphology and function.

Journal of Experimental Zoology, *204*(2), 229–237. http://doi.org/10.1002/jez.1402040210

Halberstadt, A. G., Hayes, C. W., & Pike, K. M. (1988). Gender and gender role differences in smiling and communication consistency. *Sex Roles*, *19*(9–10), 589–604. http://doi.org/10.1007/BF00289738

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning* (Vol. 1). New York: Springer series in statistics. http://doi.org/10.1007/b94608

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition. *arXiv Preprint arXiv:1512.03385v1*, *7*(3), 171–180. http://doi.org/10.3389/fpsyg.2013.00124

Hines, M. (2010). Sex-related variation in human behavior and the brain. *Trends in Cognitive Sciences*, *14*(10), 448–456. http://doi.org/10.1016/j.tics.2010.07.005

Hughes, S. M., & Bremme, R. (2011). The effects of facial symmetry and sexually-dimorphic facial proportions on assessments of sexual orientation. *Journal of Social, Evolutionary, and Cultural Psychology*, *5*(4), 214–230. http://doi.org/10.1037/h0099261

Jablonski, N. G., & Chaplin, G. (2000). The evolution of human skin coloration. *Journal of Human Evolution*, *39*(1), 57–106. http://doi.org/10.1006/jhev.2000.0403

Jannini, E. A., Blanchard, R., Camperio-Ciani, A., & Bancroft, J. (2010). Male homosexuality: Nature or culture? *The Journal of Sexual Medicine*, *7*(10), 3245–3253. http://doi.org/10.1111/j.1743-6109.2010.02024.x

Jenkinson, J. (1997). Face facts: A history of physiognomy from ancient Mesopotamia to the end of the 19th century. *The Journal Of Biocommunication*, *24*(3), 2–7.

Jernigan, C., & Mistree, B. F. (2009). Gaydar: Facebook friendships expose sexual orientation. *First Monday*, *14*(10). http://doi.org/10.5210/fm.v14i10.2611

Jones, B. C., Hahn, A. C., Fisher, C. I., Wincenciak, J., Kandrik, M., Roberts, S. C., … DeBruine, L. M. (2015). Facial coloration tracks changes in women's estradiol. *Psychoneuroendocrinology*, *56*, 29–34. http://doi.org/10.1016/j.psyneuen.2015.02.021

King, M., Semlyen, J., Tai, S. S., Killaspy, H., Osborn, D., Popelyuk, D., & Nazareth, I. (2008). A systematic review of mental disorder, suicide, and deliberate self harm in lesbian, gay and bisexual people. *BMC Psychiatry*, *8*, 70. http://doi.org/10.1186/1471-244X-8-70

Kosinski, M., Matz, S. C., Gosling, S. D., Popov, V., & Stillwell, D. J. (2015). Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines. *American Psychologist*, *70*(6), 543–556. http://doi.org/10.1037/a0039210

Kosinski, M., Stillwell, D. J., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(15), 5802–5805. http://doi.org/10.1073/pnas.1218772110

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. http://doi.org/10.1038/nature14539

Lefevre, C. E., Lewis, G. J., Perrett, D. I., & Penke, L. (2013). Telling facial metrics: Facial width is associated with testosterone levels in men. *Evolution and Human Behavior*, *34*(4), 273–279. http://doi.org/10.1016/j.evolhumbehav.2013.03.005

Lefevre, C. E., Wilson, V. A. D., Morton, F. B., Brosnan, S. F., Paukner, A., & Bates, T. C.

(2014). Facial width-to-height ratio relates to alpha status and assertive personality in capuchin monkeys. *PLOS ONE*, *9*(4), e93369. http://doi.org/10.1371/journal.pone.0093369

Little, A. C., Burriss, R. P., Jones, B. C., & Roberts, S. C. (2007). Facial appearance affects voting decisions. *Evolution and Human Behavior*, *28*(1), 18–27. http://doi.org/10.1016/j.evolhumbehav.2006.09.002

Lõhmus, M., Sundström, L. F., & Björklund, M. (2009). Dress for success: Human facial expressions are important signals of emotions. *Annales Zoologici Fennici*, *46*(1), 75–80. http://doi.org/10.5735/086.046.0108

Lombroso, C. (1911). *Criminal man*. New York: GP Putnam.

Lu, C., & Tang, X. (2014). Surpassing human-level face verification performance on LFW with GaussianFace. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence* (pp. 3811–3819). Austin, TX: AAAI Press. Retrieved from http://arxiv.org/abs/1404.3840

Lubin, G. (2016, October 12). Facial-profiling could be dangerously inaccurate and biased, experts warn. *Business Insider*. Retrieved from http://www.businessinsider.de/

Lukaszewski, A. W., & Roney, J. R. (2011). The origins of extraversion: Joint effects of facultative calibration and genetic polymorphism. *Personality and Social Psychology Bulletin*, *37*(3), 409–421. http://doi.org/10.1177/0146167210397209

Lyons, M., Lynch, A., Brewer, G., & Bruno, D. (2014). Detection of sexual orientation (“gaydar”) by homosexual and heterosexual women. *Archives of Sexual Behavior*, *43*(2), 345–352. http://doi.org/10.1007/s10508-013-0144-7

Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual Review of Psychology*, *51*, 93–120. http://doi.org/10.1146/annurev.psych.51.1.93

Meyer-Bahlburg, H. F. L. (1984). Psychoendocrine research on sexual orientation: Current status and future options. *Progress in Brain Research*, *61*, 375–298. http://doi.org/10.1016/S0079-6123(08)64448-9

Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). Deep Face Recognition. In X. Xie, M. W. Jones, & G. K. L. Tam (Eds.), *Proceedings of the British Machine Vision Conference* (p. 41.1). Swansea, UK: BMVA Press. http://doi.org/10.5244/C.29.41

Penton-Voak, I. S., Pound, N., Little, A. C., & Perrett, D. I. (2006). Personality judgments from natural and composite facial images: More evidence for a “kernel of truth” in social perception. *Social Cognition*, *24*(5), 607–640. http://doi.org/10.1521/soco.2006.24.5.607

Porter, R. (Ed.). (2003). *The Cambridge history of science: Eighteenth-century science* (Vol. 4). Cambridge, UK: Cambridge University Press.

Rhodes, G. (2006). The evolutionary psychology of facial beauty. *Annual Review of Psychology*, *57*, 199–226. http://doi.org/10.1146/annurev.psych.57.102904.190208

Riedweg, C. (2005). *Pythagoras: His life, teaching, and influence*. New York, NY: Cornell University Press.

Roberts, T., & Bruce, V. (1988). Feature saliency in judging the sex and familiarity of faces. *Perception*, *17*(4), 475–481. http://doi.org/10.1068/p170475

Rule, N. O., & Ambady, N. (2008). Brief exposures: Male sexual orientation is accurately perceived at 50 ms. *Journal of Experimental Social Psychology*, *44*(4), 1100–1105. http://doi.org/10.1016/j.jesp.2007.12.001

Rule, N. O., & Ambady, N. (2010). Democrats and Republicans can be differentiated from their faces. *PLoS ONE*, *5*(1), e8733. http://doi.org/10.1371/journal.pone.0008733

Rule, N. O., Ambady, N., Adams, R., & Macrae, C. N. (2008). Accuracy and awareness in the perception and categorization of male sexual orientation. *Journal of Personality and Social Psychology*, *95*(5), 1019–1028. http://doi.org/10.1037/a0013194

Rule, N. O., MacRae, C. N., & Ambady, N. (2009). Ambiguous group membership is extracted automatically from faces. *Psychological Science*, *20*(4), 441–443. http://doi.org/10.1111/j.1467-9280.2009.02314.x

Samochowiec, J., Wänke, M., & Fiedler, K. (2010). Political ideology at face value. *Social Psychological and Personality Science*, *1*(3), 206–213. http://doi.org/10.1177/1948550610372145

Sell, R. L., Wells, J. A., & Wypij, D. (1995). The prevalence of homosexual behavior and attraction in the United States, the United Kingdom and France: Results of national population-based samples. *Archives of Sexual Behavior*, *24*(3), 235–248. http://doi.org/10.1007/BF01541598

Skerski, J. (2011). Tomboy chic: Re-fashioning gender rebellion. *Journal of Lesbian Studies*, *15*(4), 466–479. http://doi.org/10.1080/10894160.2011.532031

Skorska, M. N., Geniole, S. N., Vrysen, B. M., McCormick, C. M., & Bogaert, A. F. (2015). Facial structure predicts sexual orientation in both men and women. *Archives of Sexual Behavior*, *44*(5), 1377–1394. http://doi.org/10.1007/s10508-014-0454-4

Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science*, *308*(5728), 1623–1626.

http://doi.org/10.1126/science.1110589

Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences*, *12*(12), 455–460. http://doi.org/10.1016/j.tics.2008.10.001

Tskhay, K. O., & Rule, N. O. (2015). Emotions facilitate the communication of ambiguous group memberships. *Emotion*, *15*(6), 812–826. http://doi.org/10.1037/emo0000077

Udry, J. R. (2000). Biological limits of gender construction. *American Sociological Review*, *65*(3), 443–457. http://doi.org/10.2307/2657466

Udry, J. R., & Chantala, K. (2006). Masculinity–femininity predicts sexual orientation in men but not in women. *Journal of Biosocial Science*, *38*(6), 797–809. http://doi.org/10.1017/S002193200500101X

UN Human Rights Council. (2015). *Discrimination and violence against individuals based on their sexual orientation and gender identity*. Retrieved from http://www.refworld.org/docid/5571577c4.html

Valentova, J. V., Kleisner, K., Havlíček, J., & Neustupa, J. (2014). Shape differences between the faces of homosexual and heterosexual men. *Archives of Sexual Behavior*, *43*(2), 353–361. http://doi.org/10.1007/s10508-013-0194-x

Whitehouse, A. J. O., Gilani, S. Z., Shafait, F., Mian, A., Tan, D. W., Maybery, M. T., … Eastwood, P. (2015). Prenatal testosterone exposure is related to sexually dimorphic facial morphology in adulthood. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1816), 20151351. http://doi.org/10.1098/rspb.2015.1351

Zebrowitz, L. A. (1997). *Reading faces: Window to the soul? New directions in social* *psychology*. Boulder, CO: Westview Press.

Zebrowitz, L. A., & Collins, M. (1997). Accurate social perception at zero acquaintance: the affordances of a Gibsonian approach. *Personality and Social Psychology Review*, *1*(3), 204–223. http://doi.org/10.1207/s15327957pspr0103_2

Zebrowitz, L. A., Collins, M., & Dutta, R. (1998). The relationship between appearance and personality across the life span. *Personality and Social Psychology Bulletin*, *24*(7), 736–749. http://doi.org/10.1177/0146167298247006

Supplementary Materials

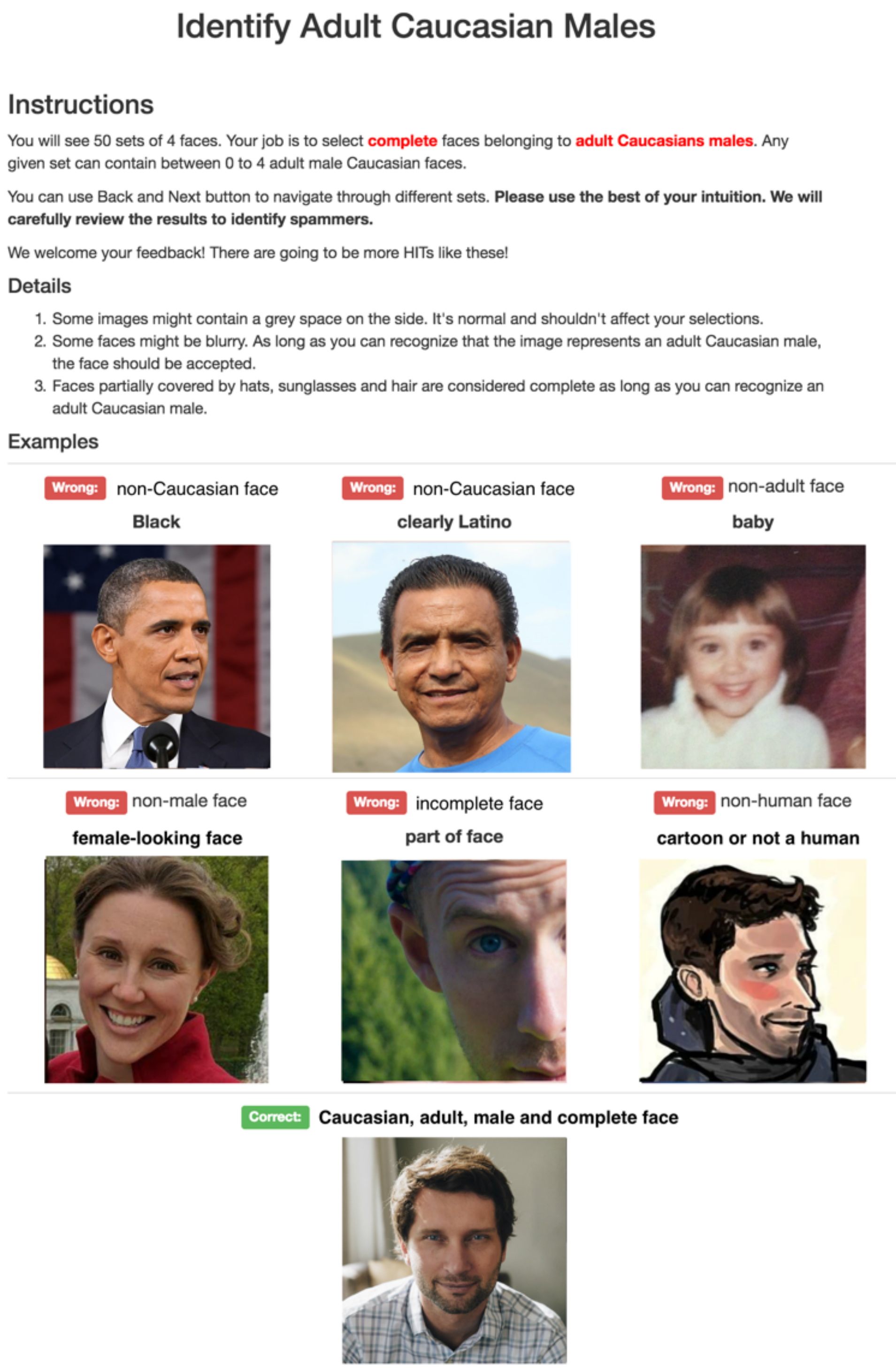


Figure S1. Instructions given to AMT workers employed to remove incomplete, non-Caucasian, nonadult, and nonhuman male faces. We used similar instructions for female faces.

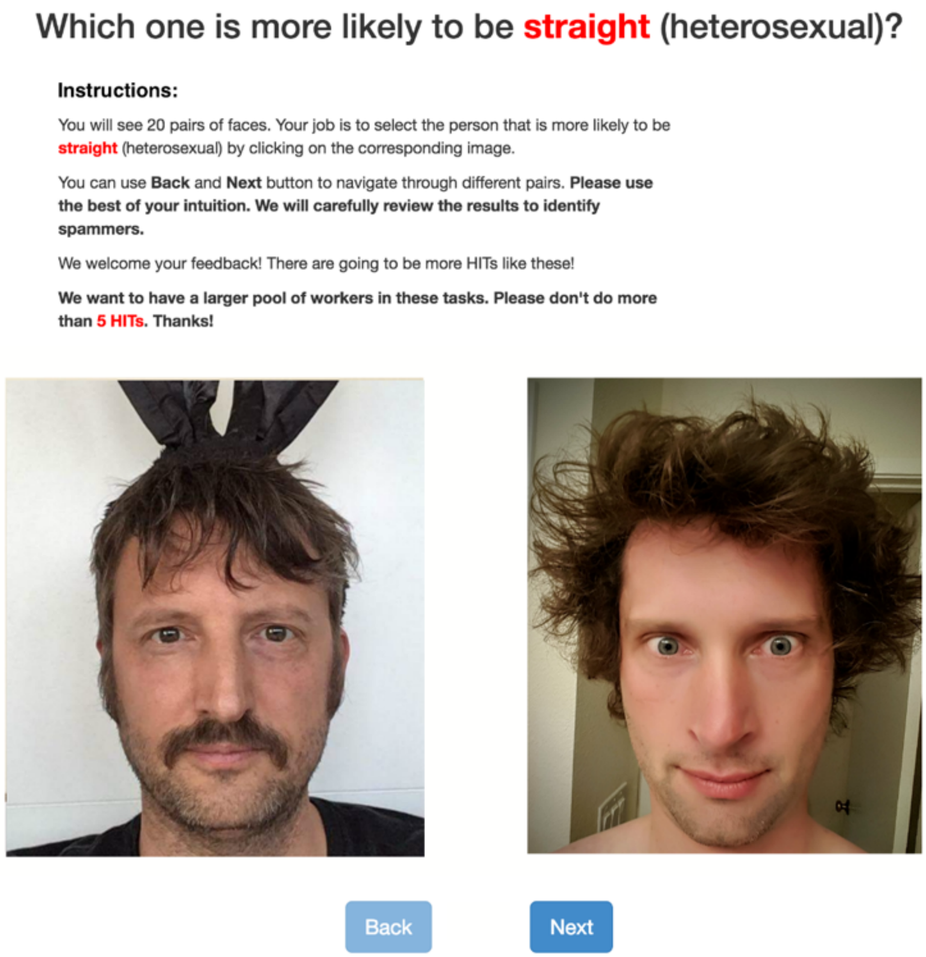


Figure S2. Instructions given to AMT workers employed to classify heterosexual and gay faces.