

1 THIS IS A PREPRINT OF THE PEER REVIEWED ARTICLE TO APPEAR IN JOURNAL OF
2 PERSONALITY AND SOCIAL PSYCHOLOGY.

3

4 THE MOST RECENT VERSION IS AVAILABLE AT <https://osf.io/zn79k/>

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

6

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

9

10 Yilun Wang, Michal Kosinski

11 Graduate School of Business, Stanford University, Stanford, CA94305, USA
12 michalk@stanford.edu

13

14

15 *The study has been approved by the IRB at Stanford University*

16

17 Citation: Wang, Y., & Kosinski, M. (in press). Deep neural networks are more accurate than
18 humans at detecting sexual orientation from facial images. *Journal of Personality and*
19 *Social Psychology*.

20

21

22

Author Note:

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

24

25 **©American Psychological Association, 2017. This paper is not the copy of record**
26 **and may not exactly replicate the authoritative document published in the APA**
27 **journal. Please do not copy or cite without author's permission. The final article is**
28 **available, upon publication, at: <http://www.apa.org/pubs/journals/psp/>**

29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

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

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

52 The science of judging one's character from their facial characteristics, or physiognomy,
53 dates back to ancient China and Greece (Jenkinson, 1997). Aristotle and Pythagoras were among
54 its disciples, and the latter used to select his students based on their facial features (Riedweg,
55 2005). Such beliefs have persisted and grown in popularity over the centuries. Robert FitzRoy,
56 the captain of the Beagle, believed that Darwin's nose revealed a lack of energy and
57 determination, and was close to barring him from the historic voyage (Glaser, 2002). Cesare
58 Lombroso, the founder of criminal anthropology, believed that criminals could be identified by
59 their facial features. He claimed, for example, that arsonists have a "softness of skin, an almost
60 childlike appearance, and an abundance of thick straight hair that is almost feminine"
61 (Lombroso, 1911, p. 51). By the eighteenth century, physiognomy "was not merely a popular fad
62 but also the subject of intense academic debate about the promises it held for future progress"
63 (Porter, 2003, p. 497).

64 Physiognomy is now universally, and rightly, rejected as a mix of superstition and racism
65 disguised as science (Jenkinson, 1997). Due to its legacy, studying or even discussing the links
66 between facial features and character became taboo, leading to a widespread presumption that no
67 such links exist. However, there are many demonstrated mechanisms that imply the opposite.
68 Such mechanisms can be arranged into three groups. First, there is much evidence that character
69 can influence one's facial appearance (e.g., Löhmus, Sundström, & Björklund, 2009; Zebrowitz
70 & Collins, 1997). For example, women that scored high on extroversion early in life tend to
71 become more attractive with age (Zebrowitz, Collins, & Dutta, 1998). Second, facial appearance
72 can alter one's character. Facial appearance drives first impressions of others, influencing our

73 expectations and behavior toward them, which, in turn, shapes their character (Berry, 1991;
74 Berry & Brownlow, 1989; Penton-Voak, Pound, Little, & Perrett, 2006; Todorov, Said, Engell, &
75 Oosterhof, 2008; Zebrowitz & Collins, 1997; Zebrowitz et al., 1998). Good-looking people, for
76 example, receive more positive social feedback, and thus tend to become even more extroverted
77 (Lukaszewski & Roney, 2011). Finally, there is a broad range of factors affecting both facial
78 appearance and one's traits. Those include pre- and post-natal hormonal levels (Jones et al.,
79 2015; Lefevre, Lewis, Perrett, & Penke, 2013; Whitehouse et al., 2015), developmental history
80 (Astley, Stachowiak, Clarren, & Clausen, 2002), environmental factors, and gene expression
81 (Ferry et al., 2014). Testosterone levels, for instance, significantly affect both: behavior (e.g.,
82 dominance) and facial appearance (e.g., facial-width-to-height-ratio; Lefevre et al., 2014).

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

96 possible that some of our intimate traits are prominently displayed on the face, even if others
97 cannot perceive them. Here, we test this hypothesis using modern computer vision algorithms.

98 Recent progress in AI and computer vision has been largely driven by the widespread
99 adoption of deep neural networks (DNN), or neural networks composed of a large number of
100 hidden layers (LeCun, Bengio, & Hinton, 2015). DNNs mimic the neocortex by simulating large,
101 multi-level networks of interconnected neurons. DNNs excel at recognizing patterns in large,
102 unstructured data such as digital images, sound, or text, and analyzing such patterns to make
103 predictions. DNNs are increasingly outperforming humans in visual tasks such as image
104 classification, facial recognition, or diagnosing skin cancer (Esteva et al., 2017; LeCun et al.,
105 2015; Lu & Tang, 2014). The superior performance of DNNs offers an opportunity to identify
106 links between characteristics and facial features that might be missed or misinterpreted by the
107 human brain.

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

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

119 Second, the widely accepted prenatal hormone theory (PHT) of sexual orientation
120 predicts the existence of links between facial appearance and sexual orientation. According to the
121 PHT, same-gender sexual orientation stems from the underexposure of male fetuses or
122 overexposure of female fetuses to androgens that are responsible for sexual differentiation (Allen
123 & Gorski, 1992; Jannini, Blanchard, Camperio-Ciani, & Bancroft, 2010; Udry & Chantala,
124 2006). As the same androgens are responsible for the sexual dimorphism of the face, the PHT
125 predicts that gay people will tend to have gender-atypical facial morphology (Bulygina,
126 Mitteroecker, & Aiello, 2006; Rhodes, 2006; Whitehouse et al., 2015). According to the PHT,
127 gay men should tend to have more feminine facial features than heterosexual men, while lesbians
128 should tend to have more masculine features than heterosexual women. Thus, gay men are
129 predicted to have smaller jaws and chins, slimmer eyebrows, longer noses, and larger foreheads;
130 the opposite should be true for lesbians. Furthermore, as prenatal androgen levels also drive the
131 sexual differentiation of behaviors and preferences during adulthood (Meyer-Bahlburg, 1984;
132 Udry, 2000), the PHT predicts that gay people may tend to adopt gender-atypical facial
133 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.

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

136 Previous empirical evidence provides mixed support for the gender typicality of facial
137 features of gay men and women. Huges and Bremme (2011) studied a sample of 60 images of
138 gay men and concluded that gay men had, on average, more feminine facial features. Lyons et al.
139 (2014) asked 120 human judges to rate the masculinity and femininity of 80 faces of men and
140 women. They found that on average, heterosexual women and gay men were rated as more
141 feminine and less masculine than lesbians and heterosexual men. However, Skorska, Geniole,
142 Vrysen, McCormick, and Bogaert (2015) used a sample of 390 photographs of men and women,
143 and found that both lesbians and gay men had more masculine faces than heterosexual women
144 and men, respectively. Valentova, Kleisner, Havlíček, and Neustupa (2014, p. 353) used a sample
145 of facial images of 40 gay and 40 heterosexual men, and found that on average, gay men had
146 relatively wider and shorter faces, smaller and shorter noses, and larger and more rounded jaws,
147 or “a mosaic of both feminine and masculine features.” Such mixed findings might be attributed
148 to the difficulty of precisely defining and measuring facial femininity. They might also be
149 attributed to the fact that the difference between gay and heterosexual faces may be too subtle to
150 be reliably detected in the small samples employed in these studies. This study aims to address
151 those limitations by using a much larger sample size and data-driven methods, including an
152 algorithm-based measure of facial femininity.

153 Finally, and perhaps most importantly, the predictability of sexual orientation could have
154 serious and even life-threatening implications to gay men and women and the society as a whole.
155 In some cultures, gay men and women still suffer physical and psychological abuse at the hands
156 of governments, neighbors, and even their own families. Perhaps due to discrimination and

157 stigmatization, gay people are also at a higher risk of depression, suicide, self-harm, and
158 substance abuse (King et al., 2008). Consequently, their well-being and safety may depend on
159 their ability to control when and to whom to reveal their sexual orientation. Press reports suggest
160 that governments and corporations are developing and deploying face-based prediction tools
161 aimed at intimate psycho–demographic traits, such as the likelihood of committing a crime, or
162 being a terrorist or pedophile (Chin & Lin, 2017; Lubin, 2016). The laws in many countries
163 criminalize same-gender sexual behavior, and in eight countries—including Iran, Mauritania,
164 Saudi Arabia, and Yemen—it is punishable by death (UN Human Rights Council, 2015). It is
165 thus critical to inform policymakers, technology companies and, most importantly, the gay
166 community, of how accurate face-based predictions might be.

167 This work examines whether an intimate psycho–demographic trait, sexual orientation, is
168 displayed on human faces beyond what can be perceived by humans. We address this question
169 using a data-driven approach. A DNN was used to extract features from the facial images of
170 35,326 gay and heterosexual men and women. These features were entered (separately for each
171 gender) as independent variables into a cross-validated logistic regression model aimed at
172 predicting self-reported sexual orientation. The resulting classification accuracy offers a proxy
173 for the amount of information relevant to the sexual orientation displayed on human faces. We
174 also explore the features employed by the classifier and examine whether, as predicted by the
175 PHT, the faces of gay men and women tend to be gender atypical. Furthermore, we compare the
176 accuracy of the computer algorithm with that of human judges. Human accuracy does not only
177 provide a baseline for interpreting the algorithm’s accuracy, but it also helps to examine whether
178 the nonstandardized facial images used here are not more revealing of sexual orientation than
179 standardized facial images taken in a controlled environment. Finally, using an independent

180 sample of gay men’s facial images, we test the external predictive validity of the classifier
181 developed here.

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

183 In Study 1a, we show that a DNN can be used to identify sexual orientation from facial
184 images. Previous studies linking facial features with sexual orientation used either images of
185 neutral² faces taken in a laboratory (e.g., Skorska et al., 2015; Valentova et al., 2014) or self-
186 taken images obtained from online dating websites (e.g., Hughes & Bremme, 2011; Lyons et al.,
187 2014; Rule & Ambady, 2008; Rule, Ambady, Adams, & Macrae, 2008). We employed the latter
188 approach, as such images can be collected in large numbers, from more representative samples,
189 and at a lower cost (from the perspective of both the participants and researchers). Larger and
190 more representative samples, in turn, enable the discovery of phenomena that might not have
191 been apparent in the smaller, lab-based samples. Additionally, using self-taken, easily accessible
192 digital facial images increases the ecological validity of our results, which is particularly
193 important given their critical privacy implications.

194 Images taken and uploaded by the participants have a number of potential drawbacks.
195 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).

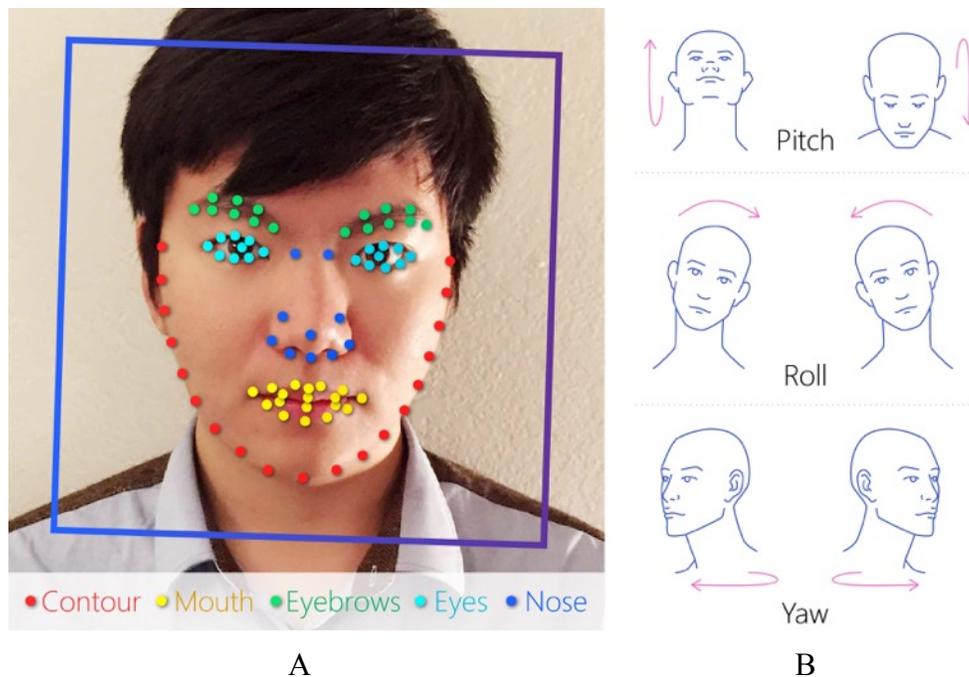
196 given that they were originally posted on a dating website, they might be especially revealing of
197 sexual orientation. We take several steps to mitigate the influence of such factors. First, the facial
198 features are extracted using a DNN that was specifically developed to focus on non-transient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

206 **Methods**

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

212

213



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

217

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

223 Based on the Face++ results, we removed images containing multiple faces, partially
 224 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>.

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

227

228 Table 1

229 *Frequencies of Users and Facial Images, and the Age Distribution in the Final Sample Used in*
 230 *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	2,037	2,878
3 images	1,363	1,367	1,058	1,951
4 images	562	731	494	1,114
5 images	219	327	223	491

231 *Note.* IQR stands for interquartile range.

232

233 Next, we employed Amazon Mechanical Turk (AMT) workers to verify that the faces
 234 were adult, Caucasian, fully visible, and of a gender that matched the one reported on the user’s
 235 profile. We limited the task to the workers from the U.S., who had previously completed at least
 236 1,000 tasks and obtained an approval rate of at least 98%. Only faces approved by four out of six
 237 workers were retained. See Figure S1 for the instructions presented to the workers.

238 Finally, we randomly removed some users to balance the age distribution of the sexual
 239 orientation subsamples and their size—separately for each gender. The final sample contained
 240 35,326 facial images of 14,776 gay and heterosexual (50/50%) men and women (53/47%; see

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

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

254 VGG-Face offers two main advantages in the context of this study. First, successful facial
255 recognition depends on the DNN's ability to detect facial features that are unlikely to vary across
256 images. Thus, VGG-Face aims at representing a given face as a vector of scores that are as
257 unaffected as possible by facial expression, background, lighting, head orientation, image
258 properties such as brightness or contrast, and other factors that can vary across different images
259 of the same person. Consequently, employing VGG-Face scores enabled us to minimize the role
260 of such transient features when distinguishing between gay and heterosexual faces. Second,
261 employing a DNN trained on a different sample and for a different purpose, reduces the risk of
262 overfitting (i.e., discovering differences between gay and heterosexual faces that are specific to

263 our sample rather than universal). We also tried training a custom DNN directly on the images in
264 our sample; its accuracy was somewhat higher, but it exposed us to the risk of overfitting.

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

269 Self-reported sexual orientation (gay/heterosexual) was used as a dependent variable;
270 4,096 scores, extracted using VGG-Face, were used as independent variables. To prevent
271 overfitting, we used a 20-fold cross-validation when estimating the predictions. The users were
272 split into 20 subsamples; one of the subsamples (test set) was put aside, while the remaining 19
273 subsamples (training sets) were used to train the prediction model. As the number of independent
274 variables was relatively large (4,096) when compared with the number of cases (7,083 in the
275 smallest training set), we used SVD to extract $n=500$ dimensions⁴ from the independent
276 variables. This helped to reduce the number of independent variables and eliminate redundant
277 information.

278 A logistic regression model was trained to classify sexual orientation (a dependent
279 variable) using 500 singular values extracted from VGG-Face scores (independent variables).
280 Least absolute shrinkage and selection operator (LASSO; Hastie, Tibshirani, & Friedman, 2009)
281 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.

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

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

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

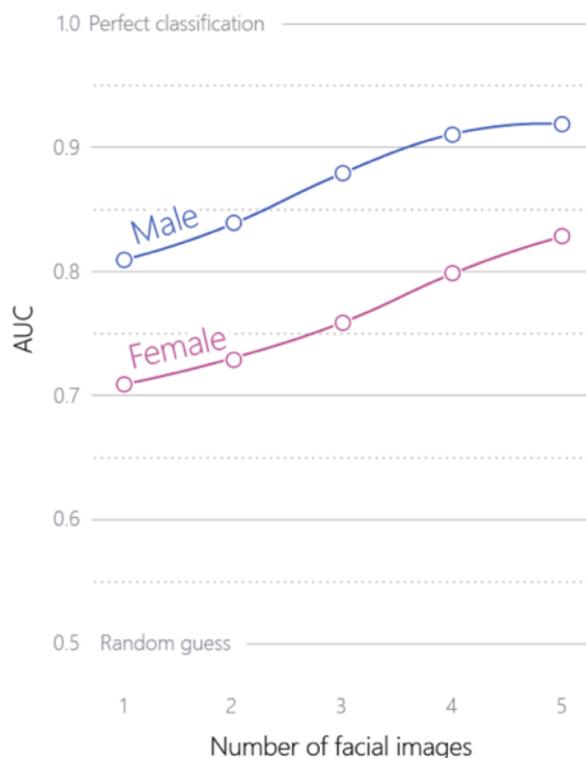
295 **Results**

296 The accuracy of predicting sexual orientation from facial images is presented in Figure 2.
297 Across this paper, the accuracy is expressed using the area under receiver operating characteristic
298 curve (AUC) coefficient. AUC represents the likelihood of a classifier being correct when
299 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.

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

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



310
 311 *Figure 2.* The accuracy of the DNN-based sexual orientation classifier against the number of
 312 images used in the classification.

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

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

321 Methods

322 **Facial images.** The results were produced separately for each gender. Facial images of
323 100 male and 100 female users were randomly drawn from the sample used in Study 1a. The
324 faces were adjusted to ascertain that a given facial feature (e.g., the mouth) was in exactly the
325 same place in all of the images. This was achieved by warping images (using piecewise linear 2D
326 transformation) to align them along nine landmarks (the left and right eye corners, left and right
327 mouth corners, nose tip, and left and right nose corners).

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

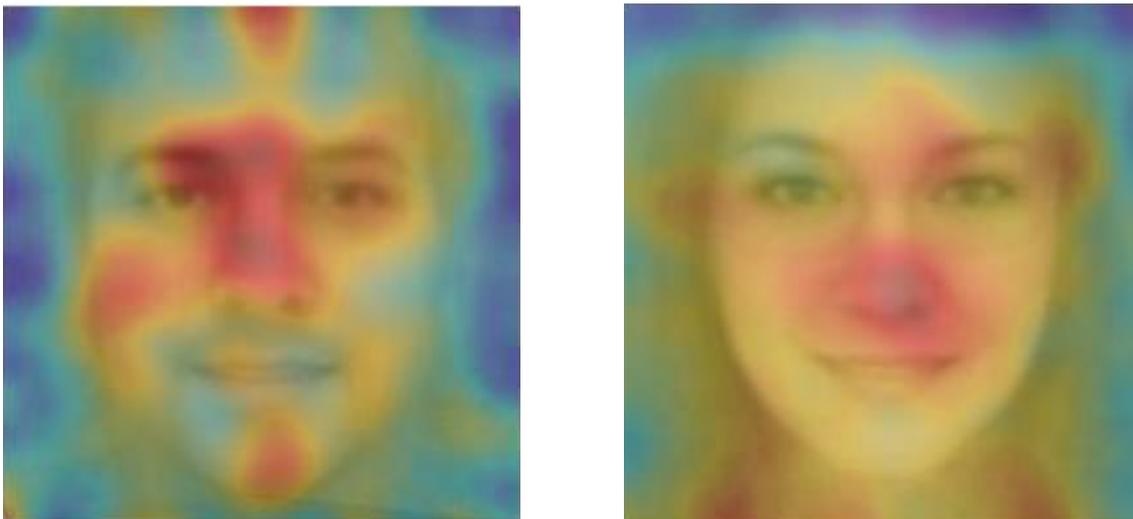
331 **Analysis.** We used the sexual orientation classifiers to estimate the probability of being
332 gay for the faces in the samples used here. Next, an area of 7 x 7 pixels in the top-left corner was
333 masked in all 100 images and the probability of being gay was estimated again. The procedure
334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

336 probability of being gay, resulting from masking a given area of the image, was used as a proxy
337 for the importance of a given area to the prediction of sexual orientation.

338 **Results**

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

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



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

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

353

354 **Study 1c: Facial Features Predictive of Sexual Orientation**

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

359 **Methods**

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

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

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

374 **Results**

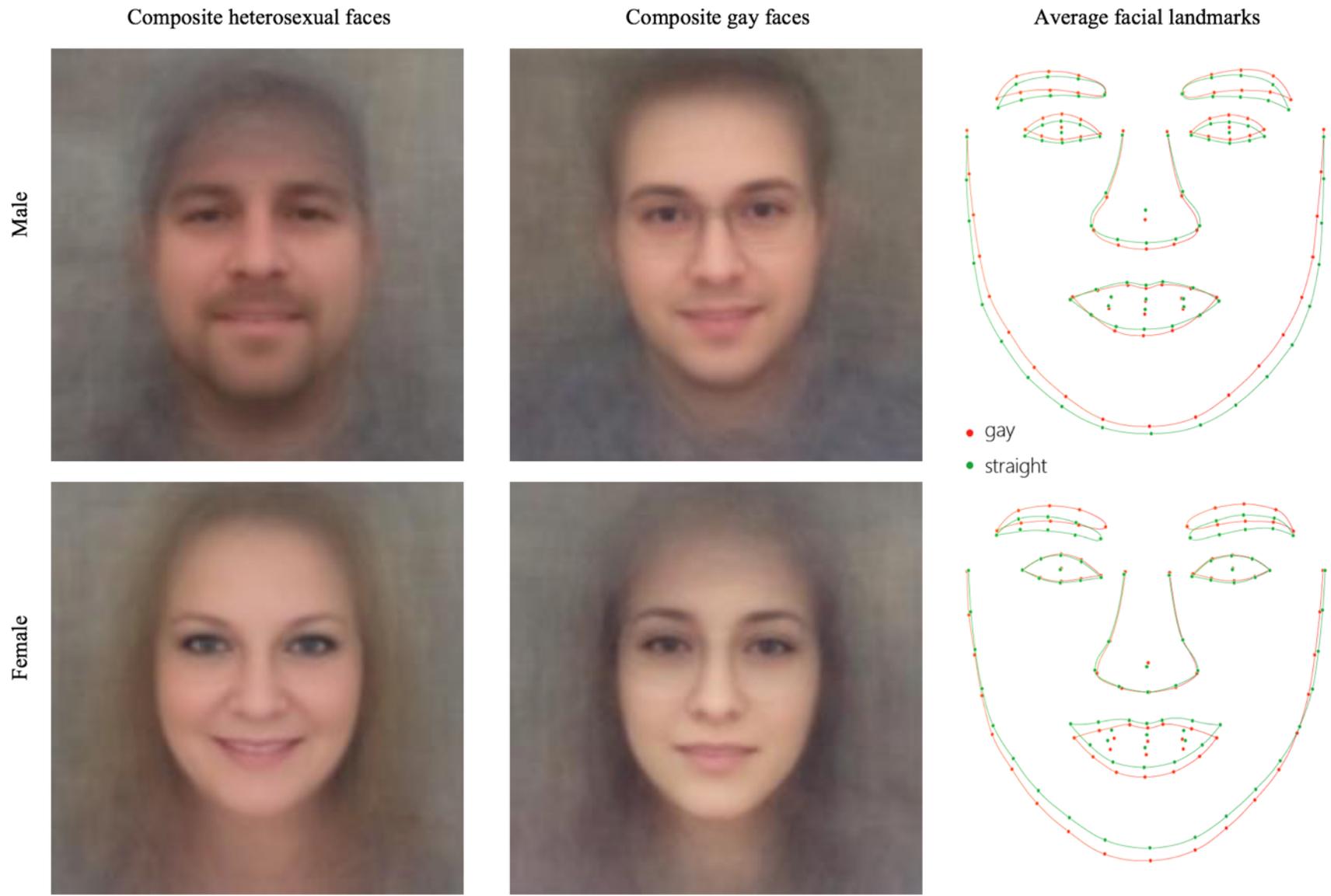
375 Figure 4 shows the average landmark locations and aggregate appearance of the faces
376 classified as most and least likely to be gay. Consistent with the PHT, gay faces tended to be
377 gender atypical. Average landmark locations revealed that gay men had narrower jaws and longer
378 noses, while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

382 The gender atypicality of gay faces extended beyond morphology. Gay men had less
383 facial hair, suggesting differences in androgenic hair growth, grooming style, or both. They also
384 had lighter skin, suggesting potential differences in grooming, sun exposure, and/or testosterone
385 levels.⁶ Lesbians tended to use less eye makeup, had darker hair, and wore less revealing clothes
386 (note the higher neckline), indicating less gender-typical grooming and style. Furthermore,
387 although women tend to smile more in general (Halberstadt, Hayes, & Pike, 1988), lesbians
388 smiled less than their heterosexual counterparts. Additionally, consistent with the association
389 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).

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

DEEP NEURAL NETWORKS CAN DETECT SEXUAL ORIENTATION FROM FACES



393

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

395 Study 2: Gender Atypicality of Gay People's Faces

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

400 Methods

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

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

411 Results

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

415 = 0.20; $p < .001$; 95% CI [0.19, 0.21]).⁷ The opposite was true for women ($r = -0.21$; $p < .001$; 95%
416 CI [-0.21, -0.20]).

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

419 **Study 3: Morphology-Based Classifier**

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

424 **Methods**

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

426 **Extracting morphological features.** We extracted morphological features from the
427 coordinates of the 83 landmarks outlining important facial features provided by Face++ (see
428 Figure 1). To subsume the shape of a given facial feature, such as the nose, we computed
429 Euclidean distances between the landmarks belonging to that feature. For example, as there are
430 10 landmarks outlining the nose (see Figure 1), its morphology was subsumed by a vector of 10
431 $\times 9 = 90$ Euclidean distances. To account for the differing sizes of the faces in facial images, the
432 distances were normalized by dividing them by the distance between the pupils.

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

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

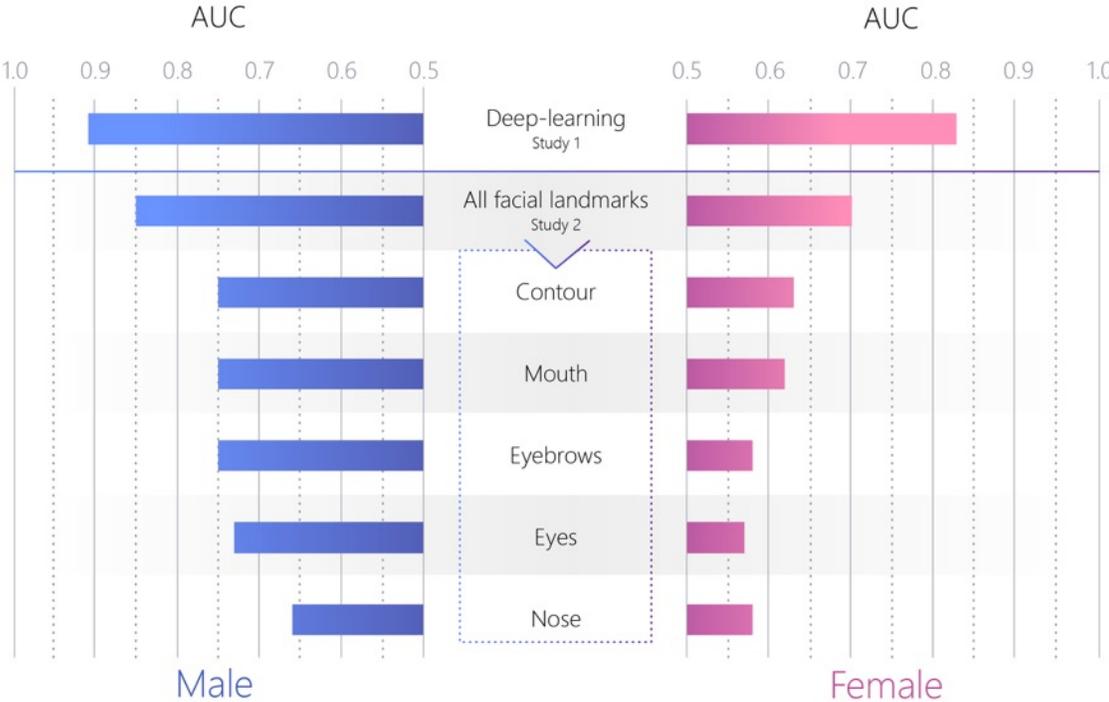
436 **Training classifiers.** The classifiers were trained, separately for each facial element and
437 for all facial landmarks combined, following a procedure similar to the one used in Study 1a.
438 Here, however, we used Euclidean distances instead of the VGG-Face scores as independent
439 variables.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

442 **Results**

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

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

455



456
457 *Figure 5.* The accuracy of the landmark-based classifiers, when provided with five images per
458 person. The accuracy of the DNN-based classifier trained in Study 1a is displayed on top of the
459 figure for comparison.

460 **Study 4: Human Judges**

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

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

472 **Methods**

473 **Facial images.** The 35,326 faces from Study 1a were randomly paired, resulting in
474 50,000 pairs for each gender (each face could be assigned to multiple pairs).

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

482 **Results**

483 Human judges achieved an accuracy of $AUC=.61$ for male images and $AUC=.54$ for
484 female images. This is comparable with the accuracy obtained in the previous studies, which
485 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.

486 2009). It is also compatible with the findings of Study 1a, which show that female faces are less
487 revealing of sexual orientation. Finally, it demonstrates that the facial images used in our study
488 were not unusually revealing of sexual orientation (at least to humans).

489 **Study 5: Beyond Dating Website Facial Images**

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

493 **Methods**

494 **Facial images.** We obtained a sample of 14,438 facial images of 6,075 openly gay men
495 from the myPersonality database (Kosinski et al., 2015). Gay males were identified using two
496 variables. First, we used the Facebook Audience Insights platform⁹ to identify 50 Facebook
497 Pages most popular among gay men, including Pages such as: “I love being Gay,” “Manhunt,”
498 “Gay and Fabulous,” and “Gay Times Magazine.” Second, we used the “interested in” field of
499 users’ Facebook profiles, which reveals the gender of the people that a given user is interested in.
500 Males that indicated an interest in other males, and that liked at least two out of the
501 predominantly gay Facebook Pages, were labeled as gay. Among the gay men identified in this
502 way, and for whom relationship data was available, 96% reported that their significant other was
503 male. Unfortunately, we were not able to reliably identify heterosexual Facebook users.

504 Those images were preprocessed and their VGG-Face scores extracted using the
505 procedure described in Study 1a. The final sample contained n=918 facial images of unique
506 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>

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

509 **Results**

510 We applied the classifier trained in Study 1a (employing the VGG-Face scores as an
511 independent variable) to distinguish between the faces of male gay Facebook users, male
512 heterosexual dating-website users, and male gay dating-website users. The classifier could
513 accurately distinguish between gay Facebook users and heterosexual dating-website users in
514 74% of cases, but was virtually unable to distinguish between gay Facebook users and gay
515 dating-website users (53%). This demonstrates that the classifier trained in Study 1a can
516 correctly identify facial images of gay men obtained in a different environment. It also shows
517 that this classifier is largely insensitive to the origin of the image, as it was unable to distinguish
518 between gay Facebook users and gay dating website users.

519 **General Discussion**

520 The findings reported in this work show that our faces contain more information about
521 sexual orientation than can be perceived or interpreted by the human brain. Study 1a showed that
522 facial features extracted by a DNN can be used to accurately identify the sexual orientation of
523 both men and women. Study 1b showed that the predictions are based on the facial area and not
524 the background. Study 1c revealed that the faces of gay men and lesbians had gender-atypical
525 features, as predicted by the PHT. This was corroborated by the results of Study 2 showing that
526 the probability of being gay was positively correlated with facial femininity among males and
527 negatively correlated with female facial femininity. The high accuracy of the classifier based on
528 the shape of facial elements, presented in Study 3, confirmed that much of the information about
529 sexual orientation is retained in fixed facial features, such as the facial contour or shape of the

530 nose. Study 4 revealed that the non-standardized facial images used in Study 1a were not
531 especially revealing of sexual orientation—at least to human judges, whose accuracy was the
532 same as in previous studies, some of which employed images of neutral faces taken in a carefully
533 controlled environment. Study 5 further corroborated these results by showing that the DNN-
534 based classifier developed in Study 1a performs similarly when presented with facial images of
535 gay men collected in a different environment.

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

551 Identifying links between facial features and psychological traits by employing
552 methodology similar to the one used here could boost our understanding of the origins and nature

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

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

575 gay people in the population correctly identified as gay). Aiming for high precision reduces
576 recall, and vice versa.

577 Let us illustrate this trade-off in a simulated scenario based on the results presented in this
578 work. We simulated a sample of 1,000 men by randomly drawing participants, and their
579 respective probabilities of being gay, from the sample used in Study 1a. As the prevalence of
580 same-gender sexual orientation among men in the U.S. is about 6–7% (Sell, Wells, & Wypij,
581 1995), we drew 70 probabilities from the gay participants, and 930 from the heterosexual
582 participants. We only considered participants for whom at least 5 facial images were available;
583 note that the accuracy of the classifier in their case reached an $AUC = .91$. If one selected 100
584 random males from this sample, 7 are expected to be gay. However, among the 100 of
585 individuals with the highest probability of being gay according to the classifier, 47 were gay. In
586 other words, the classifier provided for a nearly seven-fold improvement in precision over a
587 random draw ($47/7 = 6.71$). The precision could be further increased by narrowing the targeted
588 subsample. Among 30 males with the highest probability of being gay, 23 were gay, an eleven-
589 fold improvement in precision over a random draw ($23/2.1 = 11$). Finally, among the top 10
590 individuals with the highest probability of being gay, 9 were indeed gay: a thirteen-fold
591 improvement in precision over a random draw.

592 This study has a number of limitations. We used nonstandardized images characterized by
593 varying quality, head orientation, or facial expression. This provides for higher ecological
594 validity and a larger, more representative sample, but also introduces confounders (as discussed
595 in Study 1a). Additionally, as the images were obtained from a dating website, they might have
596 been especially revealing of sexual orientation. We believe that we sufficiently addressed this
597 problem by employing a model specifically trained to focus on non-transient facial features

598 (Study 1a), by showing that facial features enabling the prediction were consistent with the
599 theory (PHT; Studies 1c and 2), and by making sure that the images used here were not
600 substantially more revealing of sexual orientation than images of neutral faces taken in a
601 controlled setting (Study 4) or images obtained from Facebook (Study 5). Another issue pertains
602 to the quality of the ground truth: it is possible that some of the users categorized as heterosexual
603 were, in fact, gay or bisexual (or vice versa). However, we believe that people voluntarily
604 seeking partners on the dating website have little incentive to misrepresent their sexual
605 orientation. Furthermore, if some of the users were, in fact, wrongly labelled, correcting such
606 errors would likely *boost* the accuracy of the classifiers examined here. Additionally, despite our
607 attempts to obtain a more diverse sample, we were limited to studying white participants from
608 the U.S. As the prejudice against gay people and the adoption of online dating websites is
609 unevenly distributed across groups characterized by different ethnicities, we could not find
610 sufficient numbers of non-white gay participants. We believe, however, that our results will
611 likely generalize beyond the population studied here. They are consistent with the PHT of sexual
612 orientation, which was supported by variety of studies of humans and other mammals (Hines,
613 2010). As the exposure to gender-atypical androgen levels is likely to affect the faces of people
614 of different races to a similar degree, it is likely that their facial features are equally revealing of
615 sexual orientation. Finally, it is possible that individuals with more discernibly gay faces are
616 more likely to “come out.” If true, a classifier trained on the faces of openly gay users would be
617 less accurate when detecting non-openly gay individuals. While we do not have data to test this
618 hypothesis, it must be noted that coming out depends on many social, cultural, and legal factors.
619 Users who came out in our sample may wish or need to maintain their privacy in many contexts

620 and places. Thus, while some faces might be less revealing, many others may prevent their
621 owners from controlling their privacy of sexual orientation.

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

639 Some people may wonder if such findings should be made public lest they inspire the
640 very application that we are warning against. We share this concern. However, as the
641 governments and companies seem to be already deploying face-based classifiers aimed at
642 detecting intimate traits (Chin & Lin, 2017; Lubin, 2016), there is an urgent need for making

643 policymakers, the general public, and gay communities aware of the risks that they might be
644 facing already. Delaying or abandoning the publication of these findings could deprive
645 individuals of the chance to take preventive measures and policymakers the ability to introduce
646 legislation to protect people. Moreover, this work does not offer any advantage to those who may
647 be developing or deploying classification algorithms, apart from emphasizing the ethical
648 implications of their work. We used widely available off-the-shelf tools, publicly available data,
649 and methods well known to computer vision practitioners. We did not create a privacy-invading
650 tool, but rather showed that basic and widely used methods pose serious privacy threats. We hope
651 that our findings will inform the public and policymakers, and inspire them to design
652 technologies and write policies that reduce the risks faced by homosexual communities across
653 the world.¹⁰

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

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

662 studies show (e.g., Kosinski et al., 2013), such willingly shared digital footprints can be used to
663 reveal intimate traits. Thus, while we certainly need to urgently work towards policies and
664 technologies aimed at protecting privacy, in the long term the further erosion of privacy seems
665 inevitable. Consequently, the safety of gay and other minorities who may be ostracized in some
666 cultures will hinge on the tolerance of societies and governments. The postprivacy world will be
667 a much safer and hospitable place if inhabited by well-educated, tolerant people who are
668 dedicated to equal rights.

669

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

678

679 References

- 680 Allen, L. S., & Gorski, R. A. (1992). Sexual orientation and the size of the anterior commissure
681 in the human brain. *Proceedings of the National Academy of Sciences*, *89*(15), 7199–7202.
682 <http://doi.org/10.1073/pnas.89.15.7199>
- 683 Ambady, N., Hallahan, M., & Conner, B. (1999). Accuracy of judgments of sexual orientation
684 from thin slices of behavior. *Journal of Personality and Social Psychology*, *77*(3), 538–547.
685 <http://doi.org/10.1037/0022-3514.77.3.538>
- 686 Astley, S. J., Stachowiak, J., Clarren, S. K., & Clausen, C. (2002). Application of the fetal
687 alcohol syndrome facial photographic screening tool in a foster care population. *The*
688 *Journal of Pediatrics*, *141*(5), 712–717. <http://doi.org/10.1067/mpd.2002.129030>
- 689 Ballew, C. C., & Todorov, A. (2007). Predicting political elections from rapid and unreflective
690 face judgments. *Proceedings of the National Academy of Sciences of the United States of*
691 *America*, *104*(46), 17948–17953. <http://doi.org/10.1073/pnas.0705435104>
- 692 Berry, D. S. (1991). Accuracy in social perception: Contributions of facial and vocal information.
693 *Journal of Personality and Social Psychology*, *61*(2), 298–307. [http://doi.org/10.1037/0022-](http://doi.org/10.1037/0022-3514.61.2.298)
694 [3514.61.2.298](http://doi.org/10.1037/0022-3514.61.2.298)
- 695 Berry, D. S., & Brownlow, S. (1989). Were the physiognomists right? Personality correlates of
696 facial babyishness. *Personality and Social Psychology Bulletin*, *15*(2), 266–279.
697 <http://doi.org/10.1177/0146167289152013>
- 698 Bond, C. F., Berry, D. S., & Omar, A. (1994). The kernel of truth in judgments of deceptiveness.
699 *Basic and Applied Social Psychology*, *15*(4), 523–534.
700 http://doi.org/10.1207/s15324834basp1504_8

- 701 Borkenau, P., Brecke, S., Möttig, C., & Paelecke, M. (2009). Extraversion is accurately
702 perceived after a 50-ms exposure to a face. *Journal of Research in Personality*, *43*(4), 703–
703 706. <http://doi.org/10.1016/j.jrp.2009.03.007>
- 704 Brambilla, M., Riva, P., & Rule, N. O. (2013). Familiarity increases the accuracy of categorizing
705 male sexual orientation. *Personality and Individual Differences*, *55*(2), 193–195.
706 <http://doi.org/10.1016/j.paid.2013.02.023>
- 707 Brown, E., & Perrett, D. I. (1993). What gives a face its gender? *Perception*, *22*(7), 829–840.
708 <http://doi.org/10.1068/p220829>
- 709 Bulygina, E., Mitteroecker, P., & Aiello, L. (2006). Ontogeny of facial dimorphism and patterns
710 of individual development within one human population. *American Journal of Physical*
711 *Anthropology*, *131*(3), 432–443. <http://doi.org/10.1002/ajpa.20317>
- 712 Chin, J., & Lin, L. (2017, June 26). China's all-seeing surveillance state is reading its citizens'
713 faces. *The Wall Street Journal*. Retrieved from <https://www.wsj.com>
- 714 Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017).
715 Dermatologist-level classification of skin cancer with deep neural networks. *Nature*,
716 *542*(7639), 115–118. <http://doi.org/10.1038/nature21056>
- 717 Ferry, Q., Steinberg, J., Webber, C., FitzPatrick, D. R., Ponting, C. P., Zisserman, A., & Nellåker,
718 C. (2014). Diagnostically relevant facial gestalt information from ordinary photos. *eLife*,
719 *2014*(3). <http://doi.org/10.7554/eLife.02020.001>
- 720 Glaser, G. (2002). *The Nose: A Profile of Sex, Beauty, and Survival*. Simon and Schuster.
- 721 Glimcher, M. E., Garcia, R. I., & Szabó, G. (1978). Organ culture of mammalian skin and the
722 effects of ultraviolet light and testosterone on melanocyte morphology and function.

- 723 *Journal of Experimental Zoology*, 204(2), 229–237. <http://doi.org/10.1002/jez.1402040210>
- 724 Halberstadt, A. G., Hayes, C. W., & Pike, K. M. (1988). Gender and gender role differences in
725 smiling and communication consistency. *Sex Roles*, 19(9–10), 589–604.
726 <http://doi.org/10.1007/BF00289738>
- 727 Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning* (Vol. 1).
728 New York: Springer series in statistics. <http://doi.org/10.1007/b94608>
- 729 He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition.
730 *arXiv Preprint arXiv:1512.03385v1*, 7(3), 171–180.
731 <http://doi.org/10.3389/fpsyg.2013.00124>
- 732 Hines, M. (2010). Sex-related variation in human behavior and the brain. *Trends in Cognitive*
733 *Sciences*, 14(10), 448–456. <http://doi.org/10.1016/j.tics.2010.07.005>
- 734 Hughes, S. M., & Bremme, R. (2011). The effects of facial symmetry and sexually-dimorphic
735 facial proportions on assessments of sexual orientation. *Journal of Social, Evolutionary, and*
736 *Cultural Psychology*, 5(4), 214–230. <http://doi.org/10.1037/h0099261>
- 737 Jablonski, N. G., & Chaplin, G. (2000). The evolution of human skin coloration. *Journal of*
738 *Human Evolution*, 39(1), 57–106. <http://doi.org/10.1006/jhev.2000.0403>
- 739 Jannini, E. A., Blanchard, R., Camperio-Ciani, A., & Bancroft, J. (2010). Male homosexuality:
740 Nature or culture? *The Journal of Sexual Medicine*, 7(10), 3245–3253.
741 <http://doi.org/10.1111/j.1743-6109.2010.02024.x>
- 742 Jenkinson, J. (1997). Face facts: A history of physiognomy from ancient Mesopotamia to the end
743 of the 19th century. *The Journal Of Biocommunication*, 24(3), 2–7.

- 744 Jernigan, C., & Mistree, B. F. (2009). Gaydar: Facebook friendships expose sexual orientation.
 745 *First Monday, 14*(10). <http://doi.org/10.5210/fm.v14i10.2611>
- 746 Jones, B. C., Hahn, A. C., Fisher, C. I., Wincenciak, J., Kandrik, M., Roberts, S. C., ...
 747 DeBruine, L. M. (2015). Facial coloration tracks changes in women's estradiol.
 748 *Psychoneuroendocrinology, 56*, 29–34. <http://doi.org/10.1016/j.psyneuen.2015.02.021>
- 749 King, M., Semlyen, J., Tai, S. S., Killaspy, H., Osborn, D., Popelyuk, D., & Nazareth, I. (2008).
 750 A systematic review of mental disorder, suicide, and deliberate self harm in lesbian, gay and
 751 bisexual people. *BMC Psychiatry, 8*, 70. <http://doi.org/10.1186/1471-244X-8-70>
- 752 Kosinski, M., Matz, S. C., Gosling, S. D., Popov, V., & Stillwell, D. J. (2015). Facebook as a
 753 research tool for the social sciences: Opportunities, challenges, ethical considerations, and
 754 practical guidelines. *American Psychologist, 70*(6), 543–556.
 755 <http://doi.org/10.1037/a0039210>
- 756 Kosinski, M., Stillwell, D. J., & Graepel, T. (2013). Private traits and attributes are predictable
 757 from digital records of human behavior. *Proceedings of the National Academy of Sciences*
 758 *of the United States of America, 110*(15), 5802–5805.
 759 <http://doi.org/10.1073/pnas.1218772110>
- 760 LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444.
 761 <http://doi.org/10.1038/nature14539>
- 762 Lefevre, C. E., Lewis, G. J., Perrett, D. I., & Penke, L. (2013). Telling facial metrics: Facial
 763 width is associated with testosterone levels in men. *Evolution and Human Behavior, 34*(4),
 764 273–279. <http://doi.org/10.1016/j.evolhumbehav.2013.03.005>
- 765 Lefevre, C. E., Wilson, V. A. D., Morton, F. B., Brosnan, S. F., Paukner, A., & Bates, T. C.

- 766 (2014). Facial width-to-height ratio relates to alpha status and assertive personality in
767 capuchin monkeys. *PLOS ONE*, 9(4), e93369. <http://doi.org/10.1371/journal.pone.0093369>
- 768 Little, A. C., Burriss, R. P., Jones, B. C., & Roberts, S. C. (2007). Facial appearance affects
769 voting decisions. *Evolution and Human Behavior*, 28(1), 18–27.
770 <http://doi.org/10.1016/j.evolhumbehav.2006.09.002>
- 771 Löhmus, M., Sundström, L. F., & Björklund, M. (2009). Dress for success: Human facial
772 expressions are important signals of emotions. *Annales Zoologici Fennici*, 46(1), 75–80.
773 <http://doi.org/10.5735/086.046.0108>
- 774 Lombroso, C. (1911). *Criminal man*. New York: GP Putnam.
- 775 Lu, C., & Tang, X. (2014). Surpassing human-level face verification performance on LFW with
776 GaussianFace. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial*
777 *Intelligence* (pp. 3811–3819). Austin, TX: AAAI Press. Retrieved from
778 <http://arxiv.org/abs/1404.3840>
- 779 Lubin, G. (2016, October 12). Facial-profiling could be dangerously inaccurate and biased,
780 experts warn. *Business Insider*. Retrieved from <http://www.businessinsider.de/>
- 781 Lukaszewski, A. W., & Roney, J. R. (2011). The origins of extraversion: Joint effects of
782 facultative calibration and genetic polymorphism. *Personality and Social Psychology*
783 *Bulletin*, 37(3), 409–421. <http://doi.org/10.1177/0146167210397209>
- 784 Lyons, M., Lynch, A., Brewer, G., & Bruno, D. (2014). Detection of sexual orientation
785 (“gaydar”) by homosexual and heterosexual women. *Archives of Sexual Behavior*, 43(2),
786 345–352. <http://doi.org/10.1007/s10508-013-0144-7>
- 787 Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about

- 788 others. *Annual Review of Psychology*, *51*, 93–120.
789 <http://doi.org/10.1146/annurev.psych.51.1.93>
- 790 Meyer-Bahlburg, H. F. L. (1984). Psychoendocrine research on sexual orientation: Current status
791 and future options. *Progress in Brain Research*, *61*, 375–298. [http://doi.org/10.1016/S0079-](http://doi.org/10.1016/S0079-6123(08)64448-9)
792 [6123\(08\)64448-9](http://doi.org/10.1016/S0079-6123(08)64448-9)
- 793 Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). Deep Face Recognition. In X. Xie, M. W.
794 Jones, & G. K. L. Tam (Eds.), *Proceedings of the British Machine Vision Conference* (p.
795 41.1). Swansea, UK: BMVA Press. <http://doi.org/10.5244/C.29.41>
- 796 Penton-Voak, I. S., Pound, N., Little, A. C., & Perrett, D. I. (2006). Personality judgments from
797 natural and composite facial images: More evidence for a “kernel of truth” in social
798 perception. *Social Cognition*, *24*(5), 607–640. <http://doi.org/10.1521/soco.2006.24.5.607>
- 799 Porter, R. (Ed.). (2003). *The Cambridge history of science: Eighteenth-century science* (Vol. 4).
800 Cambridge, UK: Cambridge University Press.
- 801 Rhodes, G. (2006). The evolutionary psychology of facial beauty. *Annual Review of Psychology*,
802 *57*, 199–226. <http://doi.org/10.1146/annurev.psych.57.102904.190208>
- 803 Riedweg, C. (2005). *Pythagoras: His life, teaching, and influence*. New York, NY: Cornell
804 University Press.
- 805 Roberts, T., & Bruce, V. (1988). Feature saliency in judging the sex and familiarity of faces.
806 *Perception*, *17*(4), 475–481. <http://doi.org/10.1068/p170475>
- 807 Rule, N. O., & Ambady, N. (2008). Brief exposures: Male sexual orientation is accurately
808 perceived at 50 ms. *Journal of Experimental Social Psychology*, *44*(4), 1100–1105.
809 <http://doi.org/10.1016/j.jesp.2007.12.001>

- 810 Rule, N. O., & Ambady, N. (2010). Democrats and Republicans can be differentiated from their
 811 faces. *PLoS ONE*, 5(1), e8733. <http://doi.org/10.1371/journal.pone.0008733>
- 812 Rule, N. O., Ambady, N., Adams, R., & Macrae, C. N. (2008). Accuracy and awareness in the
 813 perception and categorization of male sexual orientation. *Journal of Personality and Social*
 814 *Psychology*, 95(5), 1019–1028. <http://doi.org/10.1037/a0013194>
- 815 Rule, N. O., MacRae, C. N., & Ambady, N. (2009). Ambiguous group membership is extracted
 816 automatically from faces. *Psychological Science*, 20(4), 441–443.
 817 <http://doi.org/10.1111/j.1467-9280.2009.02314.x>
- 818 Samochowiec, J., Wänke, M., & Fiedler, K. (2010). Political ideology at face value. *Social*
 819 *Psychological and Personality Science*, 1(3), 206–213.
 820 <http://doi.org/10.1177/1948550610372145>
- 821 Sell, R. L., Wells, J. A., & Wypij, D. (1995). The prevalence of homosexual behavior and
 822 attraction in the United States, the United Kingdom and France: Results of national
 823 population-based samples. *Archives of Sexual Behavior*, 24(3), 235–248.
 824 <http://doi.org/10.1007/BF01541598>
- 825 Skerski, J. (2011). Tomboy chic: Re-fashioning gender rebellion. *Journal of Lesbian Studies*,
 826 15(4), 466–479. <http://doi.org/10.1080/10894160.2011.532031>
- 827 Skorska, M. N., Geniole, S. N., Vrysen, B. M., McCormick, C. M., & Bogaert, A. F. (2015).
 828 Facial structure predicts sexual orientation in both men and women. *Archives of Sexual*
 829 *Behavior*, 44(5), 1377–1394. <http://doi.org/10.1007/s10508-014-0454-4>
- 830 Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from
 831 faces predict election outcomes. *Science*, 308(5728), 1623–1626.

- 832 <http://doi.org/10.1126/science.1110589>
- 833 Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of
 834 faces on social dimensions. *Trends in Cognitive Sciences*, *12*(12), 455–460.
 835 <http://doi.org/10.1016/j.tics.2008.10.001>
- 836 Tskhay, K. O., & Rule, N. O. (2015). Emotions facilitate the communication of ambiguous group
 837 memberships. *Emotion*, *15*(6), 812–826. <http://doi.org/10.1037/emo0000077>
- 838 Udry, J. R. (2000). Biological limits of gender construction. *American Sociological Review*,
 839 *65*(3), 443–457. <http://doi.org/10.2307/2657466>
- 840 Udry, J. R., & Chantala, K. (2006). Masculinity–femininity predicts sexual orientation in men
 841 but not in women. *Journal of Biosocial Science*, *38*(6), 797–809.
 842 <http://doi.org/10.1017/S002193200500101X>
- 843 UN Human Rights Council. (2015). *Discrimination and violence against individuals based on*
 844 *their sexual orientation and gender identity*. Retrieved from
 845 <http://www.refworld.org/docid/5571577c4.html>
- 846 Valentova, J. V., Kleisner, K., Havlíček, J., & Neustupa, J. (2014). Shape differences between the
 847 faces of homosexual and heterosexual men. *Archives of Sexual Behavior*, *43*(2), 353–361.
 848 <http://doi.org/10.1007/s10508-013-0194-x>
- 849 Whitehouse, A. J. O., Gilani, S. Z., Shafait, F., Mian, A., Tan, D. W., Maybery, M. T., ...
 850 Eastwood, P. (2015). Prenatal testosterone exposure is related to sexually dimorphic facial
 851 morphology in adulthood. *Proceedings of the Royal Society B: Biological Sciences*,
 852 *282*(1816), 20151351. <http://doi.org/10.1098/rspb.2015.1351>
- 853 Zebrowitz, L. A. (1997). *Reading faces: Window to the soul? New directions in social*

854 *psychology*. Boulder, CO: Westview Press.

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

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

861

Identify Adult Caucasian Males

Instructions

You will see 50 sets of 4 faces. Your job is to select **complete** faces belonging to **adult Caucasians males**. Any given set can contain between 0 to 4 adult male Caucasian faces.

You can use Back and Next button to navigate through different sets. **Please use the best of your intuition. We will carefully review the results to identify spammers.**

We welcome your feedback! There are going to be more HITs like these!

Details

1. Some images might contain a grey space on the side. It's normal and shouldn't affect your selections.
2. Some faces might be blurry. As long as you can recognize that the image represents an adult Caucasian male, the face should be accepted.
3. Faces partially covered by hats, sunglasses and hair are considered complete as long as you can recognize an adult Caucasian male.

Examples

<p>Wrong: non-Caucasian face</p> <p>Black</p> 	<p>Wrong: non-Caucasian face</p> <p>clearly Latino</p> 	<p>Wrong: non-adult face</p> <p>baby</p> 
<p>Wrong: non-male face</p> <p>female-looking face</p> 	<p>Wrong: incomplete face</p> <p>part of face</p> 	<p>Wrong: non-human face</p> <p>cartoon or not a human</p> 
<p>Correct: Caucasian, adult, male and complete face</p> 		

863
 864 *Figure S1.* Instructions given to AMT workers employed to remove incomplete, non-Caucasian,
 865 nonadult, and nonhuman male faces. We used similar instructions for female faces. (Faces
 866 presented here are not the actual faces used in this task.)

Which one is more likely to be **straight** (heterosexual)?

Instructions:

You will see 20 pairs of faces. Your job is to select the person that is more likely to be **straight** (heterosexual) by clicking on the corresponding image.

You can use **Back** and **Next** button to navigate through different pairs. **Please use the best of your intuition. We will carefully review the results to identify spammers.**

We welcome your feedback! There are going to be more HITs like these!

We want to have a larger pool of workers in these tasks. Please don't do more than 5 HITs. Thanks!



Back 1 / 2 Next

867

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