

Accepted refereed manuscript of: Lee AJ, Mitchem DG, Wright MJ, Martin NG, Keller MC & Zietsch BP (2016) Facial averageness and genetic quality: testing heritability, genetic correlation with attractiveness, and the paternal age effect. *Evolution and Human Behavior*, 37 (1), pp. 61-66.
DOI: <https://doi.org/10.1016/j.evolhumbehav.2015.08.003>
© 2015, Elsevier. Licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International <http://creativecommons.org/licenses/by-nc-nd/4.0/>

1 Facial averageness and genetic quality: Testing heritability, genetic correlation with attractiveness,
2 and the paternal age effect

3

4 Authors:

5 Anthony J. Lee¹, Dorian G. Mitchem^{2,3}, Margaret J. Wright⁴, Nicholas G. Martin⁴, Matthew C.
6 Keller^{2,3}, Brendan P. Zietsch^{1,4}

7

8 Author affiliations:

9

10 ¹ School of Psychology, University of Queensland, Brisbane, Queensland, Australia.

11 ² Department of Psychology and Neuroscience, University of Colorado Boulder, Boulder, Colorado,
12 United States of America.

13 ³ Institute for Behavioral Genetics, University of Colorado Boulder, Boulder, Colorado, United
14 States of America.

15 ⁴ QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia.

16

17 Corresponding authors:

18 Anthony J. Lee: anthony.lee@uqconnect.edu.au

19 Brendan P. Zietsch: zietsch@psy.uq.edu.au

20

21

22 Word Count: 5,834 Words

23

24 Keywords:

25 Mate preference; physical attractiveness; good genes; mutation load; developmental stability; twins

26

Abstract

27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45

Popular theory suggests that facial averageness is preferred in a partner for genetic benefits to offspring. However, whether facial averageness is associated with genetic quality is yet to be established. Here, we computed an objective measure of facial averageness for a large sample ($N = 1,823$) of identical and nonidentical twins and their siblings to test two predictions from the theory that facial averageness reflects genetic quality. First, we use biometrical modelling to estimate the heritability of facial averageness, which is necessary if it reflects genetic quality. We also test for a genetic association between facial averageness and facial attractiveness. Second, we assess whether paternal age at conception (a proxy of mutation load) is associated with facial averageness and facial attractiveness. Our findings are mixed with respect to our hypotheses. While we found that facial averageness does have a genetic component, and a significant phenotypic correlation exists between facial averageness and attractiveness, we did not find a genetic correlation between facial averageness and attractiveness (therefore, we cannot say that the genes that affect facial averageness also affect facial attractiveness) and paternal age at conception was not negatively associated with facial averageness. These findings support some of the previously untested assumptions of the ‘genetic benefits’ account of facial averageness, but cast doubt on others.

46 Facial averageness and genetic quality: Testing heritability, genetic correlation with attractiveness,
47 and the paternal age effect

48

49 Facial averageness is thought to be attractive in a mate (Grammer & Thornhill, 1994;
50 Komori, Kawamura, & Ishihara, 2009; Penton-Voak et al., 2001). This preference has been found
51 across cultures (Apicella, Little, & Marlowe, 2007; Rhodes, Yoshikawa, et al., 2001) and appears to
52 be more important than (and independent of) other traits such as facial symmetry or feature size
53 (Baudouin & Tiberghien, 2004; Valentine, Darling, & Donnelly, 2004). However, the mechanism
54 for this preference for facial averageness is unclear. The predominant theory is that facial
55 averageness reflects “good genes”, that is, heritable genetic quality. By mating with individuals
56 who possess good genes the associated advantages could then be inherited by offspring, increasing
57 the survival and/or reproduction of the offspring. As a result, individuals may have evolved to
58 attend to cues of genetic quality, such as facial averageness, when making mate choice decisions
59 (Gangestad & Simpson, 2000; Little, Jones, & DeBruine, 2011; Roberts & Little, 2008).

60 Facial averageness is commonly thought to represent good genes through resistance to
61 developmental instability, which is the sensitivity to perturbations during development (Polak,
62 2003). This theory stipulates that these perturbations disrupt development in random ways, which
63 can manifest in facial development as deviations from the average face shape of the population. In
64 this way, individuals who possess more average facial features are thought to have the good genetic
65 health required to withstand disruptions during development; therefore, mating with facially
66 average individuals could confer these genetic benefits to mutual offspring.

67 One source of perturbations an individual may confront during development can include
68 random environmental insults such as exposure to pathogens or diseases (Grammer & Thornhill,
69 1994; Rhodes, Zebrowitz, et al., 2001). Supporting this notion, average faces are perceived by
70 others as more healthy compared to less average faces (Grammer & Thornhill, 1994; Rhodes,
71 Zebrowitz, et al., 2001; Zebrowitz & Rhodes, 2004). Another source of perturbations may include

72 the effects of random genetic mutations. Random genetic mutations are often harmful and can
73 contribute to many forms of physical and mental health (Bray, Gunnell, & Smith, 2006). One
74 contributing factor to an individual's accumulation of genetic mutations (mutation load) is thought
75 to be paternal age at conception (Crow, 2000). This is because males continually produce sperm
76 throughout the lifespan (as opposed to women who are born with their full supply of eggs). Sperm
77 production requires continual cell divisions and chromosome replications, which is a process
78 susceptible to errors that lead to aberrations or mutations; therefore, the sperm of older males,
79 which have gone through more replications, are more likely to have accumulated more mutations
80 than the sperm of younger males. Indeed, Huber and Fieder (2014) found in a large sample ($N =$
81 8,434) that paternal, but not maternal, age at conception was negatively associated with facial
82 attractiveness, suggesting that facial information may be used as a cue of an individual's mutation
83 load.

84 Despite the popularity of facial averageness reflecting genetic quality in the literature, only
85 circumstantial evidence supports the notion that these preferences exist for indirect benefits. Also,
86 whether facial averageness confers indirect benefits is based on an assumption that has not been
87 adequately tested: if facial averageness were preferred because of genetic benefits to offspring, a
88 substantial proportion of the variance in this trait must be due to additive genetic sources.
89 Otherwise, contrary to popular theory, facial averageness could not reflect good genes as it could
90 not be inherited by offspring. Another possibility is that facial averageness represents a sexy-sons
91 trait, that is, facial averageness may have once reflected indirect benefits to offspring viability in
92 our evolutionary history but is now solely maintained by an exaggerated preference driven by genes
93 that improve offspring attractiveness (Fisher, 1930). In this case, we should still expect a heritable
94 additive genetic component.

95 Despite the importance of this assumption that facial averageness is heritable, it has never
96 been tested. Doing so would strongly inform the question of whether facial averageness reflects
97 genetic quality or is instead preferred for other reasons. For instance, facial averageness could

98 instead be preferred for more direct benefits, such as disease avoidance (assuming facial
99 averageness is in fact associated with good health). Another alternative is that preference for
100 average faces may simply reflect a more general sensory bias for prototypical faces/objects
101 (Halberstadt & Rhodes, 2000, 2003) rather than being an adaptive mate choice mechanism. Neither
102 of the latter scenarios requires a significant heritable genetic component for facial averageness,
103 whereas the good genes explanation does require it.

104 More fundamentally, it has not been well established that facial averageness is actually
105 associated with attractiveness in naturally occurring faces, which is an important prerequisite for
106 establishing its evolutionary significance. When investigating facial averageness, previous research
107 has often used computer-generated composite faces as stimuli (e.g., Apicella et al., 2007; Rhodes,
108 Yoshikawa, et al., 2001). While this has the advantage of controlling extraneous factors, composite
109 faces can also often appear artificial and also smooth/blend textural and colour imperfections,
110 spuriously increasing facial attractiveness ratings. One study that did investigate the effect of
111 natural variation in facial averageness on attractiveness was Komori et al. (2009), where objective
112 measures of facial shape averageness were computed from landmark coordinates derived from
113 facial photographs. Here a significant negative correlation was found between facial distinctiveness
114 (the inverse of facial averageness) and facial attractiveness, though these correlations were modest
115 at best ($r = -.08$ and $r = -.13$ for men and women respectively).

116 Here we compute an objective measure of facial averageness for a large sample of identical
117 and nonidentical (same-sex and opposite-sex) twins and their siblings using geometric
118 morphometrics (the statistical analysis of shape). We then use this measure in two analyses
119 designed to test predictions from the idea that facial averageness reflects genetic quality. First, we
120 extend the work of Huber and Fieder (2014) and assess whether paternal age at conception (as a
121 proxy of mutation load) is associated with facial averageness and facial attractiveness. Second, we
122 use biometrical modelling to estimate the heritability (proportion of between-individual variation
123 that is due to genes) of facial averageness in order to assess if these traits could reflect genetic

124 quality. We also test for a genetic association between facial averageness and facial attractiveness,
125 which is necessary if facial averageness is (or once was) preferred for indirect benefits.

126

127 Method

128

129 *Participants*

130 Participants were 1698 twin individuals (304 monozygotic (MZ) twin pairs, 479 dyzygotic
131 (DZ) twin pairs) and 125 of their siblings from 913 families who took part in either the Brisbane
132 Adolescent Twin Studies (BATS, $N = 1321$) located in Queensland, Australia (Wright & Martin,
133 2004) or from the Longitudinal Twin Study (LTS, $N = 502$) located in Colorado, USA (Mitchem et
134 al., 2013; Rhea, Gross, Haberstick, & Corley, 2013). For participants who were part of BATS, twins
135 were tested (and photographs taken) as close as possible to their 16th birthday ($M = 16.03$ years, SD
136 $= .46$ years) and their siblings as close as possible to their 18th birthday ($M = 17.67$ years, $SD =$
137 $.1.22$). When available, the ages of participants' parents at birth were also collected for these twins
138 (maternal age $N = 1199$, range = 17.91-42.22 years, parental age $N = 1153$, range = 17.80-60.87
139 years). Participants from the LTS were older than participants from the BATS ($M = 22.06$ years, SD
140 $= 1.29$ years).

141

142 *Facial Photographs and Landmark Coordinates*

143 For twins who were part of BATS, photographs of participants were taken between the years
144 of 1996 to 2010. In the earliest waves of data collection, photographs were taken using film
145 cameras, and later scanned to digital format. Photographs from later waves were taken on digital
146 cameras. We note that photographs of these participants were not originally taken for shape
147 analysis. As such, variation existed between photographs that could alter the shape information
148 captured by the landmarks (e.g., the participant's head angle facing the camera, or the participant's
149 facial expression). To reduce any influence this may have, photographs were rotated manually to be

150 level, and participants looking askance were removed from analysis. However, we assume that this
151 type of variation is idiosyncratic between photographs and would therefore simply add error
152 variance rather than biasing the results in any particular direction. For participants from the LTS,
153 photographs were taken between 2001-2010. Participants were asked to adopt a neutral facial
154 expression and to face the camera directly. All photographs were taken under standard indoor
155 lighting conditions.

156 Thirteen independent raters (7 males, 6 females) identified a total of 31 landmarks for each
157 face. Raters were trained for several weeks in hour-long sessions where landmarks were defined
158 using anatomical definitions. See Figure 1. for descriptions of each landmark; landmarks were
159 chosen as they were easily identifiable and would capture important shape information of each
160 facial component (e.g., eyes, nose, overall face shape). Two raters were randomly chosen for each
161 participant, and the coordinates were calculated as the mean pixel location from these two raters.

162

163 - INSERT FIGURE 1. HERE -

164

165 *Facial Averageness Scores*

166 In order to calculate scores for facial averageness, we first computed participants' facial
167 distinctiveness (the inverse of facial averageness) from landmark coordinates. We used concepts
168 from geometric morphometrics, which is the statistical analysis of shape through landmark
169 coordinates (Bookstein, 1991; Zelditch, Swiderski, Sheets, & Fink, 2004). Shape is defined as
170 differences between objects that are not due to translation, size, or rotation, and therefore
171 encapsulates all other information such as distances and angles between different landmarks.

172 A Generalised Procrustes Analysis (GPA; Zelditch et al., 2004) was conducted on raw x-
173 and y-coordinates. This procedure removes translation effects (position of the object in the shape
174 space) by standardising to a common shape space, size effects by standardising centroid size to one,
175 and rotational effects by minimising root of the summed squared distances (the total Procrustes

176 distance) between homologous landmarks between faces. This produces new coordinates that
177 purely represent shape information. For full details of GPA and shape analysis via geometric
178 morphometrics, see Zelditch et al. (2004).

179 We computed facial distinctiveness scores by comparing each individual's landmark
180 configurations with the mean coordinates of the sample using a similar method as detailed in
181 Komori et al. (2009). Since average faces are inherently more symmetrical (Rhodes, Sumich, &
182 Byatt, 1999), we control for facial symmetry by reflecting landmarks on each side of the face onto
183 the other and averaging the corresponding left-right landmark coordinates – this was done for each
184 individual and the average face. An Ordinary Procrustes Analysis was then conducted between the
185 average configuration and each individual, which compares each individual with the average face
186 configuration and calculates the total Procrustes distance between homologous landmarks. This
187 Procrustes distance for each individual is conceptually similar to a linear combination of absolute
188 deviation from the average face; thus, this value was used as the facial distinctiveness score. We
189 then reverse coded the scores so that larger scores indicated greater facial averageness. This process
190 of calculating facial averageness was done separately for males and females. Outliers on facial
191 averageness ($\pm 3 SD$ from the mean) were deleted from all analyses (14 males, 2 females).

192

193 *Ratings of Facial Attractiveness*

194 Observers rated each facial photograph on facial attractiveness. Twenty-three undergraduate
195 research assistants (10 males, 13 females; $M = 21.27$ years, $SD = 3.13$; different individuals from
196 those who identified the facial landmarks) were presented a subset of the photos in a random order
197 and rated all faces on attractiveness. Ratings were given on a 7-point scale (1 = low attractiveness, 7
198 = high attractiveness). Raters were not given instructions on how to judge attractiveness and inter-
199 rater agreement for attractiveness was moderate (intraclass correlation = .43, $p < .001$). Facial
200 attractiveness ratings computed from only male and only female raters correlated very highly with
201 facial attractiveness computed from all raters ($r = .94$ for male raters and $r = .93$ for females); given

202 the high concordance, and that the facial attractiveness scores from all raters contained substantially
203 less measurement error, we used this score for all analyses. For more detail on the rating process
204 and further analyses of observer ratings, see Mitchem et al. (2013).

205

206 *Statistical Analysis*

207 Identical twins share all their genes whereas nonidentical twins and siblings share on
208 average half of their segregating genes, while all twins/siblings completely share the family
209 environment. As such, we were able to partition the variation in facial averageness scores into three
210 of four sources: additive genetic (A, when the effects of genes on a phenotype sum additively), non-
211 additive genetic (D; when the effect on a phenotype relies on an interaction between genes, e.g.,
212 dominance or epistasis), shared environmental (C; when environmental factors are shared between
213 both twins, e.g., shared household factors), and residual (E; e.g., idiosyncratic environmental
214 sources, or measurement error) sources. C and D are negatively confounded (C works to increase
215 twin correlations, while D works to decrease the association); therefore, only one of these can be
216 estimated based on the size of the DZ twin pair correlation in relations to MZ twin pair correlation,
217 as per standard procedure (Neale & Cardon, 1992; Posthuma et al., 2003). As is standard for twin-
218 family designs, biometrical modelling was conducted using maximum likelihood modelling, which
219 determines the combination of A, C, D, and E that best matches the observed data (i.e. means,
220 variances, and twin/sibling pair correlations). For further detail of twin analysis, see (Neale &
221 Cardon, 1992; Posthuma et al., 2003). All biometric modelling was conducted in the OpenMx
222 software package. As is standard in twin modelling, differences in means and twin/sibling
223 correlations across different zygosity groups were tested by equating the relevant parameters in the
224 model and testing the change in model fit (distributed as χ^2) against the change in degrees of
225 freedom (which equals the change in the number of parameters estimated). Age and year tested
226 were included as covariates in all analyses, effectively partialling out any influence of these

227 variables. Facial attractiveness and averageness scores did not significantly differ between the
228 BATS and LTS samples; therefore, samples were combined for all analyses.

229

230

Results

231

Facial Averageness and Facial Attractiveness

232

233 If facial averageness is (or once was) preferred for potential indirect benefits, then we would
234 expect an association with rated attractiveness. As predicted, greater facial averageness was
235 positively associated with increased attractiveness rating for both females ($r = .16$, $CI = .10, .22$)
236 and males ($r = .09$, $CI = .02, .16$). These values for both men and women are similar to those
237 previously found when using geometric morphometrics to calculate facial averageness (Komori et
238 al., 2009).

239

240 Even though we find a positive correlation between facial averageness and attractiveness,
241 this apparent association could be due to some unknown third variable that is correlated with both
242 facial averageness and attractiveness. Therefore, we conducted a mediation analysis to determine
243 whether this association was specifically due to shape information. This was done by first
244 modelling via regression ratings of facial attractiveness using shape variables (i.e., the
245 decomposition of Procrustes coordinates into principal components) as the predictor variables.
246 Therefore, each individual's predicted score based on this model essentially represents their
247 attractiveness score based solely on shape information. Then, we tested whether this shape
248 component of facial attractiveness mediated the relationship between facial averageness and rated
249 facial attractiveness.

249

250 Regressions were conducted separately for males and females. To extract the shape
251 component of facial attractiveness, all shape variables that explained $> 1\%$ of the total variation in
252 face shape (15 for males, 16 for females) were entered simultaneously in the regression with rated
253 facial attractiveness as the dependent variable. Overall, these regression equations significantly

253 predicted rated facial attractiveness ($R^2 = .09, p < .001$ for males, $R^2 = .07, p < .001$ for females).
254 From the regression equation, we could compute each individual's predicted attractiveness based on
255 the individual's landmark-based face shape. This score represents the shape component of each
256 individual's facial attractiveness.

257 Contrary to predictions, the association between facial averageness and the shape
258 component of facial attractiveness was non-significant for both men and women ($r = .06, p = .093$
259 for males, $r = .01, p = .796$ for females). A follow-up mediation analyses found that the shape
260 component of facial attractiveness did not significantly mediate the association between facial
261 averageness and overall facial attractiveness for men (Sobel's $Z = 1.55, p = .119$) or women
262 (Sobel's $Z = .27, p = .785$). These results suggest that shape facial averageness may not be
263 important when evaluating facial attractiveness, and that the significant association may be
264 explained through other factors. This mediation is shown in Figure 2.

265

266 - INSERT FIGURE 2. HERE -

267

268 While not the main focus of this paper, previous work has indicated that facial averageness
269 may be associated with facial sexual dimorphism (Rhodes et al., 2007). In previous papers, we
270 computed objective scores of facial sexual dimorphism from the facial photographs and also had
271 them rated for subjective facial masculinity/femininity (for further detail, see Lee et al., 2014;
272 Mitchem et al., 2013). When comparing these scores with facial averageness scores calculated here,
273 we found no significant association with either objective sexual dimorphism ($r = -.05, CI = -.13,$
274 $.03,$ and $r = .02, CI = -.06, .12$ for males and females respectively) nor rated facial
275 masculinity/femininity ($r = .03, CI = -.04, .10,$ and $r = -.01, CI = -.08, .05$ for males and females
276 respectively). We also tested whether controlling for objective facial sexual dimorphism
277 significantly influenced the association of facial averageness and attractiveness, though this did not

278 have a substantial impact ($r = .08$, $CI = .01, .15$, and $r = .13$, $CI = 08, -.19$ for males and females
279 respectively).

280

281 *Paternal Age*

282 To assess whether facial averageness and facial attractiveness are associated with mutation
283 load, we ran a regression analysis with paternal age at birth. Similar to Huber and Fieder (2014), we
284 included participant sex, age and maternal age as covariates. We also included the extra covariate of
285 the year a participant's photo was taken. Results from the regression analyses are reported in Table
286 1. We found a positive association between paternal age at birth and facial attractiveness; this is in
287 the opposite direction to that found by Huber and Fieder (2014). We also found no significant
288 association between paternal age at birth and facial averageness, which does not support the notion
289 that facial averageness is associated with mutation load.

290

291 - INSERT TABLE 1. HERE -

292

293 *Twin Modelling*

294 Preliminary tests found that there were no significant differences between twins and siblings
295 in means and variances on facial averageness scores ($\chi^2 (2) = .12$, $p = .941$, and $\chi^2 (2) = 1.97$, $p =$
296 $.373$ for means and variances respectively) suggesting that there was nothing unusual about twins
297 on facial averageness. Also, there were no differences in facial averageness scores between men
298 and women given that they were calculated and standardised separately. Therefore, all analyses
299 conducted equated scores between twins and siblings, and between men and women. Table 2.
300 shows the twin correlations for facial averageness across different zygosity groups. Overall,
301 correlations for across all MZ twin pairs were significantly larger than that for all DZ twin pairs (χ^2
302 $(1) = 9.37$, $p < .005$) indicating genetic variation in facial averageness.

303

304 - INSERT TABLE 2. HERE -

305

306 Correlations between MZ twin pairs on facial averageness were significant, while those
307 between DZ twin pairs were not significant (as shown in Table 2.). The correlation for MZ twin
308 pairs was more than twice the correlation for DZ twin pairs; therefore, in-line with standard
309 procedure, an ADE model was estimated (Neale & Cardon, 1992; Posthuma et al., 2003). Estimated
310 components are reported in Table 3. A significant genetic component (A + D) was found,
311 suggesting that variation in facial averageness is influenced by genes; however, neither A nor D
312 was significant individually – this is a frequent consequence of the low power to statistically
313 distinguish A from D (Keller, Medland, & Duncan, 2010).

314

315 - INSERT TABLE 3. HERE -

316

317 In order to determine the common genetic variance shared between facial averageness and
318 attractiveness, we ran a common factors bivariate model. Since A and D could not be clearly
319 distinguished in the univariate model for facial averageness, we only estimated A and E
320 components in the bivariate model, in which case D variance is absorbed into the A estimate. In the
321 bivariate model, neither males nor females exhibited a significant genetic correlation between facial
322 averageness and attractiveness. This does not support the notion that facial averageness is
323 associated with genetic quality. There was, however, a significant environmental correlation
324 between facial averageness and attractiveness. The correlated factors model is reported in Table 4.

325

326 - INSERT TABLE 4. HERE -

327

328

Discussion

329

330 The predominant theory regarding preference for facial averageness is that it represents
331 genetic quality. We tested this directly by evaluating whether facial averageness has a heritable
332 component that could be passed down to offspring, and whether facial averageness is associated
333 with paternal age at birth, which is thought to be associated with mutation load. Our findings are
334 mixed with respect to our hypotheses.

335 On the one hand, we show facial averageness does have a genetic component, which is
336 necessary if facial averageness confers indirect benefits by either representing a good genes or
337 sexy-sons trait. While the estimates of additive and nonadditive genetic effects were individually
338 imprecise and differed between men, women, and the overall sample, the overall genetic component
339 (A + D) was highly significant and fairly similar in men and women. We note, however, that the
340 genetic component accounts for only around 24% of the variation in facial averageness – that is,
341 most of the variance appears to be due to non-familial factors (e.g. environmental perturbations
342 during development, as well as measurement error), and thus any interpretation supporting indirect
343 benefits should be made cautiously.

344 We also found significant phenotypic correlations between facial averageness and
345 attractiveness in both sexes, consistent with previous theory and research. If facial averageness does
346 (or once did) represent indirect benefits to offspring, then facial averageness must be preferred in a
347 partner in naturally occurring faces. Indeed, our effect sizes are similar to those previously found
348 when objective measures of averageness were computed from facial photographs (Komori et al.,
349 2009). However, we did not find significant correlations between facial averageness and the shape
350 component of facial attractiveness for either men nor women. Also, we did not find that the shape
351 component of facial attractiveness significantly mediated the relationships between facial
352 averageness (which was solely computed from shape information) and facial attractiveness ratings.
353 This gives insight into whether the shape component of facial averageness itself is important when
354 evaluating facial attractiveness, or whether other correlates, such as colour or textural information,
355 may be more important. Pertinent to this, we found that the year photographs were taken was a

356 large predictor of attractiveness rating, possibly suggesting that raters were influenced by cues such
357 as photo quality or hairstyle, when making attractiveness ratings. This is particularly important
358 given previous research has often used composite faces to assess preference for facial averageness,
359 which can confound shape averageness with the blending of idiosyncratic textural and colour
360 information.

361 On the other hand, the genetic correlation between facial averageness and attractiveness was
362 not significant in either sex or overall, meaning we cannot say that the genes that affect facial
363 averageness also affect facial attractiveness. This is contrary to what we would expect if
364 averageness reflected genetic quality. It could be that a genetic correlation exists but we did not
365 have sufficient power to detect it - the overall heritability estimates for facial averageness and the
366 phenotypic correlation between facial averageness and attractiveness were modest to begin with,
367 which suggests that the genetic correlation would be difficult to detect if it did exist. However, it
368 should be noted that the corresponding environmental correlation was significant in the overall
369 sample.

370 Furthermore, we did not see the predicted negative correlation between facial averageness or
371 facial attractiveness and paternal age, contrary to the hypothesis that the greater mutation load in
372 older sperm would be reflected in less average faces. In fact, our finding that paternal age at birth is
373 positively associated with facial attractiveness is in the opposite direction to that found in Huber
374 and Fieder (2014). A possible explanation for why we did not find an effect is that any effect of
375 increased mutation load associated with paternal age may not have a substantial effect on facial
376 attractiveness; de novo mutations are very small in number and we would expect an even smaller
377 differential between those from young and old fathers (an increase of about two mutations per year;
378 Kong et al., 2012). Indeed, it may be that ascertainment effects are generally stronger than the effect
379 of the extra mutations; that is, more attractive men might tend to have children (who inherit their
380 father's attractiveness) at a later age (perhaps due to their ability to attract younger women), thus

381 swamping any mutation load effect. Thus, paternal age at birth may not be a sensitive enough proxy
382 of mutation load to detect effects on facial traits.

383 Given that our results provide no clear support for the notion that facial averageness is
384 preferred for indirect benefits by representing either a good genes or sexy-sons trait, how might we
385 otherwise explain the association found between facial averageness and facial attractiveness
386 ratings? One possibility is that facial averageness may be preferred for more direct benefits. For
387 instance, assuming facial averageness is associated with resistance to perturbations such as
388 pathogens, individuals high in facial averageness may be less likely to succumb to illness, and
389 therefore less likely to transmit diseases to the choosing individual. Another possibility is that
390 preference may instead exist for traits correlated with shape facial averageness; this could include
391 other forms of facial averageness as discussed previously (e.g., colour averageness or textural
392 averageness), or other unrelated facial traits, such as sexual dimorphism (see Scheib, Gangestad, &
393 Thornhill, 1999). Alternatively, the association between facial averageness and attractiveness may
394 not reflect an evolved mechanism at all, but simply a more general sensory bias for prototypical
395 objects (Halberstadt & Rhodes, 2000, 2003).

396 A potential limitation is that a large proportion of photographs used in our study were of
397 twins when they were 16-years-old, which may not reflect scores on these facial attributes in
398 adulthood. However, previous theory stipulates that the effects of developmental instability should
399 occur in the early stages of life; therefore, the effect of genes of facial averageness should be
400 apparent at 16. Also, there was no significant difference in facial attributes scores between twins
401 and their older siblings, nor with the sample collected in the LTS suggesting these scores are
402 generalisable to an older population. Other limitations include standard caveats of the classical twin
403 design (Keller & Coventry, 2005; Keller et al., 2010); for instance, we are unable to fully
404 disentangle the separate effects of A and D. Further research could overcome this by including
405 other family members, such as parents.

406 In summary, our results provide mixed evidence with respect to the predominant theory that
407 facial averageness is preferred for genetic benefits to offspring. Despite finding that the objective
408 measure of facial averageness had a significant genetic component and was significantly associated
409 with facial attractiveness, the genetic component was not significantly shared between the two
410 traits, and we did not find a significant association with either facial trait and paternal age at birth.
411 Our findings support some of the previously untested assumptions of the ‘genetic benefits’ account
412 of facial averageness, but cast doubt on others. More research is needed to understand why
413 geometrically average faces are attractive.

414

415 Acknowledgements

416

417 We thank our twin sample for their participation; Ann Eldridge, Marlene Grace, Kerrie
418 McAloney, Daniel Park, Maura Caffrey, and Jacob McAloney for photograph collection and
419 processing; and David Smyth for IT support. We acknowledge support from the Australian
420 Research Council (A7960034, A79906588, A79801419, DP0212016, DP0343921, DP0664638,
421 DP1093900, FT0991360) and National Health & Medical Research Council (900536, 930223,
422 950998, 981339, 983002, 961061, 983002, 241944, 389875, 552485, 613608). AJL is supported by
423 an Australian Postgraduate Award, BPZ a Discovery Early Career Research Award, both from the
424 Australian Research Council, and MCK is supported by National Institutes of Mental Health grants
425 K01MH085812 and R01MH100141.

426

References

- 427
428
- 429 Apicella, C. L., Little, A. C., & Marlowe, F. W. (2007). Facial averageness and attractiveness in an
430 isolated population of hunter-gatherers. *Perception, 36*, 1813-1820.
- 431 Baudouin, J., & Tiberghien, G. (2004). Symmetry, averageness, and feature size in the facial
432 attractiveness of women. *Acta Psychologica, 117*, 313-332.
- 433 Bookstein, F. L. (1991). *Morphometric tools for landmark data*. Cambridge: Cambridge University
434 Press.
- 435 Bray, I., Gunnell, D., & Smith, G. D. (2006). Advanced paternal age: How old is too old? *Journal of*
436 *Epidemiology and Community Health, 60*, 851-853.
- 437 Crow, J. F. (2000). The origins, patterns and implications of human spontaneous mutation. *Nature*
438 *Reviews Genetics, 1*, 40-47.
- 439 Fisher, R. A. (1930). *The genetical theory of natural selection*: Clarendon.
- 440 Gangestad, S. W., & Simpson, J. A. (2000). The evolution of human mating: Trade-offs and
441 strategic pluralism. *Behavioural and Brain Sciences, 23*, 573-644.
- 442 Grammer, K., & Thornhill, R. (1994). Human (Homo-Sapiens) facial attractiveness and sexual
443 selection - the role of symmetry and averageness. *Journal of Comparative Psychology,*
444 *108*(3), 233-242.
- 445 Halberstadt, J., & Rhodes, G. (2000). The attractiveness of nonface averages: Implications for an
446 evolutionary explanation of the attractiveness of average faces. *Psychological Science,*
447 *11*(4), 285-289.
- 448 Halberstadt, J., & Rhodes, G. (2003). It's not just average faces that are attractive: Computer-
449 manipulated averageness makes birds, fish, and automobiles attractive. *Psychological*
450 *Bulletin & Review, 10*(1), 149-156.
- 451 Huber, S., & Fieder, M. (2014). Advanced paternal age is associated with lower facial
452 attractiveness. *Evolution and Human Behavior.*

- 453 Keller, M. C., & Coventry, W. L. (2005). Quantifying and Addressing Parameter Indeterminacy in
454 the Classical Twin Design. *Twin Research and Human Genetics*, 8(3), 201-213.
- 455 Keller, M. C., Medland, S. E., & Duncan, L. E. (2010). Are Extended Twin Family Designs Worth
456 the Trouble? A Comparison of the Bias, Precision, and Accuracy of Parameters Estimated in
457 Four Twin Family Models. *Behavior Genetics*, 40(3), 377-393.
- 458 Komori, M., Kawamura, S., & Ishihara, S. (2009). Averageness of symmetry: Which is more
459 important for facial attractiveness? *Acta Psychologica*, 131, 136-142.
- 460 Kong, A., Frigge, M. L., Masson, G., Besenbacher, S., Sulem, P., Magnusson, G., . . . Stefansson,
461 K. (2012). Rate of de novo mutations and the importance of father's age to disease risk.
462 *Nature*, 488(7412), 471-475.
- 463 Lee, A. J., Mitchem, D. G., Wright, M. J., Martin, N. G., Keller, M. C., & Zietsch, B. P. (2014).
464 Genetic factors increasing male facial masculinity decrease facial attractiveness of female
465 relatives. *Psychological Science*, 25(2), 476-484.
- 466 Little, A. C., Jones, B. C., & DeBruine, L. M. (2011). Facial attractiveness: evolutionary based
467 research. *Philosophical Transactions of the Royal Society B: Biological Sciences*,
468 366(1571), 1638-1659.
- 469 Mitchem, D. G., Purkey, A. M., Grebe, N. M., Carey, G., Garver-Apgar, C. E., Bates, T. C., . . .
470 Keller, M. C. (2013). Estimating the sex-specific effects of genes on facial attractiveness
471 and sexual dimorphism. *Behavior Genetics*, 44(3), 270-281.
- 472 Neale, M. C., & Cardon, L. R. (1992). *Methodology for genetic studies of twins and families*.
473 Boston: Kluwer.
- 474 Penton-Voak, I. S., Jones, B. C., Little, A. C., Baker, S., Tiddeman, B., Burt, D. M., & Perrett, D. I.
475 (2001). Symmetry, sexual dimorphism in facial proportions and male facial attractiveness.
476 *Proceedings of the Royal Society of London Series B-Biological Sciences*, 268(1476), 1617-
477 1623.
- 478 Polak, M. (2003). *Developmental instability: causes and consequences.*: Oxford University Press.

- 479 Posthuma, D., Beem, A. L., de Geus, E. J. C., van Baal, G. C. M., von Hjelmborg, J. B., Lachine, I.,
480 & Boomsma, D. I. (2003). Theory and practice in quantitative genetics. *Twin Research*, 6,
481 361-376.
- 482 Rhea, S., Gross, A. A., Haberstick, B. C., & Corley, R. P. (2013). Colorado Twin Registry - An
483 update. *Twin Research and Human Genetics*, 16(1), 351-357.
- 484 Rhodes, G., Sumich, A., & Byatt, G. (1999). Are average facial configurations attractive only
485 because of their symmetry? *Psychological Science*, 10(1), 52-58.
- 486 Rhodes, G., Yoshikawa, S., Clark, A., Lee, K., McKay, R., & Akamatsu, S. (2001). Attractiveness
487 of facial averageness and symmetry in non-Western cultures: In search of biologically based
488 standards of beauty. *Perception*, 30, 611-625.
- 489 Rhodes, G., Yoshikawa, S., Palermo, R., Simmons, L. W., Peters, M., Lee, K., . . . Crawford, J. R.
490 (2007). Perceived health contributes to the attractiveness of facial symmetry, averageness,
491 and sexual dimorphism. *Perception*, 36(8), 1244-1252.
- 492 Rhodes, G., Zebrowitz, L. A., Clark, A., Kalick, S. M., Hightower, A., & McKay, R. (2001). Do
493 facial averageness and symmetry signal health? *Evolution and Human Behavior*, 22(1), 31-
494 46.
- 495 Roberts, S. C., & Little, A. C. (2008). Good genes, complementary genes and human mate
496 preferences. *Genetica*, 132, 309-321.
- 497 Scheib, J. E., Gangestad, S. W., & Thornhill, R. (1999). Facial attractiveness, symmetry and cues of
498 good genes. *Proceedings of the Royal Society of London Series B-Biological Sciences*,
499 266(1431), 1913-1917.
- 500 Valentine, T., Darling, S., & Donnelly, M. (2004). Why are average faces attractive? The effect of
501 view and averageness on the attractiveness of female faces. *Psychonomic Bulletin &*
502 *Review*, 11(3), 482-487.
- 503 Wright, M. J., & Martin, N. G. (2004). Brisbane adolescent twin study: Outline of study methods
504 and research projects. *Australian Journal of Psychology*, 56(2), 65-78.

505 Zebrowitz, L. A., & Rhodes, G. (2004). Sensitivity to "bad genes" and the anomalous face
506 overgeneralization effect: cue validity, cue utilization, and accuracy in judging intelligence
507 and health. *Journal of Nonverbal Behavior*, 28(3), 167-185.

508 Zelditch, M. L., Swiderski, D. L., Sheets, H. D., & Fink, W. L. (2004). *Geometric morphometrics*
509 *for biologists: A primer*. New York and London: Elsevier Academic Press.

510

511

Figure Captions

512
513
514 Figure 1. Landmarks used to compute facial averageness from photographs.

515
516 Figure 2. Mediation analysis between computed facial averageness, rated attractiveness, and the
517 shape component of rated attractiveness.