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

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

Authors:

Anthony J. Lee[1], Dorian G. Mitchem[2,3], Margaret J. Wright[4], Nicholas G. Martin[4], Matthew C. Keller[2,3], Brendan P. Zietsch[1,4]

Author affiliations:

[1] School of Psychology, University of Queensland, Brisbane, Queensland, Australia.

[2] Department of Psychology and Neuroscience, University of Colorado Boulder, Boulder, Colorado, United States of America.

[3] Institute for Behavioral Genetics, University of Colorado Boulder, Boulder, Colorado, United States of America.

[4] QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia.

Corresponding authors:

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

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

Word Count: 5,834 Words

Keywords:

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

Abstract

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.

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

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

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

One source of perturbations an individual may confront during development can include random environmental insults such as exposure to pathogens or diseases (Grammer & Thornhill, 1994; Rhodes, Zebrowitz, et al., 2001). Supporting this notion, average faces are perceived by others as more healthy compared to less average faces (Grammer & Thornhill, 1994; Rhodes, Zebrowitz, et al., 2001; Zebrowitz & Rhodes, 2004). Another source of perturbations may include the effects of random genetic mutations. Random genetic mutations are often harmful and can contribute to many forms of physical and mental health (Bray, Gunnel, & Smith, 2006). One contributing factor to an individual's accumulation of genetic mutations (mutation load) is thought to be paternal age at conception (Crow, 2000). This is because males continually produce sperm throughout the lifespan (as opposed to women who are born with their full supply of eggs). Sperm production requires continual cell divisions and chromosome replications, which is a process susceptible to errors that lead to aberrations or mutations; therefore, the sperm of older males, which have gone through more replications, are more likely to have accumulated more mutations than the sperm of younger males. Indeed, Huber and Fieder (2014) found in a large sample (N = 8,434) that paternal, but not maternal, age at conception was negatively associated with facial attractiveness, suggesting that facial information may be used as a cue of an individual's mutation load.

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

Despite the importance of this assumption that facial averageness is heritable, it has never been tested. Doing so would strongly inform the question of whether facial averageness reflects genetic quality or is instead preferred for other reasons. For instance, facial averageness could instead be preferred for more direct benefits, such as disease avoidance (assuming facial averageness is in fact associated with good health). Another alternative is that preference for average faces may simply reflect a more general sensory bias for prototypical faces/objects (Halberstadt & Rhodes, 2000, 2003) rather than being an adaptive mate choice mechanism. Neither of the latter scenarios requires a significant heritable genetic component for facial averageness, whereas the good genes explanation does require it.

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

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

Method

Participants

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

Facial Photographs and Landmark Coordinates

For twins who were part of BATS, photographs of participants were taken between the years of 1996 to 2010. In the earliest waves of data collection, photographs were taken using film cameras, and later scanned to digital format. Photographs from later waves were taken on digital cameras. We note that photographs of these participants were not originally taken for shape analysis. As such, variation existed between photographs that could alter the shape information captured by the landmarks (e.g., the participant's head angle facing the camera, or the participant's facial expression). To reduce any influence this may have, photographs were rotated manually to be level, and participants looking askance were removed from analysis. However, we assume that this type of variation is idiosyncratic between photographs and would therefore simply add error variance rather than biasing the results in any particular direction. For participants from the LTS, photographs were taken between 2001-2010. Participants were asked to adopt a neutral facial expression and to face the camera directly. All photographs were taken under standard indoor lighting conditions.

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

- INSERT FIGURE 1. HERE -

Facial Averageness Scores

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

A Generalised Procrustes Analysis (GPA; Zelditch et al., 2004) was conducted on raw x- and y-coordinates. This procedure removes translation effects (position of the object in the shape space) by standardising to a common shape space, size effects by standardising centroid size to one, and rotational effects by minimising root of the summed squared distances (the total Procrustes distance) between homologous landmarks between faces. This produces new coordinates that purely represent shape information. For full details of GPA and shape analysis via geometric morphometrics, see Zelditch et al. (2004).

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

Ratings of Facial Attractiveness

Observers rated each facial photograph on facial attractiveness. Twenty-three undergraduate research assistants (10 males, 13 females; *M* = 21.27 years, *SD* = 3.13; different individuals from those who identified the facial landmarks) were presented a subset of the photos in a random order and rated all faces on attractiveness. Ratings were given on a 7-point scale (1 = low attractiveness, 7 = high attractiveness). Raters were not given instructions on how to judge attractiveness and inter-rater agreement for attractiveness was moderate (intraclass correlation = .43, p <. 001). Facial attractiveness ratings computed from only male and only female raters correlated very highly with facial attractiveness computed from all raters (r = .94 for male raters and r = .93 for females); given the high concordance, and that the facial attractiveness scores from all raters contained substantially less measurement error, we used this score for all analyses. For more detail on the rating process and further analyses of observer ratings, see Mitchem et al. (2013).

Statistical Analysis

Identical twins share all their genes whereas nonidentical twins and siblings share on average half of their segregating genes, while all twins/siblings completely share the family environment. As such, we were able to partition the variation in facial averageness scores into three of four sources: additive genetic (A, when the effects of genes on a phenotype sum additively), non-additive genetic (D; when the effect on a phenotype relies on an interaction between genes, e.g., dominance or epistasis), shared environmental (C; when environmental factors are shared between both twins, e.g., shared household factors), and residual (E; e.g., idiosyncratic environmental sources, or measurement error) sources. C and D are negatively confounded (C works to increase twin correlations, while D works to decrease the association); therefore, only one of these can be estimated based on the size of the DZ twin pair correlation in relations to MZ twin pair correlation, as per standard procedure (Neale & Cardon, 1992; Posthuma et al., 2003). As is standard for twin-family designs, biometrical modelling was conducted using maximum likelihood modelling, which determines the combination of A, C, D, and E that best matches the observed data (i.e. means, variances, and twin/sibling pair correlations). For further detail of twin analysis, see (Neale & Cardon, 1992; Posthuma et al., 2003). All biometric modelling was conducted in the OpenMx software package. As is standard in twin modelling, differences in means and twin/sibling correlations across different zygosity groups were tested by equating the relevant parameters in the model and testing the change in model fit (distributed as χ^2) against the change in degrees of freedom (which equals the change in the number of parameters estimated). Age and year tested were included as covariates in all analyses, effectively partialling out any influence of these variables. Facial attractiveness and averageness scores did not significantly differ between the BATS and LTS samples; therefore, samples were combined for all analyses.

Results

Facial Averageness and Facial Attractiveness

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

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

Regressions were conducted separately for males and females. To extract the shape component of facial attractiveness, all shape variables that explained > 1% of the total variation in face shape (15 for males, 16 for females) were entered simultaneously in the regression with rated facial attractiveness as the dependent variable. Overall, these regression equations significantly predicted rated facial attractiveness ($R^2 = .09$, $p < .001$ for males, $R^2 = .07$, $p < .001$ for females). From the regression equation, we could compute each individual's predicted attractiveness based on the individual's landmark-based face shape. This score represents the shape component of each individual's facial attractiveness.

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

- INSERT FIGURE 2. HERE -

While not the main focus of this paper, previous work has indicated that facial averageness may be associated with facial sexual dimorphism (Rhodes et al., 2007). In previous papers, we computed objective scores of facial sexual dimorphism from the facial photographs and also had them rated for subjective facial masculinity/femininity (for further detail, see Lee et al., 2014; Mitchem et al., 2013). When comparing these scores with facial averageness scores calculated here, we found no significant association with either objective sexual dimorphism ($r = -.05$, $CI = -.13$, .03, and $r = .02$, $CI = -.06$, .12 for males and females respectively) nor rated facial masculinity/femininity ($r = .03$, $CI = -.04$, .10, and $r = -.01$, $CI = -.08$, .05 for males and females respectively). We also tested whether controlling for objective facial sexual dimorphism significantly influenced the association of facial averageness and attractiveness, though this did not have a substantial impact (r = .08, *CI* = .01, .15, and r = .13, *CI* = 08, -.19 for males and females respectively).

Paternal Age

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

- INSERT TABLE 1. HERE -

Twin Modelling

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

- INSERT TABLE 2. HERE -

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

- INSERT TABLE 3. HERE -

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

- INSERT TABLE 4. HERE -

Discussion

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

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

We also found significant phenotypic correlations between facial averageness and attractiveness in both sexes, consistent with previous theory and research. If facial averageness does (or once did) represent indirect benefits to offspring, then facial averageness must be preferred in a partner in naturally occurring faces. Indeed, our effect sizes are similar to those previously found when objective measures of averageness were computed from facial photographs (Komori et al., 2009). However, we did not find significant correlations between facial averageness and the shape component of facial attractiveness for either men nor women. Also, we did not find that the shape component of facial attractiveness significantly mediated the relationships between facial averageness (which was solely computed from shape information) and facial attractiveness ratings. This gives insight into whether the shape component of facial averageness itself is important when evaluating facial attractiveness, or whether other correlates, such as colour or textural information, may be more important. Pertinent to this, we found that the year photographs were taken was a large predictor of attractiveness rating, possibly suggesting that raters were influenced by cues such as photo quality or hairstyle, when making attractiveness ratings. This is particularly important given previous research has often used composite faces to assess preference for facial averageness, which can confound shape averageness with the blending of idiosyncratic textural and colour information.

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

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

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

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

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

Acknowledgements

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

References

Apicella, C. L., Little, A. C., & Marlowe, F. W. (2007). Facial averageness and attractiveness in an isolated population of hunter-gatherers. *Perception, 36*, 1813-1820.

Baudouin, J., & Tiberghien, G. (2004). Symmetry, averageness, and feature size in the facial attractiveness of women. *Acta Psychologica, 117*, 313-332.

Bookstein, F. L. (1991). *Morphometric tools for landmark data*. Cambridge: Cambridge University Press.

Bray, I., Gunnel, D., & Smith, G. D. (2006). Advanced paternal age: How old is too old? *Journal of Epidemiology and Community Health, 60*, 851-853.

Crow, J. F. (2000). The origins, patterns and implications of human spontaneous mutation. *Nature Reviews Genetics, 1*, 40-47.

Fisher, R. A. (1930). *The genetical theory of natural selection*: Clarendon.

Gangestad, S. W., & Simpson, J. A. (2000). The evolution of human mating: Trade-offs and strategic pluralism. *Behavioural and Brain Sciences, 23*, 573-644.

Grammer, K., & Thornhill, R. (1994). Human (Homo-Sapiens) facial attractiveness and sexual selection - the role of symmetry and averageness. *Journal of Comparative Psychology, 108*(3), 233-242.

Halberstadt, J., & Rhodes, G. (2000). The attractiveness of nonface averages: Implications for an evolutionary explanation of the attractiveness of average faces. *Psychological Science, 11*(4), 285-289.

Halberstadt, J., & Rhodes, G. (2003). It's not just average faces that are attractive: Computer-manipulated averageness makes birds, fish, and automobiles attractive. *Psychological Bulletin & Review, 10*(1), 149-156.

Huber, S., & Fieder, M. (2014). Advanced paternal age is associated with lower facial attractiveness. *Evolution and Human Behavior*.

Keller, M. C., & Coventry, W. L. (2005). Quantifying and Addressing Parameter Indeterminacy in the Classical Twin Design. *Twin Research and Human Genetics, 8*(3), 201-213.

Keller, M. C., Medland, S. E., & Duncan, L. E. (2010). Are Extended Twin Family Designs Worth the Trouble? A Comparison of the Bias, Precision, and Accuracy of Parameters Estimated in Four Twin Family Models. *Behavior Genetics, 40*(3), 377-393.

Komori, M., Kawamura, S., & Ishihara, S. (2009). Averageness of symmetry: Which is more important for facial attractiveness? *Acta Psychologica, 131*, 136-142.

Kong, A., Frigge, M. L., Masson, G., Besenbacher, S., Sulem, P., Magnusson, G., . . . Stefansson, K. (2012). Rate of de novo mutations and the importance of father's age to disease risk. *Nature, 488*(7412), 471-475.

Lee, A. J., Mitchem, D. G., Wright, M. J., Martin, N. G., Keller, M. C., & Zietsch, B. P. (2014). Genetic factors increasing male facial masculinity decrease facial attractiveness of female relatives. *Psychological Science, 25*(2), 476-484.

Little, A. C., Jones, B. C., & DeBruine, L. M. (2011). Facial attractiveness: evolutionary based research. *Philosophical Transactions of the Royal Society B: Biological Sciences, 366*(1571), 1638-1659.

Mitchem, D. G., Purkey, A. M., Grebe, N. M., Carey, G., Garver-Apgar, C. E., Bates, T. C., . . . Keller, M. C. (2013). Estimating the sex-specific effects of genes on facial attractiveness and sexual dimorphism. *Behavior Genetics, 44*(3), 270-281.

Neale, M. C., & Cardon, L. R. (1992). *Methodology for genetic studies of twins and families.* Boston: Kluwer.

Penton-Voak, I. S., Jones, B. C., Little, A. C., Baker, S., Tiddeman, B., Burt, D. M., & Perrett, D. I. (2001). Symmetry, sexual dimorphism in facial proportions and male facial attractiveness. *Proceedings of the Royal Society of London Series B-Biological Sciences, 268*(1476), 1617-1623.

Polak, M. (2003). *Developmental instability: causes and consequences.*: Oxford University Press.

Posthuma, D., Beem, A. L., de Geus, E. J. C., van Baal, G. C. M., von Hjelmborg, J. B., Lachine, I., & Boomsma, D. I. (2003). Theory and practice in quantitative genetics. *Twin Research, 6*, 361-376.

Rhea, S., Gross, A. A., Haberstick, B. C., & Corley, R. P. (2013). Colorado Twin Registry - An update. *Twin Research and Human Genetics*, *16*(1), 351-357.

Rhodes, G., Sumich, A., & Byatt, G. (1999). Are average facial configurations attractive only because of their symmetry? *Psychological Science, 10*(1), 52-58.

Rhodes, G., Yoshikawa, S., Clark, A., Lee, K., McKay, R., & Akamatsu, S. (2001). Attractiveness of facial averageness and symmetry in non-Western cultures: In search of biologically based standards of beauty. *Perception, 30*, 611-625.

Rhodes, G., Yoshikawa, S., Palermo, R., Simmons, L. W., Peters, M., Lee, K., . . . Crawford, J. R. (2007). Perceived health contributes to the attractiveness of facial symmetry, averageness, and sexual dimorphism. *Perception, 36*(8), 1244-1252.

Rhodes, G., Zebrowitz, L. A., Clark, A., Kalick, S. M., Hightower, A., & McKay, R. (2001). Do facial averageness and symmetry signal health? *Evolution and Human Behavior, 22*(1), 31-46.

Roberts, S. C., & Little, A. C. (2008). Good genes, complementary genes and human mate preferences. *Genetica, 132*, 309-321.

Scheib, J. E., Gangestad, S. W., & Thornhill, R. (1999). Facial attractiveness, symmetry and cues of good genes. *Proceedings of the Royal Society of London Series B-Biological Sciences, 266*(1431), 1913-1917.

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

Wright, M. J., & Martin, N. G. (2004). Brisbane adolescent twin study: Outline of study methods and research projects. *Australian Journal of Psychology, 56*(2), 65-78.

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

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

Figure Captions

Figure 1. Landmarks used to compute facial averageness from photographs.

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