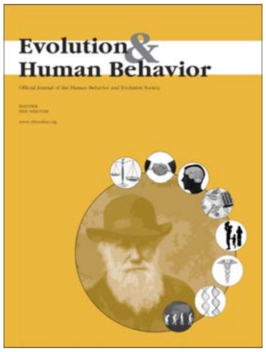
Accepted Manuscript



No Relationship Between Intelligence and Facial Attractiveness in a Large, Genetically Informative Sample

Dorian G. Mitchem, Brendan P. Zietsch, Margaret J. Wright, Nicholas G. Martin, John K. Hewitt, Matthew C. Keller

PII: S1090-5138(14)00151-2
DOI: doi: 10.1016/j.evolhumbehav.2014.11.009
Reference: ENS 5963

To appear in: *Evolution and Human Behavior*

Received date: 14 May 2014
Revised date: 21 November 2014
Accepted date: 30 November 2014

Please cite this article as: Mitchem, D.G., Zietsch, B.P., Wright, M.J., Martin, N.G., Hewitt, J.K. & Keller, M.C., No Relationship Between Intelligence and Facial Attractiveness in a Large, Genetically Informative Sample, *Evolution and Human Behavior* (2014), doi: 10.1016/j.evolhumbehav.2014.11.009

NO RELATIONSHIP BETWEEN INTELLIGENCE AND FACIAL ATTRACTIVENESS IN A LARGE, GENETICALLY INFORMATIVE SAMPLE

[Running Head: No relationship between intelligence and facial attractiveness]

Dorian G. Mitchem[1,2,*], Brendan P. Zietsch[3,4], Margaret J. Wright[3], Nicholas G. Martin[3], John K. Hewitt[1,2], Matthew C. Keller[1,2*]

1 Department of Psychology and Neuroscience, University of Colorado Boulder, Boulder, Colorado, United States of America.

2 Institute for Behavioral Genetics, University of Colorado Boulder, Boulder, Colorado, United States of America.

3 QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia.

4 School of Psychology, University of Queensland, Brisbane, Queensland, Australia.

* **Throughout the refereeing and publication process**: Correspondence concerning this article should be addressed to Dorian G. Mitchem, Department of Psychology and Neuroscience, Muenzinger Hall, 345 UCB, Boulder, CO, 80309. Email: dorian.mitchem@colorado.edu. Phone: +1 303-551-2090.

*** Post-publication**: Correspondence concerning this article should be addressed to Dorian G. Mitchem (as indicated above) or Matthew C. Keller, Department of Psychology and Neuroscience, Muenzinger Hall, 345 UCB, Boulder, CO, 80309. Email: matthew.c.keller@gmail.com. Phone: +1 303-735-5376.

Word count:
7,455 words, excluding tables and figure.

Abstract

Theories in both evolutionary and social psychology suggest that a positive correlation should exist between facial attractiveness and general intelligence, and several empirical observations appear to corroborate this expectation. Using highly reliable measures of facial attractiveness and IQ in a large sample of identical and fraternal twins and their siblings, we found no evidence for a phenotypic correlation between these traits. Likewise, neither the genetic nor the environmental latent factor correlations were statistically significant. We supplemented our analyses of new data with a simple meta-analysis that found evidence of publication bias among past studies of the relationship between facial attractiveness and intelligence. In view of these results, we suggest that previously published reports may have overestimated the strength of the relationship and that the theoretical bases for the predicted attractiveness-intelligence correlation may need to be reconsidered.

Key words: facial attractiveness; intelligence; genetic correlation; fitness trait; evolutionary genetics; twin and family study.

1. Introduction

Some evolutionary models predict that traits contributing to survival or reproductive success will tend to be positively correlated (Hansen, 2006; Rowe & Houle, 1996; but, see Falconer & Mackay, 1996; Lerner, 1954). Examples of such positive correlations have been observed in humans [IQ and sperm quality: Arden, Gottfredson, Miller, and Pierce (2009); IQ and height: Keller et al. (2013); birth rate, completed family size, and age at last childbirth: Kosova, Abney, and Ober (2009)], animals [energy storage and metabolic activity in *Drosophila melanogaster*: Clark (1990); body weight across environmental niches in *Alsophila pometaria*: Futuyma and Philippi (1987); activity metabolism and locomotor performance in *Thamnophilis sirtalis*: Garland (1988)], and plants [size and pest resistance in *Ipomoea purpurea*: Rausher and Simms (1989); life history and morphological traits in *Holcus lanatus*: Billington, Mortimer, and McNeilly (1988); life history and morphological traits in *Impatiens capensis*: Mitchell-Olds (1986)]. There are two basic types of explanation for why these correlations occur. One is that the conditions in the environment, such as pathogen levels or the availability of adequate nutrition, have similar effects on both of the correlated traits (Møller, 1997). The other is that the phenotypic correlation is caused by a correlation between the effects of the alleles influencing the two traits (Falconer & Mackay, 1996).

Genetic correlations, in turn, can come about in two principal ways. One is pleiotropy, whereby a gene affects multiple phenotypic characters. Pleiotropy is a common property of genes (Falconer & Mackay, 1996) and is a proposed explanation for genetic correlations between a large number of medical (Flint & Mackay, 2009; Solovieff, Cotsapas, Lee, Purcell, & Smoller, 2013) and psychological (Kovas & Plomin, 2006; Lee et al., 2013) traits, many of which appear to be highly polygenic (e.g., Davies et al., 2011; Purcell, Wray, Stone, & International

Schizophrenia Consortium, 2009; Stahl et al., 2012). Antagonistic pleiotropy, whereby alleles that improve one fitness-related trait deteriorate another fitness-related trait, can lead to stable genetic polymorphism and persistent negative genetic correlations between fitness-increasing traits. However, the conditions under which genetic polymorphism is maintained by antagonistic pleiotropy are restrictive (Hedrick, 1999; Prout, 2000), and most investigations in non-human animals have found positive rather than negative correlations between fitness-increasing traits (Roff, 1997). On the other hand, to the degree that the genetic variation in directionally selected traits is due to the aggregate effects of deleterious mutations across many loci (Houle, 1998), genetic correlations between fitness-increasing traits should be positive. Under this scenario, pleiotropic loci that affect two or more fitness-increasing traits should tend to harbor common alleles that are favored by selection and rare mutations that are selected against because they negatively affect both traits.

Even when the traits are affected by non-overlapping sets of genes, a second possible cause of genetic correlations is assortative mating on two or more traits simultaneously, which can lead to non-random associations between alleles at different loci (i.e., gametic phase disequilibrium; Crow & Felsenstein, 1968). To the degree that overall attractiveness is a composite of multiple sexually selected traits, positive assortment between mates on overall attractiveness necessarily implies positive cross-trait correlations between traits positively related to attractiveness. When this occurs, individuals who inherit alleles that increase the sexual attractiveness of one trait from one parent will be more likely to inherit alleles that increase the sexual attractiveness of the other trait from the other parent, leading to positive genetic correlations between sexually selected traits when scaled such that scores increase with attractiveness.

Both intelligence (Miller, 2000) and facial attractiveness (Gangestad, Thornhill, & Yeo, 1994; Thornhill & Gangestad, 1999) have been hypothesized to be sexually selected traits related to fitness, perhaps because their large mutational target sizes (Davies et al., 2011) reveal a partner's load of deleterious mutations (Gangestad & Yeo, 1997; Keller, 2007; Miller, 2000). If so, then as described above, there are two basic explanations for why facial attractiveness and intelligence might be expected to be positively genetically correlated. First, because these traits are influenced by a large number of genes, there is likely to be some degree of overlap between them. While such overlap could lead to negative genetic correlations from antagonistic pleiotropy, the restrictive conditions under which antagonistic pleiotropy can maintain negative genetic correlations at equilibrium (Hedrick, 1999; Prout, 2000) suggest that a better expectation is that pleiotropic loci lead to positive genetic correlations via transiently polymorphic, recurrent deleterious mutations that reduce both intelligence and facial attractiveness. Second, given that people rate both facial attractiveness and intelligence as desirable in romantic and sexual partners (Buss et al., 1990; Buss & Barnes, 1986; Kenrick, Sadalla, Groth, & Trost, 1990), it is also possible that cross-trait assortative mating (intelligent people choosing more facially attractive mates, and vice-versa) produces statistical associations between alleles affecting the two traits. These two possible causes of genetic correlations are not mutually exclusive; for example, Keller et al. (2013) used an extended twin-family design that accounted for the genetic effects of assortative mating to demonstrate that both processes contributed roughly equally to the genetic correlation between human height and IQ.

Several social psychological theories also predict a correlation between intelligence and facial attractiveness. For instance, status generalization theory holds that visible characteristics affecting social status, including facial attractiveness, cause perceivers to generate matching expectations about other traits of the target (L. A. Jackson, Hunter, & Hodge, 1995)—for example, more attractive individuals are assumed to be more intellectually and socially competent, to have more integrity and compassion, and so on (Eagly, Ashmore, Makhijani, & Longo, 1991; Moore, Filippou, & Perrett, 2011). Although this theory primarily predicts correlations between visible status cues and *perceived* levels of internal characteristics, Jackson et al. (1995) argue that, due to the more positive evaluations attractive individuals receive in social and intellectual contexts, they may also receive more opportunities to develop intellectual competence than unattractive individuals. Moreover, attractive individuals may form self-concepts based on social feedback that include notions of superior intellectual ability, potentially motivating intellectual achievement (L. A. Jackson et al., 1995). Thus, both social psychological and evolutionary considerations seem to predict, a priori, a positive phenotypic correlation between intelligence and facial attractiveness.

1.1 Empirical findings

A survey of the published studies on intelligence and attractiveness is summarized in Table 1.The general pattern, identifiable in Jackson et al.'s (1995) and Langlois et al.'s (2000) meta-analyses, appears to be that a small-to-moderate correlation is found in children ($\bar{r}$ = .19, weighted by sample size), but the relationship diminishes with age ($\bar{r}$ = .02, weighted by sample size). However, interpretation of these meta-analytic results is difficult, not only because meta-analyses are vulnerable to the "file drawer problem", whereby null results are less likely to be published than positive ones (Borenstein, Hedges, Higgins, & Rothstein, 2011), but also because of inconsistencies in operational definitions of intelligence and attractiveness across included studies and because many of the included studies had design flaws (e.g., non-independence of intelligence and attractiveness ratings) that could have created biases in the results (Table 1).

Several empirical studies have tested the attractiveness-intelligence correlation since these meta-analyses were published. In contrast to the pattern noted above, Zebrowitz and Rhodes's (2004) found a moderate positive correlations in both children and adults, but only among individuals with below-median attractiveness levels; averaging together the correlations in high- and low-attractiveness groups likely would have resulted in effects more consistent with the earlier meta-analyses. Similarly, Denny (2008) showed that low intelligence may predict low attractiveness in a large sample of school children but that "[f]or much of the distribution of intelligence there is no significant relationship between being attractive and intelligence" (p. 618). Kanazawa (2011) analyzed two large samples of children and young adults, including the one from Denny (2008). Controlling for parental education and income, birth weight, age at puberty, and physical health reduced but did not eliminate the association he observed between physical attractiveness and general intelligence. A serious limitation of the Denny (2008) and Kanazawa (2011) studies is that the raters of attractiveness were familiar with targets' intelligence, leading to potential rater biases that may have artificially induced the correlation under investigation [e.g., see Moore et al. (2011) for a demonstration of an intelligence 'halo' effect on perceived attractiveness.] Most recently, Kleisner, Chvátalová, and Flegr (2014), using reliable, independently collected measures of intelligence and facial attractiveness, failed to find a statistically significant correlation in either male or female young adults. However, this study, like those of Zebrowitz and Rhodes (2004) and many studies included in Jackson et al.'s (1995) and Langlois et al.'s (2000) meta-analyses, utilized a very small sample, rendering its results somewhat inconclusive.

Finally, we observe that nine of 41 previously reported correlations (22 %) were negative and that only 17 of 41 (41 %) were statistically significant. If there truly is some level of positive correlation between intelligence and facial or physical attractiveness, both of these outcomes could reflect high sampling variance due to the small samples most past studies have employed (median N = 83; Table 1). Although the literature in general seems to affirm that a correlation exists (median r = .09; Table 1), the validity of any such meta-analytic result depends both on the individual included studies utilizing appropriate research methods and on the meta-analysis as a whole being free of biases, including publication bias.

1.2 Present study

The present study is the first to utilize highly reliable and independently collected measures of facial attractiveness and general intelligence in a sample much larger than most individual studies in the past have had access to. Importantly, our study also utilizes a genetically informative twin dataset, allowing us to partition the covariation between attractiveness and intelligence into its genetic and environmental components.

2. Methods

2.1 Overview

We combined data from two twin samples to test the hypothesis that facial attractiveness and intelligence are correlated. The first sample comprised participants (n = 399) enrolled in the Longitudinal Twin Study (LTS), previously described by Rhea et al. (2006, 2012). The second sample included participants (n = 1,354) enrolled in the Brisbane Adolescent Twin Study (BATS), previously described by Wright and Martin (2004). Attractiveness data for both samples were collected at the same time and in the same way. There were some differences in the intelligence measures used in the two samples, which are detailed below.

We first tested the phenotypic relationship between facial attractiveness and intelligence using linear mixed effects regression models predicting intelligence scores from facial attractiveness, specifying family as a random factor to account for non-independence between family members. We also used the twin data to fit a structural equation model estimating the genetic and environmental components of the phenotypic correlation. Previous investigations have focused on the phenotypic relationship between attractiveness and intelligence. However, estimating the genetic and environmental components of the correlation provides a more complete test of the predicted relationship between facial attractiveness and intelligence, because opposing genetic and environmental effects might mask each other (Lande, 1982). Thus, a significant genetic or environmental correlation could be interpreted as consistent with the hypothesized relationship, even if the phenotypic correlation were non-significant.

2.2 Longitudinal Twin Study Sample

2.2.1 Participants

The Longitudinal Twin Study (LTS), located at the Institute for Behavioral Genetics in Colorado, USA, consists of monozygotic (MZ) and same-sex dizygotic (DZ) twin pairs. We used data from 399 LTS participants, including 180 complete twin pairs (58% female, 52% MZ) and 39 individuals whose co-twins' data were unavailable (36% female, 51% MZ). The Institutional Review Board at the University of Colorado approved the collection of these data, and informed consent was obtained from all participants.

2.2.2 Measures

Between 2001 and 2006, when LTS participants were aged 16 to 20 (*Mdn* = 16), they completed 11 subtests of the Wechsler Adult Intelligence Scale, 3rd edition (WAIS-III). For the present study, *IQ* was operationalized as the sum of scaled scores on all 11 WAIS-III subtests. *IQ* in the LTS sample was representative of the general population (*M* = 102.4, *SD* = 11.4; Table 2).

Between 2009 and 2012, when participants were aged 21 to 25 (*Mdn* = 22), they posed for between one and four digital photographs taken by a professional research assistant at the Center on Antisocial Drug Dependence, who ensured that photographs were standardized (participants were asked to remove glasses, adopt a neutral facial expression, and face the camera directly without turning or tilting the head). The first author and a volunteer research assistant recruited from the University of Colorado Boulder undergraduate population first inspected all photographs and removed any that were out of focus or improperly standardized (as described above), then cropped the remaining photographs until the faces occupied ~75% of total image area. After cropping, images had dimensions between 350 × 500 to 1300 × 1500 pixels. All available images of each participant were displayed simultaneously to volunteer research assistants recruited from the University of Colorado Boulder undergraduate population on computer screens within a display window with dimensions 17.2 × 22.8 cm.

Two non-overlapping groups of volunteer research assistants, who were naïve to this study's aims, rated images of LTS participants over two academic years, because we began the rating tasks while photography of participants was ongoing. The first group included eight judges (4 female, 4 male) who assigned *Attractiveness* ratings (1 = low attractiveness, 7 = high attractiveness) in the subset of LTS participants who had posed for photographs before mid-2010 (*n* = 228); fourteen additional judges (10 female, 4 male) rated factors that we hypothesized might have transient effects on facial attractiveness that could reduce the validity of our test if not statistically controlled: the amount of *Acne* (1 = no acne, 7 = heavy acne), *Grooming* (1 = carelessly groomed, 7 = carefully groomed), and *Smiling* (1 = no smile, 2 = partial smile, 3 = full smile). Ratings by judges who failed to rate all faces on a given trait were excluded from further analysis, resulting in variable numbers of male and female judges' ratings for *Acne*, *Grooming*, and *Smiling* (see Table 2 for final numbers). The second group comprised 15 judges (9 female, 6 male) who rated LTS participants who had posed for photographs between mid-2010 and 2012 (n = 171) on all four traits, beginning with *Attractiveness*, following the same protocols and using the same scales. Judges in both groups were instructed to use the entire range of trait rating scales and to distribute their ratings approximately uniformly (except for *Smiling*, because most participants did not smile). To help them calibrate their scoring and to reduce order effects, judges viewed a slideshow consisting of 50 randomly selected target faces (all male or all female) displayed for 2 seconds each prior to assigning ratings. Immediately following the slideshow, the same 50 faces were rated in randomized order. This procedure continued, alternating between sets of male and female targets, until all faces were rated. Co-twins were rated on different days to reduce rater effects (e.g., memory, expectancy) on observed similarities between twins. Inter-rater reliability was high for all rated traits (Table 2). For the present study, *Facial Attractiveness* was operationalized as the mean of *Attractiveness* ratings across all judges, although we also assessed the relationships of *IQ* with *Facial Attractiveness* as rated only by same- and opposite-sex judges. *Acne*, *Grooming*, and *Smiling* ratings were likewise averaged across judges for use in subsequent analyses.

2.3 Brisbane Adolescent Twin Study Sample

2.3.1 Participants

The Brisbane Adolescent Twin Study (BATS), located at the QIMR Berghofer Medical Research Institute in Brisbane, Australia, consists of MZ and both same-sex and opposite-sex DZ twin pairs as well as their older siblings. We used data from 1,354 BATS participants for whom acceptably high-quality photographs were available, including 589 complete twin pairs (54% female, 34% MZ), 32 individuals whose co-twins' data were unavailable (54% female, 44% MZ), and 135 non-twin siblings. The Human Research Ethics Committee at QIMR Berghofer approved the collection of these data, and informed consent was obtained from all participants.

2.3.2 Measures

Between 1996 and 2010, when BATS twins were aged 15 to 16 (*Mdn* = 16) and their siblings were aged 17 to 22 (*Mdn* = 18), participants completed five subtests of the Multidimensional Aptitude Battery (MAB). For the present study, *IQ* was operationalized as the scaled score for Full Scale IQ. BATS participants tended to have higher-than-average *IQ* (*M* = 115.3, *SD* = 13.3; Table 2).

At the same age as when IQ tests were taken, BATS participants posed for a single portrait-style facial photograph taken by a professional research assistant working for the Australian Twin Registry, initially using film cameras and later digital cameras. Film images were scanned to digital (JPEG) format. These photographs were originally intended only for identification purposes and so were less standardized than LTS photographs with respect to smiling and head tilting and turning. Therefore, after excluding out-of-focus or improperly standardized images, we rotated the remaining images as necessary to make them upright before cropping, as described above for the LTS sample. Finished images had dimensions between 200 × 500 and 300 × 600 pixels and were displayed to judges on a computer monitor within a display window with dimensions 17.2 × 22.8 cm. The same judges who rated photographs of the first group of LTS participants also rated photographs of all BATS participants on *Attractiveness*, *Acne*, *Grooming*, and *Smiling*, using the same procedures and scales. Judges rated BATS and LTS faces in separate rating sessions. Table 2 shows inter-rater reliabilities for all rated traits.

2.4 Analyses

For our primary analyses, we regressed *IQ* on *Facial Attractiveness* using mixed effects models that included family as a random effect. Next, to investigate whether the relationship between *Facial Attractiveness* and *IQ* depended on participant sex, we refit the mixed effects regression, this time including the *Facial Attractiveness* × sex interaction (where sex was contrast-coded, female = -0.5, male = +0.5). To test whether *IQ* related differentially to *Facial Attractiveness* as rated either by opposite-sex or same-sex judges, we fit two mixed effects regressions, using only ratings assigned by opposite-sex judges in one and only ratings assigned by same-sex judges in the other. Last, to test whether the relationship depended on the interaction between participant sex and rater sex, we refit the opposite-sex and same-sex ratings models, including the interaction between *Facial Attractiveness* ratings and participant sex. All models included a *Facial Attractiveness* × sample (contrast-coded, BATS = -0.5, LTS = +0.5) interaction term, to test whether the relationship differed between the LTS and BATS datasets (which had older and younger groups of participants, respectively).

Both *Facial Attractiveness* and *IQ* were rescaled to have unit variance with means of zero, allowing the slope estimates to be interpreted as partial correlations. Models controlled for participant's year of birth, age, sex, year in which intelligence was tested, SES, BMI when photographed, presence of acne (because the younger BATS participants exhibited considerable variation in facial acne), and several two-way interactions. We initially also controlled for age-squared, grooming, and smiling but dropped them from the final analyses because their effects were not significant in any model. No covariate interacted significantly with attractiveness in predicting intelligence.

We obtained *p*-values for the slope estimates from likelihood ratio tests comparing the likelihood of the full model to the likelihood of a model omitting the parameter of interest (Pinheiro & Bates, 2000). The decrement in likelihood of the reduced model relative to the full model indicates the significance of the omitted parameter, and twice the natural log of the ratio of the two models' likelihoods is asymptotically distributed as a χ^2 statistic (thus, all *p*-values reported below are two-tailed).

Finally, because our samples consisted of MZ and DZ twins and siblings, we fit a structural equation model decomposing the phenotypic correlation between *Facial Attractiveness* and *IQ* into its genetic and environmental components. Because the phenotypic correlations between MZ twins were roughly double the correlations between DZ twins for both *Facial Attractiveness* and *IQ* (e.g., see combined samples, Table 3), estimates of non-additive genetic and shared environmental effects (V_{NA} and V_C, respectively) did not approach significance in any model. We therefore present results from only the AE models, which estimate the influences of additive genetic and unique environmental effects on both traits. However, it is important to recognize that, in a twins-plus-siblings design, non-additive genetic and shared environmental effects both can contribute to estimates of additive genetic variance (V_A), which should therefore be interpreted as potentially being influenced by all three factors (i.e. by anything increasing similarity within families; Keller & Coventry, 2005). Nevertheless, extended twin family designs that produce much less biased estimates of V_A (Keller, Medland, & Duncan, 2010) suggest that the variation underlying *IQ* is mostly additive genetic in nature (Keller et al., 2013). Thus, under the assumption that shared environmental effects have minimal influence on *Facial Attractiveness*, we believe that the estimates of V_A presented here largely reflect the additive effects of genes.

We used the mixed-effects modeling package lme4 (Bates, Maechler, & Bolker, 2012) for R (version 2.15.3; R Core Team, 2013) in phenotypic analyses and the structural equation modeling package OpenMx (version 1.3.0; Boker et al., 2011) for R in twin-based analyses.

3. Results

3.1 Phenotypic analyses

As an overall estimate of the relationship between *Facial Attractiveness* and *IQ*, we obtained a standardized regression coefficient (equivalent to a partial correlation) of $\beta = .018$ ($p = .50$). This effect was similar in both sexes: when the *Facial Attractiveness* × sex interaction was included, neither *Facial Attractiveness* ($\beta = .016$, $p = .55$) nor the interaction ($\beta = -.040$, $p = .33$) was a significant predictor of *IQ*. In both models, the effect of *Facial Attractiveness* did not differ significantly between the LTS and BATS samples (*Facial Attractiveness* × sample interaction: $\beta = .021$, $p = .67$ and $\beta = .019$, $p = .70$, respectively).

We also tested whether the relationship between *IQ* and *Facial Attractiveness* was only apparent when *Facial Attractiveness* was judged by raters of the opposite sex to the target. Using this (perhaps more ecologically valid) measure of *Facial Attractiveness*, we estimated a non-significant relationship of $\beta = .012$ ($p = .64$) between *Facial Attractiveness* and *IQ* that did not differ between samples (β = .030, p = .55). When including the interaction with participant sex, neither the simple slope for *Facial Attractiveness* (β = .011, p = .68) nor the interaction (β = -.019, p = .64) was significant; the *Facial Attractiveness* × sample interaction was also not significant (β = .029, p = .55). For completeness, we then tested whether the relationship between *IQ* and *Facial Attractiveness* was only apparent when *Facial Attractiveness* was judged by raters of the same sex as the target, but none of the regression coefficients approached significance (all $p > .13$).

To ensure that our estimates of the *Facial Attractiveness-IQ* relationship were not being attenuated because we had removed the effects of variables that mediate the relationship, we refit the models above without any covariates. Zero-order slope estimates were consistent with the estimates from the full models above, whether *Facial Attractiveness* was rated by all judges (β = .012, p = .587), by opposite-sex judges (β = -.002, p = .919), or by same-sex judges (β = .024, p = .270). There was no evidence that the relationship between *Facial Attractiveness* and *IQ* differed between samples (*Facial Attractiveness* × sample interaction, all $p > .481$) or between sexes (*Facial Attractiveness* × sex interaction, all $p > .118$). Finally, in light of Zebrowitz and Rhodes's (2004) finding a correlation only among low-attractiveness individuals (see also Denny, 2008), we refit the full models from the primary analyses, this time including squared *Facial Attractiveness* terms, to test the non-linear relationship between *Facial Attractiveness* and *IQ*; none of these terms was significant (all $p > .735$). Controlling for the non-linear effects, first-order *Facial Attractiveness* slopes also remained non-significant in all models (all $p > .402$).

3.2 Biometrical analyses

Results above showed no phenotypic relationship between *Facial Attractiveness* and *IQ*.

To understand whether a positive genetic relationship between the traits may have been masked by a negative environmental one, we estimated the genetic and environmental covariances between *IQ* and *Facial Attractiveness.* Because phenotypic analyses suggested the relationship between *Facial Attractiveness* and *IQ* did not differ between samples, we combined the samples to estimate the genetic and environmental covariance components. The non-shared environmental variance in *Facial Attractiveness* could not be equated across samples without significantly reducing model fit (Table 4). However, neither the genetic covariance (cov_A = -.002, p = .95; genetic correlation, r_A = -.003) nor the environmental covariance (cov_E = .026, p = .053; environmental correlation, r_E = .088, standardized by V_E estimate in LTS; r_E = .101, standardized by V_E estimate in BATS) was significantly different from zero.

4. Discussion

Existing evolutionary genetic models predict that persistent directional selection and cross-trait assortative mating can produce positive genetic covariance between traits (Hansen, 2006), and it has been shown that IQ and height are positively correlated due to both pleiotropy and cross-trait assortative mating (Keller et al., 2013). We had expected that facial attractiveness and intelligence would be subject to even stronger cross-trait assortative mating, and in view of previous empirical results (e.g., L. A. Jackson et al., 1995; Langlois et al., 2000; Zebrowitz & Rhodes, 2004), we were confident we would observe positive genetic correlations between facial attractiveness and IQ. Our null findings were therefore unexpected. Nevertheless, several strengths of the current design give us confidence in these results. In both the LTS and the BATS, general intelligence was assessed with well-validated instruments (D. N. Jackson, 1984; Wechsler, 1981, 1997). Our measure of facial attractiveness also was highly reliable (Cronbach's α = .86-.91), and research supports the validity of subjective assessment of facial attractiveness in photographs viewed by groups of untrained judges (Langlois et al., 2000). Importantly, our intelligence and facial attractiveness assessments were made independently of each other (cf. Denny, 2008; Kanazawa, 2011), reducing the possibility that the correlation between them was biased in either direction due to confounding. The genetically informative nature of the sample ensured that we did not miss genetic correlations that were masked by environmental ones in the opposite direction. Finally, the large size of our combined sample (N = 1,753) allowed us to rule out with a high degree of certainty the existence of a sizeable correlation between facial attractiveness and general intelligence. We had 80% power to detect a true correlation of r = .07 in mixed model analyses.

On the other hand, if the true facial attractiveness-intelligence correlation in adults is approximately r = .03, as Langlois et al. (2000) suggested it is, our study was underpowered to detect it. An additional caveat is that still photographs may not provide a highly valid assessment of facial attractiveness; we attempted to lessen this potential problem by providing judges with as many photographs as were available for each participant, on the assumption that viewing multiple photographs might reduce the impact of, for example, facial expression, poor lighting, and other variable conditions captured in a single photograph. To the extent that this study's facial attractiveness ratings correlate imperfectly with participants' 'true' attractiveness (e.g., if participants were assessed by every other human in the population following face-to-face interactions), our measurements may contain error variance that would bias the estimates downward. Therefore, we remain cognizant of the possibility that these results may be false negatives due to sampling error or design faults.

In light of our findings, however, it is also possible that the previously published reports of positive correlations between intelligence and facial attractiveness were false positives. If attractiveness and intelligence are truly uncorrelated, then most investigations should not find a statistically significant correlation; due to sampling variation, the estimates from smaller studies are expected to be less precise (more variable) than those from larger studies, but their mean should be an unbiased estimate of the true correlation. However, publication bias toward significant findings (regardless of statistical power) or against null finding with sufficient power—i.e., the “file drawer problem”—could produce an over-representation of large effects estimated in underpowered studies. Consistent with this, there is a negative relationship between sample sizes and absolute effect sizes in the published studies described above, including the individual studies meta-analyzed by L. A. Jackson et al. (1995) and Langlois et al. (2000) (r = -.27, df = 28, p = .15). The plot of this trend (Figure 1) suggests that $\log_{10}$-transformed sample size best captures the linear relationship that exists between sample size and absolute effect size (r = -.41, df = 28, p = .03).

Limitations of past studies may have contributed to their reaching false conclusions, as well. For example, Zebrowitz and Rhodes (2004) analyzed seemingly reliable measures of standard psychometric intelligence and facial attractiveness rated by a large number of male and female judges which, importantly, were collected independently of each other. However, their samples were relatively small, and the effects of confounding variables such as socioeconomic status may have inflated their estimates. On the other hand, Kanazawa’s (2011) study, despite its large samples and careful statistical control of potential confounds, must be interpreted with extreme caution because intelligence and physical attractiveness were not measured independently of each other: participants were rated by their schoolteachers in one sample, and by interviewers at the conclusion of interviews in which intelligence was also measured in the other. The fact that the raters of attractiveness were in a position to be aware of the intellectual ability of those being rated means that their attractiveness assessments might have been biased such that teachers or interviewers rated more intelligent participants as more attractive (Moore et al., 2011). Many of the studies meta-analyzed by L. A. Jackson et al. (1995) and Langlois et al. (2000) operationalized intellectual competence in terms of a diverse set of academic and occupational performance measures (e.g., grades, vocabulary tests, occupational prestige, salary) that may be influenced by a variety of factors not directly related to the hypothesis under investigation. Likewise, some studies rated facial attractiveness, whereas others rated overall physical attractiveness. Furthermore, meta-analyses rely upon the published literature being an unbiased sample of all tests conducted on the question under investigation; to the degree that the published literature is biased towards statistically significant results, meta-analyses may produce spuriously high estimated effects (Borenstein et al., 2011). Thus, given our demonstration of potential publication bias in the literature (Figure 1 and Table 1), results from previous meta-analyses on the facial attractiveness-IQ association should be interpreted with caution.

If the results reported here do indeed indicate no phenotypic or genetic relationship exists between IQ and facial attractiveness, how should this be reconciled with evolutionary genetic models that appear to predict such a correlation? We discuss briefly a few possibilities. First, although people report that both facial attractiveness and general intelligence are desirable in mates (Buss et al., 1990; Kenrick, Groth, Trost, & Sadalla, 1993), there appear to be no empirical studies demonstrating cross-trait assortment involving facial attractiveness and intelligence in human couples, and it is possible either that it does not occur or that its occurrence is not widespread in the populations from which our sample was drawn. Second, if cross-trait assortative mating desists, recombination and independent assortment will reduce these associations quickly, within a few generations (Hedrick, 2011); it is possible therefore that assortment across these traits has decreased in evolutionarily recent times, although we can think of no compelling reason why this would be so. Third, in the same way that directional selection is expected to reduce the genetic variance in a single trait, it should also reduce the positive genetic covariance between a pair of traits as the fittest alleles at pleiotropic loci (i.e., alleles that improve both traits) are swept to fixation, resulting in their contributing nothing to the covariance between the traits (Falconer & Mackay, 1996; Lande, 1982). On the other hand, genetic loci harboring alleles with opposing effects on the two traits can potentially remain polymorphic much longer, for example if the deleterious (fitness-reducing) effects of both alleles are nearly equal (Hedrick, 1999). In such cases, antagonistic pleiotropy may exert a negative influence on the genetic correlation between the traits (Roff, 1996), though perhaps only transiently. Our results might therefore reflect that, on balance, the evolutionary genetic forces working to exert a positive influence on the genetic correlation between intelligence and facial attractiveness are countervailed by the forces working to exert a negative influence on this correlation.

Evolutionary theories are mute as to whether an environmental component should also exist, but it is conceivable that environmental challenges such as poor nutrition or pathogens might induce a positive correlation, for example by disrupting developmental stability (Gangestad et al., 1994; Møller, 1997). Social psychological theories imply that facial attractiveness exerts a causal influence on intelligence at the phenotypic level; if this were the case, because attractiveness is influenced by both genetic and environmental effects (McGovern, Neale, & Kendler, 1996; Mitchem et al., 2013), the cross-trait correlations measured in our twin analysis should have consisted of both a genetic and an environmental component. That we observed neither suggests that facial attractiveness and intelligence develop under the influence of independent sets of environmental factors and that the phenomena described by status generalization and expectancy theories have, at best, weak effects.

In summary, our results fail to support the hypothesis that facial attractiveness and intelligence are correlated either at the level of observable trait values or via latent genetic or environmental influences. These observations are inconsistent with the several predictions that such a correlation should exist, either for biological reasons (e.g., shared genes, cross-trait assortment, developmental stability) or for social-psychological reasons (e.g., status generalization, expectancy). Therefore, it may be necessary to reassess the theoretical bases for the expectation that a correlation exists between intelligence and facial attractiveness.

Acknowledgements

This work was supported by grants from the Australian Research Council (A79600334, A79801419, and DP0212016), the National Institutes of Health (HD010333), and the National Institute of Mental Health (MH085812 and MH63207). The authors would like to thank the Australian Twin Registry and Colorado Twin Registry twins and their families for their continued participation. We would also like to thank Marlene Grace, Ann Eldridge, Daniel Park, and David Smyth for their work with the Australian Twin Registry; the professional research assistants at the Center on Antisocial Drug Dependence for their work with the Longitudinal Twin Study; and the volunteer research assistants who assigned trait ratings.

FIGURE CAPTION

Figure 1. The estimated strength of the relationship between attractiveness and intelligence tends to decrease as $\log_{10}$ sample size increases (*r* = -.41, *df* = 28, *p* = .03), consistent with the possibility of a bias in favor of publishing small, underpowered studies that overestimated the magnitude of the correlation. In this analysis, we included only independently estimated effects. Frieze et al. (1990) and Zebrowitz and Rhodes (2004) each used only two non-overlapping samples to estimate four and eight correlations, respectively; we applied the Fisher transformation to these correlations (i.e., computed a *z* for each correlation, found the mean of all *z*s obtained from the same sample, then transformed the mean *z* back to a correlation) to arrive at independent estimates for inclusion in this analysis. We omitted Kanazawa's (2011) results as well as the third entry for Felson (1980), because attractiveness and intelligence were not assessed independently in those studies (Table 1).

References

Arden, R., Gottfredson, L. S., Miller, G., & Pierce, A. (2009). Intelligence and semen quality are positively correlated. *Intelligence, 37*(3), 277-282.

Bates, Douglas, Maechler, Martin, & Bolker, Ben. (2012). lme4: Linear mixed-effects models using S4 classes (Version R package version 0.999999-0).

Billington, H. L., Mortimer, A. M., & McNeilly, T. (1988). Divergence and genetic structure in adjacent grass populations. I. Quantitative genetics. *Evolution, 42*(6), 1267-1277.

Boker, S. M., Neale, M. C., Maes, H. H., Wilde, M. J., Spiegel, M., Brick, T. R., . . . Fox, J. (2011). OpenMx: An open source extended structural equation modeling framework. *psychometrika, 76*(2).

Borenstein, Michael, Hedges, Larry V, Higgins, Julian PT, & Rothstein, Hannah R. (2011). *Introduction to meta-analysis*: John Wiley & Sons.

Buss, D. M., Abbott, M., Angleitner, A., Asherian, A., Biaggio, A., Blanco-Villasenor, A., . . . Yang, K. (1990). International preferences in selecting mates: a study in 37 cultures. *Journal of cross-cultural psychology, 21*(1), 5-47.

Buss, D. M., & Barnes, M. (1986). Preferences in human mate selection. *J Pers Soc Psychol, 50*(3), 559-570. doi: 10.1037/0022-3514.50.3.559

Clark, Andrew G. (1990). Genetic components of variation in energy storage in Drosophila melanogaster. *Evolution, 44*(3), 637-650.

Crow, J. F., & Felsenstein, J. (1968). The effect of assortative mating on the genetic composition of a population. *Biodemography and Social Biology, 15*(2), 85-97.

Davies, G., Tenesa, A., Payton, A., Yang, J., Harris, S. E., Liewald, D., . . . Deary, I. J. (2011). Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Mol Psychiatry, 16*(10), 996-1005.

Denny, Kevin. (2008). Beauty and intelligence may–or may not–be related. *Intelligence, 36*(6), 616-618.

Eagly, A. H., Ashmore, R. D., Makhijani, M. G., & Longo, L. C. (1991). What is beautiful is good, but…: A meta-analytic review of research on the physical attractiveness stereotype. *Psychol Bull, 110*(1), 109-128.

Falconer, D. S., & Mackay, T. F. C. (1996). *Introduction to Quantitative Genetics* (4 ed.). Harlow, England: Pearson Prentice Hall.

Flint, J., & Mackay, T. F. C. (2009). Genetic architecture of quantitative traits in mice, flies, and humans. *Genome Research, 19*(5), 723-733. doi: 10.1101/gr.086660.108

Futuyma, D. J., & Philippi, T. E. (1987). Genetic variation and covariation in responses to host plants by Alsophila pometaria (Lepidoptera: geometridae). *Evolution, 41*(2), 269-279.

Gangestad, S. W., Thornhill, R., & Yeo, R. A. (1994). Facial attractiveness, developmental stability, and fluctuating asymmetry. *Ethology and Sociobiology, 15*(2), 73-85.

Gangestad, S. W., & Yeo, R. A. (1997). Behavioral genetic variation, adaptation and maladaptation: An evolutionary perspective. *Trends in Cognitive Sciences, 1*(3), 103-108.

Garland, T., Jr. (1988). Genetic basis of activity metabolism. I. Inheritance of speed, stamina, and antipredator displays in the garter snake Thamnophis sirtalis. *Evolution, 42*(2), 335-350.

Hansen, Thomas F. (2006). The evolution of genetic architecture. *Annual Review of Ecology, Evolution, and Systematics*, 123-157.

Hedrick, P. W. (1999). Antagonistic pleiotropy and genetic polymorphism: a perspective. *Heredity, 82*(2), 126-133.

Hedrick, P. W. (2011). *Genetics of Populations*: Jones & Bartlett Learning.

Houle, D. (1998). How should we explain variation in the genetic variance of traits? *Genetica, 102-103*(0), 241-253.

Jackson, D. N. (1984). *MAB, multidimensional aptitude battery: manual*: Research Psychologists Press.

Jackson, L. A., Hunter, J. E., & Hodge, C. N. (1995). Physical attractiveness and intellectual competence: A meta-analytic review. *Social Psychology Quarterly, 58*(2), 108-122.

Kanazawa, S. (2011). Intelligence and physical attractiveness. *Intelligence, 39*(1), 7-14.

Keller, M. C. (2007). The role of mutations in human mating. In G. Geher & G. Miller (Eds.), *Mating Intelligence: Sex, Relationships, and the Mind's Reproductive System*: Taylor & Francis.

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., Garver-Apgar, C. E., Wright, M. J., Martin, N. G., Corley, R. P., Stallings, M. C., . . . Zietsch, B. P. (2013). The genetic correlation between height and IQ: Shared genes or assortative mating? *PLoS Genet, 9*(4).

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. *Behav Genet, 40*(3), 377-393.

Kenrick, D. T., Groth, G. E., Trost, M. R., & Sadalla, E. K. (1993). Integrating evolutionary and social exchange perspectives on relationships: Effects of gender, self-appraisal, and involvement level on mate selection criteria. *J Pers Soc Psychol, 64*(6), 951-969.

Kenrick, D. T., Sadalla, E. K., Groth, G. E., & Trost, M. R. (1990). Evolution, traits, and the stages of human courtship: Qualifying the parental investment model. 1. 58

Kleisner, Karel, Chvátalová, Veronika, & Flegr, Jaroslav. (2014). Perceived Intelligence Is Associated with Measured Intelligence in Men but Not Women. *PLoS One, 9*(3), e81237.

Kosova, G., Abney, M., & Ober, C. (2009). Heritability of reproductive fitness traits in a human population. *Proceedings of the National Academy of Sciences*.

Kovas, Y., & Plomin, R. (2006). Generalist genes: implications for the cognitive sciences. *Trends in Cognitive Sciences, 10*(5), 198-203.

Lande, R. (1982). A quantitative genetic theory of life history evolution. *Ecology, 63*(3), 607-615.

Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., Hallam, M., & Smoot, M. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychol Bull, 126*(3), 390-423.
Lee, S., Ripke, S., Neale, B., Faraone, S., Purcell, S., & al., et. (2013). Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nature Genetics, 45*(9), 984-994.
McGovern, R. J., Neale, M. C., & Kendler, K. S. (1996). The independence of physical attractiveness and symptoms of depression in a female twin population. *The Journal of psychology, 130*(2), 209-219.
Miller, G. (2000). Sexual selection for indicators of intelligence. In G. R. Bock, J. A. Goode & K. Webb (Eds.), *The Nature of Intelligence: Novartis Foundation Symposium 233* (pp. 260-275). Chichester: Wiley.
Mitchell-Olds, T. (1986). Quantitative genetics of survival and growth in Impatiens capensis. *Evolution, 40*(1), 107-116.
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. *Behav Genet*, 1-12. doi: 10.1007/s10519-013-9627-5
Møller, A. P. (1997). Developmental stability and fitness: a review. *The American Naturalist, 149*(5), 916-932.
Moore, F., Filippou, D., & Perrett, D. (2011). Intelligence and attractiveness in the face: Beyond the attractiveness halo effect. *Journal of Evolutionary Psychology, 9*(3), 205-217. doi: 10.1556/JEP.9.2011.3.2
Pinheiro, J. C., & Bates, D. M. (2000). *Mixed effects models in S and S-PLUS*: Springer.
Prout, T. (2000). How well does opposing selection maintain variation. *Evolutionary genetics: from molecules to morphology, 1*, 157.
Purcell, S. M., Wray, N. R., Stone, J. L., & International Schizophrenia Consortium. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature, 460*(7256), 748-752.
R Core Team. (2013). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
Rausher, M. D., & Simms, E. L. (1989). The evolution of resistance to herbivory in Ipomoea purpurea. I. Attempts to detect selection. *Evolution, 43*(3), 563-572.
Rhea, S., Gross, A. A., Haberstick, B. C., & Corley, R. P. (2006). Colorado Twin Registry. *Twin Research and Human Genetics, 9*(6), 941-949.
Rhea, S., Gross, A. A., Haberstick, B. C., & Corley, R. P. (2012). Colorado Twin Registry: An update. *Twin Research and Human Genetics, FirstView*, 1-7.
Roff, D. A. (1996). The evolution of genetic correlations: An analysis of patterns. *Evolution, 50*(4), 1392-1403. doi: 10.2307/2410877
Roff, D. A. (1997). *Evolutionary quantitative genetics*: Chapman & Hall.
Rowe, L., & Houle, D. (1996). The lek paradox and the capture of genetic variance by condition dependent traits. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 263*(1375), 1415-1421.
Solovieff, N., Cotsapas, C., Lee, P. H., Purcell, S. M., & Smoller, J. W. (2013). Pleiotropy in complex traits: Challenges and strategies. *Nat Rev Genet, 14*(7), 483-495.

Stahl, E. A., Wegmann, D., Trynka, G., Gutierrez-Achury, J., Do, R., Voight, B. F., . . . Plenge, R. M. (2012). Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. *Nat Genet, 44*(5), 483-489.
Thornhill, R., & Gangestad, S. W. (1999). Facial attractiveness. *Trends in Cognitive Sciences, 3*(12), 452-460.
Wechsler, D. (1981). *WAIS-R manual: Wechsler adult intelligence scale-revised*: Psychological Corporation.
Wechsler, D. (1997). *WAIS-III: Administration and scoring manual: Wechsler adult intelligence scale*: Psychological Corporation.
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.

Table 1.

Published studies of the correlation between attractiveness and intelligence.

Study	Sample size	Age	Sex	Rating type	Attractiveness measure	Intelligence measure	*r*
Kleisner et al. (2014)	40	A	M	I	F	IQ	.21[e]
Kleisner et al. (2014)	40	A	F	I	F	IQ	.09[e]
Kanazawa (2011)	3,463	C	M + F	N	O	IQ	.30[f]
Kanazawa (2011)	5,694	C + A	M + F	N	O	IQ	.06[f]
Zebrowitz & Rhodes (2004)[a]	105	C	M + F	I	F	IQ	.20
Zebrowitz & Rhodes (2004)[b]	103	C	M + F	I	F	IQ	-.04
Zebrowitz & Rhodes (2004)[a]	83	C	M + F	I	F	IQ	.30
Zebrowitz & Rhodes (2004)[b]	100	C	M + F	I	F	IQ	.04
Zebrowitz & Rhodes (2004)[a]	115	A	M + F	I	F	IQ	.04
Zebrowitz & Rhodes (2004)[b]	121	A	M + F	I	F	IQ	.07
Zebrowitz & Rhodes (2004)[a]	64	A	M + F	I	F	IQ	.35
Zebrowitz & Rhodes (2004)[b]	70	A	M + F	I	F	IQ	.13
Langlois et al. (2000)	3,043	C	M + F	Mixed	Mixed	Mixed	.19
Langlois et al. (2000)	3,853	A	M + F	Mixed	Mixed	Mixed	.03
Jackson et al. (1995)	2,839	C	M + F	Mixed	Mixed	Mixed	.20
Jackson et al. (1995)	3,255	A	M + F	Mixed	Mixed	Mixed	.01
Gabriel et al. (1994)[c, d]	62	A	M	I	F	IQ	.01
Gabriel et al. (1994)[c, d]	84	A	F	I	F	IQ	-.02
Baugh & Parry (1991)[d]	39	A	F	I	O	A	.10
Lerner et al. (1990)[c, d]	101	C	M + F	I	F	A	.31
Frieze et al. (1990)[c]	423	A	M	I	F	O	.12[g]
Frieze et al. (1990)[c]	260	A	F	I	F	O	.03[g]
Frieze et al. (1990)[c]	452	A	M	I	F	O	.08[g]
Frieze et al. (1990)[c]	285	A	F	I	F	O	.18[g]
Dickey-Bryant et al. (1986)[c]	60	A	M	I	F	A + O	.27

Moran & McCullers (1984)[d]	37	A	M + F	I	F	IQ + A	-.43
Feingold (1982)[c, d]	75	A	M	I	F	IQ	-.13
Feingold (1982)[c, d]	75	A	F	I	F	IQ	-.10
Murphy et al. (1981)[c, d]	41	A	M + F	I	F	A	.13
Felson (1980)[c, d]	53	C	M	I	O	A	.20
Felson (1980)[c, d]	84	C	F	I	O	A	.02
Felson (1980)[c, d]	2,201	C	M	N	O	A	.22
Sparacino (1980)[c]	669	A	M	I	F	A + O	-.06
Sparacino & Hansell (1979)[c, d]	84	A	M	I	F	A	.01
Sparacino & Hansell (1979)[c, d]	83	A	F	I	F	A	-.22
Sparacino & Hansell (1979)[c, d]	55	A	M	I	F	A	.13
Sparacino & Hansell (1979)[c, d]	65	A	F	I	F	A	.09
Sparacino & Hansell (1979)[c, d]	50	A	M	I	F	A	-.30
Sparacino & Hansell (1979)[c, d]	87	A	F	I	F	A	-.02
Salvia et al. (1977)[c, d]	84	C	M + F	I	F	A	.19[h]
Hollingworth (1935)[d]	40	C + A	M + F	I	F	IQ	.33
Mohr & Lund (1933)[d]	50	A	M	I	O	IQ+A	.05
Mohr & Lund (1933)[d]	50	A	F	I	O	IQ+A	.25
Mohr (1932)[d]	25	A	M	I	O	*not available*	.28
Mohr (1932)[d]	25	A	F	I	O	*not available*	.29

Note. Age of participants: A = adults (≥16 years old), C = children. Sex of participants: M = males, F = females. Type of attractiveness rating: I = independent (raters and targets were not acquainted), N = non-independent. Attractiveness measure: F = facial attractiveness only, O = overall physical attractiveness. Intelligence measure: IQ = general intelligence, A = academic performance (e.g., grades, standardized test scores, educational attainment), O = occupational success (e.g., starting or current salary, occupational prestige). Attractiveness and intelligence measures and the type of attractiveness measure in the two meta-analyses are categorized as ‘Mixed’, reflecting the heterogeneity of methods and operational definitions employed by included studies. Table entries beginning with

Gabriel et al. (1994) were included in one or both meta-analyses. Correlations are zero-order, unless otherwise noted.

[a]Individuals with below-median attractiveness. [b]Individuals with above-median attractiveness. [c]Included in L. A. Jackson et al. (1995). [d]Included in Langlois et al. (2000). [e]Controlling for sex. [f]Controlling for sex, birth weight, age at puberty, global health, and parental income and education. [g]Controlling for years of work experience. [h]Controlling for sex and grade level.

Table 2.

Descriptive statistics for *IQ* and volunteer-rated facial measures.

Measure	LTS-1 (*n* = 228)		LTS-2 (*n* = 171)		BATS (*n* = 1,354)	
	Mean (SD)	Cronbach's α (n. judges; n. female, n. male judges)	M)	Cronbach's α (n. judges; n. female, n. male judges)	M)	Cronbach's α (n. s; n. female, n. male judges)
IQ	102.9 (11.0)	---	1 .9)	---	1 3)	---
Facial Attractiveness	3.8 (1.3)	.89 (8; 4F, 4M)	3)	.91 (15; 9F, 6M)	3)	.86 (8; 4F, 4M)
Acne	2.5 (1.0)	.91 (10; 6F, 4M)	2)	.97 (15; 9F, 6M)	3)	.94 (10; 6F, 4M)
Grooming	3.5 (1.2)	.90 (10; 8F, 2M)	3)	.94 (15; 9F, 6M)	3)	.85 (10; 8F, 2M)
Smiling	1.4 (0.4)	.91 (10; 7F, 3M)	1)	.96 (15; 9F, 6M)	2)	.98 (10; 7F, 3M)

ACCEPTED MANUSCRIPT

Table 3.

Between-twin and within-person phenotypic correlations [95% confidence intervals] for LTS, BATS, and both samples combined.

	LTS (n = 399)		BATS (n = 1,354)		Combined (N = 1,753)	
Between twins	MZ twins (93 pairs)	DZ twins (87 pairs)	MZ twins (203 pairs)	DZ twins (498 pairs)	MZ twins (296 pairs)	DZ twins (585 pairs)
Facial Attractiveness	.44 [.27, .57]	.33 [.14, .48]	.62 [.54, .69]	.28 [.20, .35]	.54 [.47, .59]	.28 [.21, .35]
IQ	.82 [.75, .86]	.53 [.37, .64]	.85 [.81, .87]	.43 [.36, .49]	.83 [.81, .85]	.43 [.36, .49]
Facial Attractiveness - IQ	.06 [-.06, .18]	.14 [.00, .27]	-.04 [-.10, .03]	-.01 [-.07, .06]	.00 [-.06, .06]	.02 [-.05, .08]
Within person						
Facial Attractiveness - IQ	.09 [-.02, .20]		.00 [-.06, .06]		.03 [-.02, .08]	

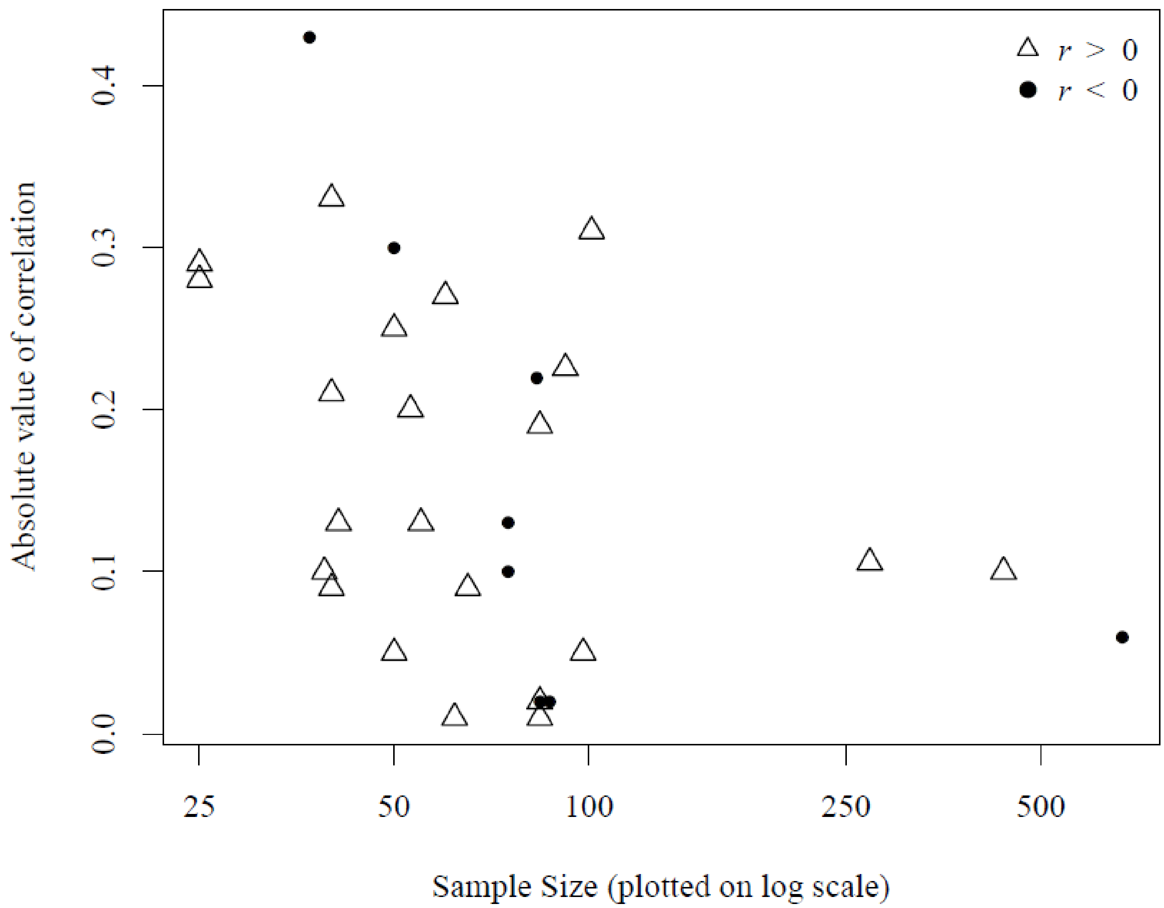
Note. Facial Attractiveness and *IQ* were residualized on year of birth, age, sex, year of IQ test, SES, BMI when photographed, acne, and several two-way interactions. In BATS and combined samples, the effective sample sizes of DZ twins given in the table header exceed the numbers of DZ twin pairs given in the text, because siblings increase the number of pairs of observations informing the correlation estimates.

Table 4.

Parameter estimates from biometrical models for LTS, BATS, and both samples combined.

	LTS (n = 399)		BATS (n = 1,354)		Combined (N = 1,753)	
Measure	V_A	V_E	V_A	V_E	V_A	V_E
Facial Attractiveness	.45 [.29, .63]	.53 [.41, 68]	.61 [.51, .72]	.39 [.33, .46]	.56 [.47, .65]	.54 [.45, .65][a] .41 [.34 .48][b]
IQ	.80 [.65, .97]	.18 [.14, .24]	.84 [.76, .94]	.15 [.13, .18]	.83 [.75, .91]	.17 [.15, .19]
Facial Attractiveness - IQ	.07 [-.04, .19]	.01 [-.05, .07]	-.04 [-.10, .03]	.03 [-.00, .06]	-.00 [-.06, .06]	.03 [-.00, .05]

Note: Additive genetic (V_A) and non-shared environmental (V_E) variance-covariance components [95% confidence intervals] for *Facial Attractiveness*, *IQ*, and the cross-trait covariation are shown. *Facial Attractiveness* and *IQ* were residualized on year of birth, age, sex, year of IQ test, SES, BMI when photographed, acne, and several two-way interactions. In the combined model, the non-shared environmental variance in *Facial Attractiveness* could not be equated across samples without significantly reducing model fit.
[a]LTS sample. [b]BATS sample.



Δ r > 0
● r < 0
Absolute value of correlation
0.0
0.1
0.2
0.3
0.4
25
50
100
250
500
Sample Size (plotted on log scale)

Figure 1