Accepted refereed manuscript of: Lee AJ, Hibbs C, Wright MJ, Martin NG, Keller MC & Zietsch BP (2017) Assessing the accuracy of perceptions of intelligence based on heritable facial features. *Intelligence*, 64, pp. 1-8. DOI: https://doi.org/10.1016/j.intell.2017.06.002

Assessing the accuracy of perceptions of intelligence based on heritable facial features.

Authors:

Anthony J. Lee[1,2], Courtney Hibbs[3,4], Margaret J. Wright[5,6], Nicholas G. Martin[7], Matthew C. Keller[3,4], Brendan P. Zietsch[2,7]

Author affiliations:

[1] Institute of Neuroscience and Psychology, University of Glasgow, Glasgow, Scotland, United Kingdom.

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

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

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

[5] Queensland Brain Institute, University of Queensland, Brisbane, Queensland, Australia.

[6] Centre for Advanced Imaging, University of Queensland, Brisbane, Queensland, Australia.

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

Corresponding authors:

Anthony J. Lee: anthony.lee@glasgow.ac.uk

Word Count: 7504 words

Keywords:

Intelligence perception; IQ; Face perception; Behavioural genetics; Attractiveness; Shape analysis;

Abstract

Perceptions of intelligence based on facial features can have a profound impact on many social situations, but findings have been mixed as to whether these judgements are accurate. Even if such perceptions were accurate, the underlying mechanism is unclear. Several possibilities have been proposed, including evolutionary explanations where certain morphological facial features are associated with fitness-related traits (including cognitive development), or that intelligence judgements are over-generalisation of cues of transitory states that can influence cognition (e.g., tiredness). Here, we attempt to identify the morphological signals that individuals use to make intelligence judgements from facial photographs. In a genetically informative sample of 1660 twins and their siblings, we measured IQ and also perceptions of intelligence based on facial photographs. We found that intelligence judgements were associated with both stable morphological facial traits (face height, interpupillary distance, and nose size) and more transitory facial cues (eyelid openness, and mouth curvature). There was a significant association between perceived intelligence and measured IQ, but of the specific facial attributes only interpupillary distance (i.e., wide-set eyes) significantly mediated this relationship. We also found evidence that perceived intelligence and measured IQ share a familial component, though we could not distinguish between genetic and shared environmental sources.

Assessing the accuracy of perceptions of intelligence based on heritable facial features.

Judgements of intelligence are made quickly and can have profound impact in various social situations. For instance, in educational settings, pre-conceived perceptions of intelligence can influence a student's academic performance (Brophy, 1983; Dunkel & Murphy, 2014; Jussim, 1989; but see Jussim & Harber, 2005). In an employment setting, interviewers are likely to seek to confirm pre-conceived intelligence evaluations, which can affect their judgement during hiring decisions (Judice & Neuberg, 1998). Perceptions of intelligence have also been found to influence leadership decisions (Spisak, Blaker, Lefevre, Moore, & Krebbers, 2014).

Perceptions of intelligence can be made based on numerous traits, such as language use (Reynolds & Gifford, 2001), body symmetry (Prokosch, Yeo, & Miller, 2005), and also facial features. Previous work investigating facial traits associated with perceptions of intelligence have implicated face height, interpupillary distance (distance between the eyes), nose size, and chin pointedness (Kleisner, Chvatalova, & Flegr, 2014), as well as eyelid openness, and mouth curvature (Talamas, Mavor, Axelsson, Sundelin, & Perrett, 2016). However, it is unclear whether these or any other facial traits are associated with actual intelligence. While some studies suggest that intelligence judgements of unfamiliar individuals based solely on facial attributes are accurate (i.e. better than chance; Carney, Colvin, & Hall, 2007; Zebrowitz, Hall, Murphy, & Rhodes, 2002), others find no relationship (Borkenau & Liebler, 1995), or that facial attributes can hinder overall accuracy (Olivola & Todorov, 2010). Other research has indicated that the relationship may be more complicated, such as being sex-dependent (Kleisner et al., 2014; Murphy, Hall, & Colvin, 2003), or age-dependent (Milonoff & Nummi, 2012). If the association between perceptions of intelligence and actual intelligence is very small, the studies to date may have been underpowered, which could explain the mixed results (see Zebrowitz et al., 2002 for a meta-analysis).

If we assume that individuals are able to judge intelligence better than chance based on facial appearance, the exact mechanism that drives this accuracy is unclear. One possibility is that intelligence is an indicator of underlying genetic quality (Haselton & Miller, 2006; Miller, 2000), which would also be associated with physical attributes, such as attractiveness (Prokosch et al., 2005; Zebrowitz & Rhodes, 2004). Such an association could be explained if the development of intelligence (and attractiveness) relies on the ability to convert energy into fitness-enhancing traits during development (Kokko, Brooks, Jennions, & Morley, 2003; Kokko, Brooks, McNamara, & Houston, 2002). Indeed, intelligence is associated with health measures (Arden, Gottfredson, & Miller, 2009), greater pathogen resistance (Eppig, Fincher, & Thornhill, 2010, 2011), and lower mutation load (Howrigan et al., 2016; Yeo, Gangestad, Liu, Wassink, & Calhoun, 2011). However, it is also possible that the accuracy of intelligence judgements is merely learnt rather than being an evolved mechanism, as previous research has found that it develops in women not at sexual maturity, but later in life (Milonoff & Nummi, 2012).

Another possibility is that intelligence and attractiveness are genetically linked, which could occur if intelligent individuals consistently mate with facially attractive partners (Kanazawa & Kovar, 2004; but see Denny, 2008; Penke et al., 2011). Some premises for this notion are supported; for instance, women rate faces manipulated to appear more intelligent as more attractive (Moore, Law Smith, & Perrett, 2014) and may also find cues to intelligence more attractive when fertile (Haselton & Miller, 2006; but see Gangestad, Thornhill, & Garver-Apgar, 2010). However, other research has found no association between facial attractiveness and intelligence (Feingold, 1992; Langlois et al., 2000; Mitchem et al., 2015), or have even suggested that facial attractiveness hinders accuracy of intelligence judgements (Talamas, Mavor, & Perrett, 2016). Pertinently, we previously found no significant genetic correlation between facial attractiveness and intelligence in the sample used in the present study (Mitchem et al., 2015). For a more nuanced discussion of the link between facial attractiveness and IQ, see Mitchem et al. (2015).

Perceptions of intelligence could also be based on more transitory facial cues (as opposed to stable characteristics). For instance, Talamas, Mavor, Axelsson, et al. (2016) suggest that perceptions of intelligence are driven by overgeneralisation of cues to tiredness, which can change quickly and can affect cognitive performance (Pilcher & Huffcutt, 1996). Indeed, facial attributes associated with tiredness (i.e., eyelid openness and mouth curvature) have been associated with perceptions of intelligence (Talamas, Mavor, Axelsson, et al., 2016). Pupil size has also been associated with intelligence, as it is thought to reflect internal mental processes (Tsukahara, Harrison, & Engle, 2016).

Regardless of the underlying mechanism, here we attempt to identify morphological cues that individuals use to make intelligence judgements based on facial information. In a large (*N* = 1660), genetically informative sample, identical and non-identical twins and their sibling had their facial photographs rated on perceived intelligence and IQ measured. If observers are able to judge intelligence accurately, we should find an association between perceived intelligence and IQ. If such a correlation exists, we will test whether various facial attributes mediate this relationship, including stable morphological facial attributes, such as face height, interpupillary distance and nose size (Kleisner et al., 2014), more transitory cues, such as eyelid openness and mouth curvature (Talamas, Mavor, Axelsson, et al., 2016), as well as predicted IQ based on overall face shape. We will also test whether perceived intelligence shares a genetic component with IQ.

Method

Participants

Participants were 1660 individual twins and their siblings from 833 families who took part in either the Brisbane Adolescent Twin Studies (BATS; Wright & Martin, 2004) or the Boulder Longitudinal Twin Study (LTS; Rhea, Gross, Haberstick, & Corley, 2013). Twins from the BATS (*N* = 1173) had photographs taken as close as possible to their 16th birthday (*M* = 16.03 years, *SD* = .46 years) while their siblings (*N* = 105) had photographs taken close to their 18th birthday (*M* = 17.81 years, *SD* = 1.08 years). Twins from the LTS (*N* = 382) were older than those from the BATS when facial photographs were taken (*M* = 22.21 years, *SD* = 1.29 years).

Photographs

For twins who were part of the BATS, photographs were taken between the years 1996 and 2010. For the earliest waves of data collection, photographs were taken using film cameras and then later scanned into a digital format. For later waves, photographs were taken using digital cameras. For twins from the LTS, digital photographs were taken between 2001-2010. Participants from the LTS were asked to adopt a neutral facial expression, while no instructions were given to participants from the BATS. All photographs were taken under standard indoor lighting conditions.

These photographs were rated on a number of traits, such as facial attractiveness, facial masculinity, and trustworthiness. For the analyses presented here, we focus on ratings of perceived intelligence (for more detail on the rating process, see Mitchem et al., 2015). For perceived intelligence, photographs were presented in a random order to one of two groups of undergraduate research assistants (21 in total; 12 Females, 9 Males; 19-30 years, median = 22 years). The two groups were based on availability as ratings were collected over multiple academic semesters. Ratings were made on a 7-point scale (1 = low in a trait, 7 = high in a trait). Mean perceived intelligence ratings between male and female raters were positively correlated ($r = .41$, $p < .001$); therefore, ratings from male and female raters were combined for further analyses. Cronbach's alpha between raters who rated the same faces was .60 for group 1 (7 raters) and .82 for group 2 (14 raters), while the intra-class correlation (i.e., the proportion of total variance in ratings that is between-faces compared to within) across all perceived intelligence ratings was .19.

Facial Metrics

In order to calculate the various facial metrics scores, we used the coordinates of 31 landmarks that were placed on each facial photograph. Two research assistants who did not give trait ratings identified 31 landmarks on each face (see Figure 1. for the locations of the landmarks). These research assistants were trained on the anatomical location of the landmarks for several sessions. The coordinate for each landmark was then calculated as the mean pixel location of the two raters.

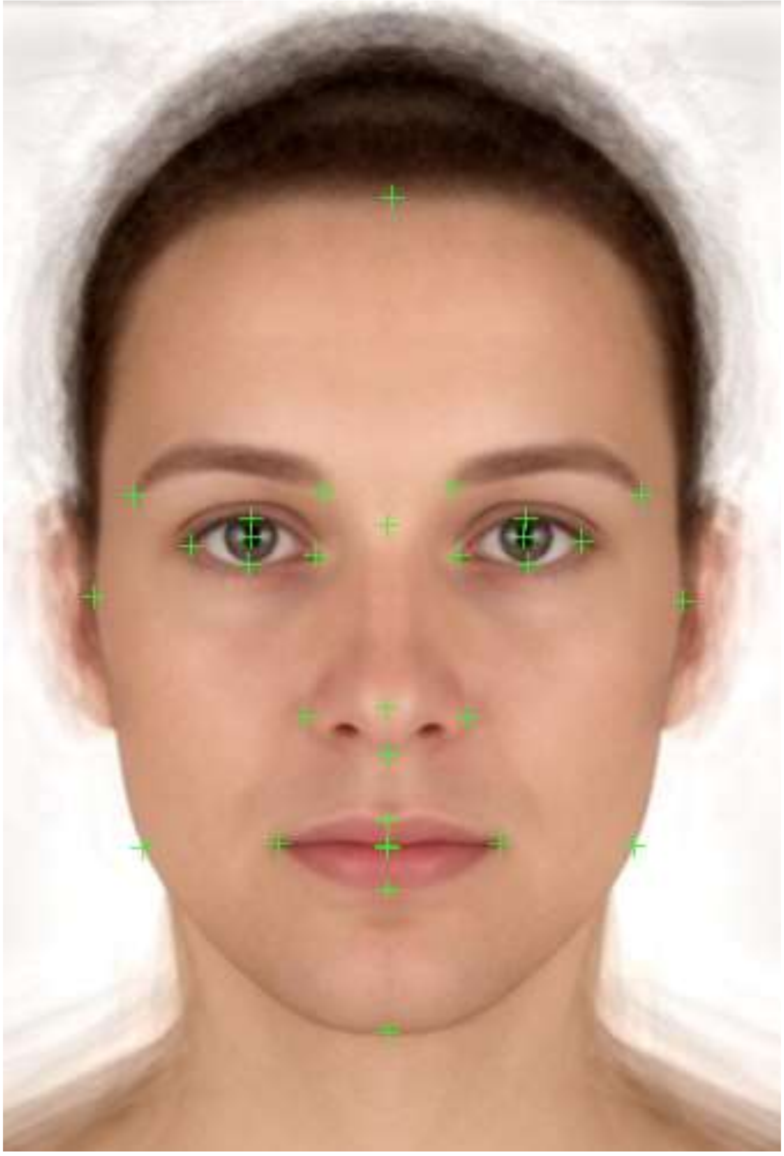

Figure 1. Locations for the 31 landmarks identified on each facial photograph.

We note that these photographs of participants were not originally taken for shape analysis. As such, the photographs vary in ways that could alter shape information not to do with anatomical shape (e.g., the participant’s head angle facing the camera, or the participant’s facial expression). Photographs were rotated to be upright prior to being rated, and overly askew faces were removed from analysis.

To calculate facial metrics, we used concepts from geometric morphometrics, which is the statistical analysis of shape (Zelditch, Swiderski, Sheets, & Fink, 2004). This was done by first running a Generalised Procrustes Analysis (GPA) to standardise the landmark coordinates and remove translation, rotation, and scale effects, essentially leaving only shape information. Two types of facial metrics were calculated using these Procrustes coordinates: a data-driven "face shape IQ" score based purely on face shape information, and specific facial metrics identified by previous research.

Face Shape IQ. From the GPA, shape variables were extracted, which are the decomposition of coordinates into principal components. Shape variables that explained more than 1% of the total variation in face shape (16 PCs) were then entered simultaneously as predictors in a regression analysis with IQ as the outcome variable. From the regression equation, we can calculate the predicted IQ score based solely on facial shape information. Overall, the regression equation significantly predicted IQ (R^2 = .02, p < .001), indicating that face shape was related to IQ. This method is described in detail in Zelditch et al. (2004) and has previously been used to assess shape components of continuous variables in face research (Lee et al., 2016). All shape analyses were conducting using the geomorph package in the R statistical software (Adams & Otárola-Castillo, 2013).

Facial Height-to-width Ratio: Face height-to-width ratio was calculated as the height of the face (distance from the centre of the hairline to the chin) divided by the width of the face (between the outer edges of the most prominent part of the cheekbones).

Interpupillary Distance: Interpupillary distance was calculated as the width between the two pupils relative to the width of the face.

Nose Size: Nose size was calculated as the height from the centre of the bridge of the nose to the bottom of the nose relative to the height of the face (forehead to chin) multiplied by width of the nose (from each nostril) relative to the width of the face. An analogous method has been previous used to calculate eye size (Cunningham, 1986; Talamas, Mavor, Axelsson, et al., 2016).

Eyelid Openness: Eyelid openness was calculated using the same method as Talamas, Mavor, Axelsson, et al. (2016), by taking the vertical distance from the centre of the pupil to the top of the eyelid and dividing it by the width from each corner of the eye. This was calculated for both the left and right eye separately and then averaged.

Mouth Curvature: Mouth curvature was calculated using the same method as Talamas, Mavor, Axelsson, et al. (2016), by taking the average height of the right and left corners of the mouth, subtracting the height of the centre of the mouth, and then dividing by the width of the mouth.

IQ

For participants in BATS, general intelligence (IQ) was measured using The Multidimensional Aptitude Battery (MAB; Jackson, 1984). The scale includes three verbal (information, arithmetic, and vocabulary) and two performance (object and spatial) sub-tests, which were combined to form a full-scale score for general intelligence. The test was administered to each participant separately using the standard MAB instructions. Participants were given 7 minutes for each sub-test, which consisted of multiple-choice questions patterned after the WAIS-R. For more details on how the MAB was administered, see Wright, Smith, Geffen, Geffen, and Martin (2000). IQ was measured on the same day as the facial photographs were taken. The mean IQ from this sample was 112.21 (*SD* = 12.80).

For participants in the LTS, when participants were aged between the ages of 16 to 20 years, they completed the Wechsler Adult Intelligence Scale, 3rd edition (WAIS-III). IQ was operationalised as the sum of the scaled scores on all 11 sub-tests of the WAIS-III. The intelligence tests for the LTS twins were taken on average 3.19 years before the facial photographs were taken (*SD* = 2.92). The mean IQ from this sample was 102.43 (*SD* = 11.53).

To combine the separate measures of intelligence so that the BATS and the LTS participants could be analysed together, IQ scores were standardised within the separate samples before being combined. Previous work has found that the MAB and the WAIS have substantial overlap on total scores (r = .81; Carless, 2000; Jackson, 1984).

Statistical Analysis

To test for the hypothesised mediated relationships, we first ran correlations between each facial metric score and both the ratings of perceived intelligence and measured IQ. If the facial metric was significantly correlated with both, we ran a mediation analysis using the mediation package in the R statistical software (Tingley, Yamamoto, Hirose, Keele, & Imai, 2014). Estimates and significance were tested using a quasi-Bayesian Monte Carlo approximation (for more information, see Imai, Keele, & Tingley, 2010).

To assess the heritability of perceived intelligence and whether it shares a genetic component with IQ, we used the classical twin design. Given that identical twins share all their genes, while nonidentical twins only share, on average, 50% of their segregating genes, and that all twins completely share family environment, we can partition the variance in any given trait into three sources: additive genetic (A), shared environmental (C), and residual (E) sources. As is standard for twin-family designs, we conducted maximum-likelihood modelling, which determines the combination of A, C, and E that best matches the observed data (for more information, see Neale & Cardon, 1992; Posthuma et al., 2003). All analyses were conducted in OpenMx package in the R statistical software (Boker et al., 2011).

Results

While analyses reported here combine male and female participants, we note that we also ran each analyses separated by sex. We found no difference in the pattern of results between males and females except where noted below. We also conducted the analyses where IQ scores were not standardised prior to combining the samples and including cohort as a binary covariate; this did not influence the pattern of results, suggesting results are not due to differences in IQ testing between samples.

Perceived Intelligence and IQ

There was a significant positive phenotypic correlation between perceived intelligence and IQ (r = .15, p < .001), which suggests that perceivers may, to some extent, be able to accurately evaluate intelligence based on facial features. We also found a significant correlation between perceived intelligence and facial attractiveness (r = .34, p < .001); however, as noted before (Mitchem et al, 2015), there is no association between measured intelligence and facial attractiveness in our data (r = .01, p = .517). Accordingly, the association between perceived intelligence and IQ remained when controlling for facial attractiveness, as well as with other facial attributes.

Even though we find a positive association between perceived intelligence and IQ, it is unclear whether this is due to any particular facial attributes. Therefore, we conducted mediation analyses, first with predicted IQ score based on overall face shape information, but also with specific facial metrics previously associated with perceptions of intelligence. As shown in Table 1., predicted IQ based on face shape was significantly correlated with perceived intelligence. All facial metrics previously found to be associated with perceived intelligence were replicated in our data in the expected direction, though of these, only taller height and wider interpupillary distance were also significantly correlated with measured IQ (see Table 1.).

Figure 2 shows the visualisations of face shape associated with perceived intelligence and IQ. Apart from the facial features identified by previous research, Figure 2 may hint at other subtle features that could be associated with perceptions of intelligence. For instance, a more upturned nose or a more square jaw could potentially be associated with lower perceptions of intelligence, though this requires further investigation. The face shape differences between low and high IQ are much subtler compared to the difference between low and high perceived intelligence.

Table 1. Correlations between eyelid openness and mouth curvature with perceived intelligence and IQ. ($N = 1660$)

	Perceived Intelligence	IQ
Predicted IQ based on face shape	.11***	.17***
Face Height	.11***	.06*
Interpupillary Distance	.08**	.06**
Nose Size	.09***	.04
Eyelid Openness	.12***	.01
Mouth Curvature	.25***	.003

* $p < .05$ ** $p < .01$ *** $p < .001$. Taller faces, wider set eyes, larger noses, more open eyes, and more curved mouths were associated with greater perceived intelligence.

Figure 2. Face shape visualisations of low (left) and high (right) perceived intelligence (top) and IQ (bottom). Each visualisation is 3SD from the mean face shape.

We ran a mediation model for each facial metric associated with both perceived intelligence and measured IQ. Table 2. reports the mediation analyses of predicted IQ based on face shape, face height, and interpupillary distance. We found significant mediation effects of predicted IQ based on face shape and interpupillary distance on the relationship between perceived and actual intelligence, suggesting that these facial metrics are used by observers to accurately estimate intelligence.

Table 2. Mediation of the association between measured IQ and perceived intelligence by face height, interpupillary distance, and predicted IQ based on face shape.

	Predicted IQ Based on Face Shape	Face Height	Interpupillary Distance
Mediation Effect	.02 [.01, .03] $p < .001$	.005 [-.0007, .01] $p = .09$	.005 [.0003, .01] $p = .03$
Direct Effect	.15 [.09, .20] $p < .001$	.16 [.11, .22] $p < .001$	.16 [.11, .22] $p < .001$
Total Effect	.17 [.11, .22] $p < .001$	.17 [.12, .22] $p < .001$	.17 [.11, .22] $p < .001$
Proportion of Total Effect via Mediation	.11 [.06, .20] $p < .001$	.03 [-.004, .08] $p = .09$	.03 [.002, .07] $p = .03$

For participants in the BATS, data on the genetic population structures determined via principal components analysis of ~600,000 single nucleotide polymorphisms (which often represents genetic ancestry; see Patterson, Price, & Reich, 2006) were available. To ensure ethnicity did not confound the association between measured IQ and perceptions of intelligence, the above analyses were also conducted only with participants in the BATS and included the first 5 ancestry principal components as covariates. The pattern of significance remained the same as reported above, with the exception that the association between perceived intelligence and nose size was non-significant.

Twin Modelling

In the following models, controlling for facial attractiveness did not change the pattern of results; therefore, we report here only the results for perceived intelligence not controlling for facial attractiveness. All models included participant age as a covariate.

There were no significant differences between twin and siblings in means and variances for perceived intelligence ($\chi^2(1) = 1.18$, $p = .552$ and $\chi^2(1) = .78$, $p = .677$ for means and variances respectively), but measured IQ had a significantly lower mean and variance in twins compared to their siblings ($\chi^2(1) = 25.70$, $p < .001$ and $\chi^2(1) = 8.42$, $p = .015$ for means and variances respectively). We tested models where means for IQ were either equated or not equated between twins and siblings; the pattern of results did not change between the two, so we report here the model where means are equated. However, men had a significantly higher mean in both perceived intelligence and IQ than women ($\chi^2(1) = 10.31$, $p = .001$ and $\chi^2(1) = 28.88$, $p < .001$ for perceived intelligence and IQ respectively) but no significant differences were found for variances of perceived intelligence and IQ between the sexes ($\chi^2(1) = .78$, $p = .500$ and $\chi^2(1) = 1.71$, $p = .191$ for perceived intelligence and IQ respectively). Therefore, means for males and females were not equated in the model.

Twin-pair correlations for perceived intelligence are reported in Table 3. Overall, for both perceived intelligence and IQ, correlations between MZ twin pairs were significantly larger than that for DZ twin pairs, which suggests that there are genetic components for both. Estimated components from ACE models for perceived intelligence and IQ are reported in Table 4. For perceived intelligence, results were inconsistent between males and females; we found with males there was a significant proportion attributable to genetic sources and not shared environmental sources, while the opposite was true for females. However, we found that there was no significant difference between male and female twin correlations on perceived intelligence within zygosity $\chi^2(2) = 2.21$, $p = .331$, and when the sexes were pooled, we found a significant genetic component of perceived intelligence. Consistent with previous findings, there was a large heritable component for IQ (Haworth et al., 2010).

Table 3. Twin-pair correlations (r and 95% CI) on perceived intelligence and IQ.

Zygosity Group	Perceived Intelligence	IQ
All Identical Twins	**.44 [.33, .55]**	**.86 [.77, .96]**
Identical Female Twins	.43 [.28, .62]	.83 [.71, .96]
Identical Male Twins	.45 [.29, .62]	.90 [.77, 1.00]
All Non-Identical Twins + Siblings	**.27 [.19, .35]**	**.44 [.37, .53]**
Non-Identical Female Twins + Siblings	.37 [.24, .50]	.51 [.40, .64]
Non-Identical Male Twins + Siblings	.28 [.12, .43]	.42 [.29, .57]
Non-Identical Opposite-Sex Twins + Siblings	.17 [.02, .31]	.43 [.32, .54]

Table 4. Proportions of variance of perceived intelligence and IQ estimated to be accounted for by A (additive genetic), C (shared environmental), and E (residual) influences.

	Perceived Intelligence			IQ		
	A	C	E	A	C	E
Females	.15 [.00, .47]	.29 [.03, .47]	.56 [.45, .68]	.57 [.40, .77]	.28 [.09, .43]	.15 [.12, .20]
Males	.47 [.04, .58]	.00 [.00, .34]	.53 [.42, .66]	.84 [.73, .89]	.02 [.00, .12]	.13 [.10, .18]
Overall	.37 [.14, .53]	.09 [.00, .25]	.54 [.46, .64]	.77 [.64, .87]	.09 [.00, .21]	.14 [.12, .17]

In order to determine if perceived intelligence and IQ share a genetic component, we ran common factors bivariate models for each sex separately and also with the sexes pooled. In the overall sample, the correlation between the genetic components for perceived intelligence and IQ was not significant (*Ar* = .06, *95% CI* = -.17, .25). The genetic correlation was also non-significant for males (*Ar* = .12, *95% CI* = -.15, .34), while no meaningful estimate could be made for females given the lack of significant A for perceived intelligence. Similarly, no meaningful shared environmental correlation could be estimated for males or the overall sample given the lack of significant C for IQ, though the shared environmental correlation was also non-significant for females (*Cr* = .34, *95% CI* = -.08, .81). These non-significant correlations are likely due to a lack of power, as running the model combining familial factors (A + C) found a significant familial correlation across all groups (see Table 5.). The residual correlation was near-zero in all cases; therefore, we can be confident that familial factors are driving the correlation between perceived and measured intelligence.

Table 5. Estimated components for the common factors models, including A + C (familial) and E (residual) components and the respective correlations for perceived intelligence and IQ.

	Perceived Intelligence		IQ			
	A + C	E	A + C	E	Familial Correlation	Residual Correlation
Female	.47 [.36, .57]	.53 [.43, .64]	.85 [.81, .88]	.15 [.12, .19]	.26 [.14, .38]	.02 [-.12, .16]
Male	.46 [.33, .57]	.54 [.43, .67]	.86 [.82, .90]	.14 [.14, .38]	.21 [.16, .34]	-.01 [-.18, .16]
Overall	.47 [.39, .54]	.53 [.46, .61]	.86 [.83, .88]	.14 [.12, .17]	.24 [.15, .33]	.002 [-.11, .11]

Discussion

First, our results support the notion that perceptions of people's intelligence based on their facial features could, in part, reflect their actual intelligence. We found a correlation between perceived intelligence and measured IQ of similar magnitude to previous research that found an association in smaller samples (e.g., Zebrowitz et al., 2002). This relationship persisted even when controlling for physical attractiveness, suggesting such a relationship was not solely driven by a halo effect, as has been proposed previously (Langlois et al., 2000; Talamas, Mavor, & Perrett, 2016). Prior research did not find an association between perceptions of intelligence and actual intelligence with adolescent faces (Zebrowitz et al., 2002); this is inconsistent with our data in which we observed a significant association despite the sample being primarily adolescents.

Further, we found that overall face shape and specific spatial measures mediated the relationship between predicted intelligence and measured IQ. This suggests that observers used face shape information to accurately judge intelligence. Of the specific facial attributes investigated, we found that taller face height, wider interpupillary distance, and larger nose size were all associated with perceptions of intelligence consistent with Kleisner et al. (2014). In addition, we found that taller face height and wider interpupillary distance were also associated with measured IQ, and that interpupillary distance significantly mediated the relationship between perceived intelligence and measured IQ. This is contrary to Kleisner et al. (2014), who found no association between measured intelligence with either face height or interpupillary distance. A likely reason for the discrepancy between Kleisner et al. (2014) and our study is that Kleisner et al. (2014) were underpowered to detect small effects, because their sample size was 80 faces compared to our 1660 faces. Indeed, the majority of previous studies would have been underpowered to detect effects as small as our results indicate, possibly explaining the mixed findings in the literature with regard to the accuracy of intelligence judgements based on facial photographs. Despite our large sample size, we note that the mediation effect for face height on the association between perceived intelligence and measured IQ fell short of significance ($p = .09$); therefore, any conclusion made about face height underlying the association is discussed tentatively.

Exactly why these stable facial features may be associated with intelligence and perceptions of intelligence is unclear. It is known that certain disorders that can involve intellectual impairment are also associated with particular facial abnormalities (e.g., Hammond & Suttie, 2012). It may be that people learn these associations from real-world observation and factor them into everyday judgements of intelligence. For example, short nose length was associated in our data with judgements of low intelligence, and short nose length is also associated with a number of disorders affecting intellectual development, including fetal alcohol syndrome, Down syndrome, Williams syndrome, Miller-Dieker syndrome, among others (e.g., see Hammond & Suttie, 2012). Further, it could be that subtle associations between face shape and measured IQ in our data reflect much milder disruptions in the same developmental pathways that are severely affected in the aforementioned disorders.

Transitory facial characteristics, such as eyelid openness and mouth curvature, were also associated with perceived intelligence, consistent with Talamas, Mavor, Axelsson, et al. (2016). Even though previous work has found an association between tiredness and cognitive ability (Pilcher & Huffcutt, 1996), we do not necessarily expect facial cues to tiredness to be associated with actual intelligence in our sample. This is because the facial photographs were not taken at the same time as when intelligence was measured, and we could expect tiredness levels to vary greatly between the two. We note, though, that upward mouth curvature and eyelid openness were still not significantly correlated with measured IQ when only considering participants from the BATS, where the facial photographs and intelligence scores were at least taken on the same day. These transitory facial characteristics had a larger effect on perceived intelligence compared to the stable features, which possibly indicates that cues to state (as opposed to trait) are weighted more heavily when making intelligence judgements. The lack of association between these cues to state and measured IQ in our sample may further muddle any association between perceptions of intelligence and actual intelligence. Note that the influences of stable and transitory facial cues are not mutually exclusive and both are likely to contribute to judgements of intelligence.

To test whether there was a genetic component to perceived intelligence, we ran quantitative genetic models. When considering the overall sample with sex pooled, we found a significant proportion of variance in perceived intelligence was attributable to genetic factors. However, when estimating the variance components for perceived intelligence separately for each sex, we found that there was a significant genetic component for males, but a significant environmental component for females. Previous research has proposed that women may place greater importance on intelligence compared to men when choosing a mate (Prokosch, Coss, Scheib, & Blozis, 2009); therefore, this sex difference could possibly reflect differential selection pressure for men (and not women) to develop facial cues to intelligence. We also did not find a significant correlation between genetic or shared environmental influences for perceived intelligence and IQ for men, women, and when sexes were pooled. However, when combining familial effects (A + C) we did find a significant familial correlation across all samples. This suggests that genetic and/or shared environmental sources that influence IQ also likely influence perceived intelligence, though our current data cannot distinguish between the two due to a lack of power. Previous research has proposed that intelligence perceptions reflects underlying genetic quality (Haselton & Miller, 2006; Miller, 2000), though the possibility that non-genetic factors could also contribute to the accuracy of intelligence perceptions has often been neglected. For instance, access to highly nutritional food during development could contribute to both cognitive development and the development of perceptible facial cues. Further research is needed to identify the underlying mechanisms that inform intelligence perceptions.

Our findings are difficult to reconcile theoretically with previous research using the same facial photos and IQ scores that found that no correlation between facial attractiveness and intelligence (Mitchem et al., 2015). Evolutionary theories suggest that it is advantageous to have an intelligent mate, so it follows that facial cues to intelligence should be attractive (Prokosch et al., 2005; Zebrowitz & Rhodes, 2004); however, other previous research on the link between attractiveness and intelligence found no association (Feingold, 1992; Langlois et al., 2000), and results are also mixed for whether perceived intelligence is preferred under contexts where genetic benefits are more beneficial (Haselton & Miller, 2006; Moore et al., 2014). An alternative possibility that has not been explored is that intelligence judgements may be advantageous in other domains, such as choosing intelligent individuals with whom to cooperate, or, during competition, estimating the formidability of opponents based on their intelligence.

Here, we have focused on facial morphology, though perceptions of intelligence are also likely to be influenced by other traits, such as body shape, movement, or contextual information (e.g., grooming and choice of clothing). Future research could investigate the accuracy of intelligence perception using other stimuli, such as body images, dynamic facial images, or face-to-face interactions. Also, future research could investigate other cognitive abilities purported to reflect genetic quality, such as musical performance, humour, or artistic skills (Miller, 2000).

Apart from the limitations already mentioned, the classical twin design also has inherent limitations, such as the inability to simultaneously estimate shared environmental (C) and non-additive genetic (D) variance. This may be particularly useful given the inconsistencies in estimated variance components in perceived intelligence between men and women, but would require additional observations from other family members (e.g., parents). Participants in our sample of facial photos were also all in late adolescence or early adulthood, at which time it is unclear whether cues to intelligence would have fully developed. Even though IQ tends to stabilise by early adolescence through to adulthood (Deary, Whiteman, Starr, Whalley, & Fox, 2004; Hertzog & Schaie, 1986), and facial dimensions are 94% of their adult size by age 16 (Edwards et al., 2007), facial cues to intelligence could develop later in life; for example, if cues to intelligence are due to repeated habitual expressions which only manifests in facial appearance over time. As such, future research should investigate intelligence perceptions in an older sample. Finally, we note again that the facial photos were not standardised; as well as precluding any absolute measures of face (or face dimension) size, this probably contributed to error which would have weakened the observed association between perceived intelligence and measured IQ.

In conclusion, we add to the literature that individuals are able, to some extent, to accurately assess intelligence based on facial photographs. In particular, our results suggest that facial shape information helps inform these judgements, and of the facial traits suggested by previous research, interpupillary distance significantly mediated this relationship (such that wide-set eyes was associated with intelligence). Also, our findings replicate previous research that identified certain facial attributes that were associated with perceptions of intelligence, including both stable cues (taller face height, wider interpupillary distance, and greater nose size) and transitory cues (eyelid openness and upward mouth curvature).

Acknowledgements

AJL has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 705478. This work was further supported by grants from the Australian Research Council (A79600334, A79801419, DP0212016, FT160100298) and the National Institute of Mental Health (MH085812 and MH63207). Thanks to Marlene Grace, Ann Eldridge, Daniel Park, David Smyth, Kerrie McAloney, Natalie Garden, and Reshika Chand; to the professional research assistants at the Center on Antisocial Drug Dependence for their work with the Longitudinal Twin Study; and to the volunteer research assistants who assigned trait ratings. And, thanks to the Queensland Twin (QTwin) Registry and Colorado Twin Registry twins and their families for their continued participation.

References

Adams, D. C., & Otárola-Castillo, E. (2013). Geomorph: An R package for the collection and analysis of geometric morphometric shape data. *Methods in Ecology and Evolution*, *4*(4), 393-399.

Arden, R., Gottfredson, L. S., & Miller, G. (2009). Does a fitness factor contribute to the association between intelligence and health outcomes? Evidence from medical abnormality counts among 3654 US Veterans. *Intelligence, 37*(6), 581-591.

Boker, S., Neale, M. C., Hermine, M., Wilde, M., Spiegel, M., Brick, T., . . . Fox, J. (2011). OpenMx: An open source extended structural quation modeling framework. *Psychometrika, 76*(2), 306-317.

Borkenau, P., & Liebler, A. (1995). Observable attributes as manifestations and cues of personality and intelligence. *Journal of Personality, 63*(1), 1-25.

Brophy, J. E. (1983). Research on the self-fulfilling propehcy and teacher expectations. *Journal of Educational Psychology*, *75*(5), 631-661.

Carless, S. A. (2000). The validity of scores on the Multidimensional Aptitude Battery. *Education and Psychological Measurement, 60*(4), 592-603.

Carney, D. R., Colvin, C. R., & Hall, J. A. (2007). A thin slice perspective on the accuracy of first impressions. *Journal of Research in Personality, 41*(5), 1054-1072.

Cunningham, M. R. (1986). Measuring the physical in physical attractiveness: Quasi-Experiments on the sociobiology of female facial beauty. *Interpersonal Relations and Group Processes, 50*(5), 925-935.

Deary, I. J., Whiteman, M. C., Starr, J. M., Whalley, L. J., & Fox, H. C. (2004). The impact of childhood intelligence on later life: Following up the Scottish Mental Surveys of 1932 and 1947. *Journal of Personality and Social Psychology*, *86*(1), 130-147.

Denny, K. (2008). Beuaty and intelligence may - or may not - be related. *Intelligence*, *36*, 616-618.

Dunkel, C. S., & Murphy, N. A. (2014). Predicting intellectual ability and scholastic outcomes with a single item: From early childhood to adulthood. *Journal of Intelligence, 2*(3), 68-81.

Edwards, C. B., Marshall, S. D., Qian, F., Southard, K. A., Franciscus, R. G., & Southard, T. E. (2007). Longitudinal study of facial skeletal growth completion in 3 dimensions. *American Journal of Orthodontics and Dentofacial Orthopedics, 132*(6), 762-768.

Eppig, C., Fincher, C. L., & Thornhill, R. (2010). Parasite prevalence and the worldwide distribution of cognitive ability. *Proceedings of the Royal Society B-Biological Sciences, 277*(1701), 3801-3808.

Eppig, C., Fincher, C. L., & Thornhill, R. (2011). Parasite prevalence and the distribution of intelligence among the states of the USA. *Intelligence, 39*(2-3), 155-160.

Feingold, A. (1992). Good-looking people are not what we think. *Psychological Bulletin, 111*(2), 304-341.

Gangestad, S. W., Thornhill, R., & Garver-Apgar, C. E. (2010). Men's facial masculinity predicts changes in their female partners' sexual interests across the ovulatory cycle, whereas men's intelligence does not. *Evolution and Human Behavior, 31*(6), 412-424.

Hammond, P., & Suttie, M. (2012). Large-scale objective phenotyping of 3D facial morphology. *Human Mutation, 33*(5), 817-825.

Haselton, M. G., & Miller, G. F. (2006). Women's fertility across the cycle increases the short-term attractiveness of creative intelligence. *Human Nature, 17*(1), 50-73.

Haworth, C. M. A., Wright, M. J., Luciano, M., Martin, N. G., de Geus, E. J. C., van Beijsterveldt, C. E. M., . . . Plomin, R. (2010). The heritability of general cognitive ability increases linearly from childhood to young adulthood. *Molecular Psychiatry, 15*, 1112-1120.

Hertzog, C., & Schaie, K. W. (1986). Stability and change in adult intelligence: 1. Analysis of longitudinal covariance structures. *Psychology and Aging, 1*(2), 159-171.

Howrigan, D. P., Simonson, M. A., Davies, G., Harris, S. E., Tenesa, A., Starr, M., . . . Keller, M. C. (2016). Genome-wide autozygosity is associated with lower general cognitive ability. *Molecular Psychiatry, 21*, 837-843.

Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods, 15*(4), 309-334.

Jackson, D. N. (1984). *Multidimensional aptitude battery manual*. Port Huron, MI: Research Psychologists Press.

Judice, T. N., & Neuberg, S. L. (1998). When interviewers desire to confirm negative expectations: Self-fulfilling prophecies and inflated applicant self-perceptions. *Basic and Applied Social Psychology, 20*(3), 175-190.

Jussim, L. (1989). Teacher expectations: self-fulfilling prophecies, perceptual biases, and accuracy. *Journal of Personality and Social Psychology, 57*(3), 496-480.

Jussim, L., & Harber, K. D. (2005). Teacher expectations and self-fulfilling prophecies: Knowns and unknowns, resolved and unresolved controversies. *Personality and Social Psychology Review, 9*(2), 131-155.

Kanazawa, S., & Kovar, J. L. (2004). Why beautiful people are more intelligent. *Intelligence, 32*(3), 227-243.

Kleisner, K., Chvatalova, V., & Flegr, J. (2014). Perceived intelligence is associated with measured intelligence in men but not women. *PLoS ONE, 9*(3), e81237.

Kokko, H., Brooks, R., Jennions, M. D., & Morley, J. (2003). The evolution of mate choice and mating biases. *Proceedings of the Royal Society B-Biological Sciences, 270*, 653-664.

Kokko, H., Brooks, R., McNamara, J. M., & Houston, A. I. (2002). The sexual selection continuum. *Proceedings of the Royal Society B-Biological Sciences, 269*, 1331-1340.

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. *Psychological Bulletin, 126*(3), 390-423.

Lee, A. J., Mitchem, D. G., Wright, M. J., Martin, N. G., Keller, M. C., & Zietsch, B. P. (2016). Facial averageness and genetic quality: testing heritability, genetic correlation with attractiveness, and the paternal age effect. *Evolution and Human Behavior, 37*, 61-66.

Miller, G. (2000). *The Mating Mind*. New York: Doubleday.

Milonoff, M., & Nummi, P. (2012). Adolescents but not older women misjudge intelligence from faces and do not consider intelligent-looking men attractive. *Annales Zoologici Fennici, 49*(5), 378-384.

Mitchem, D. G., Zietsch, B. P., Wright, M. J., Martin, N. G., Hewitt, J. K., & Keller, M. C. (2015). No relationship between intelligence and facial attractiveness in a large, genetically informative sample. *Evolution and Human Behavior*, *36*, 240-247.

Moore, F. R., Law Smith, M. J., & Perrett, D. I. (2014). Individual differences in preferences for cues to intelligence in the face. *Intelligence, 44*, 19-25.

Murphy, N. A., Hall, J. A., & Colvin, C. R. (2003). Accurate intelligence assessments in social interactions: mediators and gender effects. *Journal of Personality, 71*(3), 465-493.

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

Olivola, C. Y., & Todorov, A. (2010). Fooled by first impressions? Reexamining the diagnostic value of appearance-based inferences. *Journal of Experimental Social Psychology*, *46*(2), 315-324.

Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. *PLOS Genetics, 2*(12), e190.

Penke, L., Borsboom, D., Johnson, W., Kievit, R. A., Ploeger, A., & Wicherts, J. M. (2011). Evolutionary psychology and intelligence research cannot be integrated the way Kanazawa (2010) suggested. *American Psychologist, 66*(9), 916-917.

Pilcher, J. J., & Huffcutt, A. J. (1996). Effects of sleep deprivation on performance: A meta-analysis. *Sleep: Journal of Sleep Research & Sleep Medicine*, *19*(4), 318-326.

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.

Prokosch, M. D., Coss, R. G., Scheib, J. E., & Blozis, S. A. (2009). Intelligence and mate choice: intelligent men are always appealing. *Evolution and Human Behavior, 30*(1), 11-20.

Prokosch, M. D., Yeo, R. A., & Miller, G. F. (2005). Intelligence tests with higher g-loadings show higher correlations with body symmetry: Evidence for a general fitness factor mediated by developmental stability. *Intelligence, 33*(2), 203-213.

Reynolds, D. J., & Gifford, R. (2001). The sounds and sights of intelligence: A lens model channel analysis. *Personality and Social Psychology Bulletin, 27*(2), 187-200.

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.

Spisak, B. R., Blaker, N. M., Lefevre, C. E., Moore, F. R., & Krebbers, K. F. B. (2014). A face for all seasons: Searching for context-specific leadership traits and discovering a general preference for perceived health. *Frontiers in Human Neuroscience, 8*.

Talamas, S. N., Mavor, K. I., Axelsson, J., Sundelin, T., & Perrett, D. I. (2016). Eyelid-openness and mouth curvature influence perceived inelligence beyond attractiveness. *Journal of Experimental Psychology: General*, *145*(5), 603-620.

Talamas, S. N., Mavor, K. I., & Perrett, D. I. (2016). Blinded by beauty: attractiveness bias and accurate perceptions of academic performance. *PLoS ONE, 11*(2), e0148284.

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). mediation: R package for causal mediation analysis. *Journal of Statistical Software, 59*(5), 1-38.

Tsukahara, J. S., Harrison, T. L., & Engle, R. W. (2016). The relationship between baseline pupil size and intelligence. *Cognitive Psychology, 91*, 109-123.

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.

Wright, M. J., Smith, G. A., Geffen, G. M., Geffen, L. B., & Martin, N. G. (2000). Genetic influence on the variance in coincidence timing and its covariance with IQ: A twin study. *Intelligence, 28*(4), 239-250.

Yeo, R., Gangestad, S., Liu, J. Y., Wassink, T., & Calhoun, V. (2011). Rare copy number deletions and intelligence in schizophrenic patients and healthy controls. *Behavior Genetics, 41*(6), 943-944.

Zebrowitz, L. A., Hall, J. A., Murphy, N. A., & Rhodes, G. (2002). Looking smart and looking good: Facial cues to intelligence and their origins. *Personality and Social Psychology Bulletin, 28*(2), 238-249.

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.