Individual Aesthetic Preferences for Faces Are Shaped Mostly by Environments, Not Genes

Highlights

- Differences in how people judge face attractiveness can be reliably measured
- Individual face preferences are primarily explained by differences in environments
- In contrast, face identity recognition is explained primarily by genetic variation
- Different domains of social judgment/face perception have distinct etiologies

Authors

Laura Germine, Richard Russell, P. Matthew Bronstad, ..., Ken Nakayama, Gillian Rhodes, Jeremy B. Wilmer

Correspondence

lgermine@gmail.com

In Brief

People differ from one another in which faces they find more and less attractive. Germine et al. report that these differences can be explained by differences in our unique environments. This finding contrasts with another core aspect of the way we process faces, which is explained by differences in our genes.
Individual Aesthetic Preferences for Faces Are Shaped Mostly by Environments, Not Genes

Laura Germine,1,2,3,* Richard Russell,4 P. Matthew Bronstad,5 Gabriëlla A.M. Blokland,1,3 Jordan W. Smoller,1,3 Holm Kwok,6 Samuel E. Anthony,2 Ken Nakayama,2 Gillian Rhodes,7 and Jeremy B. Wilmer6
1Psychiatric and Neurodevelopmental Genetics Unit, Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA 02114, USA
2Department of Psychology, Harvard University, Cambridge, MA 02138, USA
3Department of Psychiatry, Harvard Medical School, Boston, MA 02115, USA
4Psychology Department, Gettysburg College, Gettysburg, PA 17325, USA
5Massachusetts Eye and Ear Infirmary, Harvard Medical School, Boston, MA 02114, USA
6Department of Psychology, Wellesley College, Wellesley, MA 02481, USA
7ARC Centre of Excellence in Cognition and Its Disorders, School of Psychology, University of Western Australia, Crawley, WA 6009, Australia
*Correspondence: lgermine@gmail.com
http://dx.doi.org/10.1016/j.cub.2015.08.048

SUMMARY

Although certain characteristics of human faces are broadly considered more attractive (e.g., symmetry, averageness), people also routinely disagree with each other on the relative attractiveness of faces. That is, to some significant degree, beauty is in the “eye of the beholder.” Here, we investigate the origins of these individual differences in face preferences using a twin design, allowing us to estimate the relative contributions of genetic and environmental variation to individual face attractiveness judgments or face preferences. We first show that individual face preferences (IP) can be reliably measured and are readily dissociable from other types of attractiveness judgments (e.g., judgments of scenes, objects). Next, we show that individual face preferences result primarily from environments that are unique to each individual. This is in striking contrast to individual differences in face identity recognition, which result primarily from variations in genes [1]. We thus complete an etiological double dissociation between two core domains of social perception (judgments of identity versus attractiveness) within the same visual stimulus (the face). At the same time, we provide an example, rare in behavioral genetics, of a reliably and objectively measured behavioral characteristic where variations are shaped mostly by the environment. The large impact of experience on individual face preferences provides a novel window into the evolution and architecture of the social brain, while lending new empirical support to the long-standing claim that environments shape individual notions of what is attractive.

RESULTS AND DISCUSSION

To understand the origins of individual preferences for certain faces, we collected face attractiveness ratings from 547 identical twin pairs (monozygotic or MZ twins) and 214 same-sex non-identical twin pairs (dizygotic or DZ twins) (see Figure 1A). Consistent with the large literature regarding universal aspects of face preferences and attractiveness [2–7], there was substantial agreement across participants about which faces were more and less attractive: mean ratings from male participants, for example, correlated at ceiling with mean face ratings from female participants, across both female faces (r(100) = 0.99; 102 female faces) and male faces (r(96) = 0.99; 98 male faces; see Figures 1B and 1C). Average aesthetic preferences, however, can mask substantial individual differences [8–10] (see Figures 1D and 1E). Selecting two participants at random produced an average of only 48% agreement (and 52% disagreement) in face preferences (see Figure 1F), even after removing apparent disagreements that could be explained away as self-inconsistency (Supplemental Information). This estimate is consistent with previous literature [11, 12] as well as with the everyday experience that on the one hand, fashion models can “make a fortune with their good looks,” while on the other hand, friends can “endlessly debate about who is attractive and who is not” [11]. To capture individual differences in face preferences, we first estimated the proportion of variation in each participant’s ratings that was unique to that participant and not explained by average ratings, based on the correlation between each participant’s ratings and the average person’s ratings [10–13]. For our subsequent analyses, we transformed correlations between individual participant and average face ratings to Z scores using Fisher’s r-to-z transformation to remove the inherent skew in correlation values [14] and then removed variance attributable to differences in response consistency or intra-individual variability (see Experimental Procedures). This procedure yielded reliable individual preference scores (IP scores; see Figure S1A) for each participant, with larger numbers indicating greater agreement between the participant’s ratings and the mean ratings for each of the faces. These IP scores for faces were only modestly correlated with IP scores for abstract objects (r(2,197) = 0.07, 95% confidence interval [CI]: 0.03–0.11) and scenes (r(2,197) = 0.28, 95% CI: 0.24–0.32; see Figure S1B) [13]. To get at face-specific preferences, we regressed out both object and scene IP. These face-specific IP scores remained highly reliable (split-half reliability = 0.88; test-retest
reliability = 0.75; see Figure S1C, Experimental Procedures, and Supplemental Information). We used standardized (Z scored) face-specific IP scores (hereafter called face IP scores) for all subsequent analyses.

Next, we estimated the contributions of genetic and environmental factors to face IP by comparing the correlation of face IP scores among MZ twins with the correlation of face IP scores among DZ twins. Although MZ and DZ twins share family environment to a similar extent, MZ twins share, on average, twice as much of their genetic variation as DZ twins. The correlations for face IP scores between MZ twins and between DZ twins can thus be used to estimate the proportion of variation in face IP that can be explained by variations in genes, shared environments, and unshared environments. We calculated a maximum likelihood correlation of 0.22 (95% CI: 0.14–0.29) for MZ twins and 0.09 (95% CI: −0.06–0.24) for DZ twins. These two correlations did not significantly differ (Fisher r-to-z transformation; p = 0.1), indicating that most of the variance in face IP is likely attributable to environmental factors. To obtain a more precise estimate of the contributions of genetic and environmental factors to face IP, we fit a standard ACE twin model that includes additive genetic influences (A), shared environmental influences (C), as well as unshared environmental influences and measurement error (E) using structural equation modeling techniques [15]. Controlling for age and sex, the ACE model attributed 22% of variance to (A) additive genetic factors, 0% to (C) shared environmental factors, and 78% to (E) individual or unshared environment/measurement error (see Figure 2B and Table 1).

We compared the full ACE model with reduced AE, CE, and E models (respectively setting the contributions of the C parameter, A parameter, and A + C parameters to zero) and determined the AE model yielded the best fit based on Akaike’s information criterion.
around each

This dissociation in heritability suggests that there are distinct genetic and etiological mechanisms underlying these two core social-perceptual phenotypes.

of the reliable variation in face recognition is due to variations in genes, and (2) most of the reliable variation in face IP is attributable to variations in environments.

estimate. Arrows indicate the upper boundary for “A” estimates, based on the test-retest reliability of each measure. Based on AE model estimates, (1) almost all

genetic and environmental contributions to both face recognition and face IP scores (face preferences) are shown, with error bars indicating 95% CIs around each estimate. Arrows indicate the upper boundary for “A” estimates, based on the test-retest reliability of each measure. Based on AE model estimates, (1) almost all of the reliable variation in face recognition is due to variations in genes, and (2) most of the reliable variation in face IP is attributable to variations in environments. This dissociation in heritability suggests that there are distinct genetic and etiological mechanisms underlying these two core social-perceptual phenotypes. See also Figure S1 and Table S2.

criterion (see Table 1 and Table S2). The AE model gave similar point estimates for both A (22%) and E (78%) parameters, but with tighter confidence intervals (see Figure 2C and Table 1). We conclude that most of the reliable variation in face IP was explained by the influence of unshared or individual environment with a relatively small contribution from genetic variation and little to no contribution from shared environment.

Our results provide a rare example of a complex, objectively measured, highly reliable, and specific behavioral characteristic that is shaped predominantly by environmental factors. Although high estimates of unshared environment contributions to social cognition and behavior are reported in twin studies [17], these estimates often occur in the context of low or unknown reliability [18, 19]. High measurement error (low reliability) spuriously reduces estimates of familial resemblance from both genetic and shared environmental factors and spuriously inflates estimates of unshared environmental contributions. Apparent examples of high unshared environment contributions are often confounded with measurement error [20]. Given the high reliability of face IP, even when estimated conservatively (via an alternate forms test-retest procedure), we conclude that the contribution of unshared environmental factors to face IP cannot be explained by unreliable measurement. Instead, our findings support the notion that individual aesthetic face preferences are truly shaped primarily by individual life experiences [21, 22].

The observed results isolate a highly specific environmental influence that impacts face IP independently of scene IP and abstract object IP. But does this environmental influence act specifically on face attractiveness judgments? Alternatively, it might act broadly on any judgment that involves a face or on any social judgment. As a strong test of specificity, we consider the case of face identity recognition. Face attractiveness judgments and face identity recognition both involve social evaluation of faces, in the visual domain. Moreover, both require processing of invariant face characteristics, which are known to rely upon inferior occipital and inferior temporal brain regions [23], and deficits in both have been found to coexist in patients [23–25]. If the etiology of face identity processing were to differ from that of face IP, then that would provide strong evidence that the observed environmental effect is specific not only to social stimuli in general (or to faces in particular, or even to judgments of invariant face characteristics) but rather to a particular subset of judgments of invariant face characteristics. We previously measured face recognition in another sample of MZ and DZ twins drawn from the same Australian Twin Registry [1] (see also [26]). While highly reliable, the face recognition measure was no more reliable than our face IP measure (Cambridge Face Memory Test scores: internal reliability = 0.89, test-retest reliability = 0.70; face IP scores: internal reliability = 0.88, test-retest reliability = 0.75). Yet despite equal precision of measurement, a sample drawn from the same population, and similarly robust evidence for independence from various non-face categories, we found little to no impact of environment on face recognition ability. Genetic variation accounted for most or all of the reliable face recognition variance, in contrast with face IP (68% versus 22% heritability; p of difference < 1E–14; see Figure 2C and Table 1). Indeed, looking across the behavioral genetic literature, face IP is among the most environmental objectively measured behavioral traits, whereas face identity recognition is among the most heritable [1, 27]. We conclude from this etiological dissociation that the observed environmental effect is highly specific to face attractiveness judgments.

Previous evidence has indicated that preferences for particular faces or face characteristics are shaped by a range of factors, including personality preferences [28], the rater’s own facial characteristics [29], features of the socioeconomic and cultural environment [30–34], previous visual experience [35–39], and history of social learning [19, 40–44]. Individual
preferences for faces are also correlated among friends and spouses [12]. In our sample, most of the variations in face preferences were explained by the contribution of unshared environment—those aspects of the environment that are unique to individuals and not shared between twins. Our data suggest that individual life history and experience are a driving force behind individual face preferences [22].

Does this mean that shared environments are not important for individual face preferences? Not necessarily. Our study was conducted with a relatively homogeneous sample of Australian twins [45]. Given the sociocultural homogeneity of our sample, the low contribution of genetic variance to face IP is particularly noteworthy: estimates of genetic contributions tend to be higher where environments are less variable [46].

We have demonstrated, in the context of a sensitive behavioral genetic investigation, that what scholars in the humanities and arts have long claimed: that, at least for faces, our environments play a substantial role in shaping our preferences and particular notions of attractiveness [21]. Our results further establish that the important environments are individual specific; that is, they are not consistent across family members. Moreover, we demonstrate a developmental dissociation in the fundamental etiology of two core domains of social perception and face processing: whereas variations in face attractiveness judgments result primarily from variations in environments, variations in face identity judgments result primarily from variations in genes. Our results provide a window into understanding the developmental and biological origins of the social brain and those aspects of our genes and environments that make us each unique.

### EXPERIMENTAL PROCEDURES

**Participants**

To understand the genetic and environmental contributions to individual face preferences, we recruited 796 twin pairs through the Australian Twin Registry [47]. We classified twin zygosity through latent class analysis via a standard self-report questionnaire [48]. After exclusions (see Supplemental Information), our final sample comprised 547 MZ twin pairs (mean age = 45.2; 415 female) and 214 same-sex DZ twin pairs (mean age = 45.8; 160 female). We also analyzed data from a set of 680 singletons (mean age = 41; 445 female) who completed the same measures as our twin sample. As the correlations between dependent measures calculated using the combined twin and singleton samples and calculating average ratings using only the singleton sample were extremely high (r's > 0.99), we combined the two samples in order to maximize the precision of our estimates. We note that our results were the same when calculated using only the twins sample, only the singletons sample, or with both samples combined. The study was reviewed and approved by the Committee for the Use of Human Subjects at Harvard University and the Australian Twin Registry. All participants gave informed consent before taking part in the study.

**Behavioral Testing and Data Analysis**

All tests were administered through our website, http://testmybrain.org [49]. Participants were sent a link to the study and participated at a time of their choosing from their own personal computers. Participants were given feedback about how their ratings compared to the average person. We have found that this feedback-as-incentive model produces high quality data that are comparable to data collected in traditional lab settings, even for demanding tests of social perception [49]. For the measures included in this manuscript, average ratings between twin participants and a separate sample tested in the lab (n = 31) were highly correlated (r = 0.96), indicating comparability between unsupervised web versus lab-based assessments.

Individual preference scores were estimated based on the correlation between a participant's ratings and the average ratings for each stimulus (see Figures 1D and 1E), transformed to z-scores using Fisher's r-to-z transformation [14]. We then regressed out Z transformed response consistency scores to produce a general face preference score that was not related to differences in intra-individual variability (see Figure S1A and Supplemental Information). Finally, we regressed out IP scores for objects and scenes (see Figures S1B and S1C). These face-specific scores were then standardized (Z scored). Summary statistics for face, object, and scene preference measures are in Table S1. We also conducted internal reliability and test-retest reliability analysis to estimate the degree to which variations in face IP scores reflect variations in stable, phenotypic characteristics (internal reliability = 0.88; test-retest reliability = 0.75; see Supplemental Information). We used standard maximum-likelihood-based behavioral genetic model-fitting procedures, implemented via OpenMx, to estimate genetic and environmental contributions to face IP (see Supplemental Information).

### SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, one figure, and two tables and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2015.08.048.

### AUTHOR CONTRIBUTIONS


### Table 1. Reliability, Twin Correlations, and Variance Component Estimates for Face IP

<table>
<thead>
<tr>
<th>Reliability</th>
<th>Internal (split-half)</th>
<th>0.88</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test-retest</td>
<td>0.75</td>
<td></td>
</tr>
<tr>
<td>Twin Correlations (95% CI)</td>
<td>MZ 0.22 (0.14–0.29)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DZ 0.09 (−0.06–0.24)</td>
<td></td>
</tr>
<tr>
<td>Model Fit: −2LL; AIC (p Value)</td>
<td>ACE 12,711.44; 9,695.44</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AE 12,711.44; 9,693.44 (p = 1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CE 12,713.54; 9,695.54 (p = 0.15)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>E 12,741.39; 9,721.39 (p &lt; 0.001)</td>
<td></td>
</tr>
<tr>
<td>Full Model: ACE Estimates (95% CI)</td>
<td>A 0.22 (0.0–0.29)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C 0 (0–0.24)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>E 0.78 (0.71–0.86)</td>
<td></td>
</tr>
<tr>
<td>Best-Fit Model: AE Estimates (95% CI)</td>
<td>A 0.22 (0.14–0.29)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>E 0.78 (0.71–0.86)</td>
<td></td>
</tr>
</tbody>
</table>

Q12

Q13

Q6

Q7

Q8

ACKNOWLEDGMENTS

This research was facilitated through access to the Australian Twin Registry, a national resource supported by a Centre of Research Excellence Grant from the National Health and Medical Research Council. This work was supported by US National Institute of Mental Health grants to L.G. (T32MH016259 and F32MH102971) and to J.W.S. (K24MH094614), the Australian Research Council Centre of Excellence in Cognition and Its Disorders (CE110001021), an ARC Professorship Fellowship to G.R. (DP0877379), an ARC Discovery Outstanding Researcher Award to G.R. (DP13012300), and a Brachman Hoffman Fellowship and Brachman Hoffman Small Grant to J.B.W.

Received: May 12, 2015
Revised: July 14, 2015
Accepted: August 20, 2015
Published: October 1, 2015

REFERENCES


