Which Facial Features Are Central in Impression Formation?

Bastian Jaeger[1,2] and **Alex L. Jones[3]**

Abstract

Which facial characteristics do people rely on when forming personality impressions? Previous research has uncovered an array of facial features that influence people's impressions. Even though some (classes of) features, such as resemblances to emotional expressions or facial width-to-height ratio (fWHR), play a central role in theories of social perception, their relative importance in impression formation remains unclear. Here, we model faces along a wide range of theoretically important dimensions and use machine learning techniques to test how well 28 features predict impressions of trustworthiness and dominance in a diverse set of 597 faces. In line with overgeneralization theory, emotion resemblances were most predictive of both traits. Other features that have received a lot of attention in the literature, such as fWHR, were relatively uninformative. Our results highlight the importance of modeling faces along a wide range of dimensions to elucidate their relative importance in impression formation.

Keywords

social perception, personality impressions, overgeneralization theory, emotional expressions, facial width-to-height ratio

People spontaneously judge others' personality based on their facial appearance (Todorov et al., 2015). For example, impressions of trustworthiness and dominance—which represent fundamental dimensions on which faces are evaluated (B. C. Jones et al., 2021; Oosterhof & Todorov, 2008)—can be formed within a few hundred milliseconds (Willis & Todorov, 2006). These impressions can be extremely consequential as they guide important decisions such as voting, criminal sentencing, and personnel selection (Olivola et al., 2014). Which facial characteristics do people rely on when forming personality impressions from faces? Previous investigations have produced a long list of facial features that are correlated with personality impressions (Hehman et al., 2019; Todorov et al., 2015). These findings provide the foundation for broader theories of social perception, which aim to explain the accuracy and functional significance of personality impressions (e.g., Carré et al., 2009; Todorov et al., 2008; Zebrowitz, 2017).

One class of characteristics that has received a lot of attention is the structural resemblance between a person's facial features and emotional expressions. Resting faces that merely resemble an expression of happiness (e.g., slightly upturned corners of the mouth) are perceived as trustworthy, whereas resting faces that resemble an expression of anger (e.g., lowered eyebrows) are perceived as dominant (Adams et al., 2012; Said et al., 2009). These findings are highlighted by overgeneralization theory, which aims to explain the functional significance of personality impressions and the cognitive mechanisms underlying impression formation (Todorov et al., 2008; Zebrowitz, 2012, 2017). Specifically, the emotion

overgeneralization hypothesis posits that, due to their relevance for social interactions, people are particularly attuned to detecting emotional expressions from faces. This sensitivity causes people to perceive emotional expressions (and associated traits) in faces that *merely resemble* an emotional expression. Thus, overgeneralization theory posits that perceived resemblances to emotional expressions are an important input in impression formation and, more generally, that personality impressions are caused by an oversensitive emotion detection system.

Other theories have focused on different features in impressions formation. For example, facial width-to-height ratio (fWHR) influences impressions of trustworthiness and dominance (Geniole et al., 2014; Ormiston et al., 2017; Stirrat & Perrett, 2010). Moreover, some have argued that fWHR is an indicator of various behavioral tendencies, such as aggression, because biological factors (e.g., testosterone) influence both facial morphology and behavioral dispositions (Carré et al., 2009; for counterarguments, see Kosinski, 2017; Wang et al., 2019). Thus, this perspective posits that fWHR is an important

[1] Vrije Universiteit Amsterdam, The Netherlands
[2] Tilburg University, The Netherlands
[3] Swansea University, United Kingdom

Corresponding Author:
Bastian Jaeger, Department of Experimental and Applied Psychology, Vrije Universiteit Amsterdam, 1081 BT, Amsterdam, The Netherlands.
Email: bxjaeger@gmail.com

input in impression formation and, more generally, that personality impressions can be accurate because facial appearance and behavioral dispositions have a common underlying cause.

Emotional expressions and fWHR occupy central roles in models of social perception, but they are only two examples from a long list of characteristics that are thought to form the basis of impression formation (for recent reviews, see Hehman et al., 2019; Todorov et al., 2015; Zebrowitz, 2017). Other overgeneralization hypotheses have been proposed, which highlight the role of babyfacedness (resemblances to neotonous facial features), attractiveness (resemblances to people with genetic anomalies or diseases), and familiarity (i.e., resemblances to familiar others) in impression formation (Zebrowitz, 2004, 2017; Zebrowitz & Collins, 1997). Moreover, studies have linked personality impressions to various other facial features such as cultural typicality (Sofer et al., 2015), race typicality (Blair et al., 2002), and skin texture (Jaeger et al., 2018).

The Importance of Different Facial Characteristics

Even though some facial characteristics occupy a more central role in theories of social perception, evidence on their *relative* importance in impressions formation remains sparse. To examine the importance of different features, previous studies have predominantly examined how one or a few features affect personality judgments.[1] This approach has two important limitations.

First, many facial characteristics are correlated, making it difficult to isolate their unique effects (A. L. Jones, 2019). For example, resemblances to emotional expressions are correlated with a variety of other features such as fWHR (Deska et al., 2018), babyfacedness (Sacco & Hugenberg, 2009), and race (Bijlstra et al., 2014). Even when one dimension of interest is manipulated, perceptions of other dimensions will also change. Manipulations of facial features that increase the perceived resemblance to a smile also change perceptions of babyfacedness and other dimensions. This raises the question whether personality impressions are indeed best explained by emotion resemblances or rather by other classes of features that are related to emotion resemblances.

Second, even when a single feature is manipulated while holding other correlated ones constant, it remains unclear how well this feature predicts impressions in real life when people are exposed to variation in facial features across many dimensions. It is possible that certain facial features are significantly related to personality impressions in highly controlled settings, but they might be poor predictors under more realistic conditions. For example, fWHR may be related to personality impressions when targets' gender, race, and approximate age are kept constant (as is often the case in social perception studies), but fWHR might not be an important cue when faces vary along many dimensions that are relevant for personality judgments. This limitation is exacerbated in studies using a two-alternative forced-choice design (Ormiston et al., 2017; Stirrat & Perrett, 2010). In this common experimental design,

a face is manipulated to score high or low on one dimension, and the two face versions are displayed side by side (e.g., high vs. low fWHR). Participants then choose the face that they perceive as scoring higher on the relevant trait. As this approach highlights even subtle differences in facial features, it can produce effects that would not be observed with more naturalistic designs (DeBruine, 2020; A. L. Jones & Jaeger, 2019).

To address these limitations, some studies have used data-driven approaches, in which a large number of low-level facial characteristics (e.g., distances between different points in the face) are used to predict personality impressions (McCurrie et al., 2017; Oosterhof & Todorov, 2008; Song et al., 2017; Vernon et al., 2014). These techniques have proven very useful, for example, for visualizing prototypical configurations of faces. However, because of their data-driven nature, it is often unclear to what extent the results support theoretical predictions about the importance of different facial characteristics. For example, data-driven methods can be used to mathematically describe and visualize what a prototypically (un)trustworthy face looks like (Dotsch & Todorov, 2012; Oosterhof & Todorov, 2008). Ratings of these prototypes might reveal that a trustworthy face scores higher on perceived femininity, babyfacedness, resemblance to a happy expression, and many other dimensions. Yet, this approach provides limited insights into the *relative* importance of different psychological variables in impression formation.

Recent evidence also supports the predictive power of theory-driven variables. When comparing the predictive power of data-driven and theory-driven models for facial attractiveness, Holzleitner and colleagues (2019) found that the performance of a complex data-driven model was matched by using five theory-driven predictors at the same time, even though in isolation, these theory-driven predictors performed poorly. This speaks to the importance of identifying and testing theoretically important predictors at the same time, rather than in isolation, in order to build parsimonious and interpretable models of social perception.

The Current Study

In sum, previous approaches provide limited insights into which facial characteristics are central in impression formation. The current study was designed to address these limitations. We extend previous work in three crucial ways.

First, the majority of prior studies only examined one feature or one class of features in isolation (e.g., Sofer et al., 2017; Stirrat & Perrett, 2010). Here, we examine and compare the relative importance of a wide range of features that are commonly studied in the literature. We test seven classes of predictors. We test the four characteristics proposed by Zebrowitz' (2012, 2017) work on overgeneralization theory: resemblances to emotional expressions (e.g., resemblance to a happy or angry expression), attractiveness, babyfacedness, and familiarity. We also test the importance of fWHR, which is another feature that has been hypothesized to form the basis of impressions (Ormiston et al., 2017; Stirrat & Perrett,

2010). Next to these theory-driven predictors, we also examine the role of a large set of demographic characteristics (e.g., gender and age) and morphological characteristics (e.g., eye size, face length, cheekbone prominence).

Second, the majority of prior work has focused on the *explanatory* power of different facial features, testing how much variance in impressions is explained by different variables. However, this might overestimate the actual importance of specific characteristics due to overfitting (Yarkoni & Westfall, 2017). In the present study, we rely on procedures from machine learning to address this issue. We use nested cross-validation and to compare the *predictive* power of different facial features (for similar applications of these methods, see Holzleitner et al., 2019; A. L. Jones & Jaeger, 2019).

Third, many prior studies were based on relatively small samples of stimuli (e.g., 50 or fewer; Carré et al., 2009; Stirrat & Perrett, 2012), which limits the generalizability of results. We therefore examine the predictors of personality impressions in a large and demographically diverse set of faces ($n = 597$). Our approach serves as a critical test of how well different characteristics—which have been theorized to be central for impression formation—predict personality impressions when faces vary along a wide variety of different dimensions.

Method

All data and analysis scripts are available at the Open Science Framework (https://osf.io/8rj7e/). We report how our sample size was determined, all data exclusions, and all measures.

Stimuli

We analyzed all 597 face images from the Chicago Face Database (Ma et al., 2015). All individuals wore a gray shirt, displayed a neutral facial expression, and were photographed from a fixed distance against a uniform background. The database provides several advantages for the purpose of the current study. First, the database contains photographs of a large and diverse set of individuals who vary on gender (51.42% female), age ($M = 28.86$, $SD = 6.30$, $min = 16.94$, and $max = 56.38$), and race (33.00% Black, 30.65% White, 18.26% Asian, and 18.09% Latino). Thus, the image set represents a wide range of facial characteristics that people are exposed to in real life.

Variables

The database contains a large number of objectively measured and subjectively rated characteristics for each target. Our aim was to predict perceptions of trustworthiness and dominance with various characteristics. We examined the predictive power of 28 facial features, which we grouped into seven classes of predictors. The first four classes represent the four overgeneralization hypotheses proposed by Zebrowitz (2012, 2017).

Emotion resemblances included six variables representing the perceived resemblance of facial features to six emotional expressions (anger, disgust, fear, happiness, sadness, and surprise). *Attractiveness* included one variable representing the perceived attractiveness of targets. *Babyfacedness* included one variable representing the perceived babyfacedness of targets. *Familiarity* included one variable representing the perceived unusualness of targets (i.e., how much the person would stand out in a crowd). *FWHR* included one variable representing the fWHR of targets. *Demographic characteristics* included four variables: gender (coded 0 for male and 1 for female), race (Asian, Black, Latino, or White, with White coded as the reference category), and age. We also included a quadratic effect for age. *Morphological characteristics* included 14 variables that were selected based on a review of the social perception literature (Ma et al., 2015): face length, face width at the cheeks, face width at the mouth, face shape (face width at the cheeks divided by face length), heartshapeness (face width at the cheeks divided by face width at the mouth), nose shape (nose width divided by nose length), lip fullness (distance between the top and bottom edge of lips divided by face length), eye shape (eye height divided by eye width), eye size (eye height divided by face length), upper head length (forehead length divided by face length), cheekbone height (distance from check to chin divided by face length), cheekbone prominence (difference between face width at cheekbones and face width at mouth divided by face length), face roundness (face width at mouth divided by face length), and median luminance of the face. Even though it is not a morphological feature, we included luminance in this group of variables, as it constitutes another objectively measured, low-level stimulus property that has been linked to personality impressions (Dotsch & Todorov, 2012; Todorov et al., 2015).

Data on gender and race were directly provided by the photographed targets, and morphological features were measured in Adobe Photoshop (Ma et al., 2015). To collect data on all other variables, Ma and colleagues (2015) presented the images to a large and demographically diverse sample of 1,087 raters ($M_{age} = 26.75$ and $SD_{age} = 10.54$; 47.47% White, 10.76% Asian, 6.81% Black, 6.62% biracial or multiracial, 5.24% Latino, 1.66% other, and 21.44% did not report; and 50.78% female, 28.33% male, and 20.88% did not report). Participants viewed the images and rated them on the dimensions of interest on a 7-point scale (ranging from, e.g., *not trustworthy at all* to *extremely trustworthy*). Participants rated a subset of 10 images (in order to reduce fatigue) on all dimensions. On average, each face image was rated by 44 independent raters ($min = 21$ raters and $max = 131$ raters). Simulation studies indicate that this number of raters is sufficient to obtain stable average ratings (Hehman et al., 2018), and the ratings showed high internal consistency (ranging from $\alpha = .896$ to $\alpha = .999$ across the dimensions; Ma et al., 2015).[2] Ratings were averaged across all raters to create a score for each face on each dimension. For example, trustworthiness ratings were averaged to create a measure of each face's perceived trustworthiness. The same steps were followed for perceptions of dominance and all other subjectively rated

characteristics. A detailed description of the variables and how they were measured is provided by Ma and colleagues (2015).

Analytic Strategy

All continuous predictors (except age) were *z*-standardized prior to analysis. We used techniques from machine learning to estimate the predictive power of different (classes of) facial characteristics. For each model, we compute the root-mean-square error (RMSE), which represents the square root of the mean squared differences between predicted and observed values. In contrast to other statistics, such as R^2, RMSE has the advantage that it is not inflated by the number of predictors. Lower RMSE values indicate better predictive accuracy. We also computed the adjusted R^2 for each model. Applying a penalty to the R^2 metric in line with the number of predictors in a model prevents, for example, that the morphology model outperforms the other models simply because it includes more predictors. We rely on cross-validation—using the caret package (Kuhn, 2008) in R (R Core Team, 2021)—to avoid the problem of overfitting, in which a model is optimized to fit a particular data set to such an extent that it does poorly in predicting novel data (Yarkoni & Westfall, 2017). In this procedure, the data are split into a training set, which is used to estimate the model, and a test set, which is used to test the predictive accuracy of the model. This procedure is then repeated with many different, random splits of the data. The models' overall predictive accuracy is assessed by averaging the observed accuracy values (e.g., RMSEs) for each split. This procedure prevents overfitting and represents a true test of the models' *predictive* (rather than explanatory) power as the models' performance is tested with new data.

Next to comparing different classes of facial characteristics, we also compared their unique predictive power by simultaneously entering all 28 characteristics into one regression model. Given that there were many substantial correlations between cues (see Figure S1 in the Supplemental Materials), ordinary linear models may result in overfitted and highly variable estimates of the true importance of the parameters. To prevent this, we relied on Elastic Net regression (Hastie et al., 2009). Elastic Nets are linear models that simultaneously (a) shrink predictors to reduce overfitting through regularization and (b) perform variable selection by setting the coefficients of uninformative parameters to zero. Thus, this approach is ideally suited to examine the relative importance of different facial characteristics in predicting personality impressions.

Results

Model Comparisons

First, we compared the predictive accuracy of different classes of facial characteristics in predicting perceptions of trustworthiness and dominance. We estimated cross-validated linear regression models (10-fold cross-validation with 100 repeats). Trustworthiness ratings and dominance ratings were regressed on seven classes of predictors (in separate

models), representing emotion resemblances, attractiveness, babyfacedness, familiarity, fWHR, demographic characteristics, and morphological characteristics.

For perceptions of trustworthiness (see Figure 1, left panel), the emotions model showed the best predictive accuracy ($M_{RMSE} = 0.285$, $SD_{RMSE} = 0.025$), followed by the attractiveness model ($M_{RMSE} = 0.331$, $SD_{RMSE} = 0.026$), the demographics model ($M_{RMSE} = 0.388$, $SD_{RMSE} = 0.030$), the babyfacedness model ($M_{RMSE} = 0.395$, $SD_{RMSE} = 0.028$), the familiarity model ($M_{RMSE} = 0.401$, $SD_{RMSE} = 0.027$), the morphology model ($M_{RMSE} = 0.407$, $SD_{RMSE} = 0.0286$), and the fWHR model ($M_{RMSE} = 0.410$, $SD_{RMSE} = 0.027$). The same pattern was found when comparing how much variance was explained by the seven models (see Figure 2, left panel). The emotions model explained most variance ($M_R{}^2 = 0.527$, $SD_R{}^2 = 0.078$), followed by the attractiveness model ($M_R{}^2 = 0.365$, $SD_R{}^2 = 0.089$), the demographics model ($M_R{}^2 = 0.122$, $SD_R{}^2 = 0.076$), the babyfacedness model ($M_R{}^2 = 0.100$, $SD_R{}^2 = 0.068$), the familiarity model ($M_R{}^2 = 0.071$, $SD_R{}^2 = 0.054$), the fWHR model ($M_R{}^2 = 0.028$, $SD_R{}^2 = 0.033$), and the morphology model ($M_R{}^2 = 0.028$, $SD_R{}^2 = 0.046$).

For perceptions of dominance (see Figure 1, right panel), the emotions model showed the best predictive accuracy ($M_{RMSE} = 0.515$, $SD_{RMSE} = 0.528$), followed by the demographics model ($M_{RMSE} = 0.535$, $SD_{RMSE} = 0.044$), the morphology model ($M_{RMSE} = 0.574$, $SD_{RMSE} = 0.047$), the babyfacedness model ($M_{RMSE} = 0.623$, $SD_{RMSE} = 0.044$), the familiarity model ($M_{RMSE} = 0.656$, $SD_{RMSE} = 0.046$), the attractiveness model ($M_{RMSE} = 0.666$, $SD_{RMSE} = 0.048$), and the fWHR model ($M_{RMSE} = 0.671$, $SD_{RMSE} = 0.047$). The same pattern was found when comparing how much variance was explained by the seven models (see Figure 2, right panel). The emotions model explained most variance ($M_R{}^2 = 0.418$, $SD_R{}^2 = 0.092$), followed by the demographics model ($M_R{}^2 = 0.371$, $SD_R{}^2 = 0.090$), the morphology model ($M_R{}^2 = 0.265$, $SD_R{}^2 = 0.097$), the babyfacedness model ($M_R{}^2 = 0.154$, $SD_R{}^2 = 0.080$), the familiarity model ($M_{RMSE} = 0.068$, $SD_{RMSE} = 0.062$), the attractiveness model ($M_R{}^2 = 0.033$, $SD_R{}^2 = 0.038$), and the fWHR model ($M_R{}^2 = 0.019$, $SD_R{}^2 = 0.025$).

Elastic Net Regression

Next, we examined the influence of all 28 facial characteristics by simultaneously entering them into one regression model. We relied on Elastic Net regression (Hastie et al., 2009), which simultaneously (a) shrinks predictors to reduce overfitting through regularization and (b) performs variable selection by setting the coefficients of uninformative parameters to zero. The model has two hyperparameters that require tuning: α, which controls the degree of shrinkage, and λ, which determines how aggressively coefficients can be set to zero. First, we relied on nested cross-validation to identify which combination of α and λ maximized the predictive fit of our models. This involved splitting the data set into 10 folds.
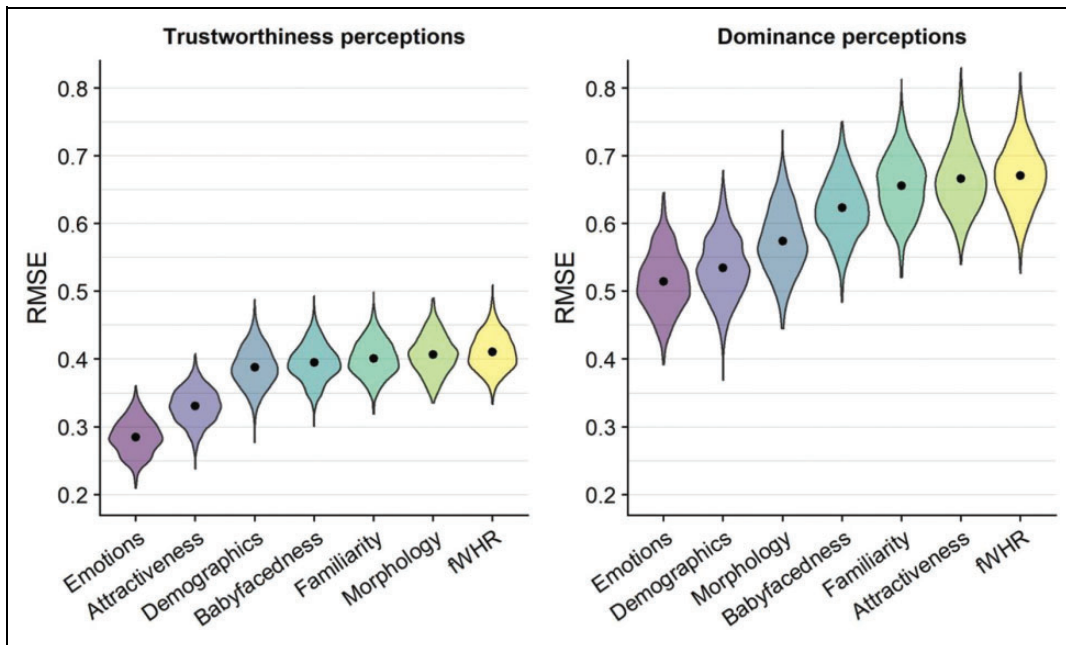
Figure 1. Predictive performance of the seven models in predicting perceptions of trustworthiness (left) and dominance (right). *Note.* Dots indicate the mean root-mean-square error (RMSE) from 10-fold cross-validation with 100 repeats.
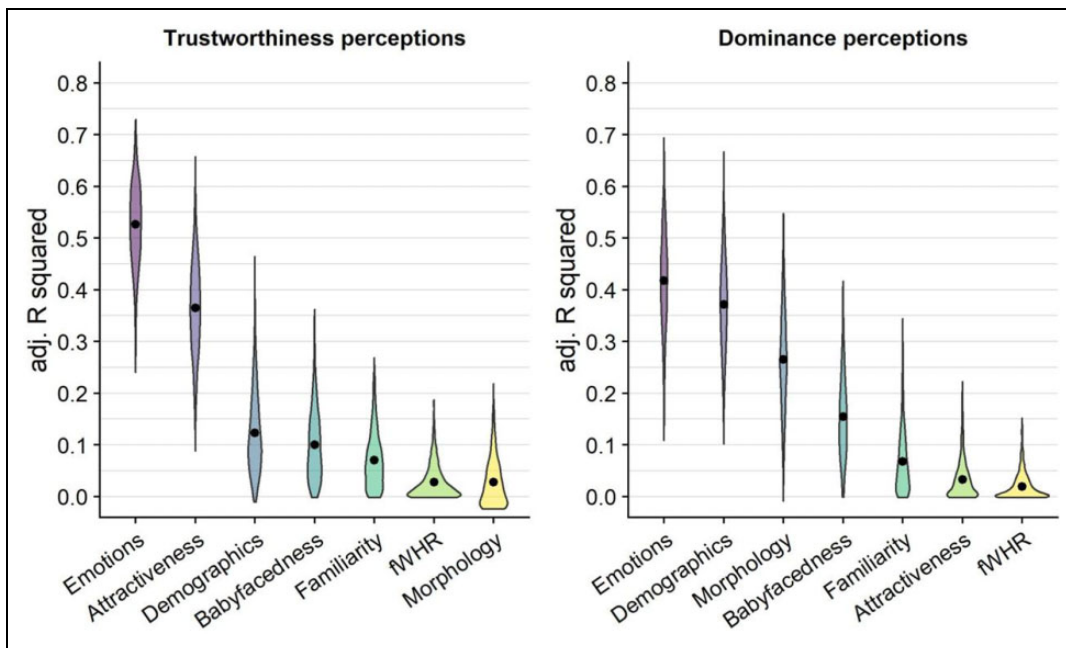

Figure 2. Performance of the seven models in predicting perceptions of trustworthiness (left) and dominance (right). *Note.* Dots indicate the mean adjusted R^2 from 10-fold cross-validation with 100 repeats.

For each split of the data, a further 10-fold grid search was carried out to derive the best hyperparameters before predicting the held out 10th fold. We repeated this process 100 times. This allowed us to identify at which levels of α and λ our models' predictive fit was maximized (i.e., RMSE was minimized). Next, we implemented models with our optimal α and λ values, again relying on 10-fold cross-validation with 100 repeats.

Our model predicted trustworthiness perceptions to within 0.23 points on a 7-point scale ($M_{RMSE} = 0.233$, $SD_{\mathrm{RMSE}} = 0.022$) and explained 67.04% of the variance ($M_{R^2} = 0.670$, $SD_{R^2} = 0.063$). We examined which facial features contributed most to the predictive accuracy of the model (see Figure 3). Resemblance to a happy facial expression ($\bar{\beta} = 0.154$) was the strongest predictor of trustworthiness perceptions. Attractiveness ($\bar{\beta} = 0.131$), being Asian ($\bar{\beta} = 0.110$),
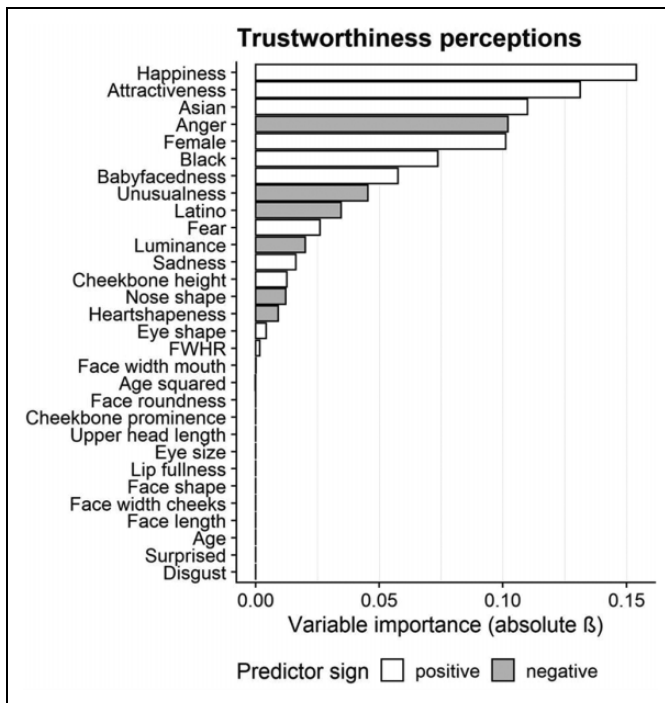
Figure 3. The relationships between facial characteristics and trustworthiness impressions. Coefficients were derived from Elastic Net models with nested cross-validation.
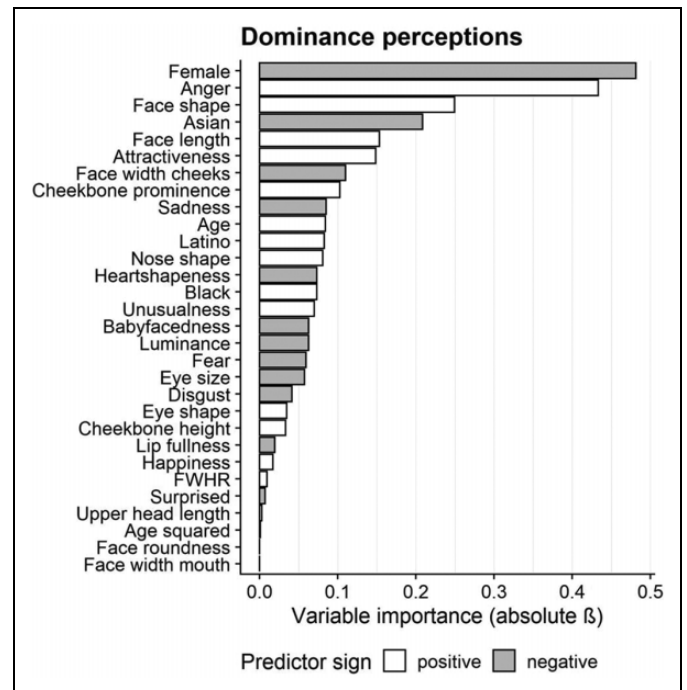

Figure 4. The relationships between facial characteristics and dominance impressions. Coefficients were derived from Elastic Net models with nested cross-validation.

resemblance to an angry facial expression ($\bar{\beta} = -0.102$), and being female ($\bar{\beta} = 0.101$) were also relatively informative predictors, whereas fWHR was relatively uninformative ($\bar{\beta} = 0.002$).

Our model predicted dominance perceptions to within 0.37 points on a 7-point scale ($M_{RMSE} = 0.370$, $SD_{RMSE} = 0.035$) and explained 68.50% of the variance ($M_R^2 = 0.685$, $SD_R^2 = 0.067$). We examined which facial features contributed most to the predictive accuracy of the model (see Figure 4). Being female ($\bar{\beta} = -0.482$) and resemblance to an angry facial expression ($\bar{\beta} = 0.434$) were by far the strongest predictors of dominance perceptions. fWHR was relatively uninformative ($\bar{\beta} = 0.002$).

General Discussion

Which facial characteristics do people rely on when forming impressions of others? Some facial features, such as resemblances to emotional expressions and fWHR, occupy a central role in theories of social perception (Todorov et al., 2008; Zebrowitz, 2017). However, it is not clear whether this focus is justified, as little is known about the relative importance of different characteristics. Faces can be modeled along many dimensions, and many facial features are correlated. Yet, prior work has mostly examined one feature or a few features in isolation. These approaches cannot provide strong evidence for the claim that people rely on certain facial features in impression formation, as it remains unclear whether people relied on the facial feature in question, or on other correlated ones. In

short, even though studies have identified a long list of facial features that are correlated with impressions, the question of which facial features are actually central in impression formation remains largely unaddressed. Here, we used methods from machine learning (i.e., cross-validation, regularization) to estimate and compare the extent to which a wide range of facial features predict trustworthiness and dominance impressions for a large and demographically diverse set of faces. We tested facial characteristics that have been theorized to be important in impression formation (resemblances to emotional expressions, attractiveness, babyfacedness, familiarity, and fWHR; Geniole et al., 2014; Stirrat & Perrett, 2010; Zebrowitz, 2017). We also tested a large set of other facial characteristics that have received less attention or are often held constant in social perception studies, even though they might be important in impression formation (e.g., gender, race, age, eye size, lip fullness).

When comparing different classes of facial features, we found that emotion resemblances were most predictive of both trustworthiness and dominance impressions, outperforming all other theory-driven models. When examining the importance of all 28 facial characteristics simultaneously, we found that perceptions of trustworthiness were best predicted by a face's resemblance to a happy expression. Emotionally neutral faces were perceived as more trustworthy when facial features resembled a facial expression of happiness. Perceptions of dominance were best predicted by targets' gender (with women being perceived as less dominant than men) and by resemblance to a facial expression of anger. Together, our results

support the notion that resemblances to emotional expressions are central for explaining how people form personality impressions from facial features. Our findings are in line with overgeneralization theory (and the emotion overgeneralization hypothesis in particular; Todorov et al., 2008; Zebrowitz, 2017), which posits that personality impressions of faces are driven by an oversensitive emotion detection system: Due to their social relevance, people even perceive emotions (and associated personality traits) in emotionally neutral faces that structurally resemble emotional expressions.

Support for the importance of other facial characteristics evoked by overgeneralization theory (i.e., attractiveness, babyfacedness, and familiarity; Zebrowitz, 2012, 2017) was mixed. Facial attractiveness was the second-most informative predictor of trustworthiness impressions, whereas babyfacedness and familiarity were less informative. None of the three characteristics were among the most informative predictors of dominance impressions.

We also found that demographic factors (i.e., gender, age, and race)—which have received less attention as predictors of personality impressions—were in some instances among the most important predictors of impressions. This highlights potential problems associated with keeping features like gender and race constant when studying social perception. Certain features may guide impression formation when demographic characteristics do not vary, but they may be uninformative when more diagnostic cues such as demographic characteristics do vary.

A wealth of studies has examined the influence of fWHR on personality judgments (e.g., Geniole et al., 2014; Ormiston et al., 2017; Stirrat & Perrett, 2010). Yet, the current results suggest that fWHR is not an informative predictor of trustworthiness or dominance impressions. When comparing the predictive fit of fWHR to the four characteristics that form the basis of overgeneralization theory, fWHR emerged as the weakest predictor. When modeled alongside all other facial features that we included in our analyses, fWHR was again among the least informative predictors. Similar results were obtained in additional analyses when examining impressions of male and female targets separately and when all other variables that included some measurement of face length or width were omitted from analyses (see Supplemental Materials). Together, these findings suggest that the importance of fWHR for impression formation may have been overstated in previous studies. Previously observed associations between fWHR and personality impressions may have been due to the fact that people rely on facial features that are correlated with fWHR, but not on fWHR per se.

Interestingly, all seven classes of predictors showed better predictive accuracy for trustworthiness perceptions than for dominance perceptions. It has been suggested that emotion resemblances are particularly important for trustworthiness impressions, whereas morphological characteristics, such as fWHR, are more important for dominance impressions (Hehman et al., 2015). The current results are not in line with this notion and suggest that emotion resemblances are the most

important determinant of both trustworthiness and dominance impressions. It should also be noted that even though emotion resemblances were the most important class of predictors, not all emotion resemblances were equally meaningful. Resemblance to a happy expression was the most important predictor of trustworthiness impressions, whereas resemblance to an angry expression was the most important predictor of dominance impressions.

Limitations and Future Directions

Despite the relatively good performance of some of our models, results also suggest that our list of relevant features was not exhaustive. Emotion resemblances explained 53% and 42% of the variance in trustworthiness and dominance perceptions. Even the optimized Elastic Net models explained around 68% of the variance, indicating there are other important factors contributing to personality impressions. Other facial features that might show independent contributions to personality impressions include skin texture (Jaeger et al., 2018; A. L. Jones et al., 2012) and perceived weight (Holzleitner et al., 2019). Examining the role of additional predictors will show how generalizable the present results are, as the relative importance of facial features ultimately depends on the specific set of features that is modeled. In order to conclusively establish that certain facial features are central in impression formation (and that observed associations are not due to other, unmeasured dimensions), faces need to be modeled along all potentially meaningful dimensions. From a practical perspective, achieving this goal may be unfeasible at best and impossible at worst. Still, future work should strive to test the relative importance of different features by comparing them against large sets of other features that have been shown to predict impressions.

Future studies could also investigate characteristics of the perceiver which explain a nontrivial amount of variance in impressions (Hehman et al., 2019). Moreover, while the current set of faces was relatively large and diverse in terms of gender, age, and race, we only examined U.S. individuals who were photographed in a controlled lab setting. Future studies could test whether the current findings replicate when using more naturalistic images of individuals from different nationalities (Sutherland et al., 2013).

Supplemental Material

The supplemental material is available in the online version of the article.

Notes

1. There are several noteworthy exceptions, with some studies examining multiple facial characteristics to test their unique effects (e.g., Berry & Zebrowitz McArthur, 1985; Blair et al., 2004; Zebrowitz & Apatow, 1983). For example, Zebrowitz and colleagues (2010) found that relationships between emotion resemblances and personality impressions still emerged when controlling for attractiveness an age.
2. Consistency estimates were based on the subset of 158 targets that were recruited for the first version of the database (as reported in Ma et al., 2015). The database was subsequently expanded multiple times. A detailed description of each update is available at https://chicagofaces.org/.

References

Adams, R. B., Nelson, A. J., Soto, J. A., Hess, U., & Kleck, R. E. (2012). Emotion in the neutral face: A mechanism for impression formation? *Cognition & Emotion*, *26*(3), 431–441. https://doi.org/10.1080/02699931.2012.666502

Berry, D. S., & Zebrowitz McArthur, L. (1985). Some components and consequences of a babyface. *Journal of Personality and Social Psychology*, *48*(2), 312–323. https://doi.org/10.1037/0022-3514.48.2.312

Bijlstra, G., Holland, R. W., Dotsch, R., Hugenberg, K., & Wigboldus, D. H. J. (2014). Stereotype associations and emotion recognition. *Personality and Social Psychology Bulletin*, *40*(5), 567–577. https://doi.org/10.1177/0146167213520458

Blair, I. V, Judd, C. M., & Fallman, J. L. (2004). The automaticity of race and Afrocentric facial features in social judgments. *Journal of Personality and Social Psychology*, *87*(6), 763–778. https://doi.org/10.1037/0022-3514.87.6.763

Blair, I. V., Judd, C. M., Sadler, M. S., & Jenkins, C. (2002). The role of Afrocentric features in person perception: Judging by features and categories. *Journal of Personality and Social Psychology*, *83*(1), 5–25. https://doi.org/10.1037/0022-3514.83.1.5

Carré, J. M., McCormick, C. M., & Mondloch, C. J. (2009). Facial structure is a reliable cue of aggressive behavior. *Psychological Science*, *20*(10), 1194–1198. https://doi.org/10.1111/j.1467-9280.2009.02423.x

DeBruine, L. M. (2020). *Composite images*. https://debruine.github.io/posts/composite-images/

Deska, J. C., Lloyd, E. P., & Hugenberg, K. (2018). The face of fear and anger: Facial width-to-height ratio biases recognition of angry and fearful expressions. *Emotion*, *18*(3), 453–464. https://doi.org/10.1037/emo0000328

Dotsch, R., & Todorov, A. (2012). Reverse correlating social face perception. *Social Psychological and Personality Science*, *3*(5), 562–571. https://doi.org/10.1177/1948550611430272

Geniole, S. N., Molnar, D. S., Carré, J. M., & McCormick, C. M. (2014). The facial width-to-height ratio shares stronger links with judgments of aggression than with judgments of trustworthiness. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(4), 1526–1541. https://doi.org/10.1037/a0036732

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning*. Springer.

Hehman, E., Flake, J. K., & Freeman, J. B. (2015). Static and dynamic facial cues differentially affect the consistency of social evaluations. *Personality and Social Psychology Bulletin*, *41*(8), 1123–1134. https://doi.org/10.1177/0146167215591495

Hehman, E., Stolier, R. M., Freeman, J. B., Flake, J. K., & Xie, S. Y. (2019). Toward a comprehensive model of face impressions: What we know, what we do not, and paths forward. *Social and Personality Psychology Compass*, *13*(2), 1–16. https://doi.org/10.1111/spc3.12431

Hehman, E., Xie, S. Y., Ofosu, E. K., & Nespoli, G. A. (2018). Assessing the point at which averages are stable: A tool illustrated in the context of person perception. https://psyarxiv.com/2n6jq/

Holzleitner, I. J., Lee, A. L., Hahn, A. C., Kandrik, M., Bovet, J., Renoult, J. P., Simmons, D., Garrod, O., Debruine, L. M., & Jones, B. C. (2019). Comparing theory-driven and data-driven attractiveness models using images of real women's faces. *Journal of Experimental Psychology: Human Perception and Performance*, *45*(12), 1589–1595. https://doi.org/10.1037/xhp0000685

Jaeger, B., Wagemans, F. M. A., Evans, A. M., & van Beest, I. (2018). Effects of facial skin smoothness and blemishes on trait impressions. *Perception*, *47*(6), 608–625. https://doi.org/10.1177/0301006618767258

Jones, A. L. (2019). Beyond average: Using face regression to study social perception. https://doi.org/10.31234/osf.io/dpmzq

Jones, A. L., & Jaeger, B. (2019). Biological bases of beauty revisited: The effect of symmetry, averageness, and sexual dimorphism on female facial attractiveness. *Symmetry*, *11*(2). https://doi.org/10.3390/sym11020279

Jones, A. L., Kramer, R. S. S., & Ward, R. (2012). Signals of personality and health: The contributions of facial shape, skin texture, and viewing angle. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(6), 1353–1361. https://doi.org/10.1037/a0027078

Jones, B. C., DeBruine, L. M., Flake, J. K., Aczel, B., Adamkovic, M., Alaei, R., Alper, S., Álvarez Solas, S., Andreychik, M. R., Ansari, D., Arnal, J. D., Babincák, P., Balas, B., Baník, G., Barzykowski, K., Baskin, E., Batres, C., Beaudry, J. L., Blake, K. R., . . . Chartier, C. R. (2021). To which world regions does the valence-dominance model of social perception apply? *Nature Human Behaviour*. https://psyarxiv.com/n26dy/

Kosinski, M. (2017). Facial width does not predict self-reported behavioral tendencies. *Psychological Science*, *28*(11), 1675–1682. https://doi.org/10.1177/0956797617716929

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, 28(5), 159–160. https://doi.org/10.1053/j.sodo.2009.03.002

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, 47(4), 1122–1135. https://doi.org/10.3758/s13428-014-0532-5

McCurrie, M., Beletti, F., Parzianello, L., Westendorp, A., Anthony, S., & Scheirer, W. J. (2017). Predicting first impressions with deep learning. In *Proceedings—Twelfth IEEE international conference on automatic face and gesture recognition, FG 2017—First international workshop on adaptive shot learning for gesture understanding and production, ASL4GUP 2017, biometrics in the wild, BWild 2017* (pp. 518–525). https://doi.org/10.1109/FG.2017.147

Olivola, C. Y., Funk, F., & Todorov, A. (2014). Social attributions from faces bias human choices. *Trends in Cognitive Sciences*, 18(11), 566–570. https://doi.org/10.1016/j.tics.2014.09.007

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, 105(32), 11087–11092. https://doi.org/10.1073/pnas.0805664105

Ormiston, M. E., Wong, E. M., & Haselhuhn, M. P. (2017). Facial-width-to-height ratio predicts perceptions of integrity in males. *Personality and Individual Differences*, 105, 40–42. https://doi.org/10.1016/j.paid.2016.09.017

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.r-project.org/

Sacco, D. F., & Hugenberg, K. (2009). The look of fear and anger: Facial maturity modulates recognition of fearful and angry expressions. *Emotion*, 9(1), 39–49. https://doi.org/10.1037/a0014081

Said, C. P., Sebe, N., & Todorov, A. (2009). Structural resemblance to emotional expressions predicts evaluation of emotionally neutral faces. *Emotion*, 9(2), 260–264. https://doi.org/10.1037/a0014681

Sofer, C., Dotsch, R., Oikawa, M., Oikawa, H., Wigboldus, D. H. J., & Todorov, A. (2017). For your local eyes only: Culture-specific face typicality influences perceptions of trustworthiness. *Perception*, 46(8), 914–928. https://doi.org/10.1177/0301006617691786

Sofer, C., Dotsch, R., Wigboldus, D. H. J., & Todorov, A. (2015). What is typical is good: The influence of face typicality on perceived trustworthiness. *Psychological Science*, 26(1), 39–47. https://doi.org/10.1177/0956797614554955

Song, A., Li, L., Atalla, C., & Cottrell, G. (2017). *Learning to see people like people*. http://arxiv.org/abs/1705.04282

Stirrat, M., & Perrett, D. I. (2010). Valid facial cues to cooperation and trust: Male facial width and trustworthiness. *Psychological Science*, 21(3), 349–354. https://doi.org/10.1177/0956797610362647

Stirrat, M., & Perrett, D. I. (2012). Face structure predicts cooperation: Men with wider faces are more generous to their in-group when out-group competition is salient. *Psychological Science*, 23(7), 718–722. https://doi.org/10.1177/0956797611435133

Sutherland, C. A. M., Oldmeadow, J. A., Santos, I. M., Towler, J., Michael Burt, D., & Young, A. W. (2013). Social inferences from faces: Ambient images generate a three-dimensional model. *Cognition*, 127(1), 105–118. https://doi.org/10.1016/j.cognition.2012.12.001

Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: Determinants, consequences, accuracy, and functional significance. *Annual Review of Psychology*, 66(1), 519–545. https://doi.org/10.1146/annurev-psych-113011-143831

Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences*, 12(12), 455–460. https://doi.org/10.1016/j.tics.2008.10.001

Vernon, R. J. W., Sutherland, C. A. M., Young, A. W., & Hartley, T. (2014). Modeling first impressions from highly variable facial images. *Proceedings of the National Academy of Sciences*, 111(32), E3353–E3361. https://doi.org/10.1073/pnas.1409860111

Wang, D., Nair, K., Kouchaki, M., Zajac, E. J., & Zhao, X. (2019). A case of evolutionary mismatch? Why facial width-to-height ratio may not predict behavioral tendencies. *Psychological Science*, 30(7), 1074–1081. https://doi.org/10.1177/0956797619849928

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science*, 17(7), 592–598. https://doi.org/10.1111/j.1467-9280.2006.01750.x

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, 12(6), 1100–1122. https://doi.org/10.1177/1745691617693393

Zebrowitz, L. A. (2004). The origin of first impressions. *Journal of Cultural and Evolutionary Psychology*, 2(1), 93–108. https://doi.org/10.1556/jcep.2.2004.1-2.6

Zebrowitz, L. A. (2012). Ecological and social approaches to face perception. In G. Rhodes, A. Calder, M. Johnson, & J. V. Haxby (Eds.), *Oxford handbook of face perception*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199559053.013.0003

Zebrowitz, L. A. (2017). First impressions from faces. *Current Directions in Psychological Science*, 26(3), 237–242. https://doi.org/10.1177/0963721416683996

Zebrowitz, L. A., & Apatow, K. (1983). Impressions of baby-faced adults. *Social Cognition*, 2(4), 315–342.

Zebrowitz, L. A., & Collins, M. A. (1997). Accurate social perception at zero acquaintance: The affordances of a Gibsonian approach. *Personality and Social Psychology Review*, 1(3), 204–223.

Zebrowitz, L. A., Kikuchi, M., & Fellous, J.-M. (2010). Facial resemblance to emotions: Group differences, impression effects, and race stereotypes. *Journal of Personality and Social Psychology*, 98(2), 175–189. https://doi.org/10.1037/a0017990

Author Biographies

Bastian Jaeger is an assistant professor at the Vrije Universiteit Amsterdam (Department of Experimental and Applied Psychology). His work focuses on social perception and decision making.

Alex L. Jones is a senior lecturer at Swansea University (Department of Psychology) in Wales. He studies face perception, first impressions, and statistical modeling.

Handling Editor: Peter Rentfrow