The functional basis of face evaluation

Nikolaas N. Oosterhof and Alexander Todorov*

Department of Psychology, Princeton University, Princeton, NJ 08540

People automatically evaluate faces on multiple trait dimensions, and these evaluations predict important social outcomes, ranging from electoral success to sentencing decisions. Based on behavioral studies and computer modeling, we develop a 2D model of face evaluation. First, using a principal components analysis of trait judgments of emotionally neutral faces, we identify two orthogonal dimensions, valence and dominance, that are sufficient to describe face evaluation and show that these dimensions can be approximated by judgments of trustworthiness and dominance. Second, using a data-driven statistical model for face representation, we build and validate models for representing face trustworthiness and face dominance. Third, using these models, we show that, whereas valence evaluation is more sensitive to features resembling expressions signaling whether the person should be avoided or approached, dominance evaluation is more sensitive to features signaling physical strength/weakness. Fourth, we show that important social judgments, such as threat, can be reproduced as a function of the two orthogonal dimensions of valence and dominance. The findings suggest that face evaluation involves an overgeneralization of adaptive mechanisms for inferring harmful intentions and the ability to cause harm and can account for rapid, yet not necessarily accurate, judgments from faces.

emotions | face perception | social cognition

The belief that the nature of the mind and personality could be inferred from facial appearance has persisted over the centuries. References to this belief can be dated back to ancient Greece, Rome, and China (1). In the 19th century, the pseudoscience of physiognomy reached its apogee. Cesare Lombroso, the founder of criminal anthropology, argued that "each type of crime is committed by men with particular physiognomic characteristics". For example, "thieves are notable for their expressive faces and manual dexterity, small wandering eyes that are often oblique in form, thick and close eyebrows, distorted or squashed noses, thin beards and hair, and sloping foreheads" (2). Lombroso provided his "scientific" testimony at several criminal trials.

Although modern science, if not folk psychology (3), has largely discarded such notions, trait evaluations from faces predict important social outcomes ranging from electoral success (4–6) to sentencing decisions (7, 8). Studies show that people rapidly evaluate faces on multiple trait dimensions such as trustworthiness and aggressiveness (9, 10). For example, trait judgments can be formed after as little as 38-ms exposure to an emotionally neutral face (10). Studies also show that the amygdala, a subcortical brain region critical for fear conditioning and consolidation of emotional memories (11), plays a key role in the assessment of face trustworthiness (12–15).

Why do mechanisms for rapid spontaneous face evaluation exist if they do not necessarily deliver accurate inferences? This apparent puzzle from an evolutionary point of view can be resolved by theories that posit that evaluation of emotionally neutral faces is constructed from facial cues that have evolutionary significance (16–18). Using a data-driven approach, the objectives of the current research were to (*i*) identify the underlying dimensions of face evaluation, (*ii*) introduce tools for formally modeling how faces vary on these dimensions, (*iii*) determine the facial features that give rise to judgments on these dimensions, and (*iv*) link the findings to a broader evolutionary context that can account for rapid yet not necessarily accurate judgments from faces.

Identifying the Underlying Dimensions of Face Evaluation. Although people evaluate faces on multiple trait dimensions, these evaluations are highly correlated with each other [supporting information (SI) Fig. S1]. To identify the underlying dimensions of face evaluation, we (*i*) identified traits that are spontaneously inferred from emotionally neutral faces, (*ii*) collected judgments on these trait dimensions, and (*iii*) submitted these judgments to a principal components analysis (PCA). At the first stage of the project, we asked 55 participants to generate unconstrained descriptions of faces (study 1). These descriptions were then classified into trait dimensions. Fourteen dimensions accounted for 68% of the >1,100 descriptions and were selected for subsequent analyses. Participants (total $n = 327$; studies 2.1 to 2.15) were then asked to judge the same neutral faces on these trait dimensions and dominance (Table S1). Dominance was included because of the central importance of this trait in models of interpersonal perception (19). For two of the traits, the interrater agreement was very low, and they were not included in the subsequent analyses. The judgments for the remaining traits were highly reliable (Cronbach's α >0.90; Table S2).

The first principal component (PC) accounted for 63.3% of the variance and the second PC accounted for 18.3% of the variance of the mean trait judgments[†]. All positive judgments (e.g., attractive, responsible) had positive loadings, and all negative judgments (e.g., aggressive) had negative loadings on the first PC (Table S3), suggesting that it can be interpreted as valence evaluation (20, 21). Judgments of dominance, aggressiveness, and confidence had the highest loading on the second PC, suggesting that it can be interpreted as dominance evaluation (19). This 2D structure of face evaluation is consistent with well established dimensional models of social perception (19, 22, 23). For example, Wiggins *et al.* (19, 23), starting with a large set of traits describing interpersonal relationships, have shown that interpersonal perception can be described by two orthogonal dimensions, affiliation and dominance, that are similar to the dimensions identified here.

Judgments of trustworthiness were closest in space to the first PC, and judgments of dominance were closest to the second PC (Fig. S2 and Table S3). To obtain a solution unbiased with respect to these two judgments, we submitted all trait judgments except trustworthiness and dominance to a second PCA. Whereas trustworthiness judgments were highly correlated with the first (0.92) but not with the second PC (−0.10), dominance judgments were highly correlated with the second (0.87) but not with the first PC (−0.20; Fig. S3). We also created PC scores weighted by the frequency of the use of the 11 traits in the unconstrained descriptions of faces (Table S1).

The correlation between trustworthiness judgments and the first (weighted) PC was not affected by the weighting procedure (0.92). The correlation between dominance judgments and the second (weighted) PC was reduced but remained highly significant (0.67, $P < 0.0001$).

To further test whether the frequency of trait use affects the PCA solution with respect to trustworthiness and dominance, we conducted a series of PCAs first using as input the five most frequently used traits and then entering additional traits according to their frequency. For all analyses, the correlation between trustworthiness judgments and the first PC was equal or >0.90 (Table S4). The correlation between dominance judgments and the second PC was 0.53 for the first five traits, increased to 0.77 for the first six traits, and reached a ceiling for the first nine traits (SI Text).

It is possible that the set of specific faces introduced biases in the estimation of the correlations between trait judgments and, ultimately, the PCs. However, the pattern of correlations between judgments of these faces was the same as the pattern of correlations between judgments of 300 computer-generated faces (SI Text and Table S5). Moreover, a PCA on a different set of trait judgments of the computer-generated faces also identified two PCs with trustworthiness and dominance judgments closest in space to these components (Table S6).

The findings suggest that the dimensions for face evaluation identified here are robust with respect to both selection of face stimuli and trait judgments. The findings also suggest that judgments of trustworthiness and dominance can be used as approximations of the underlying dimensions, valence and dominance, of evaluation of emotionally neutral faces.

Modeling of Face Trustworthiness and Face Dominance. Given these findings, we built models for representing how faces vary on trustworthiness and dominance. We used a data-driven statistical model based on 3D laser scans of faces. The shape of the faces was represented by the vertex positions of a polygonal model of fixed mesh topology (Fig. S4). These positions were subjected to a PCA to extract the components that account for most of the variance in face shape (ref. 24; SI Text). Each of the PCs represents a different holistic nonlocalized set of changes in all vertex positions. By construction, this model does not make *a priori* assumptions about the importance of specific facial parts (e.g., nose, eyebrows).

Although face texture is important for face perception, for reasons of simplicity and to avoid overfitting, we worked only with the 50 PCs that represent symmetric face shape. Using a measure of a facial attribute (e.g., face width) for a set of faces, one can construct a vector in the 50-dimensional face space (composed of the weights for each PC) that is optimal in changing this attribute. This change is reflected in linear changes in the vertex positions that define the face shape (Fig. S4). We worked with a simple linear model (SI Text). The feasibility of this linear approach has been demonstrated for modeling facial attributes such as gender, hooked vs. concave nose, and fullness of face (24). In the present studies, we used this approach to model face variations on social dimensions such as trustworthiness and dominance.

Using the face model, we randomly generated 300 emotionally neutral faces (Fig. S5) and asked participants to judge them on trustworthiness (study 3, $n = 29$) and dominance (study 4, $n = 25$). Consistent with the prior findings, the correlation between the mean trustworthiness and dominance judgments was low (-0.17; Table S5). We used the mean judgments to find vectors in the 50-dimensional face space whose direction is optimal in changing trustworthiness and dominance. Within the plane defined by the trustworthiness and dominance vectors, we rotated the dominance vector $-28°$ to make it orthogonal to the trustworthiness vector (Fig. 1A; SI Text). All behavioral studies reported below use faces that vary on the orthogonal dimensions of trustworthiness and dominance.

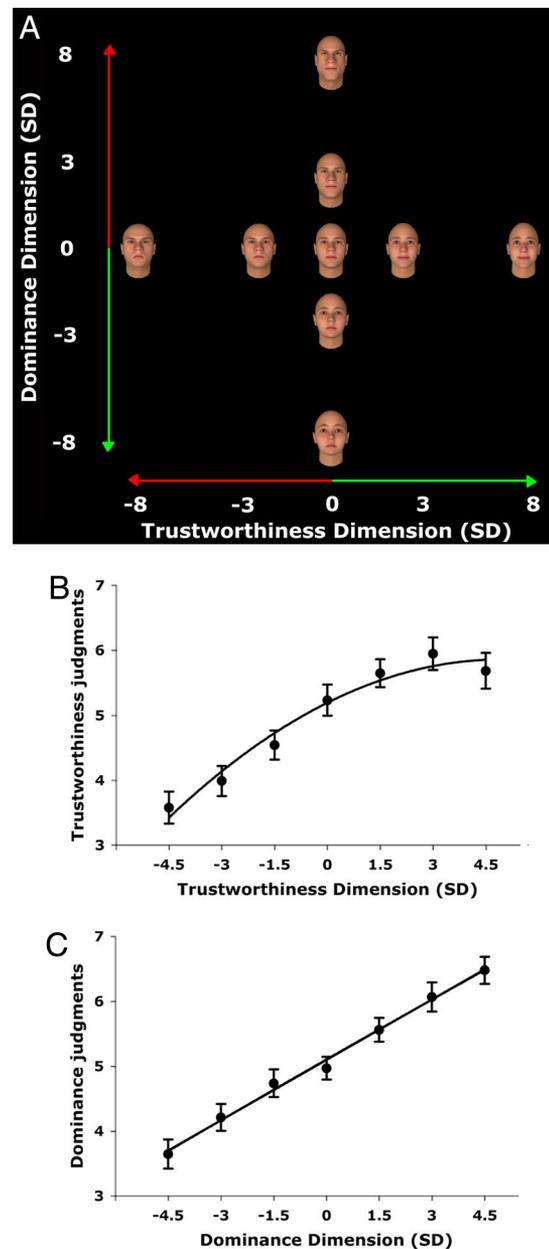To validate that the models successfully manipulate trust-

Fig. 1. A 2D model of face evaluation. (*A*) Examples of a face varying on the two orthogonal dimensions, trustworthiness and dominance. The face changes were implemented in a computer model based on trustworthiness (study 3) and dominance judgments (study 4) of 300 emotionally neutral faces. The extent of face exaggeration is presented in SD units. (*B*) Mean trustworthiness judgments of faces (study 5) generated by the trustworthiness model. (*C*) Mean dominance judgments of faces (study 6) generated by the dominance model. The judgments were made on 9-point scales. Error bars show standard error of the mean.

worthiness and dominance, we randomly generated new faces. For each face, we produced seven versions that varied on trustworthiness and seven versions that varied on dominance (-4.5, -3.0, -1.5, 0, 1.5, 3.0, and 4.5 SD on the dimensions) and asked participants to judge them on trustworthiness (study 5, $n = 17$) and dominance (study 6, $n = 17$). Trustworthiness judgments tracked the trustworthiness predicted by the model (Fig. 1B), $F(1,16) = 47.59$, $P < 0.001$ (Fisher's F test for the linear trend), although people were more sensitive to changes in trustworthiness at the low end of the spectrum than at the

Oosterhof and Todorov

high end, $F_{quadratic}(1,16) = 65.36$, $P < 0.001$[‡]. That is, although the physical distance between any two categories of faces was the same (1.5 SD), people were better at discriminating faces at the negative end of the trustworthiness dimension. Dominance judgments tracked the dominance predicted by the model in a linear fashion (Fig. 1C), $F_{linear}(1,16) = 99.31$, $P < 0.001$, $F_{quadratic} < 1$.

Revealing the Facial Cues Used for Face Evaluation. Judgments of emotionally neutral faces are based on subtle variations in the features of these faces. Using the computer models, it is possible to reveal the underlying variations that account for specific trait judgments by exaggerating the features that contribute to these judgments. As described above, the face model is holistic and not constrained by *a priori* assumptions about the importance of specific facial features. However, we can discover the important features *a posteriori* by using a linear extrapolation of the shape changes on the dimensions constructed to be optimal in representing specific face variations. In other words, in a process akin to creating a caricature of a face on the dimension of interest, by exaggerating the features specific to an evaluative dimension, we can identify the type of facial information used for this evaluation. For example, as shown in Fig. 1A, moving from the negative (−8 SD) to the positive (8 SD) extreme of the trustworthiness dimension, faces seemed to change from expressing anger to expressing happiness (Movie S1). Moving from the negative to the positive extreme of the dominance dimension, faces seemed to change from feminine and baby-faced to masculine and mature-faced (Movie S2).

We designed six studies to test whether the dimensions of trustworthiness and dominance are sensitive to different types of facial information. We randomly generated faces and produced four extreme versions of each face varying on trustworthiness and four versions varying on dominance (−8, −4, 4, 8 SD). Nineteen participants (study 7) were asked to categorize the faces as neutral or as expressing one of the six basic emotions. The faces at the center of the dimensions (0 SD) were classified as neutral (Fig. 2A). For the trustworthiness dimension, as the facial features become more exaggerated (−4 and 4 SD), the neutral categorization decreased, and fell below chance for the most exaggerated faces (−8 and 8 SD), $F_{quadratic}(1,18) = 164.82$, $P < 0.001$. As the facial features become extremely exaggerated in the negative direction (−8 SD), the faces were classified as angry, $F_{linear}(1,18) = 139.88$, $P < 0.001$, and as the features become exaggerated in the positive direction (4 and 8 SD), the faces were classified as happy, $F_{linear}(1,18) = 570.69$, $P < 0.001$. The only categorization responses that were significantly higher than chance (Table S7) were for neutral (−4, 0 SD), angry (−8 SD), and happy (4, 8 SD). This finding was replicated for a trustworthiness model based on judgments of another set of 200 faces (*SI Text* and Fig. S6).

In contrast to the findings for the trustworthiness dimension, dominance evaluation was weakly related to facial features resembling emotional expressions (Fig. 2A). The only categorization responses that were significantly higher than chance (Table S7) were for neutral (−4, 0, 4 SD), $F_{quadratic}(1,18) = 185.82$, $P < 0.001$, and fearful (−8 SD). Extremely submissive faces were classified as fearful. There was also a tendency to classify extremely dominant faces (8 SD) as angry, consistent
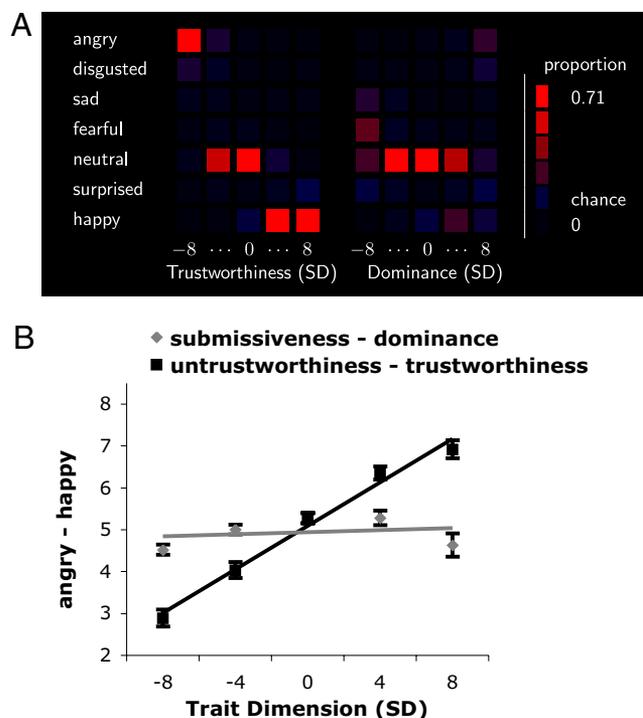

Fig. 2. Sensitivity of trustworthiness and dominance dimensions to cues resembling emotional expressions. (A) Intensity color plot showing face categorization as neutral or as expressing one of the basic emotions as a function of their trustworthiness and dominance (study 7). (B) Mean judgments of expressions of anger and happiness (study 8). The judgments were made on a 9-point scale, ranging from 1 (angry) to 5 (neutral) to 9 (happy). Error bars show standard error of the mean. The line represents the best linear fit. The x-axis represents the face exaggeration in SD units.

with prior findings (25–27), but this categorization was not significantly different from chance[§].

To directly compare the sensitivity of the valence and dominance dimensions to features resembling happy and angry expressions, nineteen participants (study 8) were asked to judge the faces on a 9-point scale, ranging from 1 (angry) to 9 (happy). Whereas face variation on the trustworthiness dimension was strongly related to these judgments (Fig. 2B), $F_{linear}(1,18) = 123.10$, $P < 0.001$, face variation on the dominance dimension was not related to the judgments[§], $F_{linear} < 1$, and the slopes of the linear trends differed significantly, $F(1,18) = 53.93$, $P < 0.001$.

Emotional expressions often signal the behavioral intentions of the person displaying the emotion (28). For example, expressions of happiness and anger signal to the perceiver that the person can be approached or should be avoided, respectively, and there is evidence that angry faces trigger automatic avoidance responses (29, 30). Although the trustworthiness model was data-driven and based on judgments of emotionally neutral faces, emotions signaling approach/avoidance behavior naturally emerged from the model when the features of the faces were exaggerated (*SI Text* and Fig. S7).

[‡]There was a slight reversal in the judgments for the two most extreme positive categories (3 and 4.5 SD), possibly because of the exaggeration of the facial features in the most extreme category. The nonlinearity of judgments was replicated for a trustworthiness model based on judgments of another set of 200 faces (*SI Text* and Fig. S6).

[§]The finding that the dominance dimension was not sensitive to features resembling angry expressions seems inconsistent with prior studies finding associations between judgments of dominance and judgments of anger (25–27). However, as described above, we rotated the dominance dimension to make it orthogonal to the trustworthiness dimension. This rotation seems to remove variations in features resembling angry and happy expressions on the rotated dominance dimension. The empirically derived dominance dimension would have been sensitive to happiness/anger information, although to a smaller extent than the trustworthiness dimension. The empirical dominance dimension passes from the fourth to second quadrant of the orthogonal space (Fig. 1A), and its sensitivity to angry/happy information can be estimated from its projection on the trustworthiness dimension.
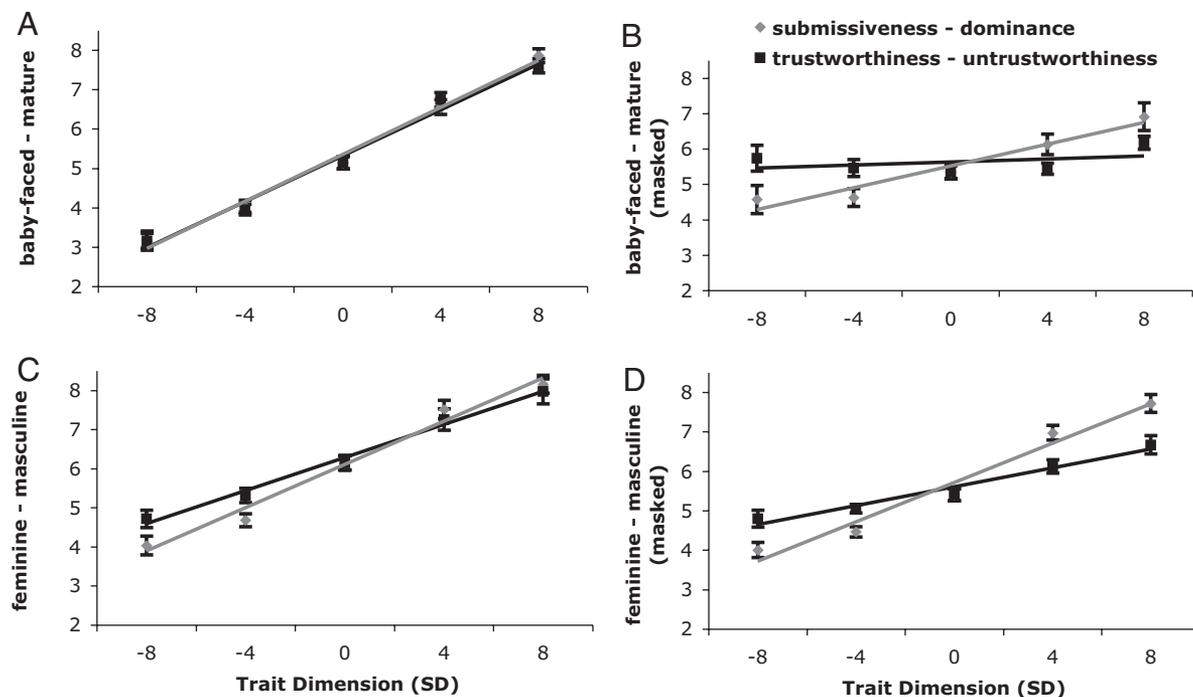
Fig. 3. Sensitivity of trustworthiness and dominance dimensions to cues related to physical strength. (*A*) Mean judgments of facial maturity (study 9) as a function of the trustworthiness and dominance of faces. The direction of the trustworthiness dimension was reversed to show that the slopes for the change from trustworthy to untrustworthy faces and the change from submissive to dominant faces were identical. (*B*) Mean judgments of facial maturity (study 10) of faces with masked internal features. The judgments were made on a 9-point scale, ranging from 1 (baby-faced) to 5 (neutral) to 9 (mature-faced). (*C*) Mean judgments of femininity/masculinity (study 11). (*D*) Mean judgments of femininity/masculinity (study 12) of faces with masked internal features. The judgments were made on a 9-point scale, ranging from 1 (feminine) to 5 (neutral) to 9 (masculine). The *x*-axis in the figures represents the extent of face exaggeration in SD units. Error bars show standard error of the mean. The lines represent the best linear fit.

The finding that the valence evaluation of faces is sensitive to features resembling emotional expressions is consistent with prior studies suggesting that trait judgments of emotionally neutral faces are an overgeneralization of perception of emotional expressions (26, 27, 31). For example, judgments of trustworthiness are negatively correlated with judgments of anger and positively correlated with judgments of happiness from emotionally neutral faces (32), as are judgments of affiliation, an attribute similar to trustworthiness (27).

Facial masculinity and maturity cues signal physical strength and the correspondent ability to cause harm. We tested whether the dominance dimension is more sensitive to these facial cues than the trustworthiness dimension. In two studies, we asked participants to rate the faces on a scale ranging from 1 (baby faced) to 9 (mature faced). Twenty-eight participants (study 9) judged the intact faces as in study 8, and 16 participants (study 10) judged the faces with their internal features masked (Fig. S8). The reason for the latter manipulation was that facial shape is one of the cues signaling facial maturity (33, 34), and the internal features of the face could trump the effects of facial shape on judgments. For example, one of the manipulations to increase baby-faced appearance is increasing the distance between the eyes and the eyebrows (34, 35), which also increases face trustworthiness (Fig. 1*A*). For the intact faces, both dimensions were related to facial maturity judgments (Fig. 3*A*), $F_{linear}(1,27) = 301.10$, $P < 0.001$, and the slopes were not significantly different from each other. However, for the faces with masked internal features, the linear trend was significant for faces that varied on the dominance dimension (Fig. 3*B*), $F_{linear}(1,15) = 16.40$, $P < 0.001$ but not for faces that varied on the trustworthiness dimension, $F_{linear} < 1$, $F(1,15) = 10.17$, $P < 0.006$, for the difference in slopes.

In the next two studies, participants were asked to rate the faces on a scale ranging from 1 (feminine) to 9 (masculine). Nineteen

participants (study 11) judged the intact faces, and 21 participants (study 12) judged the faces with their internal features masked. For the intact faces, both dimensions were related to femininity/masculinity judgments, $F_{linear}(1,18) = 142.31$, $P < 0.001$ (Fig. 3*C*), but the effect was stronger for the dominance dimension, $F(1,18) = 9.67$, $P < 0.006$, for the difference in slopes. Similarly, for the faces with masked internal features, both dimensions were related to these judgments, $F_{linear}(1, 20) = 137.77$, $P < 0.001$ (Fig. S1*D*), but the effect was stronger for the dominance dimension, $F(1,20) = 36.30$, $P < 0.001$, for the difference in slopes.

These findings suggest that, whereas the valence evaluation of faces is more sensitive to features resembling expressions signaling approach/avoidance behaviors, the dominance evaluation is more sensitive to features signaling physical strength.

Representing the Threat Value of a Face. One implication of the 2D model of face evaluation is that important social judgments can be represented within the plane defined by these dimensions. We demonstrate this with judgments of threat. Accurate assessments of threat are essential for survival (10), and threatening faces should be both untrustworthy, signaling that the person may have harmful intentions, and dominant, signaling that the person is capable of causing harm. In fact, threat judgments (study 13, $n = 21$) of the 300 faces used to create the trustworthiness and dominance models were highly correlated with both trustworthiness (-0.65, $P < 0.001$; Table S5) and dominance judgments (0.68, $P < 0.001$).

We built a threat vector based on the trustworthiness and dominance vectors by rotating the former 45° clockwise and the latter 45° counterclockwise in the plane defined by the two vectors (*SI Text* and Fig. S9). This threat vector was nearly identical to a vector based on threat judgments (study 13) of the 300 faces used to create the trustworthiness and dominance models: one SD change on the former vector corresponded to 0.98 SD change on the

latter vector. Judgments of threat (study 14, $n = 18$) tracked the threat predicted by the model (Fig. S10), F_{linear} (1, 17) = 319.30, $P < 0.001$, although people were more sensitive to changes at the high-threat end of the dimension than at the low-threat end, $F_{quadratic}$ (1, 17) = 42.93, $P < 0.001$.

Discussion

Based on behavioral studies and computer modeling of how faces vary on social dimensions, we developed a 2D model of face evaluation. Our findings suggest that faces are evaluated on two fundamental dimensions, valence and dominance. Whereas valence evaluation is an overgeneralization of perception of facial cues signaling whether to approach or avoid a person, dominance evaluation is an overgeneralization of perception of facial cues signaling the physical strength/weakness of the person. In other words, structural features that merely resemble cues that have adaptive significance (e.g., emotional expressions) give rise to trait inferences (36). Functionally, the valence and dominance related facial cues give rise to inferences about the person's intentions, harmful vs. harmless, and the person's ability to implement these intentions, respectively (cf. ref. 22). The dimensional model provides a unifying framework for the study of face evaluation and, in principle, can account for multiple trait inferences from emotionally neutral faces, as illustrated here with inferences of threat. In light of this model, the observed involvement of the amygdala in the automatic evaluation of face trustworthiness (13–15) most likely reflects general valence rather than specific trait evaluation. This hypothesis remains to be tested.

We focused on identifying the basic dimensions of face evaluation that permeate multiple social judgments from emotionally neutral faces. Although we used judgments of trustworthiness and dominance for the modeling of how faces vary on the basic dimensions, these judgments are only approximations of dimensions that encompass multiple judgments (Table S3). For example, attractiveness judgments were highly correlated with both trustworthiness judgments (0.79; Fig. S1A) and the valence component (0.81; Table S3 and Fig. S2). The very high correlation among social judgments from faces (Fig. S1) was one of the motivating reasons of the dimensional (PCA) approach.

However, it should be noted that the dimensional model is most applicable to implicit face evaluation where no specific evaluative context is provided (14). When a context makes a specific evaluative dimension relevant (e.g., competence), decisions would be most likely influenced by evaluations on this dimension. For example, in electoral decisions, voters believe that competence is the most important attribute for a politician and evaluations of competence but not trustworthiness predict electoral success (4). Similarly, in mating decisions, physical attractiveness could trump evaluations on other dimensions, including trustworthiness (37). In other words, in specific contexts, other dimensions of face evaluation may be critical for decisions.

The belief that personality can be read from a person's face has persisted over the centuries. This is not surprising, given the efficiency of trait judgments from faces and their subjectively compelling character. These compelling impressions are constructed from facial cues that have evolutionary significance (16–18). The accurate perceptions of emotional expressions and the dominance of conspecifics are critical for survival and successful social interaction (28, 38–40). In the absence of clear emotional cues broadcasting the intentions of the person, we argue that faces are evaluated in terms of their similarity to expressions of anger and happiness in an attempt to infer the person's intentions. People need to infer not only these intentions but also the ability of the person to implement these intentions. In the absence of other cues, faces are evaluated in terms of their facial maturity in an attempt to infer this ability.

The overgeneralization hypothesis (36) can account for rapid, efficient trait judgments from faces that do not necessarily deliver veridical inferences, a pattern that appears puzzling from an evolutionary point of view. Although some studies have found positive correlations between trait judgments from faces and measures of personality (41, 42), there have been other studies failing to find such correlations or finding negative correlations (3, 43, 44). Even when the correlations are positive, they are modest at best. For example, in the one study measuring the trustworthiness of actual behavior (41), judgments of honesty from faces accounted for 4% of the variance of behavior. The lack of a reliable relationship between such judgments and measures of personality is not surprising in light of the overgeneralization hypothesis. If these judgments are a measure of reading subtle emotional and dominance cues in neutral faces that are misattributed to stable personality dispositions, one should not expect that they are accurate.

Methods

Identifying Dimensions of Face Evaluation. *Participants.* Fifty-five undergraduate students participated in the first study in which we collected unconstrained face descriptions and 327 participated in the trait rating studies (2.1 to 2.15) for partial course credit or payment.
Face stimuli. We used 66 standardized faces (45) with direct gaze and neutral expressions. These were photographs of amateur actors and actresses between 20 and 30 years of age with no facial hair, earrings, eyeglasses, or visible make-up, all wearing gray T shirts.
Procedures: Free face descriptions. The 66 faces were randomly split into 11 groups of 6 faces, with the constraint that within each group, half of the faces were female faces. Based on this grouping, we created 11 questionnaire versions. At each page of the questionnaire, the face occupied slightly more than half of the page, and participants were asked to "write everything that comes to mind about this person." Six lines were provided below the face. The order of male and female faces was counterbalanced across participants. Participants were randomly assigned to one of the questionnaires (five participants per questionnaire).

In total, participants provided 1,134 person descriptions that spanned abstract trait attributes (e.g., "aggressive"), appearance statements ("unkempt"), emotional states ("sad"), preferences ("hates sports"), and miscellaneous associations. Two researchers independently classified the statements into broad categories. A third researcher resolved disagreements. Fourteen trait categories accounted for 68% of the statements (Table S1). The rest of the statements referred to physical qualities, social categories (age, sex, occupation), actions, attitudes and preferences, and emotional states (3% of the statements were unclassifiable).
Procedures: Trait-rating tasks. The 66 faces were rated on each of the traits identified in the study described above by a separate group of participants. In addition to the 14 traits, we added the trait of dominance because of its importance in models of interpersonal perception (19, 23). In all studies, participants were told that the study was about first impressions and were encouraged to rely on their "gut feeling." The faces were presented three times in three separate blocks (SI Text). Each face was presented at the center of the screen with a question above the photograph "How [trait term] is this person?" and a response scale below the photograph. The response scale ranged from 1 (Not at all [trait term]) to 9 (Extremely [trait term]). In all behavioral studies (2–14), each face was visible until the participant responded, the interstimulus interval (ISI) was 1,000 ms, and the order of faces was randomized for each participant.

Studies with Computer-Generated Faces. *Statistical model of face representation.* We used the Facegen Modeller program (http://facegen.com), Version 3.1. The face model of Facegen (24, 46) is based on a database of male and female human faces that were laser-scanned in 3D (SI Text).
Face stimuli. We generated 300 Caucasian faces using Facegen (Fig. S5). The faces were generated randomly with the following adjustments. Because a completely random face can be of any race, and we wanted to avoid judgments affected by stereotypes, we used Facegen's race controls to set the face to European. By default, the randomly generated faces are neutral. Facegen has separate controls for adding the basic emotional expressions: anger, disgust, fear, sadness, happiness, and surprise. For all of the randomly generated faces, these expressions were set to neutral. Nevertheless, to further ensure that the expressions are neutral, we set the mouth-shape control, which moves the corners of the mouth up and down, to neutral. Each face was exported to a 400 × 400 pixels bitmap with black background.
Participants and procedures (studies 3, 4, and 13). Seventy-five undergraduate students participated in the studies for partial course credit. They were asked to judge the 300 faces on trustworthiness (study 3, $n = 29$), dominance (study 4, $n = 25$), and threat (study 13, $n = 21$).

Participants were told to rely on their "gut feeling," and that there is no right

PSYCHOLOGY

or wrong answer. Each face was preceded by 500-ms fixation cross and presented at the center of the screen. The response scale ranged from 1 (Not at all [trait]) to 9 (Extremely [trait]).

The mean judgments averaged across participants were used to find dimensions of trustworthiness, dominance, and threat in the 50-dimensional face space (*SI Text*). Trustworthiness judgments (Mean = 4.75, SD = 0.66), dominance judgments (Mean = 5.17, SD = 1.05), and threat judgments (Mean = 4.81, SD = 0.91) were all reliable, α = 0.81, α = 0.92, and α = 0.87, respectively.

Model validation studies. We conducted three studies to validate the models of the orthogonal trustworthiness and dominance dimensions and the model of the threat dimension obtained from these two dimensions.

Participants and procedures (studies 5, 6, and 14). Fifty-four undergraduate students participated in the studies for partial course credit. Participants were asked to judge faces generated by (*i*) the trustworthiness model (study 5, *n* = 19), (*ii*) the dominance model (study 6, *n* = 17), and (*iii*) the threat model (study 14, *n* = 18). The procedures were the same as in studies 3, 4, and 13.

Face stimuli. We generated 20 random Caucasian faces using the same procedure as in studies 3 and 4. Using the trustworthiness model, for each face we created seven versions (−4.5, −3, −1.5, 0, 1.5, 3, and 4.5 SD). This resulted in 140 faces. The same procedures were used to generate faces that vary on dominance and faces that vary on threat.

Facial Cues Used for Valence and Dominance Evaluation of Faces. *Participants in emotion categorization study (study 7).* Nineteen adults were recruited in a shopping mall. They were paid $5 for their participation.

Face stimuli. We generated eight random Caucasian faces using the same procedures as in studies 3 and 4. Using the trustworthiness model, for each face we created four versions (−8, −4, 4, and 8 SD). The same procedures were used to generate faces that vary in dominance. This resulted in 72 faces (8 faces × 9 versions, including the 0 SD faces). The same stimuli were also used in studies 8–12.

Procedures. Participants were asked to judge the expression of each face on a seven-alternative forced-choice task. The response categories were neutral and the six basic emotions: anger, happiness, sadness, disgust, fear, and surprise. After three practice trials, each variant of each face was rated once, resulting in 72 trials total. On each trial, a face was presented with the caption ''This face shows which emotion?'' and the seven choices. Participants indicated their choice by using the number keys 1–7.

Participants and procedures (study 8): Judgments of angry–happy expressions. Nineteen adults different from the participants in study 7 were recruited in a shopping mall. They were paid $5 for their participation. Participants were asked to judge the expression of each face on a 9-point scale, ranging from 1 (angry) to 5 (neutral) to 9 (happy). After three practice trials, each variant of each face was rated once, resulting in 72 trials total. The same procedures were used in studies 9–12.

Participants and procedures (studies 9–12): Judgments of facial maturity and femininity–masculinity. Eighty-four undergraduate students participated in the studies for partial course credit or payment. Twenty-eight (study 9) were asked to judge the faces on a 9-point scale, ranging from 1 (baby-faced) to 5 (neutral) to 9 (mature-faced), and 19 (study 11) were asked to judge the same faces on a 9-point scale, ranging from 1 (feminine) to 5 (neutral) to 9 (masculine). Sixteen were asked to judge the faces with their internal features masked on the facial maturity scale (study 10, Fig. S8), and 21 were asked to judge the same faces on the femininity/ masculinity scale (study 12).

1. McNeill D (1998) in *The Face* (Little, Brown, Boston), pp 165–169.
2. Lombroso C (2006) in *Criminal Man* (Duke UnivPress, Durham, NC), p 51.
3. Hassin R, Trope Y (2000) Facing faces: Studies on the cognitive aspects of physiognomy. *J Pers Soc Psychol* 78:837–852.
4. Todorov A, Mandisodza AN, Goren A, Hall CC (2005) Inferences of competence from faces predict election outcomes. *Science* 308:1623–1626.
5. Ballew CC, Todorov A (2007) Predicting political elections from rapid and unreflective face judgments. *Proc Natl Acad Sci USA* 104:17948–17953.
6. Little AC, Burriss RP, Jones BC, Roberts SC (2007) Facial appearance affects voting decisions. *Evol Hum Behav* 28:18–27.
7. Blair IV, Judd CM, Chapleau KM (2004) The influence of Afrocentric facial features in criminal sentencing. *Psychol Sci* 15:674–679.
8. Eberhardt JL, Davies PG, Purdie-Vaughns VJ, Johnson SL (2006) Looking deathworthy: Perceived stereotypicality of Black defendants predicts capital-sentencing outcomes. *Psychol Sci* 17:382–386.
9. Willis J, Todorov A (2006) First impressions: Making up your mind after 100 ms exposure to a face. *Psychol Sci* 17:592–598.
10. Bar M, Neta M, Linz H (2006) Very first impressions. *Emotion (Washington, DC)* 6:269–278.
11. Phelps EA, LeDoux JE (2005) Contributions of the amygdala to emotion processing: From animal models to human behavior. *Neuron* 48:175–187.
12. Adolphs R, Tranel D, Damasio AR (1998) The human amygdala in social judgment. *Nature* 393:470–474.
13. Winston JS, Strange B, O'Doherty J, Dolan R (2002) Automatic and intentional brain responses during evaluation of trustworthiness of face. *Nat Neurosci* 5:277–283.
14. Engell AD, Haxby JV, Todorov A (2007) Implicit trustworthiness decisions: Automatic coding of face properties in human amygdala. *J Cognit Neurosci* 19:1508–1519.
15. Todorov A, Baron S, Oosterhof NN (2008) Evaluating face trustworthiness: A model based approach. *Soc Cognit Affect Neurosci* 3:119–127.
16. Todorov A (2004) The origins of first impressions. *J Cult Evol Psychol* 2:93–108.
17. Zebrowitz LA, Montepare JM (2006) in *Evolution and Social Psychology*, eds Schaller M, Simpson JA, Kenrick DT (Psychology Press, New York), pp 81–113.
18. Zebrowitz LA, Fellous JM, Mignault A, Andreoletti C (2003) Trait impressions as overgeneralized responses to adaptively significant facial qualities: Evidence from connectionist modeling. *Pers Soc Psychol Rev* 7:194–215.
19. Wiggins JS (1979) A psychological taxonomy of trait descriptive terms: The interpersonal domain. *J Pers Soc Psychol* 27:395–412.
20. Osgood CE, Suci GI, Tennenbaum PH (1957) *The Measurement of Meaning* (Univ of Illinois Press, Urbana).
21. Kim MP, Rosenberg S (1980) Comparison of two structural models of implicit personality theory. *J Pers Soc Psychol* 38:375–389.
22. Fiske ST, Cuddy AJC, Glick P (2007) Universal dimensions of social cognition: warmth and competence. *Trends Cognit Sci* 11:77–83.
23. Wiggins JS, Philips N, Trapnell P (1989) Circular reasoning about interpersonal behavior: Evidence concerning some untested assumptions underlying diagnostic classification. *J Pers Soc Psychol* 56:296–305.
24. Blanz V, Vetter T (1999) in *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques* (Addison Wesley Longman, Los Angeles), pp 187–194.
25. Hess U, Blairy S, Kleck RE (2000) The influence of facial emotion displays, gender, and ethnicity on judgments of dominance and affiliation. *J Nonverbal Behav* 24:265–283.
26. Knutson B (1996) Facial expressions of emotion influence interpersonal trait inferences. *J Nonverb Behav* 20:165–181.
27. Montepare JM, Dobish H (2003) The contribution of emotion perceptions and their overgeneralizations to trait impressions. *J Nonverb Behav* 27:237–254.
28. Fridlund AJ (1994) *Human Facial Expression: An Evolutionary View* (Academic, San Diego).
29. Adams RB, Ambady N, Macrae N, Kleck RE (2006) Emotional expressions forecast approach-avoidance behavior. *Motiv Emot* 30:179–188.
30. Marsh AA, Ambady N, Kleck RE (2005) The effects of fear and anger facial expressions on approach- and avoidance- related behaviors. *Emotion (Washington, DC)* 5:118–124.
31. Secord PF (1958) in *Person Perception and Interpersonal Behavior*, eds Tagiuri R, Petrullo L (Stanford Univ Press, Stanford, CA), pp 301–314.
32. Todorov A, Duchaine B (2008) Reading trustworthiness in faces without recognizing faces. *Cognit Neuropsychol* 25:395–410.
33. Keating CF, Mazur A, Segall MH (1981) A crosscultural exploration of physiognomic traits of dominance and happiness. *Ethol Sociobiol* 2:41–48.
34. Zebrowitz LA (1999) *Reading Faces: Window to the Soul?* (Westview, Boulder, CO).
35. Keating CF, Bai DL (1986) Children's attribution of social dominance from facial cues. *Child Dev* 57:1269–1276.
36. Zebrowitz LA, Montepare JM (2008) Social psychological face perception: Why appearance matters. *Soc Pers Psychol Compass* 2:1497–1517.
37. DeBruine LM (2005) Trustworthy but not lust-worthy: Context specific effects of facial resemblance. *Proc R Soc London Ser B* 272:919–922.
38. Izard CE, *et al.* (1995) The ontogeny and significance of infants' facial expressions in the first 9 months of life. *Dev Psychol* 31:997–1013.
39. Ghanzafar AA, Santos LR (2004) Primate brains in the wild: The sensory bases for social interactions. *Nat Rev Neurosci* 5:603–616.
40. Cheney DL, Seyfarth RM (1990) *How Monkeys See the World: Inside the Mind of Another Species* (Univ of Chicago Press, Chicago).
41. Bond CF, Berry DS, Omar A (1994) The kernel of truth in judgments of deceptiveness. *Basic Appl Soc Psychol* 15:523–534.
42. Berry DS (1990) Taking people at face value: Evidence for the kernel of truth hypothesis. *Soc Cognit* 8:343–361.
43. Zebrowitz LA, Voinescu L, Collins MA (1996) ''Wide-eyed'' and ''crooked-faced'': Determinants of perceived and real honesty across the life span. *Pers Soc Psychol Bull* 22:1258–1269.
44. Zebrowitz LA, Andreoletti C, Collins MA, Lee SY, Blumenthal J (1998) Bright, bad, babyfaced boys: Appearance stereotypes do not always yield self-fulfilling prophecy effects. *J Pers Soc Psychol* 75:1300–1320.
45. Lundqvist D, Flykt A, Ohman A (1998) *Karolinska Directed Emotional Faces* (Psychology Section, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm).
46. Singular Inversions (2005). Facegen Main Software Development Kit (Singular Inversions, Vancouver, BC, Canada), Version 3.1.