Alexander Todorov and
Nikolaas N. Oosterhof

Modeling Social Perception of Faces

In *Baboon Metaphysics*, a detailed investigation of the complexities of baboon life, primatologists Dorothy Cheney and Robert Seyfarth write, "Any way you look at it, most of the problems facing baboons can be expressed in two words: other baboons." This statement applies with even greater force to humans. Navigating the social world requires many cognitive feats, including keeping the identities of countless people straight, as well as the dynamics of their relationships. Our inferences about the social status, beliefs, desires, and intentions of other people determine whether we decide to approach or avoid them, what to say to them, and how to say it. Social complexity is one of the key factors driving brain evolution. Across primates, the size of the neocortex increases with the size of the group, and there is recent evidence that the quality of one's relationships has direct evolutionary benefits [1]. Maintaining suitable relationships requires sophisticated social cognition. At the basis of social cognition are the abilities to represent conspecifics as unique individuals and to perceive their intentions. In light of this, it should not come as a surprise that primates have specialized brain regions for the processing of faces and (in the case of humans) information about others' mental attributes [2].

SOCIAL PERCEPTION OF FACES

The face is our primary source of visual information for identifying people and reading their emotional and mental states. With the exception of prosopagnosics (who are unable to recognize faces) and those suffering from such disorders of social cognition as autism, people are extremely adept at these two tasks. However, our cognitive powers in this regard come at the price of reading too much into the human face. The face is often treated as a window into a person's true nature. References to this belief can be found in all ancient cultures, and the belief has persisted into modern times. The Swiss pastor Johann Kaspar Lavater, who pioneered the pseudoscience of physiognomy, described in detail how to read the true, inner nature of a person from facial features (e.g., "The nearer the eyebrows are to the eyes, the more earnest, deep, and firm the character" [3, p. 59]). Although attempts to characterize personality based on external appearance have largely fallen out of favor in science, the ideas continue to appeal at an intuitive, implicit level. Lavater was probably wrong about most of his specific claims, but research strongly supports his contention that: "Whether they are or are not sensible of it, all men are daily influenced by physiognomy." [3, p. 9]. First, people tend to agree in their social judgments based on faces, indicating that faces provide information that is consistently interpreted [4], [5]. Second, such judgments are made rapidly, without much mental effort: as little as 33 ms exposure to a face is sufficient for people to decide whether a face looks trustworthy or not [6]. Third, recent functional magnetic resonance imaging (fMRI) studies have shown that regions in the brain critical for emotion and decision making are activated when participants look at negatively perceived faces (untrustworthy and aggressive looking) even when the participants have not been asked to evaluate these faces [7]. Thus it appears that our brains automatically categorize faces. Finally, many studies have shown that social judgments based on faces predict important social outcomes, ranging from sentencing decisions to electoral success [8].

THE FACE IS OUR PRIMARY SOURCE OF VISUAL INFORMATION FOR IDENTIFYING PEOPLE AND READING THEIR EMOTIONAL AND MENTAL STATES.

STATISTICAL MODELS FOR FACE REPRESENTATION

Given the agreement in social perception of faces (see Table 1), it should be possible to model this perception. What differences in facial structure lead to appearance-based social inferences? For example, based on what perceptual information do people decide that a face looks trustworthy or untrustworthy? Human faces share the same spatial layout and differences between faces are subtle, making it difficult to characterize what differences trigger specific social inferences. In this respect, data-driven approaches that do not impose any a priori constraints on face perception can be particularly useful for modeling social perception. There are two basic tasks in these approaches: creating a statistical model of face representation and using this model to derive the changes in facial features that lead to corresponding changes in social judgments. There are several statistical approaches for characterizing the commonalities and differences among individual faces. They all

Digital Object Identifier 10.1109/MSP.2010.940006
Date of publication: 17 February 2011

attempt to reduce high-dimensional face representations [e.g., pixel values of photographs, or three-dimensional (3-D) points that define the skin surface] to a lower-dimensional "face space." The dimensions of the face space define abstract, global properties of faces that are not reducible to single features. Here we use the face space model implemented in Facegen (www.facegen.com), a derivative of Blanz and Vetter's work [9]. This model uses 50 dimensions to represent face shape and 50 dimensions to represent face reflectance (brightness, color, and texture variations on the surface map of the face). The face model in Facegen is based on a database of $N = 271$ faces laser-scanned in 3-D and subsequently aligned so that all faces share the same skin surface mesh topology (for details, see [9]). The ith face is represented by a shape vector

$$\vec{s}_i = [x_1, y_1, z_1, \ldots, x_{N_s}, y_{N_s}, z_{N_s}]^T$$

with coordinates for N_s vertices, and a reflectance vector

$$\vec{t}_i = [r_1, g_1, b_1, \ldots, r_{N_t}, g_{N_t}, b_{N_t}]^T$$

with red, green, and blue color values of the N_t pixels in the color bitmap that is projected on the skin surface (in Facegen, $N_s = 2{,}043$ and $N_t = 256 \times 256$).

The face vectors are submitted to a principal component analysis (PCA), a data-driven dimensionality reduction technique that allows for characterizing the most common variations in face shape and face reflectance. In this approach, shape variations are represented by an average face $\bar{\vec{s}} = 1/N \cdot \sum_m \vec{s}_m$ and a set of $k = 50$ orthogonal principal components (shape eigenfaces) $\vec{v}_1, \ldots, \vec{v}_k$ that have the greatest eigenvalues of the covariance matrix of the face coordinates. The shape of a face can then be approximated by a k dimensional weight vector $\vec{p}_i$, yielding shape coordinates

$$\vec{s}' = \bar{\vec{s}} + \sum_j \vec{v}_j \cdot \vec{p}_{ij} = \bar{\vec{s}} + \vec{V} \cdot \vec{p}_i,$$

where $\bar{\vec{s}}$ is the average shape and $V = [v_1 \cdots v_k]$ the matrix with principal components. Variations in reflectance are treated similarly, also with 50 components.

WITH THE AID OF A STATISTICAL FACE MODEL, IT IS RELATIVELY STRAIGHTFORWARD TO UNCOVER THE VARIATIONS IN THE STRUCTURE OF FACES THAT LEAD TO SPECIFIC SOCIAL JUDGMENTS.

Thus, faces are represented as an average face plus a weighted sum of the principal components (eigenfaces). This gives rise to the concept of face space, which is the space containing the faces that can be represented.

Assuming that shape and reflectance are approximately multinormally distributed, new faces that are plausible in the population can be generated in face space by constructing new weight vectors with random Gaussian values. A practical implication is that a virtually unlimited amount of faces can be generated using this approach, which makes it an attractive alternative to using a database of face photographs. Furthermore, as described in detail below, face properties related to shape and reflectance (the surface map of the face) can be independently manipulated. These qualities of the models allow for the constructions of vectors in face space that approximate specific social judgments and for tests of psychological hypotheses.

MODELING SOCIAL JUDGMENTS OF FACES

With the aid of a statistical face model, it is relatively straightforward to uncover the variations in the structure of faces that lead to specific social judgments [4], [10], [11]. Here, we describe models of nine different social judgments. The first task is to collect judgments of faces randomly generated by the statistical model and to show that these judgments are reliable. If the judgments are unreliable—there is a low or no agreement among judges—it is futile to try to model these judgments. As a rule of thumb, the reliability of the judgments sets the ceiling of their predictability. The second task is to test whether the statistical model of face representation can account for a meaningful proportion of the variance of these judgments. Assuming that this is the case, the third task is to construct new dimensions in face space that account for the maximum variability in the judgments. These dimensions then can be used to visualize the differences in facial structure that lead to specific judgments (Figures 1–3) and to manipulate faces along these dimensions [10], [11]. Table 1 lists nine different social judgments of 300 faces randomly generated by the statistical model described in the previous section. The most common measure of reliability used in psychological testing is Cronbach's alpha (α). This measure indicates the expected correlation between the ratings of the faces averaged across raters and the ratings of a new sample with the same number of raters. All

[TABLE 1] INTERRATER AGREEMENT AND RELIABILITY OF NINE SOCIAL JUDGMENTS OF EMOTIONALLY NEUTRAL FACES.

JUDGMENT	NUMBER OF RATERS (n)	INTERRATER AGREEMENT (r)	RELIABILITY (α)
DOMINANT	23	.36	.92
THREATENING	21	.26	.87
ATTRACTIVE	35	.23	.91
FRIGHTENING	28	.17	.84
MEAN	27	.17	.83
TRUSTWORTHY	29	.15	.81
EXTROVERTED	33	.14	.84
COMPETENT	44	.11	.84
LIKEABLE	31	.10	.76

RATERS (n) WERE ASKED TO MAKE JUDGMENTS OF 300 RANDOMLY GENERATED FACES ON A SCALE FROM 1 (NOT AT ALL [TRAIT TERM]) TO 9 (EXTREMELY [TRAIT TERM]).

judgments show sufficiently high reliability, ranging from .76 to .92. Because Cronbach's α is a function of the sample size of raters and the inter-rater agreement, it could be a misleading measure of the actual rater agreement (e.g., a large sample of raters with a low agreement will result in reliable judgments). As shown in the third column of Table 1, the interrater agreement varies as a function of the specific judgment. Whereas for some judgments, the agreement is relatively high (e.g., dominance), for others it is relatively low (e.g., likeability). As we show below, this agreement is an important constraint on the ability of statistical models to explain social judgments.

Table 2 lists the proportion of variance of the social judgments accounted for by the shape and reflectance components of the statistical model. Four things should be noted about these data. First, the model does a good job of explaining the variance of judgments. In all cases, the amount of explained variance is statistically significant. Second, there is a high correlation between the amount of variance accounted for by shape components and the amount of variance accounted for by reflectance components ($r = .86$). Third, the variance accounted for by the model that includes both shape and reflectance components is substantially smaller than the sum of the variances accounted for by shape components alone and reflectance components alone. This finding suggests that there is redundancy in shape and reflectance information. For example, a face with a dominant shape is likely to have dominant reflectance. Finally, there is a strong relationship between the inter-rater agreement in judgments (Table 1) and the amount of variance accounted for by shape and reflectance components ($r = .61$ and $r = .82$, respectively). That is, the statistical model better explains judgments for which there is a high interrater agreement. Although this is not surprising, it indicates a sensible behavior of the model.

Before we describe the construction of new social dimensions in face space, we note that introducing nonlinear, quadratic predictors in the statistical models can improve the predictability of social judgments. The quadratic models capture the intuition that extreme faces can be evaluated negatively. In fact, for seven out of nine judgments, the quadratic shape model accounted for significantly more variance than the linear model (Table 3). In contrast to shape, the quadratic reflectance model accounted for significantly more variance only for two judgments (Table 4). This finding is consistent with prior findings on attractiveness showing that averageness is more important for shape than reflectance information [12].

COMPUTING SOCIAL VECTORS IN FACE SPACE

Having shown that the statistical model of face representation accounts for meaningful variance of social judgments, we now describe the construction

[TABLE 2] PROPORTION OF VARIANCE OF SOCIAL JUDGMENTS OF FACES ACCOUNTED FOR BY SHAPE COMPONENTS, REFLECTANCE COMPONENTS, AND SHAPE AND REFLECTANCE COMPONENTS OF STATISTICAL MODEL OF FACE REPRESENTATION.

JUDGMENT	SHAPE	REFLECTANCE	SHAPE AND REFLECTANCE
DOMINANT	.751	.810	.906
THREATENING	.729	.691	.846
ATTRACTIVE	.393	.395	.603
FRIGHTENING	.498	.523	.730
MEAN	.696	.562	.811
TRUSTWORTHY	.486	.381	.640
EXTROVERTED	.692	.524	.800
COMPETENT	.355	.437	.623
LIKEABLE	.358	.329	.559

[TABLE 3] PROPORTION OF VARIANCE OF SOCIAL JUDGMENTS OF FACES ACCOUNTED FOR BY LINEAR AND QUADRATIC SHAPE COMPONENTS OF STATISTICAL MODEL OF FACE REPRESENTATION.

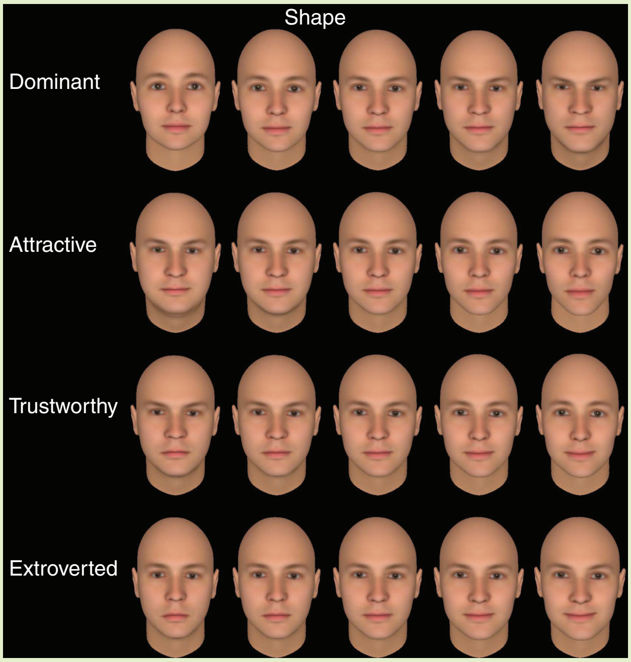
JUDGMENT	NONLINEAR MODEL	CHANGE IN ACCOUNTED VARIANCE	SIGNIFICANCE OF CHANGE
DOMINANT	.824	.073	$p < .008$
THREATENING	.784	.055	$p = .46$
ATTRACTIVE	.632	.239	$p < .0001$
FRIGHTENING	.654	.156	$p < .003$
MEAN	.758	.062	$p = .45$
TRUSTWORTHY	.674	.188	$p < .0001$
EXTROVERTED	.802	.110	$p < .0001$
COMPETENT	.612	.257	$p < .0001$
LIKEABLE	.578	.220	$p < .0002$

THE CHANGE IN ACCOUNTED VARIANCE SHOWS THE DIFFERENCE BETWEEN THE VARIANCE ACCOUNTED FOR BY THE QUADRATIC MODEL AND THE VARIANCE ACCOUNTED FOR BY THE LINEAR MODEL.

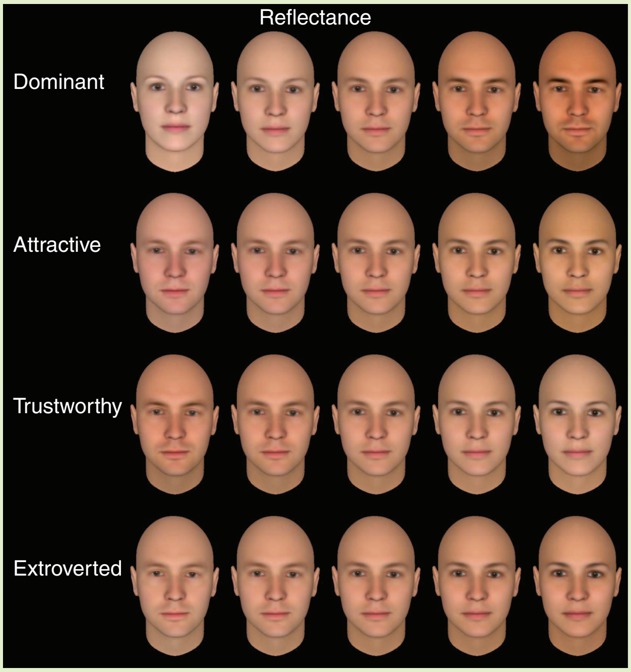
[TABLE 4] PROPORTION OF VARIANCE OF SOCIAL JUDGMENTS OF FACES ACCOUNTED FOR BY LINEAR AND QUADRATIC REFLECTANCE COMPONENTS OF STATISTICAL MODEL OF FACE REPRESENTATION.

JUDGMENT	NONLINEAR MODEL	CHANGE IN ACCOUNTED VARIANCE	SIGNIFICANCE OF CHANGE
DOMINANT	.855	.045	$P = .16$
THREATENING	.739	.048	$P = .90$
ATTRACTIVE	.530	.135	$P = .26$
FRIGHTENING	.627	.104	$P = .30$
MEAN	.646	.084	$P = .58$
TRUSTWORTHY	.502	.121	$P = .54$
EXTROVERTED	.638	.114	$P = .14$
COMPETENT	.597	.160	$P < .015$
LIKEABLE	.524	.195	$P < .010$

THE CHANGE IN ACCOUNTED VARIANCE SHOWS THE DIFFERENCE BETWEEN THE VARIANCE ACCOUNTED FOR BY THE QUADRATIC MODEL AND THE VARIANCE ACCOUNTED FOR BY THE LINEAR MODEL.



[FIG1] Variations of face shape on four social dimensions derived from judgments of dominance, attractiveness, trustworthiness, and extroversion. The perceived value of the faces on the respective dimensions increases from left to right.



[FIG2] Variations of face reflectance on four social dimensions derived from judgments of dominance, attractiveness, trustworthiness, and extroversion. The perceived value of the faces on the respective dimensions increases from left to right.

of new dimensions in face space that account for the maximum variability in the judgments.

Consider a set of randomly generated faces that have been judged on some characteristic [for example, trustworthiness rated on a scale from 1 (untrustworthy) to 9 (trustworthy) averaged over a group of participants]. A normalized linear face control $\vec{\Delta}'$ to manipulate this characteristic is constructed by

$$\vec{\Delta} = P \cdot \vec{r}, \quad \vec{\Delta}' = \vec{\Delta}/\|\vec{\Delta}\|,$$

where P_{ij} is the weight of component j for face i and $\vec{r}$ the ratings vector where the mean has been subtracted.

Intuitively, $\vec{\Delta}'$ can be considered as a normalized vector of correlations between the weights of each face component and the ratings. To justify this approach, consider that the face dimensions are, by construction, independent, and thus the obtained value for $\vec{\Delta}'$ is optimal in the least square sense.

Using the face control $\vec{\Delta}'$, an individual face with component weights $\vec{p}$ can be manipulated by α units by

$$\vec{p}\,' = \vec{p} + \alpha \cdot \vec{\Delta}'.$$

With the average shape $\vec{\bar{s}}$ and principal component matrix V described earlier, this changes the coordinates of the shape vertex components from

$$\vec{s} = \vec{\bar{s}} + \alpha \cdot \vec{\Delta}'$$

to

$$\begin{aligned} \vec{s}\,' &= \vec{\bar{s}} + V \cdot \vec{p}\,' \\ &= \vec{\bar{s}} + V \cdot (\vec{p} + \alpha \cdot \vec{\Delta}') \\ &= \vec{s} + \alpha \cdot V \cdot \vec{\Delta}', \end{aligned}$$

i.e., coordinates change linearly with changes in α. Reflectance is manipulated similarly. Face controls can be constructed for any face characteristic as long as a rating is associated with each face. Examples of face controls include hooked versus flat nose, masculine versus feminine [9], and the traits presented in this article [10], [11]. These methods uncover structural differences in appearance that predict differences in social perception. Figure 1 shows shape variations on four dimensions derived from judgments of dominance, attractiveness, trustworthiness, and extroversion, respectively. For each dimension, five versions of a face are shown, manipulated to decrease or increase its value on the respective dimension. For example, as the dominance of the face increases, the face becomes more masculine and mature. As the attractiveness increases, the face becomes thinner with higher cheekbones. As the trustworthiness increases, the face appears to express more positive emotions. As extroversion increases, the face becomes wider and happier. Figure 2 shows reflectance variations on the four dimensions. For example, as the dominance increases, the face becomes darker and more masculine. Similar darkness changes are also detectable for the other social dimensions. Figure 3 shows both shape and reflectance variations on the dimensions.

IMPLICATIONS OF FINDINGS

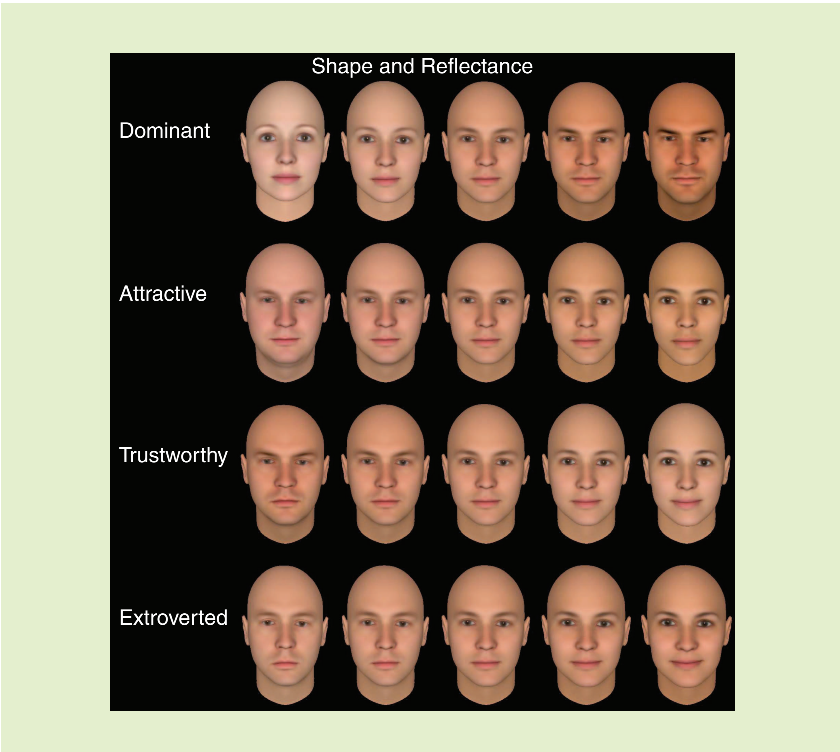
These models of social dimensions can be used to reveal the facial cues that lead to specific social judgments. For example, exaggerating the features that contribute to judgments of emotionally neutral faces reveals the underlying variations that account for these judgments. In the case of trustworthiness, although faces are perceived as emotionally neutral within the range shown in Figure 1, they are perceived as emotionally expressive outside this range [10]. Whereas faces at the extreme negative end of the dimension appear to express anger, faces at the extreme positive end appear to express happiness. In terms of social perception, these models provide clues about the basis of social inferences. Social inferences from facial appearance are based on resemblance to features that have adaptive significance—that is, to successfully navigate the social world, we need to be able to infer the emotional states, gender, and age of others [4], [5]. For example, facial expressions of emotion indicate a person's mental state and provide signals for appropriate behaviors. As a result, people with faces resembling specific emotional expressions, anger, for example, can be mistakenly judged as aggressive. When more sophisticated computer graphics and experimental methods are developed, we will have models that can be used not only to better understand social perception but also to manipulate images and create increasingly complex and lifelike avatars—knowledge that could be used for good or bad purposes. Such models can be used to manipulate images (not only of avatars but also of real people [11]) to induce specific perceptions that could influence potential important decisions ranging from consumer to voting behavior.

[THESE MODELS OF SOCIAL DIMENSIONS CAN BE USED TO REVEAL THE FACIAL CUES THAT LEAD TO SPECIFIC SOCIAL JUDGMENTS.]

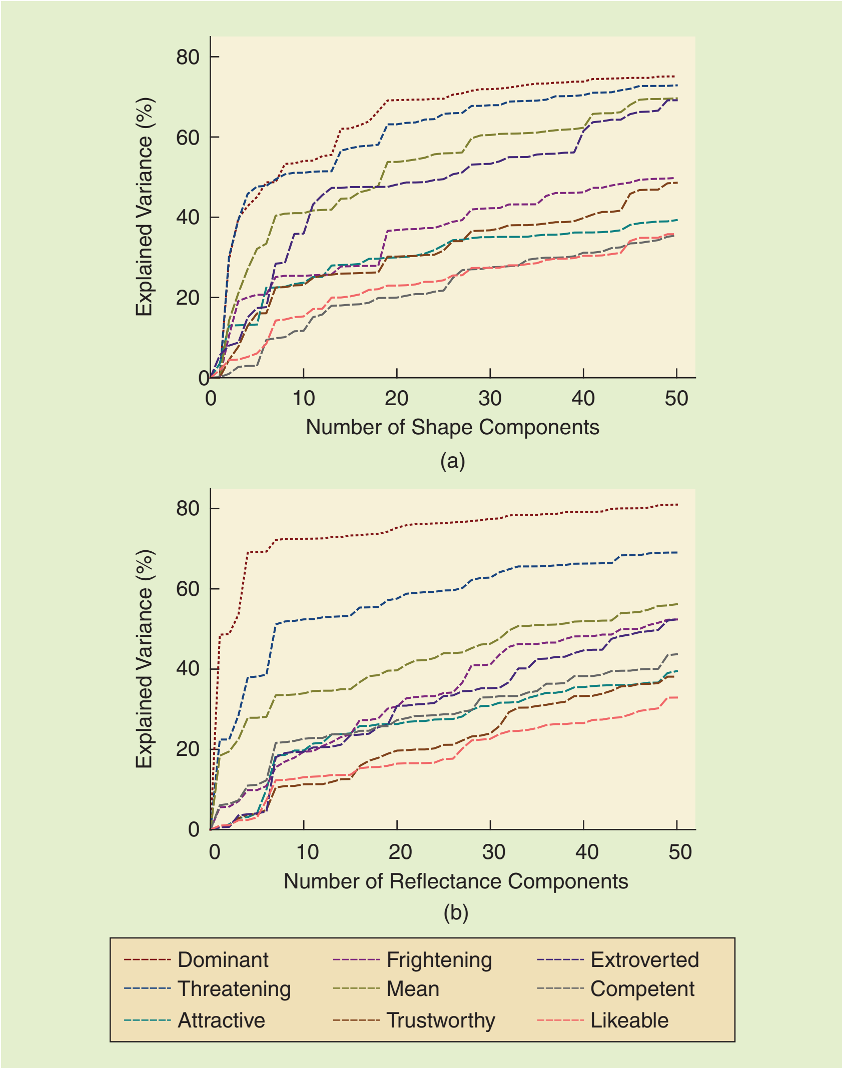
CHALLENGES AND FUTURE DIRECTIONS

One potential issue with the methods for modeling social perception of faces is overfitting. For example, here we used judgments of 300 faces to fit 50 shape and 50 reflectance parameters. Such models can perform well on the modeled set of faces but may fail to generalize to novel faces. In principle, larger training data sets should alleviate such problems.

Another potential approach is to use fewer parameters or face components. As described above, the face components were derived from a PCA, and, hence, each additional component accounts for less and less variance of facial appearance. This suggests that the first few components could capture most of the variance of social judgments. As shown in Figure 4, this is clearly the case. For example, for the shape model, the first ten components



[FIG3] **Variations of face shape and face reflectance on four social dimensions derived from judgments of dominance, attractiveness, trustworthiness, and extroversion. The perceived value of the faces on the respective dimensions increases from left to right.**



[FIG4] Explained variance of nine social judgments of faces as a function of the number of (a) shape and (b) reflectance components in the regression model.

account for more than half of the explained variance of the judgments. For the reflectance model, for many of the judgments, the first five components account for more than half of the explained variance.

Finally, the current models seem to perform rather well. First, judgments of faces manipulated by the models of these judgments agree with the models [10]. Second, the models predict judgments of novel faces. For an unrelated study, we generated another set of 300 faces that varied randomly on shape and were judged on trustworthiness and dominance. The correlations between the predicted trustworthiness and dominance scores and the judgments of these novel faces were .51 and .39 for trustworthiness and dominance, respectively, using a linear shape model, and .67 and .55, respectively, using a quadratic shape model. In principle, the models of social perception could be further improved. In addition to using larger face data sets and relying on the most informative face components, approaches that reduce the dimensionality of social judgments [10] and nonlinear approaches could be particularly fruitful.

ACKNOWLEDGMENTS

This work was supported by the U.S. National Science Foundation (0823749) and the Russell Sage Foundation. We would like to thank Jörn Diedrichsen for helpful comments on an earlier draft and Whitney Shapiro and Sara Verosky for their help.

AUTHORS

Alexander Todorov (atodorov@princeton.edu) is with the Department of Psychology, Princeton University, New Jersey.

Nikolaas N. Oosterhof (n.oosterhof@bangor.ac.uk) is with the Department of Psychology, Bangor University, United Kingdom.

REFERENCES

[1] J. B. Silk, "Social components of fitness in primate groups," *Science,* vol. 317, no. 5843, pp. 1347–1351, 2007.

[2] R. Adolphs, "The social brain: Neural basis of social knowledge," *Annu. Rev. Psychol.,* vol. 60, pp. 693–716, 2009.

[3] J. C. Lavater, *Essays on Physiognomy; for the Promotion of the Knowledge and the Love of Mankind*. London: Gale Group, 1772/1880. Eighteenth Century Collections Online. Abridged from Mr. Holcrofts translation.

[4] A. Todorov, C. P. Said, A. D. Engell, and N. N. Oosterhof, "Understanding evaluation of faces on social dimensions," *Trends Cogn. Sci.*, vol. 12, no. 12, pp. 455–460, 2008.

[5] L. A. Zebrowitz and J. M. Montepare, "Social psychological face perception: Why appearance matters," *Social Personal. Psychol. Compass*, vol. 2, no. 3, pp. 1497–1517, 2008.

[6] A. Todorov, M. Pakrashi, and N. N. Oosterhof, "Evaluating faces on trustworthiness after minimal time exposure," *Social Cogn.*, vol. 27, no. 6, pp. 813–833, 2009.

[7] A. Todorov and A. Engell, "The role of the amygdala in implicit evaluation of emotionally neutral faces," *Social Cogn. Affect. Neurosci.*, vol. 3, no. 4, pp. 303–312, 2008.

[8] C. Y. Olivola and A. Todorov, "Elected in 100 milliseconds: Appearance-based trait inferences and voting," *J. Nonverb. Behav.*, vol. 34, no. 2, pp. 83–110, 2010.

[9] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3D faces," in *Proc. 26th Annu. Conf. Computer Graphics and Interactive Techniques*, 1999, pp. 187–194.

[10] N. N. Oosterhof and A. Todorov, "The functional basis of face evaluation," *Proc. Nat. Acad. Sci. USA*, vol. 105, pp. 11087–11092, 2008.

[11] M. Walker and T. Vetter, "Portraits made to measure: Manipulating social judgments about individuals with a statistical face model," *J. Vis.*, vol. 9, no. 11, pp. 1–13, 2009.

[12] A. J. O'Toole, T. Price, T. Vetter, J. C. Bartlett, and V. Blanz, "3D shape and 2D surface textures of human faces: The role of 'averages' in attractiveness and age," *Image Vis. Comput.*, vol. 18, no. 1, pp. 9–19, 1999.

[SP]