Perception, 2008, volume 37, pages 126 – 142

doi:10.1068/p5805

Computational models of facial attractiveness judgments

P Matthew Bronstad¶, Judith H Langlois
Department of Psychology, University of Texas at Austin, Austin, TX 78712, USA
Richard Russell
Department of Psychology, Harvard University, Cambridge, MA 02138, USA
Received 18 July 2006, in revised form 24 January 2007; published online 19 December 2007

Abstract. We designed two computational models to replicate human facial attractiveness ratings. The primary model used partial least squares (PLS) to identify image factors associated with facial attractiveness from facial images and attractiveness ratings of those images. For comparison we also made a model similar to previous models of facial attractiveness, in that it used manually derived measurements between features as inputs, though we took the additional step of dimensionality reduction via principal component analysis (PCA) and weighting of PCA dimensions via a perceptron. Strikingly, both models produced estimates of facial attractiveness that were indistinguishable from human ratings. Because PLS extracts a small number of image factors from the facial images that covary with attractiveness ratings of the images, it is possible to determine the information used by the model. The image factors that the model discovered correspond to two of the main contemporary hypotheses of averageness judgments: facial attractiveness and sexual dimorphism. In contrast, facial symmetry was not important to the model, and an explicit feature-based measurement of symmetry was not correlated with human judgments of facial attractiveness. This provides novel evidence for the importance of averageness and sexual dimorphism, but not symmetry, in human judgments of facial attractiveness.

1 Introduction

Pythagoras, Heraclitus, and others believed that beauty results from good facial proportion and harmony between opposites (eg straight/curved, male/female) (Eco 2005). Partly inspired by these ancient ideas, modern researchers have identified three important factors in determining facial beauty. These factors are *averageness* (how much facial appearance deviates from the norm) (Langlois and Roggman 1990; Rhodes et al 1999; Valentine et al 2004), *sexual dimorphism* (how clearly male or female is the face) (Penton-Voak et al 2001; Perrett et al 1998), and *symmetry* of the two sides of the face (Grammer and Thornhill 1994; Rhodes et al 1999; Valentine et al 2004). The recent development of these ideas reflects a trend towards seeking biologically inspired explanations for beauty, unlike the historical tendency to pursue Platonic or religious significance in beauty (Etcoff 1999).

Averaged faces are attractive, and they are created by combining or averaging many face images (Langlois and Roggman 1990; Rhodes et al 1999; Valentine et al 2004). The more similar a face is to the averaged face, the more attractive it is predicted to be. Averaged faces are much like prototypes, the abstracted or typical members of a category. Interestingly, prototypes play a central role in recognition (Leopold et al 2001) as well as categorization of faces by gender, race, and emotional expression (Webster et al 2004). Sexual dimorphism refers to the differences between men's and women's faces. Men's jaws, brows, and noses are larger than women's (Enlow 1990), and facial pigmentation patterns are dimorphic (Bruce et al 1993). With some exceptions (Perrett et al 1998), participants prefer masculine men's faces to feminine men's faces, and especially feminine women's faces to masculine women's faces (O'Toole et al 1998; Penton-Voak et al 2001;

¶ Author to whom all correspondence should be sent at his current address: Schepens Eye Research Institute, 20 Staniford Street, Boston, MA 02114, USA; e-mail: matthew.bronstad@schepens.harvard.edu

Rhodes 2006). Symmetry refers to the degree of similarity between the right and left sides of the face; people prefer symmetric faces (Grammer and Thornhill 1994; Penton-Voak et al 2001; Rhodes et al 1999; Valentine et al 2004). The evolutionary and cognitive significance of these three factors has been the topic of considerable research, and though many researchers propose that facial appearance is an indicator of mate quality, this point is controversial (see Rhodes 2006 for review).

Although averageness, dimorphism, and symmetry are each important to attractiveness (Zebrowitz and Rhodes 2002), it is not understood how they codetermine perceptions of attractiveness in unaltered faces. Typically, researchers manipulate images to experimentally alter facial attractiveness. These studies are important; the stimuli are well-controlled, but they do not address codetermination because they examine only one or two factors concurrently. For example, increasing facial averageness by morphing a facial image with an averaged face increases the attractiveness of the resulting face, but the manipulation does not explicitly tell us how sexual dimorphism or symmetry affects the attractiveness of the face. Moreover, there are longstanding disagreements over the relative importance of the hypotheses (Rubenstein et al 2002) that are not easily resolved with image manipulations. Rather than determine the relative importance of the individual hypotheses, which will be affected by the sample of faces selected for analysis, it may instead be profitable to develop a formal account of how judgments are made. Computational modeling can be used for this end, because modeling requires a mechanistic description of the process. Modeling can also guide our understanding and intuitions about the cognitive processes underlying perception of facial beauty.

In previous attempts to create a model of attractiveness judgments, researchers predicted attractiveness ratings of faces using measurements of the distances between facial features (eg Cunningham 1986). A drawback of this method is that it represents faces with a very small quantity of data (fewer than six 2-D measures are typically used) and entire classes of facial information, such as pigmentation, are discarded. Yet this discarded information is relevant to attractiveness judgments (Russell 2003). Rhodes (2006) recently noted that current facial appearance measurement methods used in attractiveness research "...are poor, capturing only a limited part of a face's structure and nothing of its fattiness or skin quality" (page 204).

To offer a methodological improvement that addresses this concern, we used facial appearance measurements that retained nearly all aspects of facial appearance—the thousands of pixels in face images. We input images to a pattern-recognition model, encoding coarse and fine features, which include configuration, skin appearance, fattiness, and pigmentation. Such pixel-based pattern-recognition models are novel for replicating attractiveness judgments, but have successfully modeled other judgments made to faces, such as sex and race categorization (Cheng et al 2001; Furl et al 2002), and subjective judgments such as familiarity and distinctiveness (Hancock et al 1995). As will be described in greater detail in section 2 below, we chose a model that could be interrogated a posteriori, to investigate which image factors accounted for its performance, to address the question of codetermination. For comparison, we also implemented a model that used feature distances rather than pixels as inputs. To evaluate performance of all the models, we compared them with the 'gold standard' of facial attractiveness judgments—the consensus of a group of human raters. Specifically, we evaluated the performance of the models as a function of the correlation between their output and the mean attractiveness ratings given by a sample of human participants.

2 Method

2.1 *Subjects*

One hundred and two participants (fifty women), aged 18 – 75 years (mean = 34 years) rated each face for attractiveness on a 1 – 7 scale. Nearly all participants were Caucasian North Americans. We included two Asian – American participants whose attractiveness ratings were highly correlated with mean attractiveness ratings. Participants gave informed consent and methods were approved by University internal review boards.

A second group of fifteen participants (ten women) aged 19 – 34 years (mean = 25 years) rated each face for typicality and symmetry on 7-point scales. We gave participants the following instructions to rate typicality: "Faces vary in how typical they appear. Some faces look like the typical person you might see in a public place, whereas other people are very distinctive or atypical looking. Rate how typical each face looks to you". The instructions are similar to those used by Peskin and Newell (2004), in their work on familiarity and attractiveness, although their participants rated the converse of typicality (ie distinctiveness).

Rather than ask participants to explicitly rate the images for symmetry, we used a similarity task that has been previously employed for facial symmetry measurement (Penton-Voak et al 2001; Rhodes et al 2001). We constructed two mirror-image versions of every face image, left/left and right/right versions. Participants rated these image pairs for similarity. Symmetrical faces have highly similar left/left and right/right images. These similarity ratings correlate with measured symmetry (Penton-Voak et al 2001), whereas explicit judgments of symmetry made to unaltered face images do not (Scheib et al 1999).

2.2 *Images*

We analyzed 74 images of college-aged men's (38) and women's (36) faces. Photographs were taken of the students facing directly towards the camera with neutral expression. Images were 320×256 pixels in greyscale, and aligned so that iris locations coincided. Eye alignment is commonly used for face perception models and does not alter image aspect ratios (Cheng et a 2001; Hancock et al 1995). Some researchers use more exact image registration. For example, Hancock et al morphed images to precisely align facial features, though there is evidence that this is not necessary for good performance (Cheng et al). We did not observe improved performance after we altered the images so that vertical locations of mouths were consistent. An oval occluding window was applied to each image to obscure background and hair. The size of the window was kept constant for all images, and selected so that the external contour of every face was visible. Figure 1 shows an example of the stimuli and the size and dimensions of the window. Pixels within the window were analyzed.

2.3 *Statistical method*

Our model's structure is a computational implementation of our research questions. First, it discovers the appearance factors involved in attractiveness judgments. Second, it determines how to combine these factors to predict facial attractiveness. In the first stage, partial least squares (PLS) (Geladi and Kowalski 1986; McIntosh and Lobaugh 2004) analysis reduces the image data according to the attractiveness ratings data, encoding image features that predict attractiveness into several latent factors of the same size and shape as the images. PLS finds a set of common factors that decompose the independent and dependent variables X and Y, such that the factors maximize the covariance between X and Y. Here, PLS generates four factors that explain the images in terms of their mean attractiveness ratings (we observed no performance improvement for five or more factors). Each factor is an image with the same dimensions as the input faces. Each face image was then multiplied by each PLS factor, representing the image in reduced form.

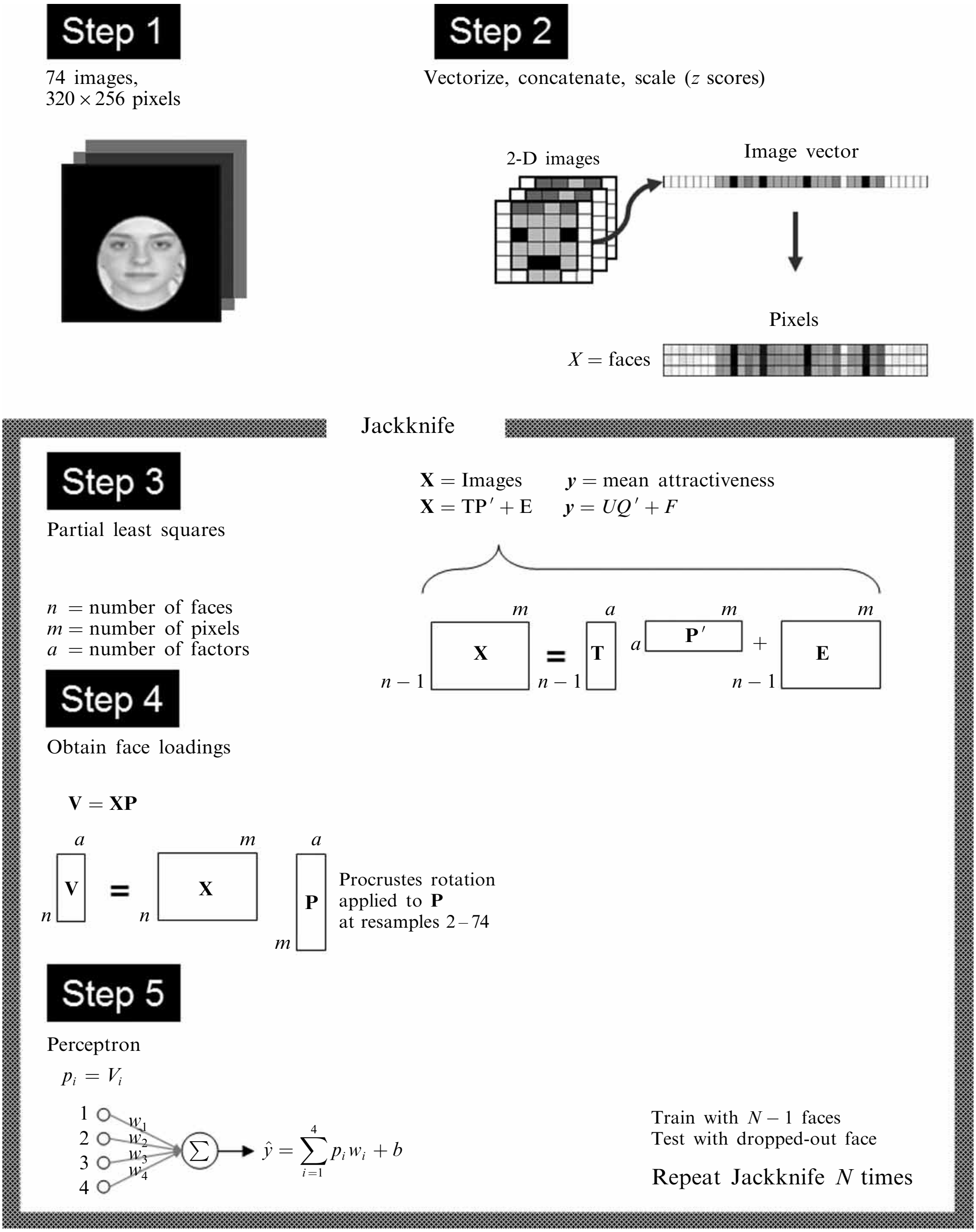

Figure 1. Visual depiction of the PLS model. Steps 1 and 2 show that images are vectorized and stored in matrix **X**. In the feature distance model, standardized feature distances are stored in matrix **X**. Vector $\boldsymbol{y}$ contains the z-scored mean attractiveness ratings. Steps 3 through 5 show the PLS statistical method, perceptron, within the jackknife. For PCA, the eigenvectors of the singular value decomposition of **X** are substituted for **V**.

In the second stage, a perceptron neural network determines how to weight the influence of each factor so that the model best predicts attractiveness judgments made of the images in the training set. We trained a perceptron (4 input cells, 1 output cell) on these factor responses; its goal was to make the best prediction of whether faces were attractive or unattractive (median-split on the mean attractiveness ratings for each face). We also used a linear neural network that treated attractiveness as a continuous measure during network training, rather than the perceptron, which treated attractiveness as a binary measure during training. The linear neural network, however, produced poor results for both pixel-based (PLS, PCA) and feature-distance models, suggesting that over-fitting may be a problem when modeling attractiveness data. Consequently, we did not continue to use the linear network.

A single image was excluded from both PLS data reduction and perceptron training so that, once trained on the remaining 73 test images, we observed the model's response to the excluded image. The simulation was repeated so that each face image was left out once during training and the model was tested on this image that it was not trained to predict (this is called *jack-knifing* or *leave-one out cross-validation*). The dropped out image was multiplied by each PLS factor, generating 4 numbers (p_i) that are multiplied by the perceptron weights (w_i). The weighted and summed factor responses give a continuous response which is the model's estimate of the attractiveness of the dropped-out image that it was not trained to predict. That is,

$$\hat{y} = \sum_{i=1}^{4} p_i w_i + b \,,$$

where b is the perceptron's beta weight. The model is shown graphically in figure 1.

PLS is, in some ways, similar to principal components analysis (PCA) used in many face-recognition algorithms (Cheng et al 2001; Furl et al 2002; Hancock et al 1995; O'Toole et al 1998). For example, both can discover components underlying variation in facial appearance, given a set of images of faces. However, there are some important differences between PCA and PLS. Most obviously, whereas PCA is an unsupervised learning method and a type of factor analysis, PLS is supervised and a form of regression.

We used PLS rather than PCA for several reasons. PLS potentially offered a more compact representation for face images—in fact PLS required fewer factors than PCA to find a solution relating variation in images to variation in attractiveness. Moreover, because PCA training is unsupervised and PCA factors are mutually orthogonal, it is unlikely that any particular PCA factor has predictive value for attractiveness ratings, but, rather, PCA may require a large combination of factors in order to function effectively. For example. Cheng et al (2001) found that optimal performance for classifying adult faces by sex was reached when using eight eigenvectors. We anticipated that sex classification is more straightforward than perception of beauty, and expected that modeling perception of beauty would be more computationally intensive. Compared to PCA, an unsupervised learning method, PLS is supervised and therefore each factor extracted in a PLS analysis has some predictive value on the dependent variable. For our purposes it was more straightforward to interpret the function of PLS factors, each of which ought to be relevant to facial attractiveness. Nonetheless, we also implemented a PCA-based model for comparison.

For all analyses, mean attractiveness ratings and images are z–scored and then scaled so that values are between -1 and 1. Procrustes rotation corrected arbitrary axis rotations and reflections of the factors across the resamples (McIntosh and Lobaugh 2004). We should note that PLS is a form of regression, and the PLS algorithm generates weights for the latent factors that are optimized for prediction of attractiveness, so the perceptron appears to be redundant. However, we employ a procrustes rotation after the PLS algorithm has generated weights; therefore those weights are no longer optimized after the rotation. We use the perceptron to find weights that are suited to the rotated factors. Use of PLS as a dimension-reduction method (followed by a secondary classifier) has been suggested as an improvement over PCA dimension reduction (Nguyen and Rocke 2002).

We constructed models in Matlab (Version 6.5, The MathWorks, Inc., Natick, MA, USA), using the neural network toolbox (Version 4.0, The MathWorks, Inc.) and the PLS toolbox (Version 3.5, Eigenvector Research, Inc., Manson, WA, USA).

2.4 *Alternative statistical models*

For comparison we also constructed models that used distances between facial features. The x and y coordinates of 64 facial feature landmarks (figure 2) were identified by hand on each face with custom software developed by Jean-Marc Fellous. We then computed a number of feature distances and facial proportions using the 64-feature landmarks. For further information see Appendices A, B, and Zebrowitz et al (2003). For testing alternative models, we used the same leave-one-out resampling method as used to validate the PLS model.

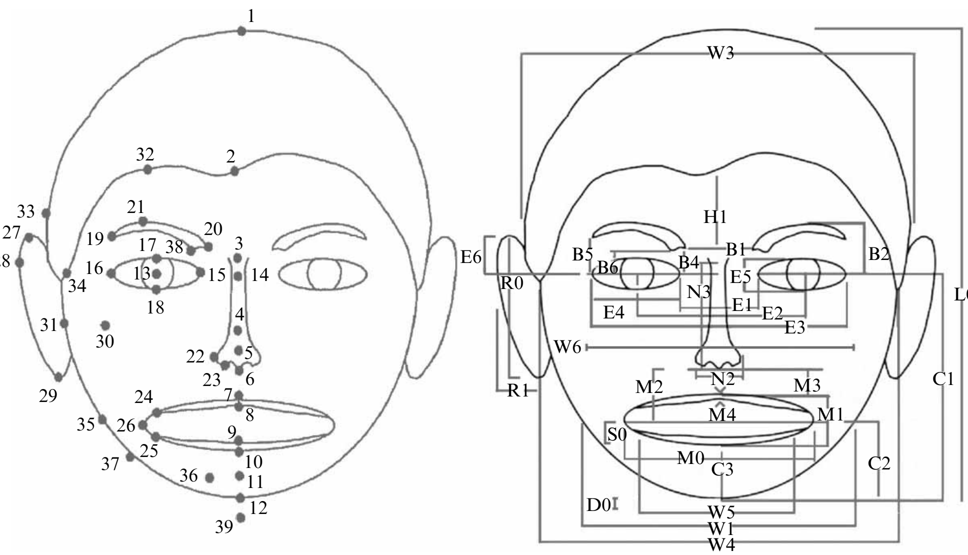

Figure 2. Facialmetric distances. (a) The locations of the feature landmarks (symmetrical landmarks are not shown on the face's other side). (b) A depiction of the feature distances used.

3 Results

3.1 *Participant ratings*

Mean attractiveness ratings ranged from 2.10 to 4.77 (mean $= 3.35$, SD $= 0.66$) for men's faces and from 1.55 to 5.02 (mean $= 3.47$, SD $= 0.84$) for women's faces. The participants' attractiveness ratings were reliable (Cronbach $\alpha = 0.986$) and the mean inter-rater correlations were similar to those reported in the literature (Langlois et al 2000; Thornhill and Gangestad 1999). We computed all pairwise correlations between non-redundant pairs of raters and used the r-to-z transformation before computing mean correlations, and the z-to-r transformation after computing the means. Average means between individual participants' attractiveness ratings were $r = 0.43$, $r = 0.42$ between pairs of men, $r = 0.44$ between pairs of women, and $r = 0.41$ between pairs of men and women raters (see figures 3c and 3d). As these differences were not significant, there is no compelling reason to treat data from male and female raters differently, so we computed means using both male and female ratings of attractiveness. Participants were correlated with the mean ratings at $r = 0.61$ for men's faces and $r = 0.72$ for women's faces (see figures 3a and 3b). We computed mean attractiveness ratings for each face and used these means with their corresponding facial images to train the computational model.

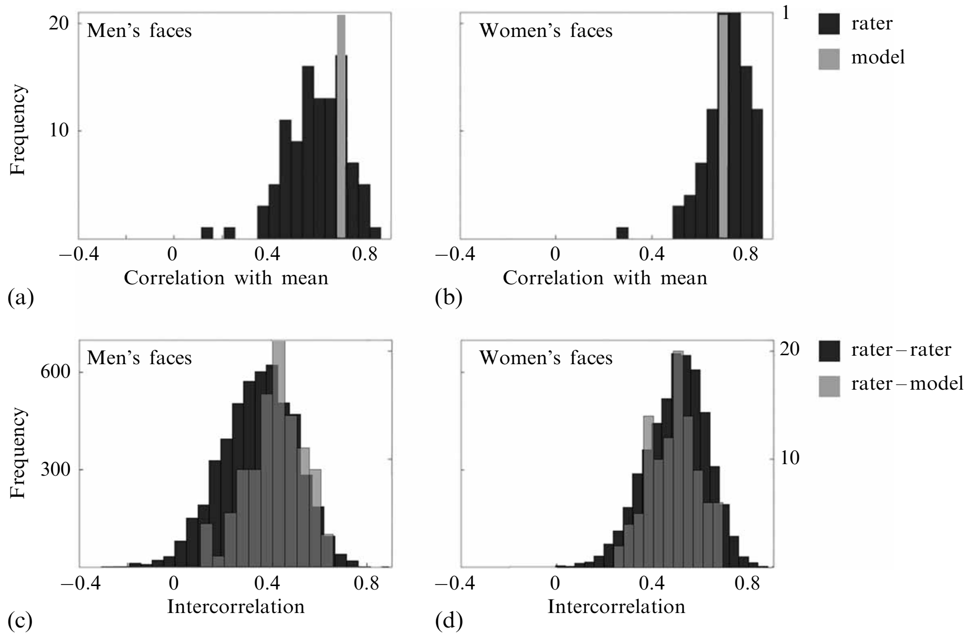

Figure 3. Comparison of human ratings and PLS model output. Upper histograms show correlations between each individual participants' ratings and mean attractiveness ratings (black) and between the model's output and mean attractiveness ratings (grey), separately for (a) men's faces and (b) women's faces. Lower histograms show correlations between raters (black) and correlations between raters and the model (grey) for (c) men's faces and (d) women's faces. Human (left) and model (right) y-axis scaling are different to facilitate comparison.

Mean typicality ratings ranged from 1.79 to 5.57 (mean = 4.23, SD = 0.98) for men's faces and from 2.07 to 5.93 (mean = 4.17, SD = 1.06) for women's faces. The ratings were reliable (Cronbach $\alpha = 0.822$, mean inter-rater correlation = 0.25). Mean symmetry ratings ranged from 2.64 to 5.92 (mean = 4.52, SD = 0.81) for men's faces and from 2.64 to 6.07 (mean = 4.49, SD = 0.80) for women's faces. The ratings were reliable (Cronbach $\alpha = 0.878$, mean inter-rater correlation = 0.34). We computed mean symmetry and typicality ratings for each face after dropping one participant's data (a different participant in each case), whose ratings were uncorrelated with the mean ratings.

Mean attractiveness and typicality were correlated, $r = 0.67$ for men's faces and $r = 0.66$ for women's faces. Attractiveness and symmetry were correlated, $r = -0.11$ for men's faces and $r = 0.07$ for women's faces. Symmetry and typicality were correlated, $r = 0.07$ for men's faces and $r = 0.22$ for women's faces.

3.2 *Model performance*

After training, we tested whether the model could predict participants' attractiveness ratings of faces on which it was not trained. The model's attractiveness ratings of novel images correlated $r = 0.70$ with mean attractiveness judgments by humans of women's faces, and $r = 0.68$ with mean attractiveness judgments of men's faces. The model's output is actually correlated with the mean ratings slightly *more* than the ratings of the average human subject, particularly for male faces (see figures 3a and 3b). Moreover, human raters and the model are inter-correlated such that it is difficult to distinguish the model's output from human attractiveness ratings (see figures 3c and 3d). Human-like performance is also observed in pixel-based face perception models of tasks such as gender classification or face recognition (Cheng et al 2001; Hancock et al 1995;

O'Toole et al 1998). The model's performance satisfies one requirement of a model of facial attractiveness perceptions: it mimics human preferences.

We used a similar pixel-based model: the standard PCA + perceptron used in many computational models of face perception (Cheng et al 2001; Hancock et al 1995; O'Toole et al 1998). The PCA + perceptron model gave optimal correlation between its output and attractiveness ratings when using eight eigenvectors (which was fewer than we anticipated would be necessary, but nonetheless twice as many as the PLS model requires). The output of the PCA model was correlated $r = 0.70$ with mean attractiveness ratings and $r = 0.92$ with the output of the PLS model. We focus our discussion of alternative models on the feature-distance model, as comparing pixel-based and feature-distance-based models is more enlightening than comparing two pixel-based models with similar output.

3.3 *Performance of alternative models*

The objective measures of facial variation typically used in psychological studies of attractiveness are the distances between facial features. Researchers have used feature distances to predict facial attractiveness for at least two decades, but there is no standard formula for combining the features into a single variable to represent or predict the attractiveness rating of a face. Cunningham (1986) reported that, for a sample of women's faces, a regression equation using the following features produced the best correlation with rated attractiveness: eye height, nose area, cheek width, smile width (multiple $r = 0.73$, although no cross-validation analysis was reported). We found that for our sample, a similar regression using highly similar features (E5, N4 × N3, W6, and M0, respectively) produced nonsignificant regression equations (all faces: $F = 1.44$, $p = 0.23$; men's faces alone: $F = 0.44$, $p = 0.78$; women's faces alone: $F = 1.19$, $p = 0.33$).

To predict facial attractiveness on the basis of facial feature distances, we first tried simply submitting all the feature distances and ratios to a perceptron. This model performed modestly, correlating with men's attractiveness $r = 0.37$, and women's attractiveness $r = 0.38$. This is considerably worse than the pixel-based model. One of the individual features was a measure of facial symmetry used by Penton-Voak et al (2001) and originally proposed by Grammer and Thornhill (1994). This symmetry measure compares the locations of the horizontal midpoints between pupils, inner and outer eye corners, cheekbones, outer points of nose, mouth, and jaw. We found that this measure correlated with rated symmetry (similarity between right/right and left/left facial images), $r = 0.40$, but not with rated attractiveness, $r = 0.05$. Interestingly, rated symmetry was also not correlated with rated attractiveness ($r = 0.02$, ns). These results are similar to those of Penton-Voak et al (2001), who found that measured symmetry correlated $r = 0.48$ with their participants' judgments of facial symmetry (again, with left/left and right/right facial images), though they found these symmetry judgments to be marginally correlated with attractiveness ($r = 0.22$). In a meta-analysis, Rhodes (2006) reports small effect sizes (R) for rated (0.30) or measured (0.19) symmetry and attractiveness.

After finding that submitting all measures to a perceptron produced poor model performance, we noted that most of the feature distances and ratios had low correlations with attractiveness (mean $r = 0.02$, SD $= 0.14$, max $r = 0.28$, min $r = -0.39$). We selected only those with correlations 0.25 or stronger and submitted those to a perceptron. These measurements were pupil height (above chin), distance between eyebrows, 'double chin' height, head length, and vertical distance of the jaw's edge from the pupil. Many of these measures are large in scale, compared to many of the other features that were measured (see Appendix B). However, we could not increase performance of the model by selecting only the most highly correlated features.

As a final attempt to improve performance of the feature-distance model, we performed a PCA on all the facial feature measurements and ratios and submitted the eigenvectors to a perceptron (there was no compelling reason to perform PLS on the features; however, PLS gave similar results). The performance of the model was optimal with the first eigenvector; correlating $r = 0.78$ with men's attractiveness and $r = 0.61$ with women's attractiveness. Interestingly, while this model was reasonably well correlated with mean attractiveness ratings, it was correlated with the PLS model $r = 0.71$ for men's faces and $r = 0.47$ for women's faces. Using a formula from Cohen and Cohen (1983) to compare two correlation coefficients drawn from the same sample, we determined that neither the PLS nor the facialmetric model was correlated significantly higher or lower with attractiveness for men's faces ($t = 1.49$, $p = 0.15$), or for women's faces ($t = 0.49$, $p = 0.49$).

We observed better-than-expected performance out of the feature-distance models, but focus our discussion on the mechanisms of the PLS model, as it uses fewer components than the PCA and is more easily interpreted than the feature-distance models. It is difficult to interpret the PCA feature-distance measures because the perceptron weights the eigenvector of more than 50 facial feature distances and proportions. However, it is important to note in the interim that human-like performance was also achieved by a model that took only feature distances as its inputs, with no representation of reflectance (pigmentation) aspects of the face, such as skin quality or coloration.

3.4 *Investigating PLS model components*

Because the PLS model gives human-like attractiveness estimates, its mechanism may help us understand how people formulate attractiveness judgments. We therefore conducted several exploratory analyses of the PLS factors. During training, the factors are sensitized to different aspects of facial appearance. To identify the aspects of facial appearance to which the factors are sensitive we examine the visual appearance of the factors (see figure 4a), whether they respond differently to men's and women's faces, how strongly each correlates with rated symmetry and typicality, and how strongly each factor affects overall output independently of the other factors.

The first two PLS factors are much like pixel-averaged faces. The first factor appears somewhat prototypical and slightly masculine, whereas the second appears the same as the average of the entire set of images. Similarly, in PCA analyses of face images the first factor is the average of the training set, and the response of this factor to images of faces (ie the product of the factor and facial image) has been proposed as an averageness measurement (O'Toole et al 1998).

To determine the degree of similarity between factor 2 and the average of the face images (the average pixel value at each pixel location in the image frame), we scaled the four PLS factors and the average face so that all values were in the range 0–255 (ie 8-bit greyscale), and then subtracted the average from each of the factors. The results of these subtractions are shown graphically in figure 4b. It can be appreciated from this figure that subtracting the average face from factor 2 resulted in a nearly null matrix, indicating that factor 2 is virtually identical to the average face.

We also determined whether any of the factor responses were correlated with rated typicality, which is a proxy measure of prototypicality or averageness. Table 1 shows correlations between individual PLS factors and rated attractiveness, typicality, and symmetry, as well as t-tests of the sexual dimorphism of each factor. Factor 2 was significantly correlated with rated typicality ($r = 0.31$, $p = 0.007$; $r = 0.23$, ns, for men; $r = 40$, $p = 0.02$, for women). As typicality and attractiveness are highly correlated ($r = 0.65$, $p < 0.001$ in this sample), it is not surprising that at least one of the factors is correlated with both attractiveness and typicality. However, the responses of factor 2

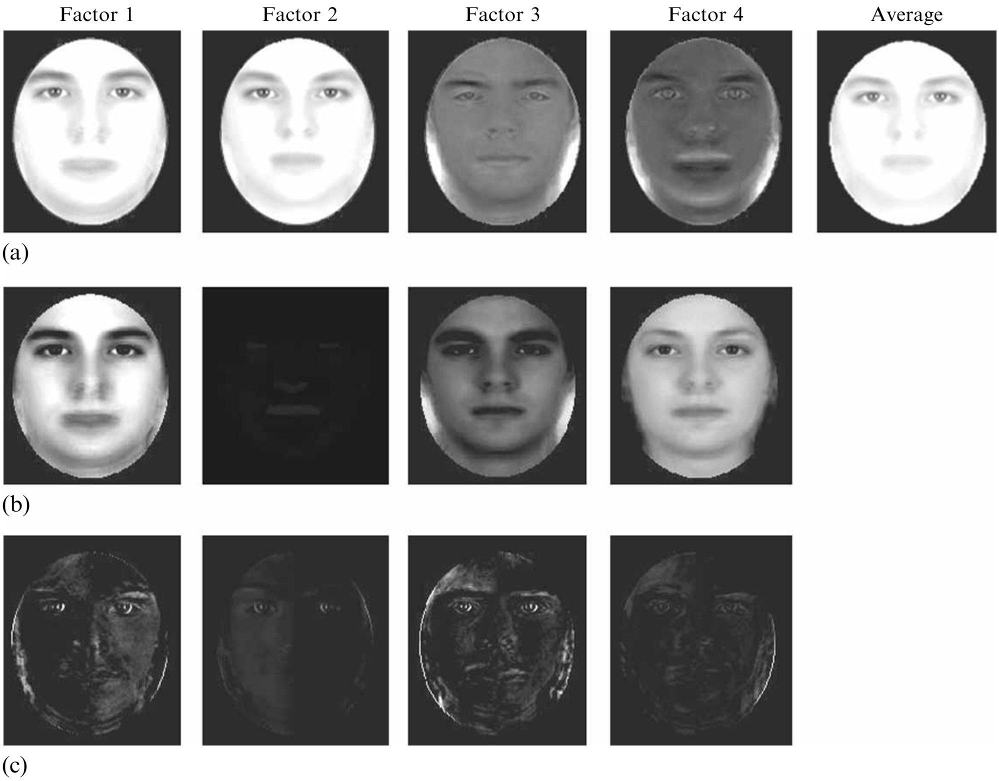

Figure 4. (a) The appearance of the PLS factors. (b) The result of subtracting the average of the 74 faces from each of the factors. That the second image from the left is blank indicates a high degree of similarity between factor 2 and the average face. (c) The result of subtracting mirror-reversed factor from itself. Light areas indicate asymmetries.

Table 1. PLS model and its individual factors. Correlations with attractiveness, typicality, and symmetry, and t-test for sexual dimorphism.

	PLS model	PLS factor 1	PLS factor 2	PLS factor 3	PLS factor 4
All faces					
PLS model	1.00	0.01	0.55*	−0.14	0.18
Attractive	0.67*	0.07	0.23*	−0.08	0.07
Typicality	0.58*	−0.04	0.31*	−0.07	0.04
Symmetry	0.05	−0.11	0.10	0.02	−0.03
Sex (t-value)	1.55^{M}	2.42*	0.55	3.15*	2.48*
Women's faces					
PLS model	1.00	−0.17	0.56*	0.12	−0.04
Attractive	0.67*	−0.03	0.26	0.02	−0.14
Typicality	0.60*	−0.19	0.40*	0.02	−0.14
Symmetry	0.22	0.07	0.08	−0.03	−0.19
Men's faces					
PLS model	1.00	0.07	0.59*	−0.27	0.32
Attractive	0.68*	0.12	0.21	−0.13	0.28^{M}
Typicality	0.58*	0.10	0.23	−0.19	0.27
Symmetry	−0.08	−0.30*	0.13	0.05	0.17

* Significant, at least $p < 0.05$, ^{M} Marginal significance (all two-tailed tests).

are more highly correlated with typicality than with attractiveness (correlations with attractiveness: $r = 0.24$, $p = 0.04$, for all faces, and $r = 0.26$ and $r = 0.21$, both ns, for men and women, respectively). That factor 2 is nearly identical to the average face and that its responses are more highly correlated with typicality than with attractiveness, particularly for women's faces, is strong evidence that factor 2 is a measure of facial averageness.

Averaged faces are often highly symmetrical so we determined whether factor 2 or any of the other factors were symmetrical by subtracting each factor from its mirror-image (horizontally reflected). The results can be seen in figure 4c. Factors 2 and 4 are fairly symmetrical, more so than factors 1 and 3, so their responses may be an indication of facial symmetry. We examined whether the responses of the factors correlated with rated symmetry of the faces and found that none were correlated significantly with facial symmetry of both men's and women's faces (all ps > 0.2), but factor 1 had a marginally negative correlation with men's rated symmetry ($r = -0.30$, $p = 0.10$), but not with women's symmetry ($r = 0.02$, ns). Although factors 2 and 4 are themselves very symmetrical, there is no compelling evidence that their responses indicate the symmetry of faces. It is possible that factor 1, on the other hand, is an indication of men's facial symmetry, though factor 1 was very weakly correlated with the output of the model, and with rated attractiveness. Additionally, rated symmetry was not correlated with attractiveness ($r = 0.02$). Thus, our results do not support the hypothesis that symmetry is an important feature in facial attractiveness.

To investigate whether the model implements a measure of sexual dimorphism, we determined whether the factor responses differentiated men's and women's faces; this is how dimorphism measurements are typically validated (Penton-Voak et al 2001). We used the product of the factor responses and the perceptron weights for each of the dropped-out images, separately for each factor, to represent that factor's response to the faces. t-Tests ($\alpha = 0.05$, df adjusted to not assume homogeneous variance) determined whether these measures of the responses of the model were sexually dimorphic. Factors 1, 3, and 4 exhibited significantly different responses for both male and female faces ($t_{72} = 2.42$, $p = 0.018$; $t_{71} = 3.14$, $p = 0.002$; and $t_{68} = 2.48$, $p < 0.015$, respectively (table 1), whereas factor 2 was not significant by sex of face ($t_{72} = 0.55$, ns).

Although factors 1, 3, and 4 exhibited significantly different output by sex of face, the output of factor 4 may be of particular interest as it was more strongly correlated with men's attractiveness ($r = -0.28$, $p = 0.09$), than with women's attractiveness ($r = 0.14$, ns). Not only do those three factors have differential responses to men's and women's faces, they are also oppositely correlated with men's and women's attractiveness. Also note how subtracting the average from factors 3 and 4 (figure 4b) results in clearly masculine (factor 3) and feminine (factor 4) images. This is very similar to O'Toole et al's (1998) analysis of PCA eigenvectors that discriminated male and female faces, in which they found that adding the first and second eigenvectors made a masculine image, whereas subtracting the second eigenvector from the first eigenvector resulted in a feminine image (the first eigenvector in their PCA analyses was in fact the average). Overall, the fact that three factors exhibit different responses to men's and women's faces is a good indication that the model uses measures of sexual dimorphism in its estimation of facial attractiveness.

We performed similar correlation analyses for the feature-distance model, which used a single eigenvector. The feature model was correlated $r = 0.47$ with typicality, and $r = 0.10$ with symmetry. These correlations are presented here for consistency with the PLS model, but are less informative. That is, because the feature model uses a single factor to predict attractiveness, these correlations are essentially the same as those among rated attractiveness with these other ratings.

4 Discussion

We have demonstrated two computational models that can predict human ratings of facial attractiveness such that the outputs of the models are indistinguishable from the ratings of our human participants. The models predicted mean ratings of male and female facial attractiveness as well or better than the average human rater. This represents an advance towards a formal understanding of facial beauty, which may have interesting practical applications.

We interpreted the PLS factors as encoding facial averageness and dimorphism based on their appearance and other properties. Factors 1, 3, and 4 exhibited significantly different responses to the faces of men and women. Factor 2 is highly similar to the average face of the set and its responses correlate more highly with facial typicality than they do with attractiveness. Responses from factor 1 were marginally correlated with rated symmetry for male but not female faces. However, a feature-distance metric that indexes facial symmetry correlated with rated symmetry but not with attractiveness. Overall, the results did not provide support for symmetry as an important factor in attractiveness. However, effect sizes for symmetry are generally very small—much smaller than those for averageness or sexual dimorphism (Rhodes 2006)—and so any one study may fail to find an effect of asymmetry owing to sampling variability.

For comparison we also presented results of a computational model that used feature distances. For this model, factor analysis (PCA) is essential for good performance from feature-distance models, which matched the pixel-based model. Variation in skin quality and fattiness is minor in the range of images we have used here, but if their variation was greater, or if the variation in age was greater, the pixel-based model would presumably have a greater advantage over the feature-distance model, because the feature-distance model does not encode variation in skin quality such as wrinkling and complexion. The feature-distance model, however, has the advantage of using extremely well-corresponded data which must, in part, account for its success in modeling attractiveness ratings. That is, because the locations of facial features are identified, they correspond to the same image features on different faces, whereas, with pixel-based analyses, features such as the mouth may not be in the same location in the image.

A second possibility for the success of the facialmetric model is that low-spatial-frequency information may be sufficient for prediction of attractiveness in some circumstances (Sadr et al 2002). In the present analyses, features with correlations of 0.25 or stronger with attractiveness were fairly large-scale features (eg head length, pupil-to-chin distance). It is hypothesized that in an early stage of object and face detection a low-spatial-frequency representation of the stimulus is sent to the orbitofrontal cortex and medial frontal cortex, and during this time it is compared to a face template (Bar et al 2006; Summerfield et al 2006). The orbitofrontal cortex is also involved in perception of facial beauty (Ishai 2007). This suggests that a code capturing the important low-spatial-frequency information in faces, properly tuned, can model facial attractiveness judgments, when variation in skin quality is minimal.

A few limitations should be noted. First, we do not intend to minimize the importance of individual differences in the perception of attractiveness, which are clearly important (Hönekopp 2006). Second, our pixel-based model cannot assign different weights to different facial areas; attractiveness is treated holistically by the model, though humans may give disproportionate weight to particular areas of the face. Third, we used a single race and age group to test our model. Although race and especially age are important factors in attractiveness judgments, our sample of faces is standard for the attractiveness literature. Finally, we do not assert that the human visual system literally implements one of the algorithms used in this study.

These methods have practical applications, particularly for automated beauty assessment, but also for greater social competence for robots and other artificial agents (Arbib and Fellous 2004). Facial beauty is a central concern of visual aesthetics, and other studies have shown the usefulness of applying computational analysis to understanding the beauty of works of art such as abstract paintings (Taylor 2002) and Zen gardens (Van Tonder et al 2002). Along with recent advances in understanding the physiological mechanisms of the perception of facial attraction (Aharon et al 2001; Shimojo et al 2003), our model is an important step towards developing a comprehensive account of beauty.

Acknowledgments. This paper was supported by a NIH grant to JHL. We appreciate help and advice from Sebastian Seung, Leslie Zebrowitz, Ted Barker, Derke Isacowitz, Estee Fleishman, Brian Richburg, and Cheryl Browne.

References

Aharon I, Etcoff N, Ariely D, Chabris C F, O'Connor E, Breiter H C, 2001 "Beautiful faces have variable reward value: fMRI and behavioral evidence" *Neuron* **32** 537 – 551

Arbib M A, Fellous J-M, 2004 "Emotions: From brain to robot" *Trends in Cognitive Sciences* **8** 554 – 561

Bar M, Kassam K S, Ghuman A S, Boshyan A M, Schmidt A M, Dale A M, Hamalainen M S, Marinkovic K, Schacter D L, Rosen B R, Halgren E, 2006 "Top – down facilitation of visual recognition" *Proceedings of the National Academy of Sciences of the USA* **103** 449 – 454

Bruce V, Burton A M, Hanna E, Healey P, Mason O, Coombes A, Fright R, Linney A, 1993 "Sex discrimination: how do we tell the difference between male and female faces?" *Perception* **22** 131 – 152

Cheng Y D, O'Toole A J, Abdi H, 2001 "Classifying adults' and children's faces by sex: Computational investigations of subcategorical feature encoding" *Cognitive Science* **25** 819 – 838

Cohen J, Cohen P, 1983 *Applied Multiple Regression/Correlation Analysis for the Behavioural Sciences* 2nd edition (Hillsdale, NJ: Lawrence Erlbaum Associates)

Cunningham M R, 1986 "Measuring the physical in physical attractiveness: Quasi-experiments on the sociobiology of female beauty" *Journal of Personality and Social Psychology* **50** 925 – 936

Eco U, 2005 *History of Beauty* (New York: Rizzoli)

Enlow D H, 1990 "Control processes in facial growth", in *Facial Growth* 3rd edition, Ed. D H Enlow (Philadelphia, PA: Harcourt Brace Jovanovich) pp 229 – 248

Etcoff N, 1999 *Survival of the Prettiest: The Science of Beauty* (New York: Doubleday)

Furl N, Phillips P J, O'Toole A J, 2002 "Face recognition algorithms and the other-race effect: computational mechanisms for a developmental contact hypothesis" *Cognitive Science* **26** 797 – 815

Geladi P, Kowalski B R, 1986 "Partial least-squares regression: A tutorial" *Analytica Chimica Acta* **185** 1 – 17

Grammer K, Thornhill R, 1994 "Human (*Homo sapiens*) facial attractiveness and sexual selection: the role of symmetry and averageness" *Journal of Comparative Psychology* **108** 233 – 242

Hancock P J B, Burton A M, Bruce V, 1995 "Pre-processing images of faces: Correlations with human perceptions of distinctiveness and familiarity", in *Proceedings of IEEE Fifth Conference on Image Processing and its Applications, Edinburgh, July 1995* pp 727 – 731

Hönekopp J, 2006 "Once more: Is beauty in the eye of the beholder? Relative contributions of private and shared taste to judgments of facial attractiveness" *Journal of Experimental Psychology: Human Perception and Performance* **32** 199 – 209

Ishai A, 2007 "Sex, beauty and the orbitofrontal cortex" *International Journal of Psychophysiology* **63** 181 – 185

Langlois J, Roggman L A, 1990 "Attractive faces are only average" *Psychological Science* **1** 115 – 121

Langlois J H, Kalakanis L, Rubenstein A J, Larson A, Hallam M, Smoot M, 2000 "Maxims or myths of beauty? A meta-analytic and theoretical review" *Psychological Bulletin* **126** 390 – 423

Leopold D, O'Toole A J, Vetter T, Blanz V, 2001 "Prototype-referenced shape encoding revealed by high-level aftereffects" *Nature Neuroscience* **4** 89 – 94

McIntosh A R, Lobaugh N J, 2004 "Partial least squares analysis of neuroimaging data: Applications and advances" *NeuroImage* **23** S250 – S263

Nguyen D V, Rocke D M, 2002 "Tumor classification by partial least squares using microarray gene expression data" *Bioinformatics* **18** 39 – 50

O'Toole A J, Deffenbacher K A, Valentin D, McKee K, Huff D, Abdi H, 1998 "The perception of face gender: The role of stimulus structure in recognition and classification" *Memory and Cognition* **26** 146 – 160

Penton-Voak I, Jones B C, Little A C, Baker S, Tiddeman B, Burt D M, Perrett D I, 2001 "Symmetry, sexual dimorphism in facial proportions and male facial attractiveness" *Proceedings of the Royal Society of London, Series B* **268** 1617 – 1623

Perrett D J, Lee K J, Penton-Voak I, Rowland D, Yoshikawa S, Burt D M, Henzi S P, Castles D L, Akamatsu S, 1998 "Effects of sexual dimorphism on facial attractiveness" *Nature* **394** 884 – 887

Peskin M, Newell F N, 2004 "Familiarity breeds attraction: Effects of exposure on the attractiveness of typical and distinctive faces" *Perception* **33** 147 – 157

Rhodes G, 2006 "The evolutionary psychology of facial beauty" *Annual Review of Psychology* **57** 199 – 226

Rhodes G, Sumich A, Byatt G, 1999 "Are average facial configurations only attractive because of their symmetry?" *Psychological Science* **10** 52 – 58

Rhodes G, Zebrowitz L, Clark A, Kalick S M, Hightower A, McKay R, 2001 "Do facial averageness and symmetry signal health?" *Evolution and Human Behavior* **22** 31 – 46

Rubenstein A J, Langlois J H, Roggman L A, 2002 "What makes a face attractive and why: The role of facial awareness in defining facial beauty", in *Facial Attractiveness: Evolutionary, Cognitive, and Social Perspectives* Eds G Rhodes, L A Zebrowitz (Westport, CT: Ablex)

Russell R, 2003 "Sex, beauty, and the relative luminance of facial features" *Perception* **32** 1093 – 1107

Sadr J, Fatke B, Massay C L, Sinha P, 2002 "Aesthetic judgments of faces in degraded images" *Journal of Vision* **2** 743a (abstract)

Scheib J E, Gangestad S W, Thornhill R, 1999 "Facial attractiveness, symmetry and cues of good genes" *Proceedings of the Royal Society of London, Series B* **266** 1913 – 1917

Shimojo S, Simion C, Shimojo E, Scheier C, 2003 "Gaze bias both reflects and influences preference" *Nature Neuroscience* **6** 1317 – 1322

Summerfield C, Egner T, Greene M, Koechlin E, Mangels J, Hirsch J, 2006 "Predictive codes for forthcoming perception in the frontal cortex" *Science* **314** 1311 – 1314

Taylor R, 2002 "Spotlight on a visual language" *Nature* **415** 961

Thornhill R, Gangestad S W, 1999 "Facial attractiveness" *Trends in Cognitive Sciences* **3** 452 – 460

Valentine T, Darling S, Donnelly M, 2004 "Why are average faces attractive? The effect of view and averageness on the attractiveness of female faces" *Psychonomic Bulletin & Review* **11** 482 – 487

Van Tonder G J, Lyons M J, Ejima Y, 2002 "Visual structure of a Japanese zen garden" *Nature* **419** 359 – 360

Webster M A, Kaping D, Mizokami Y, Duhamel P, 2004 "Adaptation to natural facial categories" *Nature* **428** 557 – 561

Zebrowitz L A, Rhodes G, 2002 "Nature let a hundred flowers bloom: The multiple ways and wherefores of attractiveness", in *Facial Attractiveness: Evolutionary, Cognitive, and Social Perspectives* Eds G Rhodes, L A Zebrowitz (Westport, CT: Ablex) pp 261 – 293

Zebrowitz L A, Fellous J M, Mignault A, Andreoletti C, 2003 "Trait impressions as overgeneralized responses to adaptively significant facial qualities: Evidence from connectionist modeling" *Personality and Social Psychology Review* **7** 194 – 215

Appendix A. Facial features that were located

top	1:	Top of the head (at midline)
hrl	2:	Top of the forehead level axis symmetry
Bgn	3:	Beginning of the nose
Cno	4:	Top of the nose ball, at midline
tpn	5:	Tip nose
tip	6:	End/edge nose
ulp	7:	Middle upper lip
iul	8:	Internal edge of the upper lip
lip	9:	Middle lower lip
ill	10:	Internal edge of the lower lip
tch	11:	Tip chin
chn	12:	Bottom chin
Ll2	41:	Left external eyelid corners (beyond eye ball)
Lly	16:	Left external eyelid corners (touching eye ball)
Lcy	13:	Left pupil
Lty	17:	Left upper eyelid (touching eye ball)
Lby	18:	Left lower eyelid (touching eye ball)
Lry	15:	Left internal eyelid corners (touching eye ball)
Lr2	40:	Left internal eyelid corners (beyond eye ball)
Rl2	40:	Right internal eyelid corner (beyond eye ball)
Rly	15:	Right internal eyelid corner (touching eye ball)
Rcy	13:	Right pupil
Rty	17:	Right upper eyelid (touching eye ball)
Rby	18:	Right lower eyelid (touching eye ball)
Rry	16:	Right external eyelid corners (touching eye ball)
Rr2	41:	Right external eyelid corners (beyond eye ball)
Mdy	14:	Pupil mid-point
Rrb	19:	Right external end eyebrows
Llb	19:	Left external end eyebrows
Rlb	20:	Right internal end eyebrows
Lrb	20:	Left internal and eyebrows
Rtb	21:	Right highest eyebrow point
Ltb	21:	Left highest eyebrow point
Ren	22:	Right edge nostril
Len	22:	Left edge nostril
Rno	23:	Right center nostril
Lno	23:	Left center nostril
Rul	24:	Right end upper lip
Lul	24:	Left end upper lip
Rll	25:	Right end lower lip
Lll	25:	Left end lower lip
Rmc	26:	Right corner mouth
Lmc	26:	Left corner mouth
Rte	27:	Right top of the ear
Lte	27:	Left top of the ear
Ree	28:	Right outermost edge ear
Lee	28:	Left outermost edge ear
Rbe	29:	Right bottom ear
Lbe	29:	Left bottom ear
Rck	30:	Right most prominent point of the cheek bone
Lck	30:	Left most prominent point of the cheek bone

Appendix A (continued)

Rf2	31: Right lateral edge of the face at cheek bone level
Lf2	31: Left lateral edge of the face at cheek bone level
Rf1	34: Right intersection eye-line hair (across pupil line)
Lf1	34: Left intersection eye-line hair (across pupil line)
Rf3	35: Right intersection mouth-line, mouth middle, edge of face
Lf3	35: Left intersection mouth-line, mouth middle, edge of face
Rch	36: Right extent chin level tip of chin
Lch	36: Left extent chin level tip of chin
Rjw	37: Right edge face (at 45 deg from mouth corner)
Ljw	37: Left edge face (at 45 deg from mouth corner)
Rbb	38: Right lowest eyebrow point (over eye)
Lbb	38: Left lowest eyebrow point (over eye)
Dch	39: Double chin (if none, equal bottom chin)

Appendix B. Facial feature distances and proportions

W4	Cheekbone width
W1	Jaw width
W5	Chin width
W6	Cheekbone width
W3	Head width
E4	Eye width
E2	Interpupil distance
E3	Distance between outer eye corners
E1	Distance between inner eye corners
B1	Minimum distance between eyebrows
N4	Nose width at nostrils
N3	Nose width not including nostrils
M0	Mouth width
R1	Width of ear to bottom of lobe
C3	Chin height
C2	Height of chin to center of mouth
L0	Head height
C1	Height of chin to pupil
E5	Eye height
B2	Pupil to top of eyebrow
B4	Pupil to medial extent of eyebrow
B5	Pupil to bottom of distant brow tip
B6	Pupil to bottom of innermost brow tip
H1	Nose to hairline
N3	Nose length
M1	Height of lips
M3	Lip to nose
M2	Center of lips to nose
M4	Height of upper lip
E6	Pupil to top of ear
R0	Length of ear
S0	Mouth to jaw
D0	Double chin height
C4	Nose tip to chin

Appendix B (continued)

C5	Top of nose to tip of nose
$= \mathrm{L0}/\mathrm{W4}$	Head length
$= \mathrm{W4}/\mathrm{W1}$	Cheekbone prominence
as per Zebrowitz et al (2003)	Facial roundness
$= \mathrm{E4}/\mathrm{E1}$	Eye spacing ratio
$= \mathrm{C2}/\mathrm{M2}$	Lower face ratio
$= \mathrm{C4}/\mathrm{C5}$	Lower face/nose length
$= \mathrm{C4}/\mathrm{H1}$	Lower face/forehead
as per Penton-Voak et al (2001)	Facial symmetry
L0	Raw vertical norm
E2	Raw horizontal norm
$\{[\frac{1}{2}(\mathrm{Ll2x} - \mathrm{Llyx}) - \mathrm{Rjwy}]^2 - [\frac{1}{2}(\mathrm{Ll2y} - \mathrm{Llyy}) - \mathrm{Ljwy}]^2\}^{1/2}$	PT37 Vert Dist from pupil
$= (\mathrm{Lry} - \mathrm{Ll2}) - (\mathrm{Lr2} - \mathrm{Lly}) + (\mathrm{Lty} - \mathrm{Ll2}) - (\mathrm{Lby} - \mathrm{Lly})$	Walleyed/Cross-eyed
$= \mathrm{Lcyx} - \mathrm{Tipx}$	(Hdisp_Nose) Tip
$= \mathrm{Lcyx} - \mathrm{Bgnx}$	(Hdisp_Nose) Top
$= (\mathrm{Lcyx} - \mathrm{Lryx}) - (\mathrm{Lcyx} - \mathrm{Lr2x})$	(Hdisp_Eye) out_to_center
$= (\mathrm{Lcyx} - \mathrm{Ltyx}) - (\mathrm{Lcyx} - \mathrm{Lbyx})$	(Hdisp_Eye) in_to_center
$= (\mathrm{Lcyx} - \mathrm{Rulx}) - (\mathrm{Lcyx} - \mathrm{Lulx})$	(Hdisp_Mouth) out_to_faceEdge
$= (\mathrm{Chny} - \mathrm{Rlby}) - (\mathrm{Chny} - \mathrm{Lrby})$	Vdisp_Nose
$= (\mathrm{Chny} - \mathrm{L12y}) - (\mathrm{Chny} - \mathrm{Llyy})$	Vertical difference of outer eye corners
$= (\mathrm{Chny} - \mathrm{Ruly}) - (\mathrm{Chny} - \mathrm{Luly})$	Vertical difference of mouth corners

ISSN 0301-0066 (print) ISSN 1468-4233 (electronic)

PERCEPTION

VOLUME 36 2007

www.perceptionweb.com