Women and Men Integrate Facial Information Differently in Appraising the Beauty of a Face

Abstract: 149 words

Main Text: 3548 words

This is an accepted manuscript of an article published by Elsevier in Evolution and Human Behavior. The final version is available at http://dx.doi.org/10.1016/j.evolhumbehav.2017.07.001.

Abstract

Facial beauty plays a crucial role in social interactions, particularly in mating and reproduction. Therefore, the perceptual and cognitive mechanisms used for facial beauty assessment should be susceptible to different evolutionary and cultural pressures across genders and thus shape different observational appraising strategies. Using a novel approach, I evaluated the observers' subjective and unique importance given to specific facial attributes: eyes, nose, lips, and hair, and their spatial organization in the process of appraising the beauty of the whole face. These importance measures reveal the modulation of the integration of attributes strategy across the gender of observers and the sex of face. The degree of agreement about the beauty of the studied facial attributes was modulated across gender of observers and, for women observers, also across sex of face. Finally, I show that beauty appraisal can be mainly explained by a simple additive manner of isolated facial attributes appraisals.

Keywords

Facial beauty; Perceptual integration; Analytical processing; Configural processing

1. Introduction

The beauty of faces is influential in many aspects of social interactions in general (Dion, Berscheid, & Walster, 1972; Little, Burriss, Jones, & Roberts, 2007) and in choice of mate in particular (Buss & Barnes, 1986; Walster, Aronson, Abrahams, & Rottman, 1966). Since the publication of Darwin's theory of natural selection (1859), the variability of perceived attractiveness has been analyzed in terms of the evolved signal content of striking phenotypic features, arguing that reproduction with a more attractive partner will increase an individual's biological fitness (Andersson, 1994; Barrett, Dunbar, & Lycett, 2002; Little, Burriss, Jones, DeBruine, & Caldwell, 2008). Choosing the right mate is crucial for successful reproduction, so reliable mechanisms for such recognition are favored by evolution. As a result, evolutionary, and maybe even cultural, pressures may act differently on women and men and, as a result, shape different observational beauty appraisal strategies across male and female genders.

In order to compare beauty appraisal strategies, one has to quantify the diagnostic dimensions of facial information that human observers use to judge the beauty of a face. Throughout history, several ideal characteristics of beauty have been suggested, mainly by formulating canons of face shapes and distances between selected facial landmarks of particularly meaningful and salient locations. The ancient Greeks believed aesthetic preferences fulfil certain geometrical conditions, such as the Golden Ratio. In the renaissance period, Neoclassical Canons were considered the ideal ratios of beautiful faces (Edler, 2001; Vegter & Hage, 2000).

This is an accepted manuscript of an article published by Elsevier in Evolution and Human Behavior. The final version is available at http://dx.doi.org/10.1016/j.evolhumbehav.2017.07.001.

Over the last few decades, many studies of facial beauty have focused on three main diagnostic dimensions: averageness, symmetry and sexual dimorphism (Gangestad, Thornhill, & Yeo, 1994; Langlois & Roggman, 1990; Perrett et al., 1998). On the other hand, the role of facial parts such as eyes, nose, and mouth, and their spatial organization and inter-attribute interactions (holistic processing) is a central issue in facial recognition research, suggesting different mechanisms and brain activation with single facial parts and their combinations (Arcurio, Gold, & James, 2012; Carey & Diamond, 1977; Farah, Wilson, Drain, & Tanaka, 1998; Gold, Mundy, & Tjan, 2012; Maurer, Grand, & Mondloch, 2002; Tanaka & Farah, 1993). The common view is that the human perceptual system integrates facial information into a gestalt whole rather than processing facial features in a non-interacting manner. The composite face effect has been used in many studies to demonstrate that facial parts cannot be perceived independently and therefore interact (Young, Hellawell, & Hay, 1987; Rossion, 2013). Nevertheless, there are some examples for which information conveyed from isolated facial parts is almost optimal when summed up in an additive manner (e.g., Maloney & Dal Martello, 2006). To date, the extent to which the impression of isolated facial parts shapes the assessment of facial beauty has not been studied.

What is the contribution of facial sub-regions and their spatial organization to the assessment of the beauty of the whole face? Pointing out the beauty of specific facial attributes is common in everyday life. The place of aesthetic characteristics of some facial attributes is well demonstrated by commonly used phrases, such as 'pretty eyes' or 'beautiful hair'. This suggests that facial beauty resides at different levels within the whole face at one level and at the level of 'facial parts' attributes at sub-levels. Nevertheless, the unique contribution of such specific sub-level attributes and the way

This is an accepted manuscript of an article published by Elsevier in Evolution and Human Behavior. The final version is available at http://dx.doi.org/10.1016/j.evolhumbehav.2017.07.001.

they are integrated to make a beauty appraisal of the whole face, have not been investigated systematically and remain obscure.

Therefore, a prospective avenue for understanding the diagnostic dimensions which humans utilize to appraise facial beauty is an approach that rigorously quantifies the importance of the beauty of facial attributes, such as facial sub-regions and their spatial organization, to the beauty impression of the whole face.

Here, I address three questions about facial attributes processing for the purpose of beauty appraisal. Firstly, is the integration of facial attributes modulated by the gender of observer and the sex of face? Secondly, to what extent are the inter-subjective facial preferences modulated across facial attributes, gender of observer and sex of face? While observers may associate a similar degree of importance with certain facial attributes, they may disagree about the level of the beauty of individual attributes. A category of attributes which has a high level of agreement within a group of observers is an indication that there is a consensus, at least to some extent, about desirable specifications, such as shape or color, in that category. Such unique specifications may reflect a reliable signal of biological fitness or alternatively a social convention. Finally, to what extent is beauty appraisal based on the additive processing of facial attributes? In the current study, I quantitatively evaluate the unique contribution of specific facial attributes to the beauty appraisal of whole faces. I use these measures to investigate how the integration strategy is modulated across the gender of observers and across the sex of face. Later, I study the modulations of inter-subjective homogeneity across the gender of observers and across the sex of face. Finally, I show that the majority of the feasible variance of beauty appraisal of the whole face is explained by the appraisal of the isolated attributes I used in the current study.

This is an accepted manuscript of an article published by Elsevier in Evolution and Human Behavior. The final version is available at http://dx.doi.org/10.1016/j.evolhumbehav.2017.07.001.

The facial phenotype is derived by the biological sex; therefore throughout this paper, I classify the face stimuli by their biological sex: female or male (Enlow, 1996). However, since it is unknown which factors shape the strategy of beauty perception, biological or cultural; I have chosen to follow the common distinction used in cross-gender studies and classify the observers by the term ‘gender’: women or men.

2. Method

2.1. Observers

Sixty four observers (32 women, M=22.8, SD=2.3 years; 32 men, M=23.8, SD=2.7 years) participated in a task rating the female face. Sixty four observers (32 women, M=22.4 years, SD=1.9 years; 32 men, M=23.8 years, SD=3.2 years) participated in a task rating the male face. This sample size was determined in advance. As a data driven study utilizing a novel method, the types of effects and their expected sizes were unknown. All observers were students at the Hebrew University of Jerusalem, with normal or corrected to normal visual acuity, who participated in the experiment for course credit or monetary reward. All observers signed an informed written consent according to the institutional review board of the Hebrew University of Jerusalem.

2.2. Stimuli and apparatus

I used two sets of frontal headshot color photographs of individuals with neutral expressions: one set of 27 Caucasian females and one set of 27 Caucasian males (all models aged between 20 and 30). The faces had similar location, size, illumination, and there were no beards, moustaches, earrings, eyeglasses, makeup, or jewellery. The resolution of all images was 350×480 pixels and the models had been instructed to assume neutral expressions. Four facial fragments were cut out from the intact faces: eyes (including eyebrows), nose, mouth, and hair (including ears, seen or occluded). An additional stimulus category denoted here as ‘configuration’, was made to capture the spatial organization of the eyes, nose and lips together with facial shape elements. I denote the latter category as ‘configuration’, however this should not be confused with the identically named term sometimes used in other studies. To create the configuration stimuli, images of the whole face were converted into greyscale (to partial out the facial coloration contribution leaving only the luminance channel), then low-pass filtered with a critical band of approximately six cycles per face width (to partial out the inner facial features specification; see Goffaux, Hault, Michel, Vuong, & Rossion, 2005), and finally cropped of hair. Figure 1 illustrates the six categories of stimuli: eyes, nose, lips, hair, configuration, and the whole face. All stimuli were presented on a 17 inch LCD screen at a viewing distance of 60cm.

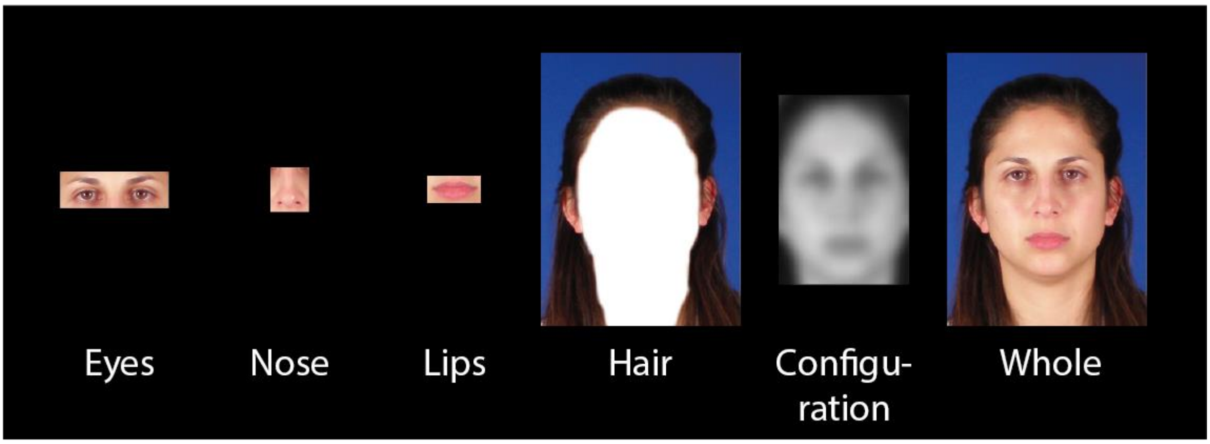

Figure 1. Stimulus categories. From left to right: eyes, nose, lips, hair (and ears), configuration, and whole face.

This is an accepted manuscript of an article published by Elsevier in Evolution and Human Behavior. The final version is available at http://dx.doi.org/10.1016/j.evolhumbehav.2017.07.001.

2.3. Procedure

Each observer participated in six different conditions, each focusing on a different category: eyes, nose, lips, hair, configuration, and whole face. The first five conditions were blocked by attribute and presented in random order of blocks and random order of individual stimuli within blocks across participants. The whole face condition was always presented as the final block in a random order of stimuli within blocks.

In each condition, pairs of images (of the same attribute and sex of face, e.g., two pairs of male noses) were presented on screen, side by side, in a random order and a random left/right juxtaposition. Participants were instructed to indicate, using a five alternative forced choice method, which of the two images they thought was more beautiful: 'the left image is much more beautiful'; 'the left image is slightly more beautiful'; 'both images are equally beautiful'; 'the right image is slightly more beautiful', and 'the right image is much more beautiful'. In most studies that address the aesthetic aspects of faces and body the term 'attractiveness' is typically used. Nevertheless, in the current study the participants were instructed to indicate the 'beauty' and not the 'attractiveness' of the face as the latter term can be interpreted also in terms of sociability and may lead to different interpretations across participants (e.g., in the case of a 'mean but beautiful' face).

3. Results

The beauty score of an individual stimulus was derived from the pairwise comparison in the following way. For each trial, if an individual stimulus was rated in a single pairwise comparison as 'much more beautiful' than the other, it got the value 2 and the other, less beautiful individual stimulus, got the value -2. In a similar way, the 'more beautiful' response yielded ratings of 1 to the more beautiful stimulus and -1 to the less beautiful stimulus. 'Equally beautiful' was evaluated as 0 for both stimuli. Figure 2A illustrates the data pre-processing stage: to obtain a unique subjective score for each individual stimuli and each observer, I averaged the responses for each observer over all comparisons in which the individual stimuli took part. To avoid heterogeneity in the use of the response scale among participants and stimulus categories, the average responses were converted to ranks over identities within each subject and each category of stimulus. This pre-processing step yielded a subjective beauty score for each individual stimulus and each observer. To measure the importance of each facial attribute to the whole face, I used the semipartial correlation between each of the attribute scores and the matching scores of the whole face (Darlington, 1990). This statistic provides some desirable properties: (i) the semipartial correlation measures the exclusive contribution of the attribute in question to the whole face appraisal whilst partialing out the rest of the facial attributes from that attribute, in other words, it measures the contribution of the specific attribute to the whole face appraisal that cannot be explained by any of the other attributes; (ii) it indicates whether the appraisal of the whole face increases or decreases with the increment of the beauty of the attribute, and (iii) it provides an intuitive interpretation of the contribution of each of the facial attributes, the square of the semipartial correlation is the increment of the explained variance of a linear model as a result of adding the attribute in question to the model.

Figure 2B depicts the computation of the distribution of importance among facial attributes for an individual observer performing judgments of a particular sex of face. For each observer, I calculated the semipartial correlation, matched by identity,

This is an accepted manuscript of an article published by Elsevier in Evolution and Human Behavior. The final version is available at http://dx.doi.org/10.1016/j.evolhumbehav.2017.07.001.

between the subjective scores of the facial attribute and the whole face. The bar graphs show the level of importance associated to each of the attributes by the individual observer for the given sex of face.

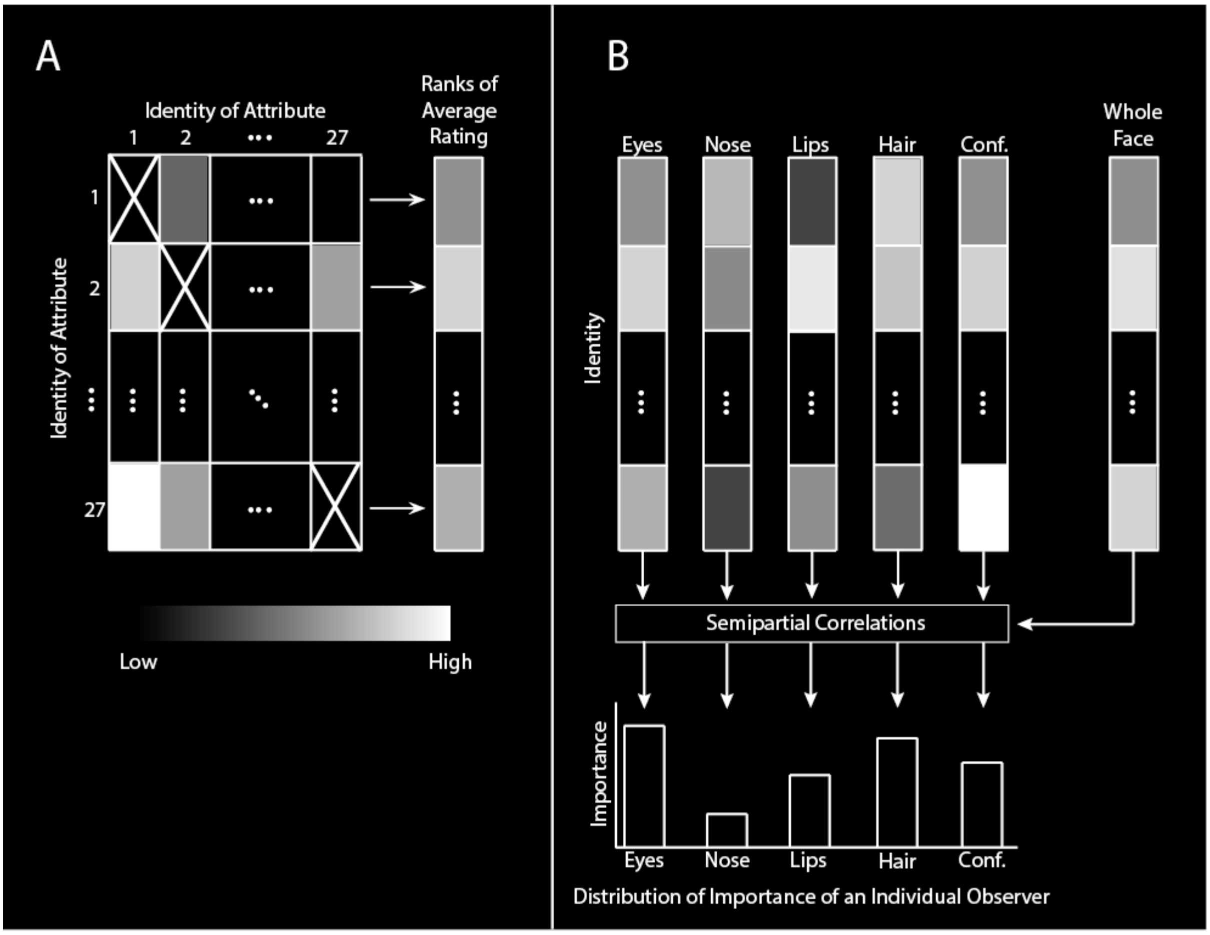

Figure 2. Illustration of data pre-processing and analysis. (A) For each observer and each stimulus category the numerical responses to pairwise comparisons between stimuli were assigned into an antisymmetric data matrix. The element D_{ij} in row i and column j is the response for the comparison between stimuli i and j ($D_{ij}>0$ means that stimulus i is more beautiful than stimulus j therefore $D_{ji}=- D_{ij}$). The level of the responses is represented by the greyscale level of rectangles. To represent the subjective score of beauty of an individual stimulus by an individual observer, I averaged all pairwise comparisons performed by the observer in question in which the stimulus took part (i.e., average along a row). The average ratings were then converted to ranks. (B) The importance of each facial attribute (from left to

This is an accepted manuscript of an article published by Elsevier in Evolution and Human Behavior. The final version is available at http://dx.doi.org/10.1016/j.evolhumbehav.2017.07.001.

right: eyes, nose, lips, hair, and configuration) to the rating of the whole face was measured by the semipartial correlation between the beauty scores of the facial attribute and the score of the whole face. This procedure yielded, for each individual observer, a vector, shown here as a bar chart, representing the distribution of importance across facial attributes.

In the following paragraphs, I analyze the following aspects of the data: (i) the modulation of the attributes integration strategy across gender of observers and across sex of face; (ii) the modulation of the degree of agreement about the beauty of the studied facial attributes across gender of observers and sex of face, and (iii) the explanation power of an additive model of the facial attributes used in the current study in terms of explained variance.

3.1. Modulation of attributes integration

Figure 3A shows the average importance of each attribute where the results are grouped into four conditions (two gender of observer x two sex of face). The height of the bars represents an average importance per attribute and condition. The error bars represent the standard error. Significant differences between conditions are represented by * (the actual numerical values are provided in Table S1 in the Supplemental Material). From now on, all statistical tests throughout this paper use a two-tailed bootstrap, N=1000 with total $p<0.05$ and simultaneous correction for multiple comparison (Mandel & Betensky, 2008).

3.1.1. Modulation across gender of observer

When judging female faces, women attached higher importance to the lips than the

This is an accepted manuscript of an article published by Elsevier in Evolution and Human Behavior. The final version is available at http://dx.doi.org/10.1016/j.evolhumbehav.2017.07.001.

men did, whilst the latter attached higher importance to the configuration ($p<0.05$ corrected). When judging male faces, women attached higher importance to the eyes than men did.

3.1.2. **Modulation across sex of face**

Women as observers attached higher importance to male eyes than to female eyes and higher importance to female lips than to male lips ($p<0.05$ corrected). Men as observers attached higher importance to female configuration than to male configuration ($p<0.05$ corrected).

3.2. Modulation of inter-subjective homogeneity

To evaluate the degree of inter-subjective homogeneity, I measured the inter-rater agreements of facial attributes and whole faces among participants. Figure 3B presents the results of these agreements, demonstrated by bar charts. The error bars represent the standard error. Significant differences between conditions are represented by * ($p<0.05$ corrected; the actual numerical values are provided in Table S2 in the Supplemental Material).

3.2.1. **Modulation across gender of observer**

When judging female faces, women held significantly higher agreement than men observers about the lips ($p<0.05$ corrected). When judging male faces, women held significantly higher agreement than men observers about the hair ($p<0.05$ corrected).

3.2.2. **Modulation across sex of face**. When judging male faces, women held higher agreements for nose and hair than the agreements they held about these attributes in female faces ($p<0.05$ corrected).

This is an accepted manuscript of an article published by Elsevier in Evolution and Human Behavior. The final version is available at http://dx.doi.org/10.1016/j.evolhumbehav.2017.07.001.

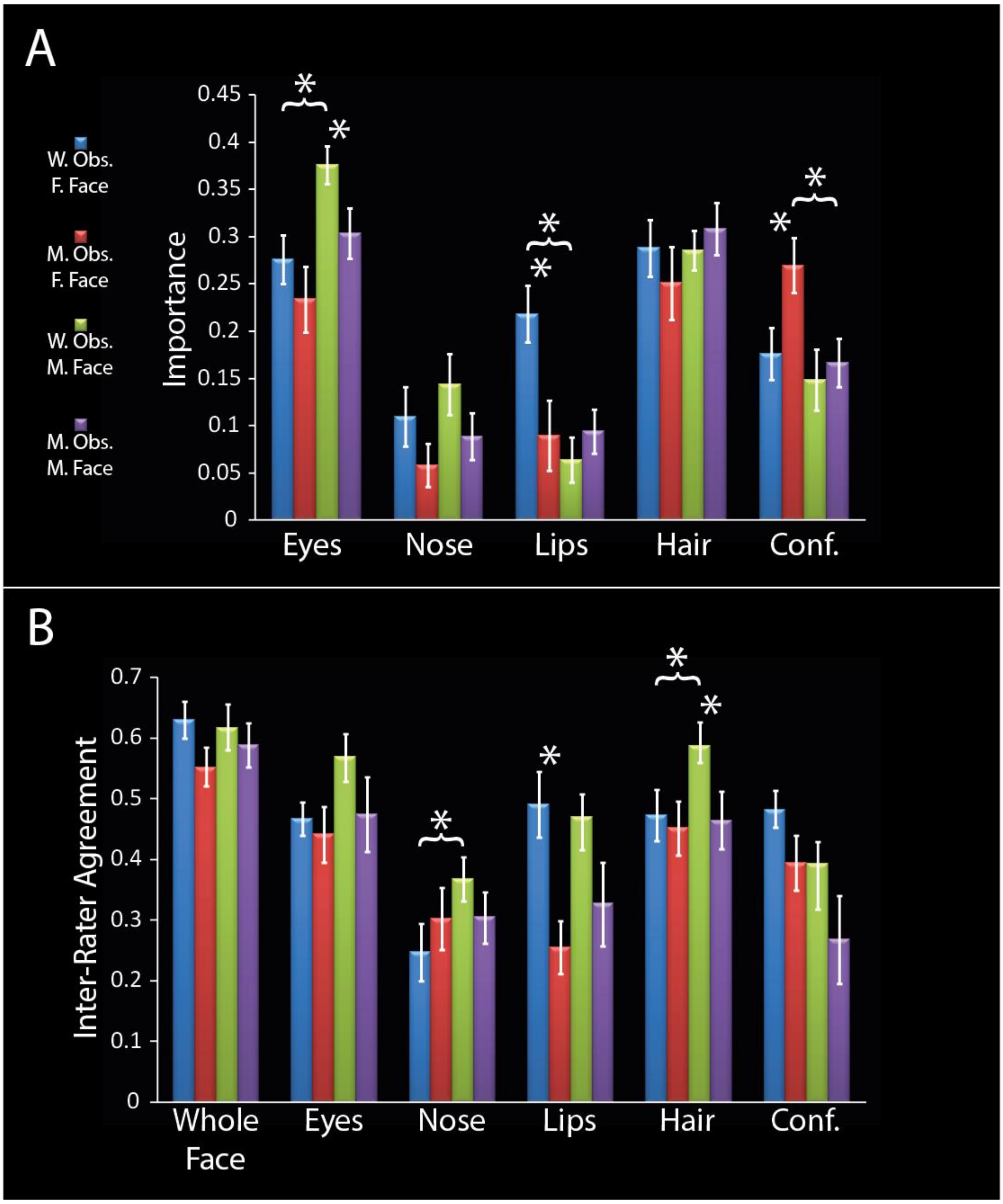

Figure 3. Results. (A) The importance of the isolated facial attributes to whole face appraisal was evaluated by semipartial correlation. The bar graph shows the average importance across observers for each facial attribute. The bar graphs are color encoded by the gender of the observer (abbreviated as 'W' or 'M' corresponding to Woman or Man, respectively) and sex of face (abbreviated as 'F' or 'M' for Female or Male, respectively). (B) Inter-rater agreements about the whole face and each of the isolated facial attributes. In both panels the error bars represent standard errors and significant differences are

This is an accepted manuscript of an article published by Elsevier in Evolution and Human Behavior. The final version is available at http://dx.doi.org/10.1016/j.evolhumbehav.2017.07.001.

indicated by * (p<0.05, bootstrap N=1000, corrected for multiple comparison).

3.3. The explanatory power of isolated facial attributes

To evaluate the total explanation power of the facial attributes to the whole face judgments, I calculated for each observer the degree of explained variance of the subjective beauty scores of the whole face, by the beauty scores of the facial attributes. To this end, I performed a linear multivariate regression, in which the subjective facial attribute scores served as the independent variables, and the whole face subjective score served as the dependent variable. The average goodness-of-fit measures over observers were as follows: women observers' appraisals of female faces R^2=0.53, women observers' appraisals of male faces R^2=0.56, men observers' appraisals of female faces R^2=0.50 and finally men observers' appraisals of male faces R^2=0.54. Importantly, the average reliability of attractiveness appraisals is known to be limited and therefore the feasible upper limit of the level of explained variance by a model of any kind is lower than R^2=1 (Oosterhof & Todorov, 2008; Willis & Todorov, 2006). Although the facial attributes used in the current study do not cover the whole face when assembled together, a simple additive model of the facial attributes appraisals still explains the majority of the feasible explained variance.

4. Discussion

In human social interaction, the beauty of the face has influential consequences for individuals and groups. The beauty of opposite-sex face is proposed to reflect, at least in part, appropriate mate choice for reproduction. Therefore it is expected that men and women should hold different strategies for beauty appraisal. In the current study

I sought to find strategy modulation in two complementary facets of beauty appraisal: (i) the importance associated by observers to certain facial attributes, and (ii) the homogeneity of inter-subjective agreements within gender about the beauty of facial attributes and whole faces.

The modulation of strategy (both association of importance and degree of subjective preference) that was found across the sex of face is not surprising. Male and female faces have different facial characteristics caused by different levels of testosterone (higher in males) and oestrogen (higher in females) and therefore different biological fitness signals (Enlow, 1996).

The modulation of strategy across the gender of observers may be due to different evolutionary pressures that shape own sex and opposite sex beauty appraisals. Another non-exclusive explanation could be different cultural pressures across genders. The modulation of the level of homogeneity of inter-subjective agreements about the beauty of facial attributes across gender of observers, suggests differences in consensus regarding prototypes of beautiful or non-beautiful facial attributes within gender. These differences may originate from evolutionary pressures that have led to different sensitivities to phenotypic signals of fitness. Alternatively, a cultural explanation is that the male and female genders have a different extent of exposure to culturally presented ideals of certain facial attributes. For example, women may have higher exposure to a specific prototype of lips as exemplified by cosmetic adverts that mainly target women.

The four linear models used in the current study explained on average the majority of the feasible variance in whole face appraisals. This, together with the fact that the facial attributes used in the current study do not cover the full face when assembled together

This is an accepted manuscript of an article published by Elsevier in Evolution and Human Behavior. The final version is available at http://dx.doi.org/10.1016/j.evolhumbehav.2017.07.001.

(i.e., full spatial frequencies and color of the cheeks and lower jaw are missing) suggests that the encoding of facial beauty at the level of isolated facial attributes provides a simple yet efficient mechanism for facial beauty processing.

Acknowledgment

This work was supported by the Louis Guttman Scholarship Fund, Israel Science Foundation and ESRC RES-060-25-0010.

Appendix A. Supplementary material

Additional supporting information may be found online.

References

Arcurio, L.R., Gold, J.M., & James, T.W. (2012). The response of face-selective cortex with single face parts and part combinations. *Neuropsychologia*, 50 (10), 2454-9.

Andersson, M. (1994). *Sexual selection*. Princeton, NJ: Princeton University Press.

Barrett, L., Dunbar, R., & Lycett, J. (2002). *Human evolutionary psychology*. Princeton, NJ: Princeton University Press.

Buss, D. M., & Barnes, M. (1986). Preferences in human mate selection. *Journal of Personality and Social Psychology*, *50*, 559–570.

Carey, S., & Diamond, R. (1977). From piecemeal to configurational representation of faces. *Science*, *195*, 312–314.

Darlington, R. B. (1990). *Regression and linear models*. New York, NY: McGraw-Hill.

Darwin, C. (1859). *On the origin of species by means of natural selection*. London: John

This is an accepted manuscript of an article published by Elsevier in Evolution and Human Behavior. The final version is available at http://dx.doi.org/10.1016/j.evolhumbehav.2017.07.001.

Murray.

Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology, 24* (3), 285–290.

Edler, R. J. (2001). Background considerations to facial aesthetics. *Journal of Orthodontics 28*(2), 159–168.

Enlow, D. H. (1996). Essential of facial growth. Philadelphia: W.B. Saunders Company.

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is “special” about face perception? *Psychological Review*, *105*, 482–498.

Gangestad, S. W., Thornhill, R., & Yeo, R.A. (1994). Facial attractiveness, developmental stability, and fluctuating asymmetry. *Ethology and Sociobiology*, *15*, 73–85.

Goffaux, V., Hault, B., Michel, C., Vuong, Q. C., & Rossion, B. (2005). The respective role of low and high spatial frequencies in supporting configural and featural processing of faces. *Perception*, *34*, 77–86.

Gold, J. M., Mundy, P. J., & Tjan, B. S. (2012). The perception of a face is no more than the sum of its parts. *Psychological Science*, *23*(4), 427–434.

Langlois, J. H., & Roggman, L. A. (1990). Attractive faces are only average. *Psychological Science*, *1*, 115–121.

Little, A. C., Burriss, R. P., Jones, B. C., & Roberts, S. C. (2007). Facial appearance affects voting decisions. *Evolution & Human Behavior*, *28*, 18–27.

Little, A. C., Burriss, R. P., Jones, B. C., DeBruine, L. M., & Caldwell, C. A. (2008). Social influence in human face preference: Men and women are influenced more for long-term than short-term attractiveness decisions. *Evolution & Human Behavior*, *29*, 140–

146.

Maloney, L. T. & Dal Martello, M. F. (2006). Kin recognition and the perceived facial similarity of children. *Journal of Vision*. 6, 1047-56.

Mandel, M., & Betensky, R. A. (2008). Simultaneous confidence intervals based on the percentile bootstrap approach. *Computational Statistics & Data Analysis, 52*, 2158–2165.

Maurer, D., Grand, R. L., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences*, *6*, 255–260.

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, *105*, 11087–11092.

Perrett, D. I., 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.

Rossion, B. (2013). The composite face illusion: A whole window into our understanding of holistic face perception, *Visual Cognition*, 21:2, 139-253.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experiment Psychology*, *46A,* 225–245.

Vegter, F., & Hage, J. J. (2000). Clinical anthropometry and canons of the face in historical perspective. *Plastic and Reconstructive Surgery*, *106*, 1090.

Walster, E., Aronson, V., Abrahams, D., & Rottman, L. (1966). Importance of physical attractiveness in dating behaviour. *Journal of Personality and Social Psychology*, *4*, 508–516.

Willis, J., & Todorov, A. (2006). First impressions - Making up your mind after a 100-Ms exposure to a face. *Psychological Science*, *17*(7), 592–598.

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception*, 16(6):747-59.