Journal of Evolutionary Psychology, 8(2010)1, 23–55
DOI: 10.1556/JEP.8.2010.1.4

DO THEY KNOW WHAT THEY LIKE? INTRA-INDIVIDUAL VARIATION OF FEMALE FACIAL PREFERENCES

K. KOŚCIŃSKI*

Institute of Anthropology, Faculty of Biology, Adam Mickiewicz University

Abstract. Although patterns of facial preferences have been examined in many studies, the intra-individual variation in assessments of facial attractiveness has been addressed far less frequently. Here we investigated the stability of women's evaluations of real and digitally manipulated male faces at monthly and yearly intervals, and we sought determinants of the stability. The following results were obtained: (1) The stability of attractiveness assessments over a year was not lower than over a month. This suggests that between-session disparity is predominantly from factors fluctuating over time with no directional trends. (2) The breakdown of a relationship resulted in an increase in the preference for friendly looking faces. (3) The change of mood positively correlated with the change of preference for good-genes facial cues, but only in paired women. This suggests the influence of mood changes on women's readiness to cheat their long-term partner in order to "gain" good genes. (4) Women that were relatively open to casual sex manifested relatively high stability of preferences for sexy looking faces. (5) The mean between-session self-correlation of attractiveness assessments amounted to 0.78 (R^2 = 61%), which is unexpectedly low given generally very efficient face processing in humans.

Keywords: facial attractiveness, human face, stability of preferences, test–retest correlation

INTRODUCTION

Although mate choice in non-human animals has always been conceived in biological terms, the partner choice and the perception of physical attractiveness in humans are traditionally regarded as being of cultural origin. Only in the 1970s did biologi-

* Address for correspondence: KRZYSZTOF KOŚCIŃSKI, Institute of Anthropology, Faculty of Biology, Adam Mickiewicz University, Umultowska 89, 61-614 Poznan, Poland; Fax: +48 061 829 5730, E-mail: koscinski@amu.edu.pl

cally oriented explanations of these human behaviors gain popularity. Nowadays, the perception of attractiveness is commonly regarded as a biological adaptation aimed at finding a partner of high biological quality (SYMONS 1995). One frequently cited argument for this claim is the universality of preferences, at least for faces (e.g. PENTON-VOAK and PERRETT 2000). Although mean evaluations of facial attractiveness calculated for different human populations, and even races, converge well with one another, indeed ($r > 0.85$, see meta-analysis in LANGLOIS et al. 2000), the agreement between two individuals, even belonging to the same population, is decisively lower ($r \approx 0.5$, see meta-analyses: FEINGOLD 1992; LANGLOIS et al. 2000). What is more, DIENER, WOLSIC and FUJITA (1995) found that inter-individual agreement is even lower when facial clips are evaluated instead of facial images ($r \approx 0.3$). If facial preferences directed an observer toward high-quality individuals, all individuals would manifest very similar preferences. They should unanimously assess a face more highly, the higher the biological quality suggested by its appearance. However, raters frequently disagree with one another substantially, and one face may be evaluated very highly by some raters, and very lowly by others (HÖNEKOPP 2006).

The mate value is, however, a more essential criterion of mate choice than the biological quality. The mate value of individual *A* for individual *B* is the benefit (or loss) for the reproductive success of individual *B* when mating with individual *A* (JONES 1996). The mate value of a specimen may vary for different opposite-sex individuals, and depend on their genetic quality, biological condition, physiological state, ecological circumstances, life experience, as well as the genetic and morphological similarity to the specimen (JENNIONS and PETRIE 1997; WIDEMO and SHÆTHER 1999). These factors also impact the perception of facial attractiveness in humans (KOŚCIŃSKI 2008), and patterns of these relationships seem to be beneficial for judges (e.g. PENTON-VOAK and PERRETT 2000; KOŚCIŃSKI 2008). Facial preferences can thus be regarded as biological adaptations, in spite of being individually diversified.

One more factor contributing to the inter-individual variation of preferences is intra-individual variation, which reflects the fact that repeat attractiveness assessments by a person may not be the same as their original assessments of these same faces. If attractiveness evaluations are taken only once from each judge, it is impossible to estimate what proportion of their inter-individual variation is attributable to the individual instability of assessments (WAGNER 1998). Theoretically, it is even possible that all individuals have identical preferences, but, in an examination, give different assessments because of random factors (then all inter-individual variation would derive from the individual instability of preferences). The instability of preferences may be strictly random, and then it is maladaptive, because it may lead to the choice of a partner of relatively low mate value. Alternatively, changes of preferences with time may be adaptive, if they are adequate to changes of the individual's internal state and their ecological situation.

There is a scarcity of data on the stability of human preferences for physical features. In several studies, the mean correlation coefficient between two assessments of a set of faces was calculated. The values obtained were from 0.71 (MAPLE et al. 2005, both assessments at one session), through 0.74 (HÖNEKOPP 2006, one-week interval) and 0.78 (BRONSTAD and RUSSELL 2007, both assessments at one session), to 0.82 (HOWELLS and SHAW 1985, five-week interval) and 0.84 (FAURE, RIEFFE and MALTHA 2002, two-week interval). No relationship between the interval length and the obtained correlation is perceptible, so differences among these estimates probably result from examining various groups of judges, using photos of different people, and from some methodological differences (e.g. HÖNEKOPP and BRONSTAD and RUSSELL applied a numeric scale, MAPLE and co-authors, and HOWELLS and SHAW – an analog scale, and FAURE and co-authors used both scales as well as the method of face ranking).

No study inquired how the preference stability depended on the length of between-session intervals, and what factors impacted the level of this stability (and consequently, whether some people manifested more stable preferences than others did). Only ROBERTS et al. (2008) found that the intra-individual correlation of male sweat scent assessments was higher in women not using hormonal contraceptives than in women who started to take the contraceptives between sessions; the authors suggested then that hormones present in the contraceptives interfere with the scent perception related to mate choice. There is also a scarcity of research on the possible impact of changes in specific factors on physical preferences. Only the menstrual cycle phase's effects on facial, bodily, scent and voice preferences have been extensively investigated (see reviews: GANGESTAD, THORNHILL and GARVER-APGAR 2005; JONES et al. 2008). Experimental studies demonstrated that the perception of facial attractiveness can be modified by previously watched faces (RHODES et al. 2003, 2005; LITTLE, DEBRUINE and JONES 2005; LITTLE and MANNION 2006; JONES and DEBRUINE 2008; RHODES, LOUW and EVANGELISTA 2009), attendant sounds (JONES et al. 2007b) or smells (TODRANK et al. 1995), an inhalation of pheromones (THORNE et al. 2002), hormones (THEODORIDOU et al. 2009) or nicotine (ATTWOOD, PENTON-VOAK and MUNAFÒ 2009), the consumption of alcohol (PARKER et al. 2008), and by social factors (PETTIJOHN and TESSER 2005; JONES et al. 2007a). However, the stimuli applied in experimental studies were artificially prepared, and their character and intensity frequently differed from what is normally encountered in real life. These studies are not then very informative about the natural variation of individual preferences. Non-experimental research on the intra-individual variation of preferences is therefore needed.

The stability of mating preferences has been repeatedly studied in non-human animals, e.g. in crickets (WAGNER, MURRAY and CADE 1995), grasshoppers (REINHOLD, REINHOLD and JACOBY 2002), guppies (KODRIC-BROWN and NICOLETTO 1997; BROOKS and ENDLER 2001), sticklebacks (BAKKER 1993), anura (JENNIONS, BECKWELL and PASSMORE 1995; HOWARD and YOUNG 1998), swallows (MØLLER 1994; WHITTINGHAM, DUNN and STAPLETON 2006) and zebra finches

(FORSTMEIER and BIRKHEAD 2004; HOLVECK and RIEBEL 2007). However, because of a specific aim in most of these studies (i.e. to estimate the possibility of further evolution of preferences), they applied a special method to measure the stability, namely repeatability. Repeatability indicates the degree of between-session intra-individual similarity of preferences in relation to inter-individual similarity of preferences (more precisely, it is the quotient of inter-individual variation of preferences by total variation, FALCONER 1970). Except for specific situations (DOHM 2002), the repeatability of a trait is not less than its heritability (FALCONER 1970), so it sets the upper estimation of the heritability, and thereby, of the possibility of further evolution of the trait. However, since the heritability may be much smaller than the repeatability, even a huge repeatability value does not prove the genetic foundation of the trait.

Because the repeatability reflects the *relation* between the intra-individual and inter-individual variation, a low value of the repeatability may result from a substantial individual instability or from high inter-individual similarity of preferences (or from both of them, WIDEMO and SÆTHER 1999). For example, BROOKS and ENDLER (2001) in their study on guppies, found that females manifested low preference repeatability for traits which are important for male attractiveness (e.g. tail size), and high repeatability for traits of little importance (e.g. black area). The reason was that females had similar preferences for the former traits and divergent ones for the latter.

Summing up, the analysis of correlations between original and repeat attractiveness assessments by a judge seems to be a more appropriate method for investigating the preference stability and its determinants than the calculation of repeatability. However, when an individual's strength of preference for a trait is studied, then the researcher possesses one preference value per session, and the calculation of the self-correlation is not possible. The measure of the preference stability may then be the repeatability or the absolute value of between-session change of the preference strength.

PRESENT STUDY

The aim of the present study was to estimate the stability of attractiveness perception of young men's faces by young women at a monthly and yearly interval, as well as to identify determinants of this stability. Although many experimental studies (see Introduction) have demonstrated that facial preferences can change in a few minutes, we felt that in natural conditions their changes were not so rapid. If we examined women at an interval of several days, the effects of menstrual cycle phase would be confounded in their assessments. As these effects were not the focus of our interest, the monthly and yearly intervals were chosen. Many individual characteristics, such as partnership status, mating preferences, use of hormonal contraceptives, and socioeconomic conditions, may change in the period of a month or year, and consequently affect facial preferences.

Both real (e.g. CUNNINGHAM, BARBEE and PIKE 1990) and digitally manipulated faces (e.g. PERRETT et al. 1998) have been repeatedly used in studies on facial preferences, and each possesses some advantages and weaknesses. The use of real faces is ecologically more relevant than of manipulated faces, i.e. they better correspond to the faces seen in real life. However, in digitally manipulated faces some features may be altered by a researcher, while the other kept fixed. This allows the role of definite features in the perception of facial attractiveness to be tested. For these reasons, both real and digitally manipulated faces were shown to the participants in the present study. Digital faces differed in some sexually dimorphic traits and were judged in contexts of short-term, long-term and friendly relationships.

Apart from attractiveness assessment, real faces were also evaluated by independent groups of judges in respect to the skin healthiness, mouth positivity, and the appropriateness for a short-term, a long-term and a friendly relationship. Skin healthiness and mouth positivity are cues to an individual's biological quality and good character, respectively, which are vital criteria in mate choice (BUSS 1999). These facial features also proved to impact attractiveness evaluations (RHODES, ROBERTS and SIMMONS 1999; KEATING and DOYLE 2002; JONES et al. 2004). Analogously, the type of a prospective relationship, short-term, long-term or friendly, effects the criteria applied to evaluate an individual and the attractiveness of their face (BUSS and SCHMITT 1993; JOHNSTON et al. 2001; BURT et al. 2007; DEBRUINE et al. 2008). We were then able to estimate the stability of attractiveness perception of a realistic set of faces (real faces), as well as the stability of preferences for objective (manipulated faces) and subjective (real faces) facial features. In each session, after facial assessments, participants filled in a questionnaire. The data obtained was helpful in seeking individual characteristics related to the magnitude of preference stability and characteristics which between-session changes alter the preference pattern. Also, the comparison between monthly and yearly groups may be helpful in identifying causes of assessment changes.

The following hypotheses on the stability of facial assessments may be formulated:

1. The mean intra-individual correlation of attractiveness assessments at a monthly interval is expected to approximate 0.75 – like in other studies examining such correlations at intervals not longer than two weeks (see Introduction).

2. The mean intra-individual correlation of attractiveness assessments at a yearly interval is expected to be lower than at a monthly interval, since the effects of various modifiers of one's taste may accumulate more over a longer period. The decrease in test–retest correlations with the increase in the between-session interval is well documented in psychological questionnaire studies (FERGUSON and TAKANE 1989). On the other hand, the yearly correlation of assessments should not be lower than the mean *inter*-individual correlation, which is about 0.5 (see Introduction) and a similar value is expected in the current sample.

3. HÖNEKOPP (2006) analyzed the inter-individual variation in attractiveness judgments and concluded that the face of almost every person is very attractive to

some observers. However, it is not known whether judges highly appreciating a face commonly regarded as unattractive would repeat their high ratings at the second examination. They might not, if the stability of their facial preferences is low. Therefore, the hypothesis that the stability of attractiveness assessments by judges of atypical taste is comparable to that displayed by judges of typical taste should be tested.

4. One may hypothesize that some faces may differ from one another only in attractiveness-irrelevant features and therefore the order of their ranks in an evaluation would be random. This possibility does not preclude, however, the criteria of facial preferences (such as healthy, sexy, or satisfied appearance of a face) being highly stable. The high stability of preferences for facial features together with relatively low stability of assessments of individual faces would support the hypothesis on the substantial attractiveness-irrelevant variation of facial appearance.

5. The long-term relationship is the kind of relationship which women seek the most and which has the greatest consequences for their reproductive success, and thereby for their evolutionary fitness (BUSS and SCHMITT 1993). We hypothesize then that women would be more stable when judging faces in the long-term relationship context in comparison to the short-term or friendly relationship. However, the effect may be absent, or even inverted, in women open to casual sex, because short-term relationships contribute substantially to their biological fitness (BUSS and SCHMITT 1993).

6. Being in a stable relationship influences the perception of attractiveness. For example, paired women prefer more masculinized male faces than single women (PENTON-VOAK et al. 1999; LITTLE et al. 2002). The preference of paired women for good-genes cues (such as masculine faces in men) is comprehended as a manifestation of their tendency to take advantage of their stable partners' care for extra-pair offspring (PENTON-VOAK et al. 1999). One may hypothesize that the formation of a new relationship or the breakdown of the existing one results in some changes in facial preference criteria. For example, when a new relationship becomes stable, some women may start to seek an affair with men possessing cues to good genes, and therefore they would prefer sexy and healthy looking male faces. The breakdown of a relationship may in turn make women feel depressed and therefore prefer friendly looking faces.

7. The mood of a person is known to influence their preferences for individuals of the opposite-sex. WONG and CUNNINGHAM (1990) demonstrated that men in an experimentally induced positive mood displayed a relatively strong preference for sexy, though emotionally cool women, to warm though less sexy ones. As regards the present study, one may expect that a woman would prefer facial cues to high biological quality more strongly, and cues to good character less strongly, in that session in which she is in a better mood.

METHODS

Participants

Two groups of young women (aged 18.4–28.9 years, $M = 21.8$) participated in this study, one group was examined at a one month interval, and the other at a one year interval. The monthly group initially comprised 96 women (Session 1 was conducted in Dec-2007 and Mar-2008), and 91 of them participated in the second session (Jan-2008 and Apr-2008). These women were students of a university in Toruń (Poland). The yearly group contained 103 women in the first session (Feb–Mar 2007) and 87 in the second one (Feb–May 2008, a few delays were caused by problems in contacting some women). They were students of a university in Poznań (Poland). Participants were recruited in student hotels and lecture buildings.

In their first sessions, both groups did not differ significantly in age (M_{MONTH} = 21.7yrs, M_{YEAR} = 21.9yrs, $p = 0.47$), the declared income, sociosexual orientation, self-rated attractiveness, requirements for a partner, and the mood (see the questionnaire description below, all ps > 0.05). During the second session, the yearly group was, obviously, a year older than the monthly group (22.7 vs. 21.8, $t_{176} = 4.84$, $p < 0.001$), which is a consequence of the study design. This group also declared higher income and higher requirements for a partner and better mood (all ps < 0.02). These changes could derive, among other things, from having ended their college and started full-time professional life. Both groups agreed strongly in their judgments of facial attractiveness ($r = 0.98$ and 0.97 in the first and the second session, respectively). Therefore, the two groups of participants may be regarded as comparable.

Several independent groups of young (20–21 years) women evaluated real faces for features other than attractiveness: skin healthiness ($n = 34$), mouth positivity ($n = 15$), appropriateness for a short-term relationship ($n = 40$), for a long-term relationship ($n = 40$), and for a friend ($n = 40$).

Stimuli

Two kinds of stimuli were used: real faces and digitally manipulated ones.

The set of real faces contained frontal photos of 30 clean shaved men (aged 19–25 years, European origin). They were photographed from a distance of three meters, without glasses or facial hair, and had a neutral expression. Such extra-facial elements like scalp hair, ears, or clothing may potentially impact the perceived facial attractiveness, while the interest of the present study focused on the attractiveness of faces themselves. For this reason, a white mask was applied to each photo so to hide all extra-facial elements. For the purposes of evaluation of skin healthiness, three rectangular fragments were extracted from each image, one from the forehead and one from each cheek. These regions were chosen for having no

cues to facial geometry. For the purposes of the mouth positivity judgments, a rectangular region containing the mouth was extracted from each photo.

Digital facial modifications were accomplished with the warping and morphing methods in software developed by the author. These methods were described, e.g. in ROWLAND and PERRETT (1995). Four facial series were produced.

Series 1. Firstly, a face with typical male proportions was manufactured by averaging the locations of 176 landmarks across images of 39 male faces. Color information (in RGB mode) of corresponding pixels was also averaged in order to make the skin texture look smooth. Finally, the face was made symmetric geometrically (with warping) and in terms of color (by averaging the colors of pixels in the left and right side of the face). Analogously, a female prototype face was manufactured from images of 56 women. Next, the shape of the prototype male face was warped toward the female prototype (feminization) or in the opposite direction (masculinization). Six faces were produced with the following levels of masculinity (%): –70, –42, –14, +14, +42, +70 (where 0% depicts typical male proportions, and –100% depicts typical female proportions). These six faces constituted Series 1 (*Fig. 1a*).

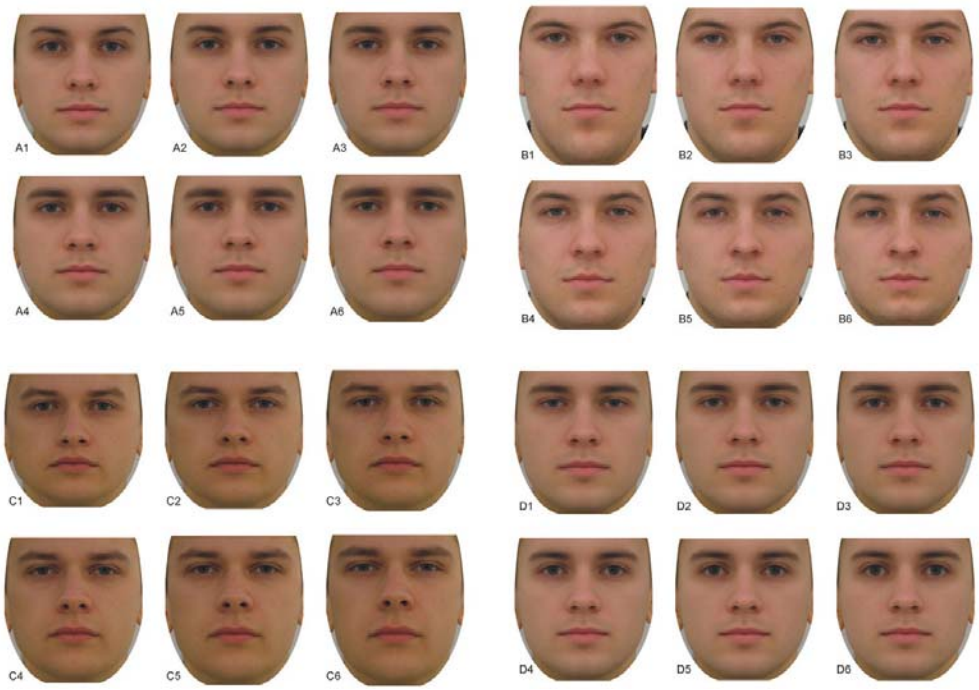

Figure 1. Four six-face series of digitally manipulated faces. Faces within a series differ in objective masculinity (series A), perceived sexiness vs. friendliness (series B – increasing friendliness, series C – increasing sexiness), and height of eye slits (series D). See the text for details

Series 2. On the basis of sexiness and friendliness judgments of 30 real male faces (see below), we chose one face which was perceived as highly sexy but not friendly, and one face of the opposite attributes. Warping the first face toward the second face produced six faces of 0, 20, 40, 60, 80, 100%. All faces had identical skin, being the average of the skins of the two basic faces (***Fig. 1b***).

Series 3 was manufactured analogously to Series 2, with two differences: (i) the two basic faces were different than those in Series 2, (ii) the highly sexy face was placed in 6th position, while in Series 2 it was in the 1st position (***Fig. 1c***).

Series 4. In the prototype male face, the height of the eye slits was changed by –20, –10, 0, +10, +20, +30% (***Fig. 1d***). All manipulated features (i.e. the masculinity, the perceived sexiness and friendliness, as well as eye height) are known determinants of facial attractiveness (e.g. CUNNINGHAM, BARBEE and PIKE 1990; KEATING and DOYLE 2002; RHODES 2006).

Images of real and manipulated faces were printed in color on glossy paper (330 DPI, 7×7 cm).

Questionnaire

Several questionnaire items concerned demographic and economic variables: the date of birth, the population of the place of origin, the total monthly income, and a 4-point assessment of the subject's own financial condition. Other questions inquired about the average length and the regularity (in days) of the menstrual cycle, the time elapsed from the last menses, and the use of hormonal contraceptives. Participants were asked whether they were in a relationship, and if so, for how long, and how committed they were to the relationship (on a 7-point scale). They also evaluated their own physical attractiveness on a special axis: They marked a point indicating the percentage of female peers being less attractive than themselves, and the percent was regarded as the self-assessment of attractiveness.

The openness to casual sex was estimated with four items selected from the 11-item Sociosexual Orientation Inventory developed by SIMPSON and GANGESTAD (1991): The number of sexual partners in the last 24 months, the frequency of sexual thoughts, and the degree of agreement with statements "Sex without love is OK", and "I would have to be closely attached to someone (both emotionally and psychologically) before I could feel comfortable and fully enjoy having sex with him or her" (the last three items were answered on 7-point scales). A factor analysis of these four items revealed one factor which is further referred to as a sociosexual orientation index (SO). SO values for the first and the second session were then averaged yielding the SO index for each woman (the Cronbach's alpha for this index was 0.80).

Finally, a short mood scale has been developed on the basis of the Depression Checklist by David Burns (BURNS 1999). Participants declared their happiness (1 – "I am very sad", 5 – "I am happy"), general self-esteem (1 – "I hate myself", 5 – "I appreciate myself very highly"), vitality (1 – "I am too exhausted to do anything", 5

– "I have very much energy"), and social life (1 – "I am uncomfortable around others and spend all of my time alone", 5 – "I have a very good social life"). The sum of these four answers (theoretically from 4 to 20) served as a mood index (the Cronbach's alphas for this index were 0.67 and 0.66 in the first and the second session, respectively).

Procedure

Each session of the monthly group had three stages: (1) ranking a series of real faces according to their perceived attractiveness, (2) choosing from a series of digitally manipulated faces the one which looked most appropriate for a lover, a spouse, or a friend, (3) filling in the questionnaire. The examination of the yearly group was identical, except that no digitally manipulated faces were presented (the author did not possess them when the group started the first session).

All 30 faces were preliminarily ranked by four young women (other than those in the main attractiveness ratings), which yielded approximate estimates of their attractiveness. Then, the set of 30 faces was divided into three 10-face series of similar distribution of attractiveness (i.e. each series contained some attractive, moderately attractive, and unattractive faces). Three different divisions of this sort were conducted, producing three sets of faces, each comprising three 10-face series. Thanks to this methodology, all series were standardized with regard to attractiveness, and the distribution of extraneous facial features were balanced across the sets of faces.

The monthly and yearly group of judges evaluated attractiveness in the following way. Each woman saw one of three stimuli sets, and the series order within the set was balanced between judges. Ten photos (i.e. one series) were taken from an envelope and laid out in front of the judge. The judge sorted the photos according to their perceived attractiveness. After completing this task, another series was laid out for evaluation, and the experimenter wrote down the sequence of the series just arranged (the photos were numbered on the back). In this way, all three facial series were sorted one by one, thus each participant assessed all 30 faces. The ranking method was chosen instead of a numerical scale, because the latter was possibly threatened with two risks. The first risk was the exposure effect: if a numerical scale had been applied, some women might have given higher attractiveness ratings at the second session than at the first session just because of being familiarized with the faces (ZAJONC 2001; RHODES et al. 2005). The second threat were possible effects of inter- and intra-individual variation in mating motivations: high levels of motivation might have made attractiveness ratings generally higher and more diversified (see BROOKS and ENDLER 2001). As opposed to numerical evaluations, the ranking method ensures the constancy of the mean and variation of assessments across all facial series, which minimizes the above threats.

The same procedure was applied for evaluations of the suitability for a short-term relationship, long-term relationship, and for friendship. Notions of short- and

long-term relationships were explicated in a way similar to that in PENTON-VOAK et al. (2003). Skin healthiness and mouth expression were assessed on a computer monitor. Skin healthiness was rated on a 5-point scale on the basis of three cuttings from the forehead and cheeks regions. Mouth positivity was rated on a scale from one ("distinct discontentment – sadness or anger") to five ("distinct contentment"). Evaluations of these five features had good reliabilities (all Cronbach's alphas ≥ 0.88), so they were averaged across all raters, yielding estimates of the skin healthiness, mouth positivity, and the suitability for a short-term, long-term and friendly relationship. For brevity, the last three characteristics will further be referred to as the lover index, the spouse index and the friend index, respectively.

The digitally manipulated faces were judged in the following way. The first series of six faces was presented (in the fixed order from version 1 to 6) in front of a judge, and the judge was asked to write down the number of the version depicting the most suitable-looking man for a short-term, long-term, and friendly relationship. Next, Series no. 2, 3 and 4 were judged in the same way. After completing this task, a participant was given the questionnaire to be filled out. At the end of the first session, phone numbers were taken from participants so as to arrange to meet at the second session.

Many studies reported the influence of the menstrual cycle phase on female preferences (see GANGESTAD, THORNHILL and GARVER-APGAR 2005). Effects of the menstrual cycle were not the focus of the present study, so we needed women to be in the same phase of their cycles in both sessions. Therefore, the planned interval between sessions in the monthly study was equal to the cycle duration declared in the first session. The actual interval differed from the planned one by no more than three days in 91% of participants. In the yearly study, however, for organizational reasons, the second sessions were arranged irrespective of the cycle phase of participants.

Analysis

Most biological and social traits are normally or roughly normally distributed, and so presumably is facial attractiveness. The rank values of facial attractiveness (from 1 to 10) collected from raters were thus transformed into standard normal values. The formula applied was $\Phi^{-1}[(rank - 3/8) / (10 + 1/4)]$, where Φ^{-1} is the inverse standard normal cumulative distribution function (BLOM 1958). The resultant values were multiplied by –1, so that the ranking number 1 (indicating the *most* attractive face) took the greatest normal value. All statistical analysis was conducted on these values.

The stability of preferences was estimated with several measures. Firstly, it was the between-session *self-correlation* of an individual's assessments, i.e. the correlation between assessments of 30 faces in the first session and their respective assessments in the second session. Secondly, the *strength of preference* for skin healthiness, mouth positivity, lover index, spouse index, and friend index was de-

termined for each judge and in each session. The strength of preference for a facial feature was calculated as the correlation (across 30 faces) between the mean evaluation of the feature by independent judges and the attractiveness assessments by a given judge. Two measures of the stability of a preference strength were applied: (1) the absolute value of the between-session *difference* in the strength of preference for a facial feature, and (2) the *repeatability* of the strength of preference for each facial feature (which approximates the correlation coefficient between the strength in two sessions, DOHM 2002). The repeatability was calculated according to LESSELLS and BOAG (1987), and its standard error according to BECKER (1984), while the repeated variable in ANOVA was the session. Analogously, the measures of the stability of the choice of digitally manipulated faces were as follows: (1) the absolute value of between-session differences in the choice, and (2) the repeatability of facial choices (which approximates the correlation coefficient between choices in two sessions).

The repeatability was also calculated in order to estimate *inter-individual agreement of tastes* – the repeated variable in ANOVA was then the judge. The value obtained indicates to what extent one judge "repeats" the judgments of another and approximates the mean correlation between attractiveness assessments by two random judges. Finally, the *preference typicality* of each judge was determined. It was computed as the correlation between 60 (30 faces × 2 sessions) facial assessments by the judge and the corresponding mean assessments by all judges.

Many statistical analyses were carried out on variables being correlation coefficients (i.e. the self-correlation, strengths of preferences for facial features, and the preference typicality). These variables were then Fisher-transformed before an analysis, to make their distributions more normal (SILVER and DUNLAP 1987). In order to determine the mean value for such a variable, the arithmetical mean was calculated from Fisher-transformed values of the variable and the mean value obtained underwent the inverse Fisher transformation (SILVER and DUNLAP 1987).

All the variables analyzed were at least approximately normally distributed (as assessed with Q-Q plots), so parametric tests known to be resistant to moderate normality violations (*t*-test, ANOVA) were applied (FERGUSON and TAKANE 1989). Alternative calculations with non-parametric tests yielded very similar results. However, the *t*-test is not resistant to normality violations when samples are small (FERGUSON and TAKANE 1989), so the Mann-Whitney test (henceforth the M-W test) was used in such situations (i.e. in the case of participants whose relationship originated or broke down between sessions, or who started or stopped using contraceptives). The popular Pearson's correlation coefficient is sensitive to outliers (FERGUSON and TAKANE 1989), so associations between two variables were tested with the Spearman's rank correlation coefficient.

RESULTS

Inter-individual Agreement

Inter-individual repeatability of attractiveness assessments were as follows: the monthly group, Session 1: $r = 0.491 \pm 0.093$, Session 2: $r = 0.427 \pm 0.098$; the yearly group, Session 1: $r = 0.501 \pm 0.092$, Session 2: $r = 0.502 \pm 0.092$ (t-test did not differentiate between these estimates). This means that a random pair of judges share about 25% (i.e. $0.5^2 \times 100\%$) of total variation in attractiveness assessments. The mean preference typicality equaled 0.70 in the monthly group and 0.73 in the yearly one. BRONSTAD and RUSSELL (2007) referred to the square of the parameter as permeation coefficient (of an individual's taste by the group standard) and estimated its value at 0.42. In the present study, these values were somewhat higher: 0.49 (0.70^2) in the monthly group and 0.53 (0.73^2) in the yearly one. Although three persons in the monthly group and two in the yearly one had a preference typicality even less than 0.3, the Grubbs' test did not regard them as outliers ($G = 3.05$, $p = 0.81$), so they were not omitted in subsequent analysis.

Self-correlation

The mean self-correlation of attractiveness assessments was $r = 0.772 \pm 0.040$ in the monthly group, and $r = 0.791 \pm 0.033$ in the yearly one. The difference between them was not significant ($t_{176} = 0.92$, $p = 0.36$). Two women turned out to be outliers according to the Grubbs' test (one in each group, $G = 4.49$, $p = 0.001$). They performed nearly identical face rankings in both sessions, so they will be excluded from further analysis. This exclusion did not noticeably alter the mean self-correlation ($M_{\mathrm{MONTH}} = 0.765 \pm 0.037$, $M_{\mathrm{YEAR}} = 0.786 \pm 0.032$, $t_{174} = 1.13$, $p = 0.26$). Therefore, the stability of preference at a yearly interval was not lower than at a monthly interval. In both groups, the mean self-correlation was about 0.78, which means that two assessments of facial attractiveness by a woman, made at a long interval, share only about 61% (i.e. $0.78^2 \times 100\%$) of variation.

The preference typicality in the first session correlated with the preference typicality in the second session ($n = 176$, $R = 0.57$, $p < 0.001$), which means that those who manifested more typical preferences in the first session were also more typical at the second session. The self-correlation was positively associated with the preference typicality in the first ($R = 0.46$, $p < 0.001$) and second ($R = 0.58$, $p < 0.001$) session. This suggests that some people have less precise apparatus for attractiveness perception, which entails both the low preference typicality and low self-correlation. A scatter plot of the self-correlation against the preference typicality confirmed that most people of low preference typicality had relatively low self-correlation (*Fig. 2*). Therefore, the low preference typicality usually resulted from an instability of taste rather than from possessing an individual, atypical, but stable taste.

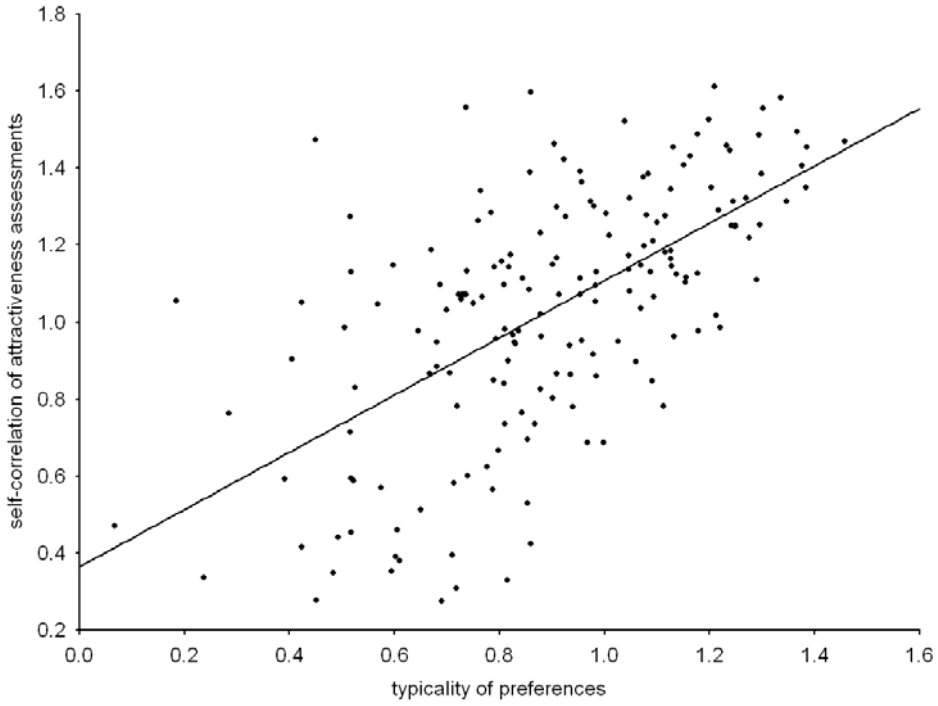

Figure 2. The typicality of preferences (the mean from two sessions) and the self-correlation of attractiveness assessments. Both variables have been Fisher-transformed ($n = 176$, $r = 0.61$, $p < 0.001$)

The Stability of Atypical Preferences

The stability of atypical preferences needs to be tested more directly, so we checked whether people who in one session strongly preferred faces commonly regarded as unattractive would also prefer them in the other session. Firstly, a group of the 10 clearly least attractive faces were identified (i.e. faces with ranks from 21 to 30). While the face ranked 20 had a mean attractiveness of –0.15, the attractiveness of the face ranked 21 equaled only –0.40. Next, we identified all cases where one of these faces was put in the top three within its ten-face series at the first session, and checked how highly the face was assessed at the second session. If all cases of high evaluations of these faces were a matter of chance, then their second-session assessments would be uniformly distributed. Alternatively, if preferences for these faces reflected an atypical taste, then the faces would also be highly appreciated in the second session.

In the first session, there were 74 cases of a strong preference (i.e. the rank 1, 2 or 3) for an unattractive face (4.2% of all 1760 assessments of these 10 faces by 176 women). If these choices were by chance (not by true preference), they would not

be systematically repeated in the second session, that is, ranks 1–5 would be as much probable as ranks 6–10. In fact, in 46 of the total 74 cases (62%), an "unattractive" face strongly preferred in the first session was also preferred in the second session (i.e. it was ranked 1–5). The probability of such a departure from the random 50% is $p = 0.047$ (the two-tailed test based on binomial distribution). This suggests some, though not substantial, stability of atypical preferences. The "other way" analysis yielded the following results: Amongst 68 (3.9%) cases of a strong preference for an unattractive face in the *second* session, a preference for the face in the first session was displayed in 40 cases, that is in 59%, and this value does not differ significantly from the random 50% ($p = 0.18$).

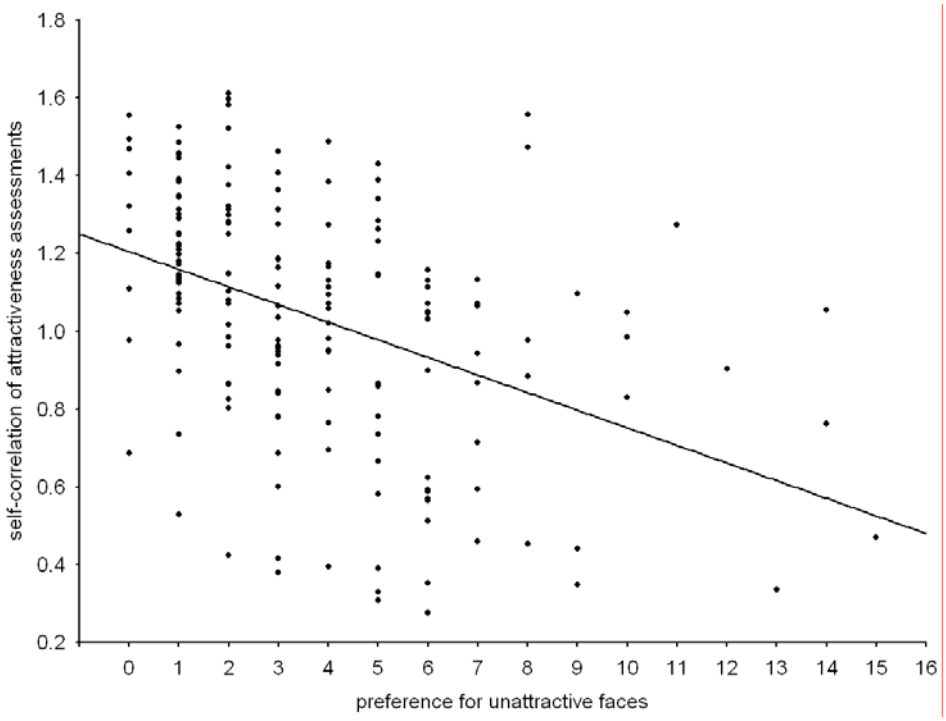

Figure 3. The preference for faces commonly regarded as unattractive and the Fisher-transformed self-correlation of attractiveness assessments ($n = 176$, $r = –0.42$, $p < 0.001$)

Next, we counted how many times each judge displayed a strong preference for any of the unattractive faces (3 top faces in each series × 3 series per session × 2 sessions give a maximum possible value of 18). This index of "the preference for unattractiveness" was negatively related to the self-correlation of attractiveness assessments ($R = –0.46$, $p < 0.001$), which suggests that individuals highly assessing faces commonly regarded as unattractive are characterized by an unstable percep-

tion of facial attractiveness. Nonetheless, a scatter plot of the self-correlation against the preference for unattractiveness demonstrates that some individuals of a relatively strong preference for unattractiveness had a low self-correlation, but in some others it was high (*Fig. 3*).

Altogether, the above analyses leads to the following conclusion: Some people perceive facial attractiveness atypically (i.e. differ substantially from the "average" person), because they possess their own, individual taste; nevertheless, in most cases, atypical attractiveness evaluations reflect low precision, and thereby low stability, of attractiveness assessments.

The Stability of Preferences for Facial Features

Mean preferences for facial features are given in the upper section of *Table 1*. As can be seen, facial cues to good genes and high biological quality (i.e. skin healthiness, the lover index, and also the spouse index) were preferred more strongly than cues to good personality (the mouth positivity and the friend index). Both groups of participants manifested similar strengths of preferences, only the yearly group preferred the skin healthiness slightly more strongly than the monthly group (M_{MONTH} = 0.42, M_{YEAR} = 0.47, t_{174} = 2.08, p = 0.04). The middle section of *Table 1* gives between-session repeatability of preference strengths. These repeatability values were similar for each facial feature, irrespective of whether it was preferred strongly or weakly, and approximated to 0.5–0.6. The monthly and yearly group did not differ in the repeatability of preference for any feature (all ps > 0.08). The bottom section of *Table 1* gives absolute values of the between-session differences in preference strengths. These values appear rather small (0.10–0.15), but noticeably greater for the lover index (0.20). Magnitudes of changes in these strengths were similar in both groups of judges (all ps > 0.3). Therefore, both the analysis of magnitudes of changes in preference strengths for facial features, and the analysis of repeatability of these preferences demonstrate that the stability of facial preference criteria was not lower at a yearly than a monthly interval.

The stability of preferences for facial features, as measured with the between-session magnitude of their changes, proved to be dependent on the sociosexual orientation index, SO (*Fig. 4*). Women of SO above zero (i.e. relatively open to casual sex, n = 72), as compared to women of SO below zero (n = 104), were characterized by higher stability of the preference for the lover index ($M_{HIGH\text{-}SO}$ = 0.17, $M_{LOW\text{-}SO}$ = 0.23, t_{174} = 2.31, p = 0.022), and lower stability of the preference for the friend index ($M_{HIGH\text{-}SO}$ = 0.13, $M_{LOW\text{-}SO}$ = 0.10, t_{174} = –2.12, p = 0.035). Although low-SO women had substantially lower stability of preference for the lover index than for the spouse index (the paired-samples t-test: M_{LOVER} = 0.23, M_{SPOUSE} = 0.16, t_{103} = 6.06, p < 0.001), and the friend index (M_{FRIEND} = 0.10, t_{103} = 6.99, p < 0.001), there were no differences between these stabilities in high-SO women (M_{LOVER} = 0.17, M_{SPOUSE} = 0.15, M_{FRIEND} = 0.13, all ps > 0.05).

Table 1. Strengths of preferences for facial features, their repeatability, and magnitudes of their between-session change

Facial feature:	Skin healthiness	Mouth positivity	Lover index	Spouse index	Friend index
Strength of preference [a]					
monthly group (n = 90)	0.421	0.133	0.675	0.629	0.355
yearly group (n = 86)	0.465	0.101	0.712	0.631	0.320
p [d]	0.039	0.084	0.064	0.906	0.054
Repeatability [b]					
monthly group	0.631	0.453	0.625	0.570	0.544
yearly group	0.566	0.604	0.583	0.555	0.505
p [d]	0.401	0.088	0.583	0.859	0.662
Magnitude of the preference change [c]					
monthly group	0.124	0.107	0.197	0.161	0.116
yearly group	0.140	0.102	0.202	0.150	0.114
p [d]	0.365	0.681	0.814	0.581	0.877

[a] For both groups pooled, *t*-test reveals significant differences between all pairs of facial features

[b] For both groups pooled, *Z*-test reveals no significant difference between any pair of facial features

[c] For both groups pooled, *t*-test reveals significant differences between all pairs of facial features, except between the friend index and skin healthiness, and between the friend index and mouth positivity

[d] *p*-levels for between-group comparisons (repeatability was tested with *Z*-test, the other values – with *t*-test)

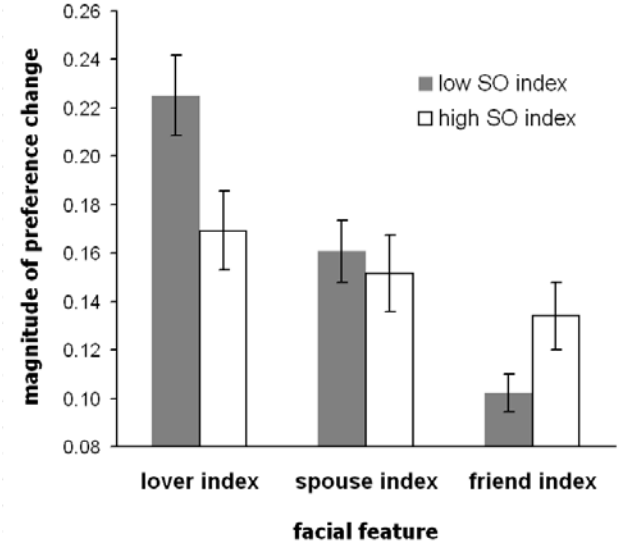

Figure 4. The magnitude of between-session change of the preference for lover index, spouse index and friend index (real faces) in groups of low and high sociosexual orientation (SO) index. Means and standard errors are presented

Additionally, an analysis of variance was conducted with the magnitude of preference strength change as the dependent variable, the facial feature as a within-subject variable (with two levels: the lover and the spouse indices) and the level of SO (below or above zero) as a between-subject variable. The analysis revealed the main effect of the facial feature ($F_{1, 174} = 22.41$, $p < 0.001$), which reflects a generally higher stability of the preference for the spouse index than the lover one, as well as the interaction between the facial feature and the SO level ($F_{1, 174} = 7.27$, $p = 0.008$), which confirms that the relation between both stabilities depended on SO.

The Stability of Judgments of Digitally Manipulated Faces

The left column of *Table 2* presents between-session repeatability of choices of digitally manipulated faces ($n = 90$, because only the monthly group judged these faces). The repeatability was always positive and usually differed significantly from zero. This means that the higher the number of the facial photo chosen by a judge in the first session (in the context of short-term, long-term or friendly relationship), the higher the number in the second session. This is evidence of some stability of preferences, although the stability is poor, since the correlations of the order of 0.2–0.3 mean that less than 10% of inter-individual variation of facial choices was stable between sessions.

Table 2. Between-session stability of choice of digitally manipulated faces in the context of short-term, long-term or friendly relationship ($n = 90$): repeatability of choices and magnitudes of their change

Series	Context	Repeatability [a]	Magnitude of choice change [b]
A	short-term	0.51***	1.29[a]
	long-term	0.26***	1.31[a]
	friendship	0.23**	1.68[b]
B	short-term	0.17*	1.38[a]
	long-term	0.28***	1.06[b]
	friendship	0.35***	1.40[a]
C	short-term	0.29***	1.32[a]
	long-term	0.52***	0.98[b]
	friendship	0.19**	1.50[a]
D	short-term	0.14	1.49[ab]
	long-term	0.23**	1.26[a]
	friendship	0.16*	1.61[b]

[a] *Z*-tests for repeatability being zero: * – $p < 0.1$, ** – $p < 0.05$, *** – $p < 0.01$

[b] Contexts with different indices within a series differ in the magnitude of choice change at the *p*-level of 0.07 (the value was chosen because some *p*-levels were between 0.05 and 0.07)

The right column of *Table 2* gives absolute values of between-session differences of choices. Within each facial series there is a tendency for greater stability of (i.e. a lower difference in) choices in the context of long-term relationship than in short-term or friendly ones. Because this tendency was frequently only marginally significant (see *Table 2*), we calculated the mean stability for each of the three contexts by averaging the respective choices across all four facial series. The stability of choices in the context of a long-term relationship was higher than in the context of a short-term relationship (the paired-samples *t*-test: $M_{LONG} = 1.15$, $M_{SHORT} = 1.37$, $t_{89} = 2.64$, $p = 0.010$), and friendly relationship ($M_{FRIEND} = 1.55$, $t_{89} = 4.41$, $p < 0.001$). The choices in the context of a short-term relationship were marginally more stable than in the friendship context ($t_{89} = 1.88$, $p = 0.063$).

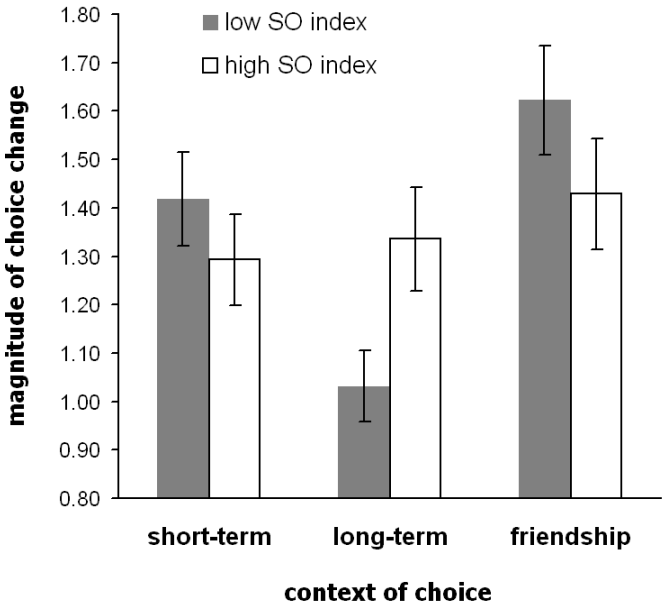

Figure 5. The magnitude of between-session change of choice in the context of short-term relationship, long-term relationship and friendship (digitally manipulated faces). Means and standard errors are presented for groups of low and high sociosexual orientation (SO) index

The sociosexual orientation index (SO) proved to moderate the relation between the choice context and its stability (*Fig. 5*). Among 55 women of SO below zero (that is, relatively closed to casual sex), the magnitudes of between-session choice changes in the marital context were clearly lower than in the lover context (the paired-samples *t*-test: $M_{LONG} = 1.03$, $M_{SHORT} = 1.42$, $t_{54} = 3.56$, $p < 0.001$), and the friendship context ($M_{FRIEND} = 1.62$, $t_{54} = 4.94$, $p < 0.001$). On the other hand, among 35 women relatively open to casual sex (SO > 0), the stabilities of choice in

these three contexts did not differ from one another (M_{SHORT} = 1.29, M_{LONG} = 1.34, M_{FRIEND} = 1.43, all *p*s > 0.2), and even the stability of lover choices was (non-significantly) higher than the stability of marital choices (p = 0.7). Direct comparison between both female groups only reveals a significant difference in the marital context – women of low SO were more stable here ($M_{LOW\text{-}SO}$ = 1.03, $M_{HIGH\text{-}SO}$ = 1.36, t_{88} = 2.42, p = 0.018).

Additionally, an analysis of variance was conducted with the magnitude of the choice change as the dependent variable, the choice context as a within-subject variable (with two levels: the short-term and long-term relationship), and the level of SO (below or above zero) as a between-subject variable. The analysis revealed the main effect of the choice context ($F_{1, 88}$ = 4.33, p = 0.040), which reflects a generally higher stability of marital than lover choices, as well as the interaction between the choice context and the SO level ($F_{1, 88}$ = 6.76, p = 0.010), which confirms that the relation between both stabilities depended on SO.

The Change of Partnership Status

In order to test whether the breakdown of a relationship impacts facial preferences, two female groups were identified: (1) 68 women who were in a relationship with the same man in both sessions (as inferred from the declared duration of the relationship), (2) 9 women who were in a relationship only during the first session, that is, their relationships broke down between sessions (7 women from the yearly group, and 2 from the monthly one). It was checked whether these groups differed from each other in terms of between-session changes in preferences for skin healthiness, mouth positivity, and the lover, spouse, and friend indices. They differed only in respect to the friendliness preference change (M-W test's Z = 3.06, p = 0.002); while the preference did not change in still-paired women (d = 0.01), it increased in women whose relationships had broken down (d = 0.15). There were similar, though only marginally significant, trends for the spouse index (*d*s were –0.03 and 0.09, respectively; M-W test's Z = 1.68, p = 0.09) and for mouth positivity (*d*s were 0.01 and 0.08, respectively; M-W test's Z = 1.32, p = 0.19). No trends were found for skin healthiness and the lover index (*p*s > 0.6).

So to check whether these female groups are comparable at all, we tested whether they differed in various dimensions at the first session. There were no significant differences in their age, the frequency of contraceptive use, the SO index, the self-assessment of physical attractiveness, or mood (all *p*s > 0.1). However, women whose relationships had ended declared a lower relationship duration in the first session (17.9 vs. 28.6 months, M-W test's Z = 2.09, p = 0.037), a lower commitment to it (3.67 vs. 5.65; M-W test's Z = 3.51, p < 0.001), and higher income (1144 vs. 778 Polish zlotys; M-W test's Z = 2.30, p = 0.022) than still-paired women. It suggests that low partner commitment and relative financial independence of woman increase the probability of a relationship breaking down.

Possible confounding effects of the relationship's duration, the woman's commitment and her income in the relationship between the relationship's survival and the enhancement of the preference for friendly looking faces should thus be eliminated. Among the women whose relationship had survived, nine individuals were identified who did not differ from women whose relationship had broken down in terms of the relationship's duration, the woman's commitment and her income (all *p*s > 0.15, according to M-W test). However, these groups differed in the change of preference for the friend index (d = –0.02 vs. 0.15, M-W test's Z = 3.05, p = 0.002). Therefore, the above variables did not mediate the relationship between the relationship's survival and the enhancement of the preference for a friendly appearance.

Further, we checked whether the start of a relationship alters facial preferences. Preference changes of women who were unpaired in both sessions (n = 81) were thus compared with preference changes of women who were single at the first session, but paired at the second session (12 women, including two from the monthly group). These groups differed only in the change of preference for the friend index (M-W test's Z = 2.69, p = 0.007): still-single women displayed a slight increase in the between-session change in the preference (d = 0.03), while the preference decreased in women who formed a new relationship (d = –0.10). Both female groups were comparable since they did not differ, at the first session, in their age, income, frequency of contraceptive use, SO index, self-assessment of physical attractiveness, or mood (all *p*s > 0.05).

The Mood Change

The between-session change of mood did not correlate with between-session changes of preferences for skin healthiness, mouth positivity, and the lover, spouse, and friend indices (all *p*s > 0.05). The partnership status, however, proved to be a strong moderator. Among women who were unpaired in both sessions (n = 81), no association of mood change with preference changes was found (all *p*s > 0.3). Yet, in women having the same partner in both sessions (n = 68), the between-session mood change positively correlated with the change of preference for skin healthiness (R = 0.31, p = 0.011), the lover index (R = 0.43, p < 0.001), and the spouse index (R = 0.34, p = 0.005), but not for mouth positivity (R = 0.13, p = 0.30), or the friend index (R = 0.08, p = 0.51). A multiple analysis of variance with changes of preferences for skin healthiness, the lover index and the spouse index as dependent variables, the partnership status as a qualitative predictor, and mood change as a covariate revealed the interaction between partnership status and mood change ($F_{3, 143}$ = 5.52, p = 0.001), which confirms the moderating effect of the partnership status. Also, in each of three univariate analyses of variance, the interaction between the partnership status and the mood change was significant (all *p*s < 0.004). Additional

tests did not reveal a moderating effect of SO index on the relationship between the mood change and preference changes.

In the present study, the mood was measured with four variables: happiness, general self-esteem, vitality, and social life. In order to establish which of these were responsible for the relationship between the mood change and preference changes in still-paired women, forward stepwise regressions were conducted with a preference change as the dependent variable, and the mood components were entered into the model as independent variables. The change of preference for skin healthiness and for the spouse index proved to depend on changes in the declared vitality and social life, while the change of preference for the lover index depended only on the change of vitality (*Table 3*).

Table 3. Between-session changes (Δs) of facial preferences as related to between-session changes (Δs) of mood indicators in women being in a relationship with the same partner at both sessions (n = 68). Results of forward stepwise regression analysis with between-session changes of four mood indicators as predictors: happiness, vitality, self-esteem and social life[a]

Criterion variable	Predictor	β	Standardized β	t	p
Δ preference for skin healthiness	Δ social life	0.08	0.28	2.45	0.017
(adjusted R^2 = 14.4%)	Δ vitality	0.07	0.26	2.32	0.023
Δ preference for lover index	Δ vitality	0.17	0.47	4.32	0.000
(adjusted R^2 = 20.9%)					
Δ preference for spouse index	Δ vitality	0.14	0.47	4.52	0.000
(adjusted R^2 = 28.0%)	Δ social life	0.08	0.22	2.11	0.039

[a] Analyses of preferences for mouth positivity and the friend index produced null models, so these are not presented here

Other results

Many other tests aimed at revealing correlates of the stability and changes of preferences were carried out. As in any relatively long series of statistical tests, some p-levels below 0.05 have been obtained. In no case, however, was an effect so clear (e.g. present in both sessions and in both female groups) that the risk of its accidental origin could be disregarded. In particular, the self-correlation of attractiveness assessments did not depend on whether a woman started to use hormonal contraceptives between the sessions, ceased to use them, still used them or still did not use them (ANOVA's p = 0.29). Thus, our results for facial preferences do not confirm those of ROBERTS et al. (2008) who found that the self-correlation of male sweat scent assessments was higher in women not using hormonal contraceptives than in women who started to take the contraceptives between the sessions (see Introduction).

DISCUSSION

The Preference Stability and the Between-session Interval

The preference stability at a yearly interval was not lower than at a monthly interval, both in terms of the self-correlation of attractiveness assessments, and in terms of repeatability and magnitudes of changes in strengths of preferences for five facial features. The self-correlation was estimated at 0.78, which agrees with results of other studies in which the between-session interval was much shorter (see Introduction). It suggests that facial preference stability is not interval-dependent; at least in the month-year range, since two examined female groups were fully comparable in terms of many questionnaire data and were investigated with the same methods. This independence is somewhat surprising, since one might expect that after longer intervals more changes in participants' internal state and living situation could have accrued, which could translate into greater changes in the attractiveness perception. The decrease in test–retest correlations with the increasing interval is well known in psychological literature and pertains to such characteristics as cognitive abilities (SCHUERGER and WITT 1989), personality (ROBERTS and DELVECCHIO 2000; GRØNNERØD 2003), and esthetic judgments of graphic patterns (HÖFEL and JACOBSEN 2003). Interestingly, however, BROOKS (1996) had very similar repeatability of female guppies' choices of males at the intervals of 30 minutes, one day and two days.

The interval independence of the facial preference stability suggests that between-session changes of attractiveness assessments were caused by quickly fluctuating factors rather than factors of directional trends over time (which effects could have accrued with time, e.g. personality development), nor rarely occurring large-effect factors (because the larger the interval between the sessions, the greater would be the probability that such a factor comes up). In the present study, the changes of two factors were found to impact facial preferences: the change of partnership status and the change of mood. Mood changes belong to the category of quickly and substantially fluctuating factors. However, a partnership status change (the formation of a relationship or its breakdown) is a relatively rare event, and therefore the probability of its occurrence increases with the time interval (here the partnership status has changed in 4% of women from the monthly group, and in 27% of women from the yearly group). The results obtained, however, suggest that even at a yearly interval the frequency of these events is too low to cause a decrease in the preference stability noticeable for the whole yearly group, and the pattern of induced changes in preference is too specific (pertains only to the friend index) to make a decrease in the stability noticeable even in the subgroup of those women whose partnership status has changed (the comparison of the self-correlation between these women and the others: $t_{174} = 0.53$, $p = 0.60$).

Determinants of the Preference Stability

The breakdown of a relationship was found to enhance the preference for friendly looking faces. This seems intuitive, yet at least two different mechanisms may underpin this effect. Firstly, after the breakdown of a relationship, a woman may be depressed and feel the need to complain to someone and to hear some consoling words. She may therefore seek friendly relationships with others. Secondly, the failure of the relationship may be a lesson for the woman to pay more attention to good character when seeking a partner. The first mechanism may be expected to operate for a relatively short time and pertain to faces of both sexes. The second mechanism may be more durable, pertain only to male faces, and may be more pronounced when the previous partner had a rough personality. Future research incorporating all these variables may establish the mechanism unambiguously. It is also difficult to say whether the observed effect of relationship breakdown on facial preferences is an evolutionary adaptation by which the sensitivity to some facial features (e.g. cues to good character) enhances after a failure (e.g. relationship breakdown), or this effect develops by social learning of associations between facial appearance and personality.

Effects of partnership status on facial preferences were found in some previous studies: PENTON-VOAK et al. (1999) and LITTLE et al. (2002) reported that women with a partner prefer more masculinized male faces than those without a partner, LITTLE et al. (2007) found that female preference for symmetry in male faces fluctuates with menstrual cycle only in paired women, and PENTON-VOAK et al. (1999) observed a trend indicating that menstrual cycle-related changes in facial masculinity preferences are greater in women having a partner. Men with lowly masculine faces are perceived to possess more pro-social personality than their more masculine counterparts (PERRETT et al. 1998; JOHNSTON et al. 2001), so the effect of relationship breakdown on the preference for friendly looking faces might have been involved in the associations between partnership status and facial masculinity preference. This supposition cannot, however, be tested because those studies lacked data on how long ago the unpaired women ended their relationships, if any. On the other hand, the effect of having a partner on women's preferences for male facial masculinity and symmetry has been commonly interpreted in terms of hunting for men with good genes (PENTON-VOAK et al. 1999; LITTLE et al. 2002, 2007). A casual contact with such a man enables the woman to "gain" good genes for her offspring and simultaneously to set the stable partner up to care for the child. In the present study, however, no increase in the preference for healthy or sexy looking faces was found among women who have entered a new relationship. The theory predicts such an increase only in women prone to cheat their stable partner. The inclination may be estimated with the sociosexual orientation index (SIMPSON and GANGESTAD 1991), which was known in the present data. Unfortunately, the number of women who have formed a new relationship was too small (n = 12) to reliably test the prediction in this study.

Unexpectedly, however, among women who have formed a new relationship, entering into the relationship was accompanied by a decrease in the preference for a friendly appearance. We may only speculate on the mechanism of this effect. Maybe women in a new relationship have their need for emotional closeness satisfied, so they feel little need to form another close relationship. It is also possible that the observed effect arose by chance. In total, ten statistical tests of the preference dependence on the change of partnership status (i.e. preferences for five facial features × two types of the partnership status change) have been conducted, so the *p*-level of 0.007 obtained does not survive the Bonferroni correction ($0.007 \times 10 = 0.07 > 0.05$). Let us add that the previously discussed effect of the *breakdown* of the relationship on the *increase* in the preference for friendly appearance had the *p*-level of 0.002, which does survive this correction ($0.002 \times 10 = 0.02 < 0.05$).

Mood changes proved another cause of preference changes over time, but only in women in a stable relationship. While mood changes in single women exerted no changes in facial preferences, the mood improvement in paired women (especially vitality) was associated with the increase in a preference for faces that looked healthy, sexy, and maritally suitable. The mood is a subjective feeling with a biological foundation; it reflects the physiological state and the individual's ecological situation, and makes them take or abandon an action (THAYER 1989). Women in stable relationships are more interested in love affairs with good-genes men than single women (see above), which explains the fact that the influence of the mood on preferences for healthy and sexy appearance existed only in paired women. The influence of mood on the preference for marital appearance may have resulted from the high correlation between sexy and marital appearance, $R = 0.88$. The association between a woman's interest in men with cues to good genes and her partnership status was also observed in other studies: paired women prefer more masculinized male faces and display greater changes in the facial masculinity preference during the menstrual cycle than single women do (PENTON-VOAK et al. 1999; LITTLE et al. 2002). Sexual desire (PILLSWORTH, HASELTON and BUSS 2004), attention to courtship language (ROSEN and LÓPEZ 2009), preference for scent of dominant men (HAVLICEK, ROBERTS and FLEGR 2005) and preference for symmetric male faces (LITTLE et al. 2007) were observed to increase in the fertile phase of the menstrual cycle, but only in women being in stable relationships. It seems difficult to explain the revealed pattern of mood-preference associations in another way than as seeking good genes when the circumstances are favorable, so the pattern supports the psychoevolutionary perspective on the facial attractiveness perception.

When assessing real faces, the stability of preference for the lover index was lower than for the spouse index; and, in the case of digitally manipulated faces, the stability of choices in the context of a short-term relationship was lower than in the context of a long-term one. These stabilities depended on the sociosexual orientation index, SO (which measured a woman's openness to casual sex). High-SO women displayed more stable preference for the lover index (in real faces) than low-SO ones, while low-SO women were more stable than high-SO ones in terms

of the partner choice for a long-term relationship (digitally manipulated faces). High-SO women are, by definition, more interested in short-term relationships than low-SO ones. Therefore, females from each group assess faces more unvaryingly in the context they are more interested in and more unvaryingly assess those facial features that are important in that context. This is in concord with other researchers' findings that high-SO women prefer masculine male faces more strongly than low-SO women (WAYNFORTH, DELWADIA and CAMM 2005), and that such faces are more preferred in the context of short-term than long-term relationships (LITTLE et al. 2002). The present results also fit well with the psychoevolutionary view of facial attractiveness perception. Natural selection may have equipped individuals of low and high SO with mechanisms of attractiveness perception that are most precise (and thereby most stable) for the aspect which is most important for their respective reproductive success. On the other hand, one cannot exclude conscious factors having contributed to the results obtained. For example, women who were not interested in short-term relationships at all might have been careless when choosing a manipulated face for a short-term relationship, which would lower their stability. While assessing real faces, women were instructed to sort the faces quickly, by their first impressions, and unreflectively. The role of conscious factors was thus here less probable but still possible, since not all women evaluated the faces unreflectively.

One consequence of the individual instability of preferences is the decrease in preference typicality. If someone assesses faces in a different way each time, their judgments will not usually be concordant with the population mean. The present study suggests that most people who display an atypical taste have rather unstable judgments and do not repeat their atypical evaluations on another occasion. Therefore, although some people perceive facial attractiveness atypically, because they possess their own, individual taste, in most cases atypical attractiveness evaluations reflect the low precision, and thereby low stability, of attractiveness assessments. This challenges the optimistic thesis of HÖNEKOPP (2006) that "(almost) all people are good-looking – at least to some" (p. 208). The fact that an unattractive face appealed to someone on one occasion does not mean that it will still be liked on another occasion.

Why are Attractiveness Assessments so Unstable?

People can discern thousands of individual faces and need only a fraction of a second to recognize facial identity, attractiveness or expression (BRUCE and YOUNG 1998). A distinction between attractive and unattractive faces is possible even in a subliminal presentation (OLSON and MARSHUETZ 2005). One might thus expect the perception of facial attractiveness to be highly precise and stable in time. The reliable assessment of attractiveness may be also expected to develop through the course of evolution, because it facilitates the choice of a partner of high mate value, and hence it is beneficial for one's reproductive success (JONES 1996; GANGESTAD

and SCHEYD 2005). Contrary to these predictions, the present study demonstrated that the self-correlation of women's attractiveness assessments of a representative set of male faces is only about 0.78, which means that as much as 40% ($1 - 0.78^2$) of individual variation in facial assessments is unstable with time. As we clarified in Introduction, however, preferences do not need to be highly stable as long as their changes are functionally associated with changes in an individual's internal state or ecological situation. Indeed, many experimental studies on the influence of visual and social stimuli on the attractiveness perception as well as observational studies on the effect of the menstrual cycle phase proved the intra-individual preference changes to be functional (see PENTON-VOAK and PERRETT 2000; KOŚCIŃSKI 2008). However, sizes of the effects revealed in those studies were small to moderate. So, if intra-individual changes of preferences are not large in experimental investigations (which intensively treated participants with artificially prepared stimuli), they would be rather unexpected to be large in natural conditions. The magnitude of currently reported instability of facial preferences needs therefore an account. Let us discuss several hypothetical explanations:

1. Participants might have felt bored during the second session, which resulted in less precise facial judgments. However, neither the preference typicality nor strength of preference for any facial feature were lower in the second session than in the first one (all *p*s > 0.09), which challenges the above conjecture.

2. From the evolutionary point of view, it is profitable for females to possess genetically diversified offspring, and therefore to mate with many males (JENNIONS and PETRIE 2000). GANGESTAD and COUSINS (2001) allow the existence of this mechanism in humans, too. If mating with many men were profitable for women, then some degree of instability of attractiveness perception might be adaptive. In such a case, however, the instability should exist only or mainly in women open for causal sex. Nonetheless, no association of the sociosexual orientation index with the self-correlation or with the stability of preferences for facial features existed in the present sample (all *p*s > 0.18).

3. In contemporary developed societies, conditions for an individual's growth and development are generally good and the prevalence of factors impairing development is relatively low. The inter-individual variation in facial appearance, including the biological quality cues, may thus be currently lower than in our ancestors (in the so-called environment of evolutionary adaptedness). This may hinder the distinction of facial attractiveness. HENNEBERG, PIONTEK and STRZAŁKO (1978) demonstrated, however, that the intra-population variation of facial dimensions did *increase* from the Neolithic Age to the present times, presumably because the stabilizing natural selection was weakening along with cultural development. On the other hand, the intensity of infectious diseases is presently generally lower than in our ancestors, which may result in a relatively small inter-individual variation of facial cues to health (mainly the skin condition) in modern populations. This hypothesis may be verified by checking whether the stability of preferences in contemporary

hunter-gatherer populations (where the intensity of development disturbing factors is substantial) is larger than in industrialized societies.

4. Maybe an observer can only discern, in a time-invariant way, attractive faces from unattractive ones, and faces within each of these groups are ranked at random. Assume that in a ten-face series, each judge consistently accepts half of the faces and perceives them as attractive, while the rest of the faces are regarded as unattractive (and apart from that they rank these faces randomly). Then the mean between-session self-correlation of attractiveness assessments would amount to 0.76 (as modeled with 10,000 random repetitions), and this agrees well with the empirically obtained value of 0.78. Of course, this does not mean that each observer perceives attractiveness according to these assumptions; it only gives an idea of the extent of the instability of the perception. It should also be stressed that these crude rules describe only the *time-invariant* perception of attractiveness; an observer may discern facial attractiveness much more precisely at an examination, but this subtlety is not stable over time. Finally, as demonstrated above, attractiveness assessment instability does not need to be random, since it may result from an individual's changes of internal state or living situation.

5. Maybe many faces differ from one another only by features irrelevant to the attractiveness assessment, and a judge would thus rank these faces at random. If so, the self-correlation of attractiveness assessments would be much smaller than unity, but the stability of preference strengths for essential facial features might still be high. In the present study, magnitudes of between-session changes in preferences for five facial features were usually in the range 0.10–0.15. Assume that in a ten-face series, judges can only discern the half of the faces with a relatively high value of a feature from the rest of the faces and rank the former higher in attractiveness than the latter (and apart from that they rank the faces randomly). Then, the mean magnitude of the preference change would be 0.10 (as modeled with 10,000 random repetitions), which roughly agrees with the empirical values obtained. Therefore, the stability of preferences for facial features is not impressive.

Summing up, it is difficult to explain the relatively low individual stability of facial preferences. Three hypotheses seem possible and should be addressed by future research: (1) preferences are relatively unstable only in industrialized populations; instead they were more stable in our ancestors, (2) preferences are unstable because individuals' internal states and living situation vary with time, (3) the large preference instability reflects the neural system's low precision of attractiveness perception; maybe in the evolutionary past there was no necessity for a higher precision of attractiveness evaluations from facial appearance?

CONCLUSIONS

The stability of physical attractiveness perception in natural conditions has so far been studied only rarely and superficially. The present study is the first to demon-

strate the importance of the attitude to casual sex, the change of partnership status and the change of mood for the stability and changes of the facial attractiveness perception. Research on intra-individual variation of facial preferences may provide some evidence for the psychoevolutionary perspective on physical attractiveness perception. One such example is the present relationship between changes of the mood and changes of preferences for facial cues to good genes, which was clear in women who were in a stable relationship, but absent in unpaired women.

The present study was preliminary and many questionnaire scales had a short form. The application of a full (11-item) Sociosexual Orientation Inventory and of more items being used for the diagnostics of the internal state changes (mood, health) and the partnership dynamics may bring many interesting results. Further research may address the stability of male assessments of female facial attractiveness as well as assessments of own-sex faces in which intra-sexual competition is involved (FISHER 2004). The stability of preferences for silhouettes, voices and scents are also worthy of scientific interest. We predict that multi-session designs, in which participants are examined many times at specific time intervals, may be more powerful than the two-session design applied in the present study. The present results may also be helpful in research not related to the stability of attractiveness perception. Since an individual's preferences manifested at an examination depend strongly on their current mood, a statistical control of the inter-individual mood variation could make other effects, essential in a given study, more clear.

ACKNOWLEDGEMENTS

The author would like to thank two anonymous reviewers for their useful comments, Jan Strzałko for his helpful discussions and Zofia Huflejt and Renata Stefańska for their help in data collection.

REFERENCES

ATTWOOD, A. S., PENTON-VOAK, I. S. & MUNAFÒ, M. R. (2009): Effects of acute nicotine administration on ratings of attractiveness of facial cues. *Nicotine and Tobacco Research*, 11, 44–48.

BAKKER, T. C. M. (1993): Positive genetic correlation between female preference and preferred male ornament in sticklebacks. *Nature*, 363, 255–257.

BECKER, W. A. (1984): *Manual of Quantitative Genetics*, Washington: Academic Enterprises, Pullman.

BLOM, G. (1958): *Statistical Estimates and Transformed Beta-variables.* New York: John Wiley & Sons, Inc.

BRONSTAD, P. M. & RUSSELL, R. (2007): Beauty is in the 'we' of the beholder: Greater agreement on facial attractiveness among close relations. *Perception*, 36, 1674–1681.

BROOKS, R. (1996): Copying and the repeatability of mate choice. *Behavioral Ecology and Sociobiology*, 39, 323–329.

BROOKS, R. & ENDLER, J. A. (2001): Female guppies agree to differ: Phenotypic and genetic variation in mate-choice behavior and the consequences for sexual selection. *Evolution*, 55, 1644–1655.

BRUCE, V. & YOUNG, A. W. (1998): *In the Eye of the Beholder: The Science of Face Perception.* Oxford University Press.

BURNS, D. (1999): *The feeling good handbook*. New York: William Morrow and Company.

BURT, D. M., KENTRIDGE, R. W., GOOD, J. M. M., PERRETT, D. I., TIDDEMAN, B. P. & BOOTHROYD, L. G. (2007): Q-cgi: new techniques to assess variation in perception applied to facial attractiveness. *Proceedings of the Royal Society B*, 274, 2779–2784.

BUSS, D. M. & SCHMITT, D. P. (1993): Sexual strategies theory: An evolutionary perspective on human mating. *Psychological Review*, 100, 204–232.

BUSS, D. (1999): *Evolutionary Psychology: The New Science of the Mind*. Boston: Allyn & Bacon.

CUNNINGHAM, M. R., BARBEE, A. P. & PIKE, C. L. (1990): What do women want? Facialmetric assessment of multiple motives in the perception of male facial physical attractiveness. *Journal of Personality and Social Psychology*, 59, 61–72.

DEBRUINE, L. M., LITTLE, A. C., PERRETT, D. I. & JONES, B. C. (2008): Social perception of facial resemblance in humans. *Archives of Sexual Behavior*, 37, 64–77.

DIENER, E., WOLSIC, B. & FUJITA, F. (1995): Physical attractiveness and subjective well-being. *Journal of Personality and Social Psychology*, 69, 120–129.

DOHM, M. R. (2002): Repeatability estimates do not always set an upper limit to heritability. *Functional Ecology*, 16, 273–280.

FALCONER, D. S. (1970): *Introduction to Quantitative Genetics*, Edinburgh: Oliver and Boyd.

FAURE, J. C., RIEFFE, C. & MALTHA, J. C. (2002): The influence of different facial components on facial aesthetics. *European Journal of Orthodontics*, 24, 1–7.

FEINGOLD, A. (1992): Good-looking people are not what we think. *Psychological Bulletin*, 111, 304–341.

FERGUSON, G. A. & TAKANE, Y. (1989): *Statistical Analysis in Psychology and Education.* New York: McGraw-Hill.

FISHER, M. L. (2004): Female intrasexual competition decreases female facial attractiveness. *Proceedings of the Royal Society of London B*, (Suppl.) 271, S283–S285.

FORSTMEIER, W. & BIRKHEAD, T. R. (2004): Repeatability of mate choice in the zebra finch: Consistency within and between females. *Animal Behaviour*, 68, 1017–1028.

GANGESTAD, S. W. & COUSINS, A. J. (2001): Adaptive design, female mate preferences, and shifts across the menstrual cycle. *Annual Review of Sex Research*, 12, 145–185.

GANGESTAD, S. W., THORNHILL, R. & GARVER-APGAR, C. E. (2005): Adaptations to ovulation. Implications for sexual and social behavior. *Current Directions in Psychological Science*, 14, 312–316.

GANGESTAD, S. W. & SCHEYD, G. J. (2005): The evolution of human physical attractiveness. *Annual Review of Anthropology*, 34, 523–548.

GRØNNERØD, C. (2003): Temporal stability in the Rorschach method: A meta-analytic review. *Journal of Personality Assessment*, 80, 272–293.

HAVLICEK, J., ROBERTS, S.C. & FLEGR, J. (2005): Women's preference for dominant male odour: Effects of menstrual cycle and relationship status. *Biology Letters*, 1, 256–259.

HENNEBERG, M., PIONTEK, J. & STRZAŁKO, J. (1978): Natural selection and morphological variability. *Current Anthropology*, 19, 67–82.

HÖFEL, L. & JACOBSEN, T. (2003): Temporal stability and consistency of aesthetic judgments of beauty of formal graphic patterns. *Perceptual and Motor Skills*, 96, 30–32.

HOLVECK, M. J. & RIEBEL, K. (2007): Preferred songs predict preferred males: consistency and repeatability of zebra finch females across three test contexts. *Animal Behaviour*, 74, 297–309.

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*, 32, 199–209.

HOWARD, R. D. & YOUNG, J. R. (1998): Individual variation in male vocal traits and female mating preferences in Bufo americanus. *Animal Behaviour*, 55, 1165–1179.

HOWELLS, D. J., SHAW, W. C. (1985): The validity and reliability of ratings of dental and facial attractiveness for epidemiologic use. *American Journal of Orthodontics*, 88, 402–408.

JENNIONS, M. D., BACKWELL, P. R. Y. & PASSMORE, N. I. (1995): Repeatability of mate choice: the effect of size in the African painted reed frog, Hyperolius marmoratus. *Animal Behaviour*, 49, 181–186.

JENNIONS, M. D. & PETRIE, M. (1997): Variation in mate choice and mating preferences: A review of causes and consequences. *Biological Reviews*, 72, 283–327.

JENNIONS, M. D. & PETRIE, M. (2000): Why do females mate multiply? A review of the genetic benefits. *Biological Reviews*, 75, 21–64.

JOHNSTON, V. S., HAGEL, R., FRANKLIN, M., FINK, B. & GRAMMER, K. (2001): Male facial attractiveness – evidence for hormone-mediated adaptive design. *Evolution and Human Behavior*, 22, 251–267.

JONES, B. C. & DEBRUINE, L. M. (2008): Adaptation reinforces preferences for correlates of attractive facial cues. *Visual Cognition*, 16, 849–858.

JONES, B. C., DEBRUINE, L. M., LITTLE, A. C., BURRISS, R. P. & FEINBERG, D. R. (2007a): Social transmission of face preferences among humans. *Proceedings of the Royal Society B*, 274, 899–903.

JONES, B. C., DEBRUINE, L. M., LITTLE, A. C. & FEINBERG, D. R. (2007b): The valence of experiences with faces influences generalized preferences. *Journal of Evolutionary Psychology*, 5, 119–129.

JONES, B. C., DEBRUINE, L. M., PERRETT, D. I., LITTLE, A. C., FEINBERG, D. R. & LAW SMITH, M. J. (2008): Effects of menstrual cycle phase on face preferences. *Archives of Sexual Behavior*, 37, 78–84.

JONES, B. C., LITTLE, A. C., FEINBERG, D. R., PENTON-VOAK, I. S., TIDDEMAN, B. P. & PERRETT D. I. (2004): The relationship between shape symmetry and perceived skin condition in male facial attractiveness. *Evolution and Human Behavior*, 25, 24–30.

JONES, D. (1996): *Physical Attractiveness and the Theory of Sexual Selection*, Ann Arbor: University of Michigan, Museum of Anthropology.

KEATING, C. F. & DOYLE, J. (2002): The faces of desirable mates and dates contain mixed social status cues. *Journal of Experimental and Social Psychology*, 38, 414–424.

KODRIC-BROWN, A. & NICOLETTO, P. F. (1997): Repeatability of female choice in the guppy: response to live and videotaped males. *Animal Behaviour*, 54, 369–376.

KOŚCIŃSKI, K. (2008): Facial attractiveness: Variation, adaptiveness and consequences of facial preferences. *Anthropological Review*, 71, 77–105.

LANGLOIS, J. H., KALAKANIS, L., RUBENSTEIN, A. J., LARSON, A., HALLAM, M. & SMOOT, M. (2000): Maxims or myths of beauty. *Psychological Bulletin*, 126, 390–423.

LESSELLS, C. M. & BOAG, P. T. (1987): Unrepeatable repeatabilities: A common mistake. *Auk*, 104, 116–121.

LITTLE, A. C., DEBRUINE, L. M. & JONES, B. C. (2005): Sex-contingent face after-effects suggest distinct neural populations code male and female faces. *Proceedings of the Royal Society B*, 272, 2283–2287.

LITTLE, A. C., JONES, B. C., BURT, D. M. & PERRETT, D. I. (2007): Preferences for symmetry in faces change across the menstrual cycle. *Biological Psychology*, 76, 209–216.

LITTLE, A. C., JONES, B. C., PENTON-VOAK, I. S., BURT, D. M. & PERRETT, D. I. (2002): Partnership status and the temporal context of relationships influence human female preferences for sexual dimorphism in male face shape. *Proceedings of the Royal Society of London B*, 269, 1095–1100.

LITTLE, A. C. & MANNION, H. (2006): Viewing attractive or unattractive same-sex individuals changes self-rated attractiveness and face preferences in women. *Animal Behavior*, 72, 981–987.

MAPLE, J. R., VIG, K. W. L., BECK, F. M., LARSEN, P. E. & SHANKER, S. (2005): A comparison of providers' and consumers' perceptions of facial-profile attractiveness. *American Journal of Orthodontics and Dentofacial Orthopedics*, 128, 690–696.

MØLLER, A. P. (1994): Repeatability of female choice in a monogamous swallow. *Animal Behaviour*, 47, 643–648.

OLSON, I. R. & MARSHUETZ, C. (2005): Facial attractiveness is appraised in a glance. *Emotion*, 5, 498–502.

PARKER, L. L. C., PENTON-VOAK, I. S., ATTWOOD, A. S. & MUNAF, M. R. (2008): Effects of acute alcohol consumption on ratings of attractiveness of facial stimuli: Evidence of long-term encoding. *Alcohol Alcoholism*, 43, 636–640.

PENTON-VOAK, I. S., PERRETT, D. I., CASTLES, D. L., KOBAYASHI, T., BURT, D. M., MURRAY, L. K. & MINAMISAWA, R. (1999): Menstrual cycle alters face preference. *Nature*, 399, 741–742.

PENTON-VOAK, I. S. & PERRETT, D. I. (2000): Consistency and individual differences in facial attractiveness judgements: An evolutionary perspective. *Social Research*, 67, 219–244.

PENTON-VOAK, I. S., LITTLE, A. C., JONES, B. C., BURT, D. M., TIDDEMAN, B. P. & PERRETT, D. I. (2003): Female condition influences preferences for sexual dimorphism in faces of male humans (Homo sapiens). *Journal of Comparative Psychology*, 117, 264–271.

PILLSWORTH, E. G., HASELTON, M. G., BUSS, D. M. (2004): Ovulatory shifts in female sexual desire. *Journal of Sex Research*, 41, 55–65.

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.

PETTIJOHN, T. F. & TESSER, A. (2005): Threat and social choice: When eye size matters. *Journal of Social Psychology*, 145, 547–570.

REINHOLD, K., REINHOLD, K. & JACOBY, K. J. (2002): Dissecting the repeatability of female choice in the grasshopper Chorthippus biguttulus. *Animal Behaviour*, 64, 245–250.

RHODES, G., HALBERSTADT, J., JEFFERY, L. & PALERMO, R. (2005): The attractiveness of average faces is not a generalized mere exposure effect. *Social Cognition*, 23, 205–217.

RHODES, G., JEFFERY, L., WATSON, T. L., CLIFFORD, C. W. G. & NAKAYAMA, K. (2003): Fitting the mind to the world: Face adaptation and attractiveness aftereffects. *Psychological Science*, 14, 558–566.

RHODES, G., LOUW, K. & EVANGELISTA, E. (2009): Perceptual adaptation to facial asymmetries. *Psychonomic Bulletin and Review*, 16, 503–508.

RHODES, G., ROBERTS, J. & SIMMONS, L, W. (1999): Reflections on symmetry and attractiveness. *Psychology, Evolution and Gender*, 1, 279–295.

RHODES, G. (2006): The evolutionary psychology of facial beauty. *Annual Review of Psychology*, 57, 199–226.

ROBERTS, B. W. & DELVECCHIO, W. F. (2000): The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin*, 126, 3–25.

ROBERTS, S. C., GOSLING, L. M., CARTER, V. & PETRIE, M. (2008): MHC-correlated odour preferences in humans and the use of oral contraceptives. *Proceedings of the Royal Society B*, 275, 2715–2722.

ROSEN, M. L. & LÓPEZ, H. H. (2009): Menstrual cycle shifts in attentional bias for courtship language. *Evolution and Human Behavior*, 30, 131–140.

ROWLAND, D. A. & PERRETT, D. I. (1995): Manipulating facial appearance through shape and color. *IEEE Computer Graphics and Applications*, 15, 70–76.

SCHUERGER, J. M. & WITT, A. C. (1989): The temporal stability of individually tested intelligence. *Journal of Clinical Psychology*, 45, 294–302.

SILVER, N. C. & DUNLAP, W. P. (1987): Averaging correlation coefficients: Should Fisher's z transformation be used? *Journal of Applied Psychology*, 72 146–148.

SIMPSON, J. A. & GANGESTAD, S. W. (1991): Individual differences in sociosexuality: Evidence for convergent and discriminant validity. *Journal of Personality and Social Psychology*, 60, 870–883.

SYMONS, D. (1995): Beauty is in the adaptation of the beholder: The evolutionary psychology of human female sexual attractiveness. In: Abramson, P. R., Pinkerton, S.D. (eds): *Sexual Nature/Sexual Culture*. Chicago: University of Chicago Press, pp. 80–118.

THEODORIDOU, A., ROWE, A. C., PENTON-VOAK, I. S. & ROGERS, P. J. (2009): Oxytocin and social perception: Oxytocin increases perceived facial trustworthiness and attractiveness. *Hormones and Behavior*, 56, 128–132.

THAYER, R. E. (1989): *The Biopsychology of Mood and Arousal*, Oxford University Press.

THORNE, F., NEAVE, N., SCHOLEY, A., MOSS, M. & FINK B. (2002): Effects of putative male pheromones on female ratings of male attractiveness: Influence of oral contraceptives and the menstrual cycle. *Neuroendocrinology Letters*, 23, 291–297.

TODRANK, J., BYRNES, D., WRZESNIEWSKI, A. & ROZIN, P. (1995): Odors can change preferences for people in photographs: A cross-modal evaluative conditioning study with olfactory USs and visual CSs. *Learning and Motivation*, 26, 116–140.

WAGNER, W. E. (1998): Measuring female mating preferences. *Animal Behaviour*, 55, 1029–1042.

WAGNER, W. E., MURRAY, A. M. & CADE, W. H. (1995): Phenotypic variation in the mating preferences of female field crickets, Gryllus integer. *Animal Behaviour*, 49, 1269–1281.

WAYNFORTH, D., DELWADIA, S. & CAMM, M. (2005): The influence of women's mating strategies on preference for masculine facial architecture. *Evolution and Human Behavior*, 26, 409–416.

WHITTINGHAM, L. A., DUNN, P. O. & STAPLETON, M. K. (2006): Repeatability of extra-pair mating in tree swallows. *Molecular Ecology*, 15, 841–849.

WIDEMO, F. & SÆTHER, S. A. (1999): Beauty is in the eye of beholder: Causes and consequences of variation in mating preferences. *Trends in Ecology and Evolution*, 14, 26–31.

WONG, D. T. & CUNNINGHAM, M. R. (1990): Interior versus exterior beauty: The effects of mood on dating preferences for different types of physically attractive women. Paper presented at the Southeastern Psychological Association, Atlanta, GA.

ZAJONC, R. B. (2001): Mere exposure: A gateway to the subliminal. *Current Directions in Psychological Science*, 10, 224–228.