Perception, 2007, volume 36, pages 1674 – 1681

doi:10.1068/p5793

Beauty is in the 'we' of the beholder: Greater agreement on facial attractiveness among close relations

P Matthew Bronstad¶
Department of Psychology MS 062, Brandeis University, Waltham, MA 02454, USA
Richard Russell
Department of Psychology, Harvard University, Cambridge, MA 02138, USA
Received 6 November 2006, in revised form 5 February 2007; published online 7 November 2007

Abstract. Scientific research on facial attractiveness has focused primarily on elucidating universal factors to which all raters respond consistently. However, recent work has shown that there is also substantial disagreement between raters, highlighting the importance of determining how attractiveness preferences vary among different individuals. We conducted a typical attractiveness ratings study, but took the unusual step of recruiting pairs of subjects who were spouses, siblings, or close friends. The agreement between pairs of affiliated friends, siblings, and spouses was significantly greater than between pairs of strangers drawn from the same race and culture, providing evidence that facial-attractiveness preferences are socially organized.

1 Introduction

Scientific interest in what makes faces more or less attractive has dramatically expanded in the last two decades. This interest has focused primarily on elucidating the factors that are consistent among raters rather than on individual differences in preferences (for a review, see Rhodes 2006). One reason for this focus on universal factors has been to demonstrate that facial-attractiveness judgments are not entirely idiosyncratic or 'in the eye of the beholder', and hence are an appropriate subject for scientific research. This effort has been quite successful, resulting in the discovery and validation of a number of universal factors of attractiveness, which can largely be summarized as averageness (Langlois and Roggman 1990), symmetry (Thornhill and Gangestad 1993), and sexual dimorphism (O'Toole et al 1998; Perrett et al 1998).

Though there is shared taste (universal standards) in facial attractiveness, people do not entirely agree about the relative attractiveness of faces. Recently, Hönekopp (2006) quantified the contributions of shared and private taste to judgments of facial attractiveness, where 'private' taste refers to the standards of particular individuals that reliably lead to disagreements with shared taste. Hönekopp found that private and shared taste explained approximately equal amounts of variance in attractiveness ratings.

This does not mean that private taste need be arbitrary. Private taste could be explained by the differential expressions of processes that vary according to developmental trajectories or personal circumstance. Indeed, several studies have indicated that individual differences in attractiveness preferences depend on hormonal (Penton-Voak et al 1999) and environmental (Perrett et al 2002) factors. It is also likely that perceptual learning of faces could play a role in shaping individual differences in facial-attractiveness preferences.

Cross-cultural studies of facial attractiveness have shown that, while there is agreement on facial-attractiveness preferences between cultures, there is greater agreement within cultures (Hönekopp 2006; Jones and Hill 1993). Given the large amount of unshared taste within a seemingly homogeneous sample (eg the Caucasian German university students of Hönekopp's study), it seems possible that attractiveness preferences

¶ Current address to which all correspondence should be sent: Schepens Eye Research Institute, 20 Staniford Street, Boston, MA 02114-2500, USA; e-mail: bronstad@schepens.harvard.edu

are also socially organized at levels smaller than an entire culture. Perhaps small groups of interacting individuals unknowingly follow a local facial-attractiveness standard that differs from their neighbors in an adjacent city or in a different subculture. Depending on one's point of view, this notion may seem either trivial or radical. The conventional wisdom among facial-attractiveness researchers leans towards the latter view.

There is little information about the social distribution of facial preferences within a given culture, so we do not know the fundamental population units at which we may 'cleave nature at its joints' in terms of attractiveness preferences. Social interactions are likely important to the social distribution of attractiveness preferences, as individuals have multiple social alliances, and have varying degrees of contact between themselves and all other individuals within a population. We explored whether there are patterns of agreement at the very smallest scale of social interaction—between pairs of individuals. The goal of this study was to determine whether pairs of people who form a close relationship (spouses, siblings, and close friends) have greater agreement in facial-attractiveness preferences than do pairs of people who are strangers.

To test the hypothesis that the facial-attractiveness ratings given by pairs of people forming a close relationship (close friends, siblings, or spouses) are more strongly correlated than the ratings of pairs of strangers, we asked pairs of closely related people to rate a set of faces for attractiveness. The correlation between two raters (the inter-rater correlation) is a measure of their agreement on facial-attractiveness preferences, and we compared the distribution of inter-rater correlations between closely related pairs with the distribution of inter-rater correlations between pairs of people who did not know each other (ie were strangers). Further, we sought to begin constraining the possible explanations for any such differences by comparing agreement between pairs of people who were and were not genetically related, and who varied in their amount of shared visual experience. To estimate the amount of shared experience of each pair, we asked participants to indicate the number of years they had known the person with whom they declared a close relation, as well as the number of years they had spent in daily contact with that person.

2 Method

2.1 *Participants*

We recruited twenty spouse pairs, twenty sibling pairs, and forty-one close friend pairs. A participant was identified as a member of more than one pair if more than one close relation of the participant also took the survey (eg, if both the spouse and sibling of the participant took the survey). We used respondent-driven sampling to locate additional individuals, and in most cases at least one member of each pair was personally known to the experimenters. In all, one hundred and thirteen (sixty female) Caucasian North Americans with a median age of 29 years (SD = 15 years, range = 18–75 years) participated. There were $(n^2 - n)/2 = 6328$ non-redundant pairs of raters. The same subjects who constituted the pairs of close relations also constituted the pairs of strangers when we correlated their ratings with those from participants whom they did not know. We identified two individuals as a pair of strangers only when it was highly unlikely that they had ever met. For example, members of the same social circle who did not know each other well were not labeled strangers. Data from rater pairs who were neither strangers nor close relations were excluded from analysis. We identified 5797 pairs of participants as strangers.

2.2 *Stimuli*

Color images of 74 Caucasian undergraduates (38 male) in frontal pose with neutral expressions served as stimuli. Images were presented at 480×480 pixels (see figure 1 for examples of stimuli).

Figure 1. Examples of stimuli (participants viewed full-color images).

2.3 *Procedure*

The study was conducted online. Participants answered a short demographic survey, giving their names and the name of a partner (ie their friend, sibling, or spouse). Participants then viewed each of the 74 faces one at a time and rated them for attractiveness on a 7 point Likert scale ('very attractive' to 'very unattractive'). Participants were instructed to complete the survey independently of their partners, and all indicated in a post-survey question that they did. We wanted to ensure that partners did not rate the faces together or otherwise affect one another's ratings. A related concern is the possibility that (since the study was not conducted in the laboratory) a subject could lie, responding to the survey once as herself/himself and a second time claiming to be the spouse, sibling, or close friend. For roughly half of the close-relation pairs ($n = 38$) we were certain that both members actually took the survey, because we communicated separately with both individuals. These pairs did not have lower agreement than the pairs for which we could not be certain that both members took the survey. This indicates that higher agreement among close relations could not be due to rogue participants pretending to be both members of a pair. Further, we did not pay the participants, so there was no motivation for them to lie about their partners participating if in fact they did not. Moreover, the subjects almost universally expressed curiosity about how similar their responses were to their partners', indicating that the subjects were actually motivated to make sure that their partners did in fact take the survey.

2.4 *Analyses*

To evaluate the hypothesis that pairs of people in close relationships have more similar facial-attractiveness preferences than pairs of people who are strangers, we compared the mean squared correlations[(1)] between pairs of close relations and those between pairs of strangers. For example, to estimate the agreement in facial-attractiveness preferences between pairs of friends, we calculated the correlation coefficients between pairs of individuals who reported that they were friends. We performed transformations to correct for skew in correlation distributions, Fisher's r to z before averaging, and z to r transformations afterwards, routine procedures for averaging correlations (Rosenthal 1991). We report r^2, which is the proportion of variance in common between the ratings of the two individuals.

[(1)] We consistently present r^2 rather than r, because it scales linearly and is more straightforward to interpret.

We used permutation tests (Good 2000) to test the null hypothesis that the distribution of correlations for different groups of individuals did not differ. Permutation tests are computationally intensive statistics, in which data from each group are combined and resamples of the same size as the original groups are drawn randomly and repeatedly from the pooled data. The resampling is done irrespective of the original category label, so that ad hoc groups are recreated from the pooled data. The proportion of times that the mean difference between the ad hoc groups exceeds the mean difference between the original groups is the p value of the test. Permutation tests show superior performance to parametric tests on small sample sizes and with non-normally distributed data.

We also examined a novel measure of shared taste—the squared correlation between individual ratings and the mean ratings of all the participants, which we term the permeation coefficient. This coefficient is a simple measure of shared taste that indicates the degree to which an individual's ratings are permeated by the shared standard of attractiveness.[2] The permeation coefficient ranges from 0 to 1, with higher values indicating greater adherence of the individual to the group standards. We also calculated Hönekopp's (2006) beholder index (*bi*) for a subset of individuals for comparison. While the beholder index is calculated for the entire sample of raters (and requires test–retest data), the permeation coefficient is calculated for each individual rater. A benefit of calculating the permeation coefficient for each rater is that a non-zero value (providing that the unsquared correlation is positive) provides evidence that the individual was performing the same task as the other subjects. This allows for the principled exclusion of individual raters with near-zero permeation coefficients, who may not have performed the same task as the other subjects.

3 Results

3.1 *Hypothesis tests*

The central results of the study, that directly address our hypothesis of greater agreement among close relations than strangers, are the mean squared correlations between pairs of raters, which are presented in figure 2 for the different categories of relationships (strangers, friends, siblings, spouses). Pairs of spouses had the highest inter-rater squared correlations (mean $r^2 = 0.27$, SE $= 0.02$), followed by sibling pairs (mean $r^2 = 0.25$, SE $= 0.03$), friend pairs (mean $r^2 = 0.23$, SE $= 0.02$), and finally, pairs of strangers (mean $r^2 = 0.17$, SE $= 0.001$), shared the lowest average inter-rater squared correlation. According to the results of two-group permutation tests the inter-rater

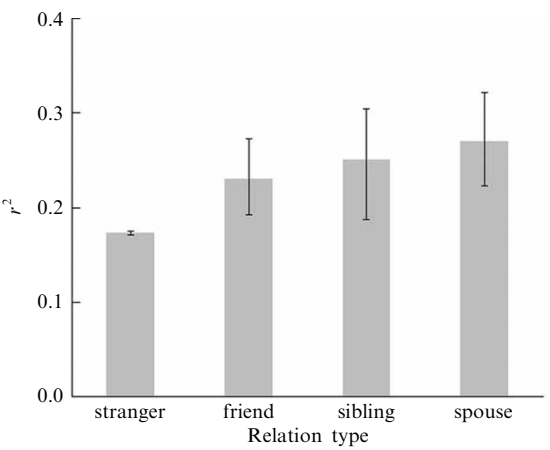

Figure 2. Mean squared inter-rater correlations (r^2) for participant relationship types. Bars represent 95% confidence intervals. The small confidence interval for strangers reflects the much larger number of stranger pairs.

[2] This was inspired by, and is similar to, the saturation coefficient (Davies 1939), which is calculated as the raters' loading on the first principal component of the ratings data. We believe that interpretation of the permeation coefficient is more straightforward.

squared correlations of the strangers were significantly smaller than were those of each of the three close-relationship types. We used 5000 resamples for each two-tailed permutation test, and each p value between stranger pairs and close-relation pairs was < 0.001. There were no significant differences between the three close-relationship types. The smallest p value among affiliated pairs was between spouses and friends ($p = 0.16$), other ps > 0.5.

It is possible that these effects are partly confounded with age. Pairs of close relations were closer in age (mean difference 3.8 years) than were pairs of strangers (mean difference 16.5 years). We computed difference in age (years) between pairs of raters, and examined the relationship between age difference and agreement on attractiveness ratings (inter-rater squared correlation). For pairs of strangers, the correlation between difference in age and agreement was $r = -0.16$ ($r^2 = 0.026$), indicating that pairs of strangers who were closer in age had more similar attractiveness preferences, with age difference accounting for slightly more than 2% of the variance. We then compared agreement among stranger pairs whose age differences were similar to the age differences in our sample of affiliates. Thus, we recomputed the stranger average using a restricted sample of stranger pairs whose age difference was 3.8 years ($n = 331$). The mean correlation of their attractiveness ratings was $r^2 = 0.19$, only slightly higher than the mean of all strangers, though there was a significant difference between the sub-sample and complete sample of strangers ($p = 0.002$). However, agreement among the sub-sample of strangers was still significantly smaller than that of the close relations ($p = 0.0002$).

3.2 *Reliability and private versus shared taste*

Attractiveness ratings were reliable (Cronbach's $\alpha = 0.96$), and the average inter-rater squared correlation was $r^2 = 0.17$, consistent with previous reports (eg Thornhill and Gangestad 1999). These measures indicate that the average agreement among participants was typical for a study of facial attractiveness. The mean permeation coefficient for all the raters was $r^2 = 0.42$ (range = 0.18 to 0.63; SD = 0.14), indicating that each rater on average shared less than half of the variance with the mean attractiveness ratings. Since permeation coefficients are not presented in contemporary studies, we calculated the permeation coefficients for two published studies (Russell 2003; Zebrowitz et al 2007), which were in the range of $r^2 = 0.3$ to 0.5. This range includes the mean permeation coefficient of the current study, and is roughly consistent with Hönekopp's (2006) findings with the beholder index. Data from one participant were removed because his permeation coefficient was close to zero (0.04), suggesting that he performed a different task than the other participants.

Twenty-four of the subjects took the survey a second time, and their test–retest squared correlations averaged $r^2 = 0.61$. There are two versions of the beholder index—bi_1, for which the differences in average attractiveness scores given by different raters are disregarded; and bi_2, for which these differences are taken to indicate meaningful inter-rater variation (Hönekopp 2006). For example, if subject A rates the faces higher on average than does subject B, this will not result in a lower bi_1 than if the average ratings of the two subjects were the same, but will result in a lower bi_2. Both beholder indices range from 0 to 1, with higher values indicating greater importance of private taste (unlike the permeation coefficient, for which higher values indicate greater importance of shared taste). We computed beholder indices $bi_1 = 0.41$ and $bi_2 = 0.65$ for these participants, according to Hönekopp's formulas. These values are similar to those Hönekopp obtained in his experiments.

3.3 *Shared experience*

To determine the relationship between shared experience and agreement on the attractiveness judgments, we calculated the correlations between the two estimates of shared

experience and the strength of the inter-rater squared correlations on the attractiveness ratings. For siblings and spouses, there was a weakly positive correlation between the number of years the pairs had spent in daily contact with each other and their agreement on the attractiveness judgments, $r^2 = 0.07$ and $r^2 = 0.09$, both ns. Though neither correlation was significant with only 20 data points each, they are in the predicted direction. There was a small positive relationship between the total number of years the spouses, $r^2 = 0.11$, $p = 0.17$, had known each other and their agreement on the attractiveness judgments, but not for the siblings, $r^2 = 0.01$, ns. There was no significant correlation with either measure of shared experience and agreement on the attractiveness judgments for close-friend pairs ($r^2 < 0.01$ for both measures).

4 Discussion

Inter-rater correlations of attractiveness ratings were significantly higher for pairs of people in close relationships than for strangers, supporting the hypothesis that people in close relationships have greater agreement in their attractiveness preferences than do strangers. While replicating the finding of shared preferences in facial-attractiveness judgments, our findings show that attractiveness standards have small-scale social organization. This also provides further evidence in support of the notion that there is large individual variation in attractiveness preferences, and that at least some of this variation is lawful.

There are several possible causes of greater agreement in facial-attractiveness preferences between individuals in close relationships than between strangers. A first possibility is that the close relations had greater agreement on attractiveness preferences because they were more similar in age than the strangers. Second, genetic factors could underlie attractiveness preferences, in which case genetically related individuals such as siblings would be expected to have greater agreement. Third, people tend to associate with others who are similar to themselves, a phenomenon called homophily (Foster 2005). Shared preferences for facial attractiveness may be a component or a result of these similarities. Consistent with this possibility, Little et al (2006) have provided evidence that people with similar preferences for mate personality have similar opposite-sex facial-attractiveness preferences. Finally, people who have greater shared visual experience could have more similar mental representations of faces. There are several perceptual learning processes by which the particular 'diet' of faces that an individual views may cause chronic shifts in face representations and attractiveness preferences. Among these processes are sensory adaptation (Leopold et al 2001; Rhodes et al 2003), mere exposure or perceptual fluency (Peskin and Newell 2004; Winkielman et al 2006; Zajonc 1968), and familiar face overgeneralization (Zebrowitz et al 2007). Through one or more of these processes, two people spending a great deal of time with one another, often viewing and recognizing the same faces, could experience similar shifts in their attractiveness preferences.

Comparison of the different groups of close relations, and of pairs with more or less shared experience can provide evidence for or against each of these possible causes of greater shared preferences between close relations than strangers. In the case of age differences, we observed that among pairs of strangers the difference in age between the pair and their agreement on attractiveness was significantly correlated, accounting for a little more than 2% of the variance. A sub-sample of strangers who were age matched to the close relations had slightly more agreement than the complete sample of strangers, but still had significantly less agreement than the close relations. From this we can conclude that similarity of age is a small factor in agreement on attractiveness preferences, but that the effect of relationship remains even when the effect of age difference is factored out. Agreement was not higher between the siblings than between the spouses or friends, which is inconsistent with the possibility that similarity in

attractiveness preferences is primarily genetic in origin. A purely genetic account would also not predict the correlation between agreement on facial-attractiveness judgments and years spent in daily contact for siblings. The number of years that two people spent in daily contact with one another was weakly correlated with the agreement between their attractiveness preferences. This is consistent with the hypothesis that shared visual experience causes people's perceptions of beauty to become more similar, but also with the hypothesis that people with more similar preferences are likely to choose to spend more time together (ie homophily). Homophily might at first appear incapable of explaining this correlation with the siblings, who (unlike spouses) did not choose to become close relations. However, beyond a certain age they do choose whether to continue interacting with each other on a daily basis—ie they choose whether to *remain* close relations. To determine whether shared visual experience, homophily, or both factors contribute to the greater agreement on attractiveness preferences among close relations, it will be necessary to investigate situations where shared visual experience is disentangled from the decision of individuals to associate with one another.

This finding is an extension of cross-cultural studies which showed that, despite agreement on attractiveness preferences between cultures, there is greater agreement within cultures (Hönekopp 2006; Jones and Hill 1993). Here we find the typical within-culture agreement between unacquainted individuals, but considerably greater agreement between individuals who do share a close relationship. An intriguing possibility is that the factors underlying greater agreement among close relations also underlie the greater agreement among unacquainted members of the same culture. An advantage of the shared visual experience account of the current findings is that (unlike the homophily account) it could potentially also explain the cross-cultural findings.

While our results specifically address facial preferences, they may be true of visual preferences more generally, which also show significant individual variation (Davies 1939). Visual experience, specifically mere exposure, has been implicated in the maintenance of artistic canons (Cutting 2003), and it is likely that experience also shapes preferences across other aesthetic domains. Visual preferences and aesthetics may be socially organized in the way that we have seen here for facial-attractiveness preferences.

Acknowledgment. We thank Ann Breckencamp and Rexanne Grove for help recruiting participants.

References

Cutting J E, 2003 "Caillebotte, French Impressionism, and mere exposure" *Psychonomic Bulletin & Review* **10** 319 – 343

Davies M, 1939 "The general factor in correlations between persons" *British Journal of Psychology* **29** 404 – 421

Foster G, 2005 "Making friends: A nonexperimental analysis of social pair formation" *Human Relations* **58** 1443 – 1465

Good P I, 2000 *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses* 2nd edition (New York: Springer)

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

Jones D, Hill K, 1993 "Criteria of facial attractiveness in five populations" *Human Nature* **4** 271 – 296

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

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

Little A C, Burt D M, Perrett D I, 2006 "What is good is beautiful: face preference reflects desired personality" *Personality and Individual Differences* **41** 1107 – 1118

O'Toole A J, Deffenbacher K A, Valentin D, McKee K, Huff D, Abdi H, 1998 "The perception of face gender: the role of stimulus structure in recognition and classification" *Memory & Cognition* **26** 208 – 224

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

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
Perrett D I, Penton-Voak I, Little A C, Tiddeman B, Burt D M, Schmidt N, Oxley R, Kinlock N, Barret L, 2002 "Facial attractiveness judgements reflect learning of parental age characteristics" *Proceedings of the Royal Society of London, Series B* **269** 873 – 880
Peskin M, Newell F N, 2004, "Familiarity breeds attraction: Effects of exposure on the attractiveness of typical and distinctive faces" *Perception* **33** 147 – 157
Rhodes G, 2006 "The evolutionary psychology of facial beauty" *Annual Review of Psychology* **57** 199 – 226
Rhodes G, 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
Rosenthal R, 1991 *Meta-analytic Procedures for Social Research* 2nd edition (Newbury Park, CA: Sage)
Russell R, 2003 "Sex, beauty, and the relative luminance of facial features" *Perception* **32** 1093 – 1107
Thornhill R, Gangestad S W, 1993 "Human facial beauty: Averageness, symmetry, and parasite resistance" *Human Nature* **4** 237 – 269
Thornhill R, Gangestad S W, 1999 "Facial attractiveness" *Trends in Cognitive Sciences* **3** 452 – 460
Winkielman P, Halberstadt J B, Fazendeiro T, Catty S, 2006 "Prototypes are attractive because they are easy on the mind" *Psychological Science* **17** 799 – 806
Zajonc R B, 1968 "Attitudinal effects of mere exposure" *Journal of Personality and Social Psychology* **9** 1 – 27
Zebrowitz L, Bronstad P M, Lee H K, 2007 "The contribution of face familiarity to ingroup favoritism and stereotyping" *Social Cognition* **25** 303 – 335

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

PERCEPTION
VOLUME 36 2007

www.perceptionweb.com