**Facial resemblance between women’s partners and brothers: Supplementary Online Material 1**

**Introduction**

Our paper indicates that women select partners who resemble their brothers. Yet people in a family resemble each other; is it possible that brother-resemblance is merely an essential corollary of parent-resemblance?

Our paper provides reasons why sibling resemblance could shape partner choice independently of parent resemblance: siblings can be a useful additional point of reference (see Introduction), and siblings and parents constitute independent influences in other species (see Discussion).

In this Supplementary Material, we additionally present empirical data for the equivalence of brother and parental influence, by comparing effect sizes between our study and previous research.

**Methods**

We searched the literature for research studies that used the same methodology as the current study to compare the facial similarity of individuals’ partners and parents: that is, they asked raters to state the similarity of a target individual compared to four other individuals, consisting of one partner and three foils. We identified three such papers: Bereczkei, Gyuris, Koves, and Bernath (2002), Bereczkei, Gyuris, and Weisfeld (2004), and Marcinkowska and Rantala (2012). For the former two, we calculated data based on the published data. For the latter, we additionally obtained data from the first author.

**Results**

In general, only summary data are available from the three previously published studies. In addition, previous analyses used a number of different methods. Furthermore, none uses multilevel models or ordinal logistic regression and it is therefore not strictly possible to determine if they are robust to an analysis that incorporates additional variability that arises from differences in the stimuli or the raters. For these reasons, the effect size comparison is derived from the best available measure. In Bereczkei et al. (2002) and Bereczkei et al. (2004), this is the proportion of occasions on which the target face (i.e. sibling or parent) was rated as most similar to the partner. In the study reported by Marcinkowska and Rantala (2012), the authors kindly provided summary data (personal communication) that included the means of the similarity ratings, on a 1-10 scale, that were generated by raters who compared each target face to the partner and to the controls. On the assumption that the highest mean similarity rating should denote the face that would be ranked most similar, it is possible to extract the required information to compare the size of the effect between studies.

The effect size is best expressed as an odds ratio, which is the odds of a correct response divided by the odds of a correct response if guessing. These odds ratios (as well as the probability and odds of correct and guess responses) are set out in Table S1. These estimates should be interpreted with some caution as they do not take account of the information available were all the rankings to be combined (see our main Results section), and there is insufficient information to provide accurate interval estimates for each effect (which would depend on the random effects in each study).

**Discussion**

The degree of imprecision for each of these estimates is unknown (depending on the consistency of raters, the variability of the face stimuli, and between-study heterogeneity) but likely to be substantial. However, the range of effects is similar for all studies with the odds ratio of 1.36 in the present study slightly below the unweighted average of 1.58 (median 1.52). That is, we find similar resemblance between brother and partner as reported previously between parents and partner.

*Table S1: Summary of data from studies on the perceptual resemblance between an individual’s partner and that individual’s family members. Decimal numbers are rounded to 2 d.p.*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Study | Pairing | Raters *n* | Stimuli *n* | Correct match ranked 1st | Proportion above chance | Odds | Chance odds | OR |
| Bereczkei et al. (2002) | wife-mother | 52 | 32 | .41 | .16 | 0.70 | 0.33 | 2.09 |
| Marcinkowska and Rantala (2012) \* | husband-parent | 120 | 31 | .39 | .14 | 0.63 | 0.33 | 1.89 |
| Bereczkei et al. (2004) | husband-father | 242 | 26 | .38 | .13 | 0.61 | 0.33 | 1.82 |
| Bereczkei et al. (2002) | wife-husband | 52 | 32 | .34 | .09 | 0.51 | 0.33 | 1.53 |
| Marcinkowska and Rantala (2012) \* | wife-parent | 120 | 39 | .33 | .08 | 0.50 | 0.33 | 1.50 |
| **Current study** | **brother-partner** | **32** | **56** | **.31** | **.06** | **0.45** | **0.33** | **1.36** |
| Bereczkei et al. (2004) | wife-husband | 242 | 26 | .31 | .06 | 0.45 | 0.33 | 1.35 |
| Bereczkei et al. (2004) | husband-mother | 242 | 26 | .27 | .02 | 0.37 | 0.33 | 1.12 |

\* Proportion estimated from average rating for each stimulus (see text).

**References**

Bereczkei, T., Gyuris, P., Koves, P., & Bernath, L. (2002). Homogamy, genetic similarity, and imprinting; parental influence on mate choice preferences. Personality and Individual Differences, 33(5), 677-690.

Bereczkei, T., Gyuris, P., & Weisfeld, G. E. (2004). Sexual imprinting in human mate choice. Proceedings of the Royal Society of London, Series B: Biological Sciences, 271(1544), 1129-1134.

Marcinkowska, U. M., & Rantala, M. J. (2012). Sexual imprinting on facial traits of opposite-sex parents in humans. Evolutionary Psychology, 10(3), 621-630.

**Facial resemblance between women’s partners and brothers: Supplementary Online Material 2**

*Stimulus and sample size decisions, data collection termination rules, and data analysis phases*

We set out below the history of the data collection and analysis, together with an explanation of stimulus and rater sample sizes.

The study design was initially carried out conceptually in the form of separate undergraduate Honours projects by CS (who used 28 brother-boyfriend pairs of stimuli taken from images published online and 16 raters) and KR (who used 20 brother-boyfriend pairs of stimuli from the volunteer sample and 30 raters). There was evidence of brother-boyfriend resemblance, along with different patterns of resemblance comparing older and younger brothers, in both of these initial studies. However, sample sizes were lower than desired; the applications for ethical review of the projects prior to the commencement of data collection stated that CS was aiming at a total of 60 brother-boyfriend pairs of stimuli, while KR was aiming at a total of 40 brother-boyfriend pairs of stimuli. We were also concerned that CS, who collected stimuli from online sources, was aware of the hypotheses when she selected the stimuli, leading to a risk of subconscious bias towards images that could support the hypotheses. Accordingly, we elected to pursue the study idea by collecting additional stimuli, and combining these with the brother-boyfriend stimuli that we had collected directly from the participants that we recruited (i.e. excluding stimuli collected online). First, TS asked AN to help boost stimuli numbers to the maximum that could be reasonably obtained from social contacts and online searches during the period when AN was timetabled to work as a Research Assistant. We managed to accumulate a total of 56 brother-boyfriend pairs of stimuli; this substantially exceeds the number of stimuli in similar previous studies (39; see Table S1). Next, AN collected and analysed ratings data from 40 raters (half male; no sex differences found) in order to fulfil the requirements of her Research Assistant assessment. Again, her analyses also found evidence of brother-boyfriend resemblance, along with different patterns of resemblance comparing older and younger brothers. We were eventually unsatisfied with that study design, because the relative age of the brothers (younger or older) was not counterbalanced across the tableaux, and we decided it was necessary to obtain new ratings of a set of tableaux where older and younger brothers were presented in a counterbalanced design. To determine rater sample size for this new round of data collection, we examined the data collected by AN. These suggested that the variance of stimuli was about twice that of the raters, meaning that we should aim for about half the number of raters as stimuli. Further, the number of raters had to be divisible by four so that all four versions of the stimuli (see above) were rated the same number of times, and so we selected a final sample size of 32 raters. The data from this final round of data collection are reported in the accompanying paper. None of the ratings data collected by CS, KR or AN, and none of the online stimuli collected by CS, were used in the analysis reported.