Article

Preference for Facial Symmetry Depends on Study Design

Anthony J. Lee *[ID], Jessica K. De La Mare, Hannah R. Moore and Pamela C. Umeh

Faculty of Natural Sciences, Division of Psychology, University of Stirling, Stirling FK9 4LA, UK;
jed00087@students.stir.ac.uk (J.K.D.L.M.); ham00083@students.stir.ac.uk (H.R.M.);
pau00007@students.stir.ac.uk (P.C.U.)
* Correspondence: anthony.lee@stir.ac.uk

Abstract: Facial symmetry is purportedly attractive, though methods for measuring preference for facial symmetry vary between studies. Some studies have used a two-alternative forced choice (2AFC) task, while others have used a ratings task. How researchers manipulate facial symmetry also varies; some studies have used faces manipulated to be more (or perfectly) symmetrical, while others have used faces manipulated to be more asymmetrical. Here, across three studies, we evaluate and compare these different methods. In Studies 1 and 2 (N = 340 and 256, respectively), we compare facial symmetry preferences as measured by the 2AFC and ratings tasks. Across both studies, we consistently found a significant preference for facial symmetry when using the 2AFC task, but not with the ratings task. Additionally, correlations between facial symmetry preferences as measured by the two tasks were weak or showed no association. In Study 3, 159 participants rated the attractiveness of faces manipulated to be either symmetrical or more asymmetrical. The asymmetrical faces were rated as significantly less attractive compared to the original faces, while the difference in attractiveness ratings between the original and symmetrical versions was comparatively much smaller. These studies suggest that preference for facial symmetry depends greatly on the study design.

Keywords: mate preference; attraction; fluctuating asymmetry; faces; two alternative forced choice; ratings

Fluctuating asymmetry refers to the small deviations from perfect symmetry in bilaterally paired structures [1]. Previous research has suggested that facial symmetry influences attractiveness, where individuals with low levels of fluctuating asymmetry tend to be reported as more attractive [2–4]. There are two main theories proposed to explain this preference for facial symmetry. The first theory, from an evolutionary perspective, is that facial symmetry is an honest cue of health and/or genetic quality. Individuals with low levels of fluctuating asymmetry are thought to have better resistance to disease which has allowed them to maintain symmetry despite environmental and genetic pressures [5,6]. As such, when selecting a mate, it is evolutionarily advantageous for individuals to search for a sexual partner that possesses this cue to heritable health, as these partners are more likely to produce high quality offspring that will survive and reproduce themselves. The second theory, from a visual perception perspective, is that humans may simply have a perceptual bias for symmetrical stimuli as they are easier for the visual system to process [7,8].

Regardless of the mechanism for facial symmetry preferences, previous research has typically used two main methods when investigating this effect; these are the two-alternative forced choice (2AFC) task and the ratings task. In the 2AFC task, participants are presented with two versions of the same face and asked to indicate which they find more attractive. Usually, one of these versions has been manipulated to be highly or perfectly symmetrical while the other is either left in its original, unaltered form (e.g., [9–11]) or has been manipulated to exaggerate its asymmetries (e.g., [3,12,13]). For the ratings task, participants are presented with faces sequentially and asked to rate how attractive they find each face on a numeric scale (e.g., a 7-point scale where 1 = unattractive, and 7 = attractive). In this task, faces are either manipulated to be more/less symmetrical (e.g., [14,15]), or

natural variation in facial symmetry is measured in unmanipulated faces (e.g., [16–18]). It is often assumed that both the 2AFC and ratings methods measure the same construct of preference for facial symmetry; however, recent work has questioned this assumption. Jones and Jaeger [19] reported divergent results when using both tasks, where a statistically significant preference for perfectly symmetrical faces was found when using a 2AFC paradigm, but no significant preference was found when participants completed a ratings task. This is consistent with a recent, large scale study using a ratings task with naturally varying faces that also failed to find a preference for facial symmetry [20].

There are numerous reasons to suspect divergent results from the 2AFC and ratings tasks. First, arguably, the 2AFC task lacks ecological validity. In reality, individuals do not select a mate by considering two nearly identical stimuli that only differ on one dimension. Instead, people make attractiveness judgements by combining many different dimensions. Scott, Clark, Boothroyd and Penton-Voak [21] suggested that as only one dimension (in their case, facial masculinity) varies between the two faces in a 2AFC task, then attention is drawn to this particular dimension that might otherwise be ignored. As such, the 2AFC task may find effects that do not influence mate choice judgements in reality. Relatedly, Lewis [22] suggested that the 2AFC task may instead measure the ability of participants to detect asymmetry in faces rather than a preference for symmetrical faces per se; this is an important distinction as ability to detect asymmetry does not necessarily indicate a preference for symmetry [23,24].

Alternatively, perhaps recent null results found with the ratings task is because facial symmetry effects are too subtle when using naturalistic faces as stimuli. Indeed, as multiple traits likely influence attractiveness ratings in a ratings task, effect sizes for facial symmetry preferences are likely to be small, and thus require more statistical power to detect. Additionally, morphometric techniques commonly used to quantify facial symmetry in naturalistic faces may not be reliable/valid. Asymmetry scores are typically calculated by summing deviations from analogous landmarks on the left and right side of the face. However, slight changes to image properties unrelated to fluctuating asymmetry (e.g., slight offsets in the orientation of the face to the camera) can drastically change these scores. If individuals can account for external factors such as head orientation when evaluating facial symmetry, then calculating asymmetry in this way from a 2D image may not be appropriate when assessing preference for facial symmetry.

In addition, little consideration is given on how symmetry is manipulated in facial images used in either the 2AFC or ratings task. As noted above, some studies investigate facial symmetry preferences by comparing original faces (with naturally occurring asymmetry) with perfectly symmetrical faces, while others use faces that have been manipulated to be less symmetrical. Often effects found from using these different methods are interpreted equivalently; however, we could expect a larger effect when using faces manipulated to be less symmetrical. Previous work has suggested that facial symmetry preferences may actually reflect an aversion to asymmetry rather than a preference for symmetry per se [25]. Facial symmetry preference could also follow a threshold model, where perception is only sensitive to large asymmetries, or minor deviations from symmetry are tolerated [25–27].

Here, across three studies, we evaluate and compare different methods used to assess preference for facial symmetry. In Studies 1 and 2, we compare facial symmetry preferences as measured by the 2AFC and ratings task. We would expect that if the 2AFC and the ratings tasks measure the same concept (i.e., preference for facial symmetry), then we should find a positive preference for facial symmetry in both tasks, as well as a strong, positive correlation between the two scores. A weak, or no association between facial symmetry preference scores may suggest the two tasks measure separate constructs. In Study 3, we compare how different methods of manipulating symmetry in facial images influences attractiveness ratings. If symmetry influences attractiveness linearly, then we would expect the same effect size when comparing between faces manipulated to be less symmetrical vs. those manipulated to be more symmetrical. Alternatively, if people have an aversion to asymmetry, or if symmetry preference follow a threshold model, then faces

that have been manipulated to be less symmetrical would have a greater influence on attractiveness ratings compared to those with higher facial symmetry.

1. Study 1

1.1. Methods

1.1.1. Participants

A total of 340 online volunteers (83 males, 257 females) were recruited from social media websites (M = 25.78 years, SD = 11.14 years). Participation was conditional on identifying as either male or female and as heterosexual. All participants were over 18 years old. There was no incentive offered to participate.

1.1.2. Measures and Procedure

Participants completed both a 2AFC and a ratings task previously used to measure preference for facial symmetry. Both tasks used White faces randomly selected from the Chicago Face Database [28] (age range = 17.55 years to 50.43 years). Each face identity was only presented to each participant once (i.e., the 20 faces included in the 2AFC task were different to the 20 faces included in the ratings task). Both measures were included in a larger survey on mate preferences and presented to participants in a random order.

Two-Alternative Forced Choice Task

Participants were presented with two versions of the same face side-by-side. One version of the face was the original face (with naturally occurring asymmetry), while the other was manipulated to be perfectly symmetrical in shape. Symmetry manipulations were done in the Webmorph program [29] according to the standard procedures [9]. Faces were manually delineated, then the face shape was morphed with a vertically reflected (mirror) version of the face; this effectively symmetrises the face shape while preserving colour and texture information (see Figure 1. for example). Participants were asked to select which face they found the most attractive. Participants rated 20 opposite-sex faces. The order of the faces presented to participants was randomised, and the position of the symmetrical face (left or right) was randomised.

Figure 1. Examples of faces manipulated to be perfectly symmetrical (**left**), unmanipulated (**middle**), and manipulated to be more asymmetrical (**right**; only used in Studies 2 and 3).

Rating Task

Participants were presented with a single face and asked to rate how attractive they found it on a 9-point scale (1 = Very Unattractive, 9 = Very Attractive). Participants were presented with 20 opposite-sex faces. For each face, facial asymmetry was measured using morphometric techniques following Komori et al. [16] and Holzleitner et al. [20]. This

involved delineating each face on 132 landmarks and calculating the Euclidian distance between the corresponding left-right Procrustes-aligned landmarks. This produced a single score for each face, representing the level of asymmetry of that face (i.e., the more asymmetrical the face, the greater the deviations between left-right corresponding landmarks, resulting in a larger asymmetry score). To aid interpretation, these scores were reverse-coded such that higher scores represented more symmetrical faces. Preference for symmetry is then assessed by comparing levels of symmetry with ratings of attractiveness provided by participants.

1.1.3. Statistical Analysis

All analyses were conducted in the R statistical software [30] using the lme4 [31] and lmerTest [32] packages. Separate analyses were conducted for males rating female faces, and females rating male faces.

Analysis 1. To assess preference for facial symmetry in the 2AFC task, we conducted a binomial mixed effects model with participants choice being the outcome variable (0 = original face chosen, 1 = symmetrical face chosen). As such, a larger intercept indicates a greater preference for facial symmetry. Random intercepts were specified for each participant and each face identity, following suggestions by DeBruine and Barr [33].

Analysis 2. To assess preference for facial symmetry in the ratings task, we conducted a linear mixed effects model with participant ratings of attractiveness for each face as the outcome variable. The z-standardised asymmetry score for each face was included as the predictor, such that a positive association represents a greater preference for symmetry. Random effects for each participant and face identity were specified, with random slopes specified maximally [34].

Analysis 3. To compare preference for facial symmetry as measured by the two tasks, for each participant and for each task, we calculated two separate scores that represented that participant's preference for facial symmetry as measured by that task. For the 2AFC task, the first score involved calculating the total proportion of trials the symmetrical face was chosen over the original face. For the ratings task, the first score involved calculating a correlation for each participant between their ratings of attractiveness and the morphometric symmetry score for each face (a greater preference for symmetry would produce a larger, positive, correlation coefficient). For the second scores for the 2AFC and ratings tasks, we used the estimated random effects for each participant from Analysis 1 and 2, respectively. Random effects for participants represent the variation in the estimated fixed effects that can be attributed to different participants; as such, participants with a more positive intercept in Analysis 1 are showing a greater preference for facial symmetry compared to the rest of the sample. Similarly, in Analysis 2, participants with a more positive slope between symmetry and attractiveness ratings are showing a greater preference for facial symmetry compared to the rest of the sample. Since both Analysis 1 and 2 also include random effects of face identity, this method has the benefit of accounting for variance in the outcome variable associated with each face identity (e.g., in the ratings task, accounting for variation in attractiveness ratings of the faces that are not due to symmetry). To compare symmetry preferences as measured by the two tasks, we conducted correlations between all four symmetry preference scores.

1.2. Results

For Analysis 1 and 2, fixed effects are reported here; for full model results, including estimated random effects, see the Supplementary Materials.

1.2.1. Analysis 1: Preference for Facial Symmetry as Measured by the 2AFC Task

For females rating male faces, the estimated fixed intercept was significant (estimate = 0.28, std. error = 0.05, z = 5.90, $p < 0.001$), which equates to the symmetrical face being chosen in 56.99% trials. For males rating female faces, while the direction

was the same, this association was not significant (estimate = 0.14, std. error = 0.07, $z = 1.85$, $p = 0.064$), and equates to the symmetrical face being chosen 53.44% of trials. See Figure 2.
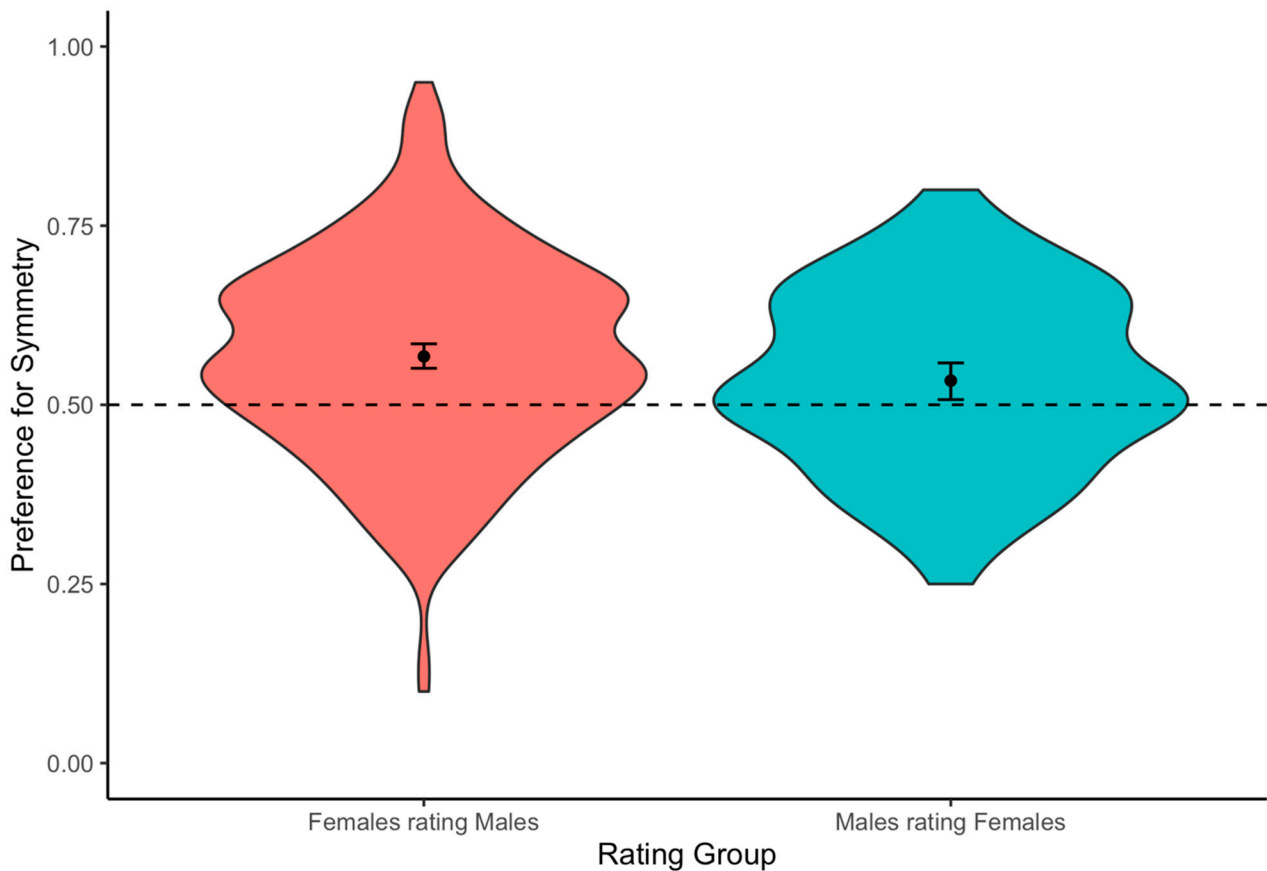

Figure 2. Proportion of times the perfectly symmetrical face was chosen over the original face in the 2AFC task. Error bars represent 95% confidence intervals.

1.2.2. Analysis 2: Preference for Facial Symmetry as Measured by the Ratings Task

For both males rating female faces and females rating male faces, the associations between facial symmetry and attractiveness ratings were non-significant (estimate = −0.28, std. error = 0.31, $t(18.16) = -0.90$, $p = 0.378$, and estimate = 0.07, std. error = 0.17, $t(18.02) = 0.42$, $p = 0.678$, respectively). See Figure 3.

1.2.3. Analysis 3: Congruency between the 2AFC Task and Ratings Task

Correlations of symmetry preference scores from the 2AFC and ratings tasks are presented in Table 1. As expected, for both males and females, symmetry preference scores derived from the same task were strongly correlated with each other. For males rating female faces, there were no significant correlations between symmetry preference as measured by the 2AFC and ratings tasks. However, for females rating male faces, symmetry preferences between the two tasks were significantly correlated, regardless of the measurement used (see Figure 4 for an example).
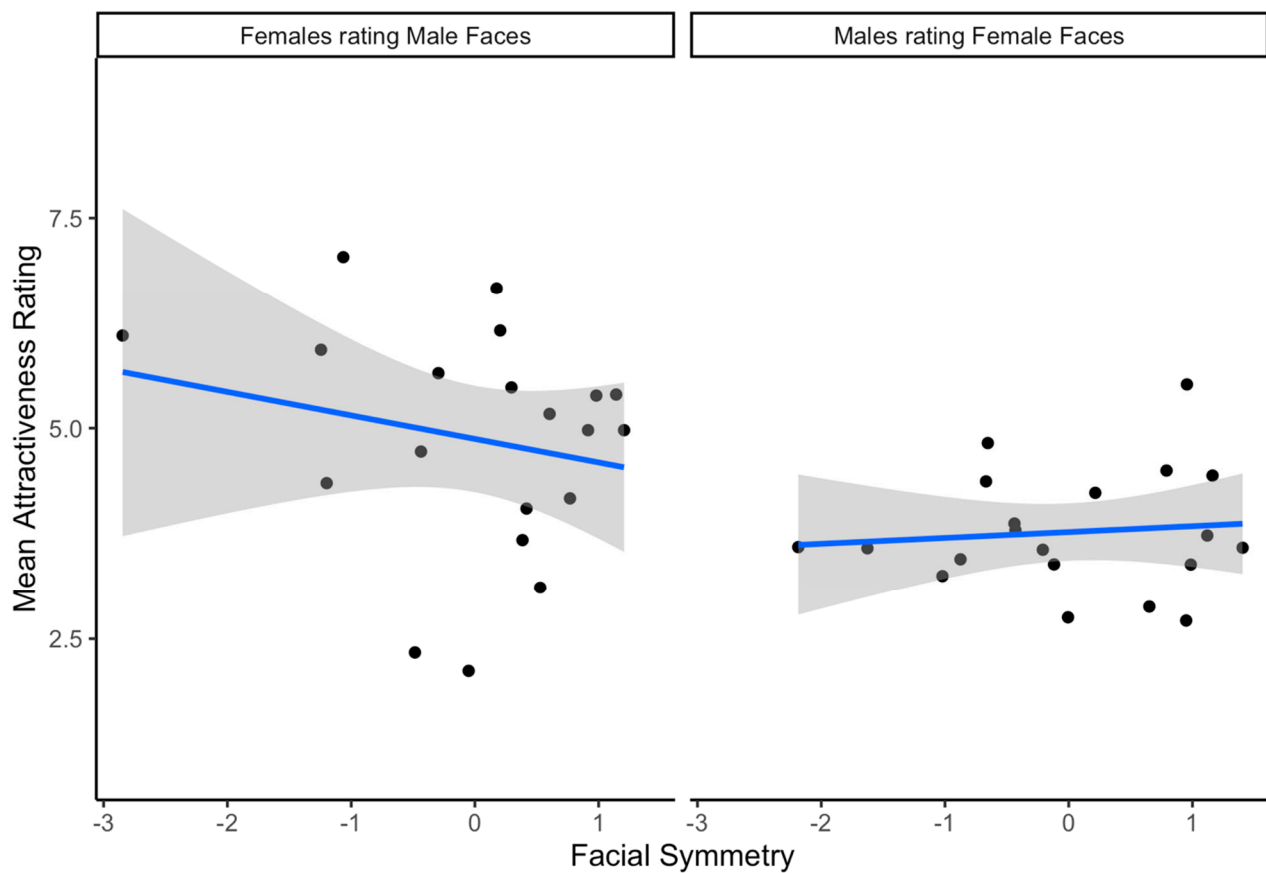
Figure 3. Associations between facial symmetry and mean attractiveness ratings for the ratings task. The grey area represents the 95% confidence interval.

Table 1. Correlation coefficients (r) between symmetry preferences measured by the 2AFC and ratings tasks. Results from males rating female faces (N = 83) are on the bottom triangle, while females rating male faces (N = 257) are on the top triangle.

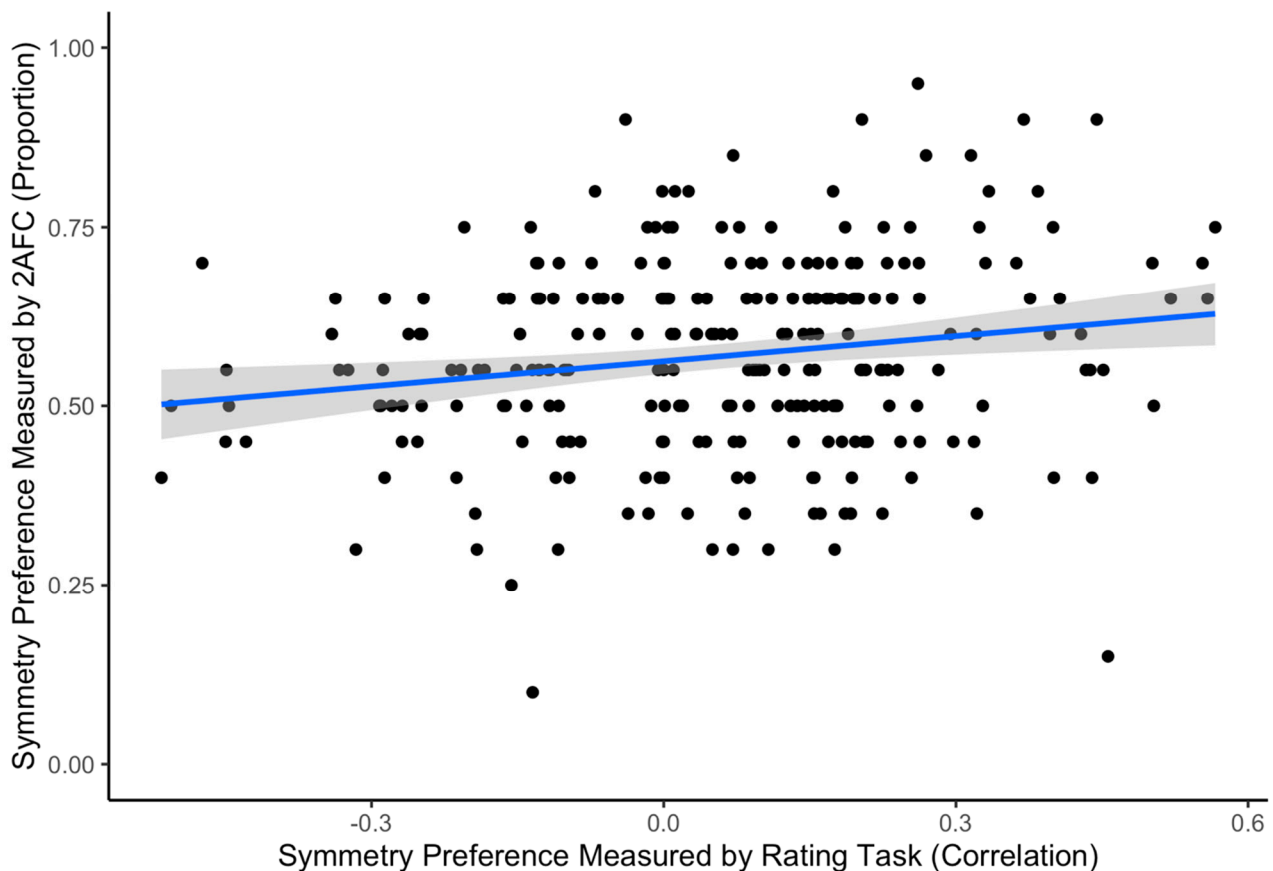|	2AFC: Proportion	2AFC: Random Effect?	Rating: Correlation	Rating: Random Effects
2AFC: Proportion		0.99 ($p < 0.001$)	0.18 ($p = 0.004$)	0.16 ($p = 0.010$)
2AFC: Random Effect	0.99 ($p < 0.001$)		0.18 ($p = 0.005$)	0.16 ($p = 0.011$)
Rating: Correlation	−0.05 ($p = 0.656$)	−0.04 ($p = 0.657$)		0.90 ($p < 0.001$)
Rating: Random Effects	−0.07 ($p = 0.503$)	−0.07 ($p = 0.503$)	0.84 ($p < 0.001$)	

Figure 4. The correlation between symmetry preference measured by the 2AFC task (using proportions) and ratings task (using correlations) for females rating male faces. The grey area represents the 95% confidence interval.

1.3. Discussion

Consistent with Jones and Jaeger [19], we detected a preference for facial symmetry when using a 2AFC task (at least for females rating male faces), but not in the ratings task. We also detected a significant positive correlation between symmetry preference as measured by the two tasks; however, this was only statistically significant for females rating male faces. Additionally, the effect size for the correlation was small, which is not consistent with the assumption that both tasks measure the same construct. While it is possible that the lack of significant correlation for males rating female faces could be due to a lack of statistical power based on the smaller sample of male participants, we note that the estimated correlation coefficients are all close to zero.

In Study 1, the symmetry manipulation for the faces in the 2AFC task involved comparing the original faces with a perfectly symmetrical version. However, some studies have instead used facial images that have been manipulated to be more asymmetrical [3,12,13]. In Study 2, we assess whether differences in facial symmetry manipulation methodology influences results.

2. Study 2

2.1. Methods

2.1.1. Participants

A total of 256 online volunteers (87 males, 169 females) were recruited from social media websites (M = 21.57 years, SD = 4.81 years). All participants reported identifying as either male or female and as heterosexual. All participants were over 18 years old. There was no incentive offered to participate.

2.1.2. Measures, Procedures, and Statistical Analysis

The measures, procedures, and statistical analyses were identical to Study 1, with two exceptions. First, the faces that were randomly chosen from the Chicago Face Database [28] for the 2AFC and ratings tasks were different to those used in Study 1 (age range = 17.55 years to 50.43 years). This was to test whether results were generalisable to a different set of faces. Second, comparisons in the 2AFC task were made between the original face and a face manipulated to be more asymmetrical. This manipulation was done by computing the linear differences between the original face and perfectly symmetrical versions, and applying those differences to the original face [12]. See Figure 1 for an example. Essentially, this exaggerates the asymmetry for each face identity and the difference between this and the original face is mathematically the same as the difference between the original face and a perfectly symmetrical face.

2.2. Results

As with Study 1, only fixed effects are reported here; for full model results, including estimated random effects, see the Supplementary Materials.

2.2.1. Analysis 1: Preference for Facial Symmetry as Measured by the 2AFC Task

For both males rating female faces (estimate = 2.76, std. error = 0.27, z = 10.34, $p < 0.001$), and females rating male faces (estimate = 2.14, std. error = 0.13, z = 16.03, $p < 0.001$), we found a significant preference for facial symmetry. This equated to the symmetrical face being chosen 94.04% and 89.45% of trials, respectively. See Figure 5.
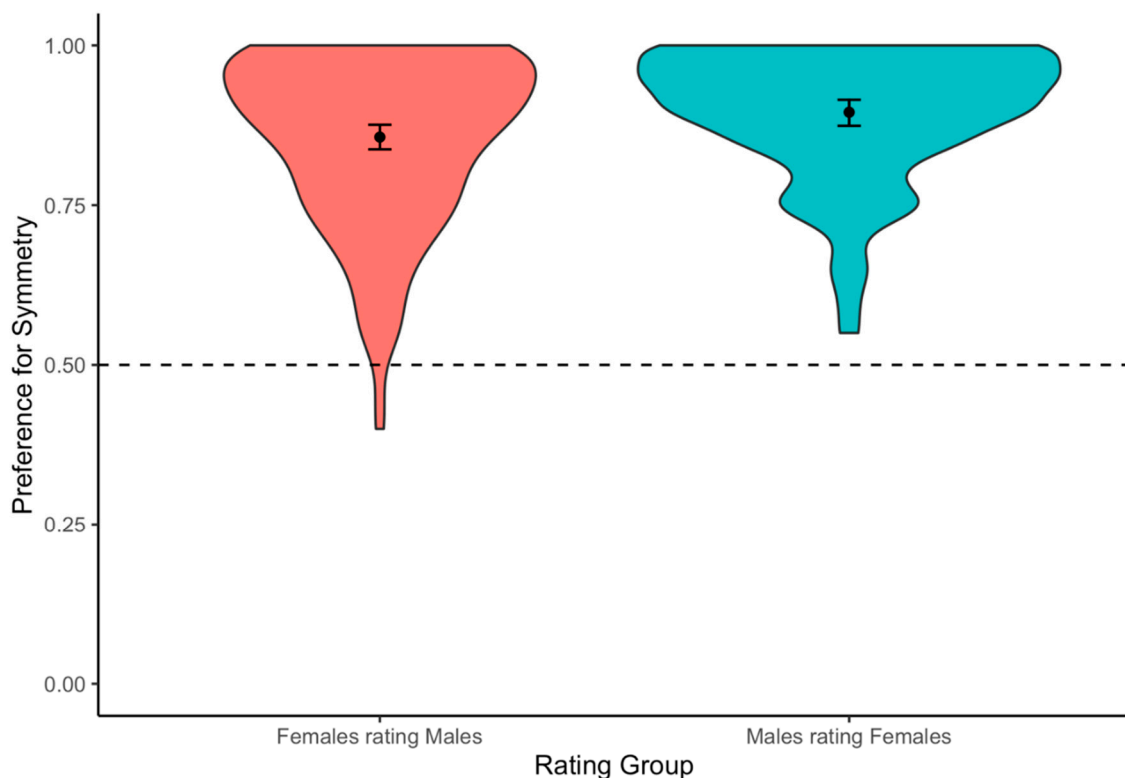

Figure 5. Proportion of times the original face was chosen over the asymmetrical face for the 2AFC task. The error bars represent the 95% confidence interval.

2.2.2. Analysis 2: Preference for Facial Symmetry as Measured by the Ratings Task

Consistent with Study 1, for both males rating female faces and females rating male faces, we found non-significant associations between facial symmetry and attractiveness

ratings (estimate = −0.04, std. error = 0.31, $t(18.00)$ = −0.12, p = 0.904, and estimate = 0.40, std. error = 0.26, $t(18.03)$ = 1.51, p = 0.149, respectively). See Figure 6.
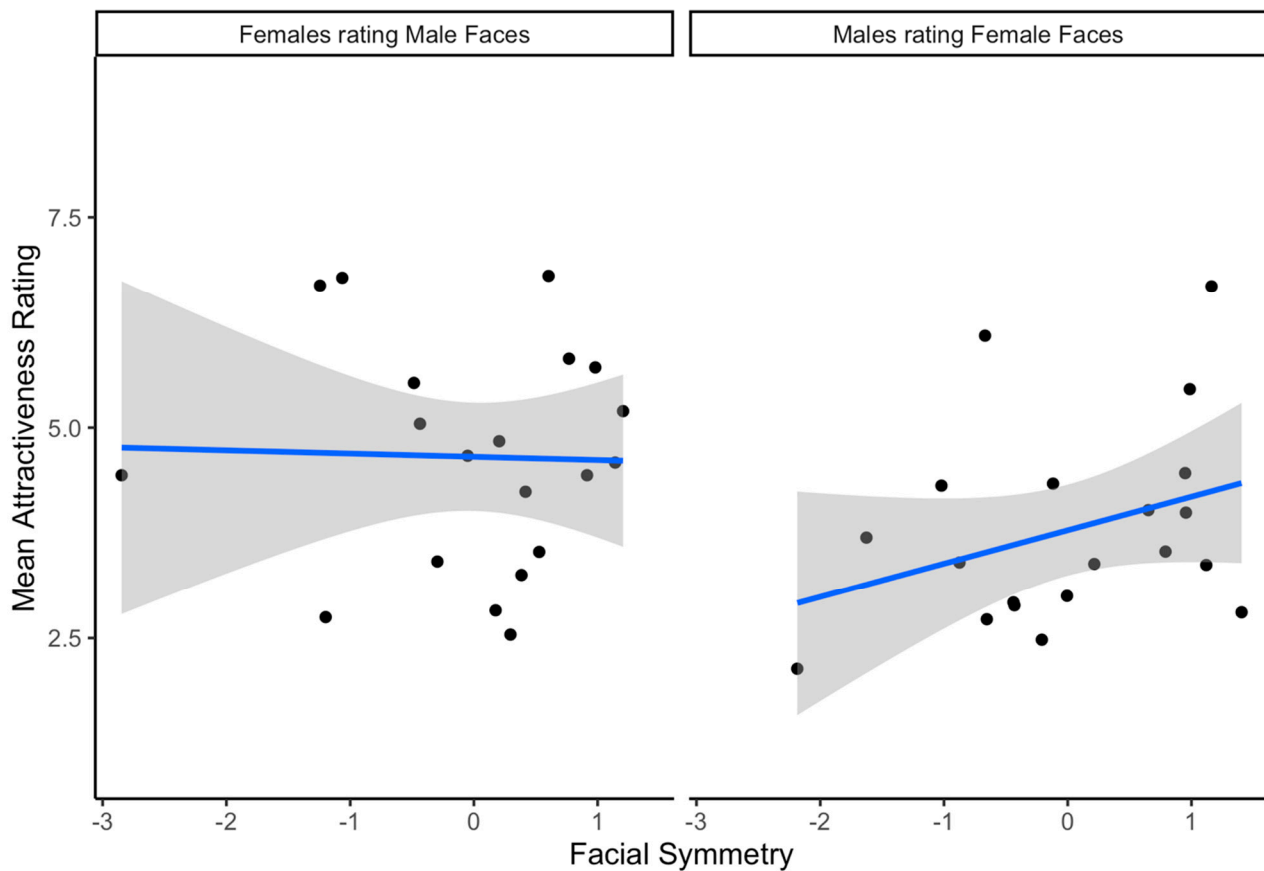

Figure 6. Associations between facial symmetry and mean ratings of attractiveness for the ratings task. The grey area represents the 95% confidence interval.

2.2.3. Analysis 3: Congruency between the 2AFC Task and Ratings Task

Correlations of symmetry preference scores from the 2AFC and ratings tasks are presented in Table 2. As with Study 1, for males rating female faces, there were no significant correlations between symmetry preference as measured by the 2AFC or ratings task. However, contrary to Study 1, symmetry preference scores between the two tasks were also non-significant for females rating male faces.

Table 2. Correlation coefficients (r) between symmetry preferences measured by the 2AFC and ratings tasks. Results from males rating female faces (N = 87) are on the bottom triangle, while females rating male faces (N = 169) are on the top triangle.

	2AFC: Proportion	2AFC: Random Effects	Rating: Correlation	Rating: Random Effects
2AFC: Proportion		0.99 ($p < 0.001$)	0.14 ($p = 0.076$)	0.06 ($p = 0.463$)
2AFC: Random Effects	0.99 ($p < 0.001$)		0.15 ($p = 0.052$)	0.06 ($p = 0.424$)
Rating: Correlation	0.09 ($p = 0.434$)	0.09 ($p = 0.360$)		0.37 ($p < 0.001$)
Rating: Random Effects	0.12 ($p = 0.271$)	0.14 ($p = 0.187$)	−0.00 ($p = 0.979$)	

2.3. Discussion

Overall, a strong preference was found for facial symmetry in the 2AFC, despite the magnitude of the facial symmetry manipulation being identical to that of Study 1. Given a strong effect was found here, while only a small effect was found in Study 1, this supports the notion that symmetry preferences may not be linear, and instead effects are stronger when original faces are paired with more asymmetrical versions compared to symmetrical versions. Consistent with Study 1, no symmetry preference was found in the ratings task.

Contrary to findings in Study 1, we did not find any significant association between symmetry preferences as measured by the 2AFC and the ratings tasks. However, we note that some of the estimated correlation coefficients were of similar magnitude to that found in Study 1; with a larger sample size (and assuming estimated effects remain the same), it is possible that these positive correlations could become significant. Additionally, preference for facial symmetry found in the 2AFC task was overall very high, which likely restricts the variation in symmetry preference scores as measured by this task. In turn, this could obscure our ability to detect a significant correlation in symmetry preferences between the two tasks if it exists. Regardless, if facial symmetry preferences are indeed correlated between the two tasks, the effect is likely to be small, further supporting the notion that both tasks measure different constructs.

Together, results from Studies 1 and 2 would indicate that when manipulating facial symmetry in images, the type of manipulation can have a drastic influence on effects. To investigate this further, in Study 3, participants rated faces for attractiveness that had been manipulated to be perfectly symmetrical or more asymmetrical, as well as the original, unmanipulated face. If symmetry preference effects are stronger between original and asymmetrical versions, then we could expect a significant difference in attractiveness between the two, but not between the original and symmetrical versions.

3. Study 3

3.1. Methods

3.1.1. Participants

A total of 159 online volunteers (78 males, 81 females) were recruited from social media websites or from Prolific.co (*M* = 28.37 years, *SD* = 8.82 years). All participants reported identifying as either male or female and as heterosexual. All participants were over 18 years old. Online volunteers did not receive incentives to participant, while participants recruited via Prolific.co received payment.

3.1.2. Measures and Procedure

Participants completed a ratings task, where faces were presented to participants sequentially and participants were asked to rate how attractive they found each face on a 7-point scale (1 = Very Unattractive, 7 = Very Attractive). Eighty faces (40 males and 40 females, age range = 18.48 years to 40.07 years) were randomly chosen from the Chicago Face Database [28] and ranged in ethnicity (Asian, Black, Latino and White). Three versions of each face were shown to participants, the original version, a version that had been manipulated to be perfectly symmetrical (as described in Study 1), and a version that had been manipulated to be more asymmetrical (as described in Study 2). Participants rated all opposite-sex faces, which resulted in each participant rating 120 faces. The order that images were shown to each participant was randomised.

3.1.3. Statistical Analysis

Data were analysed using linear mixed effects modelling in the R statistical software [30] using the lme4 [31] and lmerTest [32] packages. Separate analyses were conducted for males rating female faces and females rating male faces. In both models, the outcome variable was attractiveness ratings given by participants, while the predictor was the symmetry condition of the face (symmetrical, original, or asymmetrical). Symmetry condition was dummy coded, such that the intercept represented the mean attractiveness

rating given to the original version of faces, and estimated fixed effects represented the change in attractiveness rating for the symmetrical and asymmetrical versions of the face. Random intercepts were specified for participants and face identity, and random slopes were specified maximally [34].

3.2. Results

The estimated fixed effects for both models (males rating female faces and females rating male faces) are reported in Table 3 (for full model results, including estimated random effects, see the Supplementary Materials). For both males rating female faces and females rating male faces, the asymmetrical version ($M = 2.91$ and 2.74, respectively) was rated as significantly less attractive compared to the original version ($M = 3.32$ and 3.13, respectively). For males rating female faces, there was also a significant effect of symmetrical version ($M = 3.38$), such that the symmetrical version was rated as more attractive compared to the original. However, there was no significant difference between the original and symmetrical versions for females rating male faces ($M = 3.15$). See Figure 7.

Table 3. Fixed effects estimates from the linear mixed effects models predicting attractiveness ratings.

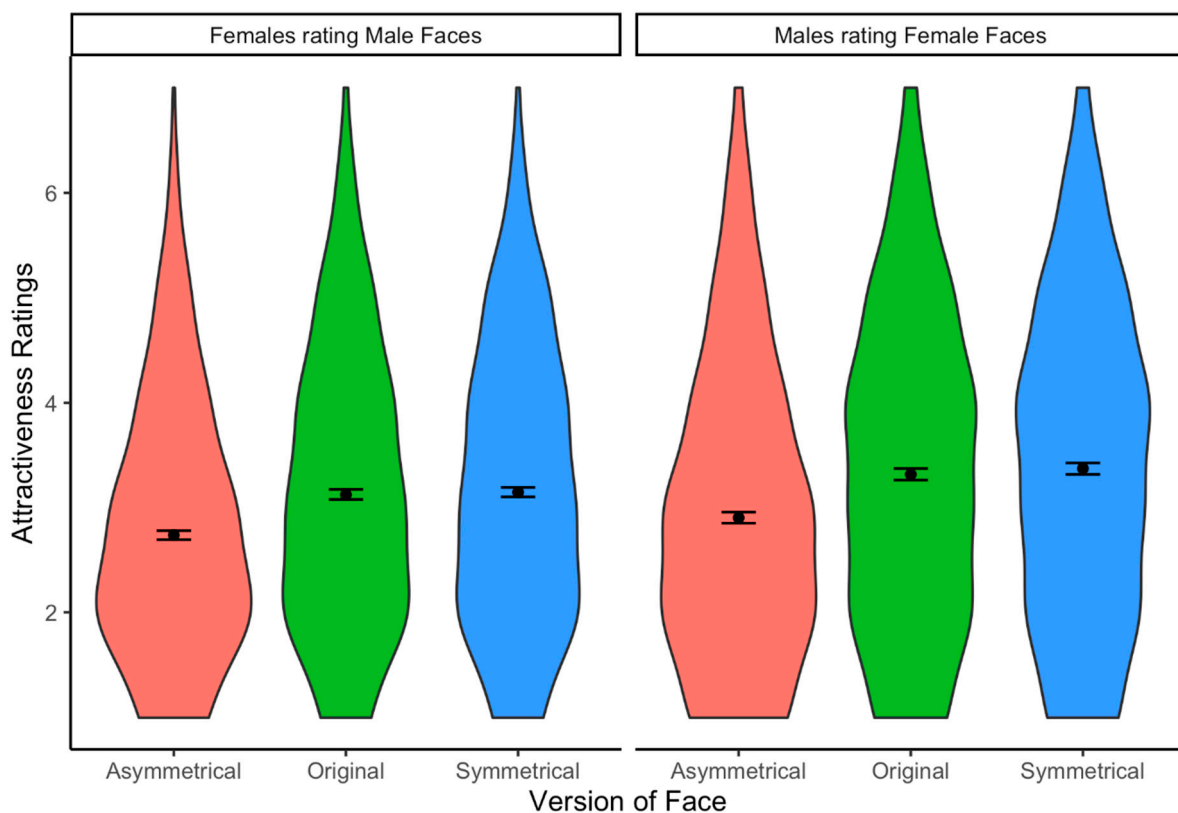|	Males Rating Female Faces			Females Rating Male Faces		
	Estimate (Std. Error)	**t (Approx. df)**	**p-Value**	**Estimate (Std. Error)**	**t (Approx. df)**	**p-Value**
Intercept	3.32 (0.15)	21.54 (92.54)	<0.001	3.13 (0.12)	27.24 (117.30)	<0.001
Symmetry Version	0.06 (0.03)	2.16 (145.19)	0.033	0.02 (0.02)	1.02 (47.23)	0.314
Asymmetry Version	−0.42 (0.05)	−8.79 (55.10)	<0.001	−0.38 (0.05)	−7.74 (79.26)	<0.001

Figure 7. Attractiveness ratings given for different versions of faces that have been manipulated on facial symmetry. The error bars represent 95% confidence intervals.

3.3. Discussion

For both males rating female faces and females rating male faces, we found a significant difference in attractiveness ratings between the original and asymmetrical versions. In comparison, the difference in attractiveness ratings between the original and perfectly symmetrical version is much smaller (or non-significant), despite the magnitude of the manipulation being mathematically identical to that between the asymmetrical and original versions. This would suggest that the influence of facial symmetry is non-linear, where aversions to facial asymmetry is much stronger compared to preferences for symmetrical features.

Since each participant saw multiple versions of the same face, it is possible that participants may have ascertained that facial symmetry was the trait of interest. If this were the case, then we would expect to find potentially exaggerated symmetry preference effects. Despite this, our findings suggest that any effect difference in attractiveness ratings between the original and perfectly symmetrical face is small, perhaps indicating these differences may not be important for attractiveness judgements.

4. General Discussion

Across three studies, we compared different methodologies used to assess preference for facial symmetry. Overall, findings suggest that results are dependent on the type of task used, as well as how facial symmetry is manipulated in stimuli.

Across Studies 1 and 2, when using the 2AFC task, we consistently found a significant preference for facial symmetry, in line with previous findings [2–4]. However, we found no association between facial symmetry and attractiveness ratings for the ratings task in both studies. These divergent findings could be explained in a few ways. First, it is possible that any preference found by the 2AFC is an artefact of the task. For instance, results from the 2AFC could be due to comparison effects, where preferences are only found when all other factors remain constant [21]. Relatedly, results could be explained by demand characteristics, where participants are easily able to determine the trait in question from a 2AFC task and respond in a way they perceive is consistent with the hypothesis. Alternatively, the lack of an association between facial symmetry and attractiveness ratings in a ratings task could be due to reduced power, as responses in a ratings task are likely influenced by many factors external unrelated to facial symmetry. Indeed, data-driven analyses have indicated that any effect of symmetry on attractiveness is likely to be small in comparison to other traits [20].

In both Studies 1 and 2, there was no consistent association between facial symmetry preference as measured by the 2AFC and that measured by the ratings task. In the instances where a significant correlation was found, the effect size estimates were small. As such, it is unlikely that both tasks measure the same construct (being preference for facial symmetry) as previously assumed, and instead the two tasks may measure separate constructs [19]. For instance, it has been recently suggested that the 2AFC may measure the ability to distinguish differences in a trait (rather than preferences for that trait; [22]). Alternatively, there may be issues related to using the ratings task, such as requiring more statistical power to detect an effect. In a similar study, DeBruine [35] examined women's preference for facial masculinity using both a 2AFC task and ratings task. Interestingly, DeBruine [35] found a large positive correlation between masculinity preferences as measured by the 2AFC and ratings tasks, but only masculinity preferences using the ratings task scores were statistically significant. As such, findings we report here may be specific to facial symmetry.

Results across the three studies also suggest that findings are dependent on how symmetry is manipulated in facial images. In Studies 1 and 3, preferences between original and perfectly symmetrical faces were either small or negligible. However, in Studies 2 and 3, attractiveness ratings between the original face and asymmetrical versions were consistently detected. This is despite the magnitude of both manipulations being mathematically the same. Theoretically, this could suggest there is a stronger aversion to the asymmetrical versions of faces more so than a preference for symmetrical faces. These

findings could be explained if facial symmetry preference follows a threshold model, where subtle asymmetries are either not important for attractiveness judgements or not perceived, while only larger deviations from symmetry are important. More generally, these results raise questions about the importance of facial symmetry on attractiveness judgements, as any preference for facial symmetry within the naturally varying range is likely to be small. Indeed, if only large numbers of both raters and stimuli are needed to detect a symmetry preference to a satisfactory level of certainty, this might suggest that symmetry may play a negligible role, if any, on facial attraction. While future research could estimate precisely how small the effect is, it may be more fruitful investing resources investigating other more important factors. Methodologically, our results suggest that more careful consideration is needed when manipulating facial symmetry in stimuli.

There are a number of issues when measuring facial symmetry preferences that are yet to be addressed. First, both Studies 1 and 2 used morphometric measurements to quantify facial asymmetry. As these scores can be greatly influenced by factors not related to fluctuating asymmetry (e.g., head orientation), this raises questions regarding the validity/reliability of this method, particularly if individuals subconsciously account for these external factors when evaluating facial symmetry. Inaccuracies in this measure may obscure any true correlation between symmetry preferences measured using both tasks. Second, our studies used only 20 face identities in each task for both Studies 1 and 2, and 40 face identities in Study 3. Lewis [36], Pollet and Little [37], and DeBruine and Barr [33] outline potential problems with using a small number of facial stimuli when investigate face perception. In particular, results may have issues generalising to other stimuli sets, though this limitation is somewhat mitigated in our studies, as each study used a different set of face identities, and data were analysed using mixed effects modelling with crossed random effects. Additionally, a small sample of stimuli reduces the reliability of the data, potentially inflating the false-positive rate. Future research should investigate facial symmetry preferences with larger stimuli set sizes. Finally, here we only investigated symmetry preferences for opposite-sex faces as this has been the focus of previous research. It is currently unclear whether results would replicate when considering symmetry preferences for same-sex faces.

In conclusion, our studies provide insight into the methods used to assess preference for facial symmetry. Namely, our studies support recent suggestions that the 2AFC task and ratings task may assess separate constructs, or perhaps that results can depend on how symmetry is manipulated in facial images. This research suggests that how we measure preferences for facial traits requires further consideration.

References

1. Van Valen, L. A study of fluctuating asymmetry. *Evolution* **1962**, *16*, 125–142. [CrossRef]
2. Grammer, K.; Thornhill, R. Human (Homo sapiens) facial attractivness and sexual selection: The role of symmetry and averageness. *J. Comp. Psychol.* **1994**, *108*, 233–242. [CrossRef] [PubMed]
3. Rhodes, G.; Proffitt, F.; Grady, J.M.; Sumich, A. Facial symmetry and the perception of beauty. *Psychon. Bull. Rev.* **1998**, *5*, 659–669. [CrossRef]
4. Little, A.C.; Jones, B.C. Evidence against perceptual bias views for symmetry preferences in human faces. *Proc. R. Soc. B Biol. Sci.* **2003**, *270*, 1759–1763. [CrossRef] [PubMed]
5. Møller, A.P.; Swaddle, J.P. *Asymmetry, Developmental Stability, and Evolution*; Oxford University Press: Oxford, UK, 1997.
6. Thornhill, R.; Gangestad, S.W. Facial attractivness. *Trends Cogn. Sci.* **1999**, *3*, 452–460. [CrossRef]
7. Little, A.C. Domain Specificity in Human Symmetry Preferences: Symmetry is Most Pleasant When Looking at Human Faces. *Symmetry* **2014**, *6*, 222–233. [CrossRef]
8. King, M.; Meyer, G.E.; Tangney, J.; Biederman, I. Shape constancy and a perceptual bias towards symmetry. *Percept. Psychophys.* **1976**, *19*, 129–136. [CrossRef]
9. Perrett, D.I.; Burt, D.; Penton-Voak, I.; Lee, K.J.; Rowland, D.; Edwards, R. Symmetry and Human Facial Attractiveness. *Evol. Hum. Behav.* **1999**, *20*, 295–307. [CrossRef]
10. Little, A.C.; Jones, B. Variation in facial masculinity and symmetry preferences across the menstrual cycle is moderated by relationship context. *Psychoneuroendocrinology* **2012**, *37*, 999–1008. [CrossRef] [PubMed]
11. Little, A.C.; Apicella, C.L.; Marlowe, F.W. Preference for symmetry in human faces in two cultures: Data from the UK and the Hadza, an isolated group of hunter-gatherers. *Proc. R. Soc. B Biol. Sci.* **2007**, *274*, 3113–3117. [CrossRef]
12. Little, A.C.; DeBruine, L.; Jones, B. Exposure to visual cues of pathogen contagion changes preferences for masculinity and symmetry in opposite-sex faces. *Proc. R. Soc. B Biol. Sci.* **2011**, *278*, 2032–2039. [CrossRef]
13. Saxton, T.; Debruine, L.; Jones, B.; Little, A.; Roberts, S.C. A longitudinal study of adolescents' judgments of the attractiveness of facial symmetry, averageness and sexual dimorphism. *J. Evol. Psychol.* **2011**, *9*, 43–55. [CrossRef]
14. Swaddle, J.P.; Cuthill, I.C. Asymmetry and human facial attractiveness: Symmetry may not always be beautiful. *Proc. R. Soc. B Biol. Sci.* **1995**, *261*, 111–116. [CrossRef]
15. Koehler, N.; Rhodes, G.; Simmons, L.W. Are human female preferences for symmetrical male faces enhanced when conception is likely? *Anim. Behav.* **2002**, *64*, 233–238. [CrossRef]
16. Komori, M.; Kawamura, S.; Ishihara, S. Averageness or symmetry: Which is more important for facial attractiveness? *Acta Psychol.* **2009**, *131*, 136–142. [CrossRef] [PubMed]
17. Penton-Voak, I.S.; Jones, B.C.; Little, A.C.; Baker, S.; Tiddeman, B.; Burt, D.M.; Perrett, D.I. Symmetry, sexual dimorphism in facial proportions and male facial attractiveness. *Proc. R. Soc. B Biol. Sci.* **2001**, *268*, 1617–1623. [CrossRef]
18. Fink, B.; Neave, N.; Manning, J.T.; Grammer, K. Facial symmetry and judgements of attractiveness, health and personality. *Pers. Individ. Differ.* **2006**, *41*, 491–499. [CrossRef]
19. Jones, A.; Jaeger, B. Biological Bases of Beauty Revisited: The Effect of Symmetry, Averageness, and Sexual Dimorphism on Female Facial Attractiveness. *Symmetry* **2019**, *11*, 279. [CrossRef]
20. Holzleitner, I.J.; Lee, A.J.; Hahn, A.C.; Kandrik, M.; Bovet, J.; Renoult, J.P.; Simmons, D.; Garrod, O.; DeBruine, L.M.; Jones, B.C. Comparing theory-driven and data-driven attractiveness models using images of real women's faces. *J. Exp. Psychol. Hum. Percept. Perform.* **2019**, *45*, 1589–1595. [CrossRef] [PubMed]
21. Scott, I.M.L.; Clark, A.P.; Boothroyd, L.G.; Penton-Voak, I.S. Do men's faces really signal heritable immunocompetence? *Behav. Ecol.* **2013**, *24*, 579–589. [CrossRef]
22. Lewis, M.B. Fertility affects asymmetry detection not symmetry preference in assessments of 3D facial attractiveness. *Cognition* **2017**, *166*, 130–138. [CrossRef]
23. Little, A.C.; Griffey, J.A.F. Preferring and Detecting Face Symmetry: Comparing Children and Adults Judging Human and Monkey Faces. *Symmetry* **2020**, *12*, 2112. [CrossRef]
24. Little, A.C.; Jones, B. Attraction independent of detection suggests special mechanisms for symmetry preferences in human face perception. *Proc. R. Soc. B Biol. Sci.* **2006**, *273*, 3093–3099. [CrossRef]
25. Zebrowitz, L.A.; Rhodes, G. Sensitivity to "bad genes" and the anomalous face overgeneralization effect: Cue validity, cue utilization, and accuracy in judging intelligence and health. *J. Nonverbal Behav.* **2004**, *28*, 167–185. [CrossRef]
26. Wang, T.T.; Wessels, L.; Hussain, G.; Merten, S. Discriminative thresholds in facial asymmetry: A review of the literature. *Aesthet. Surg. J.* **2017**, *37*, 375–385. [CrossRef]
27. Anderson, N.D.; Gleddie, C. Comparing Sensitivity to Facial Asymmetry and Facial Identity. *i-Perception* **2013**, *4*, 396–406. [CrossRef]
28. Ma, D.S.; Correll, J.; Wittenbrink, B. The Chicago face database: A free stimulus set of faces and norming data. *Behav. Res. Methods* **2015**, *47*, 1122–1135. [CrossRef]
29. De Bruine, L.M.; Tiddeman, B.P. Webmorph. 2016. Available online: http://webmorph.org (accessed on 5 February 2021).
30. R Core Team. *A Language and Environmental for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2013.
31. Bates, D.; Mächler, M.; Bolker, B.; Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **2015**, *67*, 1–48. [CrossRef]

32. Kuznetsova, A.; Brockhoff, P.B.; Christensen, R.H.B. lmerTest: Tests for Random and Fixed Effects for Linear Mixed Effect Models. 2015. Available online: https://CRAN.R-project.org/package=lmerTest (accessed on 5 February 2021).
33. De Bruine, L.M.; Barr, D.J. Understanding Mixed-Effects Models through Data Simulation. *Adv. Methods Pract. Psychol. Sci.* **2021**, *4*. [CrossRef]
34. Barr, D.J.; Levy, R.; Scheepers, C.; Tily, H.J. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *J. Mem. Lang.* **2013**, *68*, 255–278. [CrossRef] [PubMed]
35. Debruine, L.M. Evidence versus speculation on the validity of methods for measuring masculinity preferences: Comment on Scott et al. *Behav. Ecol.* **2013**, *24*, 591–593. [CrossRef]
36. Lewis, M.B. Challenges to both reliability and validity of masculinity-preference measures in menstrual-cycle-effects research. *Cognition* **2020**, *197*, 104201. [CrossRef] [PubMed]
37. Pollet, T.; Little, A. Baseline probabilities for two-alternative forced choice tasks when judging stimuli in Evolutionary Psychology: A methodological note. *Hum. Ethol. Bull.* **2017**, *32*, 53–59. [CrossRef]