Supplementary Materials

1. Participant Demographic Details

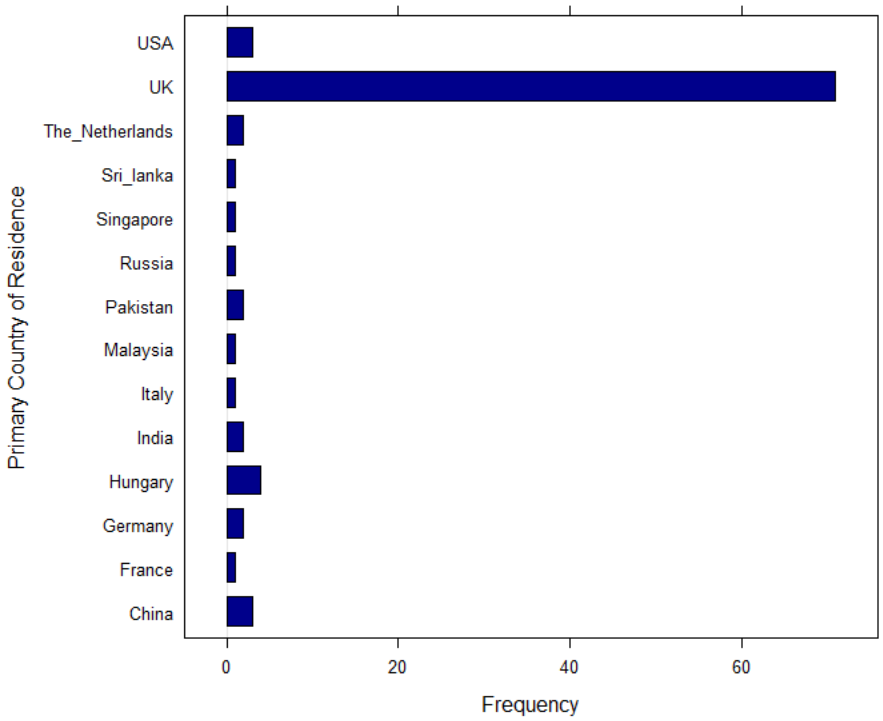

Figure 1. Bar chart showing the frequency of the primary country of residence of participants (n = 95).

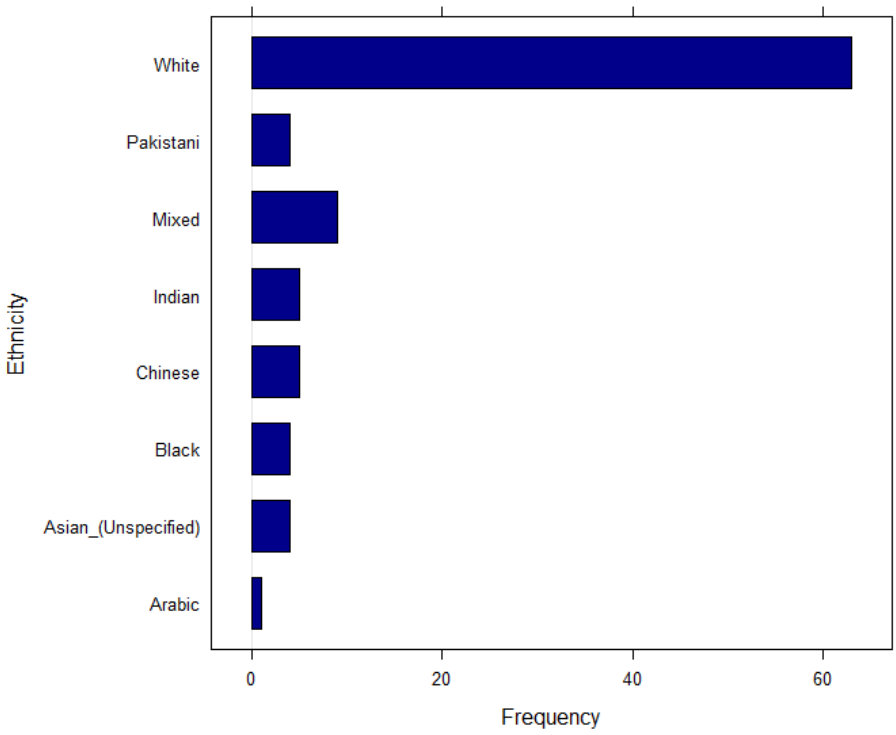

Figure 2. Bar chart showing the frequency of the ethnicities of participants (n = 95).

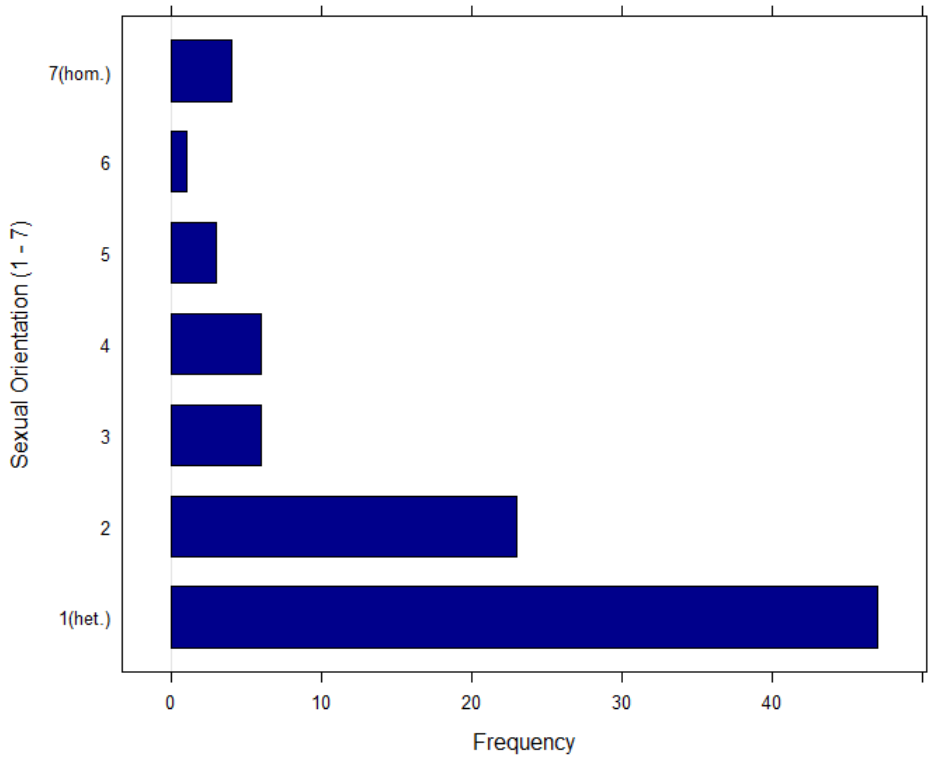

Figure 3. Bar chart showing the number of participants identifying in each sexual orientation group on a scale of 1 (completely heterosexual) to 7 (completely homosexual). As 5 participants did not disclose their sexual orientation, n = 90.

2. Face Stimulus Details

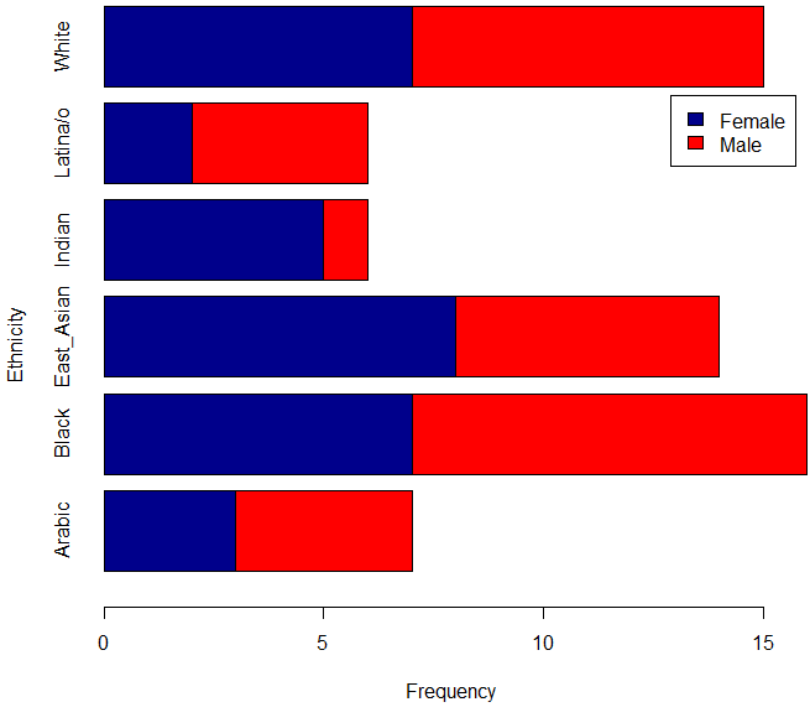

Figure 4. Bar chart showing the frequency of each gender and ethnicity amongst the pictures of faces used in this study (n = 64).

3. Instructions to Participants

Prior to the practice beauty trials, participants were presented with the following instructions:

(PART 1/3)

In this block, you will be presented with a series of faces.

For each face, you will be asked to rate how BEAUTIFUL you deem it to be on a scale of 1 (not beautiful at all) to 7 (very beautiful).

Then you will be asked to rate how certain you are of your beauty rating, again on a scale of 1 (not certain at all) to 7 (highly certain).

(PART 2/3)

Following these two ratings, you will be presented with the average beauty rating given by other participants to the same face.

In light of this information, you will be asked to give a final beauty rating for this face.

(PART 3/3)

In summary, when presented with a face, you will:

1. Give an initial beauty rating

2. Give a certainty rating

3. Be presented with the average rating of peers

4. Give a final beauty rating

For goodness trials, the following instructions were used:

(PART 1/3)

In this block, you will be presented with a series of faces.

For each face, you will be asked to rate how GOOD you deem the individual to be on a scale of 1 (not good at all) to 7 (very good). In other words, how likely are you to trust the individual?

Then you will be asked to rate how certain you are of your goodness rating, again on a scale of 1 (not certain at all) to 7 (highly certain).

(PART 2/3)

Following these two ratings, you will be presented with the average goodness rating given by other participants to the same face.

In light of this information, you will be asked to give a final goodness rating for this face.

(PART 3/3)

In summary, when presented with a face, you will:

1. Give an initial goodness rating

2. Give a certainty rating

3. Be presented with the average rating of peers

4. Give a final goodness rating

4. Gender and Sexual Orientation ANCOVA Effect Sizes

Variable	Sig.	Partial Eta Squared
Sexual orientation (cov)	.449	.007
Gender: participant (between)	.202	.020
Gender: image (within)	> .001	.384
Gender: participant X image (interaction)	.711	.002

Table 1. Effect sizes generated from the ANCOVA addressing the effect of gender and sexual orientation on ratings of beauty. Variable: type of independent variable (cov = covariate); Sig.: the p value obtained for the effect of a given variable; Partial Eta Squared: value of the effect size.

5. Intraclass Correlation Coefficient Analysis for Beauty and Goodness Rating Agreement

The means minus 1 (MM1; Vessel et al., 2018) and intraclass correlation coefficient (ICC) both reflect the degree of agreement amongst the ratings of different participants (raters). When the number of raters is high, MM1 is more sensitive to differences in agreement, whilst values of ICC are consistently high (> 0.9). This makes it hard to compare ICC values for different sets of ratings. The advantage of MM1 is that it shows the distribution of r-values for each participant (i.e. individual agreement scores) but is less accurate when the number of raters is low; also, being a relatively recent statistical analysis, it has been used less. ICC, on the other hand, has been more extensively used and is more accurate in reflecting agreement, regardless of sample size but it gives large values when the number of raters is high and large differences in agreement are reflected by small differences in number, thus obscuring numerically the magnitude of the difference (see Table 2). With this in mind, and given that the current study included data from 95 different raters, we deemed the MM1 analysis more suitable for our purposes. However, we have included the results of the ICC below. These results further highlight the dissociation in agreement amongst beauty and goodness ratings separately.

Attribute	Intraclass Correlation	Lower Bound	Upper Bound	Sig.
Beauty	.990	.987	.994	>.001
Goodness	.944	.923	.962	>.001

Table 2. Results of the intraclass correlation coefficient for agreement in beauty ratings and goodness ratings. The lower and upper bounds are the 95% confidence intervals.