Preregistration

Researcher
Kathrin Fucke

Working title
The evaluation of beauty: A mini meta-analysis of response bias in physical attractiveness ratings using three validated face databases

Introduction

"How physical attractive is this person?" These or similar questions were asked numerous times to investigate determinants and effects of physical attractiveness. Many interesting conclusions were drawn from those investigations. For example, that attractive faces are symmetric, close to the average of the population and judge and threatened more positive in many social domains (for meta-analsyses see Feingold, 1992; Langlois et al., 2000; Rhodes, 2006). Analyzing the data researchers are generally interested in the relative change of physical attractiveness ratings. The absolute value of physical attractiveness ratings is often paid less attention.

Response biases, which affect the absolute value given a person on a rating scale, are a source of measurement error and can inflate or deflate statistical results (Baumgartner & Steenkamp, 2001). Therefore, response biases in physical attractiveness ratings could threaten the validity of attractiveness research and the conclusions drawn from it. Some clues in the attractiveness literature suggest that such a response bias could exist. For example Reis et al. (1982) reported that the sample mean of physical attractiveness fell below the scale midpoint although the participants were randomly selected from the students population and should represent the spectrum of physical attractiveness. Wheeler und Kim (1997) noted that even high attractive person got low ratings and Braun, Gründl, Marberger, & Scherber, (2001) observed that people who rated the attractiveness of others expressed hard and critical comments during the rating procedure. All together, these signs point to a negative response bias when people judge the physical attractiveness of others. In the following response bias is defined as significantly deviation of sample mean from scale midpoint.

Current study

This study has two aims. The first aim is to investigate whether a response bias exists when people judge the physical attractiveness of others and, if yes, to characterize its direction and effect size. Second, I aim to evaluate dehumanization theory and sexual objectification theory to identify potential mechanisms of response bias in physical attractiveness ratings.

To accomplish this aims physical attractiveness ratings of three influential validated face databases will be analyzed: Chicago Face Database (CFD), the Face Research Lab London Set (FRLLS) and the Radboud Faces Database (RaFD). On this data analyses and mini meta-analyses will be performed.

Hypothesis
From dehumanization theory as well as sexual objectification theory can be deduced that rater see targets as objects which leads to critical and negative evaluations and low physical attractiveness ratings. Therefore, our main hypothesis is:

H1: Physical attractiveness ratings should be biased towards the negative range of rating scale and the sample mean of physical attractiveness should differ significantly from the scale midpoint.

Predictions from dehumanization and sexual objectification theory diverge concerning gender and trait effects. Their predictions are not exclusive because it's possible that their mechanisms occur simultaneously.

Hypotheses concerning gender effects of response bias:

H2a: Objectification due dehumanization should affect ratings of both genders. Therefore, the response bias should be present in ratings of women and men.
H2b: Sexual objectification should affect women more than men. Thus, gender should be a significant moderator of response bias.

Hypotheses concerning trait effects of response bias:

H3a: Objectification due dehumanization should affect a wide range of trait ratings. Hence, the negative response bias should be found in ratings of physical and nonphysical traits.
H3b: Sexual objectification should affect ratings of physical traits and nonphysical traits in different ways. Thus, the kind of trait should be a significant moderator of response bias.

Dependent variable
In this study two dependent variables will be analyzed: one is targets rated physical attractiveness; the other is targets rated trustworthy as index for nonphysical traits. Both traits were rated by a number of participants whose ratings are averaged to get a physical attractiveness measure for each target. Response bias is defined as significantly deviation of sample mean physical attractiveness values of the included databases from the rating scale midpoint.

Data bases, sample size and inclusion
Our analysis bases on data of validated face databases which are developed for scientific purposes and offer researchers free available images and norming data. Three face databases are widely used for scientific investigations and have the advantage that their norming data include ratings of physical attractiveness: Chicago Face Database, Face Research Lab London Set and Radboud Faces Database. Photographs of each database were taken under standardised conditions, e. g. similar light, background, dress, gaze direction and facial expression. Our analysis will base on ratings from photographs with direct gaze and neutral facial expression.

Chicago Face Database was developed by Ma, Correll, & Wittenbrink (2015). In this work the main image set will be included that contains photographs of 597 adults (307 females, 290 males) with different ethnic background recruited in the USA. Norming data are available for are a great variety of physical and nonphysical attributes and were assessed through independent raters. Physical attractiveness was evaluated by the question "Now, consider the person pictured above and rate him/her with respect to other people of the same race and gender." Rater answered the question regarding attractiveness on a 7-point scale ranging from 'not at all' to 'extremely'. Trustworthy was assessed by same question and scale as physical attractiveness, so it's possible to compare response biases of physical and nonphysical traits.

Face Research Lab London Set was developed by DeBruine & Jones (2017) and contains photographs of 102 adults (49 females, 53 males). Available data include self reported measures and ratings of physical attractiveness on a 7-point scale ranging from 'much less attractiveness than average' to 'much more attractive than average'. Ratings of nonphysical trait are not available for this database.

Radboud Faces Database was developed by Langner et al. (2010) and contains photographs of 67 adult and child targets. Rater judged the physical attractiveness on a 5-point scale ranging from 'unattractive' to 'attractive'. Photographs were also rated with respect to diverse features of facial expression (e. g. intensity or clarity) but not for trustworthy or other nonphysical traits.

Chicago Face Database and Face Research Lab London Set do not include stimuli of children. Because sample size of children models in Radboud Faces Database is small, only the 57 adult models (19 female, 38 male) will be included. Take together our analysis will base on N = 756 (375 women, 381 men).

Power analysis

A power analysis with G*Power 3 (Faul, Erdfelder, Lang, & Buchner, 2007) was run to compute statistical power 1-β with a given sample size, a significance level of α = 0.05 and a medium effect size of d = 0.5 (Cohen, 1992)[1] using a two sided one sample t-test. Results of power analysis are shown in

[1] We are interested in at least medium effect of response bias because this effect size could lead to a considerable amount of range restriction and threaten the validity of ratings.

Table *1*. As can be seen in this table statistical power is larger than 1-ß > 0.80 for all test with one expectation, and mostly even larger than 1-ß > 0.90. Because of differences in statistical power in our mini meta-analysis a fixed effect model will be used that gives more weight to larger databases.

Table 1

Computed achieved power 1-β for given sample size, effect size d = 0.5, α = 0.05 using a two-sided one sample t-test

Database	Gender	*N*	Power of 1-β
Chicago Face Database	female	307	1.00
Chicago Face Database	male	290	1.00
Face Research Lab London Set	female	49	0.93
Face Research Lab London Set	male	53	0.95
Radboud Faces Database	female	19	0.54
Radboud Faces Database	male	38	0.85

Notes. Total *N* = 756

Outliers and Exclusions

All available data will be included in our analysis. To do this, we need the permission to use norming data. Up to date, the permission to use norming data of Chicago Face Database is available. Permission from Radboud Faces Database is outstanding. Data from Face Research Lab London Set can be used by license CC BY 4.0. Outstanding permissions will be awaited until July 1st, 2022. Otherwise, the analysis is done without this database.

Analyses and mini meta-analysis

A two-sided one sample *t*-test will be used to examine response biases in each face database. This test compares the sample mean to the population mean μ. For our purpose μ is the value of the scale midpoint (e. g. μ = 4 when ratings were given on a 7-point scale). Effect size Cohen's *d* will be calculated to define direction and size of the response bias. A negative sign of effect size shows that the sample mean is lower than the scale midpoint, which indicates a negative response bias; a positive sign means that the sample mean is larger than the scale midpoint and indicates a positive response bias. Analsyses will be calculated separately for data base and gender resulting in six effect sizes Cohen's *d*.

Following guide of Simmons, Nelson, & Simonsohn (2021) a mini meta-analysis will be conducted to estimate the mean effect size *d* using fixed effect model that gives more weight to larger samples. First, effect size Cohen's *d* will be converted to Pearson's correlation *r* because it's simpler to analyze. Next, correlations will be transformed to Fisher's *z*. For reporting Fisher's z will be converted back to Pearson's correlation *r* and Cohen's *d*. Moderating effect of gender on response bias in physical attractiveness ratings will be evaluted.

To analyze trait specificity, we use ratings from physical attractiveness and thrusworthy of Chicago Face Database and calculate effect size Cohen's *d* and mini meta-analyze them as describe above. We evaluate the moderating effect of trait and gender on response bias in physical attractiveness ratings.

Analysis will be performed by R version 4.2.0 using packages *dplyr* (Wickham, François, Henry, & Müller, 2022), *effectsize* (Ben-Shachar, Lüdecke, & Makowski, 2020), *metafor* (Viechtbauer, 2010), *robumeta* (Fisher, Tipton, & Zhipeng, 2017) and *psych* (Revelle, 2022) will be used.

References

Baumgartner, H., & Steenkamp, J. B. E. M. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, *38*(2), 143–156. https://doi.org/10.1509/jmkr.38.2.143.18840

Ben-Shachar, M. S., Lüdecke, D., & Makowski, D. (2020). Effectsize: estimation of effect size indices and standardized parameters. *Journal of Open Source Software*, *5*(56), 2815. https://doi.org/10.21105/joss.02815

Braun, C., Gründl, M., Marberger, C., & Scherber, C. (2001). *Beautycheck - Ursachen und Folgen von Attraktivität. Projektabschlussbericht*. Abgerufen von http://www.beautycheck.de/cmsms/index.php/der-ganze-bericht

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. https://doi.org/doi.org/10.1037/0033-2909.112.1.155

DeBruine, L., & Jones, B. (2017). Face Research Lab London Set. figshare. Dataset. Abgerufen von https://doi.org/10.6084/m9.figshare.5047666.v5

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Feingold, A. (1992). Good-looking people are not what we think. *Psychological Bulletin*, *111*(2), 304–341. https://doi.org/10.1037/0033-2909.111.2.304

Fisher, Z., Tipton, E., & Zhipeng, H. (2017). *robumeta: robust variance meta-regression*. Abgerufen von https://cran.r-project.org/package=robumeta

Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., Hallam, M., & Smoot, M. (2000). Maxims and myth of beauty? A meta-analytic and theoretical review. *Psychological Bulletin*, *126*(3), 390–423. https://doi.org/10.1037/0033-2909.126.3.390

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the radboud faces database. *Cognition and Emotion*, *24*(8), 1377–1388. https://doi.org/10.1080/02699930903485076

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*(4), 1122–1135. https://doi.org/10.3758/s13428-014-0532-5

Reis, H. T., Wheeler, L., Spiegel, N., Kernis, M. H., Nezlek, J., & Perri, M. (1982). Physical attractiveness in social interaction: II. Why does appearance affect social experience? *Journal of Personality and Social Psychology*, *43*(5), 979–996. https://doi.org/10.1037/0022-3514.43.5.979

Revelle, W. (2022). *psych: procedures for psychological, psychometric, and personality research*. Abgerufen von https://cran.r-project.org/package=psych

Rhodes, G. (2006). The evolutionary psychology of facial beauty. *Annual Review of Psychology*, *57*(1), 199–226. https://doi.org/10.1146/annurev.psych.57.102904.190208

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2021). Pre-registration: why and how. *Journal of Consumer Psychology*, *31*(1), 151–162. https://doi.org/10.1002/jcpy.1208

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48.

Wheeler, K., & Kim, Y. (1997). What is beautiful is culturally good: The physical attractiveness stereotype has different content in collectivistic cultures. *Journal for Personality and Social Psychology*, *23*(8), 795–800.

Wickham, H., François, R., Henry, L., & Müller, K. (2022). *dplyr: a grammar of data*

manipulation. Abgerufen von https://cran.r-project.org/package=dplyr