

Supplementary Analysis 1: Experiment 1 - The effects of linguistic content on trait ratings

To assess whether the linguistic content of the words used in Experiment 1B systematically affected social trait evaluations (over and above the quality of the voice), we compiled arousal, valence, and dominance ratings for the words used in the study from a published database (Warriner, Kuperman & Brysbaert, 2013). Out of the 400 stimuli used in our experiment, no ratings were available for 30 of these stimuli. These stimuli were omitted from the analyses below.

To assess the effect of linguistic content on social trait evaluations, we computed the mean rating for each stimulus in our study, separately for each social trait (attractiveness, dominance, trustworthiness). We then ran three multiple linear regression analyses, one for each social trait, with the mean trait ratings per stimulus as the dependent variable and arousal, valence, and dominance ratings of the linguistic content of the stimuli as predictors.

For attractiveness ratings, the linguistic content of the words explained 3% of the variance in the trait ratings ($p = .003$). Dominance ratings of the linguistic content was a significant

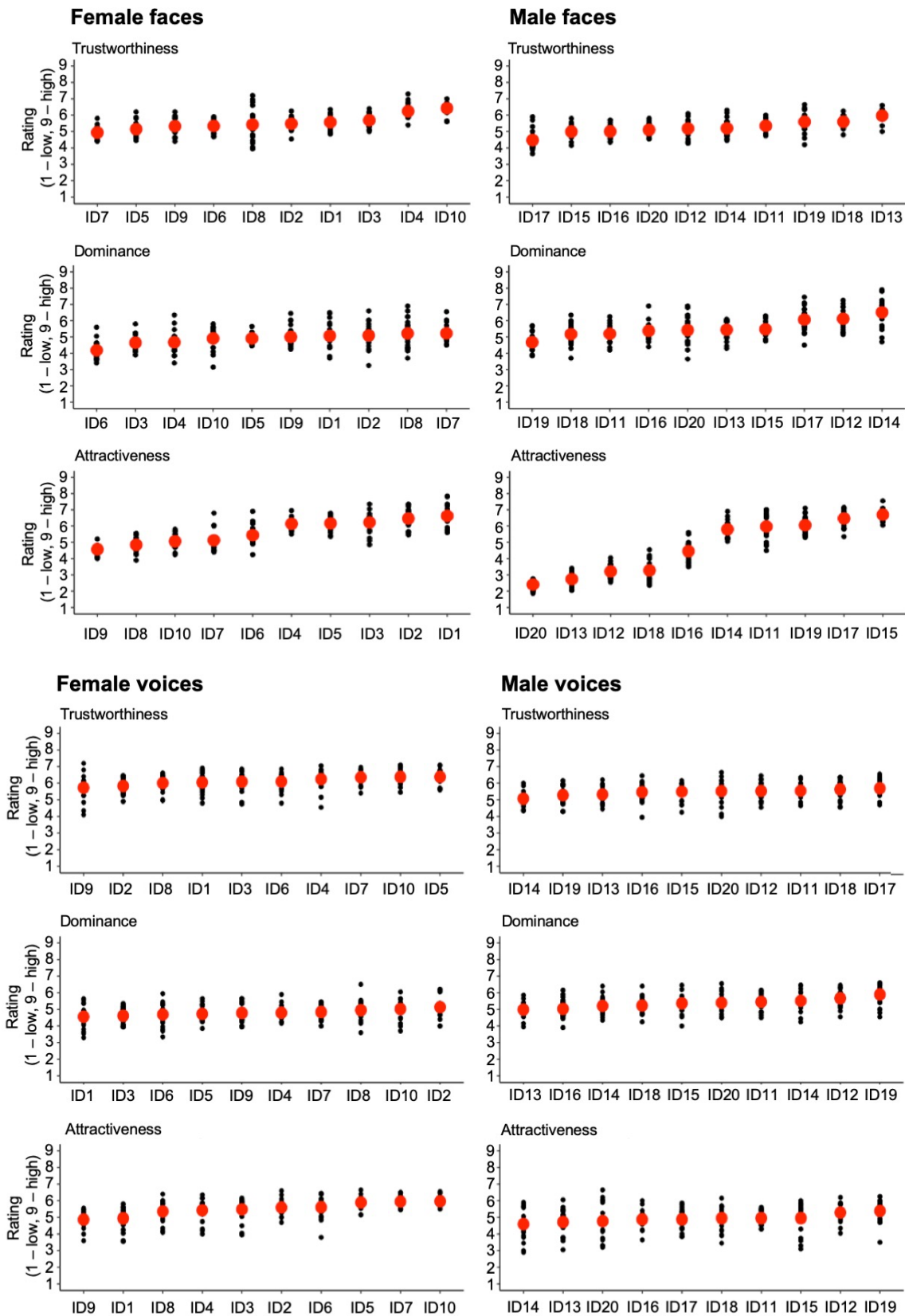
predictor for attractiveness ratings of the voices ($\beta = .18, p = .004$), arousal and valence ratings of the linguistic content did not significantly predict attractiveness ratings ($\beta < .03, p > .725$).

For dominance ratings, the linguistic content of the words explained $< 1\%$ of the variance in the trait ratings ($p = .473$). None of the ratings of the linguistic content (not even dominance ratings) significantly predicted dominance ratings of the voices ($\beta < .1, p > .128$).

For trustworthiness ratings, the linguistic content of the words explained 8% of the variance in the trait ratings ($p < .001$). Dominance ratings of the linguistic content was a significant predictor for trustworthiness ratings of the voices ($\beta = .19, p = .001$). Similarly, valence ratings of the linguistic content were a significant predictor for trustworthiness ratings, although the predictive relationship was negative, such that negatively valenced words predicted higher trustworthiness ratings of voices ($\beta = -.07, p = .044$).

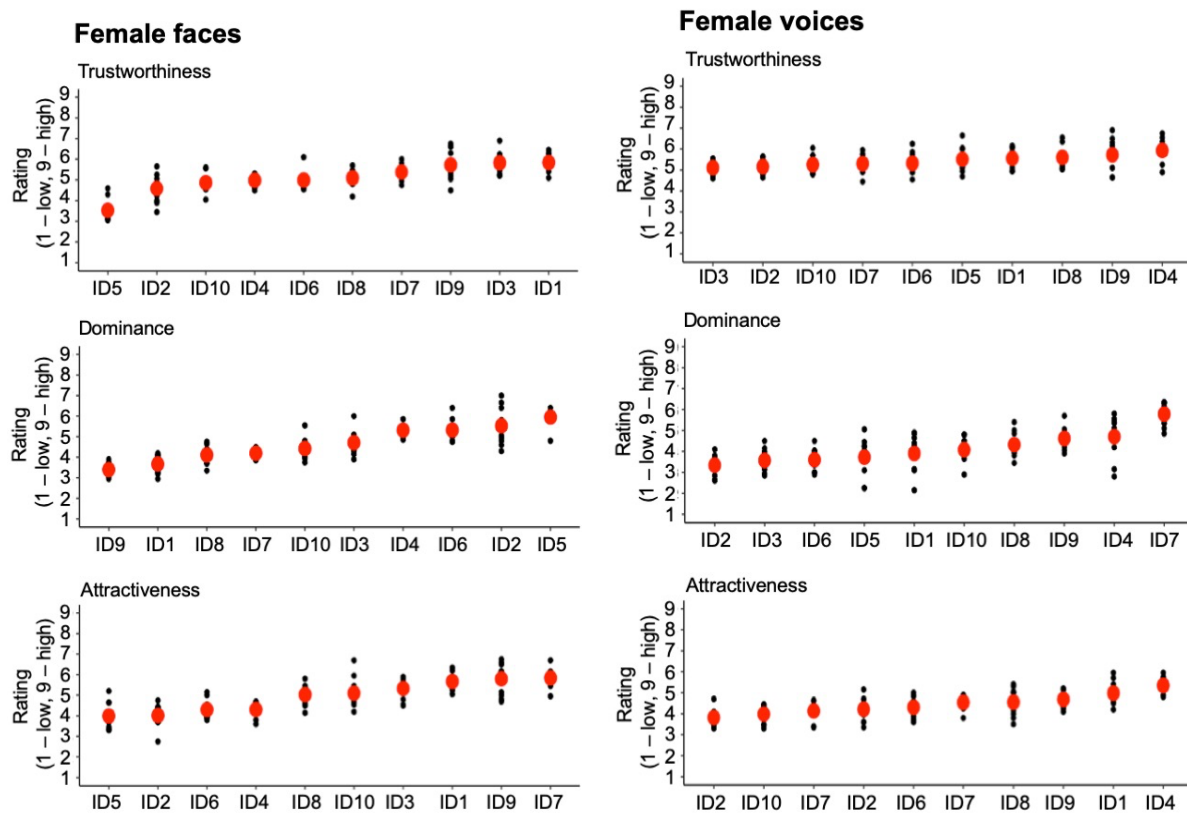
From these analyses, we conclude that linguistic content of the words can at times influence the trait evaluations of voices in our experiment. However, only a small portion of the variance is explained by the linguistic content ($< 9\%$) and the pattern of predictive relationships appears at times counterintuitive (no relationship for dominance ratings; negative relationship of valence and trustworthiness ratings). We therefore conclude that the linguistic content is unlikely to have had a systematic, interpretable effect on the social evaluations collected in this experiment.

Supplementary Figure 1: Experiment 1 - Alternative plot of the within- and between-person variability



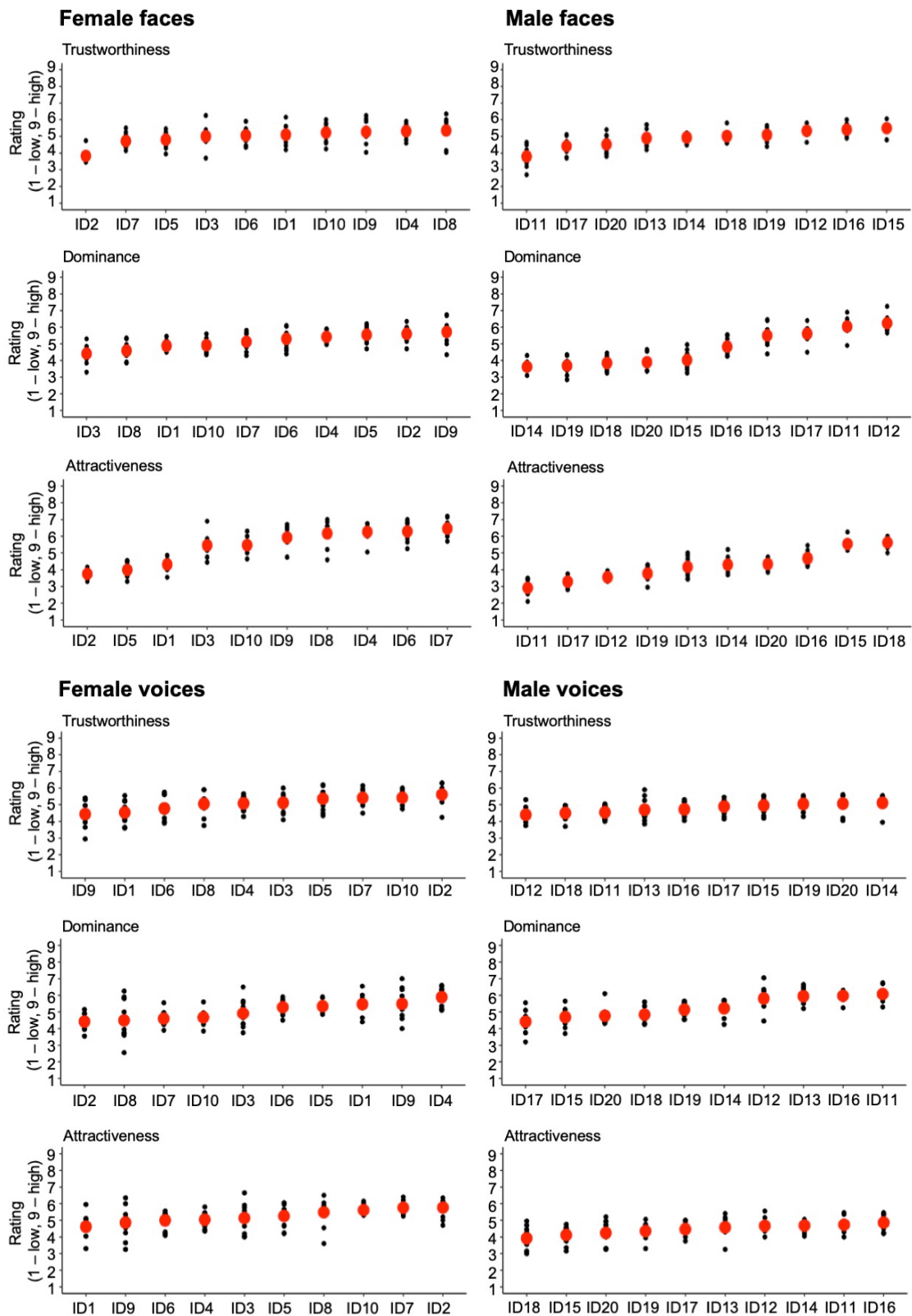
Supplementary Figure 1: Mean trait ratings plotted per item (black dots) and per identity (red dots). Higher within-person variability is indicated by a larger range in mean ratings per item. Higher between-person variability by a larger range of mean ratings per identity

Supplementary Figure 2: Experiment 2 - Alternative plot of the within- and between-person variability



Supplementary Figure 2: Mean trait ratings plotted per item (black dots) and per identity (red dots). Higher within-person variability is indicated by a larger range in mean ratings per item. Higher between-person variability by a larger range of mean ratings per identity.

Supplementary Figure 3: Experiment 3 - Alternative plot of the within- and between-person variability



Supplementary Figure 3: Mean trait ratings plotted per item (black dots) and per identity (red dots). Higher within-person variability is indicated by a larger range in mean ratings per item. Higher between-person variability by a larger range of mean ratings per identity.

References

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, 45(4), 1191-1207.
<https://doi.org/10.3758/s13428-012-0314-x>