Socialness Ratings: Data Cleaning

**Data Cleaning Procedure - Exclusion Criteria**

- completion time less than 3 SDs below the mean completion time - responded to less that 33% of the ratings task (i.e., rated less than 132 words) - responded with the same score for 25 (or more) stimuli in a row (we have 25 words per page) - responded with “I don’t know the meaning of this word” to more than 25 % of items (n = 100 out of 400 words) - correlation with the control words below .20 - correlation with the mean of the ratings of all other participants below 0.1 - items rated as “I don’t know the meaning of this word” by more than 15% of raters

**Data cleaning**

In total, we collected 241,575 observations. The data cleaning pipeline involved sequentially implementing several techniques consistent with recommendations for identifying careless or insufficient effort responders (Curran, 2016) and computer-generated random responding (Dupuis et al., 2019), as well as other data cleaning procedures used in previous word norming studies (Brysbaert et al., 2014; Pexman et al., 2019; Warriner et al., 2013). First, we removed data from participants if they completed less than 33% of the ratings task (n = 0), responded with “I don’t know the meaning of this word” for more than 25% of items (n = 8) and provided the same rating for more than 25 words in a row (n = 17). Next, we examined each participant’s ratings of the 30 control words and generated correlations with the mean ratings of those words obtained in the pilot study. We removed data from 36 participants with a correlation coefficient less than .20. We then computed the correlation between each participant’s ratings and the mean ratings of all other participants. We deleted data from 5 participants with a correlation coefficient less than .10. Finally, if more than 15% of raters reported not knowing a particular word, we removed those words from the analyses reported below. This led to the exclusion of 560 words. The final dataset comprised of 8,388 words and 202,841 observations, of which 3,542 were “I don’t know the meaning of this word” responses. Not taking into account the control words rated by all participants, each word in the final dataset had 21.92 valid ratings on average (*SD* = 1.68), ranging from 15 to 27 ratings. Overall, 7,703 (91.83%) words had at least 20 valid ratings.