Social Semantics: PILOT

Veronica Diveica

**Participants**

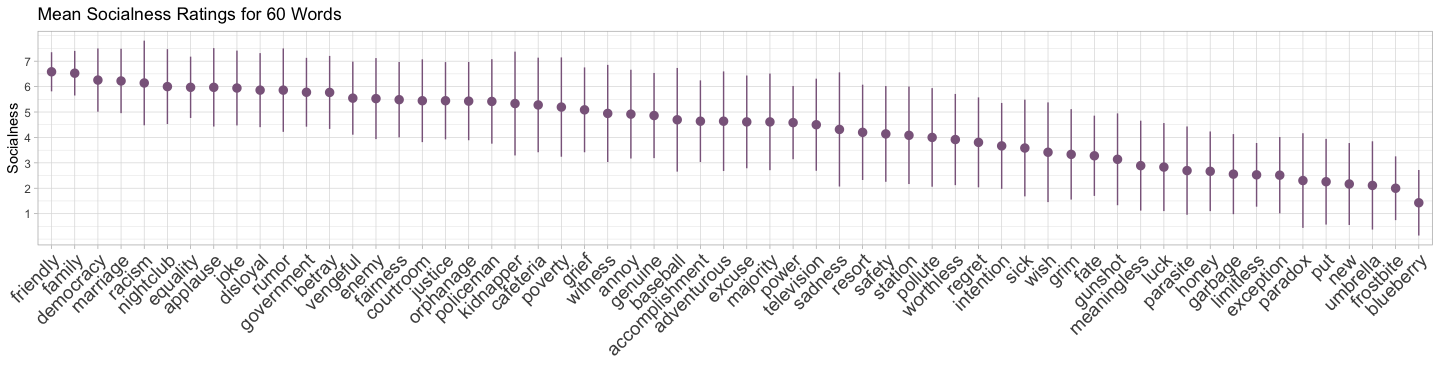
Before starting the main experiment, we tested our socialness rating task in a sample of 36 participants (23 female, 13 male; *Mage* = 22.94, *SDage* = 6.5). Participants were recruited from the participant pool at Bangor University. Participants completed the rating task in 24 minutes on average and were compensated with course credit. Of the participants, 17 saw version 1 of the instructions and 19 saw version 2.

**Materials**

We selected 60 items (including nouns, adjectives and verbs) that span the following dimensions: valence (Warriner et al., 2013), concreteness (Brysbaert et al., 2014) and social interaction (Binder et al., 2016; Troche et al., 2017). We created two versions of the instructions to assess whether wording influenced participants’ understanding of the instructions and their ratings. In version 1, socialness was defined as the degree to which a word’s meaning has a social quality, while in version 2 it was defined as the degree to which a word’s meaning has social relevance. The rest of the instructions and examples were identical in the two versions.

**Reliability**

We examined the reliability of the ratings by computing the split half reliability for the 60 words. We found a mean Spearman-Brown corrected split-half reliability of 0.97 (*SD* = 0.12) across 100 random splits, suggesting high reliability. In addition, we assessed inter-rater reliability by computing the two-way random-effects intra-class correlation coefficient (ICC) based on absolute agreement. We found an ICC(2,1) = 0.4, 95%CI [0.33, 0.48] suggesting poor to moderate reliability of individual ratings and an ICC(2, 36) = 0.96, 95%CI [0.95, 0.97] suggesting excellent reliability of the average ratings of 36 raters. Moreover, an ICC of 0.94, 95%CI [0.92, 0.96] suggested that the average ratings of 19 raters who saw the instructions subsequently used in the main experiment (version 2) were highly reliable.



**Understanding of instructions**

8.33% reported understanding the instructions moderately well, 55.56% very well and 36.11% extremely well.

**The effect of instruction version**

A Pearson’s Chi-square test of independence suggested that self-reported understanding of the instructions did not depend on the version of the instructions *χ²* (2, N = 36) = 0.92, *p* = 0.63). There was a strong positive correlation between mean socialness scores for the two instruction versions (*r* = 0.93, *p* <.001, *R2* = 0.87). Moreover, the reliability was comparable for the two versions, with a mean Spearman-Brown corrected split-half reliability of 0.93 (*SD* = 0.15) for version 1 and 0.95 (*SD* = 0.12) for version 2 (across 100 random splits). Therefore, we concluded that the wording did not significantly influence raters’ responses.