Supplementary statistics: On the interpretation of expressive adjectives: pragmatics or syntax?

1 Mixed model analysis

A binomial generalized linear mixed model was fitted using the lme4 package (Bates et al. 2015) in R (R Core Team 2020) predicting the responses by condition (i.e., left hopping, right hopping, matrix-subject, and matrix-clause reading) with random intercepts for participants and items and random slopes for participants, as shown in (1).

(1) response ~ condition + (1 + condition | participant) + (1 | item)

The exact specification is given in (2). Note that the data set as well as the R code is also available as a supplementary file.

(2)
```
library("lme4")
library("optimx")
model = glmer(condition + (1 + condition | participant) + (1 | item), family=binomial, data=data, control=glmerControl(optimizer = "optimx", calc.derivs = FALSE, optCtrl = list(method = "nlminb", starttests = FALSE, kkt = FALSE)))
summary(model)
lsmeans(model,pairwise~ condition)
```

The results of this model are shown in Table 1 (next page). The table starts with the left-hopping condition. That the matrix-clause interpretation was chosen more often to be possible than left hopping is reflected by a positive estimate in the next row. However, the matrix-clause interpretation was not judged to be significantly better than left hopping and the same applies to left hopping (last row). Only the matrix-subject interpretation differs significantly from the judgments of left-hopping. However, the matrix-subject reading was less often chosen to be possible (compared to left hopping), as reflected by the negative intercept. The different conditions were contrasted using the lsmeans package (Lenth 2016). All pairwise comparisons of conditions are shown in Table 2. The table shows that there are three statistically significant contrasts: The matrix-subject interpretation was significantly less often chosen than the left- and right-hopping option, and the matrix clause-interpretation was significantly more often chosen compared to the matrix-subject option.

	Estimate	**Std. Error**	***z* value**	**Pr(> |z|)**
(Intercept)	0.02612	0.31479	0.083	0.93386
conditionmatrixclause	0.39533	0.41066	0.963	0.33570
conditionmatrixsubject	−1.29140	0.43531	−2.967	0.00301
conditionrighthopping	0.45417	0.38180	1.190	0.23422

Table 1: Results of the binomial model.

Contrast	**Estimate**	***SE***	***z* ratio**	***p* value**
lefthopping - matrixclause	−0.3953	0.411	−0.963	0.7707
lefthopping - matrixsubject	1.2914	0.435	2.967	0.0159
lefthopping - righthopping	−0.4542	0.382	−1.190	0.6335
matrixclause - matrixsubject	1.6867	0.291	5.802	<0.0001
matrixclause - righthopping	−0.0588	0.492	−0.120	0.9994
matrixsubject - righthopping	−1.7456	0.532	−3.280	0.0057

Table 2: Pairwise comparisons of the different conditions with lsmeans.

2 Reliability coefficients

On first sight one might think that participants simply have guessed. If this would have been the case there should be no correlation between the items of each condition. The items, however, are correlated. As a measure of internal consistency the Kuder and Richardson Formula 20 (KR-20) (Kuder & Richardson 1937) and Cronbach's alpha (Cronbach 1951) was used. Both measure the reliability of the items (i.e, their consistency). All items in each condition were constructed to measure the same thing. If the items in one condition indeed measure the same thing one would expect that each participant is consistent in her/his judgments of these items (and not, for example, randomly judging left hopping to be possible in half of the cases and to be impossible in the other half). If this is the case the judgments of the items (in a specific condition for a specific participant) should be correlated. Such correlations can be measured by KR-20 and Cronbach's alpha (but note that, although widely used, reliability measures like Cronbach's alpha have also been criticized in the literature, cf. Green & Yang 2009).

KR-20 and Cronbach's alpha are similar measures, but KR-20 measures the test score reliability for binary data. For the interpretation of both the rule of thumb for Cronbach's alpha can be used as both measures ranges from 0 to 1 (although negative results are theoretically possible):

- Values between 0.7 and 0.9 are taken to show extremely high correlations (i.e., a high reliability of participants over items holding decisions fixed).
- Values between 0.6 and 0.9 are still taken to show a strong correlation.
- Values between 0.5 and 0.7 are taken to be acceptable and
- values below 0.6 are taken to show poor correlation.

It has to be noted, as already pointed out by Kuder & Richardson (1937), that KR-20 consistently underestimates reliability, thus it produces a little bit lower values. The items of the left-hopping condition revealed a KR-20 of $\rho = 0.71$ and the ones in the right-hopping condition of $\rho = 0.80$. For the CP-barrier condition a Cronbach's alpha was calculated revealing a $\rho = 0.64$. If the questions whether the matrix-subject and the matrix-clause interpretation are possible are looked at in isolation a KR-20 can also be calculated. Only looking at whether the matrix-subject interpretation was possible or not the KR-20 formula revealed a value of $\rho = 0.58$ (which is a little below the threshold) and for the matrix-clause interpretation a KR-20 of $\rho = 0.61$ was calculated. In sum, these values are in an acceptable range indicating that responses were not random, but that participants, in general, consistently judged the items in each condition and thus did not guess or make random judgments.

References

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1 48. https://doi.org/10.18637/jss.v067.i01.

Cronbach, Lee J. 1951. Coefficient alpha and the internal structure of tests. *psychometrika* 16(3). 297 334. https://doi.org/10.1007/BF02310555.

Green, Samuel B & Yanyun Yang. 2009. Commentary on coefficient alpha: a cautionary tale. *Psychometrika* 74(1). 121 135.

Kuder, G Frederic & Marion W Richardson. 1937. The theory of the estimation of test reliability. *Psychometrika* 2(3). 151 160. https://doi.org/10.1007/BF02288391.

Lenth, Russell V. 2016. Least-squares means: the R package lsmeans. *Journal of Statistical Software* 69(1). 1 33. https://doi.org/10.18637/jss.v069.i01.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/.