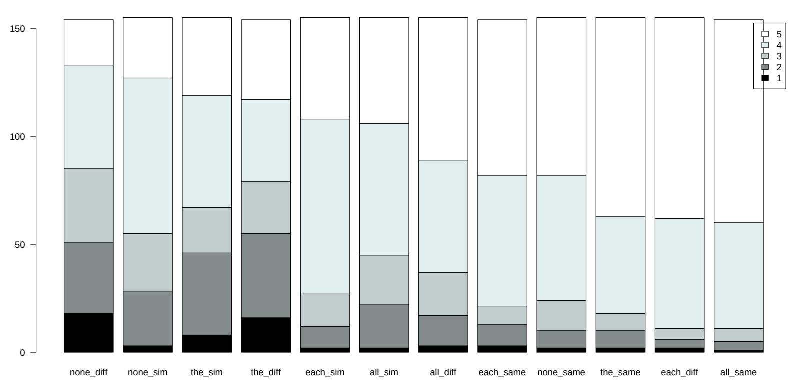
Licensing Sentence-internal Readings in English: An Experimental Study

I. The phenomena. Most, if not all, languages have lexical means to compare two elements and express identity / difference / similarity between them. English uses adjectives of comparison (henceforth AOCs) like *same*, *different* and *similar* for this purpose. Often, the comparison is between an element in the current sentence, e.g., the italicized NP *the same movie* in (1b) below, and a sentence-external element mentioned in the previous discourse, e.g., the underlined NP 'Waltz with Bashir' in (1a). However, AOCs can also compare sentence-internally, that is, without referring to any previously introduced element, as shown in (2). In this kind of cases, both compared elements are introduced in the current sentence, hence the **sentence-internal** label for this reading.

1. **a.** Arnold saw 'Waltz with Bashir'. **b.** Heloise saw *the same movie* / *a different movie*.
2. Each of the students saw *the same movie* / *a different movie*.

Some but not all semantically plural NPs are licensors of sentence-internal readings. In contrast to almost all previous semantic literature, we argue that it is incorrect to describe a particular NP as either licensing or not licensing the sentence-internal reading of a specific AOC. Licensing is more fine-grained. The gradient nature of AOC licensing in English has not been acknowledged in the previous literature (with the exception of Dotlačil 2010 for *different*) and has not been systematically studied. The main goal of this paper is to use experimental methods to begin filling this gap and establish which NPs license which AOCs and to what extent. The second goal of this paper is to show the advantage of using Bayesian statistics for analyzing data over the traditional, frequentist approach.

II. Experimental method. We used questionnaires to test people's intuitions about sentence-internal readings of 3 AOCs, namely *same*, *different* and *similar*, with 4 licensors, namely, NPs headed by *each*, *all*, *none* and *the*, i.e., $3 \times 4 = 12$ conditions in total. The *each-same* and *each-different* conditions were exemplified in (2) above. Each condition was tested four times, twice in a scenario in which the condition was most likely judged as true and twice in a scenario in which the condition was most likely judged as false; there were 32 fillers. Each item was judged with respect to (*i*) TRUTH: whether it is true, false or unknown given the accompanying scenario and (*ii*) ACCEPTABILITY: how acceptable it is on a 5-point scale (5=completely acceptable, 1=completely unacceptable). TRUTH was measured so that it can be distinguished from ACCEPTABILITY (for any given condition, true items tend to be judged more acceptable than false items). A total of 42 subjects in two undergraduate classes completed the questionnaire for extra-credit. For each subject, we randomized both the order of the scenarios in the questionnaire and the order of the items for each scenario.



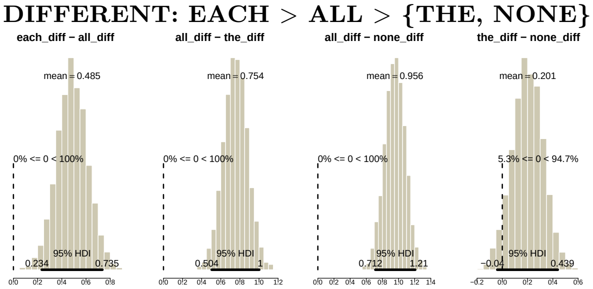
We excluded 2 subjects because of their incorrect responses to fillers and 1 because only TRUTH was completed; 1 of the remaining 39 subjects filled in only three fourths of the questionnaire. Final N(umber of obs.)=1856. Barplots of ACCEPTABILITY for the 12 conditions are shown to the left, from the least acceptable, i.e., sentence-internal reading of *different* when the licensor NP is headed by *none*, to the most acceptable, i.e., sentence-internal reading of *same* when the licensor NP is headed by *all*.

III. Statistical modeling and resulting generalizations The response variable ACCEPT(ABILITY) is ordinal, so we use ordered probit regression models to analyze the data. We have 2 fixed-effects predictors: QUANT-AOC (factor with 12 levels since we have 12 licensor-AOC combinations; reference level: the *each-different* combination) and TRUTH (factor with 3 levels: true, false, unknown; reference level: true). Our main interest is in how QUANT-AOC affects ACCEPT while controlling for / factoring out the influence of TRUTH on ACCEPT. A frequentist analysis shows that adding either of the fixed effects to the null (intercept-only) model significantly decreases deviance, but the interaction of the fixed effects does not (p=0.31). That is, licensor-AOC combinations and truth-value judgments significantly and additively influence acceptability judgments. Adding intercept-only random effects for items accounts for practically no variance, but adding random effects for subjects does (std.dev=0.56).

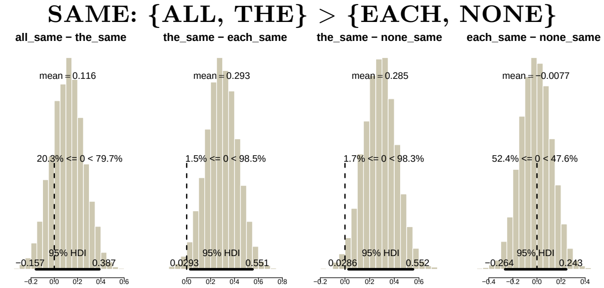
Thus, the final regression model $\mathcal{M}$ we will henceforth focus on has 2 fixed effects for QUANT-AOC and TRUTH (no interaction) and intercept-only random effects for subjects. Our primary interest is to establish which NPs license sentence-internal readings of which AOCs and to what extent. That is, we are interested in a wide range of pairwise comparisons between various licensor-AOC combinations. But doing this in the null-hypothesis significance testing framework would require an unfeasibly large amount of data to achieve significance given the necessary α-level correction for running all pairwise comparisons between the 12 licensor-AOC combinations (66 comparisons in total).

In contrast, any number of pairwise comparisons can be carried out in a Bayesian framework because we do not use p-values as a criterion for decision making. Instead, we simply study the multivariate posterior distribution of the parameters obtained given our prior beliefs, the data and our mixed-effects order probit regression model $\mathcal{M}$. Pairwise comparisons of various licensor-AOC combinations are just different perspectives on, i.e., different ways of marginalizing over, this posterior distribution (see Kruschke 2010 and references therein for more discussion). To determine whether there is a credible difference between any two conditions, we check whether 0 (=no difference) is in the 95% highest posterior density interval (HDI; basically, a 95% CI) of the difference: if 0 is outside the HDI, the two conditions are credibly different. The posterior plots for the most relevant comparisons are shown below, grouped by AOC. The resulting generalizations are summarized at the top of each set of plots, where > means the licensor(s) on the left is/are preferred to the licensor(s) on the right. The results for *different* mimic the results for *ander* ("different" in Dutch) reported in Dotlacil (2010).

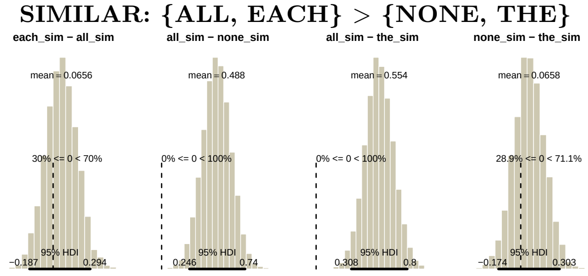
DIFFERENT: EACH > ALL > {THE, NONE}



SAME: {ALL, THE} > {EACH, NONE}



SIMILAR: {ALL, EACH} > {NONE, THE}



Priors for the Bayesian model. Independent normal priors $N(0, 1^2)$ for the non-reference levels of QUANT-AOC and TRUTH. The distribution of the subject random effects: $N(0, \sigma^2)$, with a *Unif*$(0, 5)$ prior for σ. Probit link function: the cdf of $N(0, 1^2)$. The priors for the 4 thresholds: independent normal priors with variance 1^2 centered at their MLEs -2.77, -1.88, -1.42, -0.34 to facilitate the MCMC sampling (chains: 3, iterations: 1250000 per chain, burnin: 250000, thinning: 500).

The paper discusses the theoretical significance of these findings, focusing on the fact that much of the previous semantic literature (Carlson 1987, Moltmann 1992, Beck 2000, Barker 2007, Brasoveanu 2011) does not systematically address the gradient nature of licensing sentence-internal readings.

References: [**1**] Barker, C. (2007). Parasitic Scope. *Ling. and Phil.* 30, 407-444. [**2**] Beck, S. (2000). The Semantics of Different. *Ling. and Phil.* 23, 101-139. [**3**] Brasoveanu, A. (2011). Sentence-internal *Different* as Quantifier-internal Anaphora. *Ling. and Phil.*, to appear. [**4**] Carlson, G. (1987). *Same* and *Different*. *Ling. and Phil.* 10, 531-565. [**5**] Dotlačil, J. (2010). *Anaphora and Distributivity*. PhD diss., Utrecht Univ. [**6**] Kruschke, J. (2010). Bayesian Data Analysis. WIREs Cognitive Science 1, 658-676. [**7**] Moltmann, F. (1992). Reciprocals and *Same/Different*. *Ling. and Phil.* 15, 411-462.