Exclusive disjunction: implicature or ...

Michael Franke (Universität Tübingen)
Bob van Tiel (Université Libre de Bruxelles)

Introduction. Exclusive readings of disjunctions are frequently treated as a scalar implicature arising from the competition of lexical alternatives *or* and *and*, parallel to the case of *some* and *all*. However, there is recent experimental evidence for the assumption that the strength of scalar inferences can vary substantially between different scalar items (e.g. van Tiel et al., 2016) and be affected by multiple contextual factors in subtle ways (e.g. Degen, 2015). The main question that we would like to address experimentally in this paper is therefore: is the strength of exclusive readings of disjunctions influenced by contextual factors in the way that a scalar implicature account would predict?

Many factors may have an effect on the strength of scalar inferences. We will focus here on three factors, for reasons of theoretical interest: (i) whether the scalar inference is *relevant* information for the listener (factor REL); whether the speaker is *competent*, i.e., she knows whether a stronger alternative statement would be true (factor COM), and (iii), although this is usually not discussed prominently in the theoretical literature, whether common sense expectations make the propositional content of a scalar inference *a priori likely* (factor PRI).

Different theoretical accounts of scalar implicatures may assign different weights to these factors, but, arguably, most would acknowledge that these factors play a role. For instance, standard Gricean(-like) reasoning could make relevance of information and the speaker's competence necessary conditions for the derivation of a scalar implicature, but also see a role for prior expectations. A game-theoretic or probabilistic approach might put particular emphasis on the role of prior expectations instead. In grammatical approaches potential scalar implicatures are generated as parses of sentences and some disambiguation procedure selects (contextually) preferred parses. Here, the relevance of alternatives is often flagged as important, but it is not strange to think that the disambiguation procedure is sensitive to the other factors as well: for instance, it may weigh in the contextual plausibility of ignorance inferences that different parses would give rise to (e.g. Fox, 2007).

Across many prominent theoretical frameworks, we would therefore expect that strength of scalar inferences should be higher (i) the more contextual relevance is attributed to the stronger alternative, (ii) the more likely the speaker is taken to know whether the stronger alternative is true, and (iii) the less likely it is *a priori* that the stronger alternative is true. If this pattern is empirically confirmed for some scalar inferences, but not for others that would be surprising, and potential evidence against a homogeneous theoretical treatment.

Experiment 1. We tested the effects of factors REL, COM and PRI on the robustness of exclusive readings. Materials consisted of 16 vignettes involving two characters *S* and *H*. Each vignette introduced some background scenario and, except for PRI-conditions (see below), ended with *S* uttering a sentence of the form "*A* or *B*." We classified each vignette based our own intuition as either high or low in each of REL, COM and PRI, and created 2 vignettes for all 8 theoretically possible vignette types, i.e., intuitively high/low factor combinations. Since our own intuitive classification might not accord with the perception of our experimental subjects, we obtained empirical measures for REL, COM and PRI and used these as explanatory factors for the measured strength of exclusive readings (XOR). To do so, each vignette was associated with five statements: three control statements and four types of target statements, which aimed at REL ("It is important for *H* to know whether *A* and *B*"), COM ("*S* knows whether *A* and *B*"), PRI ("If *A*, it is likely that *B* as well" and "If *B*, it is likely that *A* as well"),[1] and XOR ("From what *S* said we may conclude that not both *A* and *B*").

[1]Statements for PRI were formulated as pairs of conditionals, rather than as single statements about the conjunction "*A*

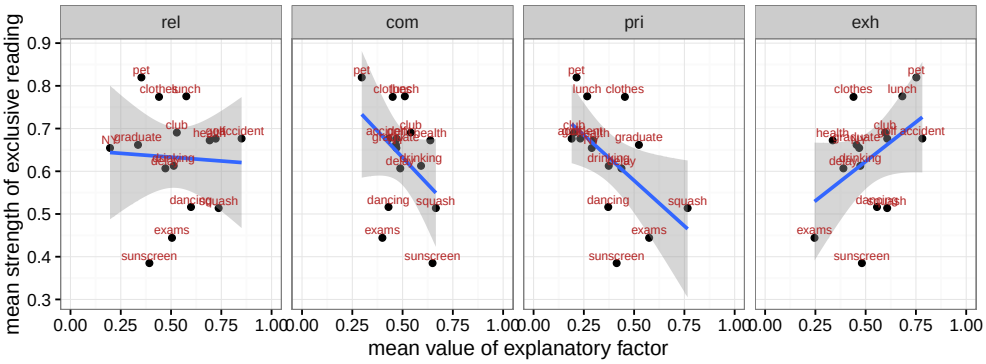

Figure 1: Per-vignette means of implicature rating vs. per-vignette means of factor ratings in Exp. 1 & 3

200 participants, recruited via MTurk, each read 8 random vignettes, one of each vignette type. Each vignette was followed by two statements: first a control statement, then one target statement out of REL, COM, PRI or XOR. (Trials probing for PRI were special: they did not show *S*'s utterance of a disjunction and presented both of its target sentences in random sequence.) Participants indicated how likely the statement was by adjusting a continuous slider bar whose endpoints were labeled "Certainly false" to "Certainly true." Responses to controls clearly showed that participants paid attention to the vignettes and understood the task (same for follow-up experiments). Slider ratings were coded as reals in $[0;1]$. Figure 1 plots per-vignette means of XOR-ratings against those of REL, COM and PRI. Our main results are foreshadowed by visual inspection: REL does not seem to be a good predictor of XOR; COM and PRI do, but the correlation between COM and XOR is not as expected.

To test factor influence, we looked at linear regression models predicting XOR ratings from by-vignette means of ratings for REL, COM and PRI as main factors. Bayes factor model comparison (e.g. Kass and Raftery, 1995), which weighs in likelihood and model complexity, clearly identified two equally favored models that were significantly better than all competitors:[2] one containing only PRI and one containing PRI and COM. Estimates of coefficients indicated that both PRI and COM had significant effects, but, critically, the effect of COM went in the opposite direction to what an implicature-based approach predicts: the perceived strength of the exclusive reading *decreased* with the perceived competence of the speaker.

How to explain the surprising negative effect of competence? Perhaps the most parsimonious explanation would be to attribute it to methodological flaws. That is, perhaps our manipulation did not measure what we assumed it measured. In order to address that possibility we replicated Exp. 1 using vignettes with *some*, which may license an upper-bounding inference to the exclusion of *all*. The status of this inference as a scalar implicature is less controversial than in the case of exclusive-*or*. Hence, *some* provides a baseline against which the results for *or* can be contrasted.

Experiment 2. Design and procedure of Exp. 2 was analogous to Exp. 1, but used different vignettes with utterances containing *some* instead of *or*. Bayes factor comparison of regression models showed that the optimal model for predicting the strength of the upper-bounding inference contained both PRI and COM. Posterior estimates of coefficients indicated, in this case, that the effect of COM went in the

and *B*," because the latter may be very unlikely in a context where also many other options *C*, *D*, ... are conceivable. As a result, we might not obtain judgements of *a priori* plausibility that compare reliably across vignettes: a low rating for "*A* and *B*" could be the result of comparing it to *C*, *D*, etc., in one case, and of considering it unlikely that both *A* and *B* hold if at least one of them does in another context. To avoid this issue, we used pairs of conditionals.

[2]Non-Bayesian analyses and more complex models involving interactions and by-participant intercepts support our main conclusions, so that we report only the simplest analyses here (similarly for follow up experiments).

"right" direction: the inference from *some* to *not all* was considered more robust if the speaker was judged to be competent about the truth of the *all*-alternative.

Taken together, the results of Exp. 1 and 2 indicate it might be a mistake to assume a homogenous theoretical treatment of exclusive-*or* and the upper-bounded inference associated with *some*. Assuming that the upper-bounded inference associated with *some* is a scalar implicature, one would then require an alternative treatment for exclusive-*or*. One such treatment is tentatively considered by Fox (2007). He suggests that exclusive-*or* may come about by exhaustifying the individual disjuncts rather than the disjunction as a whole. That is, "*A* or *B*" might be interpreted as "Only *A* or only *B*," which excludes the possibility that both *A* and *B* are true. If this account is correct, we might expect that the strength of exclusive-*or* is an increasing function of the strength of the exhaustive inference associated with an utterance of the individual disjuncts, i.e. the inferences from "*A*" and "*B*" to "only *A*" and "only *B*." This prediction was tested in Exp. 3.

Experiment 3. Exp. 3 used the 16 vignettes from Exp. 1, changing only the final utterance from a disjunction ("*A* or *B*") to an utterance of one of the disjuncts ("*A*" respectively "*B*"). Each vignette was associated with four statements: three control statements and one target statement probing for exhaustive interpretation of individual disjuncts ("From what *S* said we may conclude that *B* is not true as well" and "From what *S* said we may conclude that *A* is not true as well"). 130 participants each read 6 vignettes that were followed by one control and one target statement.

Mean ratings, per vignette, of target statements were added as an additional predictor EXH (see Figure 1). Bayes factors in favor of the most preferred models from Exp. 1 over a model that only contains EXH are about 5.24 and 5.63. This is just above the threshold of 5 for Bayes factors to indicate substantial evidence in favor of a model, as proposed by Jeffreys (1961). A model that uses both PRI and EXH as main factors is just as credible, given the data, as a model with only PRI. Our data therefore suggest that PRI might be the most reliable predictor of XOR, but EXH follows closely on its heals. The inclusion of EXH does not lead us to revise our previous conclusions about COM and REL: the former is "wrongly" correlated with XOR; REL is irrelevant.

Conclusion. The behaviour of *or* differs in some marked ways from the behaviour of *some* and more generally from what one would expect if exclusive-*or* were a scalar implicature. In particular, the strength of exclusive-*or* did not increase with the competence of the speaker about the *and*-alternative; rather, the effect was the other way around. This squares with the theoretically-motivated skepticism expressed by Geurts (2006) about the status of exclusive-*or* as scalar implicature. On the positive side, our results indicate that strength of exclusive-*or* clearly seem to be influenced by prior expectations, and these also affect the strength of *some-but-not-all*-inferences. On the other hand, our results leave open the theoretically interesting alternative that exclusive-*or* might be the outcome of reading each disjunct exhaustively. Finally, relevance failed to influence the strength of either exclusive-*or* or the upper-bounded inference associated with *some*. This finding conflicts with the received opinion about relevance, but is in line with recent experimental results suggesting a marginal role for relevance in the derivation of scalar implicature (e.g. Zondervan, 2010).