The role of relevance, competence, and priors for scalar inferences

Polina Tsvilodub, Bob van Tiel, & Michael Franke*

Abstract. Although it is often assumed that the natural language expressions 'some' and 'or' are interpreted according to their first-order logic counterparts, in certain contexts, they receive a narrower interpretation: 'some' is strengthened to 'some, but not all', and 'or' to 'or, but not both'. This process is typically explained as an instance of *scalar inference*. To test this scalar inference hypothesis, we collect experimental evidence for the effects and interactions of three factors that have been argued to affect the robustness of the scalar inferences of 'some' and 'or': the *relevance* of the stronger alternative, the speaker's *competence* about the alternative, and the *prior probability* that the alternative is true. We find that the interpretation of both triggers was affected by speaker competence, but only the interpretation of 'some' was also affected by prior probability, while relevance did not affect the interpretation of either trigger. Ultimately, our results suggest that the interdependence of the three factors is more complex than just the sum of their effects.

Keywords. scalar inference; relevance; competence; prior probability; disjunction

1. Introduction. It is often assumed that the natural language expressions 'some' and 'or' are equivalent to the existential quantifier $\exists$ and inclusive disjunction $\vee$ from first-order logic (e.g., Grice 1975). However, in certain contexts, 'some' and 'or' appear to receive a narrower interpretation than their alleged logical equivalents. For example, neither of the intuitive inferences in (1) and (2) follow from the proposed logical equivalence.

(1) Peter ate some of the doughnuts.
$\rightsquigarrow$ Peter did not eat all of the doughnuts.

(2) Peter ate a doughnut or a beignet.
$\rightsquigarrow$ Peter did not eat both a doughnut and a beignet.

The narrowing inferences of 'some' and 'or' are usually assumed to be instances of a more general type of inference called *scalar inference* (e.g., Horn 1972, Gazdar 1979, Geurts 2010). Scalar inferences are so-called because they are associated with lexical scales consisting of expressions that are, inter alia, lexicalised to the same degree and ordered in terms of logical strength. In the case at hand, 'some' is said to be associated with the scale $\langle$some, all$\rangle$ and 'or' with $\langle$or, and$\rangle$. (Positive) utterances containing a lower-ranked scalar expression may imply that the corresponding sentence with the higher-ranked expression is false.

Scalar inferences are often explained as a variety of *conversational implicature*. Conversational implicatures are inferences that can be explained on the basis of an argument that revolves around the assumption that the speaker is *cooperative*. Grice (1975) develops the notion of cooperativity by arguing that cooperative speakers tend to follow certain conversational maxims. For

*Many thanks to Natalie Clarius, Neele Witte and Elisa Kreiss for help in creating materials and to Stela Ilieva and Malin Spaniol for help with implementation during early stages of this project. Authors: Polina Tsvilodub, Osnabrück University (polina.tsvilodub@gmail.com), Bob van Tiel, Radboud University Nijmegen (bobvantiel@gmail.com) & Michael Franke, University of Tübingen (mchfranke@gmail.com).

example, cooperative speakers tend to be truthful (Quality) and informative (Quantity). Based on these two maxims, the scalar inference in (1) can be explained as follows:

1. The speaker said 'Peter ate some of the doughnuts'.
2. She could have said 'Peter ate all of the doughnuts' (Maxim of Quantity).
3. This would have been more informative, and hence cooperative.
4. So why didn't the speaker utter the more informative *alternative*?
5. Presumably, the speaker does not believe that the alternative is true (Maxim of Quality).
6. It is likely that the speaker knows whether the alternative is true or false.
7. Hence, the speaker believes that the alternative is false.

The same argument, mutatis mutandis, can be given to explain the scalar inference of 'or', as exemplified in (2). In that case, the relevant alternative is 'Peter ate a doughnut and a beignet', and the resulting inference says that the speaker believes Peter did not eat both a doughnut and a beignet. We will call this implicature-based account of the scalar inferences of 'some' and 'or' the *standard account*.

While the standard account is almost universally adopted for 'some', its application to 'or' has been more controversial (e.g., Simons 2001, Geurts 2006, Zondervan 2010). In particular, the use of 'or' tends to trigger the inference that the speaker is unsure which of the two disjuncts is true (e.g., whether Peter ate a doughnut or a beignet). This *ignorance inference* makes it a priori unlikely—though not impossible—that the speaker knows whether or not the disjuncts may be jointly true. Zondervan (2010) calls this the *speaker expertise paradox*.

In any case, according to the standard account, scalar inferences are *inferences to the best interpretation* (Atlas & Levinson 1981). The hearer interprets the speaker's utterance by trying to explain why the speaker uttered one sentence rather than a more informative alternative. In the case at hand, the most plausible hypothesis is assumed to be that the speaker believes that the alternative is false. But it is part and parcel of inferences to the best interpretation that, depending on the context, different explanations for the speaker's behaviour may emerge as the most plausible. For example, the robustness of scalar inferences has been argued to be influenced by *competence*, *relevance*, and *prior probabilities*. In the next section, we discuss these three factors in more detail. After that, we describe our experiment in which we tested the effects of these three factors on the robustness of the scalar inferences of 'some' and 'or'.

2. Factors influencing the robustness of scalar inferences.

2.1. COMPETENCE. The derivation of scalar inferences is often presented as a two-step process. First, it is inferred that the speaker avoided producing the alternative because she does not believe that the alternative is true. Second, this weak inference can be strengthened to the scalar inference that the speaker believes the alternative to be false. Crucially, the step from the weak inference to the scalar inference relies on the assumption that the speaker knows whether or not the alternative is true. This assumption is called the *competence assumption* (step 6 above) (e.g., Sauerland 2004, Soames 1982).

Goodman & Stuhlmüller (2013) experimentally investigated the effect of the plausibility of the competence assumption on the robustness of the scalar inference of 'some'. In their Exp. 1, participants were presented with vignettes in which a speaker uttered sentences containing the trig-

ger ‘some’. The vignettes were designed so as to vary with respect to the assumed knowledge of the speaker. In the *full knowledge* condition, the speaker was shown to be competent; in the *partial knowledge* condition, the competence assumption was not satisfied. Goodman & Stuhlmüller found that participants in the partial knowledge condition were significantly less likely to derive the scalar inference compared to the full knowledge condition.

The effect of the plausibility of the competence assumption on the scalar inference of ‘or’ has not been studied experimentally before. However, it has been observed that there is an apparent tension between the competence assumption for ‘or’ and its ignorance inferences. To illustrate, consider again (2). An utterance of this sentence tends to give rise to the ignorance inferences that the speaker does not know whether Peter ate a doughnut, and she does not know whether Peter ate a beignet. Given these ignorance inferences, it is difficult to imagine that the speaker is confident that Peter did not eat both a doughnut and a beignet, i.e., that the competence assumption is satisfied (Geurts 2006, Zondervan 2010). In the case at hand, that would require, e.g., that the speaker watched Peter having lunch from afar, seeing that Peter ate one and only one thing which the speaker could make out to be either a doughnut or a beignet.

2.2. RELEVANCE. An alternative explanation for the speaker’s decision to produce an informationally weaker utterance is that the speaker assumed that the hearer would not be interested in the added information expressed by the alternative. To illustrate, compare the following dialogues (from van Kuppevelt 1996):

(3) A: How many of the boys were at the party?
B: Some of the boys were at the party.

(4) A: Were some of the boys at the party?
B: Some of the boys were at the party.

A’s question in (3) makes it clear that she is interested in the precise number of boys who were at the party. By contrast, A’s question in (4) suggests that she is only interested in whether or not some of the boys were at the party. In other words, the information that not all of the boys were at the party is intuitively more relevant in the first dialogue than in the second. Consequently, in the second dialogue, B might have chosen to produce the informationally weaker ‘some’, not because she lacks evidence for the alternative containing ‘all’, but rather because she thought the hearer would have little or no interest in the extra information conveyed by the corresponding sentence with ‘all’ (the speaker might even consider the added information to be distracting for the hearer). In line with this observation, the scalar inference is intuitively more robust in the first dialogue compared to the second.

In a series of experiments, Zondervan (2010) investigated the effects of relevance on the robustness of the scalar inferences of ‘most’, which we may assume to pattern similarly to ‘some’ and ‘or’. Zondervan constructed vignettes that made the corresponding scalar inferences either relevant or irrelevant, where relevance was manipulated in various ways (e.g., by means of explicit questions, prosodic emphasis, or contextual cues). Participants were then presented with a statement containing the weaker term, even though the vignette made it clear that the statement with the stronger term was true. They had to indicate whether the statement was true or false, given the background story. Zondervan consistently found that scalar inference rates (i.e., ‘false’ responses)

were higher when the scalar inference was relevant than when it was not. However, the difference was typically rather small. For instance, in Exp. 1, the scalar inference of 'or' was drawn less than 20% more often when it was relevant (73% of the time) than when it was not relevant (55%).

2.3. PRIOR PROBABILITY. The nature of the effect of prior probability on the robustness of scalar inferences is more contentious than the effects of competence and relevance. To illustrate, consider the following sentence (from Geurts 2010):

(5) Cleo threw all her marbles in the swimming pool. Some of them sank to the bottom.

A priori, it is highly likely that all of the marbles sank to the bottom. How does this observation influence the robustness of the scalar inference? Geurts (2010) intuits that the strength of the scalar inference is unaffected by the fact that one would naturally expect all of the marbles to sink. Geurts' intuition contrast with the predictions made by the Rational Speech Act (RSA) model, a recent formalisation of the pragmatic reasoning process that underlies the derivation of scalar inferences (e.g., Frank & Goodman 2012). According to the RSA model, the prior probability of the stronger alternative should negatively correlate with the strength of the scalar inference, so that, in the example above, the scalar inference should be weak or even altogether absent.

Degen et al. (2015) experimentally investigated the effect of prior probability on the robustness of the scalar inference of 'some'. In Exp. 1, they presented participants with event descriptions such as 'John threw 15 marbles into a pool'. These event descriptions were followed by a question such as 'How many of the marbles sank?'. Participants had to indicate the probability of each possible event (e.g., one marble sinking, two marbles sinking, and so on). In Exp. 2, a different group of participants read the same event descriptions, but this time the descriptions were followed by a well-informed character producing an utterance like 'Some of the marbles sank'. Participants again had to indicate the probability of each possible event, but this time based on the utterance rather than their prior expectation. Degen et al. (2015) observed a correlation between prior probability and the strength of the scalar inference so that, when the 'all' situation was judged likely in Exp. 1, it was also judged likely in Exp. 2.

2.4. PREDICTIONS. To sum up, based on literature, we hypothesize that scalar implicatures are sensitive to these three contextual cues in the following way:

1. *Competence*: scalar inferences are more robust if the speaker is competent, i.e., knows whether or not the stronger alternative is true.
2. *Relevance*: scalar inferences are more robust if the information expressed by the stronger alternative is relevant to the hearer with respect to the purpose of the conversation.
3. *Prior probability*: scalar inferences are more robust if the information expressed by the stronger alternative is a priori likely to be false.

In this paper, we systematically investigate these effects and their interactions on the robustness of the scalar inferences associated with 'some' and 'or'. Our goals are twofold. First, we aim to obtain evidence as to which contextual factors influence the robustness of a scalar inference. Recent research has focused on variability in scalar inference rates across different scalar words (e.g., why the inference from 'some' to 'not all' is much more robust than the inference from 'pretty' to 'not beautiful', e.g., van Tiel et al. 2016). Here, we study variability in the robustness

of scalar inferences that make use of the same lexical scales. By doing so, we obtain a more direct insight into the mechanism that underlies pragmatic inferencing, which in turn may also provide important new knowledge about the factors that cause cross-scalar variability.

Second, we seek to compare the scalar inferences of 'some' and 'or' with respect to their sensitivity to the three contextual factors. If these two types of scalar inference are indeed caused by the same underlying mechanism—as the standard account argues—it is natural to expect that they are similarly sensitive to the pragmatic factors under investigation. Hence, if we observe marked differences in how competence, relevance, and prior probability affect the robustness of the scalar inferences of 'some' and 'or', that would provide at least circumstantial evidence in favour of the idea that they are aetiologically distinct, too.

To address these goals, we conduct an experiment comparing the effects of these factors for both triggers. We describe this experiment in the next section.

3. Experiment. To operationalize these factors experimentally, we designed context stories (vignettes) which we intuitively judged to score either high or low with respect to each factor of interest. That is, we manipulated the contextual *relevance* of the stronger alternative to the listener, the speaker's *competence* about the truth of the statement with 'all' or 'and', and the *prior probability* of the statement with 'all' or 'and' to be true. We studied how these manipulations influenced the the robustness of the scalar inferences of 'some' ('some but not all') and 'or' ('or, but not both'). The context stories and critical trials were designed in an analogous fashion for both triggers and varied within-subjects, so the following descriptions apply for both triggers.

3.1. MATERIALS AND PROCEDURE. This study was a $2 \times 2 \times 2 \times 2$ within-subjects rating task (relevance $\times$ competence $\times$ prior $\times$ trigger), conducted as a web-based experiment.[1]

On critical trials, participants were asked to rate four sentences, one per factor, and one containing an upper-bounded ('some') or exclusive ('or') paraphrase of the trigger. On each trial, participants read a context story, followed by a sentence presented in a blue box which was meant to elicit the likelihood rating for a given factor. The sentence had either of the following forms:

(6) *Relevance rating:* It is important to X to know whether Y.
Competence rating: Z knows whether Y.
'Or' prior rating: If A, then B. / If B , then A.
'Some' prior rating: If some Y, then all Y.
Inference strength rating: From what Z said we may conclude that W.

where X was the listener, Z was the speaker, Y was the target event in the background story and W was the event under the upper-bounded or exclusive reading. A and B were the disjuncts of Y for 'or' vignettes. For the inference strength elicitation, participants additionally read a critical utterance of the form:

(7) Z says to X: Y.

which contained the trigger ('some' or 'or'), presented below the background story in a red box, before rating the inference strength sentence.

[1]The experiment can be viewed at https://magpie-xor-some.netlify.app/. The preregistration for the experiment can be found at https://osf.io/v7zjp.

Below each sentence, participants were asked to indicate how likely it is that the statement in the blue box is true given the story, followed by a slider rating bar labeled 'certainly false' (left) and 'certainly true' (right). The slider positions were converted to 0–100 ratings.

We designed 32 different stories per trigger type ('some' vs. 'or'), resulting in a total of 64 stories. Each participant saw eight randomly sampled stories (one per prior × competence × relevance condition out of four possible stories) such that they saw four 'or' and four 'some' stories in randomized order. The assignment of conditions to the triggers was randomized between-participants.

The experiment proceeded as follows (see Fig. 1). First, participants were welcomed to the experiment and read instructions, which contained an annotated example to explain the meaning of the slider.

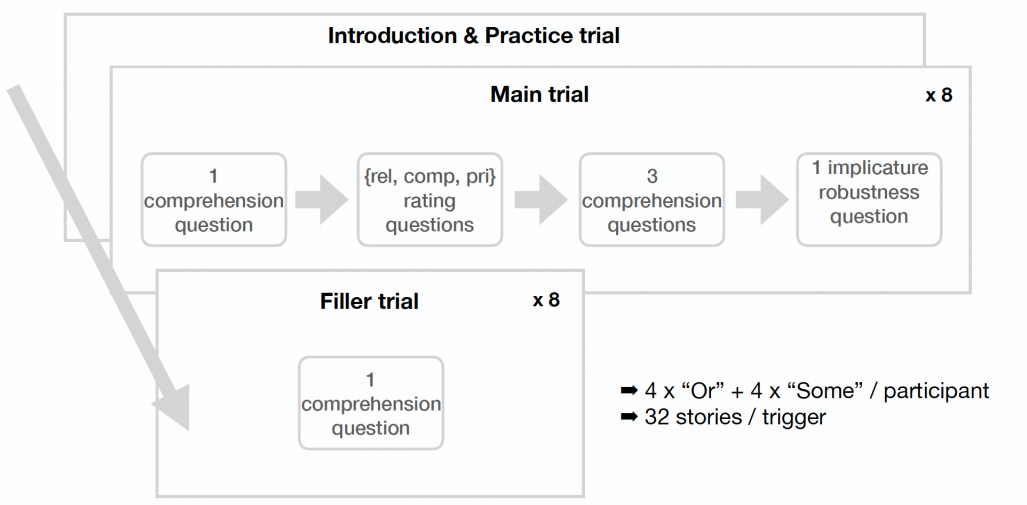

Figure 1: Experiment procedure.

The main part of the experiment consisted of eight critical vignettes, randomly shuffled with eight attention check vignettes. For each critical vignette, participants completed a block of 10 or 11 trials, consisting of a trial with a comprehension question, trials with statements eliciting relevance, competence and prior ratings (the last factor was elicited with two statements for the trigger 'or', see (6)), followed by three more trials with comprehension questions, and the critical scalar inference strength elicitation trial. Comprehension trials were visually identical to the critical trials, but the statement to be rated only referred to the content of the background stories. They were designed so as to be either clearly true, clearly false or uncertain given the background story. The four comprehension statements for one vignette were sampled at random from six possible statements (two true statements, two false statements, and two uncertain statements).

The attention checks consisted of one trial which visually matched the critical trials. On these trials, the vignettes contained a statement which wrote out what participants were supposed to answer (e.g., 'Please move the slider maximally left') in an area of text that participants had to read to complete the trial. Participants who failed more than two of these attention checks were excluded from the analysis. After the study, participants could voluntarily fill out a socio-demographic questionnaire.

3.2. PARTICIPANTS. We recruited 277 participants through the crowd-sourcing platform Prolific. Participants were restricted to those whose first language included English, who previously took part in at least five other Prolific studies, and whose approval rate was at least 0.9, according to prescreening criteria of the platform. Participants took 20 minutes on average to complete the study and were compensated £2.48 for their participation. Following our preregistered exclusion criteria, we excluded 15 participants for not indicating their native language, 3 participants for completing the study in under 8 minutes, 26 participants for failing more than two attention checks, and 27 for failing more than 20% of the comprehension questions. Due to a coding error in the comprehension questions, 206 participants were left after applying exclusion criteria, although the preregistered target sample size after exclusions was 200 subjects. In the following analyses, data from the 206 participants is analysed.[2]

4. Results. Prior to conducting statistical analyses, we preprocessed the data by standardizing (z-scoring) the responses within each factor (relevance, competence, prior) for each participant. Based on the aforementioned predictions, we expect that participants rate the scalar inference as more likely if (i) the alternative is rated as more relevant, (ii) the speaker is rated as being more competent, and (iii) the alternative is rated as a priori less likely. The remainder of this section is structured as follows: Section 4.1 provides descriptive and confirmatory analyses following preregistration and Section 4.2 provides additional exploratory analyses.

4.1. PREDICTOR RATINGS AND CONFIRMATORY ANALYSES. Descriptively, participants' factor ratings by-story agreed well with the designed classification of the stories (Fig. 2, red vs. blue color on x-axis). That is, the ratings for the relevance, competence and prior statements were not distributed uniformly across the stories, but aligned with our prior categorisations (Fig. 2, x-axis), validating our experimental manipulation of the explanatory factors.

We analysed the ratings using a Bayesian linear mixed effects model, regressing the target scalar inference ratings against the fixed effects of predictor ratings (i.e., the relevance, competence, and prior ratings elicited by the same participant for that vignette), the effect of trigger, and their two-, three-, and four-way interactions.[3] We included random intercepts and random slope effects for the main effects of trigger, relevance, competence and prior by-subject, as well as random intercepts by-vignette.[4] The categorical effect of trigger was dummy coded, using 'some' as the reference level. For all regression coefficients we used a wide and uninformative prior given by a t-distribution with mean 0, standard deviation of 2 and 1 degree of freedom. The model was fitted using the R `brms` package (Bürkner 2017).

We focus on the slope coefficients for the effects of relevance, competence and prior, once for 'some' and once for 'or'. We check whether the posterior estimate of each effect was in either one of three intervals: (1) negative effect ($\leq$ -0.05), (2) no effect (between -0.05 and 0.05), or

[2]All analyses were also conducted on data from 200 subject, excluding the last six submissions. No noteworthy quantitative or qualitative differences of the results were observed.

[3]The data and the analyses can be found under https://github.com/magpie-ea/magpie-xor-experiment

[4]Model in R syntax style: `inference-rating ~ trigger * relevance * competence * prior + (1 + trigger + relevance + competence + prior || subjectID) + (1 | vignette)`. For computational tractability reasons, the correlation of random effects by-subject was set to 0.

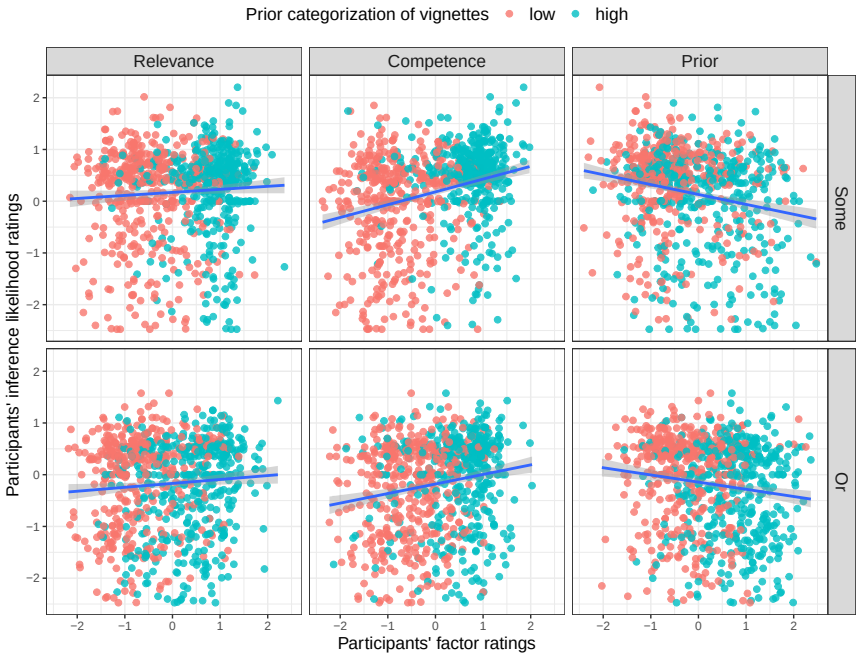

Figure 2: Ratings for relevance, competence and prior statements (x-axis) plotted against ratings for the strength of scalar inferences (y-axis). The top row shows ratings for 'some' (enriched to 'some, but not all'). The bottom row shows ratings for 'or' (enriched to 'or, but not both'). Ratings for stories initially categorised (by the experimenters) as low (red) w.r.t. a given factor are on average lower (x-axis) than for those categorised as high (blue).

(3) positive effect (≥ 0.05). We set the threshold for considering an effect as positive or negative to ±0.05 because we consider 0.05 to be the Region of Practical Equivalence (ROPE) for the effect sizes we expect (Kruschke 2014). We interpret the data as providing evidence in favor of an effect (positive, negative, no effect) if the posterior probability of the effect being true is $\geq$ 0.95 (i.e., 95% of posterior samples are in the corresponding interval). The probabilities of the respective coefficients lying in a particular interval are reported below (i.e., for instance, if P = 0.95 is reported, it means that 95% of the posterior samples of the given coefficient are in the respective interval). In particular, we speak of evidence in favour of the scalar inference account if the prior effect is negative, and the relevance and competence effects are positive. Fig. 3A shows examples of simulated posterior distributions over effect size samples which would confirm all our hypotheses (for better visual comparison to the observed results shown in Fig. 3B).

Consistent with predictions of the standard account, for the trigger 'some', we found a clear negative effect of prior probability of the stronger alternative 'all' being true, as indicated by the probability of the negative effect of prior being P = 0.999 (Fig. 3B, `Prior (some)`, orange color). Similarly, we found a clear positive effect for speaker competence (P = 1, Fig. 3B, `Competence (some)`, green color). However, we did not find a clear effect of relevance (Fig. 3B, `Relevance (some)`, split colors). If anything, the data supported the result that relevance may only marginally influence the robustness of the enriched interpretation. In con-

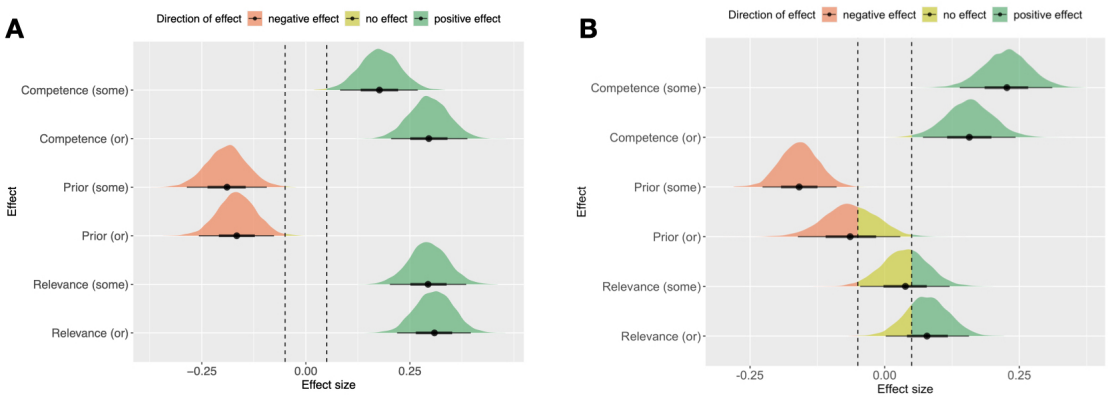

Figure 3: Distributions of the posterior samples of each factor of interest for the trigger 'some' and 'or'. The colors and the vertical dashed lines indicate the effect intervals (from left to right: negative, no, positive effect). Points indicate posterior means, thick lines indicate 66% credible intervals, thin lines indicate 95% credible intervals. A: Idealized hypothetical results implied by the standard account (for comparison to the obtained empirical results). The effect size was set to 0.25 for visualization purposes. B: Main experimental analysis results.

		negative	no effect	positive	by-subject random slope
some	relevance	0.006	0.519	0.474	0.07
	competence	0	0	1	0.13
	prior	0.999	0.001	0	0.16
or	relevance	0.001	0.204	0.795	0.07
	competence	0	0.007	0.993	0.13
	prior	0.635	0.357	0.008	0.23

Table 1: Results of the separate by-trigger exploratory models. Columns indicate probabilities of respective effects being in the indicated range, except for the column 'by-subject random slope' which shows estimates.

trast, for the trigger 'or', we only found a positive effect of competence (P = 0.993, Fig. 3B, `Competence (or)`, green color). We found no credible effects of the prior of the stronger alternative being true; results for relevance patterned with results for 'some' (Fig. 3B, `Prior (or)`, `Relevance (or)`, split colors). Therefore, the main analysis did not provide strong evidence in favor of the scalar inference account for 'or'. Comparing the overall results to 'some', they provide evidence against the identity hypothesis positing that the two triggers are interpreted via the same underlying mechanism.

4.2. EXPLORATORY ANALYSES. The descriptive results visually suggested a possible effect of prior for 'or' (see Fig. 2, lower right), which, however, was not borne out in the main analysis. To investigate the results for 'or' in more detail, we explored an analysis wherein the predictor ratings were averaged by-vignette and then regressed against the implicature strength ratings. This analysis amounts to averaging over the different participants and thereby removing possible by-subject

variability.[5] Under this model, additionally to the main results, there was a credible effect of prior probability for 'or' (P = 0.999 of a negative effect), suggesting that by-subject effects might have overridden the main effect in the main analysis. Following up, models for each trigger separately were fit (see Tab. 1). Those models included full random effects; the one for 'or' patterned with the main analysis, showing no credible prior effects (P = 0.635 of a negative effect). Yet, the largest by-participant random slope was the by-subject estimate for the effect of prior, compared to the other factors. Compared to the 'some' model, this slope estimate was also larger (Tab. 1, last column). Furthermore, an exploratory correlation analyses revealed a potential co-linearity between factors in the case of 'or' (R^2 = -0.106 for relevance and prior, R^2 = 0.127 for competence and relevance). No significant correlations were found for 'some'. Taken together, it seems that participants interpreted the prior statements for 'or' stories quite variably, which might be explained by differences in the prior expectations set up in the stories and whether these were contextual or based on world knowledge. This, in turn, might have influenced the (perceived) relevance to the listener and led to the observed correlated effects. Future research should address these aspects more systematically.

5. Discussion. Our experiment provides novel results on the effects of the factors relevance, competence and prior on the interpretation of 'some' and 'or'. However, future work may extend upon our results in several ways. First, our experiment considered how relevant the stronger alternative was to the hearer's interests. Yet other work rather focuses on relevance in terms of discourse purpose (e.g., van Kuppevelt 1996), which could be formalized, e.g., in the form of explicit questions in the background stories. Second, this paper looked at the derivation of the exclusive reading of 'or' through the lens of a scalar inference based account. However, alternative accounts derive the exclusive reading in terms of a distinctness condition, suggesting that disjunctions are infelicitous whenever the two disjuncts overlap, or either does not address the QUD (e.g., Simons 2001). Especially for the latter point, the disjuncts need to be interpreted exhaustively, which results in the exclusive reading. Our results provide no conclusive evidence with respect to this alternative account, calling for follow-up experiments manipulating exhaustivity and distinctness. Beyond that, it will be interesting to determine which other factors—e.g., typicality (van Tiel 2014), prosodic and linguistic prominence (Breheny et al. 2006), and politeness (Bonnefon et al. 2009)—might have similar effects. To sum up, our study provides new data on the effects of three contextual factors on the exclusive interpretation of 'or', and showed that they had a different effect than on the upper-bounded interpretation of 'some', calling into question the assumption that the interpretations of the two expressions are subject to the same pragmatic mechanisms.

References

Atlas, Jay David & Stephen C. Levinson. 1981. *It*-clefts, informativeness, and logical form: Radical pragmatics (revised standard version). In Peter Cole (ed.), *Radical pragmatics*, 1–62. Academic Press.

Bonnefon, Jean-François, Aidan Feeney & Gaëlle Villejoubert. 2009. When some is ac-

[5]Model in R syntax style: `inference-rating ~ trigger * relevance * competence * prior`

tually all: Scalar inferences in face-threatening contexts. *Cognition* 112(2). 249–258. https://doi.org/10.1016/j.cognition.2009.05.005.

Breheny, Richard, Napoleon Katsos & John Williams. 2006. Are generalised scalar implicatures generated by default? an on-line investigation into the role of context in generating pragmatic inferences. *Cognition* 100(3). 434–463. https://doi.org/10.1016/j.cognition.2005.07.003.

Bürkner, Paul-Christian. 2017. brms: An r package for bayesian multilevel models using stan. *Journal of statistical software* 80(1). 1–28. https://doi.org/10.18637/jss.v080.i01.

Degen, Judith, Michael Henry Tessler & Noah D. Goodman. 2015. Wonky worlds: Listeners revise world knowledge when utterances are odd. In *Proceedings of the 37th annual conference of the Cognitive Science Society*, 548–553.

Frank, Michael C. & Noah D. Goodman. 2012. Predicting pragmatic reasoning in language games. *Science* 336. 998. https://doi.org/10.1126/science.1218633.

Gazdar, Gerald. 1979. *Pragmatics: Implicature, presupposition, and logical form*. Academic Press.

Geurts, Bart. 2006. Exclusive disjunction without implicature.

Geurts, Bart. 2010. *Quantity implicatures*. Cambridge University Press. https://doi.org/10.1017/CBO9780511975158.

Goodman, Noah D. & Andreas Stuhlmüller. 2013. Knowledge and implicature: Modeling language understanding as social cognition. *Topics in Cognitive Science* 5. 173–184. https://doi.org/10.1111/tops.12007.

Grice, H. Paul. 1975. Logic and conversation. In Peter Cole & Jerry L. Morgan (eds.), *Syntax and semantics, volume 3: Speech acts*, 41–58. New York, NY: Academic Press. https://doi.org/10.1163/9789004368811_003.

Horn, Laurence R. 1972. *On the semantic properties of logical operators in English*: University of California, Los Angeles dissertation.

Kruschke, John. 2014. *Doing bayesian data analysis: A tutorial with r, jags, and stan*. Academic Press.

van Kuppevelt, Jan. 1996. Inferring from topics: Scalar implicatures as topic-dependent inferences. *Linguistics and Philosophy* 19. 393–443. http://www.jstor.org/stable/25001633.

Sauerland, Uli. 2004. Scalar implicatures in complex sentences. *Linguistics and Philosophy* 27. 367–391. https://doi.org/10.1023/B:LING.0000023378.71748.db.

Simons, Mandy. 2001. Disjunction and alternativeness. *Linguistics and Philosophy* 597–619. https://doi.org/10.1023/A:1017597811833.

Soames, Scott. 1982. How presuppositions are inherited: A solution to the projection problem. *Linguistic Inquiry* 13. 483–545. http://www.jstor.org/stable/4178288.

van Tiel, Bob. 2014. Embedded scalars and typicality. *Journal of semantics* 31(2). 147–177. https://doi.org/10.1093/jos/fft002.

van Tiel, Bob, Emiel van Miltenburg, Natalia Zevakhina & Bart Geurts. 2016. Scalar diversity. *Journal of Semantics* 33. 137–175. https://doi.org/10.1093/jos/ffu017.

Zondervan, Arjen. 2010. *Scalar implicatures or focus: An experimental approach*: Utrecht University dissertation.