

Exclusive disjunction

Michael Franke

Department of Linguistics, University of Tübingen, Germany

Bob van Tiel

Department of Languages and Literature, Université Libre de Bruxelles, Belgium

Corresponding author:

Michael Franke

Department of Linguistics

Wilhelmstraße 19

72074 Tübingen

Germany

mchfranke@gmail.com

Tel: +49-7071-2973970

Fax: +49-7071-295212

Acknowledgements:

MF's work was supported by by the Institutional Strategy of the University of Tübingen (Deutsche Forschungsgemeinschaft, ZUK 63) and the Priority Program XPrag.de (DFG Schwerpunktprogramm 1727). BT's work was supported by the F.R.S.-FNRS Research Incentive Grant F.4502.15 (principal investigator: Mikhail Kissine).

Exclusive disjunction¹

Abstract

If someone says “Donald ate a pretzel or a donut” the hearer may infer that Donald did not eat both a pretzel and a donut. This exclusive reading of “or” is often explained as a scalar implicature by comparison to an alternative utterance with “and”. We tested this explanation by investigating how the robustness of the exclusive reading of “or” is influenced by three contextual factors: relevance, competence, and prior probability. We found that only prior probability has a significant effect on the robustness of the exclusive reading, thus disconfirming the scalar implicature account. Instead, we propose that the exclusive reading of “or” is a probabilistic inference based on world knowledge.

Keywords: exclusive disjunction, conversational implicature, pragmatic reasoning, world knowledge

1 Introduction

Introductions to logic usually distinguish between two readings of “or”: an inclusive and an exclusive reading (e.g., McCawley, 1981; Copi & Cohen, 2005). The inclusive reading corresponds to the meaning of logical disjunction. According to this reading, “ A or B ” is true if at least one and possibly both of A and B are true. The exclusive reading is more strict in that it excludes the possibility that both A and B are true. So on its exclusive reading, “ A or B ” is true if exactly one of A and B is true. To illustrate, consider:

(1) Joe supports Donald or Hillary.

On its inclusive reading, this sentence is true whenever Joe supports Donald, Hillary, or both. The exclusive reading, which is arguably more prominent in this example, excludes the latter possibility. That is, on its exclusive reading, (1) is true whenever Joe supports Donald or Hillary, but not both.

The presence of these two readings might suggest that “or” is associated with two different lexical entries (e.g., Basson & O’Connor, 1960; Baum, 1996; Rescher, 1964). There are, however, good reasons to reject such a lexicalist approach. Perhaps the most compelling reason is that, in some contexts, “or” can only receive one interpretation. To illustrate, consider the following sentence, in which “or” occurs in the scope of negation:

(2) Joe does not support Donald or Hillary.

This sentence has only one interpretation, namely that Joe supports neither Donald nor Hillary (cf. Crain, 2008). This interpretation corresponds to the inclusive reading of “or”. It does not have an interpretation corresponding to the exclusive reading; in other words, it does not have a reading according to which Joe either supports both Donald and Hillary, or neither of them. Since “or” is systematically monosemous in certain contexts—negation being one of them—it follows that the two readings of “or” cannot simply be due to a lexical ambiguity.

An attractive alternative to the lexicalist approach stems from Grice’s (1975) theory of conversational implicature. According to Grice, natural conversation is governed by the assumption that speakers are cooperative; that is, they attempt to further the purpose of the discourse by means of their utterances. Speakers are cooperative by adhering to four maxims that enjoin their utterances to be truthful, informative, relevant, and clear. Sometimes hearers have to make ancillary assumptions in order to align a

speaker's utterance with the assumption of cooperativity. Such ancillary assumptions are called *conversational implicatures*. An example, from Grice's own work, is given in the following conversation:

(3) A: I am out of petrol.

B: There is a garage around the corner.

B's utterance would be irrelevant, and hence uncooperative, if he knew that the garage was closed or did not sell petrol. Since A assumes that her interlocutor is cooperative, she thus derives the conversational implicature that the garage is open and sells petrol.

Horn (1972) was the first to argue that the exclusive reading of "or" can be explained as a conversational implicature, too, based on the assumption that the primary meaning of "or" is inclusive. To illustrate, consider (1) again. Assuming that the primary meaning of "or" is inclusive, the speaker of (1) could have been more informative, and hence cooperative, by uttering the "and"-alternative "Joe supports Donald and Hillary." Why didn't she? Presumably because she does not believe that the "and"-alternative is true. This is a weak inference that is compatible with a situation in which the speaker is unsure about whether or not Joe supports both Donald and Hillary. This weak inference can be strengthened if there is reason to believe that the speaker knows whether or not Joe supports both Donald and Hillary. This assumption is often called the *competence assumption* (e.g., Geurts, 2010; Russell, 2006; Schulz & van Rooij, 2006; Zimmermann, 2000). If the competence assumption is sufficiently plausible, it follows that, according to the speaker, Joe does not support both Donald and Hillary.

This specific kind of conversational implicature is often called a *scalar implicature* because it is assumed that "or" forms a lexical scale with "and", and that alternatives are generated by substituting the scalemates "or" and "and". Other lexical scales are ⟨some, all⟩, ⟨warm, hot⟩, and ⟨intelligent, brilliant⟩.

The implicature account straightforwardly explains the absence of exclusive readings under negation. Consider (2) again. Here, unlike in the case of (1), the “and”-alternative is less informative than the utterance itself. After all, (2) is only compatible with a situation in which Joe supports neither Donald nor Hillary, whereas the “and”-alternative is also consistent with situations in which Joe supports exactly one of Donald and Hillary. So the speaker was already maximally informative and hence no implicature is derived.

Although the implicature account has since become the standard in the literature (e.g., Chevallier, Noveck, Nazir, Bott, Lanzetti, & Sperber, 2008; Chierchia, Fox, & Spector, 2012; Geurts, 2010; Sauerland, 2004), it is not without its problems, as we will see in the next section. Therefore, we set out to experimentally test the adequacy of the implicature account. To that end, we investigated the effect of three factors on the robustness of exclusive readings: (i) the *relevance* of the “and”-alternative to the hearer, (ii) the *competence* of the speaker about the truth of the “and”-alternative, and (iii) the *prior probability* that the “and”-alternative is true. These factors are usually taken to influence the robustness of scalar implicatures. Hence, if the implicature account is correct, we expect these factors to influence the robustness of the exclusive reading of “or”, too.

For comparison, we also investigated how relevance, competence, and prior probability influence the robustness of a bona fide scalar implicature, namely the inference from “some” to “not all”. To illustrate, consider:

(4) Some of my friends support Donald.

An utterance of this sentence may convey that not all of the speaker’s friends support Donald. The derivation of this upper-bounded construal runs analogous to that of the exclusive reading of “or”: the speaker could have been more informative by saying “All of my friends support Donald”. Why didn’t she? Presumably because she does

not believe that this alternative is true. If, moreover, the speaker knows whether or not all of her friends support Donald—i.e., if the speaker is competent in the relevant sense—it follows that she believes that not all of her friends support Donald.

As we will see, it turns out that there are marked differences in how relevance, competence, and prior probability influence the robustness of the exclusive reading of “or” compared to how they influence the robustness of the inference from “some” to “not all”. These findings will encourage us to reconsider the implicature account of the exclusive reading of “or”.

In the next section, we explain our three factors of interest in more detail. Afterwards, we outline the details of our experiments and discuss the results.

2 Relevance, competence, and prior probability

Relevance

Almost all theorists agree that the robustness of scalar implicatures is modulated by various extralinguistic factors (e.g., Chierchia et al. 2012; Franke 2009; Geurts 2010; Horn 1972, but see Chierchia 2004; Levinson 2000; Storto & Tanenhaus 2005). Some of these factors are specific to certain theories, while others are shared across competing theories. Relevance is of the latter kind in that it features in almost all theories of scalar inferences. However, relevance is a multivocal notion, and there are at least two construals that should be distinguished: relevance for the discourse purpose and relevance for the hearer’s interest (Geurts, 2010).

The notion of relevance for the discourse purpose can be illustrated with the following example (from van Kuppevelt 1996):

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

Discourse purposes are usually equated to questions, which can be implicit or, as in this case, explicit (Roberts, 2012). A piece of information is relevant if it contributes to answering this question. In (5), the information that not all of the boys went to the party is relevant to the discourse purpose because it narrows down the space of possible answers. Hence, it is hypothesised that “some” will be interpreted as “some but not all”.

Compare this to a situation in which B’s answer addresses the question “Were some of the boys at the party?” In this situation, the upper-bounded construal is irrelevant, since the question is already resolved when “some” is interpreted as “at least some and possibly all”. In this situation, then, it is predicted that no scalar implicature will be derived.

In this paper, however, we will mainly be concerned with the second notion of relevance: relevance to the hearer’s interest. This construal is intimately connected to *relevance theory* (Sperber & Wilson, 1995).

According to relevance theory, the relevance of an inference is determined by two factors: (i) the positive cognitive effects it has on the hearer—i.e., to what extent it makes “a worthwhile difference to the individual’s representation of the world” (Wilson & Sperber, 2002, p. 251)—and (ii) the effort needed to process that inference. In what follows, we focus on the first factor, assuming that processing effort remains invariant across the scenarios that we tested (cf. Chevallier et al. 2008 for a series of experiments on “or” that manipulate processing effort).

The central tenet of relevance theory is that communication is aimed at maximising relevance, so that hearers only derive inferences whose positive cognitive effects outweigh their processing cost. Hence, relevance theory predicts that the robustness of a scalar implicature is an increasing function of its importance to the hearer. To illustrate, consider:

(6) Donald won the primaries in Nebraska or Indiana.

According to relevance theory, the robustness of the exclusive reading of this sentence depends on how important it is to the hearer whether Donald won the primaries in both Nebraska and Indiana. Suppose, for example, that the hearer made a large bet that Donald would not win both primaries. In that case, she would be more likely to arrive at an exclusive reading than if, for example, she made a large bet that Donald would win at least one of the primaries.

Hence, if relevance theory is correct, we expect the robustness of the inference from “some” to “not all” to be positively affected by the relevance of the upper-bounded reading to the hearer. If, in addition, the implicature account is correct, we expect the same effect on the exclusive reading of “or”.

Competence

In the previous section, we have seen that competence plays an important role in the derivation of scalar implicatures. Initially, pragmatic reasoning yields a weak inference according to which the speaker does not believe that the more informative alternative is true. This weak inference can be strengthened if the competence assumption holds; that is, if it is assumed that the speaker knows whether or not the more informative statement is true. If so, it follows that, according to the speaker, the more informative statement is false.

Although the workings of the competence assumption tend to be unproblematic for most scalar expressions, they are rather precarious in the case of “or”. Indeed, the need to invoke competence to arrive at an exclusive reading of “or” gives rise to what Zondervan (2010) calls the *speaker expertise paradox* (cf. Geurts, 2006). To explain this paradox, consider (1) once again:

(7) Joe supports Donald or Hillary.

We have already seen that someone who utters this sentence is likely to convey that Joe

does not support both Donald and Hillary. Presumably, however, the utterance is also likely to convey that the speaker does not know whether Joe supports Donald and that she does not know whether Joe supports Hillary. After all, if the speaker knew that Joe supports Donald, she should have said “Joe supports Donald”, and *mutatis mutandis* if she knew that Joe supports Hillary. However, in order to arrive at the exclusive reading through pragmatic reasoning, it has to be assumed that the speaker knows whether Joe supports both Donald and Hillary.

More abstractly, in order to derive the exclusive reading of a sentence of the form “*A* or *B*”, it has to be assumed that the speaker is ignorant about the truth of *A* and *B* individually, but knowledgeable about the truth of “*A* and *B*”. Such an epistemic state is possible but intuitively improbable, which contradicts the observation that exclusive interpretations are far from uncommon.²

In order to arrive at a more decisive verdict, however, we tested the effect of competence on the robustness of the exclusive reading of “or” and the upper-bounded construal of “some”. Previous experimental research has shown that the robustness of the upper-bounded reading of “some” increases with the competence of the speaker (Goodman & Stuhlmüller, 2013). We expect to replicate this effect in our study. If, moreover, the exclusive reading is due to a scalar implicature, we expect a similar effect in the case of “or”.

Prior probability

Compared to the previous two factors, the role of prior probabilities is more controversial. Some theorists have marginalised the importance of prior probabilities (Geurts, 2010), while others have assigned them an important role in the derivation of scalar implicatures (e.g., Frank & Goodman, 2012; Franke & Jäger, 2016; Russell, 2012). According to at least some of these accounts, the robustness of a scalar implicature is an increasing function of the prior probability that it is true. To illustrate, consider the

following sentence once again:

(8) Donald won the primaries in Nebraska or Indiana.

Theories that are sensitive to prior probabilities predict that, if it is deemed extremely unlikely that Donald wins the primaries in both states, someone who hears an utterance of (8) will be more likely to arrive at an exclusive reading of “or”, compared to when the hearer is expecting Donald to win both primaries.

Interestingly, a number of theorists have argued that the robustness of exclusive readings is exhaustively determined by prior probabilities (e.g., Barrett & Stenner, 1971; Rubin & Young, 1989; Yanal, 1988). These authors hold that “or” is monosemous and always interpreted inclusively. The appearance of an exclusive reading is caused by probabilistic reasoning about possible states of the world. An especially clear example that illustrates this point is:

(9) Joe voted for Donald or Hillary.

Here, it seems that the speaker’s utterance rules out the possibility that Joe voted for both Donald and Hillary. According to the aforementioned theorists, this inference is not due to an exclusive reading of “or” but rather to an inclusive reading along with the commonsense information that one can only vote once in a democratic election. In this way, then, world knowledge is responsible for seemingly exclusive interpretations of “or”. Importantly, this position is distinct from the implicature account because it holds that the derivation of exclusive readings does not involve any reasoning about the intentions of the speaker.

Previous experimental research has confirmed that the robustness of the upper-bounded reading of “some” is positively affected by the prior probability that it is true (Degen, Tessler, & Goodman, 2015). We expect to replicate that result in our study. In addition, if exclusive readings are due to scalar implicatures, we expect a similar effect

in the case of “or”.

Predictions

In summary, we have discussed three factors that have variously been taken to increase the robustness of scalar implicatures: (i) the relevance of the more informative statement—i.e., the “and”- and “all”-alternatives—to the hearer, (ii) the competence of the speaker about the truth of the more informative statement, and (iii) the prior probability that the more informative statement is false.

In the next section, we discuss two experiments in which we measured the effect of relevance, competence, and prior probability on the robustness of, on the one hand, the upper-bounded construal of “some” (Exp. 1) and, on the other hand, the exclusive reading of “or” (Exp. 2). We hypothesise that all three factors influence the robustness of the upper-bounded construal of “some”. More concretely, we hypothesise that the upper-bounded construal of “some” should be more prominent under higher contextual relevance, higher speaker competence, and lower a priori probability of the “all”-alternative being true. If the implicature account is correct, we expect that the exclusive reading of “or” patterns with the upper-bounded construal of “some” in these respects.

Since Exp. 1 and 2 share the same methodology, we discuss them in tandem. Afterwards, in Exp. 3, we discuss and test an alternative version of the implicature account, according to which statements of the form “*A* or *B*” are read, in effect, as “*A* but not *B* or *B* but not *A*”.

3 Experiments 1 and 2

Design

The goal of Exps. 1 and 2 was to determine the effects of relevance, competence, and prior probability on the strength of the upper-bounded construal of “some” (Exp. 1) and the exclusive reading of “or” (Exp. 2). To measure these three factors, we designed different stories that systematically varied along the three relevant dimensions. We used a slider-rating task to assess participants’ intuitive judgement of all four notions of interest: strength of the inference, relevance, competence, and prior probability.

Intuitive judgements for the four notions of interests were measured between-participants: each participant only answered one of the four relevant test questions for a given story. This was to prevent cross-contamination of answers, e.g., asking first about relevance might influence subsequent answers about the strength of the inference.

Participants

203 (Exp. 1) and 200 (Exp. 2) participants were drafted on Amazon’s Mechanical Turk and paid 80 US\$ cent.³ Payment was not contingent on any of their responses. Only workers with an IP address from the United States and with a rate of accepted HITs of at least 90% were eligible for participation.

Materials

The materials in both Exp. 1 and 2 consisted of 16 short stories. Each story came with some background information and an utterance of a statement containing “some” (Exp. 1) or “or” (Exp. 2) by one of the characters in the story. For example:

Example story from Exp. 1

Lucy has to give a talk in front of a big audience of psychologists. She is going to criticize one of the dominant theories of schizophrenia. Afterwards, Jacob, who was in the audience, chatted with his neighbors.

He tells Lucy: “Some of the people enjoyed your talk.”

Example story from Exp. 2

Danny and Alex reserved a squash court but Alex still has to buy a racket and a pair of shoes. Danny is talking to Alex’s girlfriend Jill who just went to the sports store with him.

Jill says to Danny: “Alex bought a racket or a pair of shoes.”

Each story was associated with three control statements which were either certainly true, certainly false, or of uncertain truth value, given the background information. Moreover, each story was associated with four target statements gauging (i) the strength of inference, (ii) the relevance for the hearer of the truth of the stronger statement (with “all” or “or”), (iii) the competence of the speaker, and (iv) the prior probability of the stronger statement. The target statements associated with the examples above were:

Example target statements from Exp. 1

Inference: From what Jacob said we may conclude that not all of the people enjoyed Lucy’s talk.

Relevance: It is important for Lucy to know whether all of the people enjoyed her talk.

Competence: Jacob knows whether all of the people enjoyed Lucy’s talk.

Prior: All of the people enjoyed Lucy’s talk.

Example target statements from Exp. 2

Inference: From what Alex's girlfriend said we may conclude that Alex did not buy both a racket and a pair of shoes.

Relevance: It is important for Danny to know whether Alex bought both a racket and a pair of shoes.

Competence: Alex's girlfriend knows whether he bought both a racket and a pair of shoes.

Prior: If Alex bought a racket, it is likely that he also bought a pair of shoes. / If Alex bought a pair of shoes, it is likely that he also bought a racket.

Statements *Inference*, *Relevance*, *Competence* (Exps. 1 and 2), and *Prior* (Exp. 1) were single statements. Statements *Prior* (Exp. 2) were pairs of symmetric conditional statements, where each targeted the intuitive probability that, given one disjunct, the other would be true as well. We reasoned that this makes for more natural statements and that it may give us more reliable measures than having participants rate a single statement containing “and”.⁴

The stories were created to ensure sufficient variability across the three dimensions of interest (i.e., relevance, competence, and prior probability). We classified each story according to whether we intuited it to be high or low on each dimension, thus making for eight types of stories. In both experiments, we had two stories for each type. For example, we expected the example story from Exp. 1 to score high on relevance, but low on competence and prior probability, and the example story from Exp. 2 to score high on all three dimensions. A full list of the 32 stories, together with our intuitive type-classification, can be found in Appendix A and B.

Procedure

The experiments started with instructions:

In the following, you will be presented with 8 short background stories. Please read them very carefully. We ask you to rate 2 or 3 statements for each background story. Please indicate, using an adjustable slider, how likely you think a statement is true based on the background story.

Next, we presented a simple background story which was not used in the main experiment, followed by three annotated examples to illustrate the use of the slider bar. One example was clearly true, another clearly false, and the last uncertain.

In the main part of the experiment, every participant saw eight randomly sampled stories, one of each story type, in random order. Each story was followed first by one random control statement and then the statement(s) associated with one of the four factors of interest (inference strength, relevance, competence, or prior probability). Each participant rated each of the four statement types exactly twice, but never in direct succession. When the prior statements were presented, only the background story was provided, but not the utterances with “some” or “or”. This was done to make sure that answers are based on expectations about worldly events alone, unmodulated by information based on pragmatic inferences. All other question types had the background story and the utterances with “some” or “or” appear on the screen. The two prior statements in Exp. 2 were presented individually, one after the other, in random order.

Ratings of statements were elicited by asking “How likely do you think it is that the statement is true, given the information in the background story?” together with a continuous slider ranging from “certainly false” to “certainly true.” An example of a trial from Exp. 2 is given in Figure 1.

Background: Lucy has to give a talk in front of a big audience of psychologists. She is going to criticize one of the dominant theories about schizophrenia. Afterwards, Jacob, who was in the audience, chatted with his neighbours.

He tells Lucy: 'Some of the people enjoyed your talk.'

Statement: From what Jacob said we may conclude that not all of the people enjoyed Lucy's talk.

How likely do you think it is that the statement is true, given the information in the background story?



Figure 1: Example of a trial from Exp. 1

Data preparation

We coded the slider-ratings as real numbers ranging from 0 (“certainly false”) to 1 (“certainly true”). Ratings for the two conditional statements in the prior condition of Exp. 2 were averaged.

Four participants were excluded from the analyses of Exp. 1 because they were not self-reported native speakers of English; three for Exp. 2. We also removed another six (Exp. 1) and five (Exp. 2) participants for obviously deviant answers (e.g., blindly alternating between maximal agreement and maximal disagreement).⁵ Consequently, data from a total of 193 (Exp. 1) and 192 (Exp. 2) participants made it into our analysis. Unfortunately, there was a mistake in the formulation of two stories, one for each experiment. We removed all data for these stories for the analysis. (See Appendix A and B, especially footnotes 8 and 9.)

Results

Controls. Control statements were rated as expected, indicating that participants understood the task and paid attention to the background stories. For Exp. 1, the averages

over all stories of ratings for false (0.21), uncertain (0.40) and true statements (0.80) are pairwise different (two-population directed Welch’s t -test: $t \approx -4.04$, $df \approx 27.9$, $p < 0.001$ for false vs. uncertain; $t \approx -9.23$, $df \approx 27.8$, $p < 0.001$ for uncertain vs. true). Similarly, for Exp. 2, the averages over all stories of ratings for false (0.26), uncertain (0.48) and true statements (0.81) are pairwise different (two-population directed t -test: $t \approx -3.92$, $df \approx 26.6$, $p < 0.001$ for false vs. uncertain; $t \approx -7.20$, $df \approx 27.1$, $p < 0.001$ for uncertain vs. true).

Explanatory factors. We are interested in whether the ratings of statements for relevance, competence and prior probability are good explanatory factors of the strength of the upper-bounded construal of “some” (Exp. 1) or of the strength of the exclusive readings of “or” (Exp. 2). We therefore look at explanatory factors REL, COM and PRI, which are the means of the ratings, for each story, of the *Relevance*, *Competence* and *Prior* statements. Before turning to our main analysis, we begin with a few observations about these explanatory factors.

First of all, ratings of the relevant statements are not uniformly distributed across stories, but validate our intuitive pre-classification (see Figure 2, all high/low contrasts are significant). Moreover, Figure 3 shows how REL, COM and PRI correlate with the mean ratings of the strength of the upper-bounded construal of “some” and of the strength of the exclusive readings of “or”. From visual inspection, it seems that in Exp. 1 (“some”) REL and COM are not likely to be strong predictors of the strength of the upper-bounded reading, while low values of PRI seem to be correlated with high ratings, as expected. For Exp. 2 (“or”), the plots suggest that REL might not be correlated with the reported strength of exclusive readings, while there do seem to be noteworthy correlations with COM and PRI. Interestingly if exclusive readings of “or” are like bona-fide scalar implicatures, we should expect higher values of COM to be associated with higher endorsements of exclusive readings. What we observe,

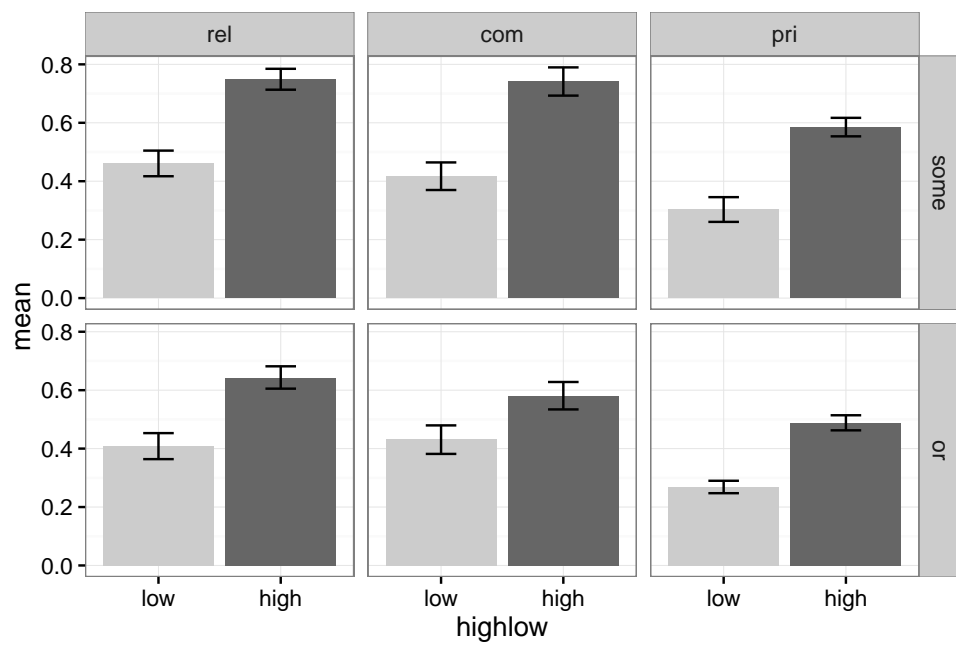


Figure 2: Ratings of statements according to intuitive pre-classification. Results for Exp. 1 (“some”) in the top row, results for Exp. 2 (“or”) in the bottom row. Error bars indicate bootstrapped 95% confidence intervals around the plotted means.

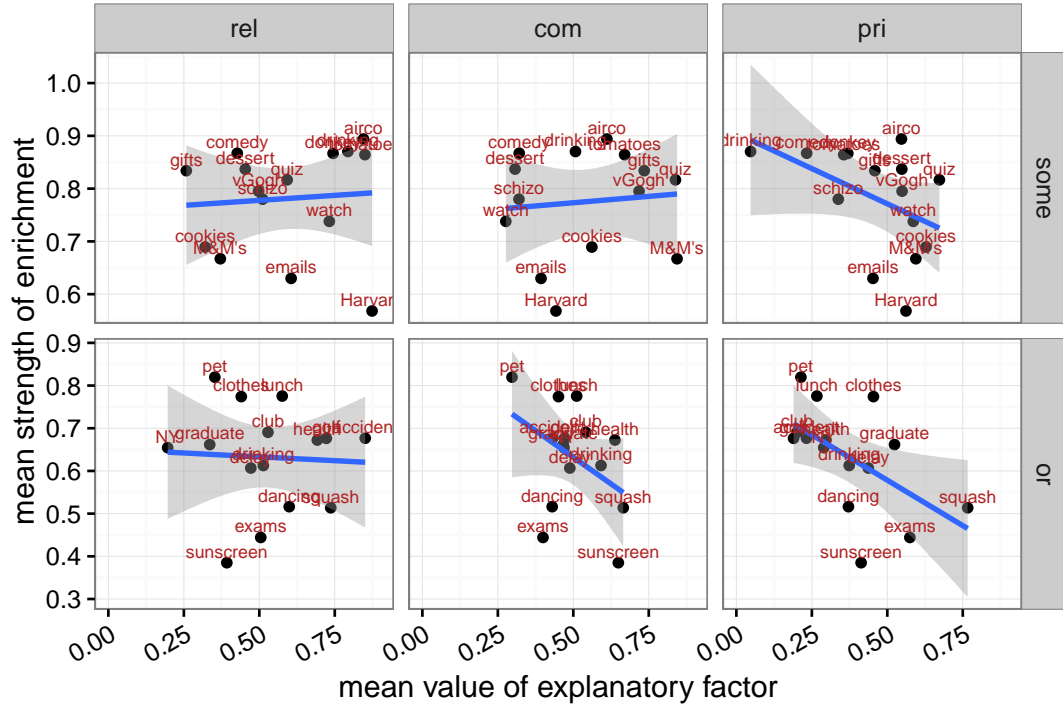


Figure 3: Relating ratings for relevance, competence and prior statements (x -axis) to ratings for the strength of pragmatic enrichments (y -axis). The top row shows ratings, averaged for each story, for Exp. 1 (“some”). The bottom row shows ratings, averaged for each story, for Exp. 2 (“or”).

however, appears to be the opposite relation.

In the light of subsequent regression models, it is important to note here that there is no significant correlation between any pair of REL, COM and PRI for neither experiment. That is, the measures of REL and COM for all stories of Exp. 1 are not correlated, neither are these of REL and PRI, et cetera.

Main analysis. To check whether factors REL, COM and PRI have an influence on the ratings of the *Inference*-statements—i.e., the strength of the upper-bounded construal in Exp. 1 and of the exclusive reading in Exp. 2—we compare regression models of different complexity. We take a Bayesian approach to model comparison here in terms of their Bayes factors (Rouder & Morey, 2012), as implemented in the *BayesFac*-

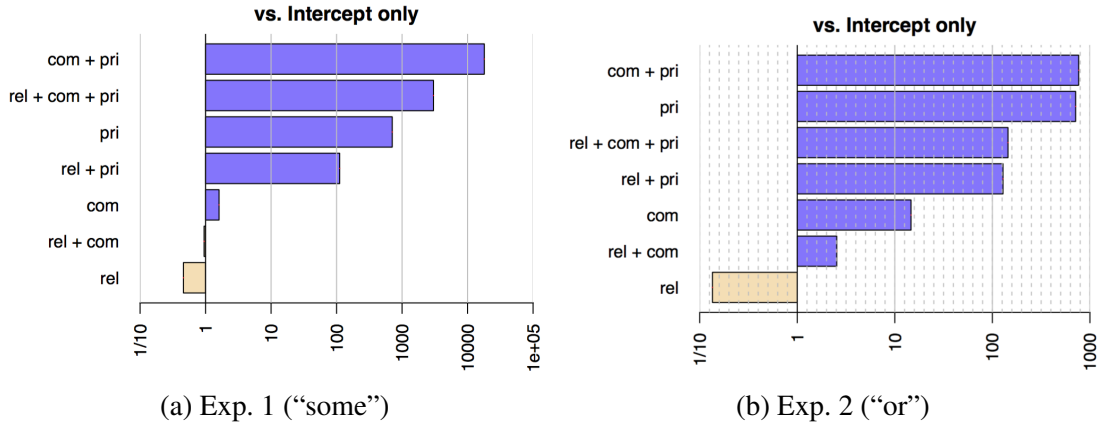


Figure 4: Bayes factor comparison of different main factor combinations. Notation like “com + pri” stands for a regression model with main factors COM and PRI. The plots show the Bayes factor on the logarithmic x -axis of the models on the left, when compared to an intercept only regression model.

tor R-package. This gives us an intuitively accessible and quite nuanced picture of the relative evidence for models of different complexity, including information about how much, e.g., the absence of a factor in a model, is supported by our data.⁶

Figure 4 shows the Bayes factors of all regression models that can be built with our three explanatory variables as main factors. The graphs give the Bayes factors of each regression model listed on the left against the intercept-only model. Given the data from Exp. 1, for example, the best model has main factors COM and PRI. It is more than 10,000 times more likely than the intercept only model. The model with the lowest a posteriori likelihood is a regression model that contains only factor REL, which is even worse than the intercept-only model. Given the data from Exp. 2, the best and worst models are the same, but there are differences as to how likely some of the other models appear to be, in the light of the data.

For Exp. 1, the best model COM + PRI is very close to six times more likely than the full model REL + COM + PRI, which in turn, is about four times more likely than the model PRI. It is common to consider Bayes factors above 3 as sufficiently noteworthy evidence in favour of a model (e.g. Jeffreys, 1961; Raftery, 1995; Lodewyckx, Kim,

Lee, Tuerlinckx, Kuppens, & Wagenmakers, 2011). The picture that emerges is that, for Exp. 1, the addition of factors PRI and COM increases the a posteriori likelihood, while the addition of factor REL actually decreases it. This suggests that the former are explanatory factors whose inclusion may make a model more complex but also leads to better predictions. In contrast, the added model complexity of including REL is not countered by a similar increase in predictive power; we are better off without factor REL.

Turning to Exp. 2, the best model COM + PRI is not noteworthy better, in terms of Bayes factors, than the second best model which only includes PRI. Still, both of these models come up noteworthy more likely, given the data, than the third best model, the full model with REL + COM + PRI: a Bayes factor of around five. The overall picture that emerges in this case is that adding REL to a model seems to decrease a posteriori likelihood, while adding COM or PRI increases it. The most important single factor clearly is PRI: when a model contains PRI, adding COM does not lead to substantial improvements. The main conclusion to be drawn from this analysis is that REL is a bad, COM an unnecessary, and PRI the best predictor of strength of exclusive readings.⁷

Finally, inspired by previous observations based on Figure 3, let us also look at the way that speaker competence influences the *Inference*-ratings in each experiment. What can be gleaned already from the plots in Figure 3, is supported by inspection of estimates of regression coefficients. Figure 5 shows posterior distributions for regression coefficients of the full model REL + COM + PRI after conditioning with the data from Exp. 1 (top row) and Exp. 2 (bottom row). We see that credible values for coefficients for COM are positive for Exp. 1, as expected, but mostly negative for Exp. 2. This means that our data suggest that the more competent the speaker was felt to be, the lower the strength of the exclusive reading. This is the inverse of what a scalar implicature account would suggest and also the inverse of what we find for the upper-bounded construal of “some”.

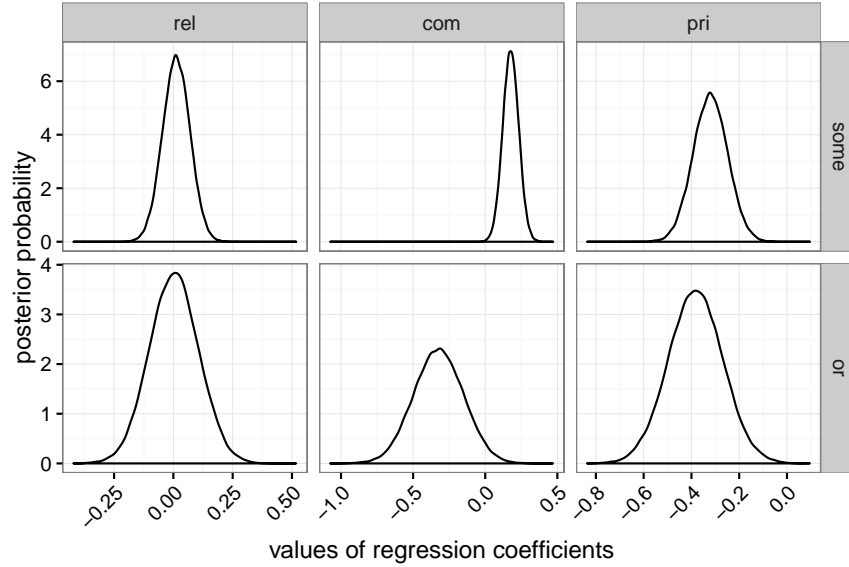


Figure 5: Density estimates of posterior over model parameter coefficients for a linear model with all three main factors.

Discussion. The main explanatory factors for *Inference*-ratings for “some” appear to be the speaker’s competence and the prior probability of the “all”-alternative. These findings are not unexpected under standard scalar implicature accounts except, perhaps, that there does not seem to be the predicted influence of relevance. As for “or”, it seems that prior probability of the stronger alternative is the main explanatory factor of *Inference*-ratings. In contrast to the results for “some”, there does not seem to be a strong effect of speaker competence on the perceived strength of exclusive readings. This is surprising under standard scalar implicature theories of exclusive “or”, even more so when we also consider the “direction” of the relation between speaker competence and inference ratings. While in Exp. 1 higher speaker competence correlated with higher scalar implicature inference rates, this relationship was reversed in Exp. 2, contrary to the expectations generated by a standard scalar implicature account of exclusive inferences.

These findings thus appear to jeopardise the standard scalar implicature account

of exclusive disjunction. Before drawing firm conclusions, however, the next section first considers an alternative approach that is similar to the standard account in that it assumes that the exclusive reading of “or” is a conversational implicature, but that is also different from the standard account in a number of important respects.

4 Experiment 3

Motivation

There is a much less prominent alternative approach to deriving exclusive disjunction readings as a conversational implicature. We will refer to it as the *distinct disjuncts* account here. It takes as its point of departure the observation that statements with “or” tend to be infelicitous when there is some overlap between the propositions that the individual disjuncts express (Hurford, 1974; Simons, 2001; Zimmermann, 2000). Hence, the following sentences appear to be pragmatically marked:

- (10) a. Ivan is an American or a Californian.
b. That painting is of a man or a bachelor.

The reason for this infelicity is that being a Californian entails being an American, and that being a bachelor entails being a man. Hence, there is overlap between the disjuncts, which violates the distinctness condition.

In addition, statements with “or” are infelicitous unless both disjuncts are answers to some question under discussion (Simons, 2001). For that reason, B’s answer in the following exchange is infelicitous:

- (11) A: Where is Donald?
B: He is in his room or fleas are annoying.

The reason for this infelicity is that it is difficult to imagine that the information that fleas are annoying provides an adequate answer to the question that A asked. Since one of the disjuncts fails to provide an adequate answer to the question under discussion, the entire statement comes out as infelicitous.

A final relevant observation is that answers to questions tend to be interpreted exhaustively. To illustrate, consider the following dialogue:

(12) A: Who does Joe support?

B: Donald.

Here, B's answer will be construed as implying that Joe supports Donald but not Hillary. In other words, B's answer is interpreted as being exhaustive. Exhaustivity in answers is usually explained as a conversational implicature along the following lines: if B knew that Joe supports, e.g., both Donald and Hillary, it would have been informative and relevant to mention that. Since he did not, the hearer may conclude that, according to B, Joe supports Donald but not Hillary.

Given that disjuncts have to provide an answer to the question under discussion, and given that answers to questions tend to be interpreted exhaustively, one might propose that disjuncts are interpreted exhaustively, too (Fox, 2007; Zimmermann, 2000). According to this proposal, a statement of the form "*A or B*" may be read, in effect, as "*A but not B or B but not A*". To illustrate, consider (1) once again:

(13) Joe supports Donald or Hillary.

According to the distinct disjuncts account, an utterance of this sentence may be interpreted, in effect, as "Joe supports Donald but not Hillary or Hillary but not Donald." This interpretation is exclusive, since the disjuncts cannot both be true.

If the distinct disjuncts account is on the right track, we should expect that the robustness of the exclusive reading increases with the robustness of the exhaustive read-

ing associated with an utterance of one of the disjuncts. In other words, the robustness of the exclusive reading of (13) should increase with the robustness of the inference from “Joe supports Donald” to “Joe supports Donald but not Hillary” and with the robustness of the inference from “Joe supports Hillary” to “Joe supports Hillary but not Donald”.

Design

The distinct disjuncts account predicts that the strength of an exclusive reading of “ A or B ” should be positively correlated with the strength of exhaustive readings that statements of individual disjuncts A and B would receive in the same context. The purpose of Exp. 3 was therefore to collect data on the strength of exhaustive readings of such single-disjunct statements in the background contexts used in Exp. 1. We would then like to investigate whether strengths of exhaustive readings are a reliable predictor of strength of exclusive readings across contexts.

Participants

Using the same selection criteria as before, 131 subjects were recruited via Amazon’s Mechanical Turk and paid US\$ 0.50 for participation.

Materials

Exp. 3 used the fifteen stories that entered the analysis of Exp. 2. For each story we consider the speaker’s utterances of single disjuncts. Concretely, where Exp. 2 had an utterance of a sentence:

Utterance of disjunction

Jill says to Danny: “Alex bought a racket or a pair of shoes.”

Exp. 3 had two single-disjunct utterances by the same speaker:

Utterance of disjunct 1

Jill says to Danny: “Alex bought a racket.”

Utterance of disjunct 2

Jill says to Danny: “Alex bought a pair of shoes.”

Each story also had corresponding statements that subjects had to rate:

Exh1

From what Alex’s girlfriend said we may conclude that Alex did not buy a pair of shoes as well.

Exh2

From what Alex’s girlfriend said we may conclude that Alex did not buy a racket as well.

Procedure

The procedure followed that of Exp. 2 very closely. After reading the slightly amended instructions and seeing examples for the use of the slider bar, each participant was presented with six randomly sampled stories. Subjects read the background story, followed by an utterance of either disjunct, randomly chosen. Subjects first rated a random control question and then rated the *Exh1* or *Exh2* statement, depending on which utterance was shown to them.

Results

Data from one subject was discarded because English was not the self-reported native language. Another four subjects were removed for bad performance on the control questions, using the same criterion as before.

Figure 6 shows the per-story means of the ratings of the *Exh*-statements plotted against the corresponding mean ratings of the *Prior*- and *Xor*-statements from Exp. 1.

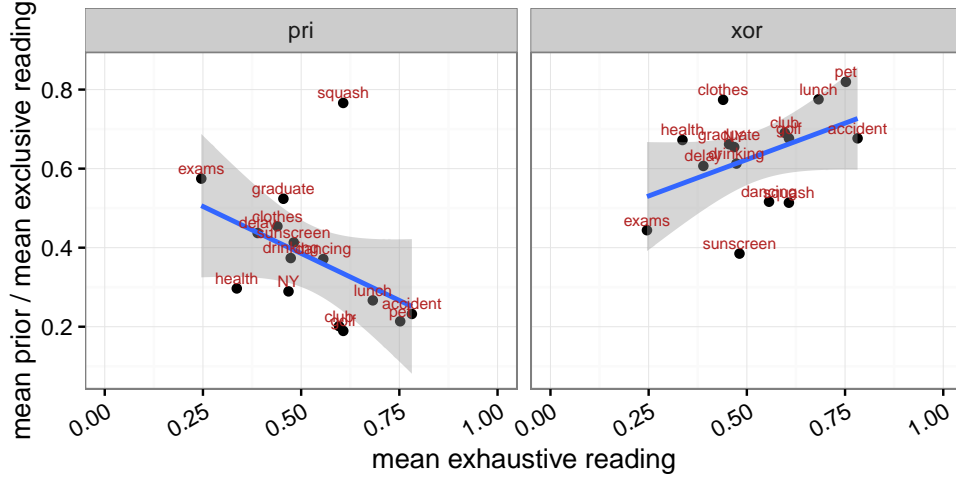


Figure 6: Means of ratings of *Exh*-statements from Exp. 3 (x-axis) vs. means of ratings of *Prior*-statements and *Xor*-statements from Exp. 2 (y-axis)

There is no significant correlation between *Prior*-ratings and *Exh*-ratings ($r \approx -0.45$, $p \approx 0.09$), suggesting that our measures of prior expectations and exhaustive strength do not coincide. Adding the per-story mean *Exh*-ratings as an additional explanatory factor EXH to the regression model comparison, we obtain the picture given in Figure 7. A model using single factor PRI to predict *Xor*-ratings is about 8.5 times more likely than a model using single factor EXH. That means that our data provides evidence for the assumption that prior expectations are a better explanatory factor of exclusive readings than the strength of exhaustive readings, thus discounting the distinct disjuncts account.

5 General discussion

Statements of the form “A or B” are sometimes interpreted as “A or B but not both”. The consensus in the recent literature is that these exclusive readings are scalar implicatures, which are inferences deriving from a pragmatic reasoning process about the speaker’s intentions. A crucial assumption in this reasoning process is that the speaker

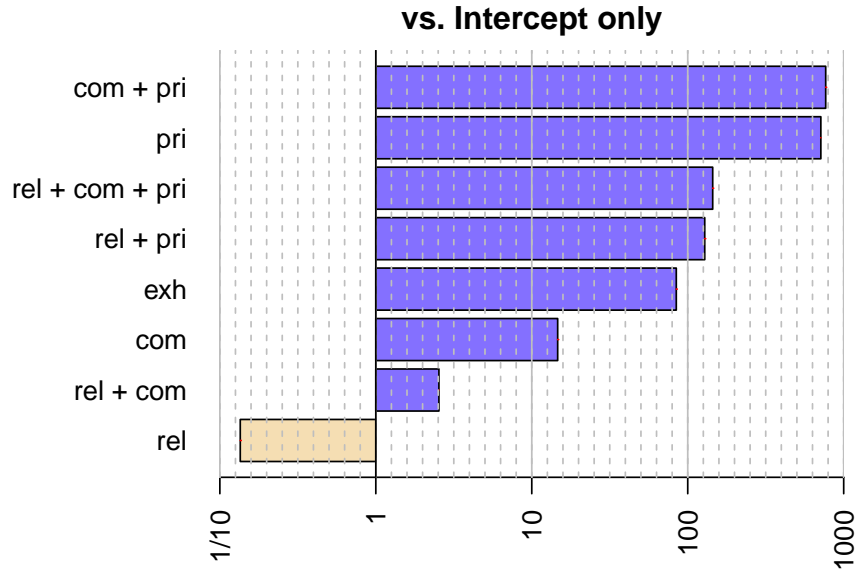


Figure 7: Bayes factor comparison of different main factor combinations, predicting the strength of exclusive disjunction readings with additional factor EXH from Exp. 3.

is taken to be knowledgeable about whether the corresponding statement with “*A* and *B*” is true. In this paper, we have presented data that challenge the implicature account. In contrast to the predictions made by this account, the robustness of the exclusive reading does not increase with the competence of the speaker about “*A* and *B*”. Instead, we observed the reverse effect.

We also tested an alternative proposal, according to which the exclusive reading is derived as a conversational implicature by exhaustifying the individual disjuncts. According to this proposal, a sentence of the form “*A* or *B*” is read, in effect, as “*A* but not *B* or *B* but not *A*”. This proposal predicts that the robustness of the exclusive reading depends on the robustness of the inference from *A* to “*A* but not *B*”. Although we did find a significant effect of this factor, it was a significantly worse predictor of the robustness of the exclusive reading than the ratings of prior probability of the “and”-alternative.

Exclusive readings and probabilistic world knowledge. Taken together, these findings suggest that the exclusive reading of “or” might not be a scalar implicature. Instead, we propose that exclusive readings could be the result of probabilistic inferences about the state of the world. In other words, the meaning of “or” is inclusive, but hearers may exclude the possibility that both disjuncts are true based on their knowledge of the world.

Such probabilistic inferences have been shown to influence various other parts of language. A case in point is quantifying expressions. In a series of experiments, Moxey and Sanford (1993) presented participants with statements such as:

- (14) a. Q people found Miss Sweden attractive.
b. Q earthquakes occurred in California in 1951.

Q was varied between ten quantifying expressions, including “a few” and “many”. Participants were instructed to estimate, e.g., the number of people who found Miss Sweden attractive and the number of earthquakes in California in 1951, based on the information expressed in the sentences. For most quantifying expressions, Moxey and Sanford observed that these estimates increased as a function of the prior probability. For example, the estimates would be higher for (14-a) than (14-b), since the prior probability that at most n people find Miss Sweden attractive is lower than the prior probability that at most n earthquakes occurred in California in 1951 (see also Pepper & Prytulak 1974).

In a similar vein, the interpretation of vague gradable adjectives, such as “big” and “small”, varies depending on our world knowledge. Kamp and Partee (1995) discuss the following minimal pair:

- (15) a. My two-year old son built a really tall snowman yesterday.
b. The D.U. fraternity brothers built a really tall snowman yesterday.

It is obvious that one’s prior expectations about the size of the snowman are lower in the case of (15-a) than in the case of (15-b), thus demonstrating that the interpretation of “tall” varies depending on one’s prior expectations. Recent probabilistic models of pragmatic language use have explored this systematic dependence on prior expectations (cf., Lassiter & Goodman, 2015; Qing & Franke, 2014; Schöller & Franke, 2015).

World knowledge and prior expectations thus influence the interpretation of various aspects of natural language. We propose that the interpretation of “or” is one such aspect as well: although, strictly speaking, a statement of the form “*A* or *B*” leaves open the possibility that both *A* and *B* are true, prior expectations can modulate the likelihood of this possibility. If our proposal is on the right track, this would have important theoretical repercussions. For instance, exclusive readings of sentences with “or” have been argued to occur in embedded positions as well, i.e., in the scope of other logical operators, as in (16), where standard accounts might not (easily) predict scalar implicatures to arise. If exclusive readings of disjunctions are no (ordinary) scalar implicatures, arguments based on their occurrence in embedded position provide little motivation for a radical conceptual rethinking of scalar implicatures *tout court* (e.g., Chierchia, 2004; Chierchia et al., 2012; Fox, 2007).

(16) If you order pizza or pasta before 7pm, you pay only \$6.50.

Competence. We compared “or” with “some”, which is often interpreted as “some but not all”. This upper-bounded construal is an uncontroversial example of a scalar implicature. Unlike the exclusive reading of “or”, the robustness of the upper-bounded construal of “some” increased with the competence of the speaker about the truth of the corresponding statement with “and”, in line with the results of Goodman and Stuhlmüller (2013).

It remains an open question for further research as to why higher ratings of speaker

competence seem to pair with lower prominence of exclusive readings. A possible line of explanation could be to dissect the relationship between world states and speaker knowledge, and experimental participants' intuitive model thereof. It is possible to imagine that world states where “*A* and *B*” is false are more likely than states where this is not so to give rise to epistemic states in which the speaker would use the statement with “or” and is ignorant about the “all”-alternative. This could be because of asymmetries between observability of positive and negative facts—it is usually easier to notice that Alex bought a racket than that he did not—or because it may require more than simple observation for an agent to form a belief that “*A* or *B*, but not both” or that “*A* iff *B*”. An integrated probabilistic model that incorporates reasoning about evidence and language use, similar to that of Goodman and Stuhlmüller (2013), which jointly infers a world state and an epistemic state based on a speaker's utterance, might shed light on the puzzling relation between competence and strength of exclusive readings that we observed. Since, presently, we can at best conceive of entirely ad hoc models of how world states and epistemic states could be related to each other, we must leave this issue for future research.

Prior probability. We observed a significant effect of prior probability: the robustness of the upper-bounded construal decreased with the prior probability that the statement with “all” was true. This effect ties in with recent experimental work from Degen et al. (2015), but speaks against observations made by Geurts (2010). Geurts discusses examples such as the following to show that the decision to derive a scalar implicature is made independently from the prior probability that the statement with “all” is true.

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

Here, according to Geurts' intuitions, the speaker communicates that some but not all of the marbles sank to the bottom, even though such a situation has an extremely low prior probability. It remains to be seen whether a probabilistic model along the lines of Degen et al. (2015) can account for the effects of priors and speaker competence that we observed here, as well as for exclusive readings of "or".

Relevance. Unlike competence and prior probability, relevance to the hearer's interest did not turn out to have any effect on the robustness of the upper-bounded construal of "some". This is problematic to relevance theory, which predicts the robustness of an upper-bounded construal to be an increasing function of its importance to the hearer. It may suggest that relevance theory should adopt a more restricted notion of relevance in terms of discourse purposes (Cummins & Rohde, 2015). Alternatively, we might hold that hearer-oriented relevance is only one specific case of discourse relevance. Accordingly, the absence of an effect of relevance manipulations in our experiment may be a result of indeterminacy as to which aspect of discourse relevance prevails. Still, if this were so, then our results would at least indicate that hearer-oriented relevance is not a strong and dominant default.

An alternative explanation for the absence of an effect of relevance, however, is that we did not manipulate how relevant the upper-bounded construal was to the participant. According to this explanation, we failed to observe a significant effect because there were no differences in how relevant the upper-bounded readings were to the participants. Even though this alternative explanation warrants closer examination, there are several reasons to be skeptical about it. Perhaps the most prominent reason is that, strictly speaking, participants did not have any reason to derive the scalar implicature in the first place. The fact that they did so nonetheless indicates, from a relevance-theoretic point of view, that they engaged with the story.

In summary, then, our results provide an insight in how various factors conspire in

shaping differences in the robustness of putative scalar implicatures. It will be interesting to determine which other factors—e.g., typicality (van Tiel, 2014), prosodic and linguistic prominence (Breheny, Katsos, & Williams, 2006), and politeness (Bonneton, Feeney, & Villejoubert, 2009)—have similar effects, and how these factors might interact. It may well be that these factors also hold the key to explaining the observation that different scalar expressions license scalar implicatures at substantially different rates (van Tiel, van Miltenburg, Zevakhina, & Geurts, 2016). Such questions, however, remain for the moment as topics for future research.

References

- Barrett, R. B., & Stenner, A. J. (1971). The myth of exclusive ‘or’. *Mind*, 80(317), 116–121.
- Basson, A. H., & O’Connor, D. J. (1960). *Introduction to symbolic logic*. Free Press of Glencoe.
- Baum, R. (1996). *Logic* (4th edition). Harcourt Brace.
- Bonneton, J.-F., Feeney, A., & Villejoubert, G. (2009). When some is actually all: scalar inferences in face-threatening contexts. *Cognition*, 112(2), 249–258.
- Breheny, R., Katsos, N., & Williams, J. (2006). Are generalized scalar implicatures generated by default? An online investigation into the role of context in generating pragmatic inferences. *Cognition*, 100(3), 434–463.
- Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon’s Mechanical Turk: a new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, 6(1), 3–5.

- Chevallier, C., Noveck, I. A., Nazir, T., Bott, L., Lanzetti, V., & Sperber, D. (2008). Making disjunctions exclusive. *The Quarterly Journal of Experimental Psychology*, 61(11), 1741–1760.
- Chierchia, G. (2004). Scalar implicatures, polarity phenomena and the syntax/pragmatics interface. In A. Belletti (Ed.) *Structures and beyond*, (pp. 39–103). Oxford: Oxford University Press.
- Chierchia, G., Fox, D., & Spector, B. (2012). The grammatical view of scalar implicatures and the relationship between semantics and pragmatics. In P. Portner, C. Maienborn, & K. von Steubner (Eds.) *An international handbook of natural language meaning*, (pp. 2297–2332). Berlin: Mouton de Gruyter.
- Copi, I. M., & Cohen, C. (2005). *Introduction to logic* (12th edition). Prentice Hall.
- Crain, S. (2008). The interpretation of disjunction in universal grammar. *Language and Speech*, 51(1-2), 151–169.
- Cummins, C., & Rohde, H. (2015). Evoking context with contrastive stress: effects on pragmatic enrichment. *Frontiers in Psychology*, 6.
- Degen, J., Tessler, M. H., & Goodman, N. D. (2015). Wonky worlds: listeners revise world knowledge when utterances are odd. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.) *Proceedings of the 37th annual conference of the Cognitive Science Society*, (pp. 548–553). Austin, TX: Cognitive Science Society.
- Fox, D. (2007). Free choice and the theory of scalar implicatures. In U. Sauerland, & P. Stateva (Eds.) *Presupposition and implicature in compositional semantics*, (pp. 71–120). Houndmills: Palgrave Macmillan.

- Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, 336, 998.
- Franke, M. (2009). *Signal to act: game theory in pragmatics*. Ph.D. thesis, University of Amsterdam.
- Franke, M., & Jäger, G. (2016). Probabilistic pragmatics, or why Bayes' rule is probably important for pragmatics. *Zeitschrift für Sprachwissenschaft*, 35(1), 3–44.
- Geurts, B. (2006). Exclusive disjunction without implicature.
- Geurts, B. (2010). *Quantity implicatures*. Cambridge: Cambridge University Press.
- Goodman, N. D., & Stuhlmüller, A. (2013). Knowledge and implicature: modeling language understanding as social cognition. *Topics in Cognitive Science*, 5(1), 173–184.
- Grice, H. P. (1975). Logic and conversation. In P. Cole, & J. L. Morgan (Eds.) *Syntax and semantics, volume 3: Speech acts*, (pp. 41–58). New York: Academic Press.
- Horn, L. R. (1972). *On the semantic properties of logical operators in English*. Ph.D. thesis, University of California, Los Angeles. Distributed by Indiana University Linguistics Club.
- Hurford, J. (1974). Exclusive or inclusive disjunction. *Foundations of Language*, 11, 409–411.
- Jeffreys, H. (1961). *Theory of Probability*. Oxford: Oxford University Press, 3rd ed.
- Kamp, H., & Partee, B. (1995). Prototype theory and compositionality. *Cognition*, 57(2), 129–191.
- Lassiter, D., & Goodman, N. D. (2015). Adjectival vagueness in a Bayesian model of interpretation. To appear in: *Synthese*.

- Levinson, S. C. (2000). *Presumptive meanings: the theory of generalized conversational implicature*. Cambridge, MA: MIT Press.
- Lodewyckx, T., Kim, W., Lee, M. D., Tuerlinckx, F., Kuppens, P., & Wagenmakers, E.-J. (2011). A tutorial on bayes factor estimation with the product space method. *Journal of Mathematical Psychology*, 55, 331–347.
- McCawley, J. D. (1981). *Everything that linguists have always wanted to know about logic but were ashamed to ask*. University of Chicago Press.
- Moxey, L. M., & Sanford, A. J. (1993). Prior expectation and the interpretation of natural language quantifiers. *European Journal of Cognitive Psychology*, 5(1), 73–91.
- Pepper, S., & Prytulak, L. S. (1974). Sometimes frequently means seldom: context effects in the interpretation of quantitative expressions. *Journal of Research in Personality*, 8(1), 95–101.
- Qing, C., & Franke, M. (2014). Gradable adjectives, vagueness, and optimal language use: A speaker-oriented model. In J. Grieser, T. Snider, S. D’Antonio, & M. Wiegand (Eds.) *Proceedings of SALT 44*, (pp. 23–41).
- Raftery, A. E. (1995). Bayesian model selection in social research. In P. V. Marsden (Ed.) *Sociological Methodology*, (pp. 111–196). Cambridge: Blackwell.
- Rescher, N. (1964). *Introduction to logic*. St. Martin’s Press.
- Roberts, C. (2012). Information structure in discourse: towards an integrated formal theory of pragmatics. *Semantics and Pragmatics*, 5(6), 1–69.
- Rouder, J. N., & Morey, R. D. (2012). Default Bayes factors for model selection in regression. *Multivariate Behavioral Research*, 47, 877–903.

- Rubin, R., & Young, C. M. (1989). *Formal logic: a model of English*. Mayfield.
- Russell, B. (2006). Against grammatical computation of scalar implicatures. *Journal of Semantics*, 23(361–382).
- Russell, B. (2012). *Probabilistic reasoning and the computation of scalar implicatures*. Ph.D. thesis, Brown University.
- Sauerland, U. (2004). Scalar implicatures in complex sentences. *Linguistics and Philosophy*, 27(3), 367–391.
- Schnoebelen, T., & Kuperman, V. (2010). Using Amazon Mechanical Turk for linguistic research. *Psihologija*, 43(4), 441–464.
- Schöller, A., & Franke, M. (2015). Semantic values as latent parameters: Surprising *few & many*. In S. D’Antonio, M. Moroney, & C. R. Little (Eds.) *Proceedings of SALT 45*, (pp. 143–162).
- Schulz, K., & van Rooij, R. (2006). Pragmatic meaning and non-monotonic reasoning: The case of exhaustive interpretation. *Linguistics and Philosophy*, 29, 205–250.
- Simons, M. (2001). Disjunctions and alternativeness. *Linguistics and Philosophy*, 24, 597–619.
- Sperber, D., & Wilson, D. (1995). *Relevance: communication and cognition* (2nd edition). Blackwell.
- Sprouse, J. (2011). A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavior Research Methods*, 43(1), 155–167.
- Storto, G., & Tanenhaus, M. K. (2005). Are scalar implicatures computed online? In E. Maier, C. Bary, & J. Huitink (Eds.) *Proceedings of Sinn und Bedeutung 9*, (pp. 431–445). Nijmegen: Nijmegen Centre for Semantics.

- van Kuppevelt, J. (1996). Inferring from topics: scalar implicatures as topic-dependent inferences. *Linguistics and Philosophy*, 19(4), 393–443.
- van Tiel, B. (2014). Embedded scalars and typicality. *Journal of Semantics*, 31(2), 147–177.
- van Tiel, B., van Miltenburg, E., Zevakhina, N., & Geurts, B. (2016). Scalar diversity. *Journal of Semantics*, 33(1), 137–175.
- Wilson, D., & Sperber, D. (2002). Relevance theory. *UCL Working Papers in Linguistics*, 14, 249–290.
- Yanal, R. J. (1988). *Basic logic*. Thomson.
- Zimmermann, T. E. (2000). Free choice disjunction and epistemic possibility. *Natural Language Semantics*, 8, 255–290.
- Zondervan, A. (2010). *Scalar implicatures or focus: an experimental approach*. Ph.D. thesis, Utrecht University.

A Stories from Experiment 1

1. **airco** (*Rel+* / *Com+* / *Pri+*)

Harry’s job is to inspect hotels of an international franchise in order to check whether everything is according to the high standards of the organization. He reports to his boss Mr Jaynes after a visit to the San Diego hotel. Harry reports to Mr Jaynes: “Some of the rooms had working air conditioning.”

2. **quiz** (*Rel+* / *Com+* / *Pri+*)

George and Amber are playing a quiz. George gets to ask Amber ten questions on

a topic of George's choice. George picked "Star Trek" as a topic but he does not know Amber is actually a big "Star Trek" fan. Amber just answered the final question. George says: "You got some of the answers right."

3. NBA (*Rel+* / *Com+* / *Pri-*)

Jason Barley and Richard Trellis are TV experts engaged in a live discussion of the current NBA season which is nearing its end. They are debating about whether Greg Jones should be the Most Valuable Player of the season. Jason Barley thinks so, but Richard Trellis is less convinced. Jason Barley says: "Greg Jones secured victory for his team during the last seconds of some of the decisive playoff matches."⁸

4. donkey (*Rel+* / *Com+* / *Pri-*)

Grace and Oliver are playing a game of pin the tail on the donkey. Oliver has hung up a large number of pictures of donkeys and Grace has to pin the tail on as many of them as possible while wearing a blindfold. Grace just pinned the last tail. Oliver says: "You put some of the tails in the right place."

5. watch (*Rel+* / *Com-* / *Pri+*)

Mr Tobler works for a Swiss factory which produces luxury watches. Today he ran several tests with a new prototype in order to make sure that it meets the very high requirements on product quality that the company aims for. At this late stage in the project nobody expects any problems. But quality control is very important for the company's image. Mr Tobler's boss Mr Papenhauer is curious about the results and tries to call Mr Tobler but he has just left. Therefore Mr Tobler's apprentice answers the phone. The apprentice has only attended the first half of the testing procedure. The apprentice says: "The prototype passed some of the tests scheduled for today."

6. Harvard (*Rel+* / *Com-* / *Pri+*)

Ethan is hoping to go to Harvard so he has to do well on his final exams. Fortunately he has had a month of spare time during which he studied day and night. His friend Eleanor just heard rumors from her fellow students. She says: “You passed some of the exams.”

7. tomatoes (*Rel+* / *Com-* / *Pri-*)

Ted is a chef at a three star restaurant. He asks his sous chef Gilbert to check today’s purchase of fresh produce. Gilbert looks at the first crate of tomatoes, and comes running back. Gilbert says: “Some of the tomatoes are rotten.”

8. schizo (*Rel+* / *Com-* / *Pri-*)

Lucy has to give a talk in front of a big audience of psychologists. She is going to criticize one of the dominant theories about schizophrenia. Afterwards, Jacob, who was in the audience, chatted with his neighbours. He tells Lucy: “Some of the people enjoyed your talk.”

9. gifts (*Rel-* / *Com+* / *Pri+*)

Leonard celebrated his 5th birthday with his best friends today. All of his friends brought him presents. The presents were given to him while everybody was sitting at the table after a nice piece of chocolate cake. When Suzanne’s mother comes to take her daughter home from the party, Leonard’s mother tells Suzanne’s mother about the party. Leonard’s mother says: “Leonard unwrapped some of the presents right away.”

10. M&M’s (*Rel-* / *Com+* / *Pri+*)

Henry is in the hospital because he broke his arm in a skiing accident. His aunt Bethany visited him yesterday and brought him some fruit and a bag of M&Ms. Later, Henry

tells his father: “I liked some of the M&Ms.”

11. vGogh (*Rel– / Com+ / Pri–*)

The mayor joins the opening of a new van Gogh exhibition in the city’s Museum of Modern Art. The curator welcomes him warmly and gives him a private tour through the exhibition. The curator says: “Some of van Gogh’s paintings are among the most valuable paintings in the world.”

12. drinking (*Rel– / Com+ / Pri–*)

Edward is in a crowded bar IDing the patrons who are drinking alcohol. There are new laws and if there is even one incident of underage drinking today Edward will close down the bar. Later he tells the owner of the bar: “Some of the patrons drinking alcohol were underage.”

13. dessert (*Rel– / Com– / Pri+*)

Jeremy is asking his friend Claire for her opinion about the newly opened gourmet restaurant on 5th Street. Claire has been there twice already and she is very enthusiastic. Claire says: “Some of their desserts are fantastic.”

14. cookies (*Rel– / Com– / Pri+*)

Martha received a couple of chocolate cookies even though neither she nor her husband eats chocolate. This evening the two of them are going out for dinner leaving Emily to babysit their children. Emily is a voracious eater and chocolate lover. When she comes home, Martha sees some crumbs on the couch. She says to her husband: “Emily ate some of the cookies.”

15. comedy (*Rel– / Com– / Pri–*)

Lennard is a very popular stand-up comedian. He is currently on tour. Both shows in Seattle were sold out. 500 tickets were sold for each show. He phones his wife after the second show in Seattle. Lennard says: “Some of the people who came yesterday also came to today”’s show.”

16. emails (*Rel-* / *Com-* / *Pri-*)

Kate has problems with her laptop but she needs to let Joshua know she will not be able to make it to their meeting today. Therefore she sends the same email numerous times from different email accounts. Later that day, Joshua phones Kate up and says: “I received some of your emails.”

B Stories from Experiment 2

1. health (*Rel+* / *Com+* / *Pri+*)

Mrs Gibbs is worried about her husband”’s health. Her friend Cindy, who is a waitress at a local restaurant, served Mrs Gibbs’ husband yesterday. Cindy says to Mrs Gibbs: “Yesterday your husband had a steak or a beer.”

2. squash (*Rel+* / *Com+* / *Pri+*)

Danny and Alex reserved a squash court but Alex still has to buy a racket and a pair of shoes. Danny is talking to Alex’s girlfriend Jill who just went to the sports store with him. Jill says to Danny: “Alex bought a racket or a pair of shoes.”

3. lunch (*Rel+* / *Com+* / *Pri-*)

During summer camp, every student was allowed at most one main dish for lunch. But some students queued up twice. Mrs Sanders was with her son Ted during the whole lunch break yesterday. A teacher approaches Mrs Sanders to find out how her

son behaved during lunch. Mrs Sanders says to the teacher: “Ted had pizza or pasta yesterday.”

4. accident (*Rel+ / Com+ / Pri-*)

Pete was playing on the schoolyard during lunch break when he fell off the slide. His teacher drove him to the hospital immediately and called Pete’s mother from there. The teacher told Pete’s mother: “Pete broke his arm or his leg.”

5. clothes (*Rel+ / Com- / Pri+*)

Brad is notorious for his shabby clothes. His friends Gina and Mandy make frequent jokes about it, much to Brad’s distress. Gina heard that he finally went shopping the other day. Gina tells Mandy: “Brad bought a sweater or a pair of jeans.”

6. exams (*Rel+ / Com- / Pri+*)

Recently, Gigi took exams in Physics and Chemistry. She studied day and night and even paid for private tutoring. Carrie just heard some rumors from her fellow students. Carrie tells Gigi’s best friend: “Gigi passed her Physics exam or her Chemistry exam.”

7. dancing (*Rel+ / Com- / Pri-*)

Many people came to Carl’s garden party yesterday, including his ex-wives Sue and Mary. It was terribly crowded everywhere. John helped out in the kitchen where it was much more quiet during most of the evening. Afterwards John told Carl’s wife: “Carl danced passionately with Sue or Mary.”

8. golf (*Rel+ / Com- / Pri-*)

Leo is sponsoring a golf tournament. In particular, he provides a luxurious car for any hole-in-one. When Leo arrives at the tournament, his friend Hans comes up to him to

tell about the events so far. Hans tells Leo: “I heard someone just hit a hole-in-one on the first or second hole.”

9. order (*Rel-* / *Com+* / *Pri+*)

Jake and Bill went jogging together in the afternoon. Afterwards they were very hungry and went to a restaurant. Jake later told Bill’s wife: “Bill ordered a starter or a side dish with his main course.”⁹

10. sunscreen (*Rel-* / *Com+* / *Pri+*)

Chloe is telling her father about her holidays with her two friends Susan and Amelia. Chloe forgot to bring suntan lotion. Chloe tells her father: “Susan or Amelia brought some extra suntan lotion I could use.”

11. club (*Rel-* / *Com+* / *Pri-*)

Tony loves to dance to electronic music, preferably Techno. The simpler and more monotonic, the better. He contemplates going to club Macabre tonight, but he has no idea what music they play. He phones his friend Rob who occasionally works at club Macabre as a bartender. Rob tells Tony: “They play Jazz or Metal tonight.”

12. drinking (*Rel-* / *Com+* / *Pri-*)

Lily is worried that her husband Eric is still drinking despite the doctor’s orders. She phones up Eric’s drinking buddy Bill who went to a bar with him last night. Bill tells Lily: “Your husband drank whisky or vodka.”

13. delay (*Rel-* / *Com-* / *Pri+*)

Kevin, Cory and Joshua are 7-year-olds from the same neighborhood. They are friends but argue a lot. Yesterday Joshua, who has very strict parents, was later than usual to

arrive at their hang-out. Kevin thinks that Joshua doesn't like to play with them anymore. Cory says to Kevin: "Joshua had to eat lunch or do his homework."

14. graduate (*Rel-* / *Com-* / *Pri+*)

Nico and Rick are talking about their mutual friend Gerald. Gerald has been studying for ages but Nico heard he recently graduated. Nico says to Rick: "His mother or his father attended the ceremony."

15. NY (*Rel-* / *Com-* / *Pri-*)

Jimmy skipped school to go on a day-trip to New York City last week. Having to be back in the afternoon, he barely had one hour to spend in the city. Jimmy's younger brother, who is very talkative even if he doesn't know what he is talking about, told the neighbors about the trip afterwards. Jimmy's younger brother told the neighbors: "Jimmy visited the Statue of Liberty or the Empire State Building."

16. pet (*Rel-* / *Com-* / *Pri-*)

Richard and Tom are talking about Harold. Harold has been wanting a pet for a long time but his girlfriend wouldn't let him. Richard heard she finally gave in. Richard says to Tom: "Harold bought a cat or a dog."

List of Figures

1	Example of a trial from Exp. 1	16
2	Ratings of statements according to intuitive pre-classification. Results for Exp. 1 (“some”) in the top row, results for Exp. 2 (“or”) in the bottom row. Error bars indicate bootstrapped 95% confidence intervals around the plotted means.	18
3	Relating ratings for relevance, competence and prior statements (<i>x</i> -axis) to ratings for the strength of pragmatic enrichments (<i>y</i> -axis). The top row shows ratings, averaged for each story, for Exp. 1 (“some”). The bottom row shows ratings, averaged for each story, for Exp. 2 (“or”).	19
4	Bayes factor comparison of different main factor combinations. Notation like “com + pri” stands for a regression model with main factors COM and PRI. The plots show the Bayes factor on the logarithmic <i>x</i> -axis of the models on the left, when compared to an intercept only regression model.	20
5	Density estimates of posterior over model parameter coefficients for a linear model with all three main factors.	22
6	Means of ratings of <i>Exh</i> -statements from Exp. 3 (<i>x</i> -axis) vs. means of ratings of <i>Prior</i> -statements and <i>Xor</i> -statements from Exp. 2 (<i>y</i> -axis) .	27
7	Bayes factor comparison of different main factor combinations, predicting the strength of exclusive disjunction readings with additional factor EXH from Exp. 3.	28

Notes

¹Word count (including appendices and references): ca. 12,500

²Here is a small detour into basic modal logic to better understand what the relationship is between ignorance about proposition “ A and B ”, on the one hand, and, on the other hand, ignorance about proposition A and ignorance about proposition B . Let us assume that for proposition P , the formula $\Box P$ represents that “the agent believes P ”, formula $\Diamond P$ represents that “the agent considers P possible” and formula $?P$ stands for “the agent is ignorant about P ”. As usual, we say that $\Box P$ is true in world w iff all worlds accessible from w make P true; $\Diamond P$ is true in w iff there is an accessible world from w that makes P true; $?P$ is true in w iff $\Diamond P$ and $\Diamond \bar{P}$ are, where \bar{P} is the negation of P . With this, it is not difficult to see that the two propositions of interest $?(A \wedge B)$ (“ignorance about proposition “ A and B ”) and $?A \wedge ?B$ (“ignorance about A and ignorance about B) are *logically independent* in the standard sense that any truth-value combination of both is logically possible. To show this, it suffices to give examples for worlds with the desired truth-value combinations. Individuating a world w with the set of accessible worlds, which in turn can be individuated with the truth-values these worlds assign to A and B respectively, we can write, for instance $\{AB, A\bar{B}, \bar{A}B, \bar{A}\bar{B}\}$ for a world w which accesses three types of worlds: one in which both A and B are true, one in which A is true and B is false, and one in which A is false and B is true. The following table shows four types of worlds that prove logical independence:

world	$?(A \wedge B)$	$?A \wedge ?B$
$\{AB, A\bar{B}, \bar{A}B, \bar{A}\bar{B}\}$	1	1
$\{AB, A\bar{B}\}$	1	0
$\{\bar{A}\bar{B}, \bar{A}B\}$	0	1
$\{AB\}$	0	0

³Mechanical Turk is a website where workers perform so-called “Human Intelligence Tasks” (HITs) for financial compensation. It has been shown that the quality of data gathered through Mechanical Turk equals that of laboratory data (e.g., Buhrmester, Kwang & Gosling 2011, Schnoebelen & Kuperman 2010; Sprouse 2011).

⁴In order to ascertain that nothing hinges on this decision, we conducted a follow-up experiment in which we paid 70 participants on Amazon’s Mechanical Turk 50 US\$ cent to judge the stories from Exp. 2 followed by a control statement and the following target statement:

(18) It is likely that Alex bought both a racket and a pair of shoes.

Each participant saw six randomly sampled stories, which were presented without the statement containing “or”, just as in Exp. 2.

We excluded one participant for not identifying as a native speaker of English and another one for bad performance on the control questions, using the same criterion as for Exp. 2. The mean ratings, taken for each story, for both types of statements which targeted prior probability were highly correlated ($r \approx 0.89, p < 0.0001$). For all of the analyses reported in the main text, nothing of substance changes if we include the ratings for (18) instead.

⁵The formal criterion for exclusion was having a *deviance score* greater than a fixed threshold, where the deviance score of a participant is the sum of the absolute differences between the expected answers for all control questions (0, 0.5, or 1) and the subjects’ answers. We set the threshold of exclusion to the mean plus twice the standard deviation of all deviance scores. This exclusion criterion was also used in all other experiments.

⁶All conclusions of theoretical relevance are also supported by more traditional, frequentist regression analyses in terms of significance of factors and model comparison by AICs.

⁷This general conclusion is also vindicated by more complex analyses that would take interactions and random effects for participants into account.

⁸This item was removed from the analysis. We observed that the average robustness rating for the upper-bounding construal of “some” for this item (39%) was substantially below the corresponding ratings for the other items (58% or more). In addition, almost half of the participants indicated that the probability of the upper-bounded reading was zero, while none did so for the next lowest scoring item. On closer examination, we noticed an ambiguity in the statement measuring the robustness of the upper-bounding construal:

(19) From what Jason Barley said we may conclude that Greg Jones did not secure victory for his team during the last seconds of all of the decisive playoff matches.

This statement may be read as implying that Greg Jones did not secure victory during the last seconds of *any* of the decisive playoff matches, which is manifestly false. In order to determine if participants indeed succumbed to this ambiguity, we tested an alternative formulation:

(20) From what Jason Barley said we may conclude that not all of the decisive playoff matches were secured during the last seconds by Greg Jones.

Importantly, this formulation is equivalent to the desired reading of (19) but lacks the ambiguity because

the negation directly modifies the noun rather than the verb phrase.

We presented 40 participants on Mechanical Turk with the NBA story, followed by three questions. The first two were control questions. The third one was the target question. Half of the participants saw (19) and the other half (20). The procedure was the same as in the original experiment, except that responses were made by marking a value on a seven-point scale rather than by means of a continuous slider. The results for the formulation in (19) (28%) were in line with the results of the original experiment (39%). However, the results for the formulation in (20) were significantly higher (66%, $t(29) = 4.28$, $p < .001$), thus indicating that participants did not arrive at the desired reading of (19). Hence, we decided to remove this item from the analysis.

⁹ We excluded this item, because of an unfortunate mistake in the formulation of the disjunctive utterance which included the phrase “with his main course”. As this phrase was mistakenly not included in the formulation of the target statements concerning relevance, competence and prior, it is not guaranteed that the information obtained for different target statements is coherent.