Symmetric alternatives and semantic uncertainty modulate scalar inference

Brandon Waldon and **Judith Degen**
{bwaldon, jdegen}@stanford.edu
Stanford University Department of Linguistics, 450 Jane Stanford Way
Stanford, CA 94305 USA

Abstract

Scalar inferences are commonly assumed to involve both literal semantic interpretation and social cognitive reasoning. However, the precise way to characterize listeners' representation of context - including the space of possible utterance alternatives as well as the space of possible conventional meanings associated with linguistic forms - is a matter of ongoing debate. We report a partial replication of a scalar inference priming study by Rees and Bott (2018), introducing a novel baseline condition against which to compare behavior across different priming treatments. We also investigate the effect of raising participants' awareness of communicatively stronger alternatives that explicitly encode an exhaustive meaning (e.g. *some but not all* with respect to *some*). Our results suggest that exhaustive alternatives (which are 'symmetric' to canonical alternatives) can modulate the availability and strength of scalar inferences, and that semantic uncertainty is an independent channel through which scalar inferences are modulated. We discuss implications for theories of pragmatic competence.

Keywords: experimental pragmatics; implicature; priming; adaptation; computational pragmatics

Introduction

Linguistic communication depends on a listener considering not only what a speaker has said (e.g. *The movie was good*) but also the utterance alternatives a speaker could have produced but chose not to (*The movie was excellent*). This process yields inferences beyond the literal meaning of the original utterance, including scalar inferences (e.g. *The movie was good, but not excellent*). Scalar inferences, rather than being deterministically derived, are modulated by features of the linguistic and extra-linguistic context (Horn, 1972; Grice, 1975; Levinson, 2000). One long-discussed feature is the role that the unsaid alternatives play (Hirschberg, 1985; Fox & Katzir, 2011; Degen & Tanenhaus, 2015). The traditional view holds that the weak scalar item competes with communicatively stronger alternatives retrieved from a lexicalized or contextually-parameterized scale (e.g., ⟨all, some⟩). This view additionally characterizes scalar inference in terms of negation of the alternatives, which has implications for how theoreticians characterize the alternative set: *all* canonically competes with *some*, but (on this view) the upper-bounded form *some but not all* cannot: if both were alternatives to *some*, then their joint negation would lead to contradiction - the alternatives negate one another in meaning (the so-called symmetry problem - see Katzir, 2007). Thus, on this view, the alternative set is assumed to be highly constrained.

This view has recently come under scrutiny. For example, it has been shown experimentally that number terms interfere with the computation of scalar implicatures from *some* to *not all* (Degen & Tanenhaus, 2015, 2016). Moreoever, recent probabilistic computational approaches to pragmatics (Goodman & Stuhlmüller, 2013; Franke & Jäger, 2016) discard the 'negation-of-alternatives' view of scalar inference and thus allow for the possibility that the set of alternatives is relatively large (and possibly even includes symmetric alternatives). However, the experimental pragmatic literature on scalar inference has largely adopted the traditional view's assumptions regarding the alternative set, and few studies have considered whether non-'canonical' alternatives additionally affect the strength and availability of these inferences.[1]

Another assumption of the traditional view is that speakers and listeners have no uncertainty regarding the conventions - the map from linguistic forms to meanings - of their shared language. For example, *some* is assumed to have a fixed *at least some* literal meaning, and it is via reasoning about communicatively stronger alternatives that this meaning is possibly enriched to an exhaustive (upper-bounded) *some but not all* meaning in context. A different view posits that interlocutors have a priori semantic uncertainty (that is, a priori uncertainty regarding these form-to-meaning mappings) and that this uncertainty is updated over the course of interaction.[2]

Rees and Bott (2018) help to address this issue in a series of priming experiments. The authors use a paradigm similar to that of Bott and Chemla (2016), in which one priming treatment involved exposure to utterances containing canonical alternative expressions (following the authors, 'alternative' priming), whereas another involved explicitly associating weak scalar items with scenes that the items exhaustively describe ('strong' priming). The authors found that interpretation of weak scalar expressions on target trials differed substantially after canonical alternative priming and strong priming relative to 'weak' priming (which involved explicitly associating weak scalar items with a non-exhaustive interpretation). However, there was no detectable difference in behavior on critical trials between alternative priming and strong priming conditions, leading the authors to conclude that "the rate of scalar implicature is determined entirely by the salience of the [strong] alternative" (2018: 14).

We extend Rees and Bott (2018)'s paradigm in order to a) address the question of symmetric alternatives and their

[1] Though see Nicolae and Sauerland (2015) and Bott and Chemla (2016) for exceptions.

[2] We stay uncommitted as to the proper way to spell out this view formally. One possibility - which we explore later on in the paper - is to posit that lexical items such as *some* directly map to either a *some but not all* meaning or to an *at least some* meaning. One could also posit that *some* is fixed to an *at least some* reading on its semantics but that listeners have a priori uncertainty as to whether *some* occurs in the scope of a semantic enrichment operator which triggers the exhaustive interpretation (e.g. the exhaustivity operator proposed by Chierchia et al., 2012).

©2020 The Author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

role in scalar inferences; as well as to b) address the question of whether or not scalar inferences are modulated by updating listeners' a priori semantic uncertainty. We find evidence that raising awareness of symmetric alternatives affects the strength and availability of scalar inferences, and that the extent to which the symmetric alternative competes with the weak scalar item is a function of the alternative's morphosyntactic complexity. Lastly, we find that weak and strong priming give rise to different patterns of behavior than do conditions designed only to raise awareness of alternative forms, which we take as evidence that semantic uncertainty is an independent channel for modulating scalar inference.

Experiment

Methods

Participants We recruited 480 participants online through Amazon Mechanical Turk (MTurk - US IP addresses; minimum 95% prior approval rating). Participants were paid $2.20, and average completion time was about 13 minutes.[3]

Materials and procedure Priming and target trials of the experiment are schematized in Table 1. Priming trials consisted of a sentence or a math task (see our example Baseline priming trial, Figure 1), plus a forced choice between two images. Target trials followed two priming trials and involved a forced choice between an image and the option to select "Better Picture?". The linking assumption for behavior on target trials was as follows: if participants interpreted the sentence as conveying an exhaustive meaning, they selected "Better Picture?"; otherwise, they selected the other image.

The experiment featured sentences containing constructions from one of three expression categories: *some*, cardinal numbers, or simple existential declaratives of the form *There is an X* (coded as 'ad-hoc' for the purposes of the study). For each expression category, there was a construction for which we assumed the availability of a lower-bounded, non-exhaustive interpretation: that is, *some* is compatible with the meaning *some and possibly all*; the cardinal numeral expression *four* is compatible with an *at least four* meaning; and *There is an X* is compatible with *There is an X and possibly more*. *Some of the symbols are X*, *There are four X*, and *There is an X* were the resulting construction templates that participants read on critical trials.

For each expression category, we identified an expression that was communicatively stronger than the weak expression: the canonical alternative expressions *all* (cf. *some*), *six* (cf. *four*), and *There is an X and a Y* (cf. *There is an X*). Each weak expression was furthermore associated with a semantically exhaustive meaning expressible in three different ways, which differed in length and hence in presumed production cost. The three construction types were obtained by adding *only*; by conjoining the weak expression with the negation of communicatively stronger alternative at the sub-clausal level; and by conjoining the weak expression with the negation of communicatively stronger alternative at the clausal level (see Table 1 for examples of constructions).

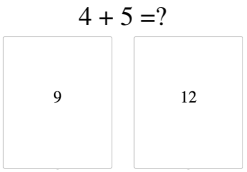


Figure 1: Example of a Baseline priming trial.

Weak priming trials featured a weak scalar expression (e.g., *some*) and a forced choice between a) an image compatible with the expression's non-exhaustive interpretation and b) an image that made the sentence false. Strong priming trials also featured the weak scalar expression but involved a forced choice between a) an image compatible with its non-exhaustive interpretation and b) an image compatible with its exhaustive interpretation. Alternative priming trials featured an alternative expression (either a *canonical* alternative or an *exhaustive* one, whose meaning included negation of the meaning of the canonical alternative and hence was a symmetric alternative) and a forced choice between an image that made the sentence true and one that made the sentence false. Prime type (Weak, Strong, Baseline, Alternative) was a within-subjects manipulation, while alternative type (*canonical*, *exhaustive only*, *exhaustive sub-clausal*, and *exhaustive clausal*) was a between-subjects manipulation.[4]

For each non-Baseline prime type, we assumed an optimal 'correct' choice which served as the basis of our exclusion criteria: on Weak priming trials, the correct choice was the image compatible with the expression's non-exhaustive interpretation; on Strong priming trials, it was the image compatible with the exhaustive interpretation; on Alternative trials, it was the image that made the sentence true.

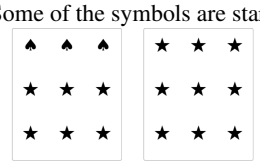
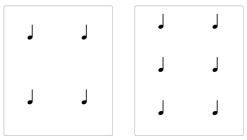
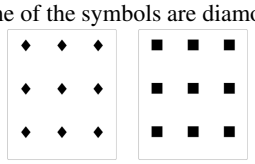
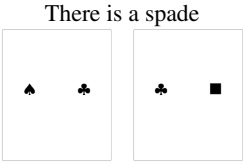
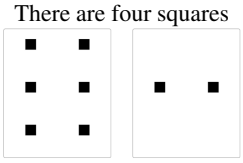
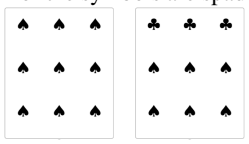
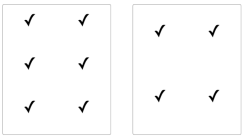
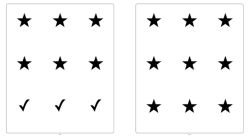
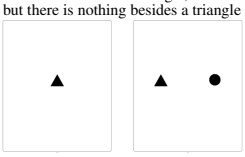
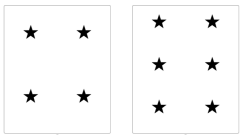
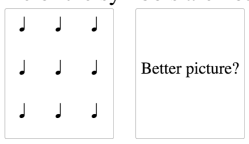
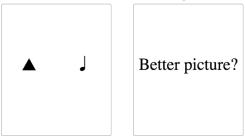
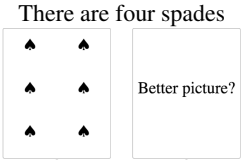
In each between-subjects condition, there were 12 types of priming blocks (3 expression categories * 4 prime types). Participants saw four blocks of each type, which were balanced for position of the correct image in the preceding two priming trials (left-left, right-right, left-right, or right-left), for a total of 48 experimental blocks. Participants also saw 48 filler blocks, which matched the structure of the experimental blocks except that the third trial was a choice between an image which made the sentence false and "Better Picture?".

Participants were told "You will read a number of English sentences. For each sentence, you will be asked to select one of two pictures. Please select the image that best reflects the meaning of the sentence. Most pictures contain either sym-

[3]Methods, exclusions, and analyses for the experiment were pre-registered through the Open Science Foundation, available at `https://osf.io/p4nj5`. Data and code are available at `https://github.com/bwaldon/symalts/`.

[4]The present study differs from that of R&B in that we introduce the Baseline prime type for all participants, as well as alternative type as a between-subjects manipulation (including novel *exhaustive* alternative forms). Our addition of the Baseline priming blocks (and a matching number of fillers) means that our participants saw 25% more trials than did R&B's participants. Anticipating noise from data collected online, we assigned 120 participants to each condition (whereas R&B recruited 100 participants for their in-person study).

Table 1: Example priming and target trials for each priming type (rows) and expression category (columns).

	some	ad-hoc	number
Prime: Strong	Some of the symbols are stars 	There is a club	There are four notes 
Prime: Weak	Some of the symbols are diamonds 	There is a spade 	There are four squares 
Prime: Alternative (*canonical*)	All of the symbols are spades 	There is a note and a triangle 	There are six ticks 
Prime: Alternative (*exhaustive* -*only*[i]; -*sub-clausal*[ii]; -*clausal*[iii])	[i]Only some of the symbols are stars [ii]Some but not all of the symbols are stars [iii]Some of the symbols are stars, but not all of them are stars 	[i]There is only a triangle [ii]There is a triangle but nothing else [iii]There is a triangle, but there is nothing besides a triangle 	[i]There are only four stars [ii]There are four but no more than four stars [iii]There are four stars, but there are no more than four stars 
Target	Some of the symbols are notes 	There is a triangle 	There are four spades 

bols or numbers. In some cases, one picture may contain the text 'Better Picture?'. This option should be selected if you don't feel the other picture sufficiently captures the sentence meaning." Participants were prevented from starting the experiment until they correctly selected images on two practice trials, on which they received feedback.[5]

Predictions

If participants' awareness of alternative forms (via direct exposure) is heightened on Alternative priming trials, and this heightened awareness triggers Gricean counterfactual pragmatic reasoning on target trials regarding the non-production of the alternative, then we predict greater rates of "Better Picture?" selection relative to Baseline in the case of *canonical* Alternative priming; and lower rates for *exhaustive* Alternative priming. We identify in particular a **symmetric alternative** hypothesis, which (contra the traditional view) is associated with the latter prediction regarding *exhaustive* Priming.

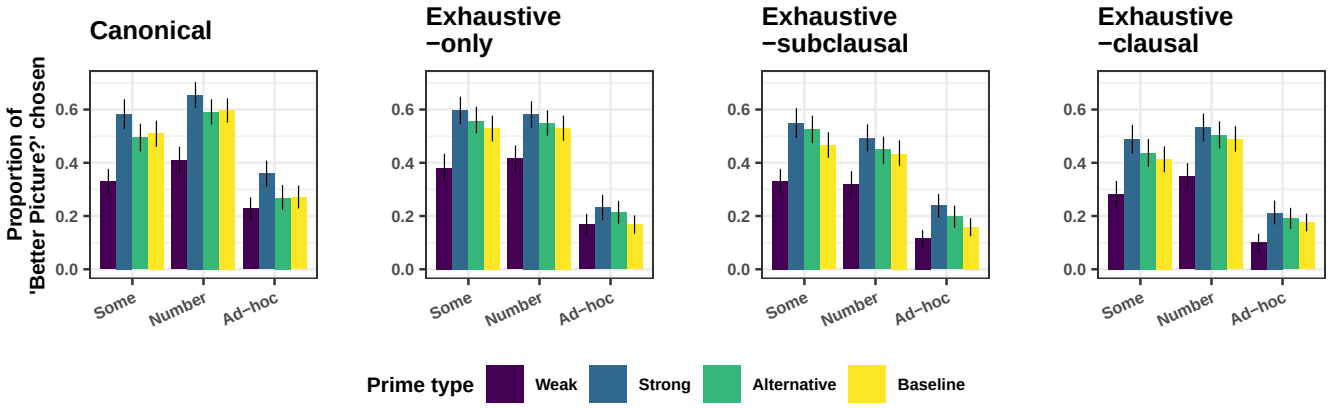
Effects of Strong and Weak priming beyond any observed effects of Alternative priming are consistent with predictions of the **semantic uncertainty** hypothesis, which posits that scalar inference is driven at least in part by a priori uncertainty about whether weak scalar items are exhaustive on their literal semantic meaning. The semantic uncertainty hypothesis accounts for these patterns in the data by positing that on target trials, participants consider not only of the various forms that can be used in the experiment but also of the conventions of use of those forms. The hypothesis is consistent with the possibility that exposure to *canonical* alternative forms may increase the rate of "Better Picture?" selection relative to Baseline, though the null hypothesis is that no differences exist between *canonical* Alternative priming and Strong priming, both of which we expect to raise rates of "Better Picture?" selection if they indeed modulate behavior.

Results

23 sets of responses were removed due to multiple participation. 6 participants were excluded for reporting a native lan-

[5]Practice trial 1 featured *Many of the symbols are X*, plus a choice between an image where 8/9 symbols were X and an image where 3/9 symbols were X; practice trial 2 read *There is an X above a Y* and was a choice between a false picture and "Better Picture?".

Figure 2: Proportion of "Better Picture" selection by expression category, prime type, and between-subjects condition (e.g. 'Canonical' = condition where the alternative type was *canonical*). Error bars indicate 95% bootstrapped confidence intervals.



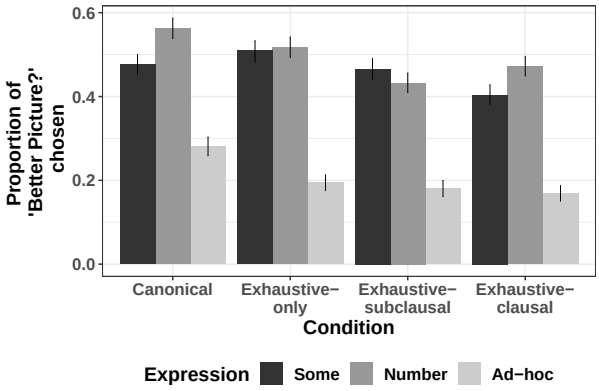
guage other than English. We excluded 9.5% of the remaining experimental blocks due to incorrect answers on priming trials. Proportions of "Better Picture?" selection on target trials are displayed for each between-subjects condition in Figure 2. For the data from each between-subjects condition separately, we conducted a Bayesian mixed effects logistic regression analysis predicting log odds of "Better Picture?" selection from fixed effects of expression, prime type, and their interaction; with random by-subject and by-item intercepts and slopes for expression, prime type, and their interaction - the maximal random effects structure justified by the design (*some* was the reference level for the expression factor; Baseline was the reference level for prime).[6]

Model results are shown in Table 2. In the *canonical* priming group, there was evidence of a negative main effect of Weak priming relative to Baseline (0 not included in the 95% credible interval of estimate) and conflicting evidence of a positive main effect of Strong priming ($P(\beta > 0) > 0.95$, 0 inside the 95% credible interval); however, there was no evidence of a main effect of Alternative priming. There was evidence of an effect of both Weak and Strong priming in the other three between-subjects conditions and evidence of Alternative priming in the *exhaustive-sub-clausal* group (where the effect was positive). There was no evidence of interactions between expression and prime type except in the *canonical* group, where there was evidence of a positive interaction between the *ad-hoc* expression category and weak priming (mean est. = 1.08, 95% CI = [0.18, 1.91]).

Figure 3 displays variation in behavior across the between-subjects conditions. There is a numeric downward trend in the rate of "Better Picture" selection across conditions: first between the condition featuring *canonical* Alternative priming and the one featuring *exhaustive-only* priming, and then a further drop when the morphosyntactic complexity of the exhaustive form increases. In post-hoc analyses, we subset

[6]Bayesian analyses were performed using `brms` in R (Bürkner, 2017), with default parameter priors and four 3000-iteration chains.

Figure 3: Proportion of "Better Picture?" selection, aggregated across between-subjects condition (x-axis labeled according to type of Alternative prime seen in that condition). Error bars indicate 95% bootstrapped confidence intervals.



these data by expression type and conducted Bayesian mixed effects logistic regressions predicting "Better Picture?" selection from a fixed effect of condition, with a random by-item intercept and random by-item slope for condition as well as a random by-participant intercept (the reference level for between-subjects condition was the condition with *canonical* alternative priming).There was a main effect of *exhaustive-sub-clausal* condition on the data subset to the number expression type (mean est. = -1.06, 95% CI = [-1.91, -0.25]). For the ad-hoc expression type, there were main effects of the *exhaustive-sub-clausal* and *exhaustive-clausal* conditions.[7]

In a second post-hoc analysis, we investigate the possibility that the similarities between Baseline behavior and behavior after exhaustive Alternative priming are due to individual variation.[8] Perhaps observation of overtly exhaustive forms

[7]Model coefficients, data subset to ad-hoc expression type:
Exhaustive-only: Mean est. = -0.92, 95% CI = [-1.87, 0.02]
Exhaustive-sub-clausal: Mean est. = -1.13, 95% CI = [-2.12, -0.10]
Exhaustive-clausal: Mean est. = -1.03, 95% CI = [-2.02, -0.09]

[8]We thank an anonymous reviewer for this suggestion.

Table 2: Mean estimates, 95% credible intervals, and posterior coefficient probabilities (probability that coefficient differs from 0 in predicted direction), for main effects of prime type (relative to Baseline), across conditions.

Alt. type	**Strong prime**	$P(\beta > 0)$	**Weak prime**	$P(\beta < 0)$	**Alt. prime**	
canonical	0.50 [-0.02, 1.01]	0.97	-1.83 [-2.37, -1.30]	1	-0.29 [-0.78, 0.19]	$P(\beta > 0) = 0.12$
exh-only	0.60 [0.08, 1.12]	0.99	-1.70 [-2.31, -1.10]	1	0.05 [-0.43, 0.53]	$P(\beta < 0) = 0.41$
exh-sub-clausal	0.47 [0.004, 0.94]	0.98	-1.44 [-2.02, -0.91]	1	0.49 [0.007, 0.98]	$P(\beta < 0) = 0.02$
exh-clausal	0.55 [0.07, 1.05]	0.99	-1.52 [-2.12, -0.95]	1	0.31 [-0.16, 0.78]	$P(\beta < 0) = 0.10$

primes some speakers to parse subsequent target sentences with a covert *only* or other exhaustivity operator,[9] while other speakers take the exhaustive sentences to be pragmatic alternatives to the target sentence. This account would explain why mean exhaustive Alternative priming behavior is similar to Baseline, on the assumption that roughly equal numbers of participants fall into either behavioral category (with perhaps slightly greater numbers of covert-exhaust primers).

To investigate, we examine behavior of participants for which we have 4 non-excluded observations of target trial behavior after both Baseline and exhaustive Alternative priming. We consider the change in the number of "Better Picture?" responses between these two priming conditions. A small proportion (13%) of participants exhibit lower rates of "Better Picture?" response after exhaustive Alternative priming relative to their Baseline behavior; this proportion is comparable to the proportion of participants who exhibit less "Better Picture" selection after Strong priming relative to Baseline across all four between-subjects conditions (12%, a result we attribute to experimental noise). 21% of participants exhibit more "Better Picture?" response after exhaustive Alternative priming, and this proportion is greater than the proportion of participants who exhibit more "Better Picture" selection after Weak priming (6%, also attributable to experimental noise). Thus, we have preliminary evidence that the small positive effect of exhaustive Alternative priming overall is driven by a small number of participants, rather than by sizable amounts of priming in either direction.[10]

Discussion

The condition of the experiment featuring *canonical* Alternative priming can be considered a partial replication of R&B's Experiment 1. Recall that R&B posit that Alternative and Strong priming do not differ in extent to which they can modulate scalar inference. The results from the *canonical* Alternative condition of our experiment are in slight tension with those results: we find at least weak evidence of modulation from Strong priming but not from Alternative priming.

This result generalizes to our novel experimental conditions in which we replace *canonical* Alternative priming with *exhaustive* Alternative priming. In those conditions, one might expect that calling attention to relatively 'marked' pragmatic alternatives can more effectively modulate subsequent scalar inference behavior than can raising the salience of putatively unmarked alternative forms, if the latter are already a priori likely pragmatic alternatives for the participant. For example, immediately after exposure to two sentences of the form *Only some of the symbols are X*, a participant might be relatively more likely to find the non-production of *only* meaningful on the target trial - and hence interpret *some* non-exhaustively on that trial. In fact, the data trend in the opposite direction: there is a numeric increase in "Better Picture?" selection after Alternative priming relative to Baseline in exhaustive priming conditions - and we find evidence for an effect in the case of the *exhaustive-sub-clausal* priming group. However, in all three exhaustive priming groups, we find evidence of Strong priming in the predicted direction. We take this as evidence that (non-deterministic) expectations regarding the form-to-meaning mapping of weak scalar expressions play an independent role in modulating scalar inference.

However, there is evidence that raising awareness of alternatives - including symmetric alternatives - also modulated scalar inference. Relative to participants who saw *canonical* alternatives, participants exposed to *exhaustive-subclausal* forms were less likely to provide a response indicative of scalar inference on number trials. This pattern was observed on ad-hoc expression trials for two groups exposed to exhaustive alternative forms. We would predict this if the exhaustive form is active as an alternative whose non-production indicates an intention to convey a non-exhaustive meaning.

A formal model of behavior in this paradigm, then, should allow for pragmatic inferences to be modulated via two distinct channels: one in which participants' expectations about the form-to-meaning mapping in the lexicon is updated, as in the case of Strong and Weak priming in the experiment; and one that allows for update of participants' expectations regarding the linguistic forms they might encounter in context (accounting for the changes in behavior across conditions).

The Lexical Uncertainty Rational Speech Act model of pragmatic competence proposed by Bergen et al. (2016) is one such formalism. On this model, pragmatic speakers are modeled as a probability distribution S_1 over utterances $u \in U$ given intended meanings $m \in M$. S_1's production probability of utterance u given intended meaning m and possible lexicon $\mathcal{L}$ is a function of utterance cost $C(u)$ and of the probability that a literal listener L_0 with prior meaning expectations P

[9] Such an analysis could help to account for the within-block effect of exhaustive priming in the *exhaustive-sub-clausal* condition.

[10] More details are available in the GitHub repository. See: `bwaldon.github.io/symalts/results/switching.html`

would conclude m given u and lexicon $\mathcal{L}$. Pragmatic interpretation of an utterance u is modeled as a probability distribution L_1 over meanings given observation of that utterance. The likelihood with which L_1 associates meaning m with utterance u is a function of the expectation that S_1 produces u to signal m and prior meaning expectations P – marginalized over a prior probability distribution over possible lexica P_Λ.

$$L_0(m|u,\mathcal{L}) \propto \mathcal{L}(u,m)P(m)$$
$$S_1(u|m,\mathcal{L}) \propto e^{log(L_0(m|u,\mathcal{L}))-C(u)}$$
$$L_1(m|u) \propto \sum_{\mathcal{L}\in\Lambda} S_1(u|m,\mathcal{L})P(m)P_\Lambda(\mathcal{L})$$

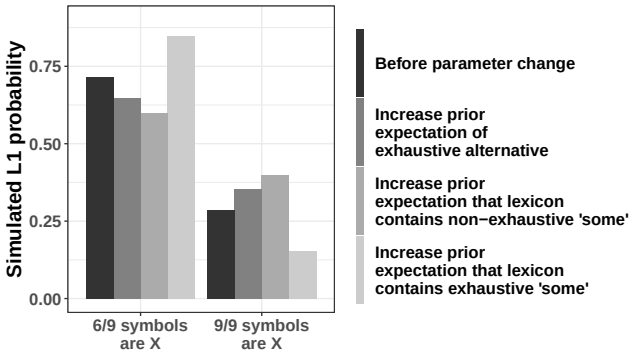
Lexical uncertainty models have been previously applied to analyze interpretation of *some*, including to analyze ignorance implicatures (by Bergen et al., 2016) as well as embedded implicatures (by Potts et al., 2016). Following those authors, we posit that listeners' lexical uncertainty includes uncertainty as to whether *some* should map to a non-exhaustive *some and possibly all* meaning or to an exhaustive *some and not all* meaning. On Weak and Strong priming trials of our experiment, participants map *some* to one of these two possible meanings, leading to a corresponding update of the lexicon distribution P_Λ. However, there is no uncertainty regarding the semantic representation of *some*'s canonical and exhaustive alternatives, including *all* and *some but not all*, which are unambiguous. Observation of these utterances thus would not update P_Λ but might update C, the listener's representation of the speaker's utterance cost function.

Qualitative patterns of expected behavior are schematized in Figure 4 for the case of interpretation of *Some of the symbols are X*.[11] On trials featuring constructions from this expression category, participants saw cards in which 9/9 symbols were of type X, cards in which only 6/9 were, or (on filler trials) cards in which 0/9 were. We assume that a participant has a uniform prior over these three potential meanings. In the 'default' case, (in the figure: 'Before parameter change'), we additionally assume a uniform prior over expected utterances *some*, *all*, *some but not all*, and *none* and a uniform prior over two lexica - one ($\mathcal{L}_J$) in which *some* maps to a literal exhaustive meaning and one ($\mathcal{L}_K$) in which it maps to a lower-bounded one (we assume both lexica contain a normal semantics for the other items). By decreasing C(*some but not all*) - that is, by increasing the listener's prior expectation of the pragmatic speaker producing the exhaustive alternative - we can cause a relative decrease in the strength of the listener's belief that *Some of the symbols are X* conveys an exhaustive, 6/9 symbol meaning. A similar pattern of results is achieved by increasing the listener's prior expectation of $\mathcal{L}_K$ over $\mathcal{L}_J$. The pattern is reversed when we increase prior expectation of $\mathcal{L}_J$ over $\mathcal{L}_K$: in this case, *Some of the symbols are X* is more strongly associated with the exhaustive meaning.[12]

[11] We present this schematic for illustrative purposes only. In our study, we do not find evidence that the presence of symmetric alternatives decreases exhaustive interpretation of *some* (though we find evidence for this in the case of the other two expressions).

[12] This analysis cannot, on its own, account for why we observed more modulation after Weak priming than after Strong priming.

Figure 4: Simulated L_1 probabilities given observation of *Some of the symbols are X*, with parameter changes to the lexical uncertainty RSA model as indicated in the paper.



General discussion

Our findings suggest the need to relax a traditional assumption regarding the set of alternatives active in scalar inference. Moreover, we argue that our results are best understood as listeners incrementally updating their uncertainty about speaker production behavior and their uncertainty about the underlying semantic representation of scalar expressions.[13] That is, we argue that the observed results reflect *adaptive* linguistic processes rather than simple bottom-up priming.

Of course, the context of our experiment leaves underspecified *who* the participant is adapting to (cf. Yildirim et al., 2016 and Schuster & Degen, in press, whose tasks involve introducing participants to speakers with more elaborate identities and motives). An adaptation analysis of our data would be better motivated if we could replicate our findings in a more naturalistic setting. We leave this to future work.

However, the differences in behavior across our four conditions suggest that one cannot analyze the priming blocks in isolation, and that a more nuanced story of incremental adaptation is warranted to understand the full pattern of results. We find evidence of participants updating their expectations about alternative utterance forms when considering participants' behavior in aggregate across the entire experiment. At the same time, replicating the results of both Bott and Chemla (2016) and Rees and Bott (2018), we find evidence of more 'local' within-block Strong and Weak priming.

Finally, our results constitute just one of a growing number of challenges to the traditional understanding of scalar inference as a categorical phenomenon driven by a small set of linguistic alternatives and a fixed semantic lexicon. Even for the case of scalar inference - one of the most well-studied phenomena in the pragmatics literature - patterns of data such as ours underscore the need for a departure from the traditional view and towards a richer, more gradient understanding of the relationship between language, context, and inference - as offered by contemporary probabilistic pragmatic frameworks.

[13] For a similar argument regarding modulation of interpretation of *might* and *probably*, see Schuster and Degen (in press).

Acknowledgements

We gratefully acknowledge Leyla Kursat and Benjamin Sparkes for their assistance in implementing the experiment. We also wish to thank Cleo Condoravdi, Daniel Lassiter, Christopher Potts, our three anonymous Cog Sci reviewers, the interActive Language Processing Lab at Stanford (ALPS), and the audience at CAMP3 for feedback and discussion. This work was supported by a National Science Foundation Graduate Research Fellowship (# 2019289423, to BW).

References

Bergen, L., Levy, R., & Goodman, N. (2016). Pragmatic reasoning through semantic inference. *Semantics and Pragmatics*, *9*(20).

Bott, L., & Chemla, E. (2016). Shared and distinct mechanisms in deriving linguistic enrichment. *Journal of Memory and Language*, *91*, 117-140. doi: 10.1016/j.jml.2016.04.004

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. doi: 10.18637/jss.v080.i01

Chierchia, G., Fox, D., & Spector, B. (2012). Scalar implicature as a grammatical phenomenon. In K. von Heusinger, C. Maienborn, & P. Portner (Eds.), *Semantics: An international handbook of natural language meaning* (pp. 3–2297). De Gruyter Mouton.

Degen, J., & Tanenhaus, M. K. (2015). Processing scalar implicature: A constraint-based approach. *Cognitive Science*, *39*(4), 667-710. doi: 10.1111/cogs.12171

Degen, J., & Tanenhaus, M. K. (2016). Availability of alternatives and the processing of scalar implicatures: A visual world eye-tracking study. *Cognitive science*, *40*(1).

Fox, D., & Katzir, R. (2011, Mar 01). On the characterization of alternatives. *Natural Language Semantics*, *19*(1), 87–107. doi: 10.1007/s11050-010-9065-3

Franke, M., & Jäger, G. (2016). Probabilistic pragmatics, or why bayes' rule is probably important for pragmatics. *Zeitschrift für sprachwissenschaft*, *35*(1), 3–44.

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: Vol. 3: Speech acts* (p. 41-58). New York: Academic Press.

Hirschberg, J. L. B. (1985). *A theory of scalar implicature* (Unpublished doctoral dissertation). University of Pennsylvania.

Horn, L. (1972). *On the semantic properties of logical operators in english* (Unpublished doctoral dissertation). UCLA.

Katzir, R. (2007). Structurally-defined alternatives. *Linguistics and Philosophy*, *30*(6), 669–690. doi: 10.1007/s10988-008-9029-y

Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. MIT press.

Nicolae, A., & Sauerland, U. (2015). A contest of strength: or versus either–or. In *Proceedings of sinn und bedeutung 20.*

Potts, C., Lassiter, D., Levy, R., & Frank, M. C. (2016). Embedded implicatures as pragmatic inferences under compositional lexical uncertainty. *Journal of Semantics*, *33*(4).

Rees, A., & Bott, L. (2018). The role of alternative salience in the derivation of scalar implicatures. *Cognition*, *176*, 1–14.

Schuster, S., & Degen, J. (in press). I know what you're probably going to say: Listener adaptation to variable use of uncertainty expressions. *Cognition*.

Yildirim, I., Degen, J., Tanenhaus, M. K., & Jaeger, T. F. (2016). Talker-specificity and adaptation in quantifier interpretation. *Journal of memory and language*, *87*, 128–143.