Alternatives, Priming, and Adaptation in Implicature

Brandon Waldon

Draft for Review - 12/4/2018

This paper investigates the nature of the alternative utterances that are active in a listener's pragmatic enrichment of a given utterance. Focusing on the case of scalar implicature, I first review traditional attempts to skirt the so-called *symmetry problem* by imposing constraints on the set of computed utterance alternatives. On the traditional view of scalar implicature as the categorical negation of communicatively stronger alternatives, symmetry is broken so long as constraints on alternative computation prevent the derivation of symmmetric alternatives. This view is contrasted with recent models of pragmatic competence which allow for a second avenue for symmetry breaking: compute symmetric alternatives, but prioritize some alternatives over others in pragmatic enrichment. I then investigate how the latter type of model thus allows for novel qualitative predictions in implicature priming experimental paradigms, and in the case of Rational Speech Act models of scalar implicature (c.f. Goodman and Stuhlmüller 2013) allows for the specification of explicit linking hypotheses of implicature priming understood as the behavioral reflex of listeners incrementally adapting their expectations of speakers in an unfolding discourse. The remaining sections report experiments designed to investigate the nature of this adaptation.

Alternatives and Scalar Implicature

The notion of an *utterance alternative* is central to pragmaticists' understanding of the type of inference known as *scalar implicature*, as the following theory-neutral definition illustrates:

(1) An utterance U conveys a scalar conversational implicature iff there are alternative utterances U' that are at least as relevant as U in the discourse and that are communicatively stronger[1] than U. (The content of this implicature will depend on the context, the nature of the utterance competition, and other pragmatic factors.) (Potts 2015: 179)

On theories which follow the framework of Grice (1975) in representing conversational implicature as a confluence of linguistic competence and social cognition, a scalar implicature is conveyed when a hearer engages in counterfactual reasoning about possible stronger alternative utterances a speaker could have said but did not. This counterfactual reasoning

[1]Note that this definition leaves underspecified the exact type of communicative 'strength' that we care about in the case of scalar implicature. As discussed below, many theorists have operationalized strength as logical entailment, but there are competing proposals (cf. Hirschberg 1985).

depends on the hearer's belief that the speaker is behaving cooperatively - as well as on the hearer's belief that the speaker is aware that the hearer believes the speaker to be cooperative. The following Gricean-inspired analysis of scalar implicature is adapted from Potts (2015):

(2)
- Let q = the proposition denoted by *John ate some of the cookies*
- Assume a speaker Ellen asserts *John ate some of the cookies* to a listener Kyle

i. *Contextual premise:* Ellen and Kyle know that John is able to eat any portion of some contextually-salient batch of cookies.
ii. *Contextual premise:* Ellen is maximally-informed with respect to how many cookies John ate.
iii. Assume that Ellen is cooperative and obeys the Gricean maxims of quality and quantity.[2]
iv. Then she will assert what is maximally relevant, informative and true.
v. An alternative proposition p (*John ate all of the cookies*) is more informative and (at least) as relevant in this context than is q.
vi. Therefore, Ellen must be experiencing a clash between the maxims: she cannot assert p because she lacks sufficient evidence for p.
vii. By (ii), she must lack evidence because it is false.

Implicature: it is not the case that John ate all of the cookies.

The aforementioned analysis of scalar implicature contrasts with more recent grammaticist analyses, whereby scalar implicature is associated with an inference which arises from a semantic operator whose input is a proposition p and whose output is a proposition computed from p and a set of alternatives to p. The following analysis of scalar implicature is adapted from one proponent of the grammaticist view, Chierchia (2006). The analysis must first define a function $||\cdot||^{ALT}$ which can map propositions to sets of propositions[3]; for example,

(3) $||$John ate some of the cookies$||^{ALT}$ = {*John ate some of the cookies*, *John ate all of the cookies*}

The aforementioned semantic operator takes as its input a proposition and returns a proposition which is a conjunction of the input proposition and the statement that the input proposition is the strongest true alternative to itself:

(4) $\mathrm{O}_C(\mathrm{p}) = p \wedge \forall q \in C[q \rightarrow p \subseteq_c q]$, where C = the output of $||\mathrm{p}||^{ALT}$

Given how we specify the semantic operator O_C in (4) and how we specify the alternatives to *John ate some of the cookies* in (3), the analysis predicts that *John ate some of the cookies*

[2] Grice's maxim of quantity: "Make your contribution as informative as is required. Do not say more than is required." The maxim of quality: "Contribute only what you know to be true. Do not say false things. Do not say things for which you lack evidence."

[3] One account offered by Chierchia (2006), $||\cdot||^{ALT}$ is the function that "associates any item [not necessarily a proposition] with its scalar alternatives" (Chierchia 2006: 546). The ability to specify alternatives for sub-sentential items - and thus to recursively define propositional alternatives - is necessary to capture embedded scalar inferences within Chierchia (2006)'s framework. But crucially, Chierchia leaves underspecified how exactly the alternatives of any item are computed.

under the scope of O_C should give rise to the inference that it is not the case that John ate all of the cookies:

(5) O_C(*John ate some of the cookies*)
$= \Big[$ *John ate some of the cookies* $\wedge$
$\forall p : p \in ||\text{John ate some of the cookies}||^{ALT} \wedge \text{p} \rightarrow$
John ate some of the cookies $\subseteq$ p $\Big]$
Pragmatically-enriched meaning: *John ate some of the cookies* $\wedge \neg$(*John ate all of the cookies*)

One commonality between Gricean and grammaticist approaches is that neither, on its own, is a fully-specified theory of how to compute the alternatives of a given utterance. While both theories claim that a scalar implicature will be conveyed by an utterance in case there are pragmatically-active stronger alternatives to that utterance, this alone is not sufficient to avoid what Fox and Katzir (2011), following von Fintel (2002), call the *symmetry problem*. The symmetry problem is illustrated as follows (discussion adapted from Fox and Katzir 2011):

(6)
- Let q = the proposition denoted by *John ate some of the cookies.* Let p_1 = the proposition denoted by *John ate all of the cookies*, which is stronger with respect to logical entailment than q.
- Let p_2 = the proposition denoted by *John ate <u>some but not all</u> of the cookies*, which is also stronger with respect to logical entailment than q.
- q $\rightarrow p_1 \vee p_2$.
- p_1 and p_2 contradict one another.
- q seems to implicate $\neg p_1$, but this is only possible if p_2 is not activated as an alternative to q.

By virtue of the fact that p_1 and p_2 contradict one another, we call p_1 and p_2 *symmetric alternatives* of q. The symmetry problem is twofold. First, we want a way to favor p_1 as an alternative to q over p_2: $q \wedge \neg p_1$ corresponds to the desired 'some but not all' strengthening of q in the context of utterance provided in, e.g., (2), while $q \wedge \neg p_2$ expresses a proposition that is truth-conditionally equivalent to *John ate all of the cookies* (clearly not the desired strengthening of q in the context of 2)[4]. Second, uttering q cannot implicate both $\neg p_1 \wedge \neg p_2$ because this conjunction is a contradiction. The avoidance of these two issues - conveying illicit scalar implicatures and conveying contradictory meanings - is a main desideratum of any theory of alternatives. That is, a successful theory of alternatives **breaks** symmetry in a manner which delivers the right predictions regarding the nature of the scalar implicature(s) conveyed by a given utterance.

[4]Note furthermore that there is nothing inherent to scalar implicature as we have defined it in (1) that seems to favor p_1 as an alternative over p_2; for example, p_1 being more relevant than p_2 in the context of utterance. von Fintel (2002) [http://web.mit.edu/24.954/www/files/24.954.lecturenotes.pdf] remarks that in a context where *How many of the boys are at the party?* is under discussion, *all of the boys are at the party* seems to be just as relevant as *some but not all of the boys are at the party.*

By and large the most popular and traditional class of theory has sought to break symmetry by placing constraints on the computation of certain alternatives in pragmatic interpretation. On theories of scalar implicature such as those discussed above - whereby strengthened meanings are derived via the categorical negation of communicatively stronger utterance alternatives - this is the most straightforward (and perhaps the exclusive) option for symmetry breaking. In fact, the symmetry problem is a problem just in case one's theory exhibits this categorical property shared by both mainstream Gricean and grammaticist theories. I first review theories of alternatives proposed in the backdrop of this widely-held assumption about the nature of the scalar implicature inference.

Breaking symmetry via constraints on alternatives

Horn (1972) proposes an account of scalar implicature whereby the calculation of alternatives is driven by conventionalized scalar relationships between lexical items. According to Horn, these scales are always ordered according to the relative logical strength of the lexical items on the scale. Utterance alternatives are then computed by taking the original utterance and replacing an lexical item from that utterance with one of that lexical item's stronger scalemates. This provides a straightforward way of understanding why *John ate all of the cookies* may be an alternative to *John ate some of the cookies*: an item in the original utterance, *some*, is replaced with another item, *all*, which stands in an asymmetric logical entailment relationship to *some*. That is, Horn proposes that <some, all> is a conventionalized lexical scale from which alternatives may be computed.

Horn proposes several other lexical scales which are similar to the <some, all> scale in that they represent an ordering of logical entailment. These scales include the scale of cardinal numerals, <or, and>, <might, must>, <pretty, beautiful>, and <warm, hot>. Importantly, <some, some but not all> is not taken by Horn to be a scale, even though the two items stand in an asymmetric logical entailment relationship to one another. The same can be said of potentially problematic scalar orderings such as <or, either-or> or <N, exactly N> (for any cardinal number N), all of which would lead to a similar symmetry problem as outlined above. Symmetry is thus broken on the grounds that only select alternatives are activated in pragmatic inference - and the alternatives which we happen to activate lead to empirically satisfying results (*some* implicates not *all*, *or* implicates not *and* and does not implicate not *either-or*, etc).

Though this may at first seem to have the air of stipulation, Horn provides some support for the idea that while, for example, <some, all> has conventionalized as a lexical scale, <some, some but not all> has not. Evidence comes from focus or contrastive discourse environments of the form in (7), where a scalar inference of the form $\neg Y$ is canceled, suspended, or reinforced in the context of asserting X (discourse frames adapted from Horn and Abbott 2012 as cited in Collins 2017):

(7) i. not only X but Y
  ii. X and for all I know Y
  iii. X if not Y

iv. X or even Y
v. X, indeed/in fact Y
vi. not even X, let alone Y
vii. Y, or at least X

For any two items X and Y, their acceptability in environments such as those in (7) is evidence that those two items are, to borrow the language of Horn and Abbott (2012), "natural paradigmatic alternatives" (334). The acceptability of (8) and (9) supports this analysis for *some* with respect to *many* and *all*, while (10) points against a similar analysis for *some* and *some but not all*:

(8) i. not only some but all
ii. some and for all I know all
iii. some if not all
iv. some or even all
v. some, indeed/in fact all
vi. not even some, let alone all
vii. all, or at least some

(9) i. not only some but many
ii. some and for all I know many
iii. some if not many
iv. some or even many
v. some, indeed/in fact many
vi. not even some, let alone many
vii. many, or at least some

(10)# i. not only some but some but not all
ii. some and for all I know some but not all
iii. some if not some but not all
iv. some or even some but not all
v. some, indeed/in fact some but not all
vi. not even some, let alone some but not all
vii. some but not all, or at least some

Application of this test with *or/either-or* as well as *N/exactly-N* similarly results in infelicity, as in (10).

One major revision to this account comes from Gazdar (1979), who argues that the scales from which alternatives are computed should rather be understood as conventionalized scales of *semantic representations* rather than lexical scales. One of Gazdar's key observations was that by positing scales at the level of the lexicon rather than at a deeper level of representation, the analysis put forward by Horn (1972) fails to account for the *nondetachability* of scalar inferences: independent of the individual lexical choices one makes in forming an utterance, we expect regularity in the implicatures that semantically-equivalent utterances will give rise to. For example, in (11), both response (a) and response (b) may give rise to the implicature that for all A knows, *from both courses* is not certain (example from Hirschberg 1985: 70, her example 73):

(11) A: I think you would have to get it from the instructor for the course...
B: For which course?
a. A: Possibly from both courses.
b. A: Maybe from both courses.

If scales are lexical, as Horn (1972) takes them to be, then we miss the generalization that non upper-bounded expressions of possibility such as *possibly* and *maybe* are associated with inferences of non-maximal certainty. Rather, we are put in a position of having to stipulate that *possibly* and *maybe* each reside on a lexical scale with *certain.* Gazdar's solution remedies this problem, but opens a new one: at *which level* of semantic representation are the scales derived? Gazdar is hesitant to claim that scales are derived at the level of semantic interpretation, on the grounds that conversational implicatures in general may vary, even between truth-conditionally equivalent sentences. For example, a sentence involving the verb phrase connective *and* may give rise to different temporal conversational implicatures depending on the order of the conjuncts. Thus, for Gazdar, an intermediate notion of semantic representation - somewhere between semantic interpretation and lexical representation (where, for example, information such as order of verb phrase conjuncts is preserved) - is the level at which scales should be derived.

However, Hirschberg (1985) observes that in a number of cases (including upper-bounding inferences with *some*, *or*, and cardinal numbers), truth-conditional semantic interpretation is an acceptable level at which to compute utterance alternatives. Moreover, temporal conversational implicatures involving *and* are qualitatively different from scalar implicatures in that the latter, but not the former, presumably relies on a clash in the Gricean maxims of quantity and quality in order for the inference to come about: the temporal *and*-implicature is presumably a manner implicature, and not a quantity one. While Hirschberg concedes that "a general representation of conversational implicature must accommodate conversational implicatures that rely upon the Maxim of Manner for their interpretation" (70), she remarks that "Gazdar's specific claim about scalar quantity implicature" - in particular the need to compute scalar and non-scalar alternatives at the same level of semantic representation - "is unclear" (ibid).

Hirschberg's second observation with respect to the analysis put forth by Gazdar (1979) is that he, like Horn, must continue to operate on the assumption that scales are somehow "simply given" (70). That is, there remains a lack of a fully descriptively-adequate analysis of where scales come from and of how to characterize their constraints. Though logical entailment appears to be an operable notion in scalar inference, Hirschberg observes that this not a sufficient - or even necessary - condition for deriving scales. Hirschberg offers a number of semantic orderings which support scalar inferences and notes that "[w]hile many might be defined by some notion of entailment, a number cannot" (84).

For example, orderings such as <dating, married, engaged>, <sick, dying>, <condone, promote> (e.g. violence), and <misdemeanor, felony> all appear to be pragmatically-active scales, which is supported by the fact that they licit in the discourse environments from (7).[5] However, there are no logical entailment relations **between** any of the items in these orderings. Rather, Hirschberg argues that a more general and more context-sensitive notion of communicative strength is needed.

Horn (1989), cited in Matsumoto (1995), presents an analysis specifically designed to constrain scales such that orderings such as <some, some but not all>, <or, either-or> and <N, exactly-N> are ruled out for the purposes of pragmatic inference. The constraint proposed by Horn builds off an observation that linguistic expressions may have one of three monotonicity properties: they may be upward monotone, as in the case of the generalized quantifier *some men*; downward monotone, as in the case of *no men*; or nonmonotone, as in the case of *exactly three men*. Horn (1989) introduces a constraint on alternatives whereby alternatives may not be calculated from scales where the expressions are nonuniform in their monotonicity behavior; moreover, nonmonotone expressions may never participate in scales.

Matsumoto caches out Horn (1989)'s observation as a more generalized conversational condition on pragmatic competition, as articulated below (from Matsumoto 1995: 25):

(12) Conversational Condition: The choice of W instead of S must not be attributed to the observance of any information-selecting Maxim of Conversation other than the Quality Maxims and the Quantity-1 Maxim (i.e., the Maxims of Quantity-2, Relation, and Obscurity Avoidance, etc.).[6]

As illustration of this principle in action, Matsumoto provides examples such as the following (taken from Matsumoto 1995: 45-46):

(13) a. “Three men came."
b. ‘(The speaker believes) it is not the case that exactly three men came.’
c. ‘(The speaker believes) exactly three men came.’
d. “Exactly three men came."

(14) a. “Bill met John on the way home."
(where the speaker knows that meeting Mary is of concern to the hearer)

[5]Hirschberg, for her part, takes issue with Horn's diagnostics for scales which support scalar inference. As she notes, Horn's discourse environments “will identify scales such as *only vote/ did vote* from [(1-a)] and *bald/exists* from [(1-b)]", which are clearly not pragmatically-active orderings:

(1) a. Only Muriel voted for Humphrey if even she did.
b. The king of France is bald if there is a king of France.

Of course, if we continue to assume that scales must be orderings of logical entailment, then the data in (1) do not pose a problem for Horn.

[6]Matsumoto makes a distinction between two quantity maxims: a Quantity-1 maxim, whereby speakers are directed to “make your contribution as informative as is required", and a Quantity-2 maxim: “do not make your contribution more than is required in the context of the exchange". That is, the Quantity-1 and Quantity-2 maxim taken together simply amount to Grice's original quantity maxim.

b. ‘(The speaker believes) is not the case that Bill met John but not Mary on the way home.’

c. ‘(The speaker believes) Bill met only John on the way home.’

d. “Bill met John but not Mary (or any other person) on the way home."

Matusmoto remarks that while the (d) sentences asymmetrically entail the (a) sentences above and are thus stronger than those sentences, the (b) sentences are not available implicatures. Rather, it seems that only the (c) sentences are available inferences. Rather than stipulate a monotonicity constraint on scales in the vein of Horn (1989), Matsumoto offers the following explanation: in asserting (13-a), the only utterances which may act as alternatives are those which themselves would license quantity implicatures had they been asserted (e.g. *four men came*, *five men came*, etc.). (13-d) is not such a candidate, as *exactly N* does not license scalar implicatures. A similar pattern of reasoning holds for (14): *Bill met Mary* may license a quantity implicature, unlike *Bill met John but not Mary*, hence why the former, but not the latter, may be activated as an alternative to the original utterance.

One commonality shared among *some* with respect to *some but not all*, *or* with respect to *either-or*, and *N* with respect to *exactly N* is that the latter is morphosyntactically more complex than the former. It may thus be tempting to rule out these pairings as pragmatically-active contrasts simply on the basis that alternatives may not be structurally more complex than the original utterance. However, Matsumoto (1995) presents data which problematizes this view, as in the following example (taken from Matsumoto 1995: 44):

(15) a. It was warm yesterday, and it is a little bit more than warm today.

b. It was a little bit more than warm yesterday, and it is a little bit more than warm today.

In the below example, the (b) sentence is intuitively more complex than the (a) sentence; but nonetheless, (b) appears to be a salient alternative to (a), in that (a) appears to give rise to the inference that it is not the case that it was a little bit more than warm yesterday. Matsumoto takes this as evidence that alternatives cannot be ruled out due to relative structural complextity.

Katzir (2007) responds in turn that a theory of alternatives computed from scales - even when those scales are constrained following Matsumoto (1995) or Horn (1989) - appears to yield inadequate empirical coverage. In support of this claim, Katzir first offers data such as the following as evidence that a scale-based account undergenerates (examples from Katzir 2007: 677):

(16) a. I doubt that exactly three semanticists will sit in the audience.

b. I doubt that three semanticists will sit in the audience.

(17) a. If we meet John but not Mary it will be strange.

b. If we meet John it will be strange.

(18) a. Everyone who loves John but not Mary is an idiot.

b. Everyone who loves John is an idiot.

According to Katzir (2007), in (16)-(18), the assertion of the (a) sentence seems to give rise to the inference that the (b) sentence is unassertable (if not necessarily untrue). However, the scales such as <three, exactly three> and <John, John but not Mary> are ruled out on either Matsumoto (1995)'s or Horn (1989)'s analyses. Moreover, Katzir offers data such as the following to suggest that a scale-based account sometimes overpredicts the prevalance of scalar inferences; for example, the (a) sentences below do not implicate the negation of the respective (b) sentence (examples from Katzir 2007: 684):

(19) a. A man came to every party.
b. A tall man came to every party.

(20) a. Each reporter talked to a candidate.
b. Each reporter talked to a candidate who sang.

(21) a. John is sure that many dogs will be sold.
b. John is sure that many dogs with long tails will be sold.

In (19) through (21), we have examples of potential orderings - including <a man, a tall man>, <a candidate, a candidate who sang>, <many dogs, many dogs with long tails> - which are neither blocked by the uniformity of monotonicity principle proposed by Horn (1989) nor by the Conversational Condition proposed by Matsumoto (1995). However, as Chris Potts (p.c.) points out, it does not necessarily follow that the analyses put forward by these authors predict illicit scalar inferences in these cases, provided that we assume listeners reason about only those alternatives that are at least as relevant in context as the original utterance (and it is easy to imagine a context where, for example, interlocutors are interested in communicating whether and how many men came to a party, with no regard for the tallness of those men). In fact, on Katzir's own account, alternatives to a given utterance should only be those that are *weakly assertable* in context, where weak assertability is defined partly in terms of discourse relevance. So it is not clear whether (19) through (21) motivate a new analysis of alternatives.

Katzir (2007), for his part, takes the data in (16) through (21) to be a challenge for existing scale-based accounts of alternative computation. Instead, he proposes the following principles from which alternatives are calculated:

(22)
- SUBSTITUTION SOURCE: Let ϕ be a parse tree. The substitution source for ϕ, written as L(ϕ), is the union of the lexicon of the language with the set of all subtrees of ϕ.
- STRUCTURAL COMPLEXITY: Let ϕ, ψ be parse trees. If we can transform ϕ into ψ by a finite series of deletions, contractions, and replacements of constituents in ϕ with constituents of the same category taken from L(ϕ), we will write $\psi \lesssim \phi$ If $\psi \lesssim \phi$ and $\phi \lesssim \psi$ we will write $\psi \sim \phi$ If $\psi \lesssim \phi$ but not $\phi \lesssim \psi$ will write $\psi < \phi$.
- STRUCTURAL ALTERNATIVES: Let ϕ be a parse tree. The set of structural alternatives for ϕ, written as $A_{str}(\phi)$,, is defined as $A_{str}(\phi) := \{\phi' : \phi' \lesssim \phi\}$

Katzir's analysis thus allows for the alternatives of an utterance to be those structures of equal or lesser structural complexity than the original utterance. Katzir's definition of a *SUBSTITUTION SOURCE* provides that alternatives derived from substituting material in the original utterance with other material from that same utterance is a neutral operation with regards to structural complexity. This circumvents the original problem posed by Matsumoto (1995) in example (15): because *a little bit more than warm* is a subtree of (15-a), it is in the substitution source with respect to that sentence. This allows for the construction of (15-b) as an alternative, where *a little bit more than warm* simply replaces *warm* in the first conjunct of the sentence.[7]

The commonality in the above theories of alternatives is the overarching assumption that we must block the computation of symmetric alternatives.

Breaking symmetry with probabilistic pragmatics

In the accounts discussed above, symmetry is broken by restricting the set of alternatives against which an utterance is compared. In this section, I review a probabilistic formalism of pragmatic competence - the Rational Speech Act model - and discuss how it affords a wider range of options for symmetry breaking.

As it is presented by Goodman and Stuhlmüller (2013) in those authors' analysis of scalar implicature, the Rational Speech Act model is an iterative, probabilistic model of pragmatic competence in which utterance interpretation is modeled as a distribution of possible meanings given an observation of that utterance. Just as with more traditional Gricean accounts, scalar implicature in the RSA framework is, at its core, the product of counterfactual reasoning about alternative utterances that a speaker might produce (but does not, on the assumption that the speaker is a cooperative interlocutor). However, the RSA framework explicitly models cooperative interlocutors as agents whose language production and comprehension is a function of Bayesian probabilistic inference regarding other interlocutors' expected behavior in a discourse context.

Specifically, in the RSA framework the probability with which L_1 concludes meaning s from an utterance u depends upon a prior probability distribution of potential states of the world P_w, and upon reasoning about the communicative behavior of a speaker S_1. S_1 in turn is modeled as a continuous probabilistic distribution over possible utterances given an intended meaning the speaker intends to communicate. This distribution is sensitive to a rationality parameter α, the production cost C of potential utterances, and the informativeness of the utterance, quantified via a representation of a literal listener L_0 whose interpretation of an utterance is in turn a function of that utterance's truth conditional semantics $[[u]](s)$ and her prior expectations about the state of the world $P_w(s)$.

(23) $P_{L_1}(s|u) \propto P_{S_1}(u|s) * P_w(s)$
$P_{S_1}(u|s) \propto exp(\alpha(log(L_0(s|u)) - C(u)))$

[7] See Fox and Katzir (2011) for an attempt to extend this theory to one of alternative computation in focus constructions.

$$P_{L_0}(s|u) \propto [[u]](s) * P_w(s)$$

One consequence is that in Goodman and Stuhlmüller (2013)'s RSA analysis of scalar implicature, individuals never categorically draw (or fail to draw) scalar inferences. For example, upper-bounded readings of *some* are represented in RSA as relatively lower posterior conditional probability of an *all* meaning on the P_L distribution given an utterance of *some*, compared to the prior probability of that meaning.

An example: exclusive *or*

The RSA model of pragmatic competence models exclusivity inferences with *or* as a negative change from the prior to posterior probability of an inclusive *and* meaning given an observation of an utterance of *or*. The model first declares a possible set of possible world states. For simplicity, I assume that the set of possible worlds contains only a world in which some statement X holds, a world in which some statement Y holds, and a third world in which both statements are jointly true. We assume that listeners have a uniform prior expectation of being in any of these three world states:[8]

```
var states = [['X'],['Y'],['X','Y']]

var statePrior = function() {
  return uniformDraw(states);
};
```

Next, we add a space of possible utterances that a speaker might produce. In this example, speakers are only capable of producing utterances of the form X *and* Y or X *or* Y; that is, we have implicitly constrained the model such that the only possible alternative to *or* is *and*. Furthermore, these two utterances are assumed to have a uniform cost, such that the prior expectation of hearing either of these utterances is uniform:

```
var utterances = ["or", "and"];

var cost = {
  "or": 1,
  "and": 1,
};

var utterancePrior = function() {
 var uttProbs = map(function(u) {return Math.exp(-cost[u]) }, utterances);
 return categorical(uttProbs, utterances);
};
```

A truth conditional semantics of the two utterances is then specified. *Or* assumes the meaning

[8] The code presented here is excerpted from a model I've coded in WebPPL, which is accessible here: http://bwaldon.github.io/ad_qp/models/bareor.js. The model is executable in the online WebPPL editor: http://webppl.org/.

of logical disjunction (i.e. it is logically consistent with conjunction), while *and* assumes the meaning of conjunction:

```
var literalMeanings = {
    and: function(state) { return state.includes("X") &&
    state.includes("Y"); },
    or: function(state) { return state.includes("X") ||
    state.includes("Y"); },
};
```

Lastly, models of a pragmatically-competent listener and speaker, as well as of a literal listener are specified as according to the equations in (23). For the pragmatic speaker, we set the rationality term α to 1 and keep it constant for the following discussion. On these assumptions, a pragmatically competent listener's expectation of being in the inclusive ['X','Y'] world state shifts from 33.33% to roughly 11.11% upon observation of an utterance of *or*. This negative change from prior to posterior expectation of the inclusive world state is how RSA captures the pragmatic exclusivity inference of *or*.

Note that in the above example, we have stipulated that there is only one alternative to *or* (namely *and*), but this need not be the case in order to break symmetry. Consider a second model, in which listeners also reason about a speaker's potential use of the exclusive disjunctive construction *either-or*.[9]The model differs minimally from the first model, where no *either-or* is present. The major changes are listeners have equal expectation of seeing any of three utterances (*or*, *and*, and *either-or*), and we must also specify the semantics of *either-or* (the rest of the model remains the same):

```
var utterances = ["or", "eitheror", "and"];

var cost = {
  "or": 1,
  "eitheror": 1,
  "and": 1,
};

var literalMeanings = {
    and: function(state) { return state.includes("X") &&
    state.includes("Y"); },
    or: function(state) { return state.includes("X") ||
    state.includes("Y"); },
    eitheror: function(state) { return (state.includes("X") ||
    state.includes("Y")) &&
    !(state.includes("X") && state.includes("Y"));
};
},
```

On these assumptions, a pragmatically competent listener's expectation of being in the

[9]This second WebPPL model is accessible here: http://bwaldon.github.io/ad_qp/models/eitheror.js

inclusive ['X','Y'] world state shifts from 33.33% to roughly 23.8% upon observation of an utterance of *or*. This negative change from prior to posterior expectation of the inclusive world state is smaller than the change observed in the previous model; the important point, however is that symmetry is broken even if *either-or* is allowed to enter into pragmatic competition with *or*. The change in prior to posterior expectation can be made greater by changing the cost term on utterances, such that the morphosyntactically more complex *either-or* is costlier than *or*.[10]

This paper is not the first to acknowledge RSA's potential to break symmetry yet allow for the computation of symmetric alternatives: as Bergen et al. (2016) observe, RSA "places weaker requirements on the set of alternative utterances. [In the case of pragmatic interpretation of *some*], the model can include both *all* and *some but not all* as alternatives, and still derive the correct implicatures" (16). As a formal and quantitative model of pragmatic competence, RSA provides a means of exploring the consequences of relaxing constraints on the computation of alternatives. As Bergen et al. (2016) also observe, we can increase the predicted strength of exhaustive inference in these models by introducing a cost asymmetry between upper-bounding strong alternatives and the more morphosyntactically complex alternatives which themselves encode exhaustive meanings.

This method of breaking symmetry - compute symmetric alternatives, but prioritize some alternatives over others in pragmatic enrichment - marks a substantial departure from the categorical theories of scalar implicature described above, which require a restrictive theory of alternative computation in order to skirt the symmetry problem. Uncovering empirical domains which adjudicate between these two approaches to alternative computation - and thus between categorical and gradient theories of scalar implicature - has been the subject of recent work in the experimental and computational pragmatics literature: in particular, probabilistic lexical uncertainty models of pragmatic competence (Bergen et al. 2016, Potts et al. (2016)) - which model listeners as agents with a priori uncertainty as to the literal semantic content of weak scalar terms - have been shown to uniquely predict empirically-attested interpretations of weak scalar items in embedded contexts. In what follows, I explore whether priming behavior with regards to scalar implicature can serve as an another adjudicating empirical domain.

Priming and scalar implicature

Whether and how scalar implicatures can be 'primed' is an established area of experimental pragmatics research. The literature on scalar implicature priming takes as its point of departure the observation that upper-bounded inferences of weak scalar items are not uniformly available across all populations of speakers - or across all lexical scales and discourse contexts (Papafragou and Musolino 2003, van Tiel et al. (2016), Doran et al. (2012)). The contextual conditions under which upper-bounded interpretations can (or cannot) be activated has informed debates regarding the mechanisms underlying pragmatic competence. In an early

[10] The reader my verify this by experimenting with cost terms in the attached code (see the previous footnote).

example from the acquisition literature, Papafragou and Musolino (2003) report that while native Greek-speaking children tend to accept underinformative uses of weak scalar items on the *<all, some>*, *<three, two>*, and *<finish, start>* scales more frequently than adults, rate of rejection increases when children are supplied a context which makes the strong items on these scales more discourse-salient. This finding is corroborated more recently by Skordos and Papafragou (2016): in one experiment reported by the authors, children are more likely to reject underinformative *some* when first asked to assess sentences containing the strong scalemate *all*; rejection rates further increase in a subsequent experiment when the incorrectness of sentences of the form *all of the X are Y* stems from the quantifier being too strong to describe a visual scene (only **some**, not all, of the X are Y) - rather than the predicate being incorrect (all of the X are **Z**, not Y). The authors argue that acceptance of underinformative *some* is thus linked to children's "failure to recognize that the [strong] scalar terms constitute relative alternatives" (2016: 8).

Modulating the discourse-salience of strong alternatives has also been a central theme in the more recent adult literature on scalar implicature priming. Bott and Chemla (2016) find evidence that adults' rejection of underinformative *some*, cardinal numbers, and ad-hoc existential constructions (*There is an X...*) can be modulated by training adults to associate these lower-bounded expressions with visual representations for which the expression is an underinformative descriptor - e.g. *there are six clubs* given a scene where there are in fact nine clubs - as well as visual representations for which the expression exhaustively describes the scene - e.g. *some of the symbols are clubs* given a scene where some but not all of the symbols are clubs. In addition to priming implicature behavior within a given scale, the authors find some evidence that priming lower/upper-bounded interpretations of a weak scalar item influences interpretations of items in other scales. Building on this result, Rees and Bott (2018) ask whether scalar implicature priming occurs exclusively via activation of stronger alternatives or whether the activation of an independent implicature-calculating mechanism contributes to priming. To investigate this, the authors use a paradigm similar to that of Bott and Chemla (2016) in which some participants are exposed to utterances containing strong alternative expressions, whereas others are trained to associate weak scalar items with scenes which the items exhaustively describe. Rees and Bott (2018) find no difference in behavior on critical trials between these two groups of participants, leading the authors to conclude that "the rate of scalar implicature is determined entirely by the salience of the [strong] alternative" (2018: 14).

Perhaps because of the commonly-held assumption that an utterance's alternatives constitute a restricted, semantically-consistent set rather than a broader range of potentially symmetric alternatives, most investigations into scalar implicature priming examine the effect of contextually activating one alternative in particular - the strong, upper bound-encoding alternative (e.g *all* with respect to *some*). However, there are important exceptions to this generalization. Degen and Tanenhaus (2015), for example, find evidence that raising the discourse salience of cardinal number expressions can cause numerals to enter into pragmatic competition with *some* and affect subsequent naturalness ratings and reading times for *some*. The authors take these findings as evidence that the alternative set for utternaces containing *some* must at least be capable of enrichment, such that the set contains *all*-sentences as well as sentences which replace the *some* for a cardinal number expression. Moreover, Nicolae and Sauerland

(2015) report a series of experiments which they argue demonstrate that upper-bounded interpretation of *or* can be modulated by making the explicitly exhaustive *either-or* contextually salient. In one between-subjects experiment from that paper, participants who were asked to rate the likelihood of an exclusivity reading given sentences containing *or* behaved no differently from participants in a separate experimental condition who were asked to rate the likelihood of an exclusivity reading with *either-or*. However, in a second experiment, where participants saw critical trials containing both *or* and *either-or*, there was an observed strength asymmetry between the two lexical items, where *or* was interpreted less exclusively than *either-or*.

These results suggest that pragmatic interpretation can be modulated by raising the salience of linguistic constructions not commonly assumed to be within the set of alternatives for an utterance - including, in the case of Nicolae and Sauerland (2015), a symmetric alternative (*either-or*) to *or*'s putative scalemate *and*. This already problematizes the aforementioned assumption that an utterance's alternatives are relatively few and are semantically consistent, unless we provide additional stipulations regarding the contextual availability of alternatives outside this constrained set. If it can be specifically shown that raising the salience of semantically upper-bounded, putative symmetric alternatives (e.g *some but not all* with respect to *some*) can modulate interpretation of weak scalar items, then we problematize the project of specifying the constraints on alternative computation in order to break symmetry.

However, the current state of the literature on implicature priming leaves open several questions to be addressed in future empirical work. First, do the priming phenomena reported in past papers on a relatively small set of putative lexical scales (e.g. <some, all>, cardinal numbers) generalize to other, less-researched putative scales, including scales built from gradable adjectives and ad-hoc lexical scales? That is, does raising the salience of a stronger, upper bound-encoding alternative prime upper-bounded interpretations across a broad range of weak scalar items? Second, what is the nature of priming behavior across a range of lexical scales when a semantically upper-bounded, putative symmetric alternative is made salient rather than the putative conventional strong alternative? It is possible that the result of Nicolae and Sauerland (2015) generalizes across scales, whereby the symmetric alternative appears to compete pragmatically with the weak scalar form and thus lead to less exhaustive interpretations of that form. However, it is also a priori possible that raising the salience of symmetric alternatives leads to priming behavior in the other direction (i.e. *more* exhaustive interpretations), which is less suggestive of the idea that the semantically upper-bounded form behaves as a pragmatic alternative to the weak scalar item. Rather, such behavior would suggest that raising the salience of semantically upper-bounded forms primes listeners to interpret lower-bounded weak scalar items similarly. Lastly, what types of experimental tasks elicit implicature priming behavior? In all of the studies mentioned in this section (with the exception of Nicolae and Sauerland 2015), priming involves associating forms of interest with contextually-specified meanings. It remains an open question as to whether or not simply raising the salience of forms of interest, without biasing participants towards a particular interpretation of those forms, is a sufficient means of priming. Experiments 1 and 2 of this paper are designed to address the above questions.

Experiments 1 and 2

Experiments 1 and 2 investigate the capacity of semantically upper bound-encoding forms (putative **strong scalar alternatives**) and semantically upper-bounded forms (symmetric to the strong scalar alternatives) to prime subsequent interpretation of weak scalar items. The hypotheses are below:

The *null hypothesis:* for a given scale, prior exposure to [semantically upper bound-encoding / semantically upper-bounded] forms does not affect the degree to which weak scalar items receive an upper-bounded interpretation.

The *meaning blocking hypothesis*: for a given scale, prior exposure to [semantically upper bound-encoding / semantically upper-bounded] forms causes weak scalar items to compete pragmatically with those forms, leading to divergence in the interpretation of weak scalar items and the form of interest.

The *meaning priming hypothesis*: for a given scale, prior exposure to [semantically upper bound-encoding / semantically upper-bounded] forms causes weak scalar items to be interpreted more similarly to those forms (i.e. convergence in the interpretation of weak scalar items and the form of interest).

Experiment 1 investigates the effects of prior exposure to strong scalar alternatives on interpretation of weak scalar items, while Experiment 2 (deployed at the same time as Experiment 1) investigates the effects of prior exposure to semantically upper-bounded forms, henceforth referred to as **exhaustive alternatives**. The experiments reported here additionally explore whether simply raising the salience of linguistic forms can prime subsequent interpetation of weak scalar items, in contrast to the vast majority of implicture priming studies conducted to date, in which participants are trained to associate forms with particular meanings before critical trials.

Experiment 1

Methods

Participants

93 participants (US IP Addresses, prior approval rating $> 95\%$) were recruited on Amazon Mechanical Turk. Participants were paid \$1.20 for a task which took, on average, roughly 7 minutes to complete.

Materials

Each trial consisted of a **bolded** sentence (the target item), attributed to a named speaker, followed on a separate line by the text "How likely is it that [the named speaker] meant to suggest", followed on a separate line by an *italicized* sentence (the 'interpretation sentence').

Each trial also contained a continuous scale bounded by the endpoints "very unlikely" (on the left end of the scale) and "very likely" (on the right end of the scale). Critical trials of the experiment differed from priming and filler trials of the experiment in that the bolded sentence was also colored red, as can be seen in Figures 1 and 2.[11].

As Figures 1 and 2 illustrate, the target item on priming trials (in the priming condition of the experiment) contained a strong scalar construction (e.g. *all*). Participants in the priming condition saw three priming trials in a row, where each target item contained the same lexical item. Participants in the no-prime condition of the experiment did not see a scalar item on priming trials. In both condiitons, however, the four priming trials were immediately followed by the identical critical trial, which contained a weak lexical item (e.g. *some*) from the same scale as the strong scalar construction seen in the priming condition. Moreover, in all conditions, participants saw the same interpretation sentences on both priming trials and critical trials, which always conveyed an exhaustive meaning from the scale of interest (e.g. *not every* for the <all, some> scale). Between participants, order of presentation of priming trials was always consistent for a given block of trials (i.e. order of priming trials did not vary within a block).

The three priming trials, plus the one critical trial and a subsequent random filler item, constituted one block of trials. The name of the speaker was kept consistent within a single block, but names were randomly distributed between blocks such that no speaker saw the same name in more than one block. The experiment determined the distribution of names randomly every time the experiment was opened by a participant; that is, though Figures 1 and 2 happen to show the name "Henry" for the *some* block, this name had an equal chance of association with any other experimental block. In total, participants saw six blocks, constructed using items from the following scales: <all, some>, <and, or>, <is, looks like> (an ad-hoc scale), <delicious, palatable>, and <impossible, hard>. There were two lists of trials, such that each participant saw three no-prime condition blocks and three priming condition blocks (presented in random order). In total, participants saw 32 trials, including two training items (explained below).

'Some'

- Stronger alternative: 'all'
- Interpretation question: "not every…"
- Example sentence, priming condition: **Sally saw all of her former classmates at her high school reunion**
- Example sentence, no prime condition: **Sally saw her favorite teacher, Mr. Meyer, at her high school reunion.**
- Example interpretation question: *Sally didn't see every former classmate of hers at the reunion*

'Or'

[11] These and all other figures can be found in the appendix of the paper

- Stronger alternative: ‘and’
- Interpretation question: “only one of these things…”
- Example sentence, priming condition: **Peter inherited the painting and the wardrobe from his grandmother**
- Example sentence, no prime condition: **Peter inherited the painting from his grandmother, whereas his aunt Jill inherited the wardrobe.**
- Example interpretation question: *Peter inherited only one of these things from his grandmother*

‘Looks like’

- Stronger alternative: ‘is’
- Interpretation question: “something other than…”
- Example sentence, priming condition: **It’s raining outside.**
- Example sentence, no prime condition: **It’s snowing outside.**
- Example interpretation question: *The weather outside is something other than rain*

Cardinal numbers (i.e. 34)

- Stronger alternative: ‘35’
- Interpretation question: “no more than 34…”
- Example sentence, priming condition: **35 books were stolen from the local library yesterday.**
- Example sentence, no prime condition: **There was a break-in at the local library yesterday.**
- Example interpretation question: *No more than 34 books were stolen from the local library yesterday*

‘Palatable’

- Stronger alternative: ‘delicious’
- Interpretation question: “not exceptionally tasty…”
- Example sentence, priming condition: **Guillermo’s brewery has just released a new beer that is delicious.**
- Example sentence, no prime condition: **Guillermo’s brewery has just released a new beer.**
- Example interpretation question: *Guillermo’s brewery has just released a new beer that is not exceptionally tasty*

‘Hard’

- Stronger alternative: ‘unsolvable’
- Interpretation question: “not impossible to…”

- Example sentence, priming condition: **The homework that Professor Bridges assigns is unsolvable.**
- Example sentence, no prime condition: **The homework that Professor Bridges assigns is always due the Monday after it is assigned.**
- Example interpretation question: *The homework that Professor Bridges assigns is not impossible to finish*

As mentioned above, each block also contained a randomly-distributed filler trial. These filler trials served as controls, in that for half of the trials, the target item uncontroversially implied the interpretation sentence, while for the other half of the trials, the target item clearly did not imply the interpretation sentence. In the former case, a response in the lower half of the scale was clearly unjustified; in the latter, a response in the upper half of the scale was similarly a priori dispreferred. Divergent behavior on more than one of these trials led to exclusion of a participant's data. Data were also excluded from participants who reported their native language to be something other than English.

Procedure

Participants were shown the following instructions:

In this experiment, you will see up to 7 pairs of sentences. For each pair, please judge how likely it is that the speaker of the **bolded** sentence meant to imply the sentence in *italics*.

"Indicate your judgment by clicking on the sliding scale that will appear below the sentences. Don't think too deeply about the sentences, just provide your first intuition about what the speaker meant!"

After clicking through these instructions to start the experiment, participants were informed that they would be given two practice trials which contained the following sentence pairs:

Practice trial 1:

A handful of people showed up to the meeting.

Nobody showed up to the meeting.

Practice trial 2:

Judith's hometown has a population of less than 10,000.

Fewer than 10,000 people live in Judith's hometown.

If a participant provided a response within the upper portion of the scale on the first practice trial, and/or if she provided a response within the lower portion of the scale for the second practice trial, the following message appeared:

"Not quite! Remember that if the implied meaning is likely, your answer should be on the higher end of the scale. If the implied meaning is unlikely, your answer should be on the lower end of the scale."

In these cases, participants had to move the slider to the opposite end of the scale before continuing. There was no feedback on subsequent trials. In order to move to the next trial, participants had a to make a selection on the scale.

Predictions

The hypotheses specified above make differing predictions as to how behavior should change on critical trials between prime and no-prime conditions of the experiment (and for a given scale).

The *null hypothesis* predicts that between prime and no-prime conditions of the experiment, there should be no change in behavior on critical trials. That is, for example, *some* should be rated just as likely to communicate *not every* regardless of whether or not the preceding three trials contain the item *all* in the bolded sentence.

The *meaning blocking hypothesis* predicts that there should be a change in behavior on critical trials, such that weak scalar items in the priming condition are rated more exhaustively in the priming condition than in the no-priming condition of the experiment (due to divergence in the interpretation of the weak scalar item and interpretation of its stronger scalemate).

The *meaning priming hypothesis* predicts the opposite change, namely that weak scalar items in the priming condition are rated less exhaustively in the priming condition than in the no-priming condition of the experiment (due to convergence in the interpretation of the weak scalar item and interpretation of its stronger scalemate).

Results

No participants were excluded for reporting a native language other than English; however, 14 participants were excluded for failing more than one of the exclusion trials described above, meaning that the subsequent analyses include data from 79 participants.

Figure 3 displays the by-condition mean exhaustivity ratings on critical trials, broken down by scale type. A visual inspection of the results suggests that behavior between conditions is highly overlapping, with the possible exception of the <looks like, is> and <palatable, delicious> scales. This observation is further corroborated by a linear mixed effects regression analysis performed using the `lme4` package in R. Response on the sliding scale, with the endpoints "Very Unlikely" and "Very Likely" coded as 0 and 100, respectively, was predicted from fixed effects of condition (No-prime or Prime), scale, and their interaction, along with a random by-participant slope for condition. The reference level fixed effect of condition was the No-prime condition, and the reference level for the fixed effect of scale was the <hard, unsolvable> scale. The results of this analysis did not yield a significant effect of condition ($\beta = 1.044$, SE = 6.13, p = 0.86) but yielded signficant effects of scale type ($p < 0.01$) for all scales except the <palatable, delicious> scale (p = 0.27). The interaction of condition and scale type was not significant in any case except the <looks like, is> scale, for which the interaction approached significance ($\beta = -15.75$, SE = 9.04, p = 0.08).

Figure 4 illustrates the by-subject change in behavior in Experiment 1, with points (connected by a line) representing an individual participant's change in response between priming trials (expressed as a composite mean value of response on priming trials) and the corresponding critical trial of a given block of the experiment. Broadly, we see that in the Prime condition, participants provide low exhaustivity ratings on priming trials, which is to be expected, as the sentences interpreted on these trials contain scalar items which semantically encode the upper bound. There is, in general, substantially more variability on in priming trials in the No-prime condition: this is also not surprising, as the sentences interpreted in these conditions do not contain lexical material which semantically encodes the upper bound. Lastly, Figure 5 illustrates how behavior changes sequentially within a given block on the experiment (that is, from the first prime through the critical trial of the experiment). From this figure, it is particularly clear that mean exhaustivity ratings were low overall on priming trials in the Prime condition and highly variable in the No-prime condition, with the important exception of the *or* block. This exception, however, is not surprising in light of the fact that all priming trials in the No-prime condition of the *or* block contained a connective (*whereas*) that presupposes the exhaustive meaning probed by the interpretation question.

Discussion

The results of Experiment 1 are generally surprising in light of the previous experimental pragmatics literature on implicature priming. While previous work had reported that raising the salience of putative strong alternatives to weak scalar items modulates subsequent interpretation of those items (in a way supporting the meaning blocking hypothesis, where weak scalar items are interpreted more exhaustively relative to a neutral baseline condition), this was not an effect observed in Experiment 1 - even on the relatively well-researched <some, all> and cardinal number scales. A change in behavior on critical trials between No-prime and Prime conditions of the experiment was only observed to a marginal degree with the *looks like* block, which involves the construction of a putative ad-hoc contextual scale with *looks like (an X)* and *is (an X)*. To the extent that priming with the strong alternative *is* in the Prime condition marginally decreased exhaustivity ratings of *looks like* relative to the No-prime condition, there is marginal support for the meaning priming hypothesis in this instance. The inability of this paradigm to replicate an attested priming effect suggests that raising salience of putative alternative forms through exposure alone (and without additional form-meaning association tasks) is generally insufficient to modulate subsequent interpretation of lower-bounded weak scalar items, at least in the case of the scales explored here.

Experiment 2

While the Prime condition of Experiment 1 involved forms which semantically encode the upper bound of a putative lexical scale, Experiment 2 (reported below) investigated whether and how raising the salience of exhaustive, semantically upper-bounded forms modulates interpretation of weak scalar items.

Methods

Participants

107 participants (US IP Addresses, prior approval rating > 95%) were recruited on Amazon Mechanical Turk. Participants were paid \$1.20 for a task which took, on average, roughly 7 minutes to complete.

Materials and Procedure

The materials in Experiment 2 were identical to the materials from Experiment 1, except that blocks in the Prime condition featured priming trials with lexical items which semantically encoded an exhaustive, upper-bounded meaning on the scale of interest. The sentences in the No-prime condition, the critical trials in all conditions, and the interpretation questions on all trials remained consistent between these two experiments, as did the filler trials.

'Some'

- Exhaustive alternative: 'some but not all'
- Interpretation question: "not every…"
- Example sentence, priming condition: **Sally saw some but not all of her former classmates at her high school reunion**
- Example sentence, no prime condition: **Sally saw her favorite teacher, Mr. Meyer, at her high school reunion.**
- Example interpretation question: *Sally didn't see every former classmate of hers at the reunion*

'Or'

- Exhaustive alternative: 'or, but not both'
- Interpretation question: "only one of these things…"
- Example sentence, priming condition: **Peter inherited the painting or the wardrobe from his grandmother, but not both**
- Example sentence, no prime condition: **Peter inherited the painting from his grandmother, whereas his aunt Jill inherited the wardrobe.**
- Example interpretation question: *Peter inherited only one of these things from his grandmother*

'Looks like'

- Exhaustive alternative: 'looks like, but isn't'
- Interpretation question: "something other than…"
- Example sentence, priming condition: **It looks like it's raining outside, but it isn't.**

- Example sentence, no prime condition: **It's snowing outside.**
- Example interpretation question: *The weather outside is something other than rain*

Cardinal numbers (i.e. 34)

- Exhaustive alternative: 'exactly 34'
- Interpretation question: "no more than 34…"
- Example sentence, priming condition: **Exactly 34 books were stolen from the local library yesterday.**
- Example sentence, no prime condition: **There was a break-in at the local library yesterday.**
- Example interpretation question: *No more than 34 books were stolen from the local library yesterday*

'Palatable'

- Exhaustive alternative: 'palatable, but not delicious'
- Interpretation question: "not exceptionally tasty…"
- Example sentence, priming condition: **Guillermo's brewery has just released a new beer that is palatable, but not delicious.**
- Example sentence, no prime condition: **Guillermo's brewery has just released a new beer.**
- Example interpretation question: *Guillermo's brewery has just released a new beer that is not exceptionally tasty*

'Hard'

- Exhaustive alternative: 'hard, but not unsolvable'
- Interpretation question: "not impossible to…"
- Example sentence, priming condition: **The homework that Professor Bridges assigns is hard, but not unsolvable.**
- Example sentence, no prime condition: **The homework that Professor Bridges assigns is always due the Monday after it is assigned.**
- Example interpretation question: *The homework that Professor Bridges assigns is not impossible to finish*

Procedure was identical to the procedure of Experiment 1.

Predictions

The *null hypothesis* predicts that between prime and no-prime conditions of the experiment, there should be no change in behavior on critical trials. That is, for example, *some* should be rated just as likely to communicate *not every* regardless of whether or not the preceding three trials contain the item *some but not all* in the bolded sentence.

The *meaning blocking hypothesis* predicts that there should be a change in behavior on critical trials, such that weak scalar items in the priming condition are rated less exhaustively in the priming condition than in the no-priming condition of the experiment (due to divergence in the interpretation of the weak scalar item and interpretation of the corresponding exhaustive, semantically upper-bounded lexical item).

The *meaning priming hypothesis* predicts the opposite change, namely that weak scalar items in the priming condition are rated more exhaustively in the priming condition than in the no-priming condition of the experiment (due to convergence in the interpretation of the weak scalar item and interpretation of the exhaustive form).

Results

Two participants were excluded for reporting a native language other than English; additionally, three participants were excluded for failing more than one of the exclusion trials described above, meaning that the subsequent analyses include data from 102 participants. Figure 6 displays the by-condition mean exhaustivity ratings on critical trials, broken down by scale type. A visual inspection of the results again suggests, as with Experiment 1, that behavior on critical trials between conditions is highly overlapping. A linear mixed effects regression analysis (identical in structure to the analysis performed for Experiment 1) did not yield a significant effect of condition ($\beta = 0.159$, SE = 5.51, p = 0.98). There were signficant effects of scale type ($p < 0.01$) for all scales except the <palatable, delicious> scale (p = 0.60) and the <looks like, is> scale (p = 0.13). The interaction of condition and scale type was not significant for any scale.

Figure 7 compares behavior on critical trials between Experiment 1 and Experiment 2. Response on critical trials was predicted with a linear mixed effects model from fixed effects of experiment (Experiment 1 or Experiment 2), scale, and their interaction, along with a random by-participant slope for condition. The reference level fixed effect of experiment was Experiment 1, while the reference level for the fixed effect of scale was the <hard, unsolvable> scale. The results of this analysis did not yield a significant effect of experiment ($\beta = 1.044$, SE = 6.13, p = 0.86). The interaction of experiment and scale type was not significant in any case except the <looks like, is> scale, for which the interaction was significant ($\beta = 22.10$, SE = 8.16, p = 0.007).

General Discussion - Experiments 1 and 2

As with Experiment 1, the paradigm employed in Experiment 2 generally failed to show priming effects on critical trials. Unlike Experiment 1, where an interaction of the *looks like* block with condition type was marginally significant, none of the scale-condition interactions in Experiment 2 even approached significance; however, subsequent comparison of critical trial behavior in the *looks like* block between Experiments 1 and 2 did reveal a signficant difference, with *looks like* being interpreted more exhaustively after exposure to the strong alternative form *is* than after exposure to the exhaustive form *looks like, but isn't.* Thus, the *looks like* blocks were overall much more sensitive to the experimental manipulations explored here than were the other blocks. Importantly, the *looks like* blocks differ from the rest of the blocks in Experiments 1 and 2 in that there is no semantic entailment relationship between being an X and looking like an X. Thus, for proponents of theories of alternatives derived from scales, any contrast between *is an X* and *looks like an X* is an ad-hoc construct of the local experimental context and not a conventionalized scale as exemplified by, e.g., <some, all> or the scale of cardinal numbers.

That this highly context-sensitive scale uniquely exhibits priming behavior in this generally insensitive paradigm is not necessarily surprising on this view: simply raising the salience of linguistic forms (without further training participants to associate those forms with certain meanings) might be insufficient for modulating interpretation of a given utterance in case the forms already conventionally contrast with that utterance; however, this type of priming might be reasonably expected to be more effective just in case the priming establishes (rather than reinforces) pragmatic competition between linguistic forms. Thus, proponents of categorical theories of implicature, who rely on highly-constrained theories of alternative computation in order to avoid the symmetry problem, explain the difference between the *looks like* blocks and other blocks of the experiments as a difference of kind: the <looks like, is> scale is simply not a scale in the same way the <some, all> scale is.

Proponents of gradient theories of implicature, such as the probabilistic Rational Speech Act model, may alternatively claim that the difference between the *looks like* scale and other scales explored here is not a difference of kind but of degree: whereas *is an X* is an a priori relatively unlikely utterance for a speaker to produce in a context where she produces *looks like an X*, the strong 'scalar' alternatives in the other blocks are a priori very likely in contexts where the corresponding weak form is produced, to the point that subsequent mentions of the strong alternative do little to additionally raise the salience of these forms (and hence do little to modulate interpretation of the weak scalar item).

Future work should investigate priming behavior in this paradigm across a broader range of putative lexical scales - especially scales where there is evidence that there is relatively faint pragmatic competition between weak scalar items and stronger alternative forms. Three of the six scales explored in Experiments 1 and 2 - <some, all>, <hard, unsolvable>, and <palatable, delicious> - were demonstrated by van Tiel et al. (2016) to exhibit relatively robust implicature behavior; that is, participants in their studies interpreted the weak form as implicating the negation of the strong form relatively frequently compared to, e.g., the scales <special, unique>, <pretty, beautiful>, and <intelligent, brilliant>. Where pragmatic competition between weak and strong forms on a putative scale is not particularly robust a priori, we might predict the paradigm explored here to be more sensitive.

Lastly, it must be mentioned that because every scale block had only one critical item, (lack of) priming effects is confounded with possible artifacts of the individual items used to elicit exhaustivity judgments. Future work should address this confound by eliciting judgments across a broader range of items for a given scale.

Symmetry and ignorance

Experiments 1 and 2 investigated the nature of the set of alternatives active in pragmatic interpretation by exploring how exposure to particular forms (including forms not commonly assumed to be in the set of alternatives of a given utterance) modulates subsequent interpretation of utterances. While the paradigm generally failed to show priming effects (even for putative lexical scales such as <some, all> and cardinal numbers - where priming behavior is attested in other paradigms), the results offer the preliminary and speculative conclusion that for two linguistic forms already exhibiting a very high degree of pragmatic competition, merely raising the salience of one through explicit mention has little to no effect on interpretation of the other form.

In what follows, I motivate and report the results of a third experiment designed to investigate the nature of pragmatic reasoning about utterance alternatives. This experiment takes as its point of departure a specific empirical prediction of categorical theories of scalar implicature with respect to the type of inferences derived when symmetric utterance alternatives are made contextually salient.

According to this strain of theory (Sauerland 2004, Katzir (2007), Nicolae and Sauerland (2015)), activation of symmetric communicatively-stronger alternatives should yield the inference that the speaker is not in a position to assert either alternative, thereby blocking any implicature regarding the truth/falsity of those alternatives. This claim is advanced by Sauerland (2004), who argues that "being opinionated about one of the two [symmetric] alternatives means being opinionated about the other, and the hearer can conclude that the speaker is not opinionated about either. In other words, the speaker did not have access to the relevant evidence after all, and so was not in a position to believe either [of the alternatives]. The speaker is simply ignorant of the relevant facts" (2004: 674). Motivation for claim comes from data such as (24), which according to Katzir (2007) gives rise to an uncertainty inference regarding the first conjunct (i.e. speaker does not know whether John spoke to all of the girls yesterday) (example taken from Katzir (2007), his example 14a):

(24) John talked to some of the girls yesterday, and he talked to some but not all of the girls today.

On Katzir's analysis, (24) activates the following symmetric alternatives, and it is precisely the cancellation of these symmetric alternatives in tandem which gives rise to the uncertainty inference:

(25) John talked to **all** of the girls yesterday, and he talked to some but not all of the girls today.

(26) John talked to **some but not all** of the girls yesterday, and he talked to some but not all of the girls today.

Katzir derives (26) and (26) as alternatives to (24) in the following way: first, (25) activated as an alternative to (24) because it is structurally simpler than (and as relevant as) (24).

Furthermore, (26) is an alternative to (24) by virtue of the same structural-complexity principle (because *some but not all* is a subtree of 24, it may substitute *some* in the first conjunct of 24 to yield the alternative 26). Because Katzir's account specifically derives symmetric alternatives in this instance, the analysis must explain why (24) does not have the feel of contradiction (as we might expect if the pragmatically-enriched meaning of 24 included the negation of both 25 and 26). Following Sauerland (2004) and Fox (2007), Katzir argues that "no strengthening [of this kind] takes place, since it is impossible to strengthen... without contradicting the original assertion" (2007: 682).

A similar claim is advanced more recently by Nicolae and Sauerland (2015), who claim that the putative symmetric alternative *either-or* may displace *and* as the alternative to *or*, hence the strength asymmetry between *either-or* and *or* observed in their experiment (if *and* functions as the alternative to *or*, then no strength asymmetry is predicted between *or* and a lexical item such as *either-or* which semantically encodes the upper-bound - assuming a categorical view of scalar implicature). However, inference regarding *either-or* is, according to the authors, merely that the speaker is unsure whether or not the exclusive meaning holds: if the inference were the negation of the *either-or* alternative, one would expect *or* to be interpreted as conjunction, which the authors find highly unlikely given their experimental results (*or* was still rated by most participants as more likely than not expressing an exclusive meaning - just slightly less likely overall than *either-or*). Though Nicolae and Sauerland (2015) do not attempt to generalize their analysis beyond the case of *or*, the claim is again that activating a symmetric alternative gives rise to ignorance inferences regarding the truth/falsity of stronger meanings.

However, as noted earlier, gradient theories of scalar implicature allow for more nuanced predictions regarding the inferences listeners draw when symmetric alternatives are activated. For example, the RSA analysis of *or* explored above demonstrated that activating - and raising the contextual salience of - the symmetric alternative *either-or* can be modeled in such a way that one predicts attenuated exhaustivity inferences (rather than ignorance inferences regarding the truth/falsity of the exhaustive state). In general, the finding is that in RSA, the presence of symmetric alternatives does not necessarily inhibit pramgatic inference in the form of posterior belief update of the likelihood of world states consistent with those stronger alternatives.

The experiment reported below explores whether or not this is a desirable feature of a model of pragmatic competence by considering how interpretation of *looks like an x* changes when two symmetric alternatives (*is an X*, *isn't an X*) are made to compete with *looks like* in an experimental setting. As with Experiments 1 and 2, the scale built from *looks like* is a putative ad-hoc scale, i.e. not one built from conventionalized contrasts between *looks like* and other items (and not one that can be derived from entailment relationships between *looks like an X* and, e.g., *is an X*). However, one finding of note from Experiments 1 and 2 was that this putative ad-hoc scale was sensitive to the experimental manipulation in a way that other putative conventionalized lexical scales (e.g. cardinal numbers) were not, which suggests that *looks like* is an appropriate scale to explore in experimental paradigms where behavioral changes across conditions of an experiment could be relatively subtle.

Moreover, this particular ad-hoc scale, and interpetation of utterances of the form *It looks* *like an X* in particular, have already been explored in a series of studies which report effects of speaker prosody (i.e. verb-focus *It LOOKS like an X* vs noun-focus *It looks like an* x) and prevalance of alternative linguistic forms in the experimental context (e.g. *It's an X*) (Kurumada et al. 2012, Kurumada et al. (2014)). For example, Kurumada et al. (2012) report the results of a forced-choice task whereby participants choose between two images on a screen - one target image corresponding to an imageable high-frequency noun and one competitor image visually similar to the target - given that a speaker in the experiment has produced an utterance of the form *It looks like a [target]* (with either verb-focus or noun-focus prosody). In their Experiment 2, they report that selections of the competitor on these *looks like* trials increase when a subset of trials contain utterances of the form *It's a [target]*. The authors claim that raising the salience of *It's a [target]* raises the strength of pragmatic competition between this form and *It looks like a [target]*; thus, *It looks like a [target]* indicates that the speaker was not in a position to use the simpler, communicatively stronger form (hence implicating that the correct picture is not the target). Experiment 3 of the present (reported below) explores the consequences of introducing the symmetric alternatives *It's a [target]* and *It's not a [target]* in this paradigm and in particular investigates the effect on interpretation of *It looks like a [target]*.

Experiment 3

Experiment 3 investigates two competing hypotheses of how activating symmmetric alternatives modulates pragmatic interpretation of utterances, building off the discussion from above:

The *ignorance* hypothesis: activation of symmetric alternatives to a given utterance yields ignorance inferences.

The *modified inference* hypothesis: activation of symmetric alternatives to a given utterance gradiently modulates the strength of pragmatic inference but does not yield ignorance inferences.

Methods

Experiment 3 modifies and extends the forced-choice task of Kurumada et al. (2012) whereby participants choose between two images given an utterance produced by a speaker (their Experiment 2). Whereas in that experiment the utterances were presented as recorded audio, stimuli in this experiment are presented as text. The images used in the present study are the same used by Kurumada et al. (2012) in their original design.

Participants

120 participants (US IP Addresses, prior approval rating > 90%) were recruited on Amazon Mechanical Turk. Participants were paid $0.85 for a task which took, on average, roughly 3.5

minutes to complete.

Materials

24 imageable high-frequency nouns (the same nouns used by Kurumada et al.) were embedded into the following sentence frames: "It is an X", "It looks like an X", and "It isn't an X". These nouns were associated with a target and competitor image, as in Kurumada et al.'s original study. Participants were randomly assigned to one of four conditions, which were created from these items: in the **control condition**, all items were of the form "It looks like an X"; in the **target condition**, 8 of the 24 items in the control condition were replaced by sentences of the form "It is an X"; in the **not-target condition**, 8 of the 24 items in the control condition were replaced by sentences of the form "It isn't an X"; in the **symmetric condition**, 4 of the 24 items were of the form "It is an X" while another 4 of the 24 were of the form "It isn't an X". (The original Kurumada et al. study contained only the first two conditions in order to investigate the effect of "It is an X" on interpretation of "It looks like an X"). In the target, non-target, and symmetric conditions, it was the same 8 items which changed from a *looks like* trial to another trial type. Trials were presented in random order; however, in all four conditions, the 6 of the 8 items whose presentation changes across condition were shown in the first half of the experiment.

In addition, every condition contained the same four filler trials which contained instructions (presented in italics) to click on the right/left image on the screen. Divergent behavior on more than one of these trials led to exclusion of a participant's data. Data were also excluded from participants who reported their native language to be something other than English.

Procedure

As in the study reported by Kurumada et al. (2012), participants were provided a cover story in which a school teacher described objects from an encyclopedia which contained pictures that were occluded to the students in his class. The teacher, named Mr. Davis in the current study, produced utterances of the form [*It looks like an X/It is an X/It isn't an X*] to convey what he saw on the page. Participants were instructed to provide their answer as to what the teacher saw by selecting one of two images (the target and competitor image) on their screen. Participants were also told: "At times, you will see instructions, served in italics, which explicitly instruct you to select one of the two images. These serve as attention checks; failing too many of them can affect your compensation!" (However, all participants were paid $0.85 regardless of their performance on the filler trials). There was no feedback between trials, mirroring the procedure of Kurumada et al. A sample trial of the experiment is shown in Figure 8.

Predictions and linking assumptions

The first prediction is that the experiment will replicate one of the main findings from Kurumada et al. (2012), namely an increase in the proportion of competitor pictures chosen in the target condition relative to the control condition (on "It looks like an X" trials). In the target condition, the pragmatic alternative "It is an X" is made more salient; participants in the target condition are thus aware of the speaker's ability to refer to the target image using the simpler and communicatively stronger utterance "It is an X". As a result, "It looks like an X" will convey the speaker's intention to identify the competitor image more often in the target condition than in the control condition. The linking assumption is that interpreting "It looks like an X" as "It isn't an X" is associated with selecting the competitor image rather than the target image.

I furthermore predict a decrease in the proportion of competitor pictures chosen in the not-target condition relative to the control condition, by similar reasoning. In the not-target condition, the pragmatic alternative "It isn't X" is made more salient; participants in the not-target condition are thus aware of the speaker's ability to refer to the competitor image using the simpler and communicatively stronger utterance "It isn't an X". As a result, "It looks like an X" will convey the speaker's intention to identify the competitor image less often in the not-target condition than in the control condition.

However, the hypotheses described above make different predictions regarding behavior in the symmetric condition. On its own, the *ignorance* hypothesis is underspecified regarding its empirical predictions, but I propose two models which link the ignorance inference to behavior in a priori plausible ways:

First, in the `ignorance-1` model, selection of target and competitor images on *looks like* trials is at chance; that is, when symmetric alternatives are activated, there is an 0.5 probability that participants will choose the target image when the utterance is of the form "It looks like a [target]".

Conversely, in the `ignorance-2` model, participants rely on their a priori expectations of the speaker's intended referent, which may not be uniform over the two images. That is, the probability of a participant selecting a target image in the symmetric condition is equal to the probability of a participant selecting the target image given no linguistic input.

Finally, the `modified-inference` model is a Rational Speech Act model of pragmatic competence whereby listeners are represented as probabilistic distributions over possible intended referents (either the target image or competitor image) given observation of a particular utterance. This distribution is a function not only of a soft-optimal pragmatic speaker intending to communicate an intended referent given possible utterances but also of participants' a priori expectations of the speaker's intended referent, as well as the relative nameability of the images on the screen of every trial. In the model, the nameability of a referent R inversely affects the cost of an utterance $U(R)$ containing R, such that less nameable R is a priori, the more costly we expect $U(R)$ to be. For the `modified-inference` model, I assume a linking function from pragmatic competence to behavior whereby selection of the target image on *looks like* trials will occur if the model of interest returns `true` when sampling from a Bernoulli distribution with probability equal to the posterior probability of the target image as intended referent given observation of the utterance "It looks like an X". The model is further explicated in the 'Model comparison' section below and can be accessed here: [github link].

Norming studies

Prior expectation norming study

The `ignorance-2` and `modified-inference` models specified above make reference to participants' priori expectations of the speaker's intended referent. In one norming study, the *prior expectation* norming study, 30 participants (Amazon Mechanical Turk, same geographic and approval rating criteria as above) are shown a variant of the main experiment, where instead of utterances of the form [*It looks like an X*/*It is an X*/*It isn't an X*], the teacher's utterance is obscured to the participant (i.e. the participant reads only **mumble mumble** on every trial). The proportion of target vs. competitor selections in this norming study served as an experimentally-elicited proxy for the likelihood of participants selecting the target or competitor image absent lingusitic input from the speaker. Participants in this norming study did not participate in the other studies of Experiment 3 and were compensated $0.85. A sample trial from this study is presented in Figure 9.

Nameability norming study

Additionally, the `modified-inference` model makes reference to the a priori nameability of the objects on the screen of a given trial. The nameability of all objects displayed in the experiment is elicited experimentally with the *nameability* norming study, in which 30 participants (Amazon Mechanical Turk, same geographic and approval rating criteria) are sequentially shown the 48 images from Experiment 3 and asked to name the object they see. On every trial of this study, participants were given the option to respond that they did not know what the object was. Participants in this norming study did not participate in the other studies of Experiment 3 and were compensated $1.40 for a task which lasted roughly 7 minutes on average. A sample trial from this study is presented in Figure 10.

Results (main study)

Data from one participant were excluded because that participant reported a native langauge other than English. A further four participants were excluded for failing one or more of the filler trials of the experiment. Therfore, subsequent analyses include data from 115 participants.

Figure 11 displays the mean proportion of competitor images selected on the 16 *looks like* trials shared across the four conditions of the experiment, while Figure 12 shows mean behavior across conditions broken down by item. A visual inspection of Figure 12 suggests that the pattern of results seen in Figure 11 is fairly robust across individual items. Pairwise mixed effects logistic regression analyses were performed using the `lme4` and `lsmeans` libraries in R to investigate log likelihood of choosing the competitor image from a fixed effect of experiment condition, a random intercept for participant, and a by-item random slope for condition. The pairwise comparisons were significant between the target and control conditions ($\beta = 4.04$, SE = 1.27, P = 0.008) and between the symmetric and control conditions ($\beta = 4.07$, SE = 1.18, P = 0.003), and the comparison approached significance between the not-target condition and control condition ($\beta = 2.89$, SE = 1.15, p = 0.058). No other pairwise comparisons approached significance. A subsequent analysis of log likelihood of choosing the competitor image from fixed effects of experiment condition and order, along with a by-participant random slope for order and a by-item random slopes for order and condition, revealed a negligible marginally significant ordering effect in the experiment ($\beta = -0.11$, SE = 0.06, P = 0.083).

Interim discussion - Experiment 3

The results replicate one of the main findings of Kurumada et al. (2012), namely an asymmetry in response between the control and target conditions. When participants see trials with "It's a [target]" forms, they are more likely to interpret "It looks like a [target]" as conveying that the intended referent is the competitor image. However, a second prediction of the experiment - decreased selections of the competitor image in the not-target condition - is not borne out. Rather, selections of the competitor image increase in this condition, as in the target condition. One interpretation of these results is that in the not-target condition, exposing participants to the form "It's not a [target]" also activates the structurally-simpler positive form "It's a [target]". On this account, this structurally-simpler form is more salient in interpretation of *looks like* than the negative form - despite the fact that it is never explicitly mentioned in the not-target condition.

Though the result in the not-target contradicts the original prediction, it is nonetheless clear that participants are sensitive to the presence of "It's not a [target]" forms in this experimental design. Taken in tandem with the difference in behavior in the target and control conditions, this suggests that it is valid to analyze the symmetric condition of the experiment as one where symmetric alternative forms have been activated in interpretation of *looks like*. In what follows, I report a comparison of the `ignorance-1`, `ignorance-2`, and `modified-inference` models of behavior in this condition, given the results of the main study and the two norming studies described above.

Model comparison

This section reports the results of a Bayesian data analysis procedure designed to infer the relative probability that the data in the symmetric condition were produced by the `ignorance-1`, `ignorance-2`, and `modified-inference` models. The analysis assumes a uniform prior over hypotheses. Prior expectation of selections of target/competitor image were computed for the `ignorance-2` and `modified-inference` models using the data from the prior expectation norming study. Using the data from the nameability norming study, nameability of images was determined by computing, for every image, the proportion of participants responding that they did not know how to name the image. This proportion was then subtracted from 1 to create a nameability score for the `modified-inference` model.

The `modified-inference` model assumed the following set of possible alternative utterances: {[*It looks like*/*It is*/*It isn't*] a [target], [*It looks like*/*It is*/*It isn't*] a [competitor]}, along with a uniform distribution over possible cost values for each of these utterances. For every item in the experiment, the cost of {[*It looks like*/*It is*/*It isn't*] a [target]} was assumed to be inverseley proportinal to the nameability of the target image of that item; similarly, the cost of {[*It looks like*/*It is*/*It isn't*] a [competitor]} was assumed to be inversely proportional to the nameability of the competitor. The model furthermore assumed a uniform distribution over possible values of the alpha rationality parameter.

The Bayesian data analysis infers the posterior likelihood of the models by comparing, for every *looks like* trial in the symmetric condition, the model-predicted proportion of selections of the competitor image against the observed proportion in the dataset. The posterior likelihood of the `ignorance-1` model is exceedingly high (probability of virtually 1), compared to virtually zero probability for either the `ignorance-2` or `modified-inference` models. That is, given the results of Experiment 3, the results strongly support the *ignorance* hypothesis (conceived as `ignorance-1`) over the *modified inference* hypothesis explored above. The results of this analysis are unchanged when cost and alpha parameters of the `modified-inference` model are sampled from a joint posterior distribution of model parameter values given the data from the symmetric condition (rather than being sampled from uniform distributions); results are also unchanged when the ignorance models are compared against the `modified-inference` model that has been parameterized with the maximum a posteriori values of the joint parameter distribution.

Exploratory data analysis: comparing different instantiations of the RSA model

Though the comparison between the `ignorance1`, `ignorance2`, and `modified-inference` models assigns zero posterior likelihood to the `modified-inference` model, a caveat of this approach is that the `modified-inference` model will be dispreferred to the ignorance models due to Bayesian Occam's Razor considerations, namely the fact that the `modified-inference` model is vastly more complex than either of the ignorance models (for more discussion of this point, c.f. Goodman et al. (2016)). In what follows, I compare the `modified-inference` model to two modifications:

The `modified-inference-2` model specifies a global uniform cost term for all utterances. The cost of an utterance *U(R)* with a noun denoting referent *R* is inversely proportional to the nameability of R, but for two syntactically distinct utterances *U(R)* and *U'(R)* both containing *R*, the cost term will be equal.

The `modified-inference-3` model specifies a unique cost parameter for the sentential frames

[*It looks like/It is/It isn't*] a [target] and [*It looks like/It is/It isn't*] a [competitor] but does not take into account the nameability of the target or competitor. That is, the cost of *U(R)* and the cost of *U(R')*, where *U(R)* and *U(R')* contain nouns which denote distinct referents but are otherwise syntactically identical, will be identical in cost.

A Bayesian data analysis was performed to infer the relative posterior likelihood of these three models, given uniform priors over the three models and given uniform distributions over each models' cost and alpha parameters. The posterior distribution was computed with WebPPL using Markov Chain Monte Carlo sampling. As with the model comparison above, this analysis yields a categorical preference for one of the models, namely the `modified-inference` model - with virtually zero probability assigned to either of the two other models. Despite the relative complexity of the`modified-inference` model, which specifies unique cost parameters for each utterance (informed in part by the relative a priori nameability of referents in the utterance), the analysis suggests that this added complexity yields superior empirical coverage over simpler variants.

Exploratory data analysis: comparing inferred utterance cost across conditions

Above, I note that implicature priming behavior can be modeled in the RSA framework as the consequence of listeners incrementally adapting their expectations about the alternative utterances produced by a speaker (and the relative salience of those alternatives). The RSA framework as it is discussed here specifies that listeners' prior expectations about speakers' production behavior is a function of listeners' subjective understanding of the relative production cost of possible utterances. As listeners are exposed to more sentences of the form, e.g., 'It looks like a [target]', we might reasonably expect on this view that listeners' understanding of the production cost of this utterance decreases correspondingly. Converseley if participants never see the form 'It's not a [target]' (as they don't in the target and control conditions of the experiment), we might expect listeners to associate this utterance with relatively higher cost than in the not-target and symmetric conditions.

Using data from each condition of the experiment, joint posterior distributions of alpha parameter values and utterance costs from the `modified-inference` model were computed in WebPPL from uniform prior distributions for each parameter, using Markov Chain Monte Carlo sampling. Figures 13, 14, 15 and 16 compare the inferred values across the four experimental conditions. The figures illustrate some clear cases where exposure to a particular utterance form is associated with lower inferred cost of that form: for example, the peak of the posterior distribution of the cost value for 'It's a [target]' is lowest in the target condition, where participants saw the most exemplars of this form. The peak of this cost distribution is higher in the not-target condition, where participants see the negated form 'It's not a [target]' but not the positive form, and even higher in the control condition, where participants see neither 'It's a [target]' nor 'It's not a [target]'. Similarly, in the not-target condition, where participants saw the most 'It's not a [target]' exemplars, the analysis infers the lowest cost for 'It's not a [target]'.

However, this general negative correlation of utterance cost and utterance frequency in the experiment breaks down in some notable respects. For example, the inferred cost distribution of ‘It’s a [target]’ is bimodally distributed close to the endpoints of the range of possible parameter values. There is also a unique bimodal distribution for the cost of ‘It looks like a [target]’ for the not-target condition. Moreover, the analysis infers substantially lower alpha parameter distributions in the target and symmetric conditions of the experiment than in the control and not-target conditions; that is, the analysis infers that listeners in the former two conditions represent the speaker in the experiment (i.e. the teacher, Mr. Davis) as substantially less optimal than in the latter two conditions. Note that while the overall proportion of competitor image selections are highly overlapping in the target, not-target, and symmetric conditions, this analysis suggests that these similar behavioral patterns are in fact the product of differently-parameterized representations at the cognitive level. However, future work should investigate whether these behavioral patterns, and the inferred parameter values seen here, replicate in different settings (for example, with linguistic contrasts other than *looks like / is*).

Discussion - Experiment 3

Experiment 3 took as its point of departure a previously reported study on the context sensitivity of interpretation of *It looks like an X*.

Replicated main effect - not only that, explained main effect as the product of listeners adapting to speaker by updating representation of utterance cost

Got the same thing in the not-target condition: one predicition not borne out, but we infer that cost of ‘It’s a target’ and ‘It’s not a target’ lower in this condition compared to the control condition. Upon seeing the more complex form’It’s not a target’, listeners adapt production expectations of that utterance as well as the simpler positive form.

But, model comparison vastly prefers the simplest model over the rsa model.

General discussion and conclusion

This paper considered two approaches to the symmetry problem of scalar implicature:

References

Leon Bergen, Roger Levy, and Noah Goodman. Pragmatic reasoning through semantic inference. *Semantics and Pragmatics*, 9, 2016.

Lewis Bott and Emmanuel Chemla. Shared and distinct mechanisms in deriving linguistic enrichment. *Journal of Memory and Language*, 91:117 – 140, 2016. ISSN 0749-596X. doi:

https://doi.org/10.1016/j.jml.2016.04.004. URL http://www.sciencedirect.com/science/article/pii/S0749596X16300109. New Approaches to Structural Priming.

Gennaro Chierchia. Broaden your views: Implicatures of domain widening and the "logicality" of language. *Linguistic Inquiry*, 37:535–590, 2006.

James Collins. *Structure-Sensitive Interpretation: A Case Study in Tagalog*. PhD thesis, Stanford University, Stanford, CA, August 2017.

Judith Degen and Michael K. Tanenhaus. Processing scalar implicature: A constraint-based approach. *Cognitive Science*, 39(4):667–710, 2015. doi: 10.1111/cogs.12171. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/cogs.12171.

Ryan Doran, Gregory Ward, Meredith Larson, Yaron McNabb, and Rachel E Baker. A novel experimental paradigm for distinguishing between what is said and what is implicated. *Language*, pages 124–154, 2012.

Danny Fox. Free choice and the theory of scalar implicatures. In *Presupposition and implicature in compositional semantics*, pages 71–120. Springer, 2007.

Danny Fox and Roni Katzir. On the characterization of alternatives. *Natural Language Semantics*, 19(1):87–107, Mar 2011. ISSN 1572-865X. doi: 10.1007/s11050-010-9065-3. URL https://doi.org/10.1007/s11050-010-9065-3.

G. Gazdar. *Pragmatics: Implicature, Presupposition and Logical Form*. Acad. Press, 1979. ISBN 9780122784507. URL https://books.google.com/books?id=CtnWAAAAMAAJ.

Noah D Goodman and Andreas Stuhlmüller. Knowledge and implicature: Modeling language understanding as social cognition. *Topics in cognitive science*, 5(1):173–184, 2013.

Noah D Goodman, Joshua B. Tenenbaum, and The ProbMods Contributors. Probabilistic Models of Cognition. http://probmods.org/v2, 2016. Accessed: 2018-12-1.

H. P. Grice. Logic and conversation. In Peter Cole and Jerry L. Morgan, editors, *Syntax and Semantics: Vol. 3: Speech Acts*, pages 41–58. Academic Press, New York, 1975. URL http://www.ucl.ac.uk/ls/studypacks/Grice-Logic.pdf.

Julia Bell Hirschberg. A theory of scalar implicature (natural languages, pragmatics, inference). 1985.

Laurence Horn. *On the Semantic Properties of Logical Operators in English*. PhD thesis, UCLA, 1972.

Laurence Horn. *A Natural History of Negation*. University of Chicago Press, 1989.

Laurence Horn and Barbara Abbott. <the, a>: (in)definiteness and implicature. In William P. Kabasenche, Michael O'Rourke, and Matthew H. Slater, editors, *Reference and Referring*. MIT Press, Cambridge, MA, 2012.

Roni Katzir. Structurally-defined alternatives. *Linguistics and Philosophy*, 30(6):669–690, Dec 2007. ISSN 1573-0549. doi: 10.1007/s10988-008-9029-y. URL https://doi.org/10.1007/s10988-008-9029-y.

Chigusa Kurumada, Meredith Brown, and Michael Tanenhaus. Pragmatic interpretation of contrastive prosody: It looks like speech adaptation. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 34, 2012.

Chigusa Kurumada, Meredith Brown, Sarah Bibyk, Daniel F Pontillo, and Michael K Tanenhaus. Is it or isn't it: Listeners make rapid use of prosody to infer speaker meanings. *Cognition*, 133(2):335–342, 2014.

Yo Matsumoto. The conversational condition on horn scales. *Linguistics and Philosophy*, 18 (1):21–60, 1995.

Andreea Nicolae and Uli Sauerland. A contest of strength: or versus either–or. In *Proceedings of Sinn und Bedeutung 20 (SuB 20)*, 2015.

Anna Papafragou and Julien Musolino. Scalar implicatures: experiments at the semantics–pragmatics interface. *Cognition*, 86(3):253–282, 2003.

Christopher Potts. Presupposition and implicature. In Shalom Lappin and Chris Fox, editors, *The Handbook of Contemporary Semantic Theory*, pages 168–202. Wiley-Blackwell, 2 edition, 2015. doi: 10.1002/9781118882139.ch6.

Christopher Potts, Daniel Lassiter, Roger Levy, and Michael C Frank. Embedded implicatures as pragmatic inferences under compositional lexical uncertainty. *Journal of Semantics*, 33 (4):755–802, 2016.

Alice Rees and Lewis Bott. The role of alternative salience in the derivation of scalar implicatures. *Cognition*, 176:1–14, 2018.

Uli Sauerland. Scalar implicatures in complex sentences. *Linguistics and philosophy*, 27(3): 367–391, 2004.

Dimitrios Skordos and Anna Papafragou. Children's derivation of scalar implicatures: Alternatives and relevance. *Cognition*, 153:6–18, 2016.

Bob van Tiel, Emiel Van Miltenburg, Natalia Zevakhina, and Bart Geurts. Scalar diversity. *Journal of Semantics*, 33(1):137–175, 2016. doi: 10.1093/jos/ffu017. URL http://dx.doi.org/10.1093/jos/ffu017.