Alternatives, Priming, and Adaptation in Implicature:

An Empirical Investigation of Categorical and Probabilistic Theories of Pragmatic Competence

Brandon Waldon

3/22/2019

Introduction

Successful linguistic communication depends on a listener considering not only what a speaker has said (e.g. "John ate some of the cookies") but also the utterance alternatives a speaker could have produced but chose not to (*John ate all of the cookies*). This process yields inferences beyond the literal meaning of the original utterance (*John ate some, but not all, of the cookies*). Scalar implicature - the type of inference exemplified by the *not all* inference in the preceding discussion - is inextricably linked to the notion of an utterance alternative, as the following theory-neutral definition of scalar implicature demonstrates:

> An utterance U conveys a scalar conversational implicature[1] iff there are alternative utterances U' that are at least as relevant as U in the discourse and that are communicatively stronger[2] 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)

Though the idea that scalar implicature depends on utterance alternatives is uncontroversial, the precise nature of the relationship between alternatives and the inferential mechanisms giving rise to implicature remains hotly contested in the pragmatics literature. By and large, most mainstream theories of implicature are **categorical** theories in two important respects: first, these theories assume that alternative-hood is a binary, categorical feature of any utterance *U'* with respect to another utterance *U* - *all*, for example, is an alternative to *some*; *none* is not. Second, these theories conceptualize scalar inference as the categorical negation of meanings encoded by utterance alternatives, which proceeds either via counterfactual social cognitive inference (c.f. Grice 1975) or a semantic operator (c.f. Chierchia 2006).

---

[1] Or, alternatively phrased, 'A speaker who produces U conveys a scalar conversational implicature'. In what follows, I use the term 'implicature' to denote both a speaker-oriented action of conveying and the listener-oriented action of inferring what the speaker intended to convey.

[2] Note that this definition leaves under-specified 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).

A more recent framework of pragmatic analysis - the Rational Speech Act framework (Frank and Goodman 2012, Goodman and Frank 2016) - abandons both of these categorical commitments in favor of gradient **probabilistic** views, both regarding the nature of alternative-hood and regarding the nature of scalar inference. First, rather than simply associating an utterance U with a set of alternatives, listeners are modeled in standard RSA as reasoning about speakers for whom linguistic behavior - i.e. their selection of an utterance to produce from the set of utterance alternatives - is underlyingly stochastic. In encountering an utterance U, listeners consider that the speaker may have produced alternative U' or U'', but the a priori probabilities of the speaker producing either of these alternatives need not be equal (contra the categorical view). The inference mechanism is also probabilistic on this account: scalar inferences are conceptualized as gradient, negative changes in the a posteriori probabilities of meanings encoded by utterance alternatives.

There are possible hybrids of the above views: for example, one might claim that while the alternative-hood is a binary feature, scalar inference is best represented probabilistically. Conversely, one might claim that scalar inference proceeds via categorical inference mechanisms but at the same time argue for a rich, probabilistic representation of utterance alternatives a la RSA. In this paper, I explore this space of possible analyses and argue that theories which represent alternatives probabilistically allow for the set of utterance alternatives to be large and logically-inconsistent with one another, contra theories which represent alternatives categorically (and regardless of how the inference mechanism is conceived). I argue that this is because the latter type of theory commits itself to a small, logically-consistent set of alternatives in order to solve what has become known in the pragmatics literature as the **symmetry problem**.

This theoretical split is interesting insofar as recent empirical evidence puts pressure on the idea that the set of alternatives is, in fact, small and logically-consistent. In particular, a growing body of psycholinguistic research on scalar implicature priming suggests that the availability and strength of scalar inferences can be modulated by raising the salience of linguistic forms not commonly assumed (on the categorical view) to be in the set of alternatives. After I review this literature and its implications for a theory of alternatives, I motivate and report the results of three experiments designed to further expand this empirical picture. I'll argue in particular from the results of my Experiment 3 that a probabilistic theory of alternatives best captures the full pattern of observations, and that the descriptive power of this theory is complemented by a probabilistic theory of the inference mechanism underlying scalar implicature.

Along the way of this paper, I'll also make the claim that a fully probabilistic theory best allows for the specification of explicit linking hypotheses which represent implicature priming as the behavioral reflex of listeners' expectations of speakers incrementally adapting in an unfolding discourse, where this adaptation is expressed as an attenuation in the strength and availability of scalar inferences. The paper thus contributes to a growing body of literature on the nature of listener adaptation and its role in pragmatic interpretation.[3]

---

[3]Links to all experiments, data, and analyses reported in this paper may be found at https://github.com/bwaldon/alts.

Alternatives and the symmetry problem

As mentioned above, most mainstream theories of implicature are categorical both in their understanding of alternative-hood and of the inference mechanism underlying scalar implicature. For ease of reference in what follows, I call this class of implicature theory the Traditional theory:

> The Traditional theory: The inference mechanism underpinning implicature is the categorical negation of utterance alternatives; alternative-hood is a binary, categorical feature of alternatives.

For example, according to theories which follow the framework of Grice (1975) in representing conversational implicature as the 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 produced but did not. The set of alternative utterances is, itself, unstructured - insofar as it contains all and only the utterances which possess the feature of being an alternative to the produced utterance (and contains no other information to distinguish between alternatives[4]). The following Gricean-inspired analysis of scalar implicature is adapted from Potts (2015):

- 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.[5]
  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 shares commonalities with more recent grammaticist analyses, whereby scalar implicature is associated with an inference triggered by a semantic operator whose input is a proposition p and whose output is a proposition

---

[4](Though, e.g., form-based or context-specific considerations such as manner or discourse relevance may prune or expand the set of alternatives active in pragmatic interpretation.)

[5]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."

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[6]; for example,

(1) $||$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:

(2) $O_C(p) = p \wedge \forall q \in C[q \to p \subseteq_c q]$, where $C =$ the output of $||p||^{ALT}$

Given how we specify the semantic operator O_C in (2) and how we specify the alternatives to *John ate some of the cookies* in (1), the analysis predicts that *John ate some of the cookies* 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:

(3) O_C(*John ate some of the cookies*)
$= \Big[$ *John ate some of the cookies* $\wedge$
$\forall p : p \in ||$John ate some of the cookies$||^{ALT} \wedge p \to$
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, but both assume that alternative-hood is a binary, categorical feature: either *U'* is an alternative to *U*, or it isn't. Both theories claim that a scalar implicature will be conveyed by an utterance in case there are pragmatically-active stronger alternatives to that utterance, but 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):

(4) • 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.

---

[6]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's framework. But crucially, Chierchia leaves under-specified how exactly the alternatives of any item are computed.

- 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.
- $q \rightarrow p_1 \lor p_2$.
- 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: $q \land \neg p_1$ corresponds to the desired 'some but not all' strengthening of q in the context of utterance provided in, e.g., (2), while $q \land \neg p_2$ 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). Second, asserting q cannot implicate both $\neg p_1 \land \neg p_2$ because this conjunction is a contradiction. Avoiding these two issues - illicit inferences and contradictory meanings - is a main desideratum of any theory of scalar inference. That is, a successful theory **breaks** symmetry in a manner which delivers the right predictions regarding the nature of the scalar implicature(s) conveyed by a given utterance.

Categorical theories of scalar inference such as those discussed above - whereby pragmatically-enriched meanings are derived via the categorical negation of communicatively stronger utterance alternatives - thus require an auxiliary theory of utterance alternatives in order to break symmetry. If we furthermore conceive of alternative-hood as a binary feature of the relationship between two utterances (as is the case with <u>Traditional</u> theories), then the theory of alternatives must have something to say about why, e.g., p_1 is in the set of alternatives but p_2 - and perhaps other, equally-relevant utterance alternatives - are not. In other words, one way to break symmetry (given a categorical theory of scalar inference) is to posit that the set of alternatives contains a relatively small number of utterances which are logically consistent with one another, such that negating each alternative does not yield a contradiction, as above. Below, I review attempts in the pragmatics literature to break symmetry in exactly this way.

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 a 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 (5), 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):

(5) 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 (5) is evidence that those two items are, to borrow the language of Horn and Abbott (2012), "natural paradigmatic alternatives" (334). The acceptability of (6) and (7) supports this analysis for *some* with respect to *many* and *all*, while (8) points against a similar analysis for *some* and *some but not all*:

(6) 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

(7) 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

(8)# 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

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 is 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 (9), 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):

(9)   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 built from contrasts between lexical items, 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 distinct 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, engaged, married>, <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 (5).[7] However, there are no logical entailment relations between any of the items in these orderings (e.g. one being in the state of *dying* does not entail that one is or was *sick*). 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 cashes out Horn (1989)'s observation as a more generalized conversational condition on pragmatic competition, as articulated below (from Matsumoto 1995: 25):

---

[7]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.

(10) 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.).[8]

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

(11) 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."

(12) a. "Bill met John on the way home."
         (where the speaker knows that meeting Mary is of concern to the hearer)
      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 (11-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.). (11-d) is not such a candidate, as *exactly N* does not license scalar implicatures. A similar pattern of reasoning holds for (12): *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):

(13) 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.

---

[8]Matsumoto, following Grice, 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.

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 complexity.

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):

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

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

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

According to Katzir (2007), in (14)-(16), 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 prevalence of scalar inferences; for example, the (a) sentences below do not implicate the negation of the respective (b) sentence (examples from Katzir 2007: 684):

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

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

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

In (17) through (19), 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 (17) through (19) motivate a new analysis of alternatives.

Katzir (2007), for his part, takes the data in (14) through (19) to be a challenge for existing scale-based accounts of alternative computation. Instead, he proposes the following principles from which alternatives are calculated:

(20)
- 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 (13): because *a little bit more than warm* is a subtree of (13-a), it is in the substitution source with respect to that sentence. This allows for the construction of (13-b) as an alternative, where *a little bit more than warm* simply replaces *warm* in the first conjunct of the sentence.[9]

Mapping the space of scalar implicature theories

The above review shows that specifying an adequate theory of alternatives has been an active domain of pragmatics research for decades; however, this work has overwhelmingly been done in the Traditional theoretical paradigm. That is, research on alternatives has been driven largely by one understanding of the nature of scalar inference, namely that these inferences come about via negation of stronger alternatives, and by one understanding of the nature of alternatives, namely that alternative-hood is a binary, categorical feature of the relationship between two utterances. Thus, in all of the accounts discussed above, it is assumed that symmetry must be broken by restricting the set of alternatives to which an utterance is contrasted in pragmatic interpretation. As soon as our assumptions regarding the nature

---

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

of the alternatives are relaxed, the need for a small, logically consistent set of alternatives ordered on a scale is obviated, as I demonstrate below.

First, I claim that conceptualizing scalar inference as a (non-categorical) posterior update in the likelihood of meanings expressed by alternatives is, in itself, insufficient for symmetry breaking (given the presence of symmetric alternatives in the alternative set). Using the analytic machinery of the Rational Speech Act (RSA) framework, I explore and compare three classes of theories against <u>Traditional</u> theories to demonstrate that this is the case:

> The <u>Categorical Alternatives, Probabilistic Inference</u> theory: The inference mechanism underpinning implicature is a soft posterior update on the likelihood of world states; alternative-hood is a binary, categorical feature of alternatives.
>
> The <u>Fully Probabilistic</u> theory: The inference mechanism underpinning implicature is a soft posterior update on the likelihood of world states; the representation of utterance alternatives is non-trivially probabilistic[10].
>
> The <u>Probabilistic Alternatives, Categorical Inference</u> theory: The inference mechanism underpinning implicature is the categorical negation of utterance alternatives; the representation of utterance alternatives is non-trivially probabilistic.

While a <u>Categorical Alternatives, Probabilistic Inference</u> theory will fail to break symmetry given symmetric alternatives, a <u>Fully Probabilistic</u> theory will accomplish this given additional assumptions about listener's a priori expectations regarding the speaker's choice of utterance. I then show that, in fact, representing alternatives in this way obviates the need for a probabilistic representation of the scalar inference, if our desire is merely to break symmetry (that is, a <u>Probabilistic Alternatives, Categorical Inference</u> theory will also accomplish this). However, the additional probabilistic machinery of the <u>Fully Probabilistic</u> theory allows for the specification of explicit and quantitative linking hypotheses from pragmatic competence to behavior in the implicature priming paradigms with which the experimental portion of the paper is concerned. The following section recapitulates the main features of the RSA framework before exploring these claims in more depth.

Probabilistic pragmatics and the Rational Speech Act framework

As it is presented by Goodman and Stuhlmüller (2013), the Rational Speech Act model is an iterative, probabilistic model of pragmatic competence in which utterance interpretation is represented as a distribution over 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.

---

[10]"Non-trivially" in the sense that listeners are modeled as having non-uniform prior expectations about the speaker's choice of utterance.

Specifically, in the RSA framework the probability with which a pragmatic listener L_1 concludes meaning s from an utterance u depends upon a prior probability distribution over 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 $P_w(s)$.

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

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.

Exploring an example: exhaustive *some*

The RSA model of pragmatic competence models 'not-all' inferences with *some* as a negative change from the prior to posterior probability of an *all* meaning given an observation of an utterance of *some*. The model first declares a possible set of possible world states. For simplicity, I assume a set of two possible worlds: one in which one there is exactly one entity x, and one in which there are two x entities. We assume that listeners have a uniform prior expectation of being in one of these two world states. [11]

```
var states = [['x'],['x','x']]

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

I first use RSA to explore the implications of a Categorical Alternatives, Probabilistic Inference theory, whereby scalar inference is conceptualized as a soft posterior update on the likelihood of world states but alternative-hood is represented categorically. To accomplish this, 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 *some* or *all*; that is, we have implicitly constrained the model such that the only possible alternative to *some* is *all* (this simplifying

---

[11]The code presented here is excerpted from models I have coded in the probabilistic programming language WebPPL Goodman and Stuhlmüller (2014). The models, linked in footnotes in this section, are executable in the online WebPPL editor: http://webppl.org/.

assumption is made here only to demonstrate the basic mechanism of the RSA model).[12] Furthermore, these two utterances are assumed to have a uniform cost, ensuring that the listener's prior expectation of hearing either of these utterances is uniform (this effectively makes alternative-hood a binary, categorical feature of utterances in this model: either an utterance is an alternative to *some*, and thus it is included in the set `utterances` and assigned a uniform cost of 1, or it is not in the set):

```
var utterances = ["some", "all"];

var cost = {
  "some": 1,
  "all": 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. *Some* assumes the meaning of existential quantification (i.e. it is logically consistent with worlds which contain at least one entity), while *all* assumes the meaning of universal quantification (all of the possible entities in our model must be in the world of evaluation to return `true`):

```
var literalMeanings = {
  all: function(state) { return state.length == 2; },
  some: function(state) { return state.length >= 1; },
};
```

Lastly, models of a pragmatically-competent listener and speaker, as well as of a literal listener are specified as according to the equations in (21). For the pragmatic speaker, we arbitrarily set the rationality term α to 1 for the time being.

The listener is represented as having uniform prior beliefs about the likelihood of being in either the 'some-but-not-all' world (i.e. ['x'] - some but not all of the possible x entities are in the world) or the 'all' world ['x','x']. That is, the prior probability of being in any particular one of these two worlds is roughly 0.5. However, upon observation of the utterance *some*, the listener considers that, although the utterance is truth-conditionally consistent with the ['x','x'] world, an equally costly alternative utterance - *all* - would have uniquely selected this world had this been the speaker's intended meaning. That is, the conditional probability of the speaker producing *all* given an intended meaning of ['x','x'] is higher than the conditional probability of *some* given this same intended meaning.

Because the speaker is assumed to be a (soft) optimal communicator, the listener concludes that since the speaker produced *some*, it is unlikely that the speaker intended to communicate the meaning represented by ['x','x']. The listener's posterior probability of being in the ['x','x']

---

[12]Try this model here: https://bwaldon.github.io/alts/toymodels/type2_ver1.js

thus decreases (on these assumptions, from 0.5 to 0.25). This negative change from prior to posterior expectation of the *all-* world state is how RSA captures the pragmatic exhaustivity inference of *some*.

Note that in the above example, we have stipulated that there is only one alternative to *some* (namely *all*). Introducing a second alternative - the symmetric alternative *some but not all* - leads to the symmetry problem absent additional assumptions about the nature of pragmatic competition between *some* and its alternatives:[13]

```
var utterances = ["some","all","somenotall"];

var cost = {
  "some": 1,
  "all": 1,
  "somenotall" : 1,
};

var literalMeanings = {
  all: function(state) { return state.length == 2; },
  some: function(state) { return state.length >= 1; },
  somenotall: function(state) { return state.length == 1;},
};
```

With the above enrichment to the model, the listener's pragmatic enrichment of *some* is qualitatively different. The prior probability of being in either the ['x'] world or the ['x','x'] world is still 0.5; however, upon observation of the utterance *some*, the listener reasons backward about the speaker's decision not to use either of the two symmetric alternatives. The listener considers that unlike *some*, which is compatible with both world states, both *all* and *some but not all* uniquely select one of these two states. Moreover, the two alternatives are equally costly for the speaker to produce as *some* (and equal in cost to one another). Therefore, the listener does not update her beliefs about the likelihood of either world state: in this model, the listener's posterior probability distribution over world states is uniform, just like the prior distribution.

The Categorical Alternatives, Probabilistic Inference model thus fails to break symmetry given symmetric alternatives in the alternative set: though the resulting inference is not a contradiction (as with categorical models of scalar inference), the model does not predict a decrease in the posterior likelihood of the *all* world state, as we would intuitively expect given observation of an utterance of *some*. This failure illustrates an important fact regarding probabilistic models of pragmatic competence: conceptualizing scalar inference as a decrease in the posterior likelihood of certain meanings (rather than the categorical negation of those meanings) is not sufficient to break symmetry.

The issue lies with the fact that the representation of the set of alternatives requires additional probabilistic enrichment, if we would like to keep symmetric alternatives in the alternative

---

[13]Try this update to the model here: https://bwaldon.github.io/alts/toymodels/type2_ver2.js

set. In the above model, alternatives to *some* are assumed to be equally costly as one another and equal in cost to *some*. Thus, when the listener encounters an utterance of *some*, she reasons that for either possible state of the world the speaker intended to communicate, there was a strictly more optimal utterance (i.e. equal in production cost, greater in communicative strength) that could have been used to uniquely pick out that state of the world. Due to the uniform cost of the alternatives, either alternative was a priori just as likely to be produced as the other, and this uniform prior probability leads to the fact that exposure to an utterance of *some* does not change the listener's information state.

The solution is to change the model to a Fully Probabilistic model[14] by representing the speaker as producing the two alternatives with different a priori probabilities. Because *some but not all* is morphemically more complex than either *some* or *all*, it is reasonable to think that *some but not all* is associated with a higher production cost than either of the two other utterances:

```
var utterances = ["some", "all","somenotall"];

var cost = {
  "some": 1,
  "all": 1,
  "somenotall" : 2,
  };
```

Now, when the listener in the model encounters an utterance of *some*, she considers the fact that neither of the two communicatively-stronger alternatives were produced; however, one of these alternatives (*some but not all*) was a priori relatively unlikely to be produced in the first place, by virtue of the fact that it is relatively costly for the speaker to produce. Thus, *some but not all* is not necessarily a more optimal utterance choice over *some* to convey the exhaustive ['x'] world state, and the probability of the ['x'] world state increases a posteriori - from 0.5 to roughly 0.63. Correspondingly, the probability of the ['x','x'] state decreases - from 0.5 to roughly 0.37. Thus, by enriching the notion of alternative-hood such that some alternatives are a priori more likely than others, we break symmetry.

The above example illustrates that, within a framework that models scalar inference as a probabilistic rather than categorical information update, the set of alternatives may contain logically inconsistent expressions (e.g. *all* and *some but not all*) as long as listeners are modeled as possessing non-uniform a priori expectations about the likelihood with which speakers produce those alternatives. In fact, as I now show, this observation about symmetric alternatives extends to categorical models of scalar inference as well. We continue to utilize WebPPL and the RSA framework to model a listener whose representation of a speaker's possible utterance choices are graded, non-uniform and probabilistic (as in the Fully Probabilistic theory) but whose inferences are effectively categorical, in that posterior probabilities of meanings expressed by alternatives shift to zero rather than undergoing a softer negative posterior shift (as in the Traditional theory explored in the previous section).

---

[14]Viewable here: https://bwaldon.github.io/alts/toymodels/type3.js

16

That is, we change the model to a Probabilistic Alternatives, Categorical Inference model. This is achieved by representing the speaker as being hyper-rational, i.e. by raising the alpha rationality parameter from the Fully Probabilistic model to an exceedingly high value. Given this change to the model, the posterior probability of the ['x','x'] world state given observation of *some* drops to effectively zero (so long as we maintain a cost asymmetry between *all* and *some but not all*). We thus show that symmetry is broken even if we do not conceive of scalar implicature as a softer posterior update on the likelihood of certain world states.

This method of breaking symmetry - compute symmetric alternatives, but introduce an asymmetry in the a priori expectation of either alternative being produced - marks a substantial departure from the categorical notion of alternative-hood, which requires a restrictive theory of alternative computation in order to skirt the symmetry problem. 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).

Moreover, Bergen et al. (2016) also observe, if both alternatives and the scalar inference mechanism are represented probabilistically (as in RSA), we can modulate the predicted strength of the scalar inference by shifting the cost asymmetry between upper-bounding strong alternatives and the more morphosyntactically complex alternatives which themselves encode exhaustive meanings. We set the alpha rationality parameter back to 1 (that is, we return to a Fully Probabilistic model) but change the cost asymmetry between *all* and *some but not all* by subtracting cost from the latter alternative:

```
var utterances = ["some", "all","somenotall"];

var cost = {
  "some": 1,
  "all": 1,
  "somenotall" : 1.5,
};
```

Recall that, upon observation of *some* given the Fully Probabilistic model, the posterior shift in the likelihood of the ['x','x'] state decreases from 0.5 a priori to roughly 0.37 when the cost of *some but not all* is set to 2. As the cost of *some but not all* decreases in the model, the listener's a priori expectation of the speaker producing that alternative increases, thereby attenuating the strength of the exhaustive interpretation of *some*. For illustration, when we decrease the relative cost of *some but not all* from 2 to 1.5, the magnitude of the negative change in the likelihood of the ['x','x'] state becomes smaller - from 0.5 to roughly 0.42 a posteriori.

Of course, when the inference mechanism is represented categorically rather than probabilistically, as in the Probabilistic Alternatives, Categorical Inference model, no such attenuation effect is observed: the posterior likelihood of ['x','x'] world state is 0 given observation of *some*, as long as *all* is represented as less costly than *some but not all*. Because

the output of the Fully Probabilistic model is - conversely - gradient and quantitative, it allows for the specification of a range linking hypotheses from competence to behavior unavailable to Probabilistic Alternatives, Categorical Inference theories. In particular, the Fully Probabilistic model allows us - contra the Probabilistic Alternatives, Categorical Inference model - to explicitly model how repeated exposure to particular alternative forms gradiently and incrementally modulates the interpretation of an utterance over time, by linking this exposure to changes in the listener's representation of the utterance cost function and to subsequent attenuation of the posterior change in the likelihood of particular world states.[15] In the third experiment of this paper, I will compare models which represent the inference mechanism behind scalar implicature either as a gradient, probabilistic update (as in RSA) or as a categorical update (as in standard Gricean and grammaticist theories) to a listener's information state.

However, the first two experiments of this paper will focus on the competing ways to represent of the set of alternatives. Whether or not we commit to a categorical or probabilistic theory of the inference mechanism, we have seen that the theories which represent alternative-hood categorically (i.e. Traditional and Categorical Alternatives, Probabilistic Inference theories) will fail to break symmetry if symmetric alternatives are included in the set of alternative utterances. The first experiment of the paper will investigate whether it is tenable to assume that the set of alternatives to a given utterance is, in fact, small and logically consistent, or whether there is empirical evidence to suggest that utterances compete pragmatically not only with putative scale-mates (e.g. *all* with respect to *some*) but also with utterances (e.g. *some but not all*) which would give rise to the symmetry problem given categorical theories of alternative-hood (but not when the theory of alternatives is enriched probabilistically, as in Fully Probabilistic and Probabilistic Alternatives, Categorical Inference theories).

Experiments 1 and 2 are priming experiments designed to measure whether, and if so how, exposure to linguistic stimuli modulates the subsequent interpretation of other linguistic forms (relative to an experimentally-elicited baseline). Before presenting the design and results of those experiments, I situate the motivation of those experiments in the context of the existing experimental pragmatics literature on priming as it relates to scalar implicature.

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 interpretations of weak scalar items are not

---

[15]One might imagine a modification of the Probabilistic Alternatives, Categorical Inference model whereby listeners' computation of the implicature is stochastic, i.e. subject to some noise parameter. Note that this modification predicts a bimodal distribution in listener inferences, with some set of listeners inferring the negation of stronger alternatives while others exhibit no such posterior update. That is, the 'attenuation' effect modelable from a Fully Probabilistic theory would still not be modelable given this type of change to the Probabilistic Alternatives, Categorical Inference model. In the discussion of Experiments 1 and 2, I consider further possible modifications to non-Fully Probabilistic theories and their ability to capture adaptation effects in implicature priming paradigms.

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 implicature inferences can (or cannot) be activated has informed debates regarding the mechanisms underlying pragmatic competence, including the types of linguistic forms that compete pragmatically with one another in interpretation. In an early 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 subsequently affect how *some* is processed and interpreted. The authors take these findings as evidence that the alternative set for utterances 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 problematizes the categorical view of alternative-hood, which (as discussed above) is committed to the idea that the set of utterance alternatives is relatively small and semantically-consistent. To account for the data presented by Degen and Tanenhaus (2015) and Nicolae and Sauerland (2015), categorists must provide additional stipulations regarding the contextual availability of alternatives outside this constrained set (as well as the ability of these context-dependent alternatives to displace the putative conventional ones), or they must explain away these particular data as reflecting idiosyncratic features of the lexicon. If it can be shown across a range of putative scales that raising the salience of the semantically upper-bounded, symmetric alternatives (e.g *some but not all* with respect to *some*) can modulate interpretation of weak scalar items, then we raise a complication for analyses which breaking symmetry via constraints on alternatives (absent ad-hoc stipulations for these analyses).

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 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, namely that the symmetric alternative appears to compete pragmatically with the weak scalar form and thus leads to less exhaustive interpretations of that form (i.e. raising salience of the alternative "blocks" the association of the weak scalar item with the meaning expressed by the alternative - call this the *meaning blocking hypothesis*).

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 simply "primes" listeners to associate the upper-bounded meaning with the weak scalar form (call this the *meaning priming hypothesis*).

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. Recall that Fully Probabilistic theories, unlike the other three theory types explored above, allows for the specification of quantitative linking functions between a.) repeated exposure to linguistic form U' (which competes pragmatically with another form U) and b.) modulated subsequent interpretation of linguistic form U, conceptualized as a listener updating her beliefs about the relative speaker cost of producing U'. Thus, empirical evidence that this type of exposure does indeed yield priming behavior would provide indirect support for a Fully Probabilistic theory, which is uniquely capable of capturing this phenomenon quantitatively. Experiments 1 and 2 of this paper are designed to address this and the above questions.

Experiments 1 and 2

Experiments 1 and 2 investigate the ability of semantically upper bound-encoding forms (putative strong scalar alternatives) and semantically upper-bounded exhaustive 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.[16]

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 inter-

---

[16]As will be shown below, the particular nature of the anticipated divergence will depend on the forms to which participants are exposed. In the case of semantically upper bound-encoding forms, we anticipate weak scalar items to not convey an upper bound-encoding meaning, i.e. to be interpreted more exhaustively. The anticipated pattern is the opposite in the case of the semantically upper-bounded forms: we anticipate weak scalar items to not convey the upper-bounded meaning, i.e. to be interpreted less exhaustively.

preted 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 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 interpretation of weak scalar items, in contrast to the vast majority of implicature priming studies conducted to date, in which participants are trained to associate forms with particular meanings before critical trials[17].

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 control trials of the experiment in that the bolded sentence was also colored red, as can be seen in Figures 1 and 2, in order to emphasize to participants that linguistic material in the critical trial differed from the material seen in the previous three priming trials.

---

[17]Experiments 1 and 2 were preregistered through the Open Science Framework at https://osf.io/7u8zd/ and https://osf.io/pv3mb/, respectively. A stable version of the experiment program for Experiments 1 and 2 is viewable at https://bwaldon.github.io/alts/experiments/Experiments1-2/index.html. Note that Experiments 1 and 2 were deployed simultaneously in the same experiment script.
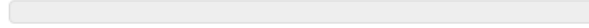
Henry: "Sally saw all of her former classmates at her high school reunion."

How likely is it that Henry meant to suggest:

Sally didn't see every former classmate of hers at the reunion.

very unlikely   very likely

[Continue]

Figure 1: A sample priming trial in Experiment 1

Henry: "Marc and Melissa sent some of their children to private school."

How likely is it that Henry meant to suggest:

Marc and Melissa didn't send every child of theirs to private school.

very unlikely   very likely

[Continue]

Figure 2: A sample priming trial in Experiments 1 and 2

As Figures 1 and 2 illustrate, the target item on priming trials (in the Prime condition of the experiment) contained a strong scalar construction (e.g. *all*). Participants in the No-prime condition of the experiment did not see a scalar item on priming trials. In both conditions, however, three 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 Prime condition. Moreover, in both conditions, participants saw the same interpretation sentence on both priming trials and critical trials, which always conveyed an <u>exhaustive</u>, upper-bounded meaning from the scale of interest (e.g. *not every* for the <all, some> scale, *not both* on the <and, or> scale). The interpretation question never used the same lexical material from priming trials of the Prime condition; for example, in the *some* block, Prime condition trials featured the strong alternative *all*, but the corresponding interpretation question on these trials asked about the likelihood of an intended "not every" meaning rather than a "not all" meaning. The concern was that if the interpretation question was also phrased as, e.g., "not all", participants' responses on these trials would be driven primarily by the presence of negation in the interpretation question rather than on the meaning of the lexical material in the bolded sentence.

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 control item), constituted one block of trials. A set of priming and critical trials from one block - the *or* block in the No-prime condition - are shown in Figure 3:

Figure 3: A block of trials from the No-prime condition

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>, <delicious, palatable>, <impossible, hard>, and <is, looks like>. The scales were selected to span a range of pragmatic contrasts, from scales commonly treated as lexicalized (i.e. <all, some> as well as <and, or>) to scales commonly treated as more context-dependent (i.e. <delicious, palatable> and <impossible, hard>) to a scale commonly treated as being a completely ad-hoc contextual construction (i.e. <is, looks like>). There were two lists of trials, such that each participant saw three No-prime condition blocks and three Prime 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, Prime 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'

- Stronger alternative: 'and'

24

- Interpretation question: "only one of these things. . . "
- Example sentence, Prime 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, Prime 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, Prime 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, Prime 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, Prime 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 control trial. 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 dispreferred. Divergent behavior on more than one of these trials (divergent in the sense that the participant's response was on the opposite side of the scale as expected) 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 32 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 Prime condition are rated more exhaustively (i.e. higher on the sliding scale) than in the No-prime 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 Prime condition are rated less exhaustively 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.
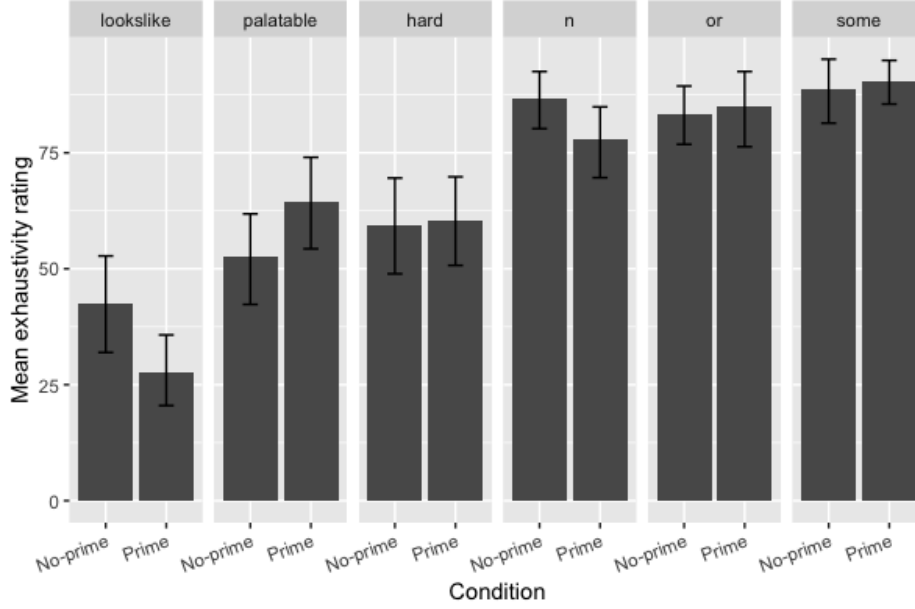
Figure 4: Mean interpretation ratings by condition (Experiment 1)

Figure 4 displays the by-condition mean interpretation ratings on critical trials, broken down by scale type. A visual inspection of the results suggests that behavior between conditions is highly similar, 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 by-participant random intercept and by-participant random slope for condition.[18] The binary condition variable was centered (with reference level No-prime), and the reference level for the fixed effect of scale was the <all, some> scale - the scale with the highest overall mean interpretation rating.

The results of this analysis did not yield a significant effect of condition ($\beta = 1.64$, SE = 6.13, t = 0.268, p = 0.79[19]) but yielded significant negative effects of scale type (p < 0.01) for all pairwise comparisons to the <all, some> scale except in the case of the <and, or> scale (β = -5.64, SE = 4.14, t = -1.363, p = 0.75) and the cardinal number scale (β = -7.41, SE = 3.96, t = -1.873, p = 0.42), where the effects were insignificant.[20] Lastly, the interaction of condition and scale type in the linear mixed effects model specified above was not significant in any case except the <looks like, is> scale, for which there was a marginally significant negative effect (β = -16.35, SE = 9.04, t = -1.809, p < 0.08).

---

[18]There were no by-item random effects, because every scale was associated with just one critical item.

[19]P values were obtained using the 'lmerTest' package in R.

[20]Pairwise comparisons were conducted by employing Tukey's method (allowing for the same by-participant variability as in the linear mixed effect model described above), using the 'lsmeans' package in R.
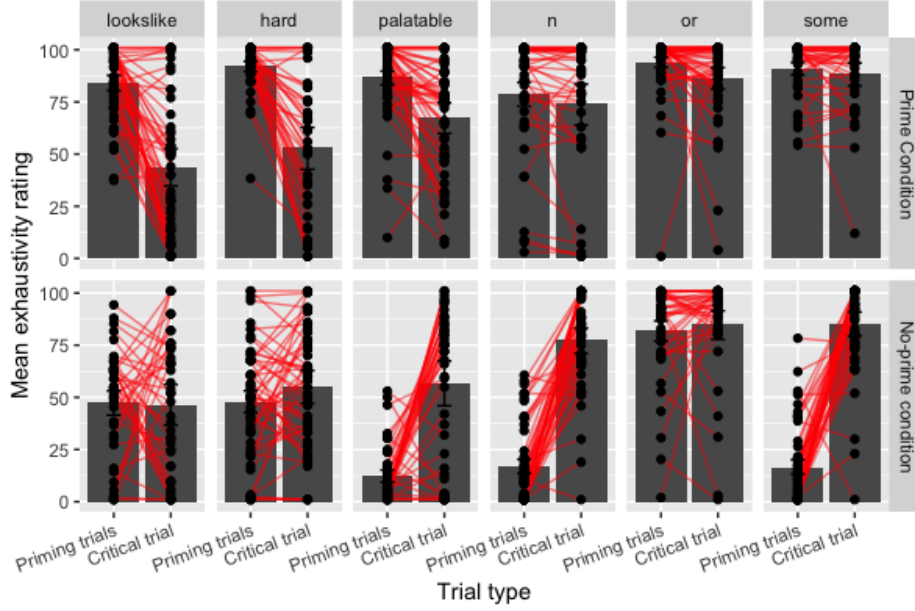
Figure 5: By-subject variation in behavior (Experiment 1)

In order to see whether individual participants in general made a distinction between priming and critical trials in the experiment, Figure 5 illustrates the by-subject variation 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 numerical response on the three 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 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.
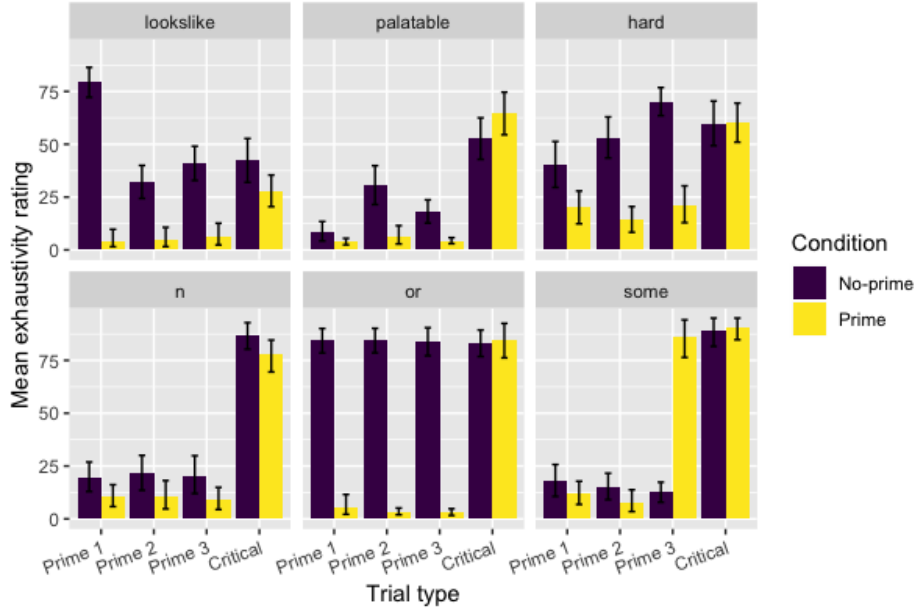
Figure 6: Sequential change in response (Experiment 1)

Lastly, in order to see whether there were notable patterns in behavior as participants complete the experiment, Figure 6 illustrates how mean response 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 did not reflect results reported in previous research 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 (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 an X* in the Prime condition marginally decreased exhaustivity ratings of *looks like an X* relative to the No-prime condition, there is marginal support for the meaning priming hypothesis in this instance. Of course, the observance of

null effects in this experiment can provide no evidence for or against null hypotheses; at best, one can speculate that raising salience of putative alternative forms through exposure alone (and without additional form-meaning association tasks) could not - at least in this instance - modulate subsequent interpretation of lower-bounded weak scalar items.

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 (e.g. *some but not all*) modulates interpretation of weak scalar items (e.g. *some*). Recall that this question is particularly of interest given the space of theories explored above because empirical evidence that upper-bounded forms compete pragmatically with weak scalar items problematizes implicature theories according to which alternative-hood is represented categorically.

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 control trials.

'Some'

- Exhaustive alternative: 'some but not all'
- Interpretation question: "not every. . . "
- Example sentence, Prime 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, Prime 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, Prime 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, Prime 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, Prime 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, Prime 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 Prime condition are rated less exhaustively than in the No-prime 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 Prime condition are rated more exhaustively than in the No-prime 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, eleven participants were excluded for failing more than one of the exclusion trials described above, meaning that the subsequent analyses include data from 94 participants.
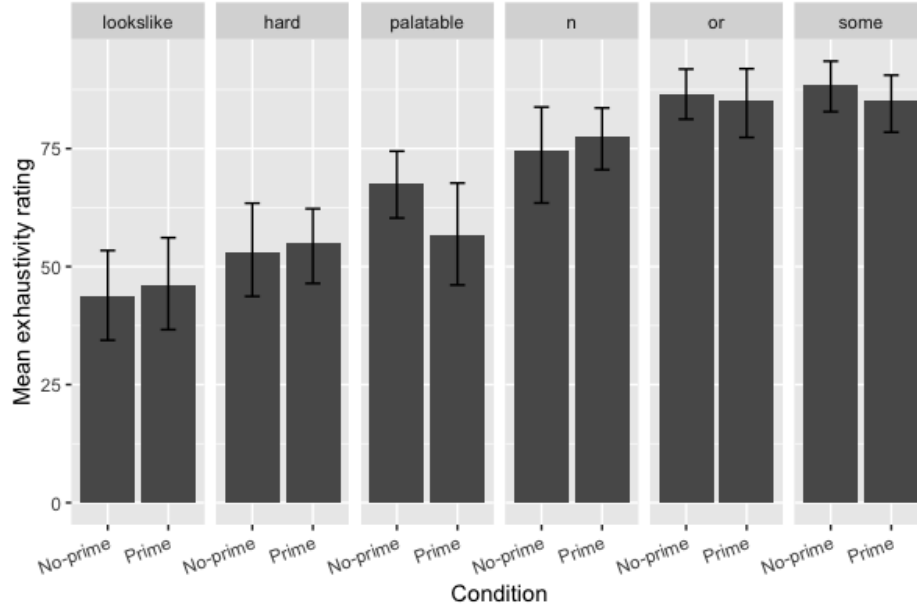
Figure 7: Mean interpretation ratings by condition (Experiment 1)

Figure 7 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 similar. A linear mixed effects regression analysis (identical in structure to the analysis performed for Experiment 1) did not yield a significant effect of condition (β = 3.45, SE = 5.76, t = 0.60, p = 0.55). Pairwise comparisons between <all, some> and the various other scale types were significant (p < 0.01) except in the case of comparison to the <and, or> scale (β = -0.95, SE = 3.82, t = -0.25, p = 0.99) and comparison to the cardinal number scale, which was significant to a lesser degree (β = -10.88, SE = 3.68, t = -2.95, p < 0.05). The interaction of condition and scale type did not approach significance for any scale.
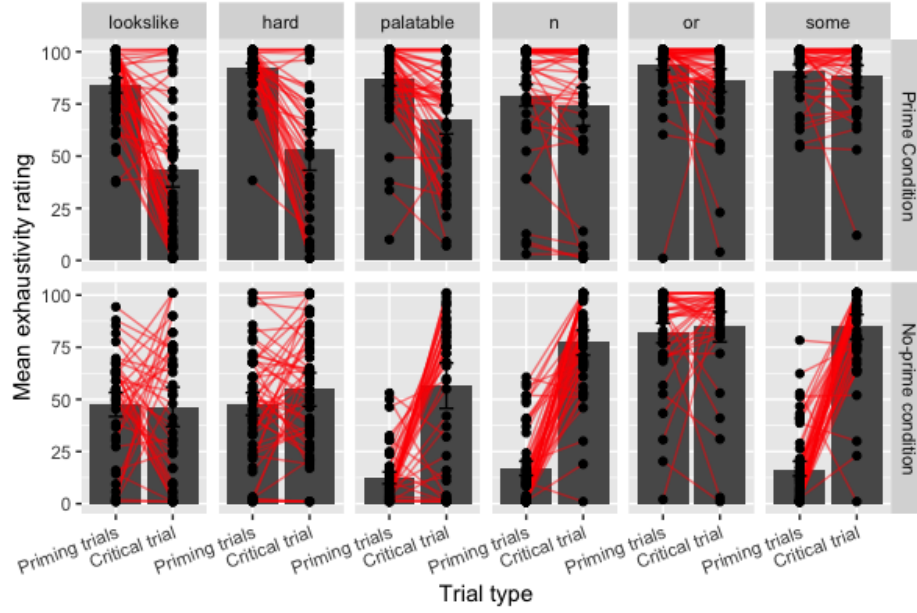
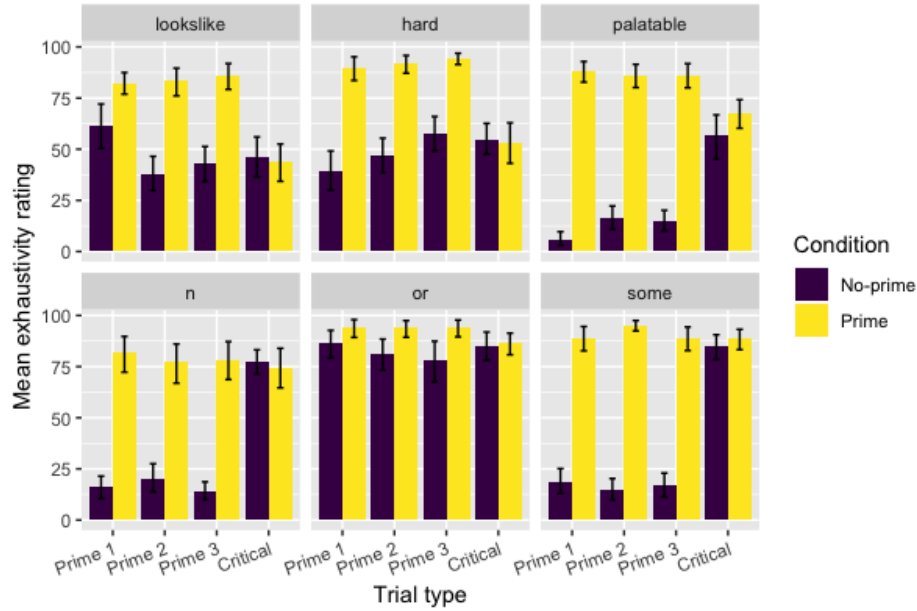Figure 8: By-subject variation in behavior (Experiment 2)

Figure 9: Sequential change in response (Experiment 2)

Figure 8 shows the by-subject change in behavior in Experiment 2, while Figure 9 illustrates how mean response changes sequentially within a given block in the experiment. The figures reveal patterns in response qualitatively similar to those observed in Experiment 1, with the obvious exception that on priming trials in the Prime condition of Experiment 2, overall exhaustivity ratings are quite high.
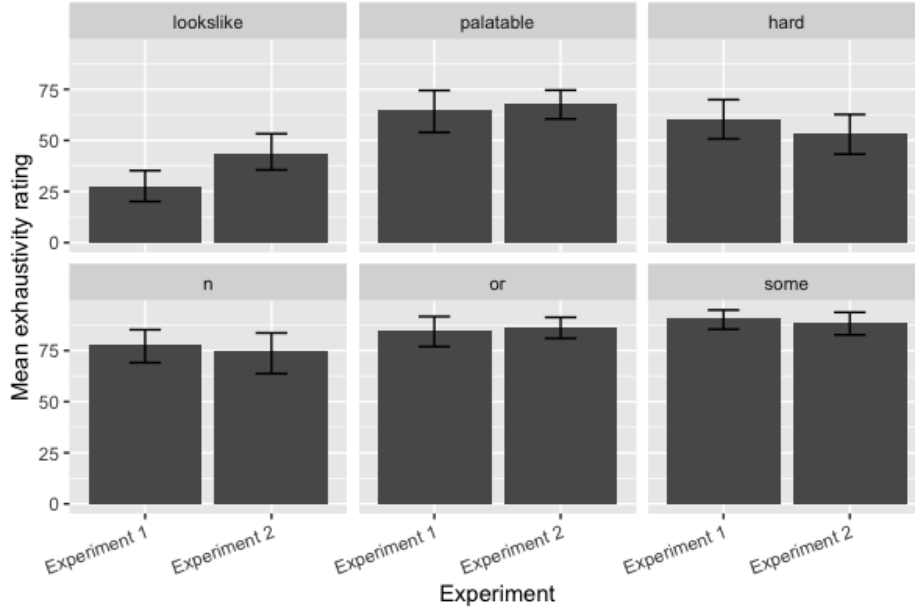
Figure 10: Comparison of behavior on critical trials in Prime conditions (Experiments 1 and 2)

Lastly, Figure 10 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 intercept. The reference level for the experiment variable - which was centered - was Experiment 1, while the reference level for the fixed effect of scale was the <all, some> scale. The results of this analysis did not yield a significant effect of experiment (β = -1.89, SE = 5.95, t = -0.318, p = 0.75). 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 (β = 18.051, SE = 8.30, t = 2.174, p = < 0.04).

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, comparison of critical trial behavior in the *looks like* block between Experiments 1 and 2 did reveal a significant 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 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 <u>being</u> an X and <u>looking</u> like an X.

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 these 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 alternatives who rely on highly-constrained theories of alternative computation in order to avoid the symmetry problem (i.e. proponents of Traditional and Categorical Alternatives, Probabilistic Inference theories) might 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 conventionalized scale in the same way the <some, all> scale is.

Proponents of probabilistic theories of alternatives, including Fully Probabilistic and Probabilistic Alternatives, Categorical Inference theories, may conversely 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>.[21] It is possible that Experiments 1 and 2 yielded ceiling effects in priming behavior with lexical items for which pragmatic contrasts with communicatively stronger alternatives are already sufficiently salient to give rise to exhaustive interpretations. Where pragmatic competition between weak and strong forms on a putative scale is not particularly robust a priori, we might conversely predict the paradigm explored here to be more sensitive, by activating and/or strengthening pragmatic contrasts not already salient to the listener.

---

[21]Van Tiel et al. specifically find that *some* is interpreted more exhaustively than *hard*, which in turn is interpreted slightly more exhaustively than *palatable* - a result replicated in Experiment 1 and partially replicated in Experiment 2 (where the slight asymmetry in exhaustivity between *hard* and *palatable* reverses).

It must also 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.

Revisiting categorical vs. probabilistic theories of alternatives

The point of departure for Experiments 1 and 2 was the discussion of categorical vs. probabilistic theories of pragmatic alternatives in the preceding section of the paper. In that discussion, I claim that probabilistic theories are uniquely positioned to break symmetry given symmetric alternatives in the alternative set, and that evidence of weak scalar items competing with symmetric alternatives could offer initial empirical support for representing the set of alternatives as being logically-inconsistent. A major wrinkle in this theoretical discussion is that symmetry can be broken while preserving the categorical notion of alternative-hood, provided we stipulate that listeners may have uncertainty concerning the set of alternatives about which she should reason. That is, we can build a model with two (logically-consistent) sets of alternatives - `['some','all']` and `['some','somenotall']`, where we assume that the latter set of alternatives is a priori less likely to be the set of alternatives from which we sample in computing the `utterancePrior`. In fact, this enrichment obviates the need to posit a difference in cost between *some* and *some but not all* if our goal is simply to break symmetry:

```
var cost = {
  "some": 1,
  "all": 1,
  "somenotall" : 1,
  };

var utterancePrior = function() {
  var utterances = flip(0.75) ? ["some","all"] : ["some","somenotall"]
  var uttProbs = map(function(u) {return Math.exp(-cost[u]) }, utterances);
  return categorical(uttProbs, utterances);
  };
```

Moreover, this model allows us to represent incremental adaptation cashed out as incremental changes to the prior over possible alternative sets: as the argument to the above `flip()` function approaches 1 (that is, as the `['some','somenotall']` alternative set becomes less likely a priori), we predict more exhaustive interpretations of *some*.

On one hand, given a Gricean understanding of pragmatic reasoning, positing the existence of this kind of higher-order uncertainty is not all that stipulative: we standardly assume that listeners must decide which set of alternatives are relevant for the purposes of a given linguistic exchange with an interlocutor, and it is reasonable to think that this process of resolving pragmatic relevance is not fully deterministic. A possible conceptual objection to introducing an asymmetric 'relevance' prior over possible alternative sets in this particular way can be

found in von Fintel (2002) [http://web.mit.edu/24.954/www/files/24.954.lecturenotes.pdf], who 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 intuitively to be just as relevant as *some but not all of the boys are at the party.* But the fact remains that it is - at least computationally - possible to build a model with the same qualitative features of the *Fully Probabilistic* model whereby alternative-hood is still a fundamentally categorical feature of the model.

I assume for what precedes and follows in this paper that this kind of higher-order uncertainty about linguistic alternatives is unavailable in the space of possible theories I explore; however, I concede that future work will have to engage seriously with this possibility.

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 effects of priming were not detected in this paradigm (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 speculation (unfortunately based on a null effect rather than positive empirical support) that for two linguistic forms already exhibiting a very high degree of pragmatic competition with one another, merely raising the salience of one through explicit mention has little 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 <u>Traditional</u> and <u>Categorical Alternatives, Probabilistic Inference</u> theories of scalar implicature (those with represent the set of alternatives categorically) with respect to the type of inferences derived when symmetric utterance alternatives are made contextually salient.

According to <u>Traditional</u> theories (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 this claim comes from sentences such as (22), 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):

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

On Katzir's analysis, (22) activates the following symmetric alternatives, and it is precisely the cancellation of these symmetric alternatives in tandem which gives rise to the uncertainty inference:

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

(24) 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 (24) and (24) as alternatives to (22) in the following way: first, (23) is activated as an alternative to (22) because it is structurally simpler than (and as relevant as) (22). Furthermore, (24) is an alternative to (22) by virtue of the same structural-complexity principle (because *some but not all* is a subtree of 22, it may substitute *some* in the first conjunct of 22 to yield the alternative 24). Because Katzir's account specifically derives symmetric alternatives in this instance, the analysis must explain why (22) does not have the feel of contradiction (as we might expect if the pragmatically-enriched meaning of 22 included the negation of both 23 and 24). 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). For the purposes of the following discussion, call this hypothesis the *ignorance hypothesis.*

The ignorance hypothesis 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, according to the authors, the resulting interpretation of *or* after this displacement is merely that the speaker is unsure as to whether 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.

Recall from the WebPPL-based analysis of Categorical Alternatives, Probabilistic Inference theories that this class of theory also gives way to the ignorance hypothesis: given symmetric alternatives in the alternative set, the posterior distribution over world states given observation of the utterance of interest will be identical to the prior distribution (all else equal). However, as noted earlier, gradient, probabilistic theories of alternative-hood allow for more nuanced predictions regarding the inferences listeners draw when symmetric alternatives are activated. For example, the Fully Probabilistic analysis of *some* explored

above demonstrated that activating - and raising the contextual salience of - the symmetric alternative *some but not all* 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 with Fully Probabilistic theories (as well as Probabilistic Alternatives, Categorical Inference theories), the presence of symmetric alternatives does not necessarily inhibit pragmatic inference in the form of posterior belief update of the likelihood of world states consistent with those stronger alternatives. For the purposes of the following discussion, call this competing the *modified inference* hypothesis.

The experiment reported below explores the ignorance and modified-inference hypotheses 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 interpretation 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* x̲) and prevalence of alternative linguistic forms in the experimental context (e.g. *It's an X*) (Kurumada et al. 2012, Kurumada et al. (2014)). 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, the authors 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 paper (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 modifies and extends the forced-choice task of Kurumada et al. (2012) whereby participants choose between two images - a target and competitor image - given an utterance

of 'It looks like a [target]' produced by a speaker (their Experiment 2). A sample trial of the experiment is shown in Figure 11 below.
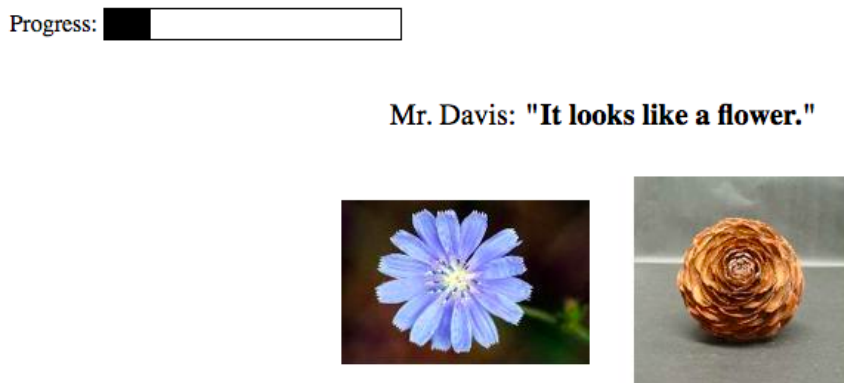

Figure 11: A critical trial from Experiment 3

The crucial extension of the present study is the inclusion of a condition in which two symmetric, communicatively stronger expressions ('It's a [target]' and 'It isn't a [target]') are made salient on non-critical trials of the experiment. It investigates two competing hypotheses of how activating symmetric alternatives modulates pragmatic interpretation of utterances, building off the discussion from the previous section.

The *ignorance* hypothesis: activation of symmetric alternatives to a given utterance yields the inference that the speaker is ignorant as to the truth/falsity of stronger meanings. In the context of this experiment, the hypothesis is that listeners draw an ignorance inference (i.e. fail to resolve whether the speaker intended the target or competitor image) given observation of 'It looks like a [target]' in the condition of the experiment where symmetric alternatives 'It's a [target]' and 'It isn't a [target]' are made salient through explicit mention.

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. In the context of this experiment, the hypothesis is that listeners do not draw an ignorance inference in the condition where symmetric alternatives 'It's a [target]' and 'It isn't a [target]' are activated: listeners resolve the speaker's intended meaning, but the activation of symmetric alternatives modulates listeners' interpretation relative to situations where only one (or neither) of these alternatives are made contextually salient through explicit mention.

Methods

Whereas in Kurumada et al. (2012), 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[22].

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 high-frequency nouns capable of being represented visually (the same nouns used by Kurumada et al.) were embedded into the following sentence frames: *It's 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 a [target]'; in the **target condition**, 8 of the 24 items in the control condition were replaced by sentences of the form 'It's a [target]'; in the **not-target condition**, 8 of the 24 items in the control condition were replaced by sentences of the form 'It isn't a [target]'; in the **symmetric condition**, 4 of the 24 items were of the form 'It's a [target]' while another 4 of the 24 were of the form 'It isn't a [target]'. (The original Kurumada et al. study contained only the first two conditions in order to investigate the effect of 'It's a [target]' on interpretation of 'It looks like a [target]'). 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 (i.e. selecting the wrong image) 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

---

[22]Experiment 3 was preregistered through the Open Science Framework at https://osf.io/64esy/. A stable version of the experiment program for Experiment 3 is viewable at https://bwaldon.github.io/alts/experiments/Experiment3/main/index.html.

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.

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 a [target]' trials). In the target condition, the pragmatic alternative 'It's a [target]' 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's a [target]'. As a result, 'It looks like a [target]' 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 a [target]' as 'It isn't a [target]' 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 a [target]'. As a result, 'It looks like a [target]' 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 under-specified 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 `rsa` 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 dis-

tribution 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 a priori nameability of the images on the screen of every trial. That is, the `rsa` model is a Fully Probabilistic model, in that it represents both scalar inference and the alternative set probabilistically.[23]

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 a speaker to produce. For the `rsa` 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 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 a [target]'.

Norming studies

No-prompt norming study

The `ignorance-2` and `rsa` models specified above make reference to participants' priori expectations of the speaker's intended referent. In one norming study, the *no-prompt* 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 linguistic 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 12 below.[24]

---

[23]Nameability data is elicited in a norming study reported below.

[24]Additionally, a stable version of the experiment program for the no-prompt norming study can be viewed at https://bwaldon.github.io/alts/experiments/Experiment3/noprompt/index.html.

Progress:

Mr. Davis: *mumble mumble*

What was Mr. Davis looking at?

Figure 12: A sample trial from the no-prompt norming study (Experiment 3)

Nameability norming study

Additionally, the `rsa` 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 13[25].

---

[25] A stable version of the experiment program for the no-prompt norming study can be viewed at https://bwaldon.github.io/alts/experiments/Experiment3/naming/index.html.

Figure 13: A sample trial from the nameability norming study (Experiment 3)

Results (main study)

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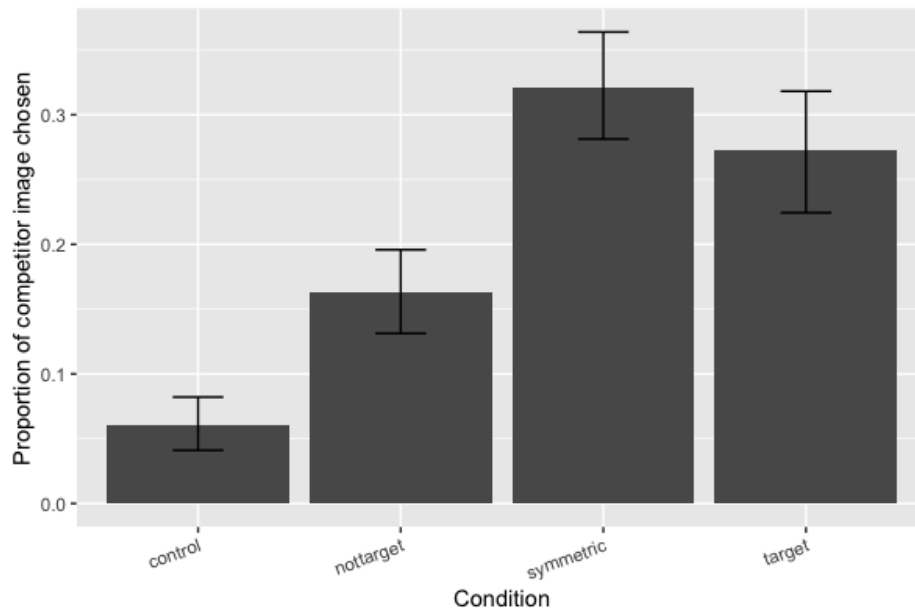

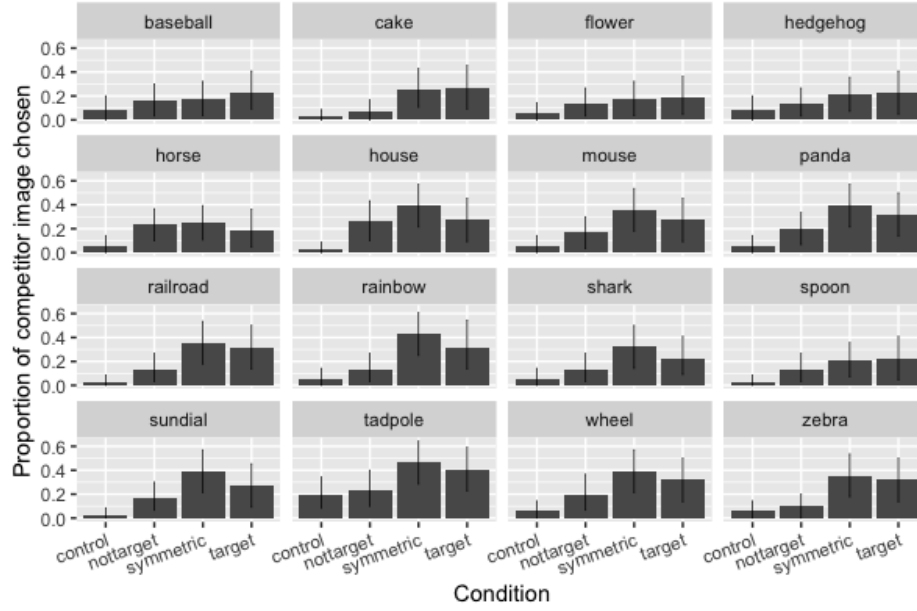Figure 14: Proportion of competitor image chosen on 'looks like' trials (Experiment 3)

Figure 15: Proportion of competitor image chosen on 'looks like' trials (By item, Experiment 3)

Figure 14 displays the mean proportion of competitor images selected on the 16 *looks like* trials shared across the four conditions of the experiment, while Figure 15 shows mean behavior across conditions broken down by item. A visual inspection of Figure 15 suggests that the pattern of results seen in Figure 14 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, a random intercept for item, 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, z ratio = 3.19, P = 0.008) and between the symmetric and control conditions ($\beta = 4.07$, SE = 1.18, z ratio = 3.44, P = 0.003), and the comparison approached significance between the not-target condition and control condition ($\beta = 2.89$, SE = 1.15, z ratio = 2.51, 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, z value -1.734, 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

48

- 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 relative to the control condition (as in the target condition - though to a lesser degree). 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 `rsa` 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 `rsa` models. The analysis assumes a uniform prior over hypotheses. Prior expectation of selections of target/competitor image were computed for the `ignorance-2` and `rsa` 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 `rsa` model.

The `rsa` model assumes the following set of possible sentential frames: {*It looks like an X/It is an X/It isn't an X*}, along with a bounded uniform distribution over possible cost values for each of these frames. For every item in the experiment, the cost of {[*It looks like/It is/It isn't*] a [target]} is assumed to be proportional to the cost of the sentential frame and inversely proportional 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 assumes a bounded uniform distribution over possible values of the alpha rationality parameter.

```
var cost_lookslike = uniformDrift({a: 0, b: 5, w: 1})
var cost_not= uniformDrift({a: 0, b: 5, w: 1})
var cost_is = uniformDrift({a: 0, b: 5, w: 1})

var cost = {
  "istarget": cost_is / d.target_nameability,
```

```
    "nottarget" : cost_not/ d.target_nameability,
    "looksliketarget": cost_lookslike / d.target_nameability,
    "iscompetitor" : cost_is / d.competitor_nameability,
    "notcompetitor" : cost_not/ d.competitor_nameability,
    "lookslikecompetitor": cost_lookslike / d.competitor_nameability,

  var alpha = uniformDrift({a: 0, b: 5, w: 1})


  };
```

The Bayesian data analysis employs a Markov Chain Monte Carlo algorithm in WebPPL (2000 samples, burn-in rate of 20,000) to infer the relative posterior likelihood of the three models by comparing, for every *looks like* trial in the symmetric condition, each model's predicted proportion of selections of the competitor image against the observed proportion in the data set[26] 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 `rsa` 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 `rsa` 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 `rsa` model that has been parameterized with the maximum a posteriori values of the joint parameter distribution[27].

Exploratory data analysis: comparing different instantiations of the RSA model

Though the comparison between the `ignorance1`, `ignorance2`, and `rsa` models assigns zero posterior likelihood to the `rsa` model, a caveat of this approach is that the `rsa` model will be dis-preferred to the ignorance models due to Bayesian Occam's Razor considerations, namely the fact that the `rsa` model is vastly more complex than either of the ignorance models (for more discussion of this point, c.f. Goodman et al. (2016)). Against the original `rsa` model, five modifications were compared which varied along these two dimensions, as well as according to whether or not item nameability was factored into the computation of utterance cost.

First, the assumption of 6 utterance alternatives was relaxed, to allow for the exploration of models in which there are only 3 alternatives, namely {[It looks like/It is/It isn't] a [target]}.

---

[26]The model comparison is viewable here: https://bwaldon.github.io/alts/results/Experiment3/model_comparison.js. This and the other model comparison scripts reported for Experiment 3 are executable in the WebPPL editor.

[27]These two modifications to the Bayesian data analysis may be seen at: https://bwaldon.github.io/alts/results/Experiment3/model_comparison2.js.

Second, in 6-alternative models, I allowed for a unique cost parameter for each of these alternatives, rather than three cost parameters (as in the original `rsa` model). Lastly, I allowed for models in which the cost of an alternative is not sensitive to the a priori nameability of items in the experiment.

The full space of model comparison is schematized below (note that `6alt_3cost_N` corresponds to the original `rsa` model):

Model name	Number of alternatives	Number of cost parameters	Nameability?
3alt_3cost_N	3	3	Yes
3alt_3cost_noN	3	3	No
6alt_3cost_N	6	3	Yes
6alt_3cost_noN	6	3	No
6alt_6cost_N	6	6	Yes
6alt_6cost_noN	6	6	No

The posterior predictive checks of these various models are compared in the figure displayed in Appendix 1, with error bars denoting 97.5% confidence intervals. In terms of correlation of predictions to observations, the best models were the `3alt_3cost_N` models and `6alt_6cost_N` models (highlighted in red), with the `3alt_3cost_N` model categorically preferred over the `6alt_6cost_N` model in a subsequent model comparison (posterior likelihood of the `3alt_3cost_N` model is virtually 1, compared to virtually zero probability for the `6alt_6cost_N` model).

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. Conversely 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 separately, utterance costs from the `3alt_3cost_N` model were computed in WebPPL from uniform prior distributions for each cost parameter along with the maximum a posteriori alpha parameter value inferred from the entire data set. The figures in Appendix 2 compare the inferred cost values of utterances in the `3alt_3cost_N` model for the four experimental conditions.

While the analysis assumes that the absolute cost of 'It's a [target]' stays constant across

the four conditions, the analysis importantly predicts shifts in the cost of 'It's not a [target]' and 'It looks like a [target]' relative to 'It's a [target]', in line with the overall changes in participant behavior observed in the experiment.

First, between the control condition and target condition, we observe a decrease in the relative cost of 'It's a [target]' with respect to 'It's not a [target]'. Recall that in the target condition, participants see several exemplars of 'It's a [target]'. The model explains the relative increase in selections of the competitor image on 'looks like' trials (in the target condition) via participants' relatively higher prior expectation of observing the sentence 'It's a [target]', an utterance that uniquely identifies the target referent. Moreover, increases in proportion of competitor images selected on 'looks like' trials are similarly observed in the not-target and symmetric conditions of the experiment. Again, the model illustrates these increases as (downward) shifts in the relative cost of 'It's a [target]' compared to 'It's not a [target]', though to varying degrees (reflecting the fact that the shifts in competitor selection were non-uniform across these conditions).

Note that while the overall proportion of competitor image selections is highly similar in the target, not-target, and symmetric conditions, this analysis suggests that these similar behavioral patterns are in fact the product of differently-parameterized cognitive representations, with listeners updating their expectations of speaker behavior differently according to experimental condition. However, future work should investigate whether these behavioral patterns, and the inferred shifts in parameter values seen here across conditions, replicate with different linguistic stimuli (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 from Kurumada et al. (2012) on the context sensitivity of interpretation of *It looks like an X*. The experiment replicated a main effect of that study, namely that exposure to *It's an X* forms causes *It looks like an X* to more frequently convey a 'not an X' meaning. In the analysis of Experiment 3, I proposed a Rational Speech Act model of this phenomenon and showed by way of Bayesian data analysis that the data from the experiment are consistent with the idea that listeners adapt their interpretation of *It looks like an X* in the experiment by updating their mental representation of utterance costs (specifically, by virtue of the fact that 'It's a [target]' is associated with lower production cost in the target condition than in the control condition).

One main prediction for Experiment 3 is not borne out: there was no evidence that participants in the not-target condition interpreted *It looks like an X* more frequently as *It's an X* than in the control condition. The Bayesian data analysis nonetheless suggests that in this condition, listeners update their expectations about speaker utterance production relative to the control condition - as is the case in the target and symmetric conditions.

The model comparisons between the `rsa` model and two ignorance models vastly prefer the simplest account of behavior in the symmetric condition, namely that participants answer

at chance on *looks like* trials. This is consistent with the idea that, upon exposure to both 'It's a [target]' and 'It's not a [target]' forms, participants consider to the linguistic signal of 'It looks like a [target]' to be insufficient to resolve the speaker's intended referent: if the speaker has two simpler, communicatively stronger utterances at her disposal, then what can we make of her decision to use a relatively complex, communicatively weak utterance? The `ignorance-1` model caches this out as a random choice between competitor and target images in the experiment.

However, the an RSA model of behavior in this paradigm - in particular the `3alt_3cost_N` model explored in the previous sections, has several advantages over these simpler models. First, unlike the `ignorance-1` and `ignorance-2` models, which are designed exclusively to be models of behavior in one condition of the experiment, these models easily generalize to other conditions of the experiment without additional stipulations and is able to make quantitative predictions about behavior in each condition. This is because the linking function from competence to behavior proposed for this model - namely that selections of the target image are a function of the pragmatic listener distribution over speaker intended referents - is defined for all conditions of the experiment. Moreover, as a general model of behavior in the entire experimental paradigm (and as a Fully Probabilistic model, whereby both the nature of pragmatic inference and the alternative set are represented probabilistically), an RSA model allows for a quantitative explication of change in behavior across the four conditions, conceived as listeners adapting their expectations of the speaker's linguistic behavior via an update on the representation of the speaker's cost function on utterances.

General discussion and conclusion

This paper considered two analytic approaches to the symmetry problem of scalar implicature: one whereby computation of symmetric alternatives is blocked (necessary to break symmetry given a Traditional or Categorical Alternatives, Probabilistic Inference theory), and one whereby symmetric alternatives are computed but the production probabilities of these alternatives are represented asymmetrically (an option available to Fully Probabilistic and Probabilistic Alternatives, Categorical Inference theories). I argue that one advantage of adopting a Fully Probabilistic theory is that implicature priming behavior - namely the change in listeners' interpretation of weak scalar items given exposure to pragmatically-contrasting linguistic material - can be modeled as listeners' incremental update to their expectations of the types of utterances their interlocutors produce. Three experiments were deployed in order to investigate the nature of the set of alternatives active in interpretation and to explore how listeners' representation of this set can change in a local experimental setting. Along the way, I noted important some important limitations of these studies and avenues for future research, which I recapitulate here.

First, Experiments 1 and 2 explored a novel priming paradigm which revealed itself to be insensitive across the vast majority of scales tested. However, lack of evidence for an effect is not evidence for lack of an effect, and implicature priming is attested in paradigms which more explicitly train participants to associate forms of interest with specific meanings. For

these reasons, future work should explore the possibility that different configurations of the paradigm in Experiments 1 and 2 could give rise to different patterns of results. Above, I note that this paradigm was most effective in modulating behavior on a putative ad-hoc scale, and that future work should investigate whether the priming behavior observed for the *looks like / is* scale generalizes to other ad-hoc scales and/or to putative lexical scales where there appears to be relatively weak pragmatic competition between forms absent priming (van Tiel et al. (2016)). In Experiments 1 and 2, behavior on critical trials was also confounded with item-specific behavior, as every block of trials for a given scale had the same critical trial.

Experiment 3 shows some promising first results that implicature priming can indeed be analyzed as listeners updating their expectations about the utterances their interlocutors produce. Even in the condition of the experiment where symmetric alternatives are made salient, the Bayesian parameter estimation of an RSA model discussed above suggests that this model breaks symmetry by assigning different costs to the alternatives. These results should be confirmed through the testing of a broader range of linguistic constructions, including putative lexical scales.

The context sensitivity of pragmatic inferences, including implicatures, remains under-explored in the experimental pragmatics literature, yet the various pragmatic theories discussed here afford themselves of different empirical predictions in this respect. Therefore, if we should want to ultimately adjudicate between categorical and more gradient conceptions of implicature, future work should continue to attempt to shed light on how language, pragmatic reasoning, and discourse dynamics interact.

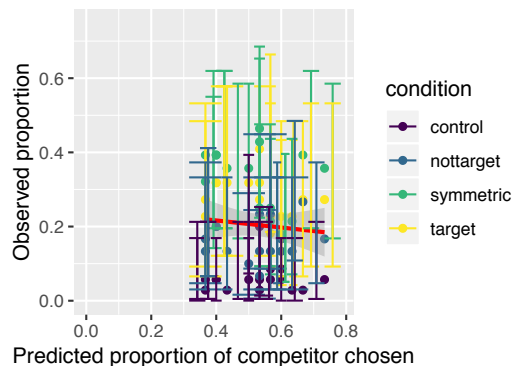# Appendix 1: Predictive checks of models compared for Experiment 3

3alt_3cost_N :
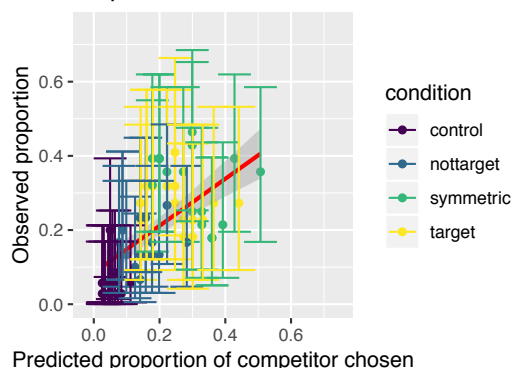Adj R2 = 0.40088 Intercept = 0.080023
Slope = 0.65396 P = 1.19e−08
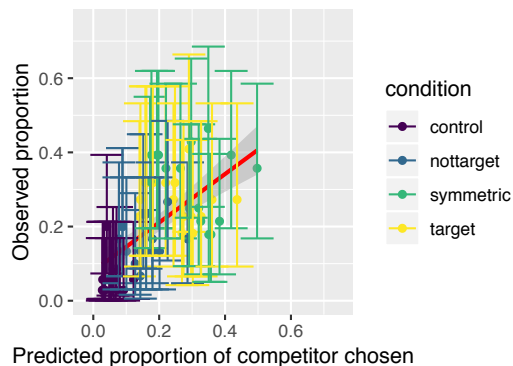
6alt_3cost_noN :
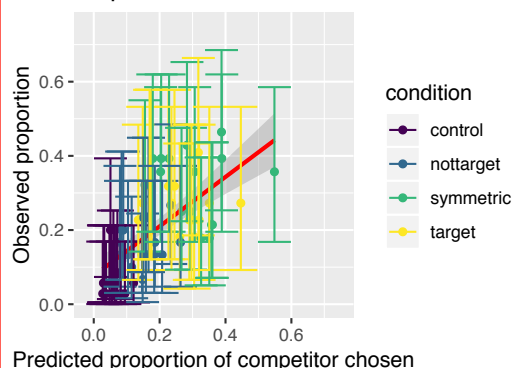Adj R2 = −0.0087788 Intercept = 0.2545
Slope = −0.095174 P = 0.504

3alt_3cost_noN :
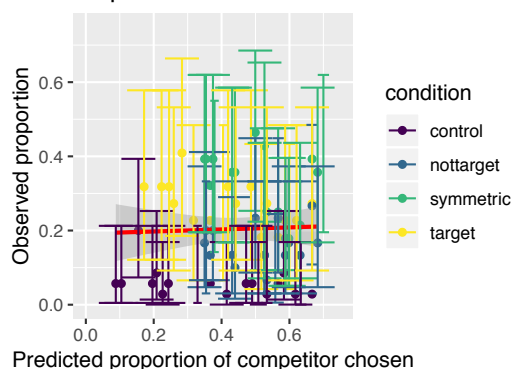Adj R2 = 0.3874 Intercept = 0.084747
Slope = 0.63346 P = 2.4094e−08

6alt_6cost_N :
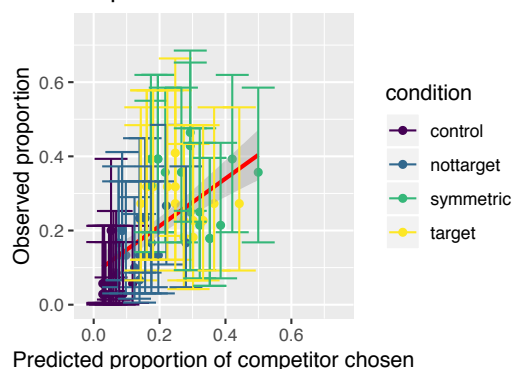Adj R2 = 0.42753 Intercept = 0.076684
Slope = 0.66626 P = 2.8219e−09

6alt_3cost_N :
Adj R2 = −0.014798 Intercept = 0.19183
Slope = 0.027702 P = 0.77649
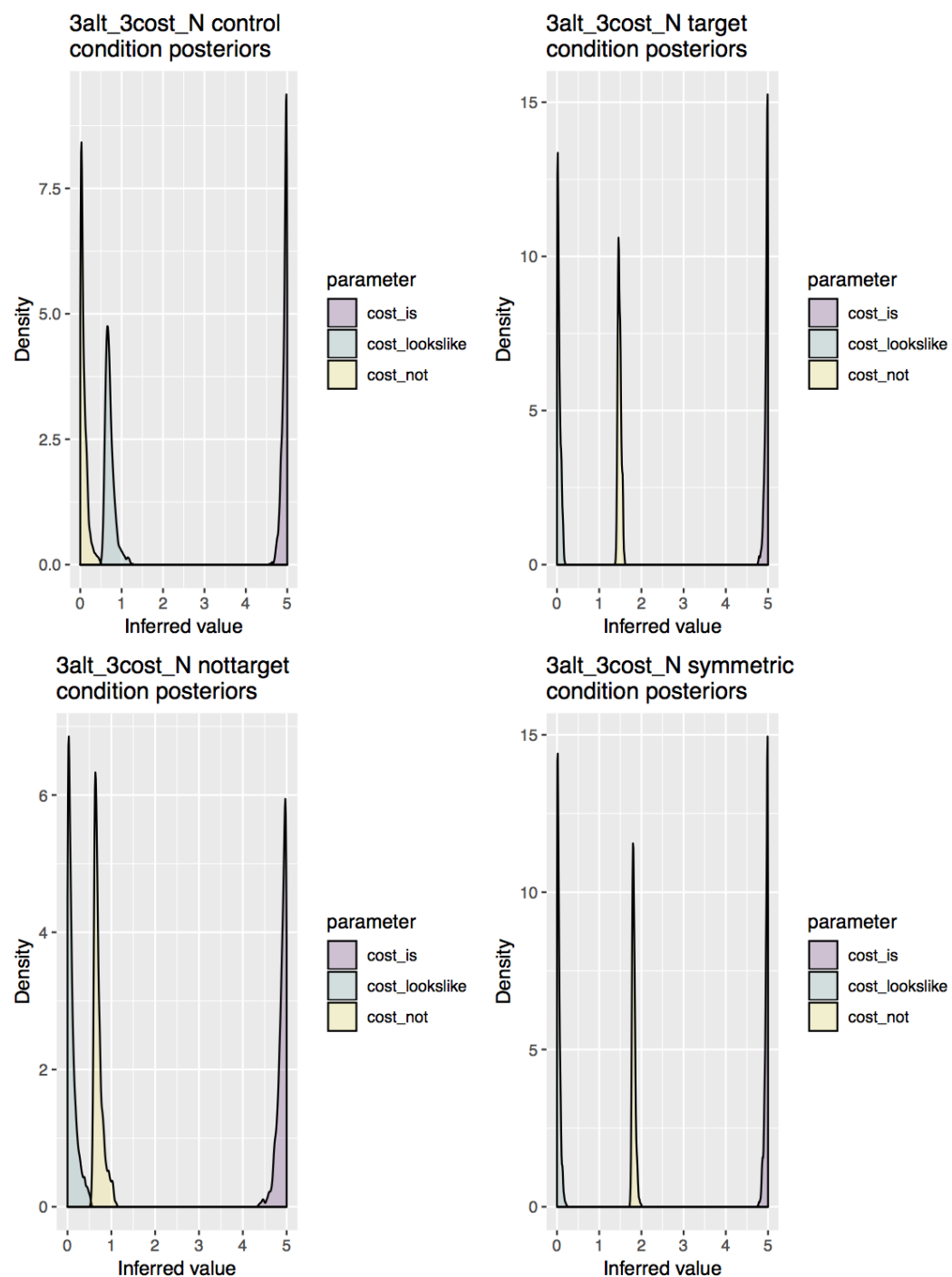
6alt_6cost_noN :
Adj R2 = 0.37952 Intercept = 0.084195
Slope = 0.63976 P = 3.6135e−08

Appendix 2: Posteriors over parameters for the 3alt-3cost-N model

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.

Michael C Frank and Noah D Goodman. Predicting pragmatic reasoning in language games. *Science*, 336(6084):998–998, 2012.

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 Michael C Frank. Pragmatic language interpretation as probabilistic inference. *Trends in cognitive sciences*, 20(11):818–829, 2016.

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 and Andreas Stuhlmüller. The Design and Implementation of Probabilistic Programming Languages. http://dippl.org, 2014. Accessed: 2019-1-16.

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.

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.