How are alternatives 'activated' in pragmatic inference?

Brandon Waldon

***Abstract*—This is a first pass at a proposal for my QP, including the motivation for the project, proposed methods, and a proposed timeline for the completion of the project.**

I. Introduction

Pragmatic enrichment is a key ingredient of linguistic meaning: in uttering U, what a speaker conveys to a listener is comprised of much more than U's entailments. But the nature of pragmatic enrichment has been the object of intense study - as well as the focus of intense debate - in linguistics. For example, upon hearing U (e.g. 'John ate some of the cookies'), how does our listener come to infer $\neg U'$ (e.g. $\neg$(John ate all of the cookies))?

Researchers disagree over the proper way to analyze this phenomenon. A simplified version of the disagreement is as follows: on one side, the Griceans claim that the inference of $\neg U'$ arises due to the listener reasoning counterfactually about why the speaker did not assert the possible alternative U'. On the other side, the grammaticalists argue that the inference of $\neg U'$ arises due to a semantic operation that negates truth-conditionally stronger alternatives of U, including U'.

Regardless of which side one takes in this debate, one needs a theory of the alternative utterances to U which enter into this type of pragmatic reasoning. Researchers typically agree that salient alternatives include sentences which asymetrically entail the original sentence. In the example above, U are U' are differentiated by two lexical items - *some* and *all* - which are said to exist on the same logical scale, and it is this scalar relationship that is said to drive the computation of U' as a salient alternative to U as well as the subsequent inference of $\neg U'$ - hence the term scalar implicature as a commonly-accepted name for this type of inference.

However, that a lexical item L exists on the same scale as some stronger lexical item L' does not guarantee that an utterance containing L will give rise to an implicature containing L'. Levinson (2000) observes that the pragmatically-active lexical contrasts to L are typically monolexemic if L is monolexemic; Fox (2007) and Chierchia et al. (2008) similarly claim that L' will not factor into the computation of alternatives if it is more complex than L.

Furthermore, researchers have noted that apparently, only certain scale-mates to L are privileged enough to enter into the computation of alternatives, even when those scale-mates are equivalent in complexity. For example, Davis (2014) notes Horn's (1972, 1979) claim that *all, most, many,* and *some* form an ordered scale of logical entailment. But citing Levinson (2000), Davis also notes the observation that despite occupying a common logical scale with *all, ..., some*, *several* does not enter into the computation of alternatives to *some*; that is, 'Some S are P' does not implicate 'It is not the case that several S are P'" (Davis 2014).

Taking stock, Davis argues that conversational implicature is constrained by conventions, including conventional lexical contrasts to a given lexical item L which govern the computation of alternatives to utterances containing L. Though there is apparently some systematicity to these conventions (e.g. logical scale relationships, complexity, as discussed above):

> "[T]here is no more reason to expect that our conversational implicature conventions are completely systematic than there is to expect that lexical conventions are... . For example, the regular pattern for adjectives in English is that of *tall, taller, tallest*. But there are exceptions, such as *good, better, best*. No one expects that anything other than a historical explanation exists for these facts." (Davis 2014).

Thus, if we accept that the activation of alternatives is conventional, then we will want an account of those conventions which may or may not include explanatory generalizations. Ultimately, we will also want an account of what I will for the time being call **ad-hoc alternatives**: alternatives to U which are not readily accounted for by our current understanding of implicature conventions and which *prima facie* seem more context-sensitive than what I will for the time being call **conventional alternatives** (what we 'normally' think of as the alternatives to U).

We may ultimately wish to reject this dichotomy (and below, I raise an objection to it). But as an illustration of what I mean by "ad-hoc" versus "conventional" alternatives, consider the findings of Nicolae & Sauerland (2016), who report that "the difference in strength [between *or* and *either-or*] only arises when the two forms are both used; in isolation both disjunctions exhibit the same level of exclusivity" (2016: 1). In one between-subjects experiment, 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* (see Figure 1 for a sample critical trial from this experiment). 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 (see Figure 2 for their results).

To explain these results, the authors claim that absent exposure to *either-or*, participants contrast *or* with its Horn scale-mate *and*, which leads to a pragmatically-derived ex-

clusivity inference that is truth-conditionally equivalent to the semantic exclusivity entailment of *either-or*. However, in a context where *either-or* has been made salient, participants make *either-or* (rather than *and*) the object of contrast with *or*. Viewed in this way, *either-or* is a more context-sensitive and **ad-hoc** lexical contrast to *or* than is *and*, the **conventional** default lexical contrast. Note furthermore that *either-or* is more complex than *or*, contra the generalizations discussed by Levinson (2000), Fox (2007), and Chierchia et al. (2008).

Relatedly, Kurumada et al. (2012) report the results of a experiment in which participants were presented with two images on a screen: "a *target picture* (e.g. a zebra) and a *competitor* (e.g. an okapi, a four-legged animal with black and white stripes only on its legs)" (2012: 649). Participants were asked to select the intended referent between the two images after listening to a sentence uttered by a pre-recorded speaker. In one condition, the critical trials always involved sentences of the form "It looks like a/n [target noun]"; in a second condition, some of the critical trials involved sentences of the form "It is a/n [target noun]". Kurumada et al. report that exposure to sentences of the form "It is an X" had an effect on participant's behavior on "It looks like an X" trials: "Introducing a stronger statement ('It is an X') increases the proportion of contrastive implicatures for the 'It looks like an X' construction" (2012: 652). Here, again, there seems to be something 'ad-hoc' about *It's an X* as an alternative to *It looks like an X*, given its context sensitivity - and perhaps also in light of the fact that there is no entailment relation between the two sentences.

In light of the discussion above, I've formulated three questions which I hope will guide my QP research:

A. Activation of 'ad hoc' alternatives: absolute or gradient?

According to Nicolae & Sauerland's (2016) analysis of *or* and *either-or*, "in a context where both *or* and *either-or* are used to express disjunction, the two disjunctions enter into a competition with each other such that the alternative relevant when computing the meaning of *or* is not *and* but rather *either-or*" (2016: 12). Formally, in such a context, *or* is associated with the following alternatives:[1]

Alt($\mathbf{K}(p \vee q)$ = $\{\mathbf{K}(p \vee q)$, $\mathbf{K}Exh(p \vee q)\}$

... Where *Exh* is an exhaustivity operator (following Chierchia 2006) and $\mathbf{K}Exh(p \vee q)$ is the LF of *either-or*. Importantly, on this analysis, the activation of the ad-hoc alternative *either-or* is an absolute affair: that is, *either-or* is either in the set of alternatives (having displaced *and*) or it isn't.

But this is not the only possibility: we could also imagine that the accessibility of *either-or* as an alternative to *or* is a more complex function of the context. For example, we might imagine that the frequency with which a speaker uses *either-or* affects subsequent interpretation of *or* (beyond whatever frequency of use of *U'* that is necessary for *U'* to enter into pragmatic competition with *U*). This hypothesis **H1** would contrast with a null hypothesis **H0** which is also a prediction of the Nicolae & Sauerland (2016) account of the phenomenon at hand:

H0: Given an utterance *U* and an ad-hoc alternative *U'* to *U*, interpretation of *U* is not affected by the frequency with which a speaker employs *U'*. (Interpretation is affected only by whether *U'* is understood by the listener to be an alternative to *U*.)

H1: Given an utterance *U* and an ad-hoc alternative *U'* to *U*, interpretation of *U* is affected by the frequency with which a speaker employs *U'*.

Neither Kurumada et al. (2012) nor Nicolae & Sauerland (2016) report on the presence/absence of any ordering effects in their respective experiments that would help to adjudicate between **H0** and **H1**. However, in both experiments, participants were exposed to multiple primes of the ad-hoc alternative, so it should be possible to reanalyze the original data or run replication studies to determine whether the number of prior exposures to the ad-hoc alternative modulated the effect on interpretation of the the utterance.

Should we reject **H0** in favor of **H1** (that is, should we find contra Nicolae and Sauerland that the activation of ad-hoc alternatives is gradient rather than absolute), the Rational Speech Act (RSA) framework offers the means with which to formalize the phenomenon. Recall that the RSA framework models a pragmatically competent listener L_1 as reasoning about a speaker S_1 intending to communicate a state of affairs to a listener L_0 whose interpretation of an utterance is in turn a function of that utterance's truth conditional content her prior beliefs about the state of the world:

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

The goal is to show how a pragmatically-competent L_1's probabilistic distribution of possible world states given an utterance of U changes as exposure to some alternative U' increases. By way of example, consider again the phenomenon observed by Kurumada et al. (2012), in which exposure to sentences of the form "It is a/n [target noun]" appears to have increased P_{L_1}(not an x | "It looks like an x"), the probability that the speaker's intended referent with "It looks like a/n [target noun]" was the contrast image rather than the target image. Within the RSA model, there are four parameters which we can change to reflect an increase in P_{L_1}(not an x | "It looks like an x"):[2]

1) A change in the world state prior, namely an increase in P_w(intended referent is the competitor image).
2) A change in a prior probability distribution on utterances, namely an increase in prior probability of the

[1] The **K** operator, a covert doxastic operator, is employed in order to prevent exhaustified *or* from receiving a conjunctive interpretation, given *either-or* as an alternative. But otherwise, the details are irrelevant to the discussion here.

[2] A working WebPPL implementation of an RSA model which models a trial of the Kurumada et al. experiment as a signaling game is available at https://bwaldon.github.io/files/kurumada2012.js and executable in the online WebPPL editor: http://webppl.org/.

speaker saying the ad-hoc alternative, "It's a/n [target noun]"
3) Equivalently, a reduction in the speaker cost of uttering "It's a/n [target noun]"
4) An increase in the rationality parameter α.

There is certainly much more to be said here. For one thing, one outstanding issue is that we don't have a full understanding of the empirical picture regarding ad-hoc alternatives (specifically whether their activation in context is an absolute or gradient affair). But the main upshot for the time being is that this is a possible area where RSA could provide us qualitatively superior empirical coverage than existing formalisms of pragmatic competence.

B. 'Ad-hoc' alternatives vs. 'conventional' alternatives: a meaningful distinction?

Another qualititatively unique feature of the RSA framework is that there is no formal distinction between 'conventional' and 'ad-hoc' alternatives as I have defined them here. In the RSA, the distinction is most naturally represented as one of degree rather than one of kind: that is, possible utterances may vary by their prior probability or cost, but they are never categorically absent from pragmatic competition with a given utterance. Contrast this with a Nicolae & Sauerland-style analysis, where certain alternatives to an utterance are conventional defaults (e.g. *and* as an alternative to *or*), while others (e.g. *either-or*) are made available given special contextual conditions. Furthermore, on a Nicolae & Sauerland-style analysis, we need a theory of the set of 'ad-hoc' alternatives which can be contextually activated: does it consist of the logically stronger alternatives to U, possibly including sentences more complex than U? This might be adequate in the case of *or* / *either-or*, but it will not explain the effect found by Kurumada et al. (2012) (there is no entailment relationship between "looks like an X" and "is an X": one can be a police officer but not look like a police officer).

At a deeper level, how justified is the 'ad-hoc' vs. 'conventional' distinction? How can we empirically investigate whether this distinction is justified? If some alternatives to U are indeed 'conventional', then it would be surprising if their activation had similar context-dependence to what we have thus far called 'ad-hoc' alternatives, including prior use in the discourse:

H0: Given an utterance *U* and a putatively conventional alternative *U'* to *U*, interpretation of *U* is not affected by the frequency with which a speaker employs *U'* prior to uttering *U*.

H1: Given an utterance *U* and a putatively conventional alternative *U'* to *U*, interpretation of *U* is affected by the frequency with which a speaker employs *U'* prior to uttering *U*.

Consider the case of Nicolae & Sauerland's experiment with *or* and *either-or*: what if we were to similarly find that exposure to *or*'s conventional alternative *and* had an effect on interpretation of *or*? This would be strongly suggestive of **H1** and hence of reason to rethink the ad-hoc/conventional distinction.

II. Methods

We first want to investigate *Question A*: whether the activation of 'ad-hoc' alternatives is absolute or gradient. As a first pass at answering this question, it should be possible either to run replications of Kurumada et al. (2012) and Nicolae & Sauerland (2016) or reanalyze the results of those original studies to look for ordering effects. In both studies, participants provided interpretations of U after seeing N instances of an ad-hoc lexical contrast to L in U, for varying values of N.[3]

Answering *Question B*, whether the ad-hoc/conventional distinction is meaningful for alternatives, requires a new experimental design. The proposed experiment involves taking the basic methodology of Nicolae & Sauerland (2016) (i.e. asking participants to rate the likeliness of a pragmatically-enriched meaning of a given utterance) and running more experimental conditions.

- Condition 1: Following Nicolae & Sauerland (2016), a condition in which participants interpret an utterance without prior exposure to (conventional or 'ad-hoc') alternatives to that utterance.
- Condition 2: Following Nicolae & Sauerland (2016), a condition in which participants interpret an utterance with prior exposure to a plausible ad-hoc alternative.
- Condition 3: A novel condition in which participants interpret an utterance with prior exposure to a conventional alternative.

The proposed experiment also involves investigating the interpretation of a broader set of lexical items than simply *or*, including:

- *Some*, its putative conventional contrast *all*, and a plausible ad-hoc contrast *some but not all*.
- Adjective pairs, e.g. *good*, a putative conventional contrast *great*, and a plausible ad-hoc contrast *only good*.
- *Looks like* and two plausible ad-hoc contrasts *is* and *is not*.

A. Pilot study

In a pilot experiment,

[3]Chigusa Kurumada (pc.) has offered to either rerun the original experiment of Kurumada et al. (2012) or to provide me with the dataset from that experiment.

"Sandy bought a dress or a shirt."
suggests
Sandy didn't buy both a dress and a shirt.

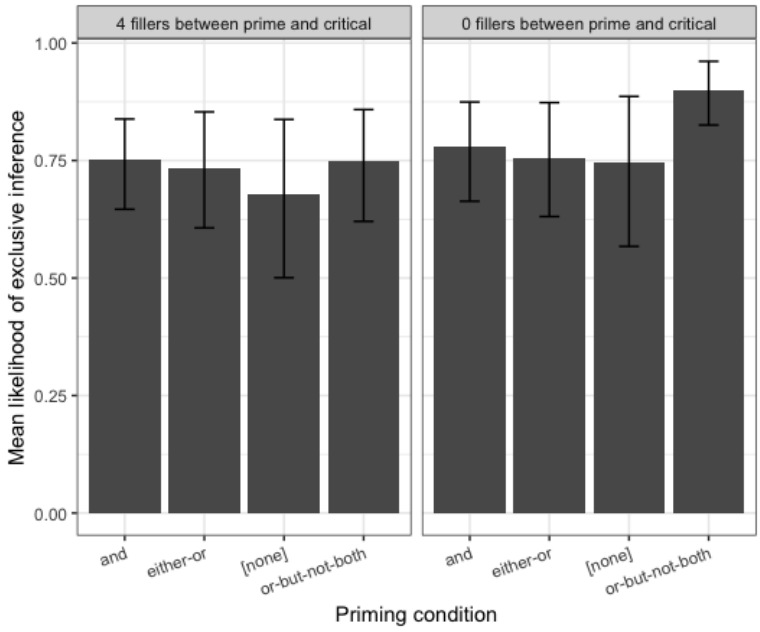
very unlikely ○ ○ ○ ○ ○ ○ ○ very likely

Fig. 1. A sample critical trial from Nicolae & Sauerland's (2016) experiments.

(12) *Average response by condition*

	OR	EITHER–OR
Experiment 1	5.671	5.773
Experiment 2a	5.655	5.946

Fig. 2. The results of Nicolae & Sauerland's (2016) Experiment 1 and Experiment 2a. Experiment 1 was a between-subjects manipulation in which participants saw exclusively *or* or *either-or* trials. In Experiment 2a, participants saw both types of trials. The table shows mean ratings on a 1-7 scale, with 1 = *very unlikely* and 7 = *very likely*.



III. Proposed timeline

This quarter, I'm in two classes which can

IV. Conclusion

The conclusion goes here.

References

[1] H. Kopka and P. W. Daly, *A Guide to LaTeX*, 3rd ed. Harlow, England: Addison-Wesley, 1999.