Haddock Descriptions

Helena, Roger, Liz

September 20, 2019

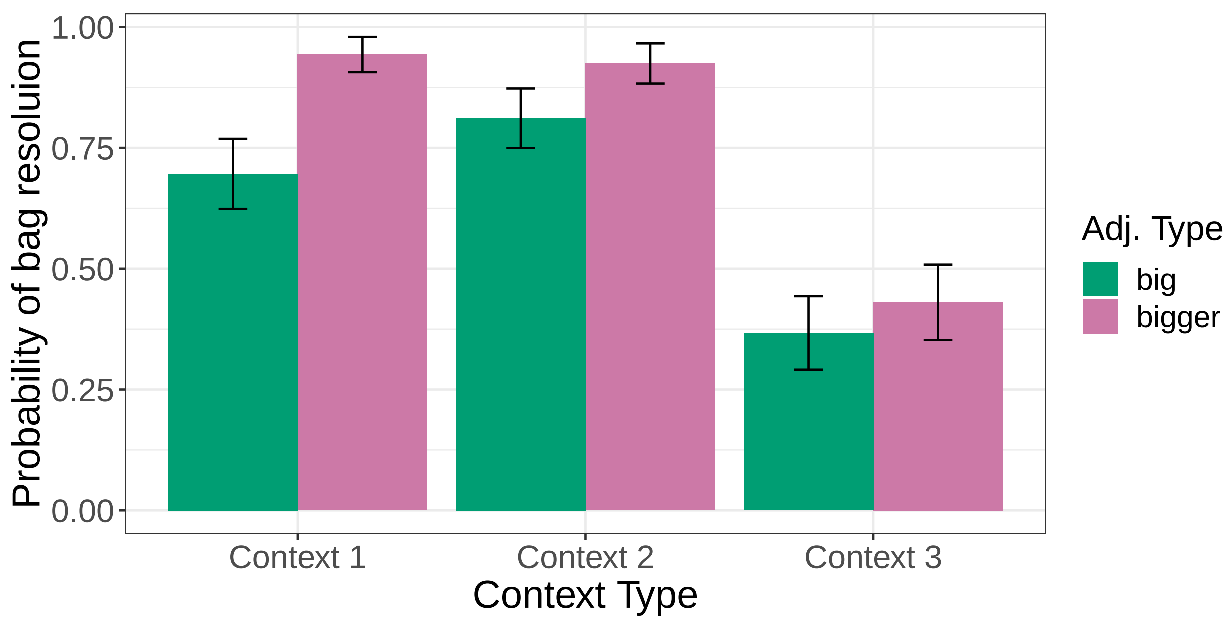
1 Experiment

1.1 Design

1.2 Results

this section is outdated

Experimental results are plotted in Figure XX, which plots for each of the three contexts tested the probability of bag resolution (proportion of clicks to the rabbit in the medium bag vs. clicks on the the rabbit in the medium box). The yellow line corresponds to chance after correcting for responses that did not correspond to any of these two objects (a total of X).



Type of analysis: results for subset of data corresponding to the first two conditions: posterior mean estimate and 95% credible interval: $\beta = -1.25[-2.63, 0.02]$

2 Semantic Assumptions

2.1 Definite Article

2.2 Gradable Adjectives

2.3 Comparative

3 Computational Model

We implement a one level RSA model where the pragmatic listener jointly infers a referent, a threshold and a context, where the latter can be narrowed down to accomodate the semantic requirements of the speaker's utterance.

3.1 Literal Listener

The Literal Listener infers (assigns posterior probability) a referent r given a description d, a context C and a threshold θ used in the interpretation of the relative adjective *big* or its comparative form *bigger* (see semantics in the previous section). This is done proportionally to whether the description d is true of r in C for the threshold value d times the prior probability of r. Put it differently, the literal listener discards potential referents that do not satisfy the semantic requirements of the description, and assigns posterior probability to each remaining regerent that is modulated by its prior probability.

$$L_0(r \mid d, C, \theta) \propto [\![d]\!]^{C,\theta}(r) \cdot P(r) \tag{1}$$

In this model, both θ and C are treated as lifted variables. This means that the value of both θ and C, whose value is not resolved at the fist level of the model (i.e., at the Literal Listener level), but rather 'lifted' all the way up to the Pragmatic Listener.

We assume a flat prior over referents. However, this distribution is undefinable if there is no referent in C satisfying d, i.e., if there is presupposition failure of existence and/or uniqueness.

$$P(r) = \begin{cases} \epsilon & \text{if } r = \mathsf{fail} \\ \text{uniform} & \text{otherwise} \end{cases} \tag{2}$$

Technically, we achieve this effect by posing a special fail referent, with prior probability ϵ (see 3).

3.2 Speaker

The speaker softMax agent chooses a description d given the referent context and threshold that she wishes to convey by maximizing the likelihood of the LIteral Listener inferring the right referent, while minimizing production cost. . The former in terms of informativity the latter in terms of production cost assigns probabilities to descriptions d given intended referent r, context C and threshold θ that reflect both the probability of the literal listener picking the referent and the cost (length) of d.

$$S_1(d \mid r, C, \theta) \propto \exp(\alpha \times \ln(L_0(r \mid d, C, \theta)) - \mathsf{cost}(d)) \tag{3}$$

Utility of u for the Speaker is proportional to its informativity to the literal listener minus the utterance cost cost $C(u)$.

The utterance cost comparative 1.5, positive 1 and 0.5 if the utterance does not contain an adjective.

Informativity is quantified as negative surprisal (or positive log probability) of the referent in the posterior.

Finally, we assume the rationality parameter $\alpha = 1$. Ensures that the speaker is only quasi-deterministic

Questions to consider or to keep in mind: Posterior robabilities go down if we have both pos and comparative compete with each other. We currently do not use a *silence* utterance among the alternatives.

3.3 Pragmatic Listener

Assigns probabilities to referents r given (partially masked) description d, reflecting the probability of the speaker using (any resolution of) d to describe r in any context C with any threshold θ. A context is defined as any element powerset of the xxx (excluding the empty set).
Marginalizing over C, θ, and N_2:

$$\begin{aligned} &L_1(r \mid d = N_1 \text{ in the (Adj) } \textsf{[masked]}) \propto \\ &\textstyle\sum_C \sum_\theta \sum_{N_2} S_1(d = N_1 \text{ in the (Adj) } N_2 \mid r, C, \theta) \cdot P(r \mid C) \cdot P(\theta \mid C, d) \cdot P(C) \end{aligned} \tag{4}$$

For models without context coordination, the context is $C = \{r_1, r_2, r_3, r_4, r_5\}$. Models that allow for context coordination are such that a cotext C' is set to $C' \subseteq \mathcal{P}(C)$.

We have considered two types of priors over contexts: either a flat prior, or a prior with different degrees of skewedness that assigns higher probability to bigger contexts.

Prior over thresholds is flat, but only thresholds that are instantiated by CC in context + 1.

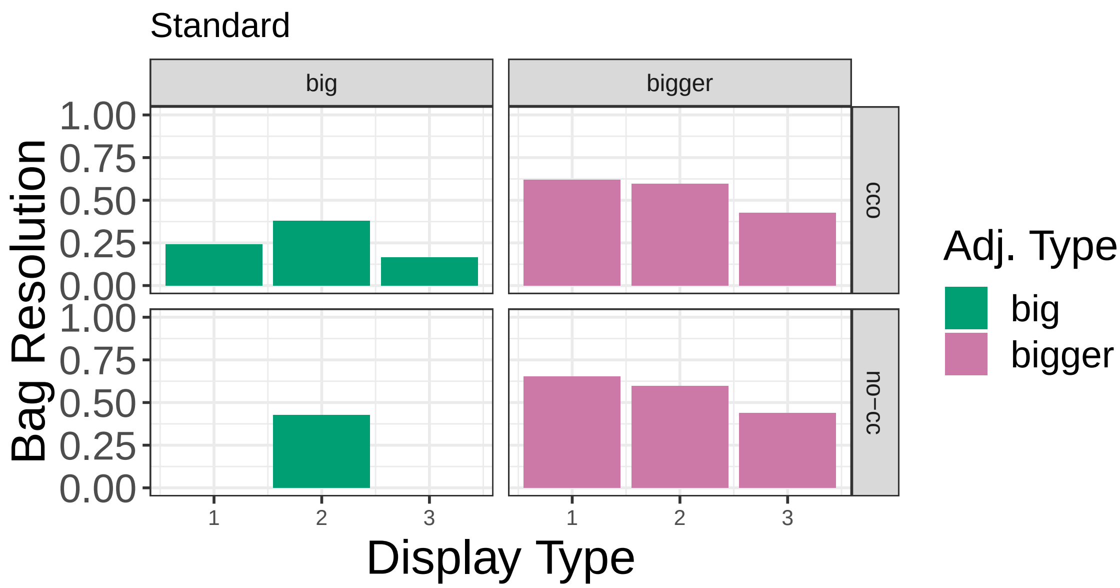
Open questions:

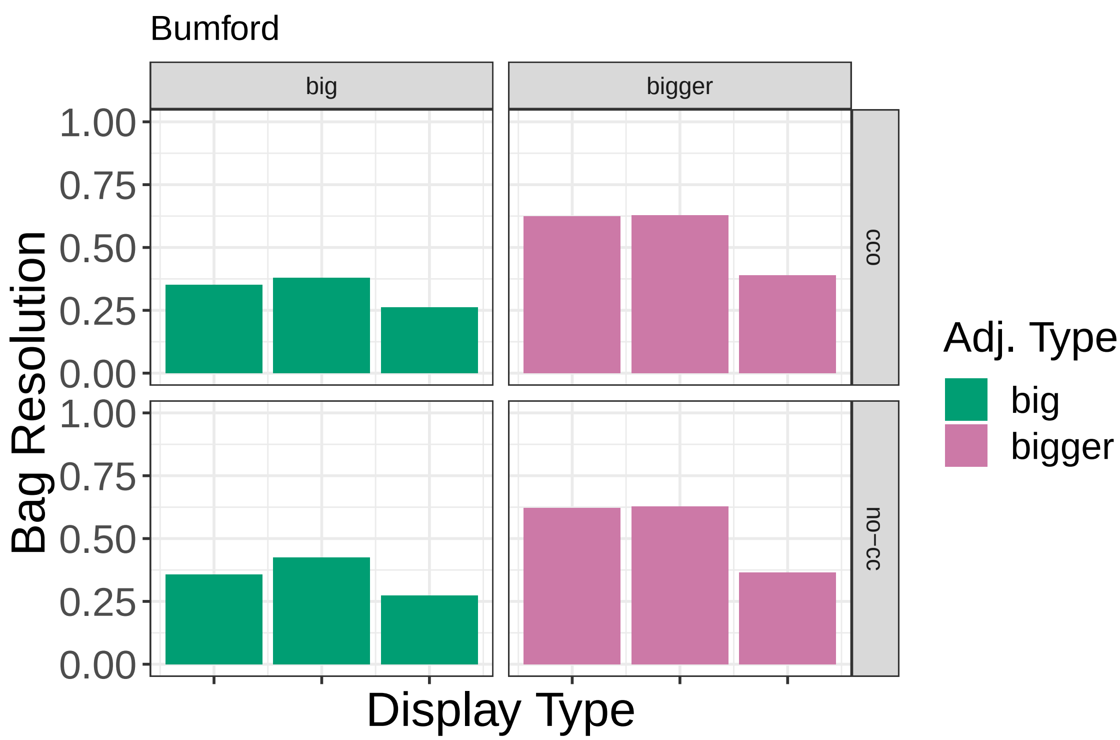
1. As of now, skewness is only in context prior. Do we also want to use a skewed prior as part of the referents given context probability?

```
var referentsPriorGivenContext = function(context) {
  return Infer({method: "enumerate"}, function() {
     flip(0.01) ? rU : uniformDraw(context)
  });
};
```

2. Alternative ways of implementing skeweness?
3. Do we want flat prior over thresholds? As of now it is flat.

3.4 Simulations





3.5 Pragmatic Listener 2

Pragmatic Listener Full Utterance

$$L_1(r \mid d) \propto \sum_C \sum_\theta S_1(d \mid r, C, \theta) \cdot P(r \mid C) \cdot P(\theta \mid C, d) \cdot P(C) \tag{5}$$

Pragmatic Speaker

$$S_2(u \mid r) \propto \exp(\alpha \times \ln(L_1(r \mid d)) - \textsf{cost}(d)) \tag{6}$$

Second Level Pragmatic Listener (L2)

$$L_2(r \mid d = N_1 \text{ in the (Adj) } \textsf{[masked]}) \propto \sum_{N_2} S_2(u \mid r) \cdot P(r) \tag{7}$$

Questions to consider or to keep in mind: No inference over thresholds or contexts at L2?