Rational Speech Acts

...or why this presentation took me a thousand hours to make

Saujas Vaduguru

IIIT Hyderabad

2019

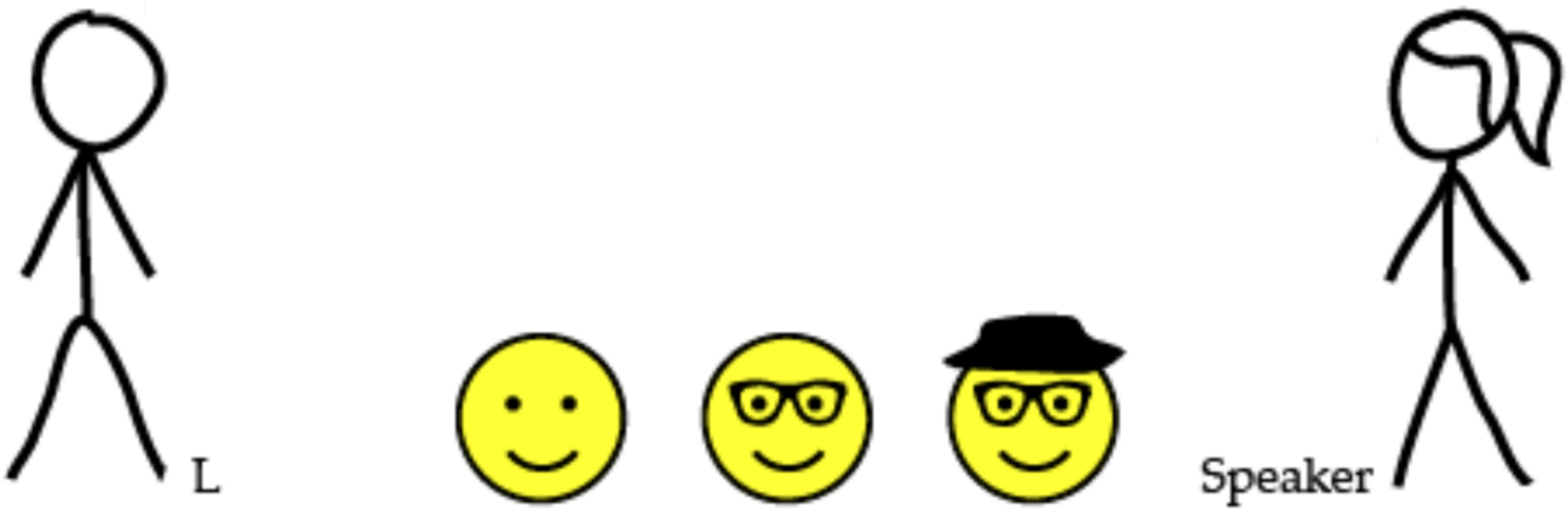
Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged.

– H. P. Grice's Cooperative Principle [2]

Grice's Cooperative Principle

- Speakers cooperatively choose their utterances to convey a meaning
- Listeners work backwards from the utterance to infer the speaker's goal
- Conversational maxims
  - *Quality*: Be truthful, and accurate
  - *Quantity*: Be only as informative as required
  - *Relation*: Be relevant
  - *Manner*: Be perspicuous

A Simple Reference Game



- Referents/Worlds
  - N: No glasses or hat
  - G: Glasses, but no hat
  - HG: Glasses and hat
- Utterances
  - *'none'*
  - *'glasses'*
  - *'hat'*

What would a Gricean Speaker and Listener do?

- Let's say our speaker wanted to refer to the G, but no hat
- The Speaker could just say *'glasses'*
- But, there are **two** potential referents
- The Listener is under the assumption that the speaker is cooperative
- Since a cooperative speaker would be informative, the listener assumes the speaker would have specified the hat if she was referring to HG
- The Listener then ***infers*** that the speaker is referring to G

So did Grice solve pragmatics?

- Grice lays out a powerful framework
- But the predictions are qualitative
- Qualitative predictions don't sit well with empirical testing

Can we do better?

Bayes to the rescue!

- We can model the Speaker and Listener as agents who perform Bayesian inference!
- Bayesian, or probabilistic, models have been used to understand the interplay between structured representations, and statistical information [1]
  - We can use a 'structured representation' to account for Gricean maxims
  - Bayesian inference allows for quantitative predictions, and experimental testing

- The Listener infers the state of the world w from the Speaker's utterance u using Bayes' rule

$$P_L(w|u) = \frac{P_S(u|w)P(w)}{\sum_{w'} P_S(u|w')P(w')}$$

- $P_L(w|u)$ is the probability that the listener infers that the speaker is referring to world w when she says u
- $P(w)$ is the probability of referring to world w, and tells us about the 'salience' of a particular world

- The Speaker can be viewed as a utility-maximising agent
  - Longer utterances are more expensive
  - But, they are more beneficial to the Listener
- To put in an equation

$$P_S(u|w) = \frac{exp(\alpha U(u,w))}{\sum_{u'} exp(\alpha U(u',w))}$$

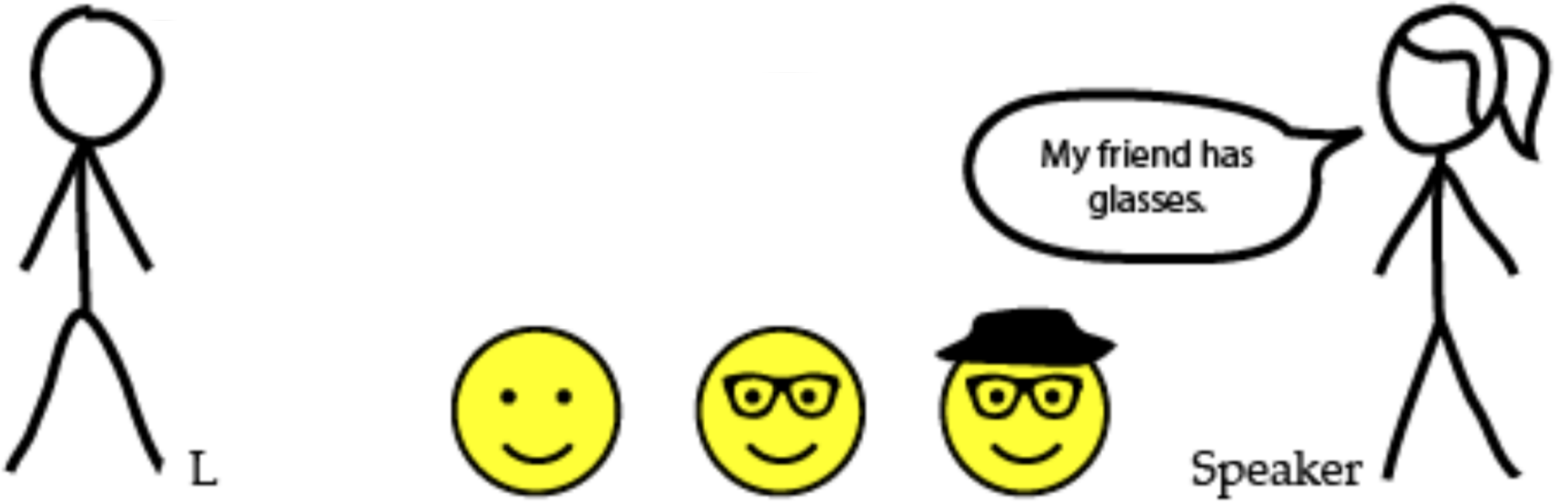
- $P_S(u|w)$ tells us the probability of
- $U(u,w)$ tells us about the utility of saying utterance u about the world w
- α represents how pragmatic the speaker is
  - For this presentation, we assume α to be 1 everywhere

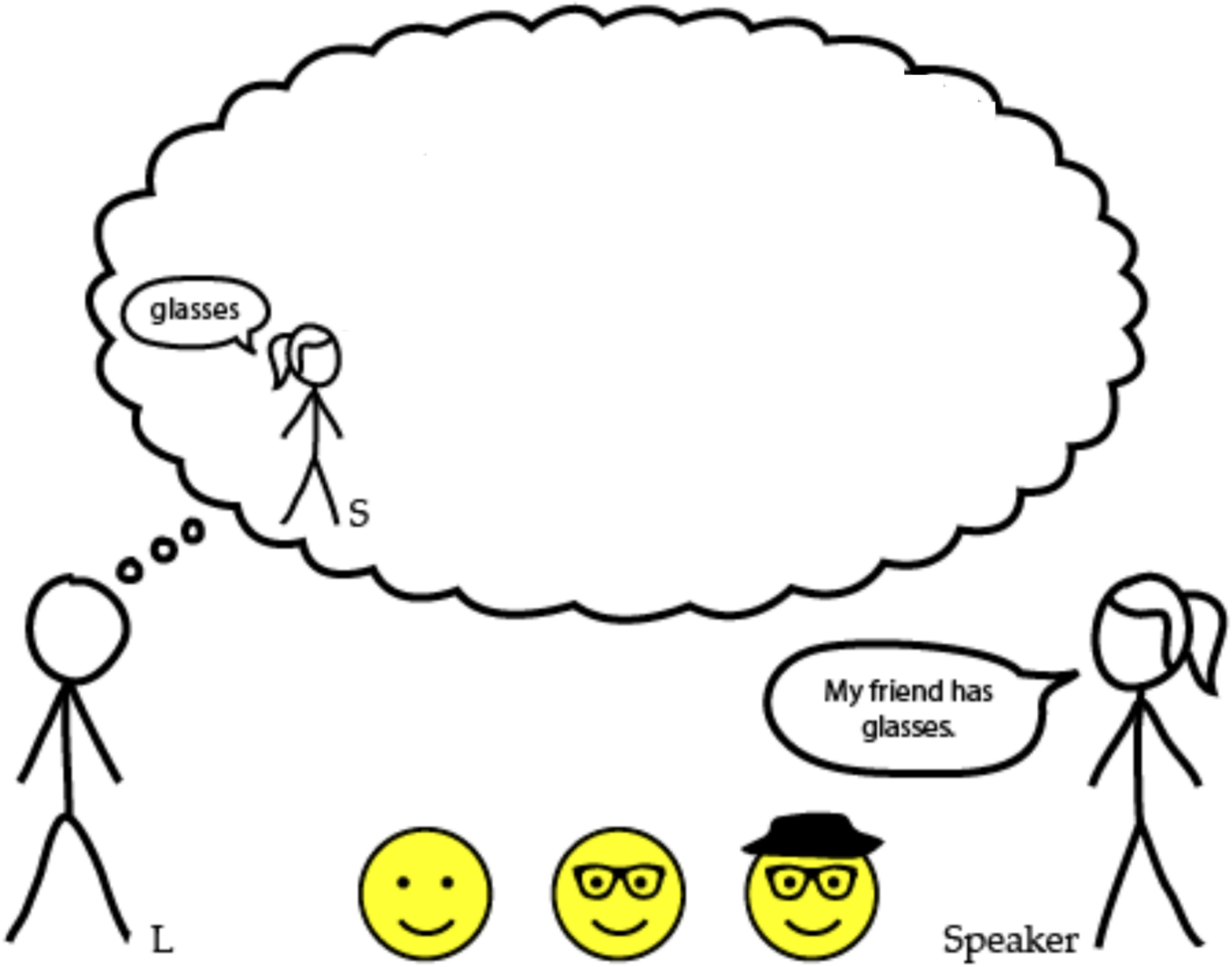
- We terminate the recursion now, by using the beliefs of the literal listener to determine the utility of the utterance
- A literal listener is an agent that isn't pragmatic
  - Given an utterance like *'glasses'* from the game, he will consider both G and HG to be equally likely
-

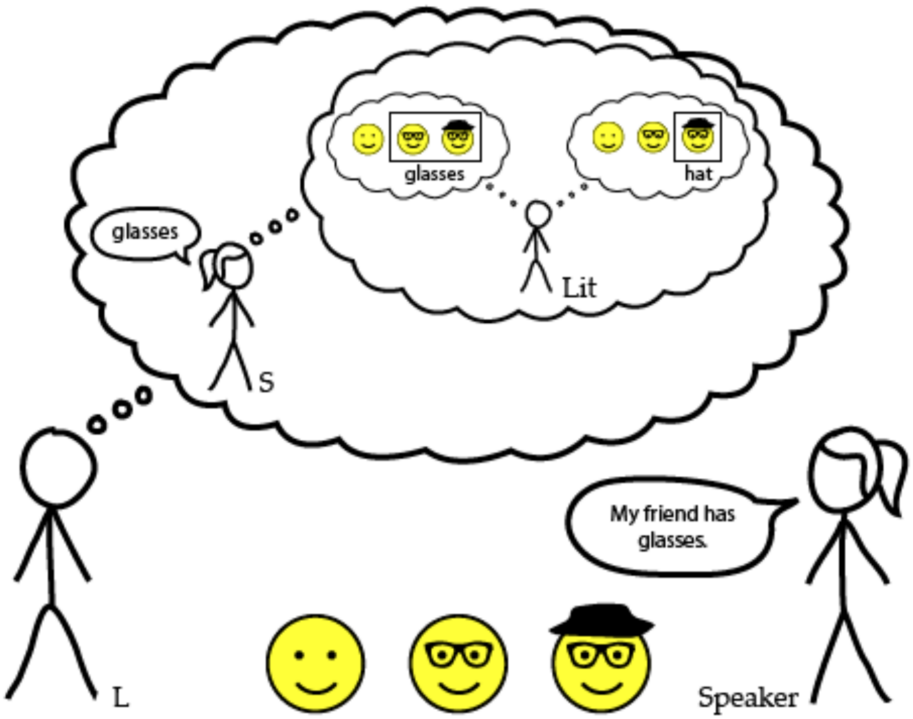
$$U(u,w) = log\, P_{Lit}(w|u)$$

$$P_{Lit}(w|u) = \frac{[\![u]\!](w)}{\sum_{w'} [\![u]\!](w')}$$

- $[\![u]\!](w)$ evaluates to true if utterance u holds in world w, false otherwise


My friend has glasses.
L
Speaker

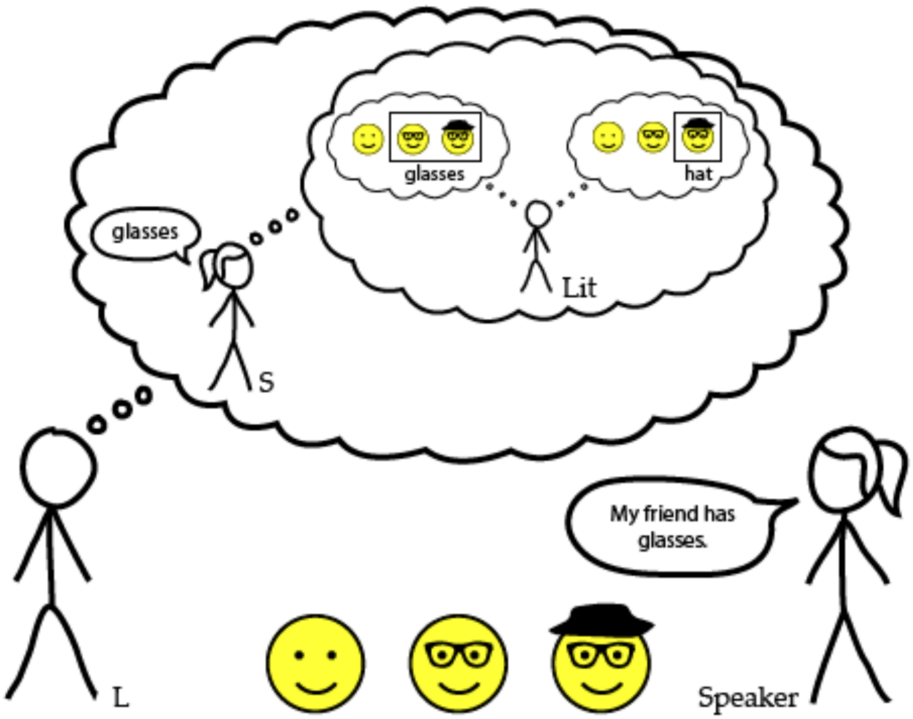

glasses
S
My friend has glasses.
L
Speaker



$$P_{Lit}(w|u) = \frac{[\![u]\!](w)}{\sum_{w'} [\![u]\!](w')}$$

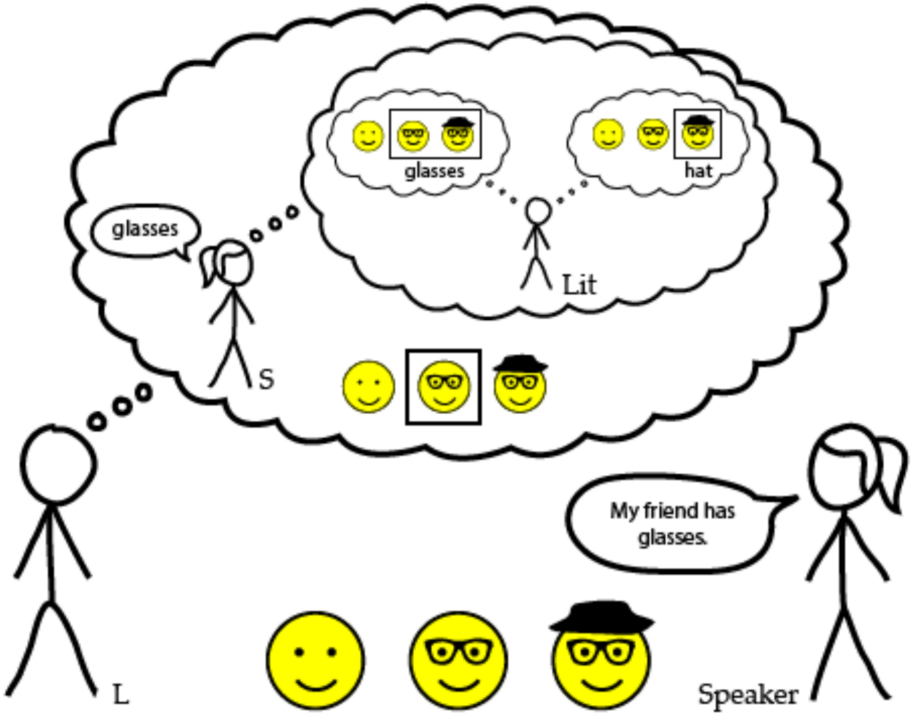
$$U(u,w) = log\, P_{Lit}(w|u)$$

P_{Lit}	HG	G
'glasses'	$\frac{1}{2}$	$\frac{1}{2}$
'hat'	1	0



$$P_S(u|w) = \frac{exp(U(u,w))}{\sum_{u'} exp(U(u',w))}$$

P_S	'glasses'	'hat'
HG	$\frac{1}{3}$	$\frac{2}{3}$
G	1	0



[4]

$$P_L(w|u) = \frac{P_S(u|w)P(w)}{\sum_{w'} P_S(u|w')P(w')}$$

P_L	HG	G
'glasses'	$\frac{1}{4}$	$\frac{3}{4}$
'hat'	1	0

We now have a model...

- How does the model work on more complex tasks than toy reference games?
- Now that we have a quantitative model, how does it compare to human cognition?

...and some exciting possibilities

We take a look at two applications of this model, and how they compare to human performance:

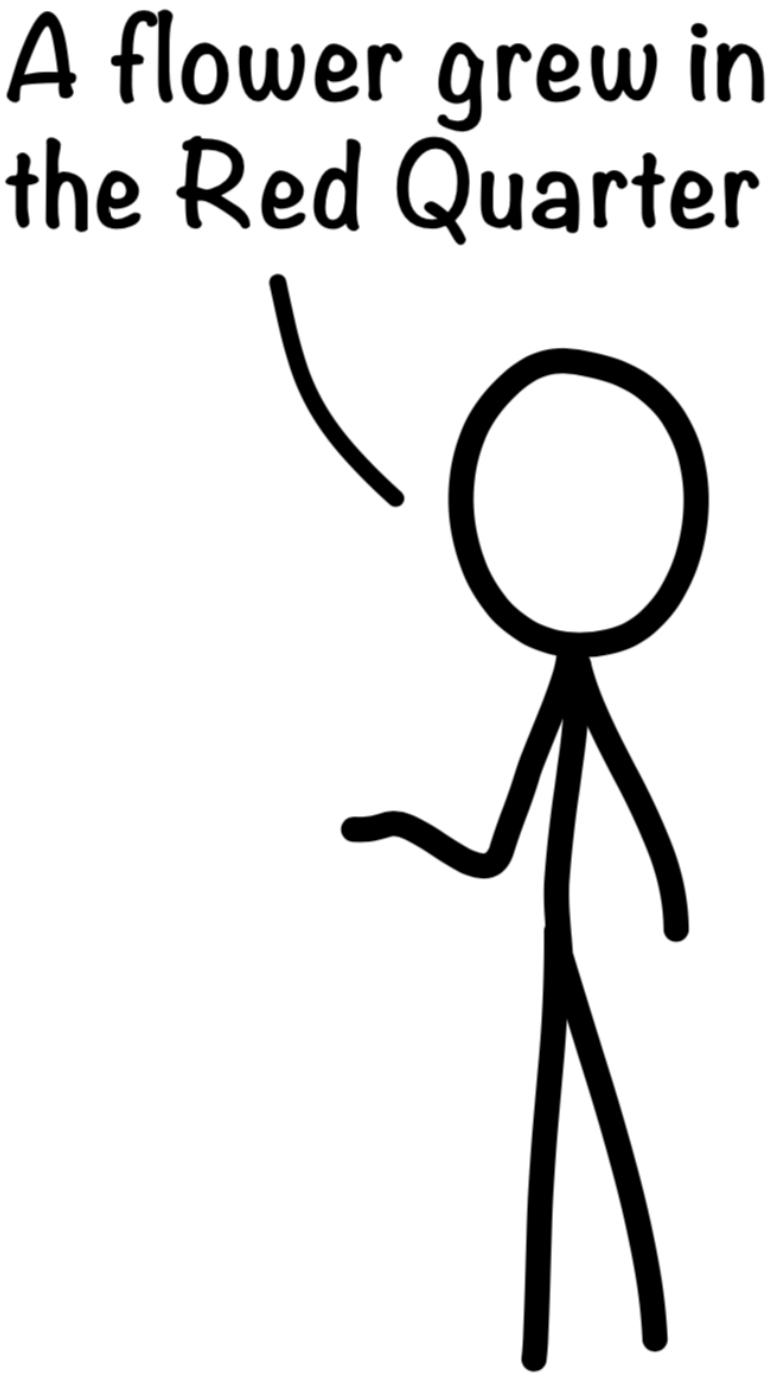
- Spatial implicature [5]
- Number words [3]

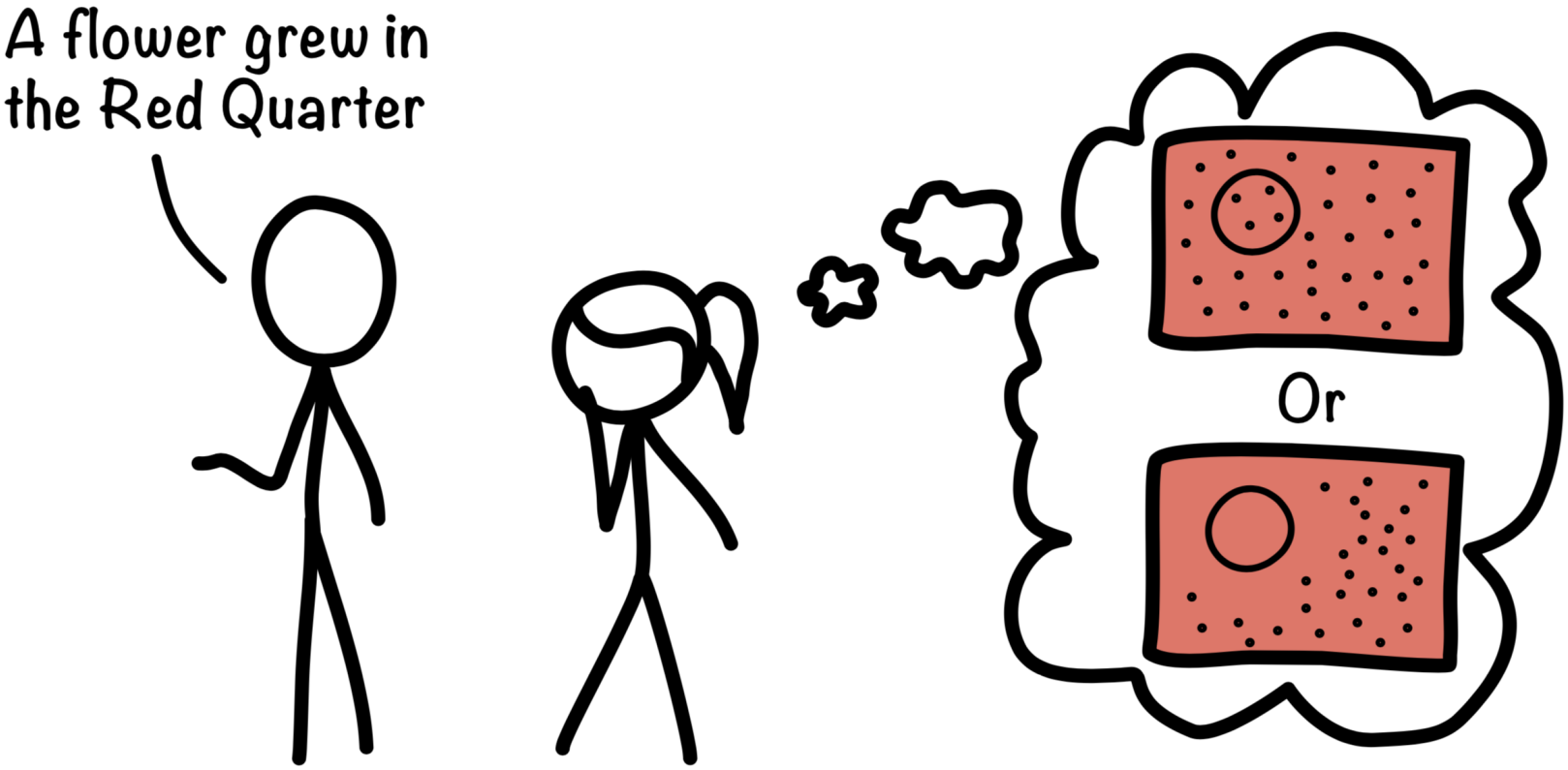
Spatial implicature

- Space is continuous, but language discrete
- For accurate communication of spatial relations, we have to rely on pragmatics
  - Let's see why...

The World

- In this world, there is a special flower called the gold lily that grows at a particular location
- To describe the location, we can use the prepositions *{in, near, near_edge}*, and the locations *{The City, The Red Quarter, The Blue Quarter, The Victory Plaza}*
- To model the World, we can think of it as a 2D plane, with the location of the gold lily being a particular point, and each of the locations being a set of points on the plane


A flower grew in
the Red Quarter


A flower grew in the Red Quarter
Or

RSA for spatial implicature: Literal listener

- $P_{Lit}(w|u) = \frac{[\![u]\!](w)}{\sum_{w'} [\![u]\!](w')}$
  - u still refers to the utterance, but w now represents the location of the gold lily
- $[\![u]\!]$ also means something different depending on the preposition and the location R that make up u
  - $[\![in\ R]\!](w)$ means $w \in R$
  - $[\![near_edge\ R]\!](w)$ means $distance(w, boundary(R)) < d$
  - $[\![near\ R]\!](w) = [\![near_edge\ R]\!](w) \wedge \neg [\![in\ R]\!](w)$

RSA for spatial implicature: Literal listener

- The definition of *near* can depend on the object, we allow the literal listener to choose whichever value of d makes the utterance true
- $$P_{Lit}(w|u) = \sum_d \frac{[\![u]\!]^d(w)}{\sum_{w'} [\![u]\!]^d(w')} P(d)$$

  where $P(d)$ is a uniform distribution between 10 and 30 units, and $[\![u]\!]^d(w)$ is $[\![u]\!](w)$ evaluated with distance d for *near*

RSA for spatial implicature: Speaker

- $$P_S(u|w) = \sum_{w', |w'-w|<\epsilon} \frac{P_{Lit}(w'|u)P(u)}{P(w')}$$
- P(w) represents the probability of the lily being at the location w, and is a uniform distribution
- The speaker is happy as long as the the listener infers a location that is close enough, so we evaluate for all points within ϵ of the true location (ϵ is set to 20 units)

RSA for spatial implicature: Pragmatic listener

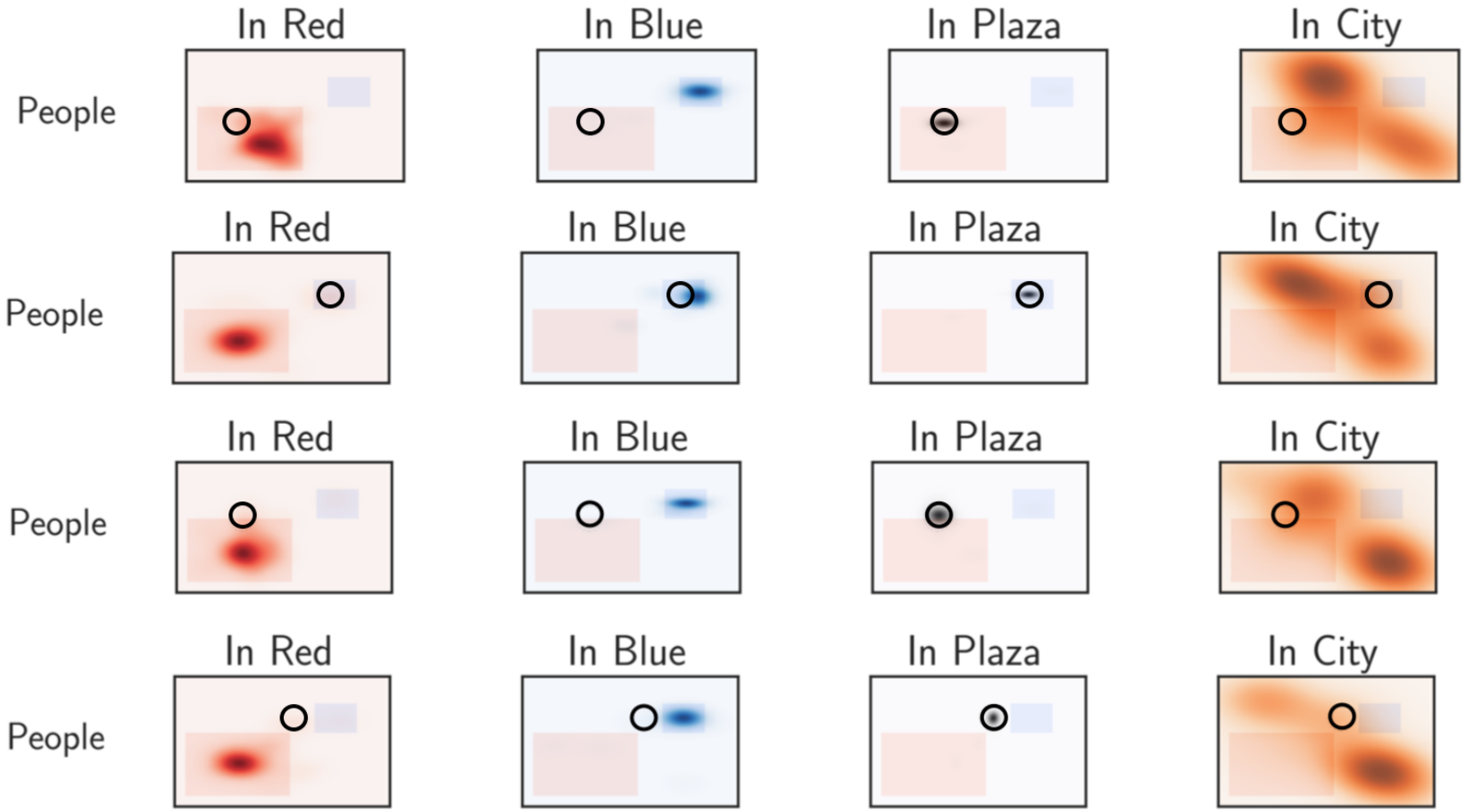
- $$P_L(w|u) = \frac{P_S(u|w)P(w)}{P(u)}$$
- $P(u)$ is the probability of the utterance being u, and is a uniform distribution

Experiments with people

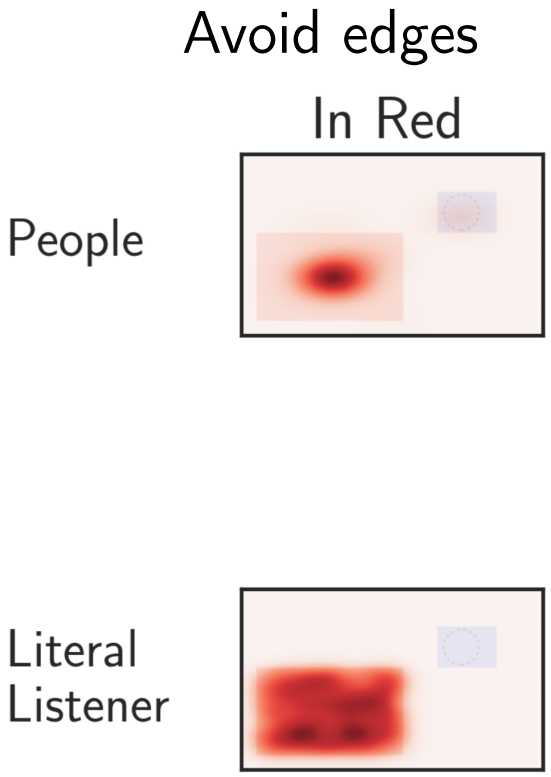
- Participants presented with the world, and a hint from a speaker
- The hint is constructed using a (w, l) pair, drawn from $word \times location$, where $word = \{in, near\}$, and $location = \{Red\ \ Quarter, Blue\ \ Quarter, Victory\ \ Plaza, City\}$
- The participant has to indicate where they think the gold lily is
- Participants were shown all combinations of (w, l) except $(near, City)$ for 4 different configurations of the city
  - 28 prompts in random order

Human responses: Qualitative results

In X means *In X but not in Y*



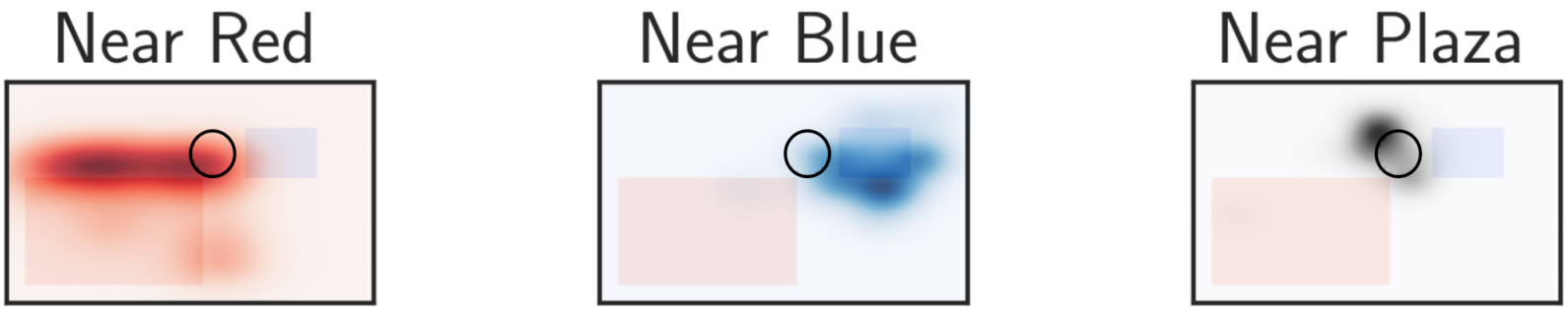
Human responses: Qualitative results



- When there are no intersections between regions, people don't guess uniformly
- For comparison, a literal listener does guess uniformly

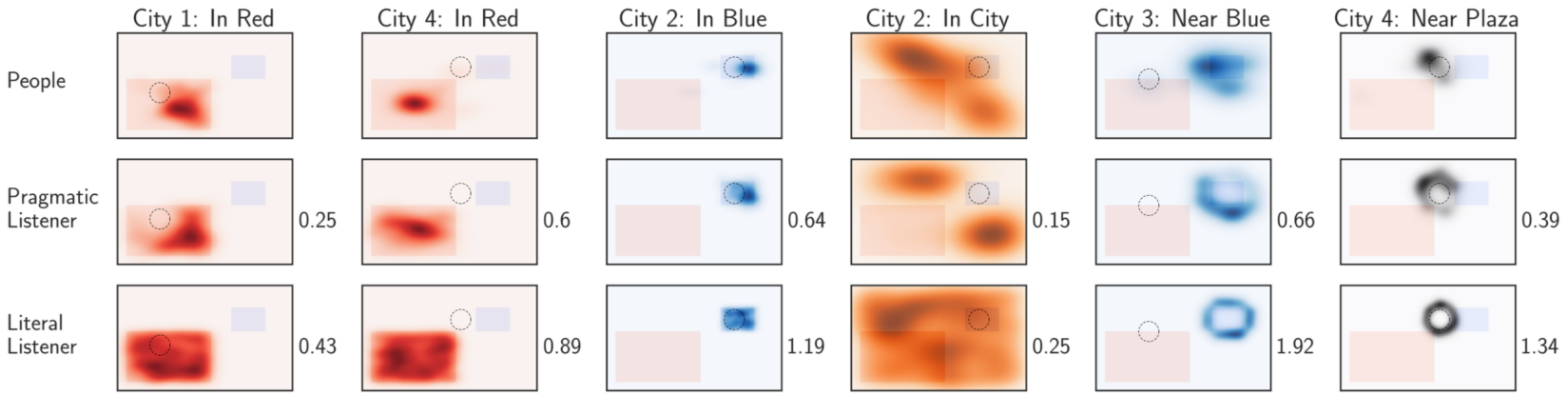
Human responses: Qualitative results

Near is non-uniform on edges



- The distribution of responses to *near* depends on context

Comparing people and pragmatic listeners

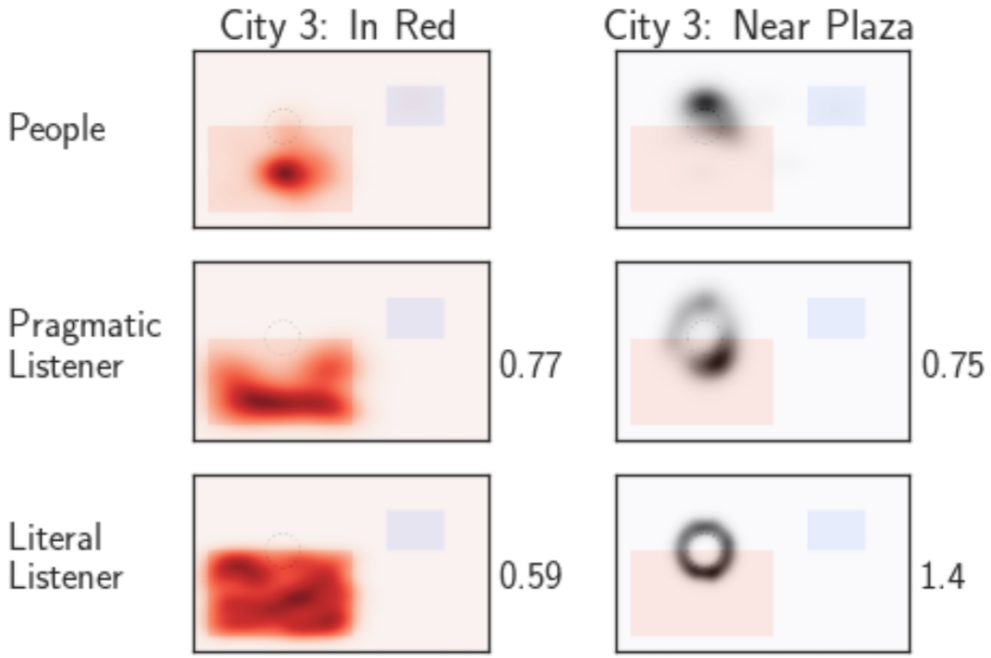


Comparing people and pragmatic listeners

- The ‘distance’ between the distributions for people, pragmatic listeners and people can be measured using the Kullback-Liebler divergence between them
- In 26 out of the 28 cases, it was found that the distance between the pragmatic listener and people was less than the distance between the literal listener and the pragmatic listener

Comparing people and pragmatic listeners

- Here's where it doesn't do so well



Takeaways

- Overall, the pragmatic listener does a good job capturing how humans reason pragmatically
- Avenues to explore further are:
  - Working with a larger, more natural vocabulary
  - Integrating the speaker's knowledge of the listener
  - Lifting the pragmatic slack, which is the allowance for many different values of distance of *near* in the literal listener

Number words

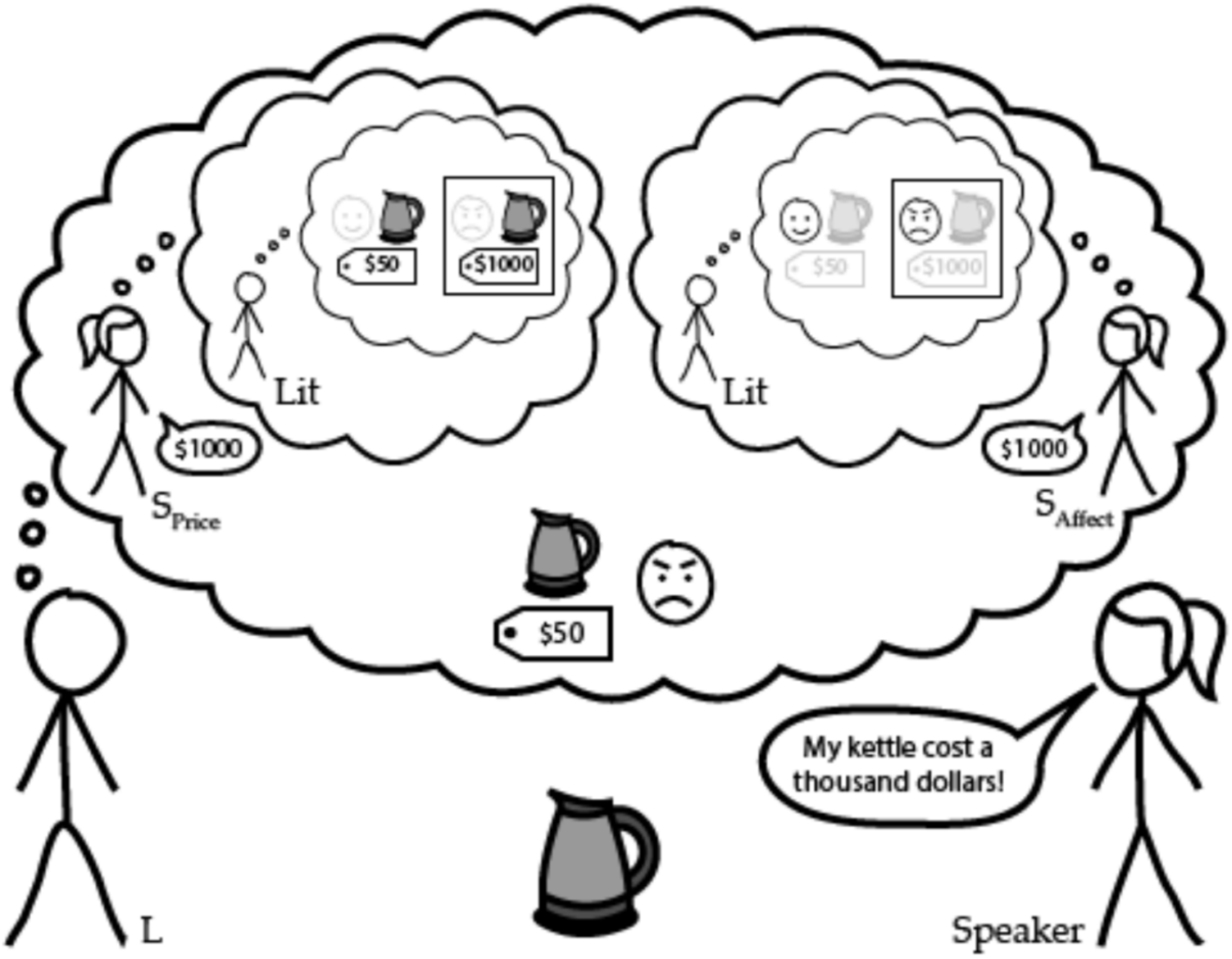
- When non-literal number words are used, the literal meaning directly contradicts the intended meaning
- Despite this, people understand what utterances like *'This presentation took me a thousand hours to make'* mean
  - It took a long time, but not actually a thousand hours

Number words

- This isn't supported by our current version of RSA, which helps resolve incomplete, but not contradictory meanings
- We can modify RSA by reasoning probabilistically about the speaker as well, and modelling the speaker trying to achieve goals along different dimensions

New Dimensions

- We introduce uncertainty about the speaker by introducing different dimensions, and having the pragmatic listener reason about which dimension/goal the speaker is trying to convey
- In the exact model we discuss, the dimensions are the price state and affect, the 'emotional response'


$50
$1000
Lit
$50
$1000
Lit
$1000
$1000
S Price
S Affect
$50
My kettle cost a thousand dollars!
L
Speaker

The Model: Literal listener

$$P_{Lit}(x|u) = \sum_{s',a'} \delta_{x=g(s',a')} P_{Lit}(s',a'|u)$$

where x is the topic of the discussion (a function of the dimensions the speaker is trying to optimise), and g is a function that maps meanings ($(state, affect)$ pairs) to topics (the goal function).

$$P_{Lit}(s,a|u) = \begin{cases} P_A(a|s) & if \;\; s = u \\ 0 & otherwise \end{cases}$$

where s is the number value, and a is the affect value (the other dimension), and P_A is the affective prior distribution (more on this soon).

The Model: Speaker

- Speaker's utility is a softmax decision rule over the utility of the utterances

$$P_S(u|s,a,g) \propto exp(U(u|s,a,g))$$
$$U(u|s,a,g) = log(P_{Lit}(g(s,a)|u)) - C(u)$$

  where C(u) is the cost of the utterance u

-

$$P_S(u|s,a,g) \propto \sum_{s',a'} \delta_{g(s,a)=g(s',a')} P_{Lit}(s',a'|u) \cdot exp(-C(u))$$

The Model: Pragmatic listener

The listener now accounts for the knowledge of the speaker's priors – the affective (P_A) and state (P_{state}) priors.

$$P_L(s, a|u) \propto \sum_g P_{state}(s) P_A(a|s) P_G(g) P_S(u|s, a, g)$$

where $P_G(g)$ is the probability of the goal being g, which is uniform.

Hyperbole and Halo

- We seek to model hyperbole – speakers exaggerating the values they are referring to – and halo – speakers interpreting 'sharp' numbers as round
- So, we formulate the goal functions as follows:

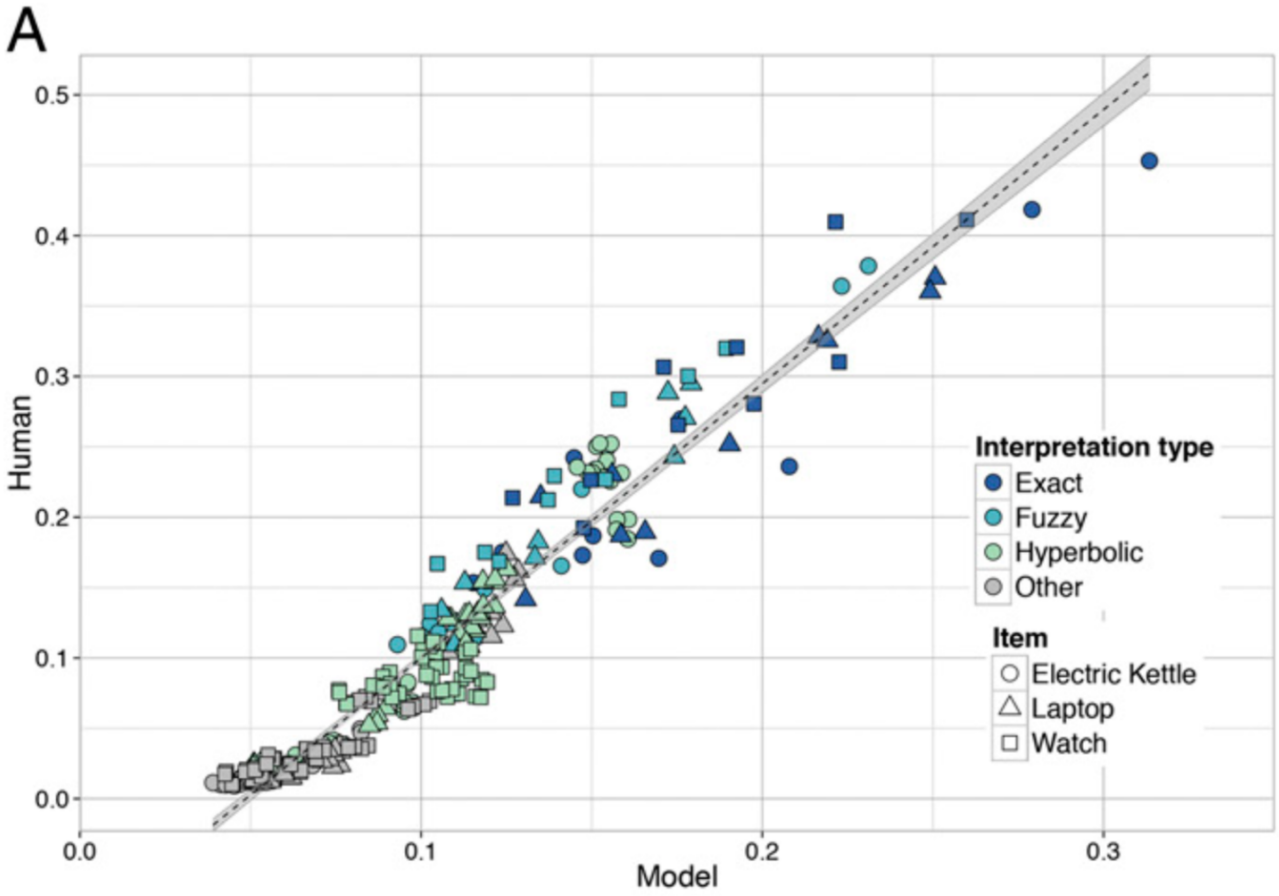
$$r_s(s,a) = s \ , \ r_a(s,a) = a \ , \ r_{s,a}(s,a) = s,a$$
$$f_{exact}(s) = s$$
$$f_{approx}(s) = Round(s)$$
$$f(s) = f_{exact}(s) \ or \ f(s) = f_{approx}(s)$$
$$g(s,a) = r_s(f(s),a) \ or \ g(s,a) = r_a(f(s),a) \ or \ g(s,a) = r_{s,a}(f(s),a)$$

Experiments[1]

- To test for hyperbole, participants were given a price, and asked if it was too expensive
- To test for halo, participants were given a price, and asked to say how likely it cost exactly that price
  - The price state is drawn from $\{50, 51, 500, 501, 1000, 1001, 5000, 5001, 10000, 10001\}$
- Additional experiments were conducted to create price state and affect priors
  - Participants rated the likelihood of an item having a certain cost to determine the price state prior
  - Participants rated the likelihood of an item being expensive to determine the affect prior

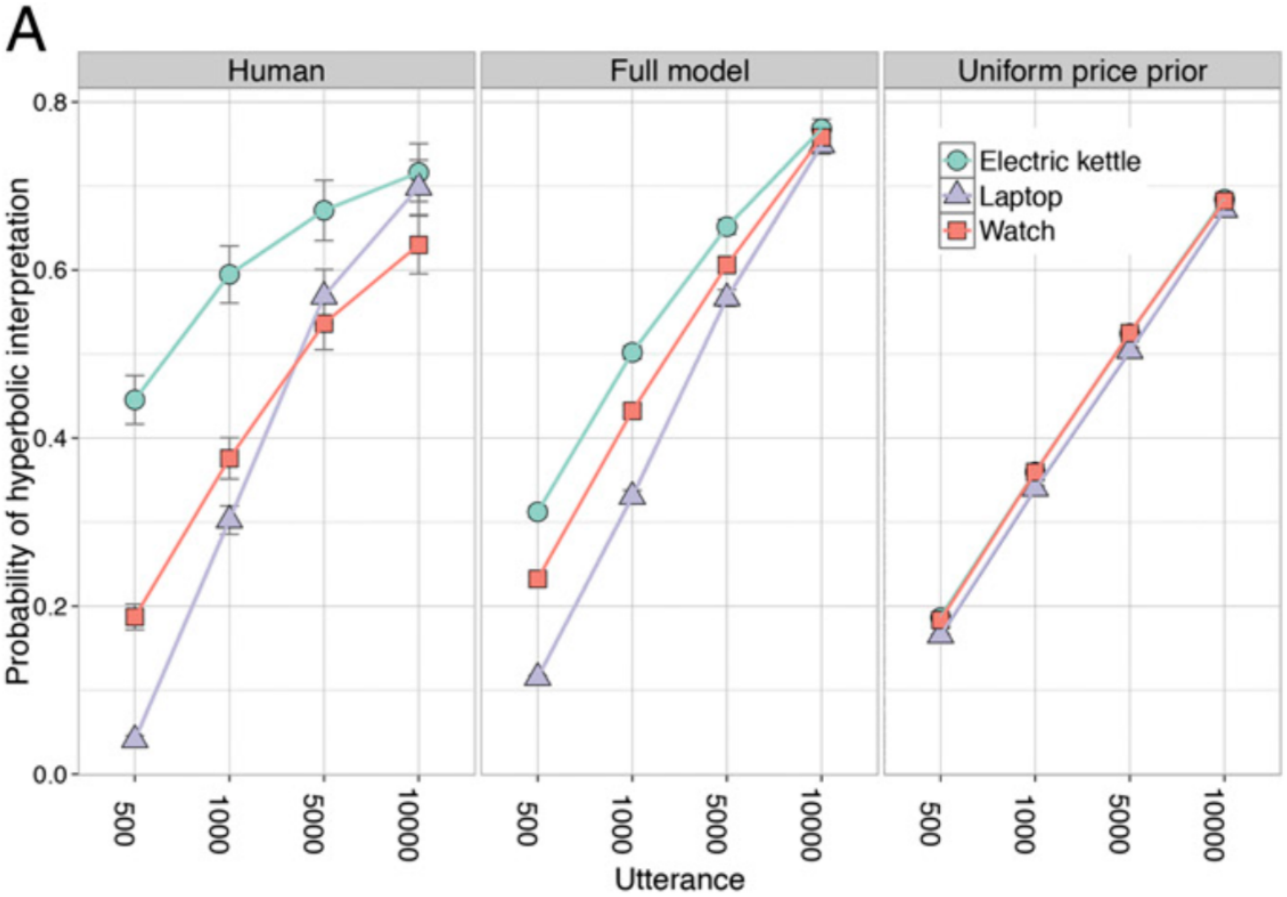
[1] Exact details of the experiments are a lot more nuanced. Please refer to the original paper for full details.

Comparing Humans and the Model



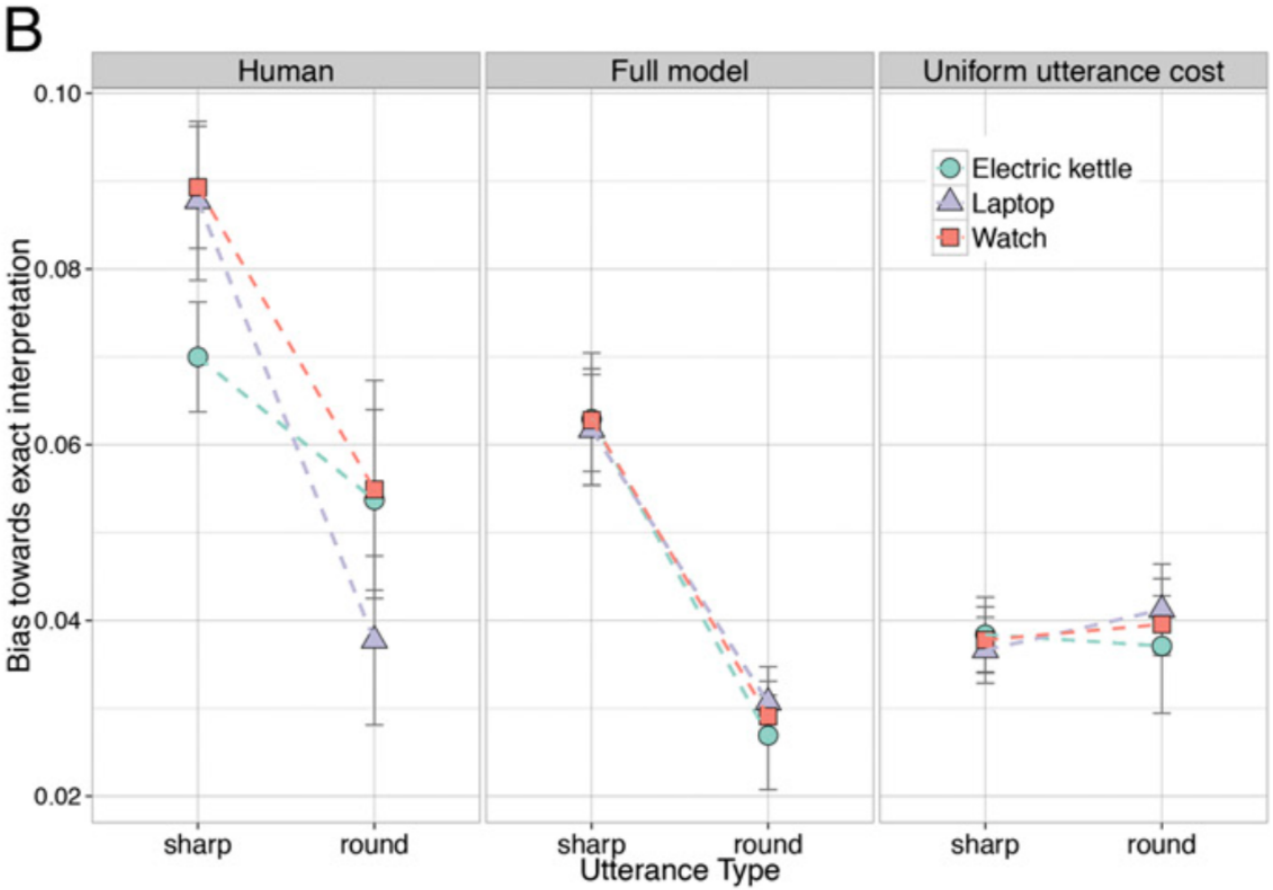
- There's a strong correlation (0.968) between predictions of the model and human responses

Comparing Humans and the Model: Hyperbole



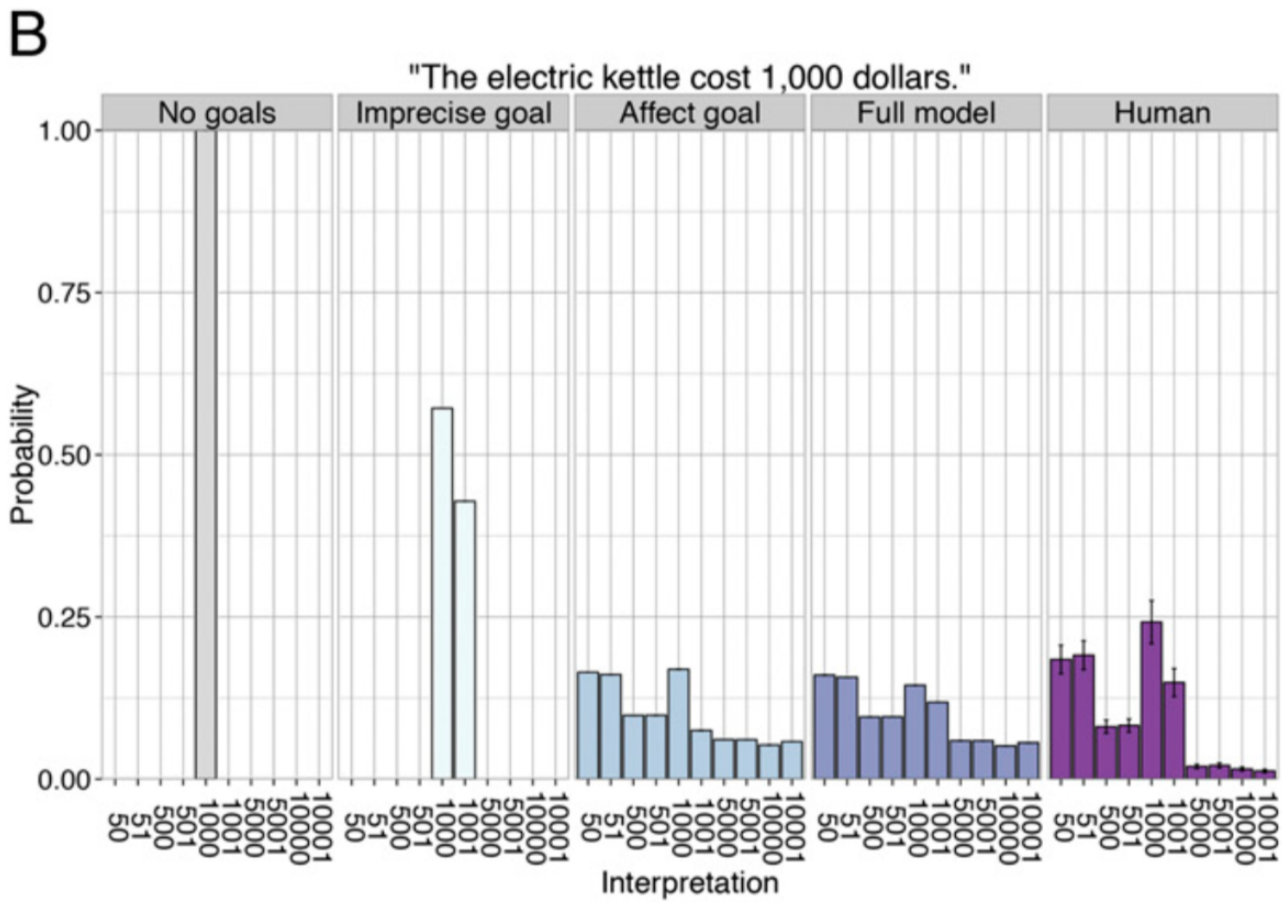
- Costs that are 1 apart collapsed for analysing hyperbole

Comparing Humans and the Model: Halo



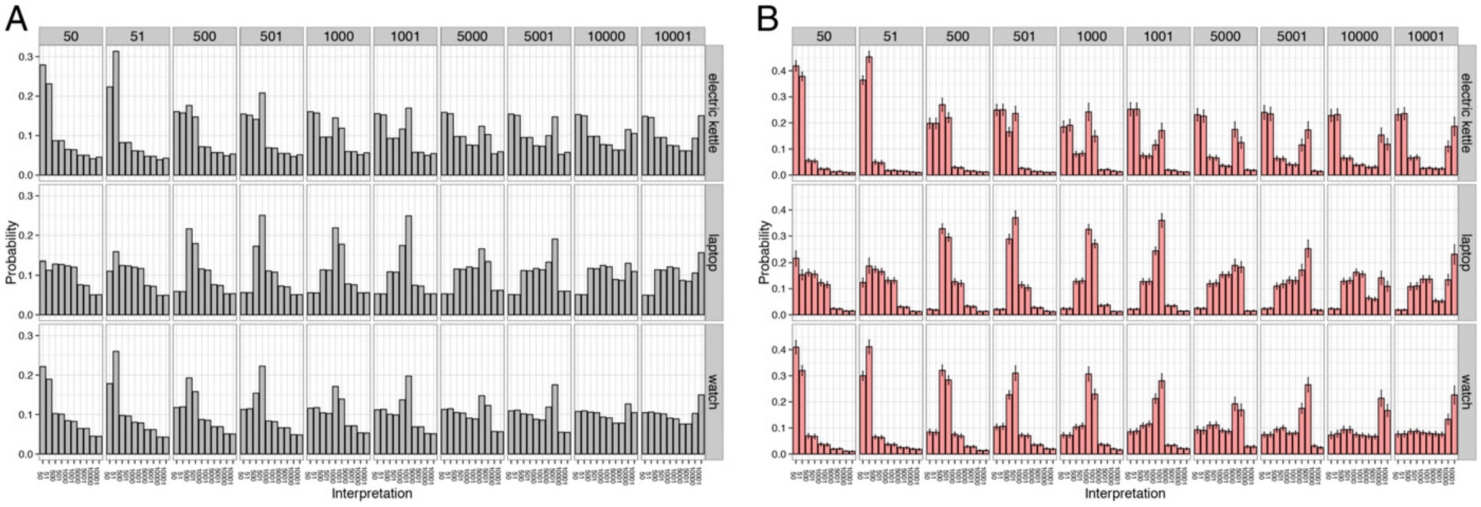
- RSA model is biased to interpreting sharp utterances exactly, like humans

Studying the Model: Dimensions



- Both affect and precision goals bring the model close to human behaviour

Comparing the full distributions



- Left panel is the full distributions across utterances for the model, and the right panel is the distributions for humans

Takeaways

- We see than an extension to the RSA model allows us to correctly model more complex pragmatic reasoning
- We can apply the framework to a speaker trying to optimize multiple goals
- This model performs very similarly to humans in interpreting number words

[1] Noah D Goodman and Michael C Frank.
Pragmatic language interpretation as probabilistic inference.
Trends in Cognitive Sciences, 20(11):818–829, 2016.

[2] H Paul Grice.
Logic and conversation.
1975, pages 41–58, 1975.

[3] Justine T Kao, Jean Wu, Leon Bergen, and Noah D Goodman.
Nonliteral understanding of number words.
Proceedings of the National Academy of Sciences, 111(33):12002–12007, 2014.

[4] Christopher Potts.
Linguist 130a - introduction to the rational speech acts model.
YouTube https://www.youtube.com/watch?v=bPd6CNy5UqA.

[5] Tomer D. Ullman, Yang Xu, and Noah D. Goodman.
The pragmatics of spatial language.
In *Proceedings of the Thirty-Eighth Annual Conference of the Cognitive Science Society*, 2016.