Computational Pragmatics

Nicholas Tomlin

w/ slides from Daniel Fried

Reasoning About Alternatives

Core Idea:

Large chunks of linguistic understanding can be attributed to reasoning about alternatives. E.g., if a speaker says X but not Y, then perhaps Y isn't true, or the speaker doesn't want to talk about Y.

Reasoning About Alternatives

Core Idea:

Large chunks of linguistic understanding can be attributed to reasoning about alternatives. E.g., if a speaker says X but not Y, then perhaps Y isn't true, or the speaker doesn't want to talk about Y.

Example:

"I didn't steal your car."

Reasoning About Alternatives

Core Idea:

Large chunks of linguistic understanding can be attributed to reasoning about alternatives. E.g., if a speaker says X but not Y, then perhaps Y isn't true, or the speaker doesn't want to talk about Y.

Example:

"I didn't steal your car."

Conveyed meaning:

Someone stole your car, but it wasn't me.

Reasoning About Alternatives

Core Idea:

Large chunks of linguistic understanding can be attributed to reasoning about alternatives. E.g., if a speaker says X but not Y, then perhaps Y isn't true, or the speaker doesn't want to talk about Y.

Example:

"I <u>didn't</u> steal your car."

Conveyed meaning:

Contrary to what you think, I did not steal your car.

Reasoning About Alternatives

Core Idea:

Large chunks of linguistic understanding can be attributed to reasoning about alternatives. E.g., if a speaker says X but not Y, then perhaps Y isn't true, or the speaker doesn't want to talk about Y.

Example:

"I didn't <u>steal</u> your car."

Conveyed meaning:

I did something to your car, but not stealing it. E.g., I just borrowed it.

Reasoning About Alternatives

Core Idea:

Large chunks of linguistic understanding can be attributed to reasoning about alternatives. E.g., if a speaker says X but not Y, then perhaps Y isn't true, or the speaker doesn't want to talk about Y.

Example:

"I didn't steal <u>your</u> car."

Conveyed meaning:

I stole somebody else's car.

Reasoning About Alternatives

Core Idea:

Large chunks of linguistic understanding can be attributed to reasoning about alternatives. E.g., if a speaker says X but not Y, then perhaps Y isn’t true, or the speaker doesn’t want to talk about Y.

Example:

“I didn’t steal your car.”

Conveyed meaning:

I stole something you own, but not your car.

Overview of Pragmatic Phenomena

"I ate some of the curry."

- *There is some curry leftover.*

Scalar Implicature

"The car was stolen."

- *The speaker doesn't know, or doesn't want to tell, who stole it.*

Conversational Implicature

"I stopped going to the office."

- *I used to go to the office.*

Presupposition

Scalar Implicature

The New York Times
@nytimes

We've deleted an earlier tweet and updated a sentence in our article that implied that only "some experts" view the ingestion of household disinfectants as dangerous. To be clear, there is no debate on the danger.

9:17 AM · Apr 24, 2020 · Twitter Web App

4.7K Retweets **22K** Likes

Scalar Implicature

Q: Does *some* mean *not all*?

A: Not always:

- "Some of the students were late for class; in fact, they all were."
- "I'd be much happier if some grocery stores had eggs in stock."

We call this *implicature*. The implicature occurs because a rational listener might assume that the speaker would have said *all* if they meant to, since *all* is the more informative choice.

Implicature ≠ Entailment

Implicatures are cancellable:

> “Some of the students were late for class; in fact, they all were.”

But presuppositions and entailments aren’t:

> “I stopped going into the office; in fact, I’ve never been there before.”
>
> “I stopped going into the office; in fact, I didn’t stop going in.”

Implicature ≠ Entailment

This distinction even shows up in perjury law (Bronston v. United States):

Q. “Do you have any bank accounts in Swiss banks, Mr. Bronston?”

A. “No, sir.”

Q. “Have you ever?”

A. “The company had an account there for about six months, in Zürich.”

Q. “Have you any nominees who have bank accounts in Swiss banks?”

A. “No, sir.”

Q. “Have you ever?”

A. “No, sir.”

Additional Phenomena

"The investor is a shark."

- *The investor is cunning/aggressive.*

Metaphor

"He went to the bank, the grocery store, and the mall."

- *He visited each place in that order.*

Ordering

"Class will begin at 2pm."

- *Class will begin around 2pm.*

Loose Use

Gricean Maxims

Grice (1975) claims that speakers and listeners typically follow four maxims for cooperative communication.

1. Quantity – be as informative as possible, give as much information as needed, but no more
2. Quality - be truthful, and don't give information that is false or unsupported by evidence
3. Relation – be relevant, and say things that are pertinent to the discussion
4. Manner – be clear, brief, and orderly as possible; avoid unnecessary prolixity

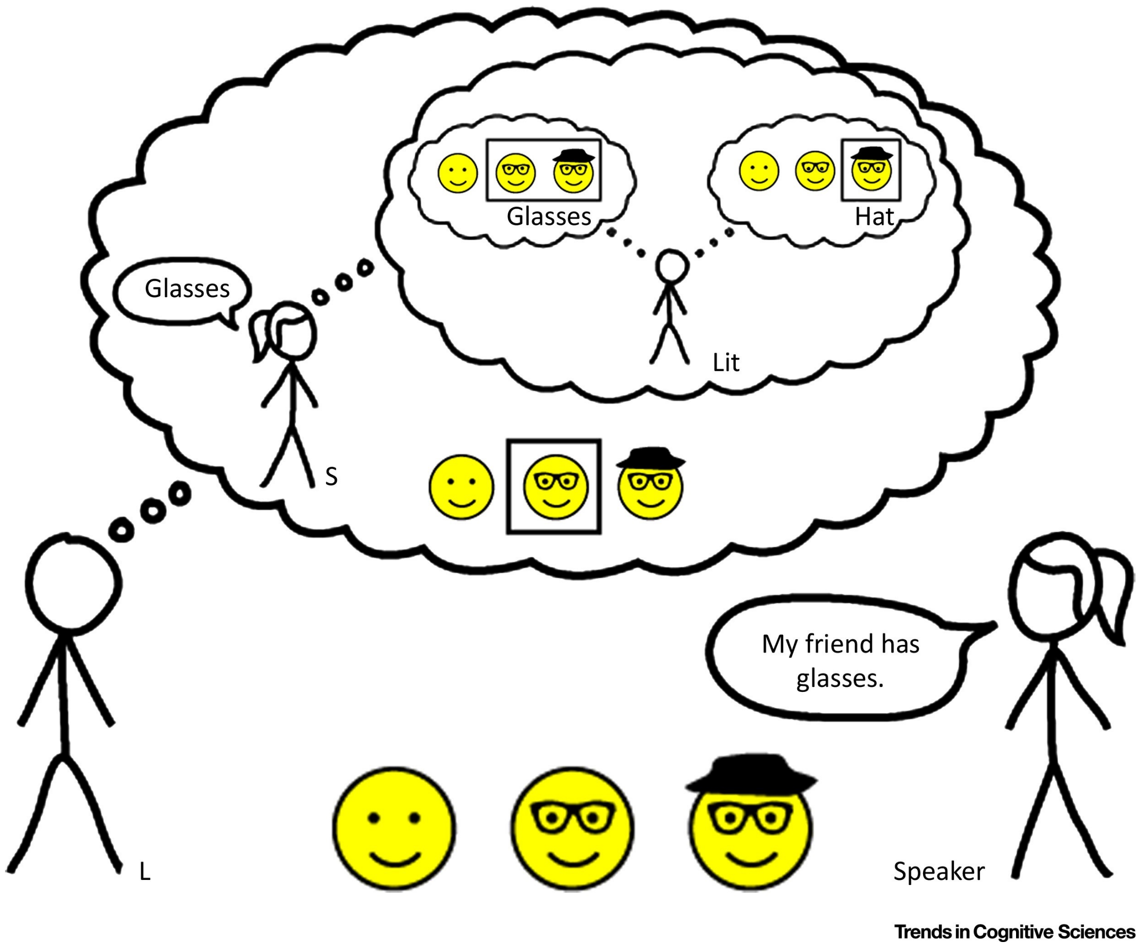
The Cooperative Principle

The Cooperative Principle (Grice 1975):

Make your 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.

Rational Speech Acts (RSA) Model

- Recursive reasoning between speakers and listeners about utterances and intentions
- Meant to operationalize the cooperative principle



Rational Speech Acts (RSA) Model

Base listener:

$$L_0(w, \mathrm{L} \mid \mathrm{msg}) \propto \mathrm{Lex}(\mathrm{msg}, w) \cdot \mathrm{P}(w)$$

Pragmatic speaker:

$$S_1(\mathrm{msg} \mid w, L) \propto \exp \lambda(\log L_0(w, \mathrm{L} \mid \mathrm{msg}) - \mathrm{C}(\mathrm{msg}))$$

Pragmatic listener:

$$L_1(w, \mathrm{L} \mid \mathrm{msg}) \propto \mathrm{S}_1(\mathrm{msg} \mid w, L) \cdot \mathrm{P}(w)$$

Rational Speech Acts (RSA) Model

Base listener:

State prior

$$L_0(\,w, \mathrm{L} \mid \mathrm{msg}\,) \propto \mathrm{Lex}(\mathrm{msg}, w) \cdot \mathrm{P}(w)$$

Lexicon

Pragmatic speaker:

Utterance cost

$$S_1(\,\mathrm{msg} \mid w, L\,) \propto \exp \lambda(\log L_0(\,w, \mathrm{L} \mid \mathrm{msg}\,) \; - \mathrm{C}(\mathrm{msg}))$$

Pragmatic listener:

State prior

$$L_1(\,w, \mathrm{L} \mid \mathrm{msg}\,) \propto \mathrm{S}_1(\mathrm{msg} \mid w, L) \cdot \mathrm{P}(w)$$

Rational Speech Acts (RSA) Model

Sample RSA Calculation: *Look at the man who is wearing glasses.*

	Glasses	Hat
	**1**	**0**
	**1**	**1**

L_2

S_1

L_0

Lex

Rational Speech Acts (RSA) Model

Sample RSA Calculation: *Look at the man who is wearing glasses.*

	Glasses	Hat
	**0.5**	**0**
	**0.5**	**1**

L_2

S_1

L_0

Lex

Rational Speech Acts (RSA) Model

Sample RSA Calculation: *Look at the man who is wearing glasses.*

	Glasses	Hat
	**1**	**0**
	**0.33**	**0.67**

L_2

S_1

L_0

Lex

Rational Speech Acts (RSA) Model

Sample RSA Calculation: *Look at the man who is wearing glasses.*

	Glasses	Hat
	**0.75**	**0**
	**0.25**	**1**

L_2

S_1

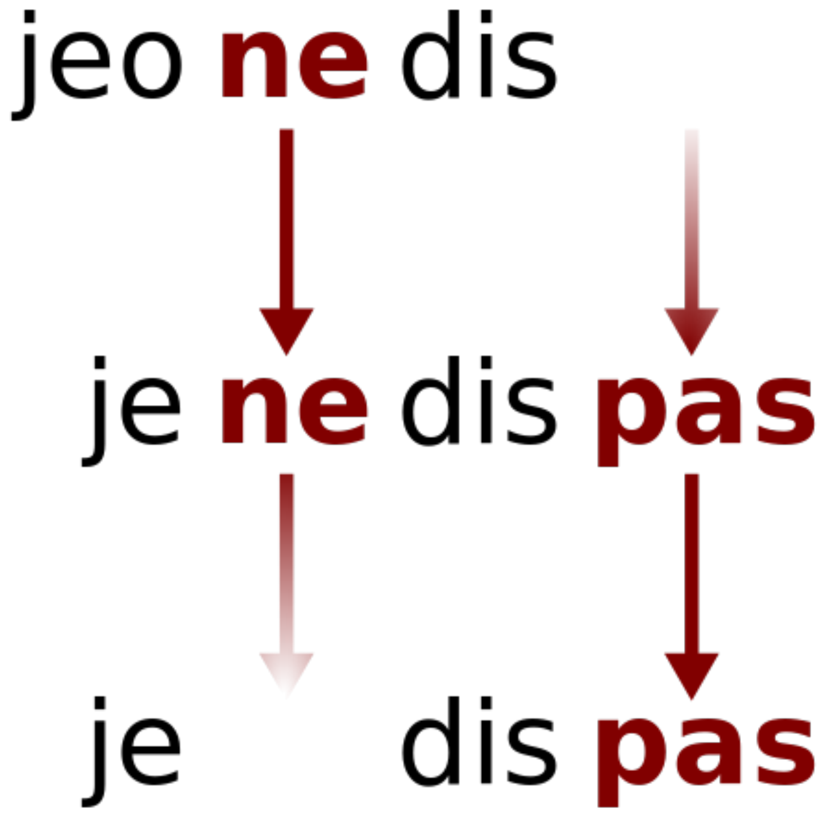
L_0

$\mathbf{Lex}$

Diachronic Pragmatics

Jespersen's Cycle describes a historical model of negation marking

Can be modeled as a pragmatic phenomena (Lund, et al. 2019) with a tradeoff between informativity (quantity) and brevity (manner)



Rational Speech Acts (RSA) Model

Base listener:

$$L_0(w, \mathrm{L} \mid \mathrm{msg}) \propto \underbrace{\mathrm{Lex}(\mathrm{msg}, w)}_{\text{Lexicon}} \cdot \overbrace{\mathrm{P}(w)}^{\text{State prior}}$$

Pragmatic speaker:

$$S_1(\mathrm{msg} \mid w, L) \propto \exp \lambda(\log L_0(w, \mathrm{L} \mid \mathrm{msg}) - \overbrace{\mathrm{C}(\mathrm{msg})}^{\text{Utterance cost}})$$

Pragmatic listener:

$$L_1(w, \mathrm{L} \mid \mathrm{msg}) \propto \mathrm{S}_1(\mathrm{msg} \mid w, L) \cdot \overbrace{\mathrm{P}(w)}^{\text{State prior}}$$

Issues with the RSA Model

Some issues with the Frank & Goodman (2012) model:

- Requires explicit lexicon for semantic evaluation
- Requires normalization over small set of alternative utterances and alternative meanings
- Doesn't account for real-world pragmatic phenomena like over-informative referring expressions, anticipatory implicatures, etc.
- No model of topic relevance

Learning in the RSA Model

Monroe & Potts (2015) propose a differentiable RSA model, without a fixed lexicon:

- Feature representation $\varphi(\text{msg}, w, L)$ and parameters θ, e.g.:

$$S_0(\text{msg} \mid w, L; \theta) \propto e^{\varphi(\text{msg}, w, L)}$$

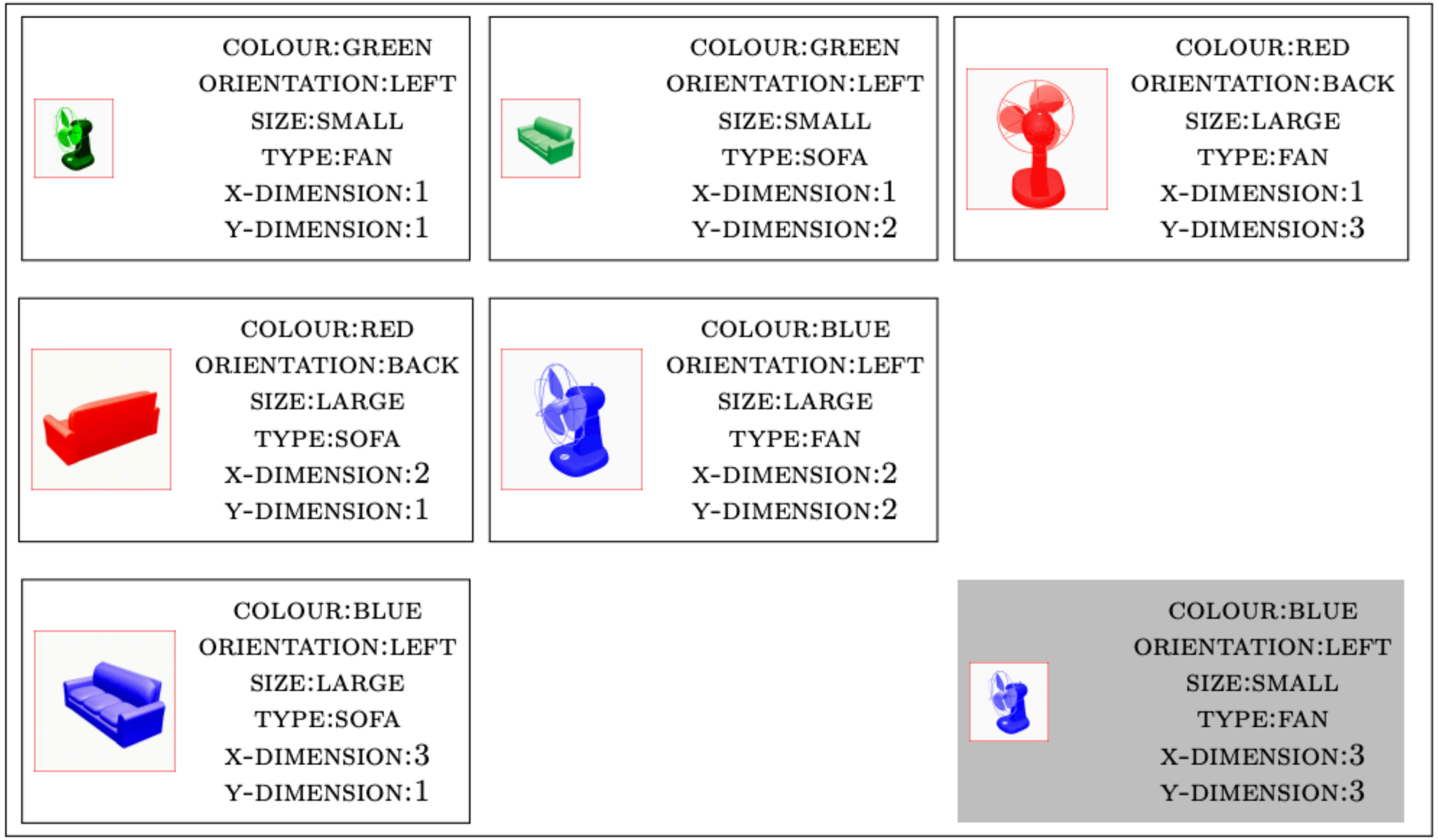
- Continue for layered models, and maximize probability of learned text under S_2 model

Learning in the RSA Model

Evaluate on TUNA Corpus of referring expressions:

- Given list of items with attributes and a target referent, generate a list of attributes needed to distinguish target item
- Modify feature representation by generating feature combinations
- Measure performance with multiset Dice

TUNA Corpus of Referring Expressions



Utterance: "blue fan small"
Utterance attributes: [*colour:blue*]; [*size:small*]; [*type:fan*]

Results on TUNA Corpus

Model	Furniture Acc.	Furniture Dice	People Acc.	People Dice	All Acc.	All Dice
RSA s_0 (random true message)	1.0%	.475	0.6%	.125	1.7%	.314
RSA s_1	1.9%	.522	2.5%	.254	2.2%	.386
Learned S_0, basic feats.	16.0%	.779	9.4%	.697	12.9%	.741
Learned S_0, gen. feats. only	5.0%	.788	7.8%	.681	6.3%	.738
Learned S_0, basic + gen. feats.	**28.1%**	**.812**	17.8%	.730	***23.3%***	***.774***
Learned S_1, basic feats.	23.1%	.789	11.9%	.740	17.9%	.766
Learned S_1, gen. feats. only	17.4%	.740	1.9%	.712	10.3%	.727
Learned S_1, basic + gen. feats.	***27.6%***	.788	**22.5%**	**.764**	**25.3%**	**.777**

Issues with the RSA Model

Some issues with the Frank & Goodman (2012) model:

- ~~Requires explicit lexicon for semantic evaluation~~
- Requires normalization over small set of alternative utterances and alternative meanings
- Doesn't account for real-world pragmatic phenomena like over-informative referring expressions, anticipatory implicatures, etc.
- No model of topic relevance

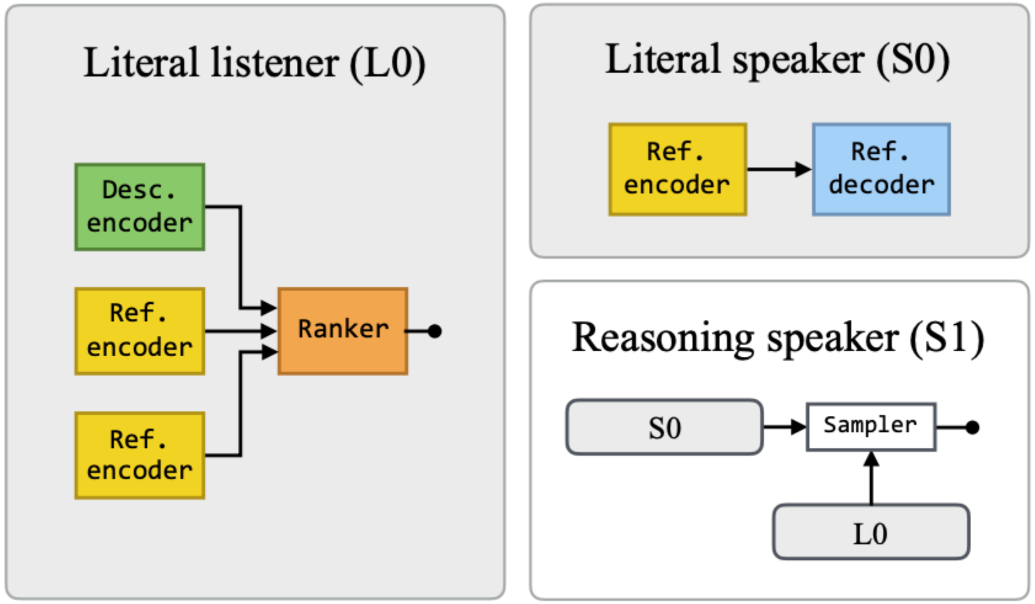
Neural RSA (Andreas & Klein, 2016)

Applies sampling-based method to address normalization over theoretically infinite set of potential utterances. Focuses on reference game task shown below:

(a) target (b) distractor

the owl is sitting in the tree

(c) description



Neural RSA (Andreas & Klein, 2016)

Despite worries about normalizing over entire set of potential utterances, the required number of samples levels off:

# samples	1	10	100	1000
Accuracy (%)	66	75	83	85

Table 1: S1 accuracy vs. number of samples.

Colors in Context

A brown dog and a tan one

[Young, et al. 2014; McMahan & Stone 2014]

Colors in Context

~~A brown dog and a tan one~~

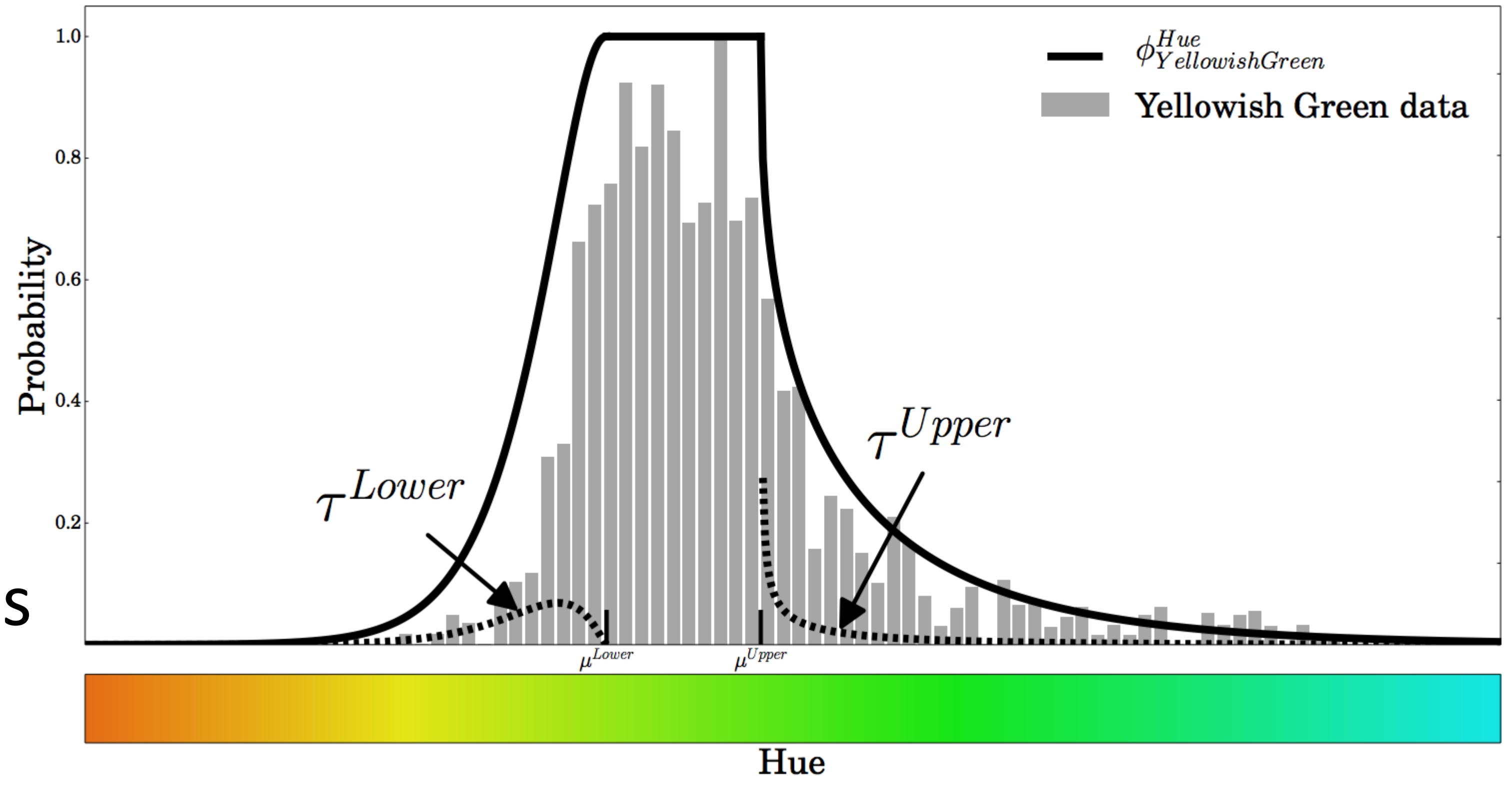
A tan dog and a white one

[Young, et al. 2014; McMahan & Stone 2014]

Colors in Context

- When we say "yellowish-green", what does that mean?

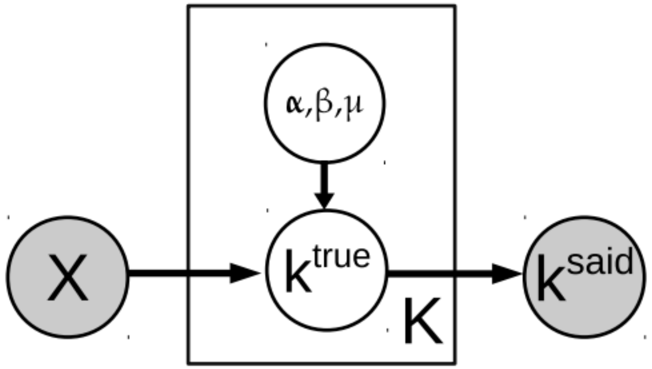
- Color descriptions governed by perception as well as *availability*: how commonly it is used (yellowish green vs. chartreuse)



McMahan and Stone (2014)

Colors in Context

- $P(k_{true} \mid X)$: distribution parameterized in HSV space as follows: there are certain ranges where a color can "definitely apply", others where it can apply



- $P(k_{said} \mid k_{true})$: captures availability; prior towards common colors

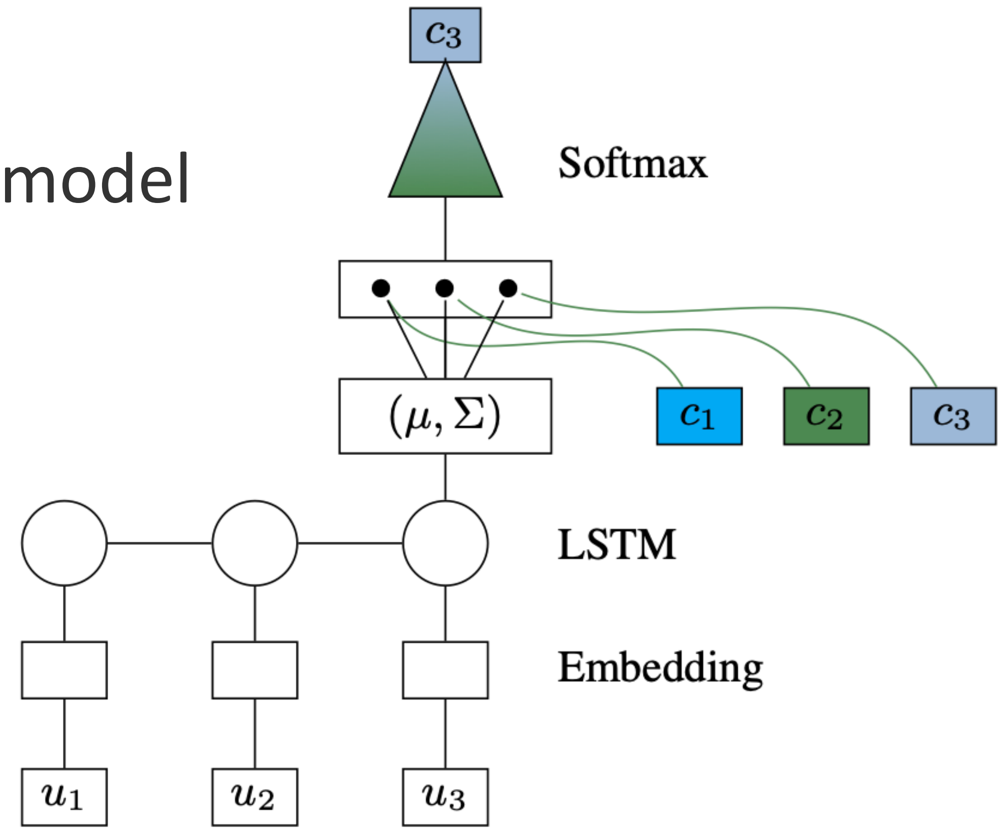
- Model combines language / reasoning with basic perception — characteristic of grounding

McMahan and Stone (2014)

Colors in Context

From the listener perspective: sample generations from a base speaker language model

	Context			Utterance
1.				darker blue
2.				Purple
3.				blue
4.				blue



Monroe, et al. (2017)

Issues with the RSA Model

Some issues with the Frank & Goodman (2012) model:

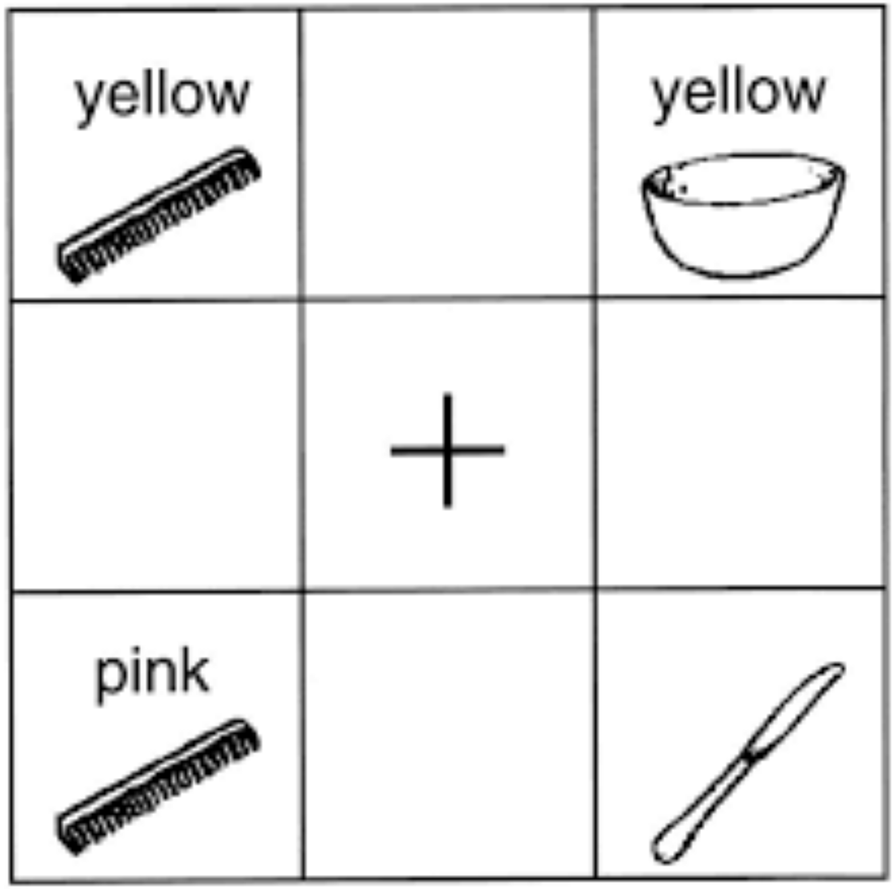
- ~~Requires explicit lexicon for semantic evaluation~~
- ~~Requires normalization over small set of alternative utterances and alternative meanings~~
- Doesn’t account for real-world pragmatic phenomena like over-informative referring expressions, anticipatory implicatures, etc.
- No model of topic relevance

Incremental Pragmatics

Incremental pragmatics is a well-motivated mechanism of human language processing.

Sedivy, et al. (1999):

- Target: “Touch the yellow bowl.”
- Before the word “bowl” is uttered, participants look more toward the comb instead of the bowl



Incremental RSA (Cohn-Gordon, et al.)

Cohn-Gordon, Goodman, & Potts (2018): Pragmatically Informative Image Captioning with Character-Level Inference

Cohn-Gordon, Goodman, & Potts (2019): An Incremental Iterated Response Model of Pragmatics

Pragmatic Image Captioning

Task: given multiple images, one of which is the target, write a caption to distinguish the target image from the others

Approach:

- Instead of sampling utterances, normalize over all possible characters and distractor images
- Use beam search decoding to generate optimal captions

Pragmatic Image Captioning

S_0 caption: the dog is brown
S_1 caption: the head of a dog

Pragmatic Image Captioning

S_0 caption: a double decker bus
S_1 caption: a red double decker bus

Issues with the RSA Model

Some issues with the Frank & Goodman (2012) model:

- ~~Requires explicit lexicon for semantic evaluation~~
- ~~Requires normalization over small set of alternative utterances and alternative meanings~~
- ~~Doesn't account for real-world pragmatic phenomena like over-informative referring expressions, anticipatory implicatures, etc.~~
- No model of topic relevance

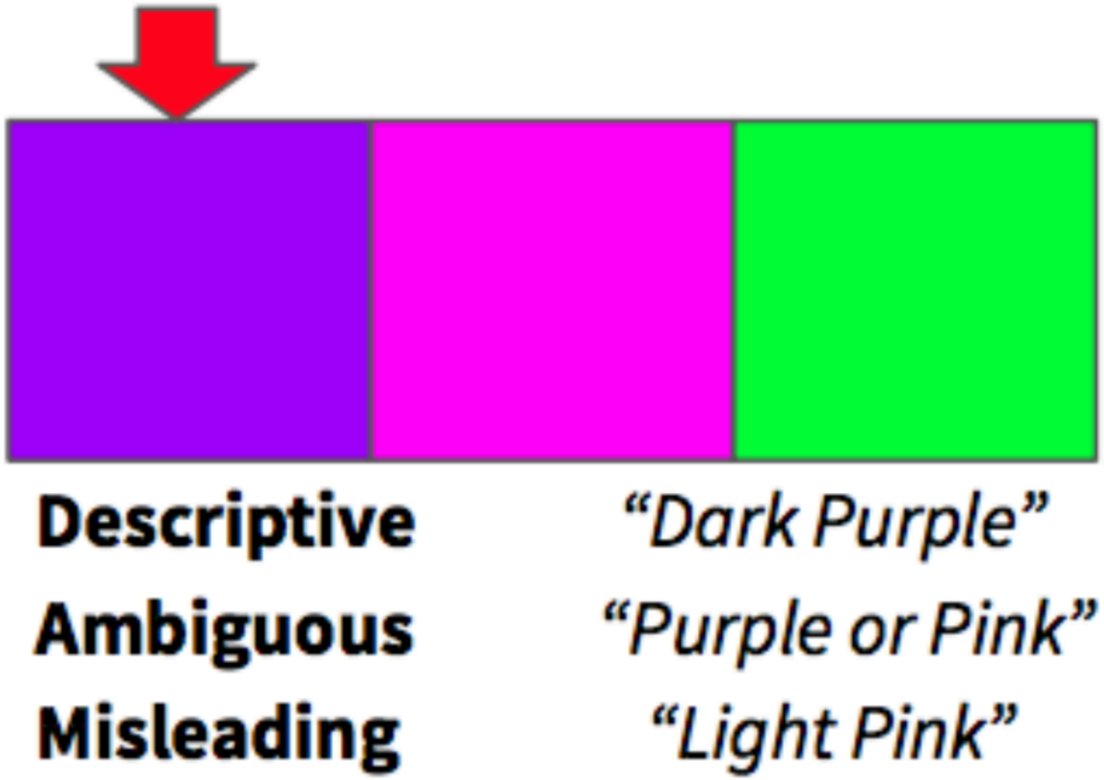
Issues with the RSA Model

Some issues with the Frank & Goodman (2012) model:

- ~~Requires explicit lexicon for semantic evaluation~~
- ~~Requires normalization over small set of alternative utterances and alternative meanings~~
- ~~Doesn't account for real-world pragmatic phenomena like over-informative referring expressions, anticipatory implicatures, etc.~~
- No model of topic relevance (no general solution yet)

RSA for NLG Evaluation (Newman, et al.)

Motivation: n-gram overlap evaluation metrics like BLEU and ROUGE don't capture utterance semantics or speaker intentions.



Task: Colors in Context (Monroe, et al. 2017)

Comparison of NLG Methods

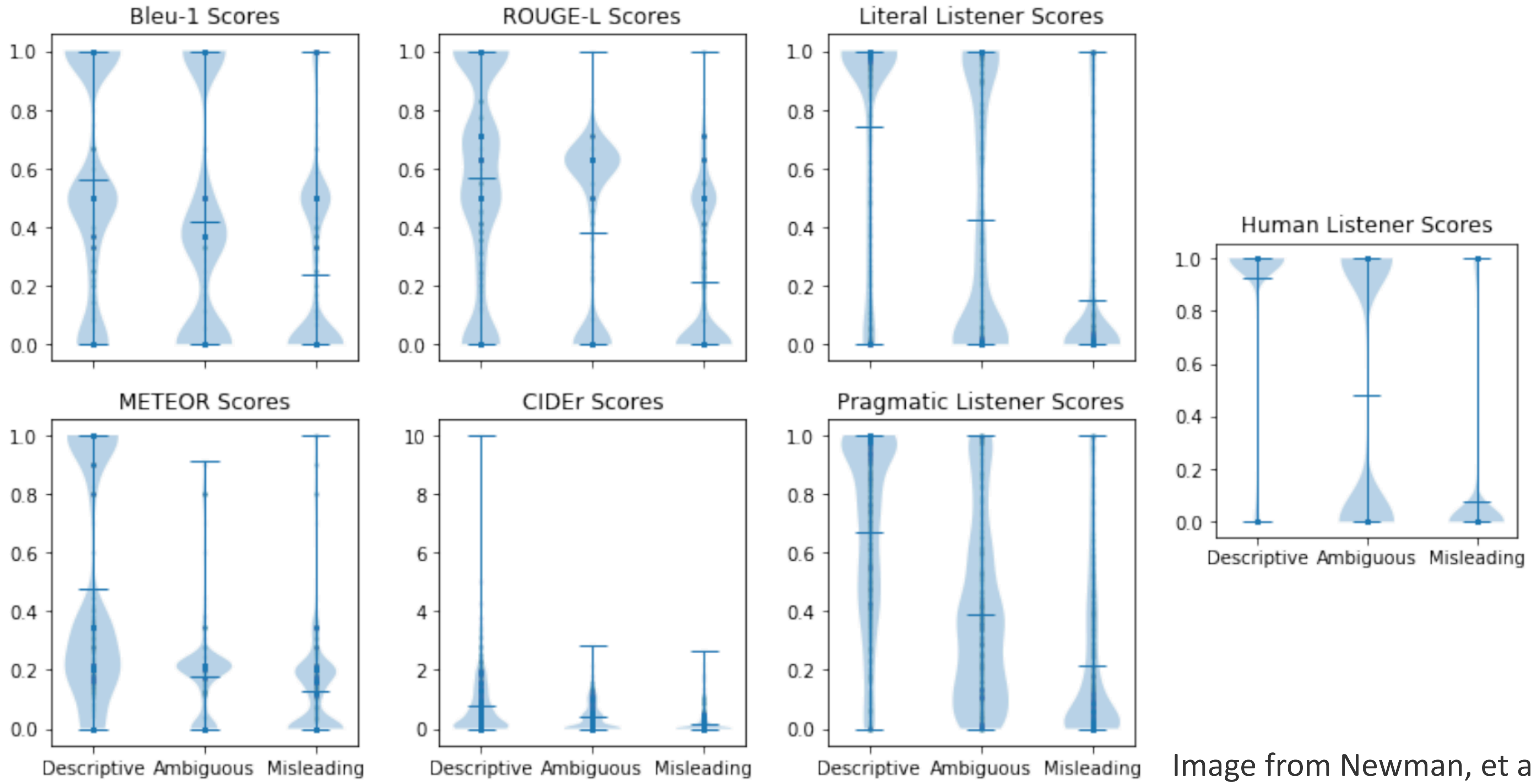


Image from Newman, et al. (2019)

Comparison of NLG Methods

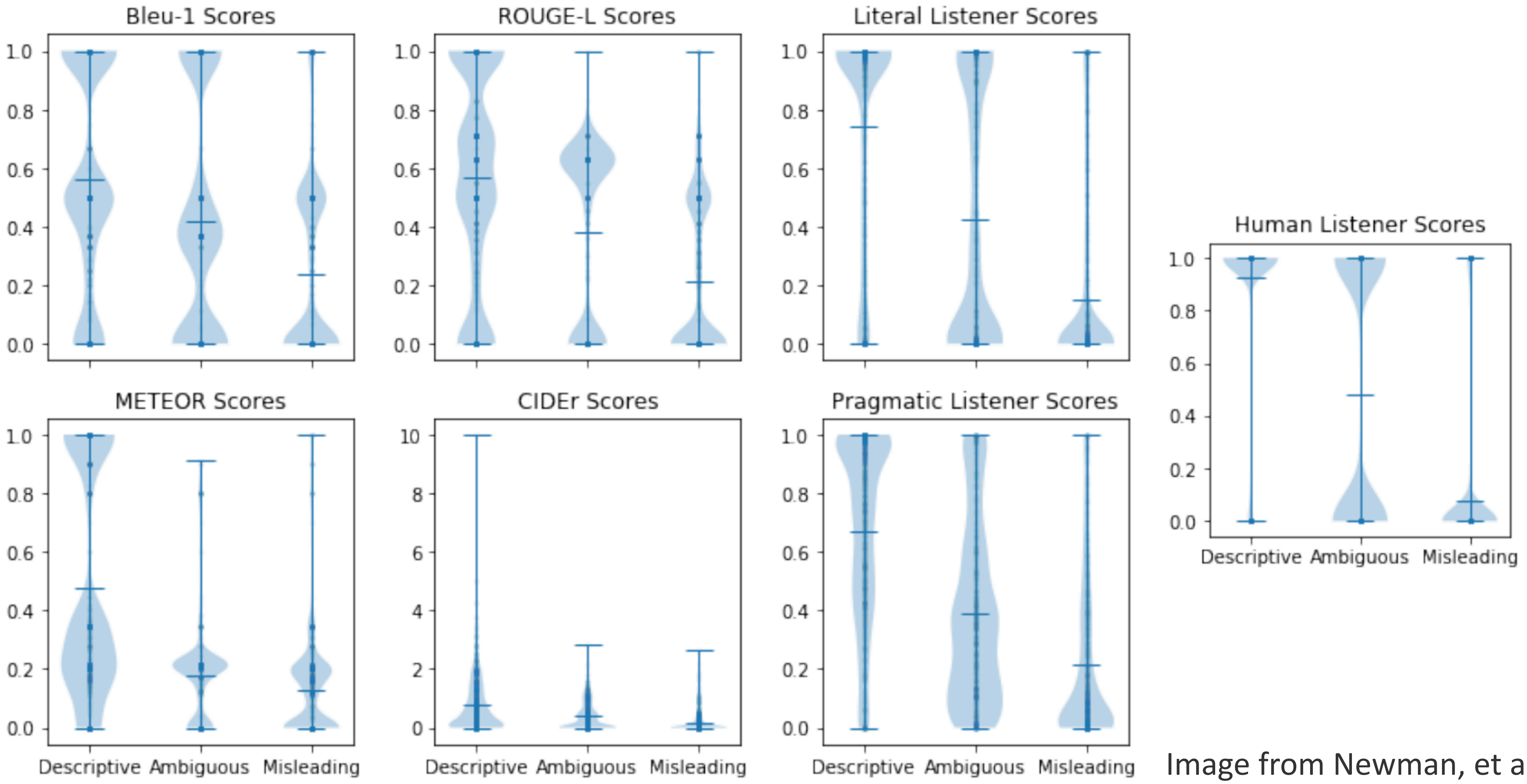


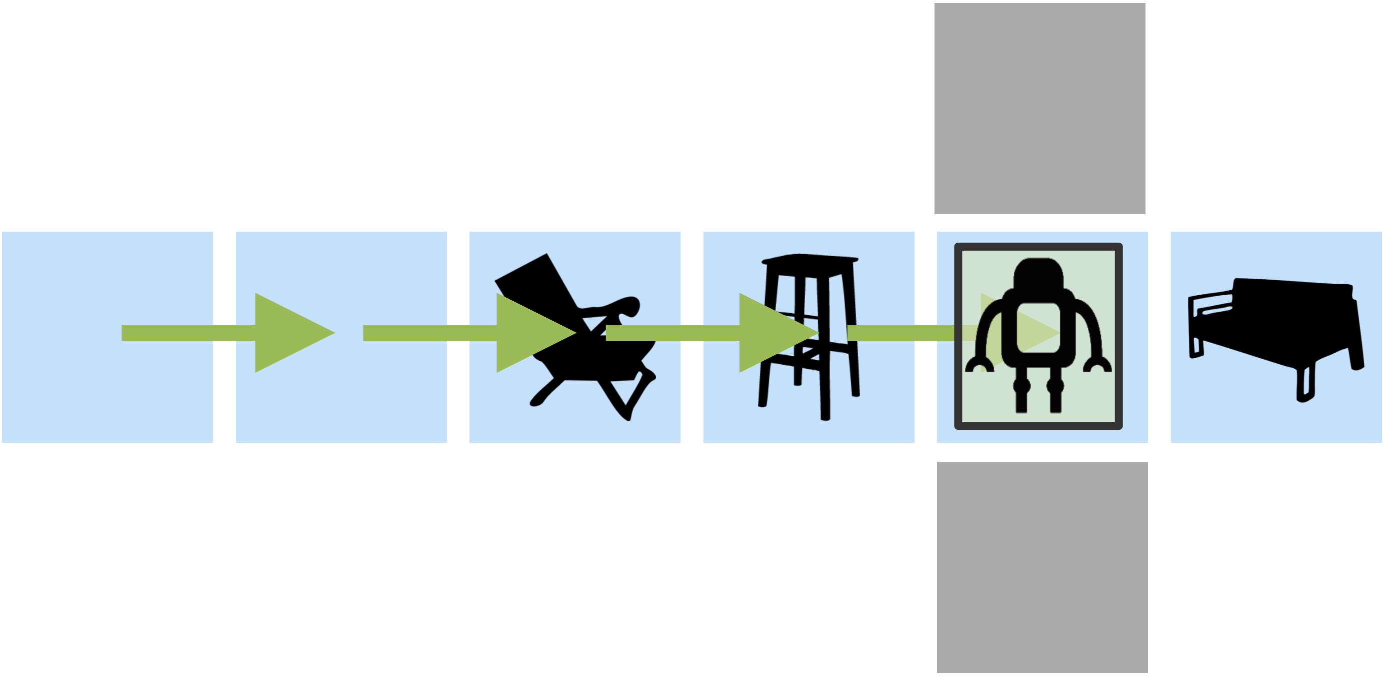
Image from Newman, et al. (2019)

Further Directions for RSA

- RSA for machine translation (Cohn-Gordon & Goodman 2019)
- RSA for summarization (Shen, et al. 2019)
- Neural RSA without sampling (McDowell & Goodman 2019)
- RSA-type models for dialogue faithfulness (Kim, et al. 2020)
- Explaining linguistic phenomena with RSA (Bergen, et al. 2016)

Instruction Generation

Input actions:



Output Instruction: *go forward to the grey hallway*

For SAIL: Daniele et al. 2016

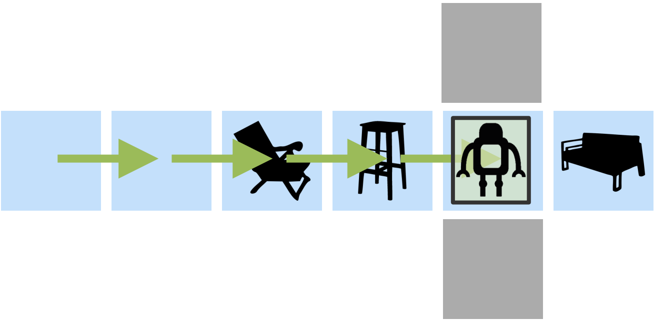
Neural Instruction Following

Inputs

Outputs

go forward to the grey hallway

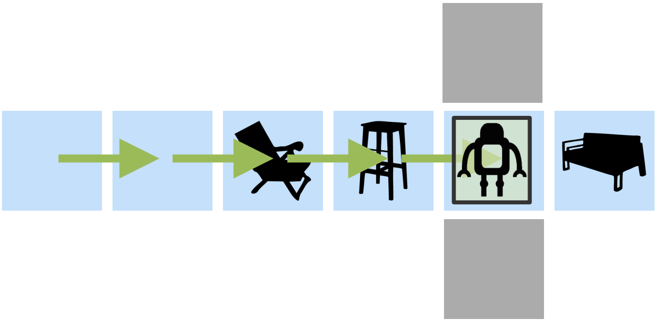
Listener



$P_L(a \mid i)$

Instruction, i

Actions, a



Speaker

go forward to the grey hallway

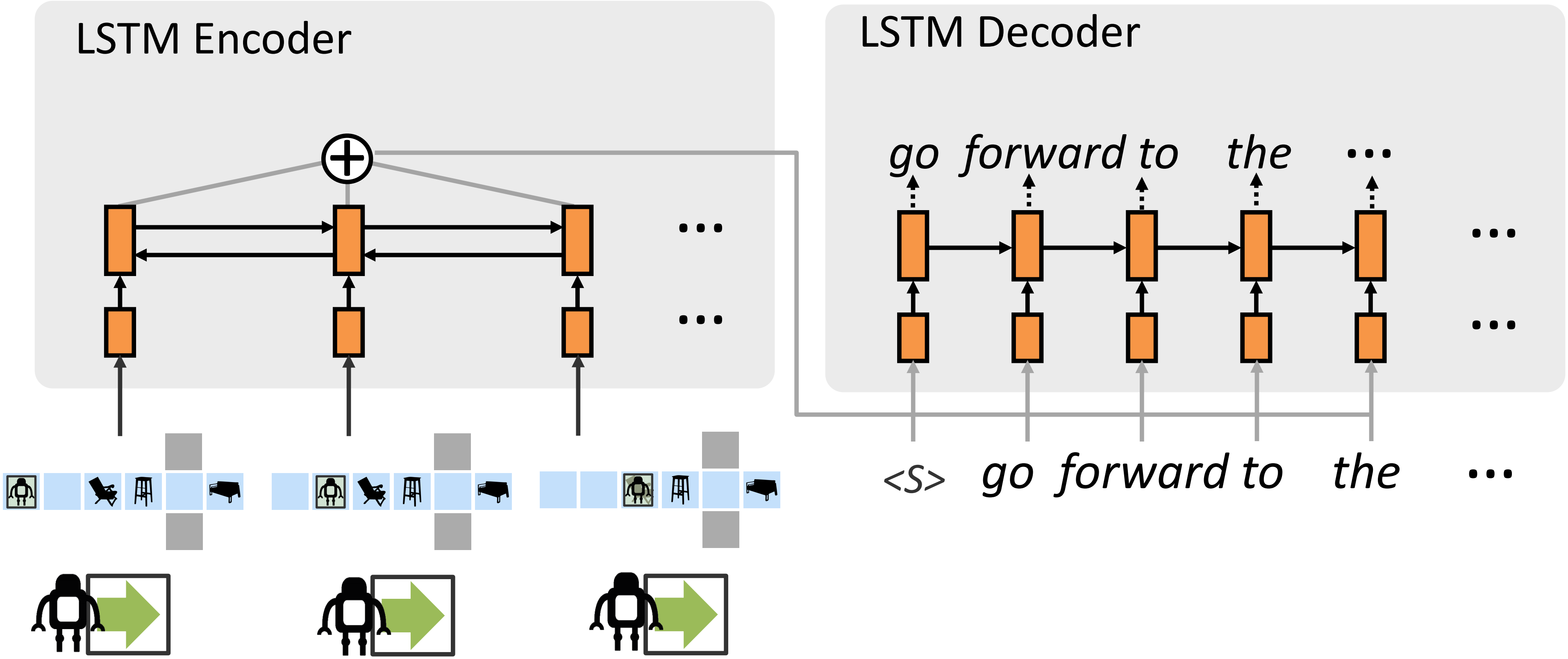
$P_S(i \mid a)$

Instruction, i

Actions, a

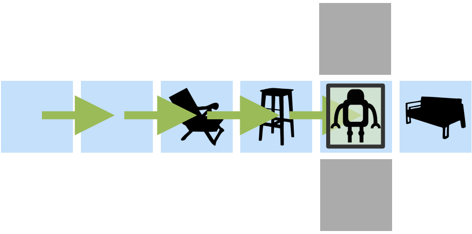
A Neural Speaker

Base Speaker $P_S(i \mid a) = \prod_t P_S(i_t | i_{1:t-1}, a)$



Fried et al. 2018

Building a Pragmatic Speaker



Base Speaker

proposes

0.4

walk forward past the stool

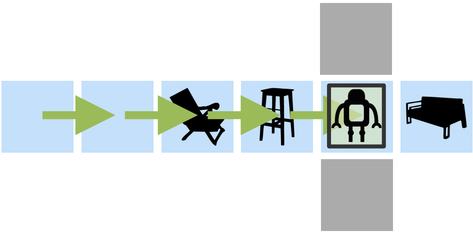
go forward four segments to the intersection with the bare concrete hall

Base Listener

rescores

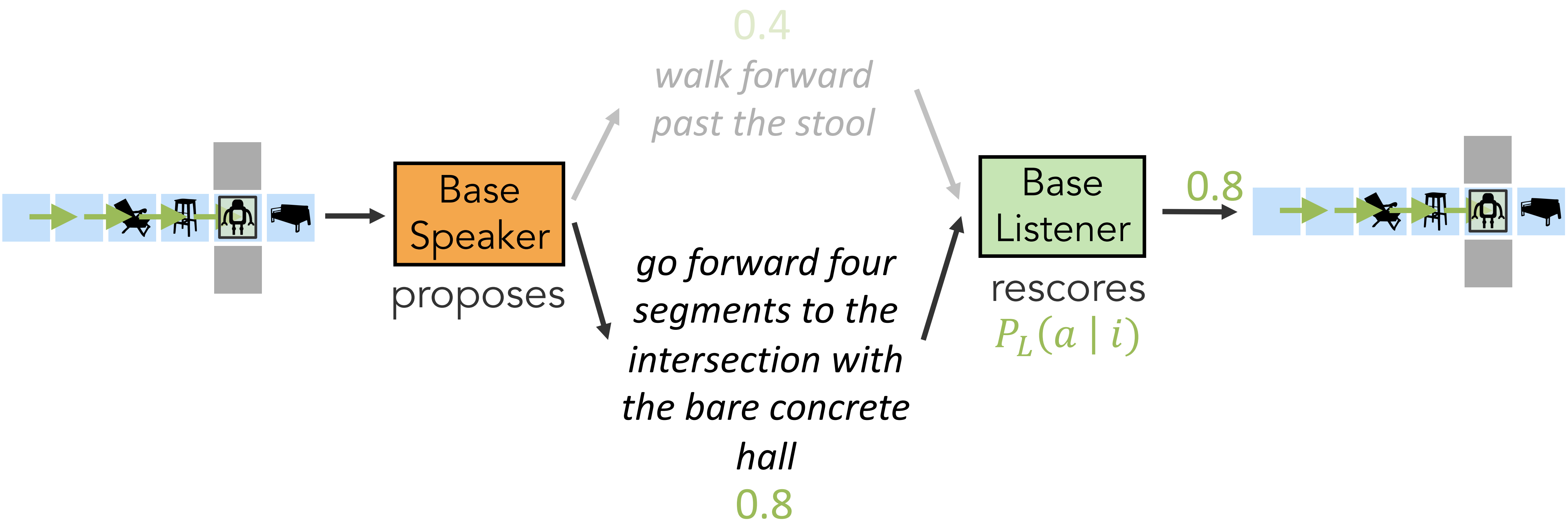
$P_L(a \mid i)$

0.4



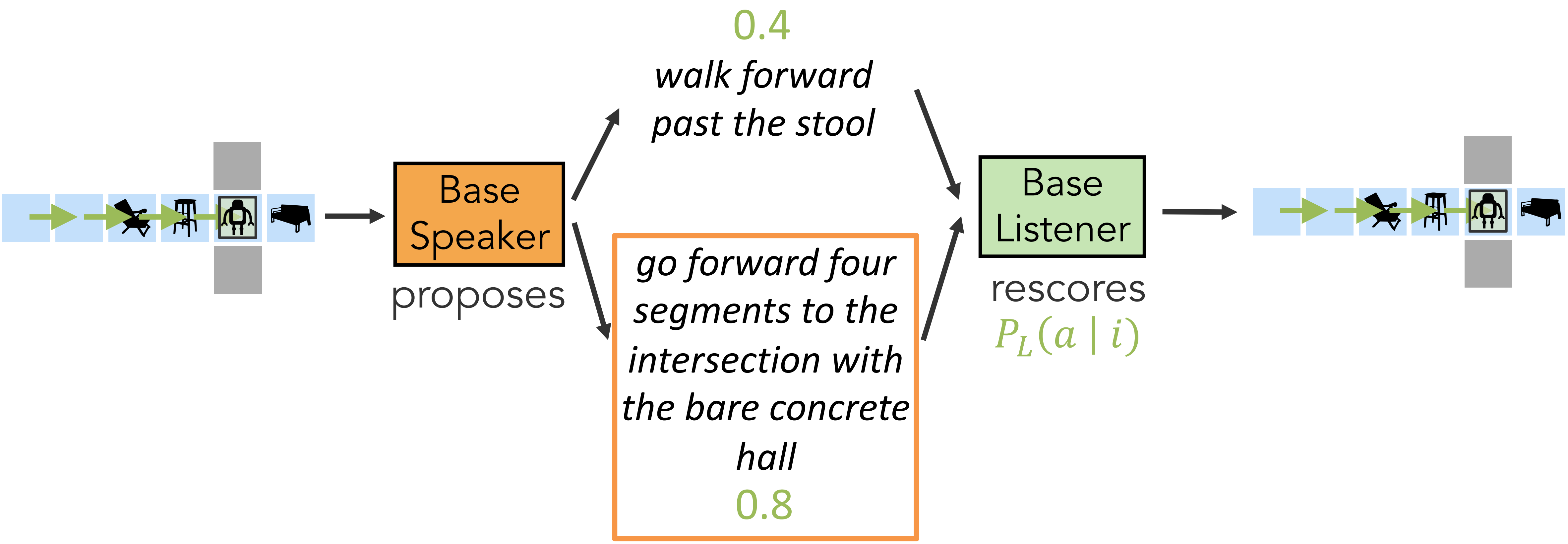
Fried et al. 2018

Building a Pragmatic Speaker



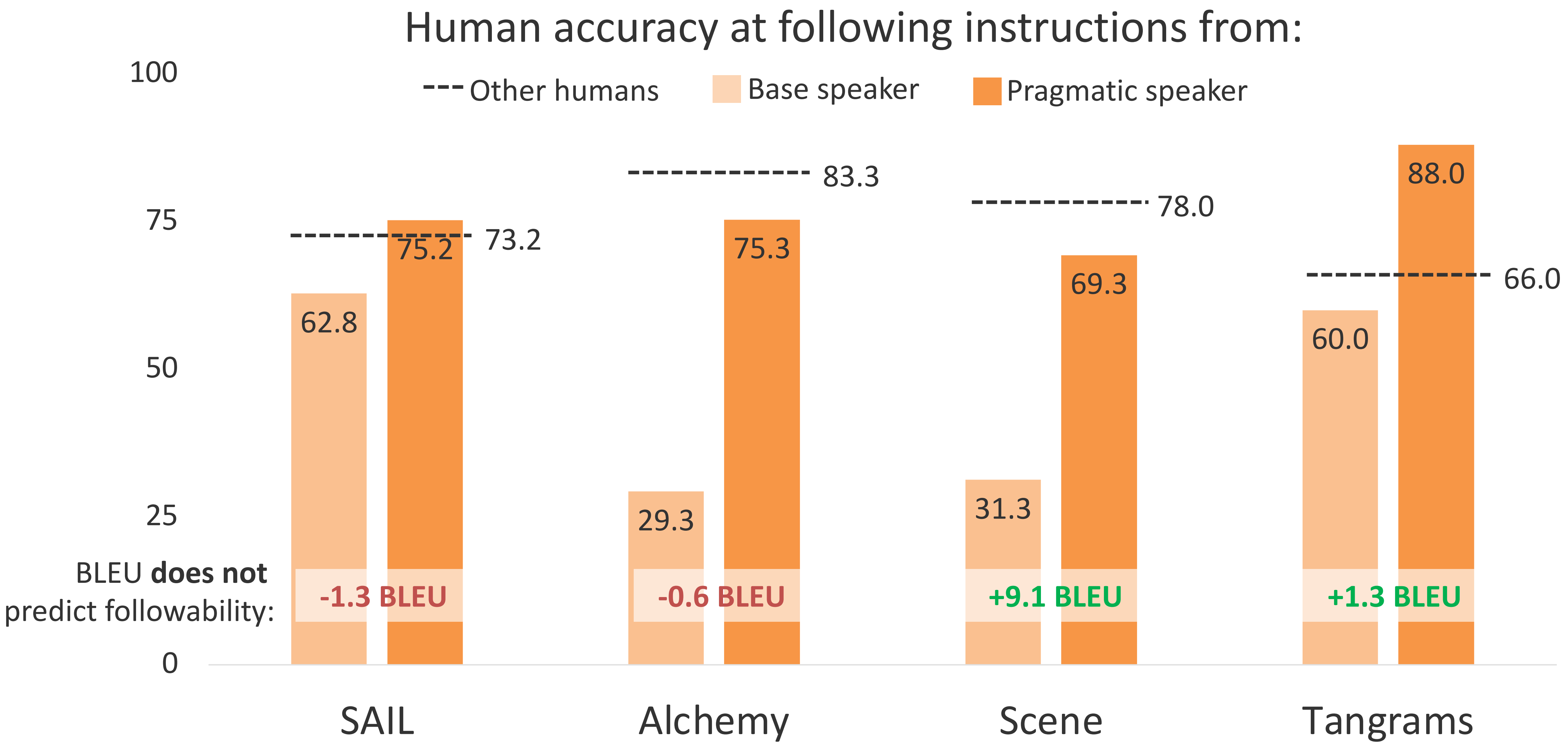
Fried et al. 2018

Building a Pragmatic Speaker

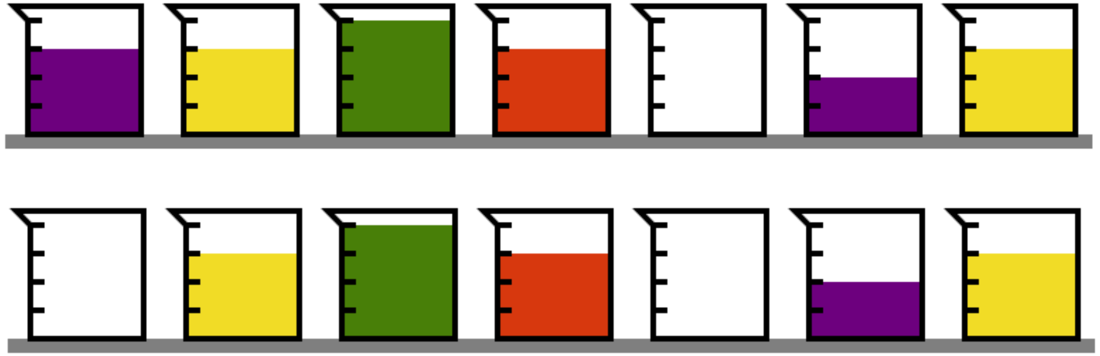


Fried et al. 2018

Building a Pragmatic Speaker

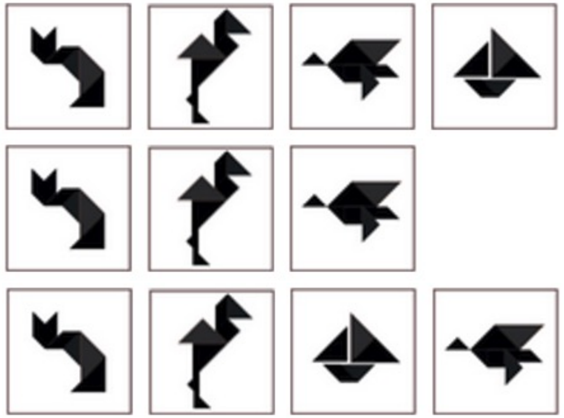


Pragmatics and Communicative Success



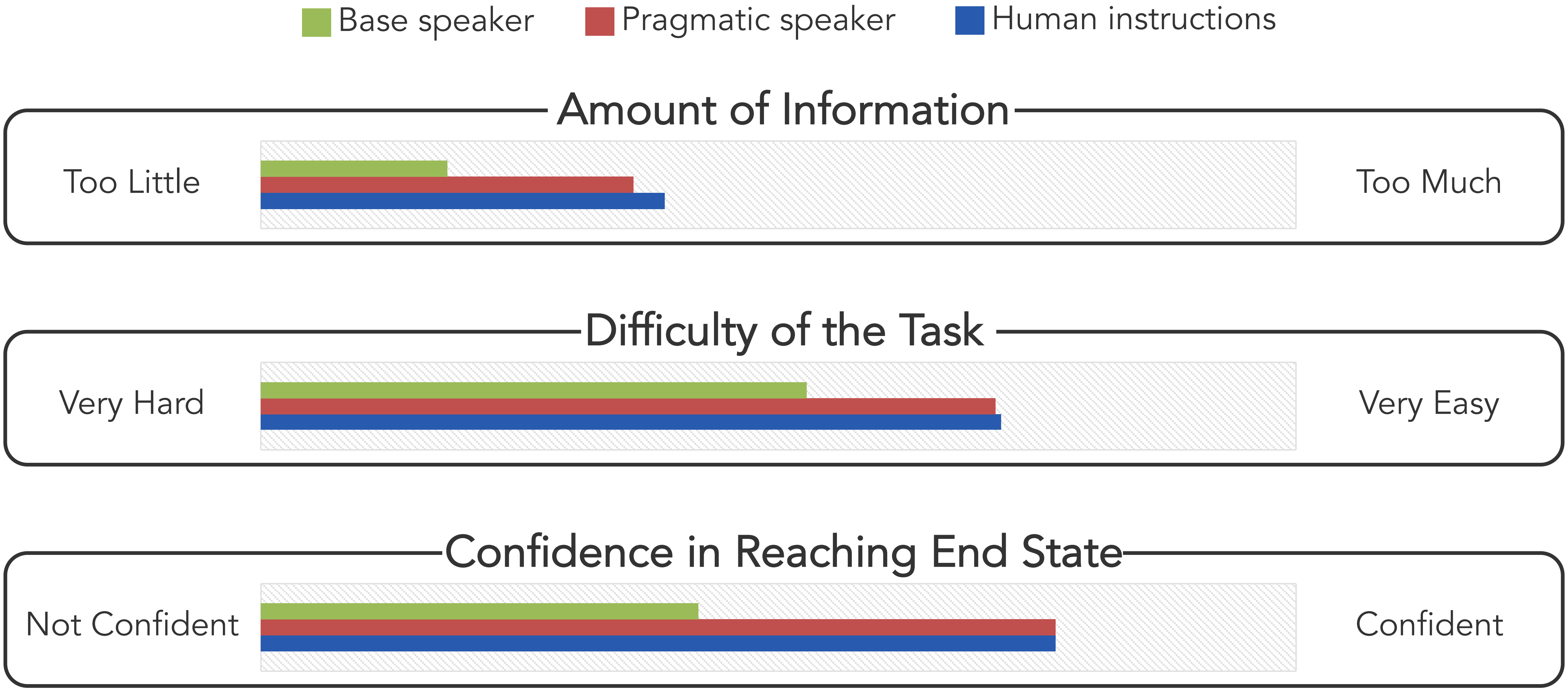
Base Speaker	*throw out the purple chemical*	✗
Pragmatic Speaker	*throw out the first purple chemical*	✓
Human	*remove all the purple chemical from the beaker on the far left*	✓

Pragmatics and Communicative Success



Base Speaker	*remove the last figure* *add it back*	✗
Pragmatic Speaker	*remove the last figure* *add it back in the 3rd position*	✓
Human	*take away the last item* *undo the last step*	✗

Pragmatics and Communicative Success



Averaged from 3 or 5 point Likert scales [Daniele et al. 2017]. Differences between base and pragmatic all statistically significant by χ^2 on counts.

Visually-Grounded Instructions

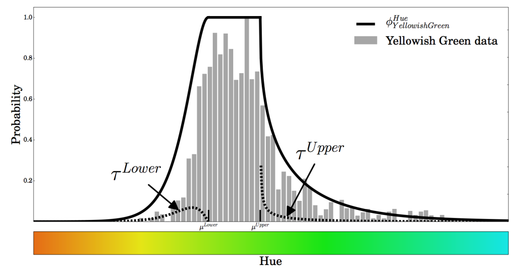
Human Description:
walk through the kitchen. go right into the living room and stop by the rug.

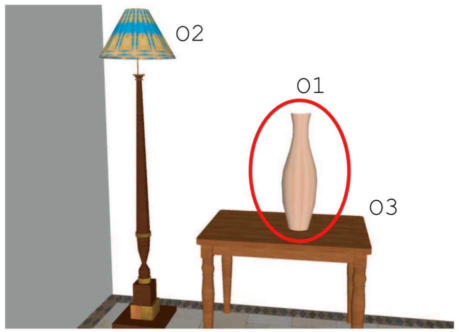
Base Speaker:
walk past the dining room table and chairs and wait there .

Pragmatic Speaker:
walk past the dining room table and chairs and take a right into the living room. stop once you are on the rug.

Connections to Semantic Parsing

- Each grounding framework requires mapping natural language to something concrete (distribution in color space, object, action sequence)
- Sometimes looks like semantic parsing, particularly when language -> discrete output
- Using linguistic structure to capture compositionality is often useful





go	to	the	chair
S	AP/NP	NP/N	N
$\lambda a.move(a)$	$\lambda x.\lambda a.to(a,x)$	$\lambda f.\iota x.f(x)$	$\lambda x.chair(x)$

$$\text{the chair} \xrightarrow{>} NP: \iota x.chair(x)$$

$$\text{to the chair} \xrightarrow{<} AP: \lambda a.to(a, \iota x.chair(x))$$

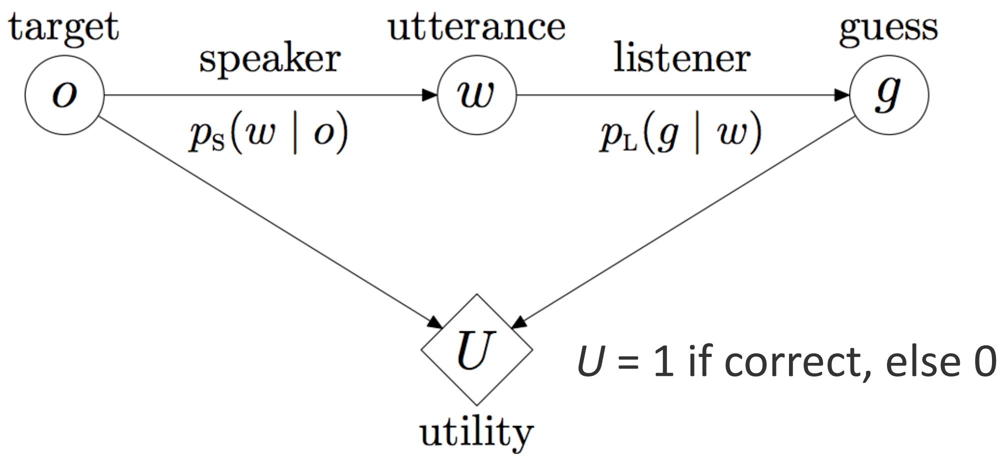
$$S\backslash S: \lambda f.\lambda a.f(a) \wedge to(a, \iota x.chair(x))$$

$$\xrightarrow{<} S: \lambda a.move(a) \wedge to(a, \iota x.chair(x))$$

Spatial Relations

- Two models: a speaker, and a listener
- We can compute expected success:

$$\mathrm{EU}(\mathrm{S}, \mathrm{L}) = \sum_{o,w,g} p(o) p_{\mathrm{S}}(w|o) p_{\mathrm{L}}(g|w) U(o,g)$$



- Modeled after cooperative principle of Grice (1975) : listeners should assume speakers are cooperative, and vice-versa
- For a fixed listener, we can solve for the optimal speaker, and vice-versa

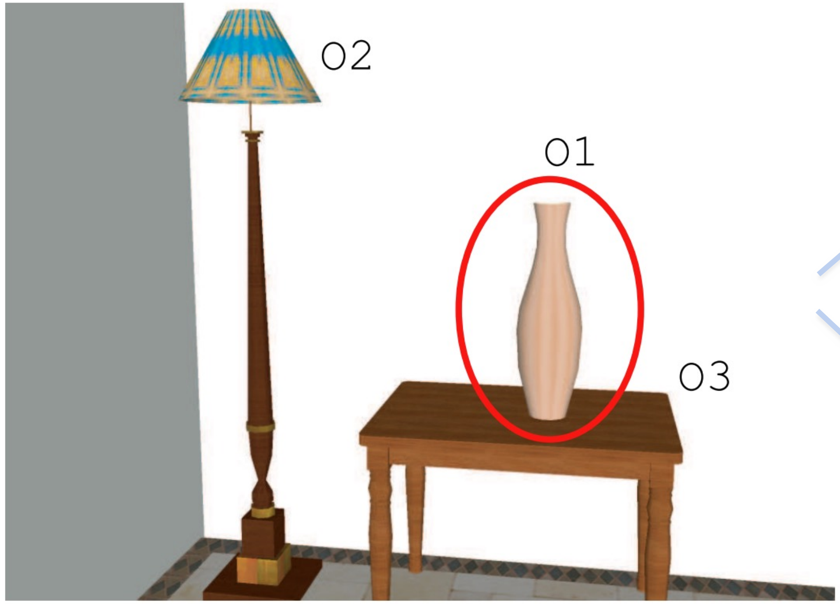
Golland et al. (2010)

Spatial Relations

- For a fixed listener, **L**, and a uniform prior p(o), we can solve for the optimal speaker, **S(L):**

$$S(L)(o) = argmax_w p_L(o|w)$$

- Visualize as a game tree:



S(L):

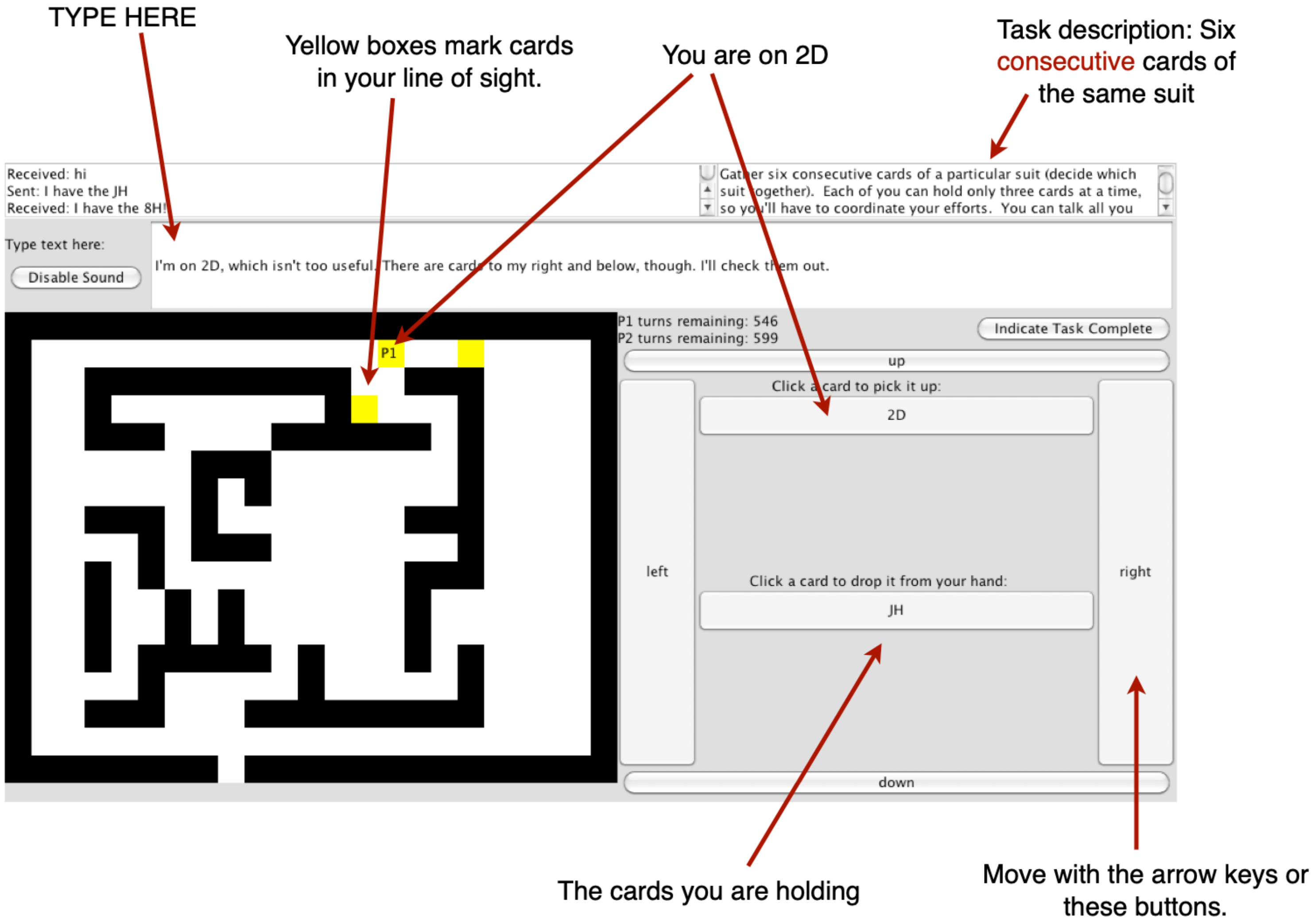
- Right of O2 → **L**
  - O1 (p_L=0.5)
  - O3 (p_L=0.5)
- On top of O3 → **L**
  - O1 (p_L=1.0)

Golland et al. (2010)

Challenge Tasks: Cards Corpus

Two players navigating a partially-observable environment must coordinate to collect a straight of cards

[Potts 2012; Vogel, et al. 2013]



Challenge Tasks: Among Us