Language Grounding

slides from: Greg Durrett, Daniel Fried, Chris Potts, Nick Tomlin

Announcement: Future Panels

- November 16: NLP beyond text
- November 21: NLP beyond English (remote only!)
- November 28 and 30: General topics in NLP and its future
- Add your questions here anytime! https://app.sli.do/event/xtY1jFFA5h9Ld1xcrsLYju

What is Language Grounding?

- Tying language to non-linguistic things (e.g. a database in semantic parsing)
- The world only looks like a database some of the time!
- Some settings depend on grounding into perceptual or physical environments:

"Add the tomatoes and mix"

"To get to BART, cross the street and keep going south toward the tall buildings…"

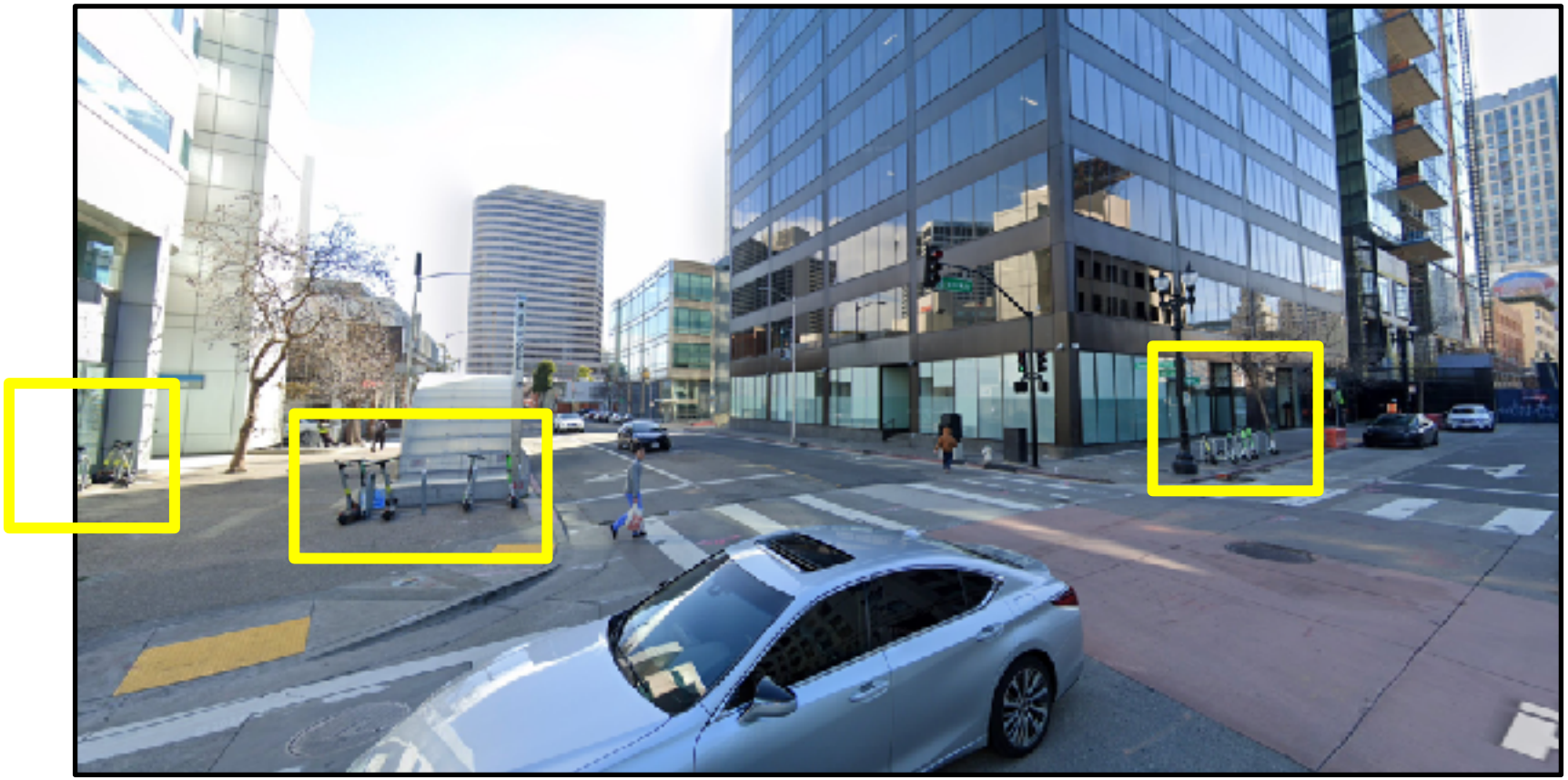
Grounded Semantics

What things, actions, etc. in the world does language refer to?

"The entrance is to the right of the bare tree in the sun"

Pragmatics

How does context shape the interpretation of language?



“The entrance is behind the scooters”

Using Language

Saying something will often… produce certain consequential effects upon the feelings, thoughts, or actions of the audience.

[*How to Do Things with Words.* Austin, 1962]

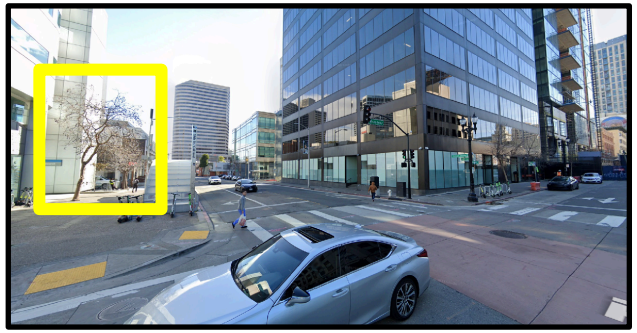
Our talk exchanges … are cooperative efforts… One of my avowed aims is to see talking as purposive, indeed rational, behavior.

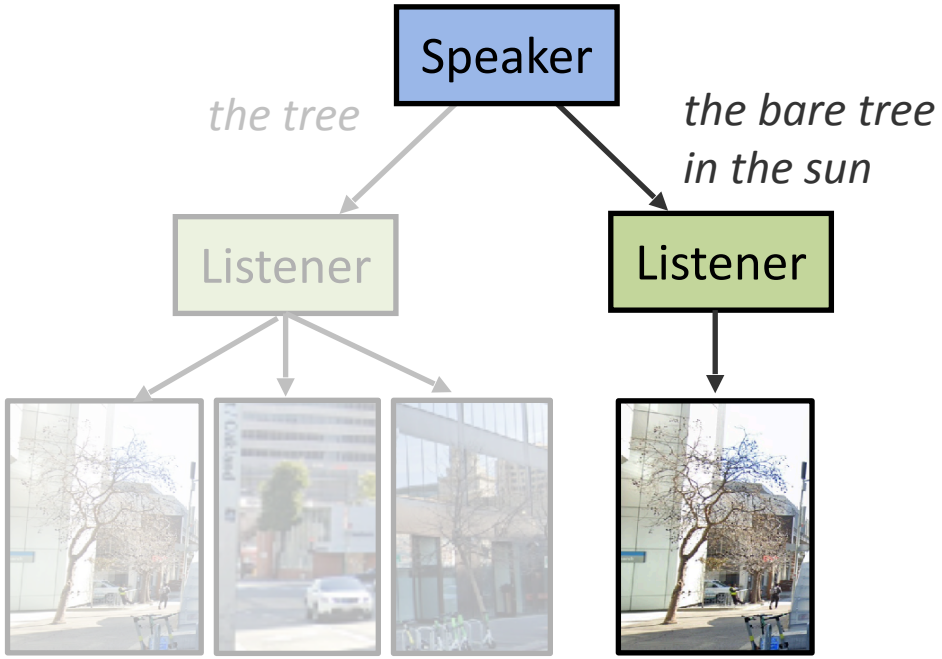
[*Logic and Conversation.* Grice, 1975]

Language is an act people take to produce effects on others and the world!

Using Language

Generation

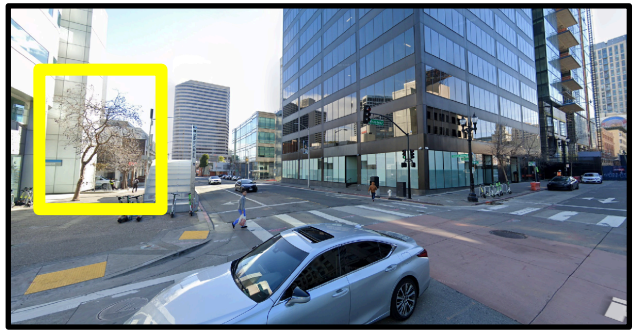


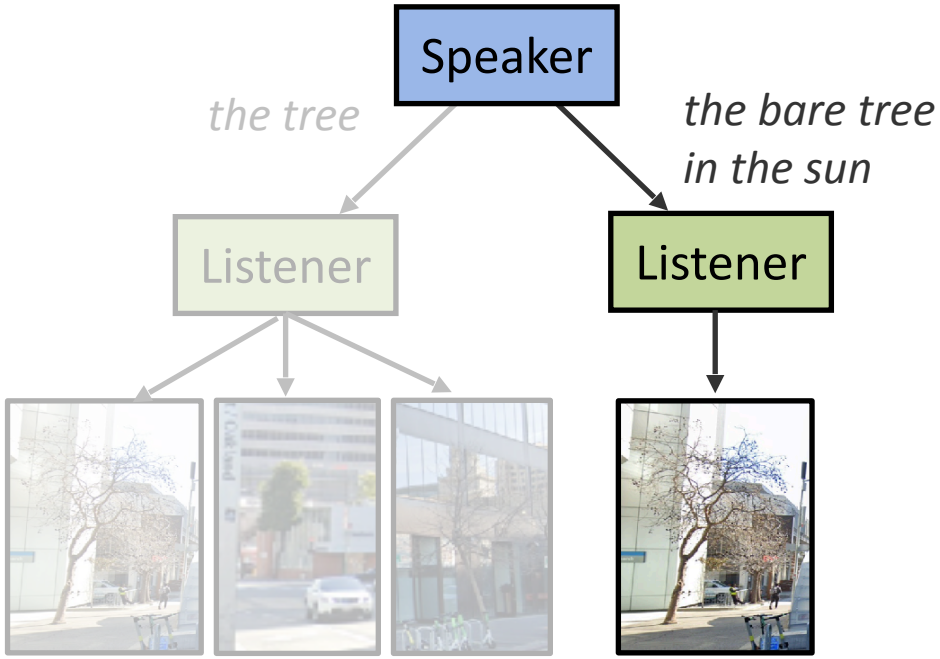


[e.g. Lewis 1969; Golland et al. 2010;
Frank and Goodman 2012; Degen et al. 2013]

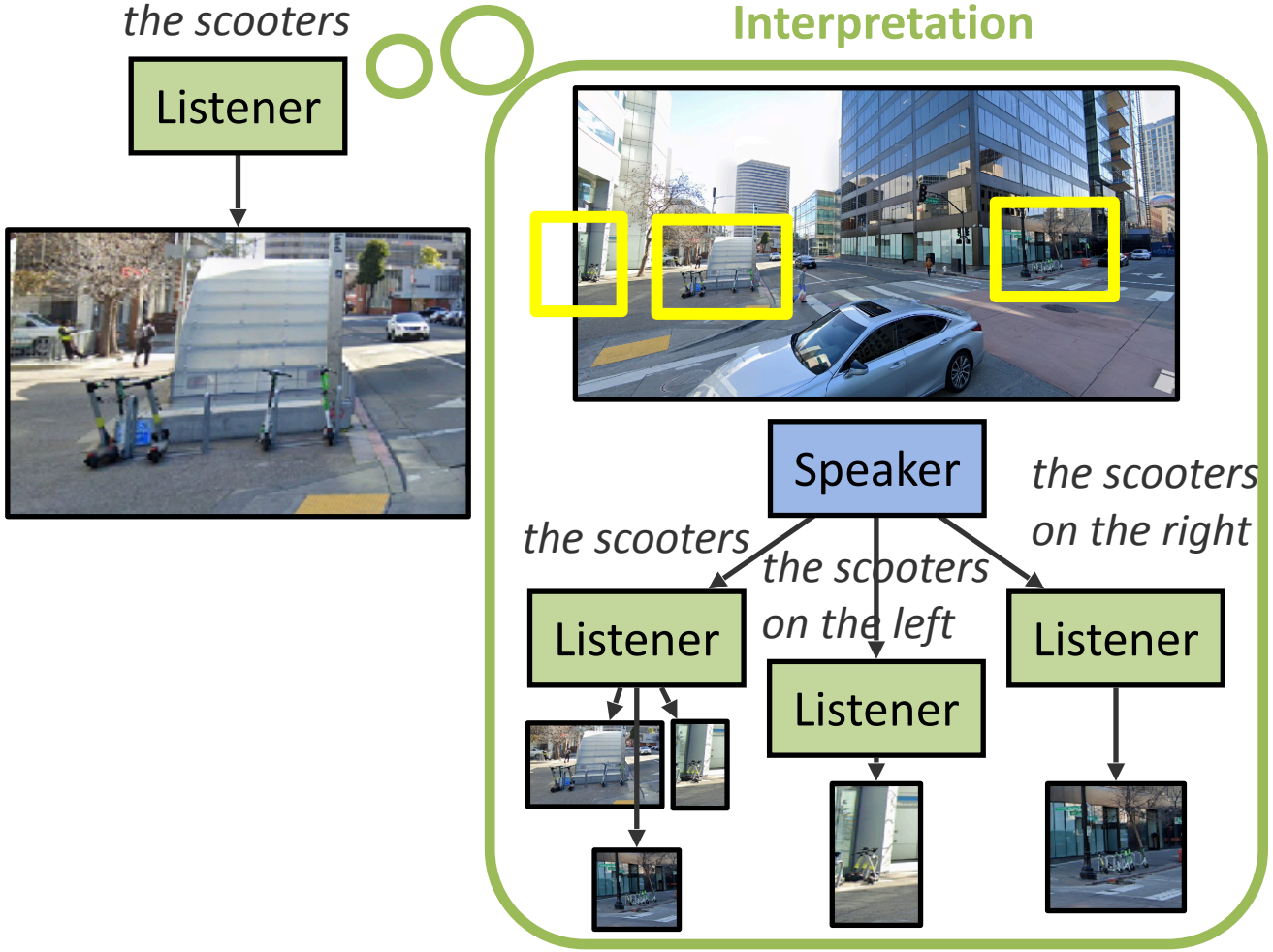
Using Language

Generation





Interpretation



[e.g. Lewis 1969; Golland et al. 2010; Frank and Goodman 2012; Degen et al. 2013]

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 scooter."

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 scooter."

Conveyed meaning:

Someone stole your scooter, 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 didn't steal your scooter."

Conveyed meaning:

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

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 scooter."

Conveyed meaning:

I did something to your scooter, but didn't steal it. E.g. 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> scooter."

Conveyed meaning:

I stole somebody else's scooter.

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 scooter."

Conveyed meaning:

I stole something you own, but not your scooter.

Language is Contextual

- Some problems depend on grounding references to context
- Indexicals and *Deixis*: “pointing or indicating”(e.g. pronouns, “this”, “that”, “here”, “now”)
  - *I am speaking*
  - *We won* (a team I’m on, OR a team I support)
  - *I am here* (in my house; in this Zoom room)
  - *We are here* (pointing to a map)
  - *I’m in a class now*
  - *I’m in a graduate program now*
  - *I’m not here right now* (voicemail greeting)

Language is Contextual

- Some problems depend on grounding into speaker intents or goals:
  - “Can you pass me the salt”

    -> please pass me the salt
  - “Do you have any kombucha?” // “I have tea”

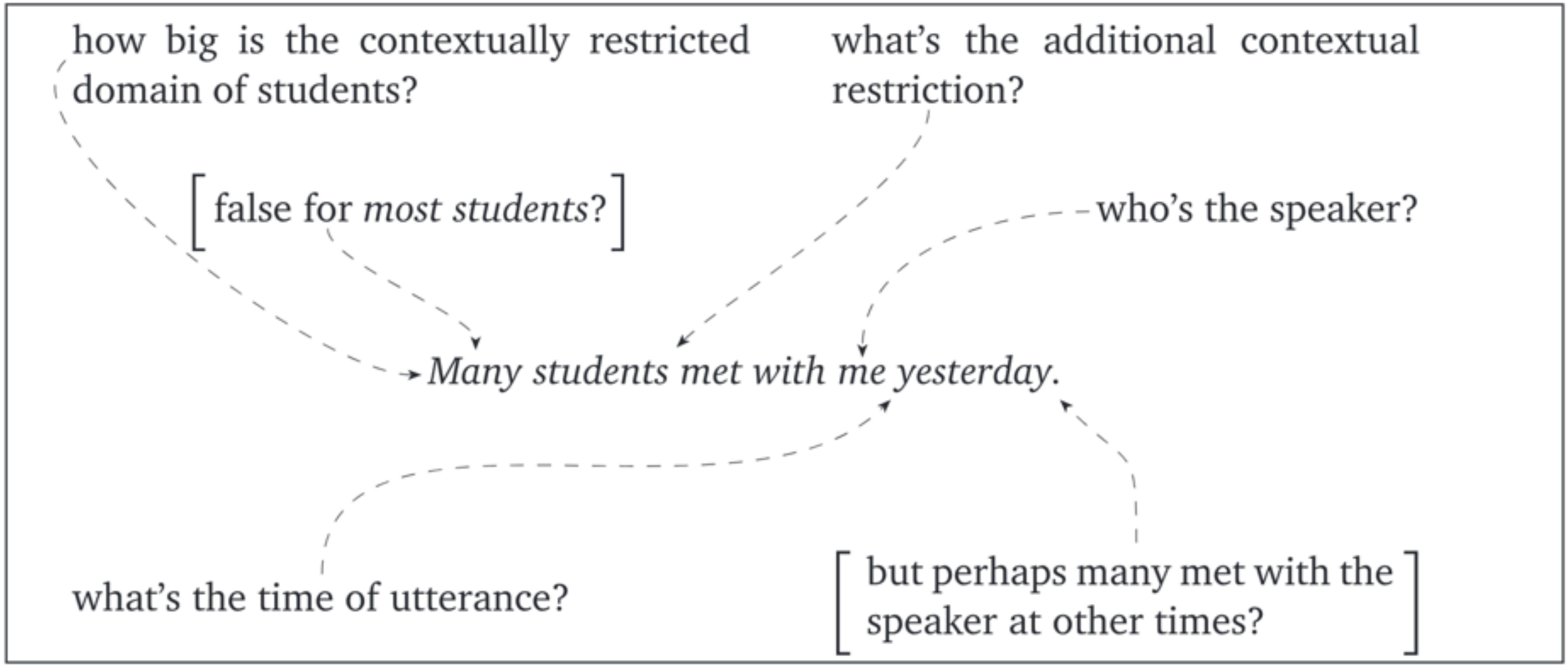
    -> I don’t have any kombucha
  - “You’re fired!”

    -> *performative*, that changes the state of the world

Language is Contextual

- Scope or type of answers: *Where are you from?*
  - Athens, Ohio (issue: hometown)
  - The U.S. (issue: nationality)
  - Berkeley (issue: affiliation)
  - Planet Earth (issue: intergalactic meetings)

Language is Contextual



Language is Contextual

- Children learn word meanings incredibly fast, from incredibly little data

  - Regularity and contrast in the input signal
  - Social cues
  - Inferring speaker intent
  - Regularities in the physical environment

Tomasello et al. 2005, Frank et al. 2012, Frank and Goodman 2014

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."

Language is a rational action in a cooperative game.

Gricean Maxims

Grice (1975) claims that many of these phenomena are explained by the tensions between the following ***maxims***:

1. **Quantity** – be as informative as possible, give as much information as needed, but no more. (*"The car was stolen."*)
2. **Quality** - be truthful, and don't give information that is false or unsupported by evidence. (*"Did you invite A and B?" // "I invited B."*)
3. **Relation** – be relevant, and say things that are pertinent to the discussion. (*"I'm out of gas" // "There's a station round the corner."*)
4. **Manner** – be clear, brief, and orderly as possible; avoid unnecessary prolixity. (*"He overslept and failed the test."*)

Implicature

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.

Conversational Implicature

"The car was stolen."

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

"Did you invite Alice and Bob?" // "I invited Alice."

- *The speaker didn't invite Bob.*

"I'm out of gas." // "There's a station round the corner."

- *You can get gas there (e.g. it's open).*

"He overslept and failed the test."

- *Those events happened in that order.*

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."

Conversational Implicature

Speaker S saying utterance U to listener L conversationally implicates q if, and only if,

Conversational Implicature

Speaker S saying utterance U to listener L conversationally implicates q if, and only if,

1. S and L mutually, publicly presume that S is obeying the cooperative principle.

Conversational Implicature

Speaker S saying utterance U to listener L conversationally implicates q if, and only if,

1. S and L mutually, publicly presume that S is obeying the cooperative principle.
2. To maintain 1 given U, it must be supposed that S thinks that q.

Conversational Implicature

Speaker S saying utterance U to listener L conversationally implicates q if, and only if,

1. S and L mutually, publicly presume that S is obeying the cooperative principle.
2. To maintain 1 given U, it must be supposed that S thinks that q.
3. S thinks that both S and L mutually, publicly presume that L is willing and able to work out that 2 holds.

Conversational Implicature

Ann: What city does Paul live in?
Bob: Hmm . . . he lives in California.

Conversational Implicature

Ann: What city does Paul live in?
Bob: Hmm . . . he lives in California.

Conversational implicature: Bob does not know which city Paul lives in.

Conversational Implicature

Ann: What city does Paul live in?
Bob: Hmm . . . he lives in California.

Conversational implicature: Bob does not know which city Paul lives in.

1. Contextual premise: Ann and Bob are planning a trip, and both are open to visiting Paul.

Conversational Implicature

Ann: What city does Paul live in?
Bob: Hmm . . . he lives in California.

Conversational implicature: Bob does not know which city Paul lives in.

1. Contextual premise: Ann and Bob are planning a trip, and both are open to visiting Paul.
2. Assume Bob is cooperative at least insofar as he is forthcoming about where Paul lives.

Conversational Implicature

Ann: What city does Paul live in?
Bob: Hmm . . . he lives in California.

Conversational implicature: Bob does not know which city Paul lives in.

1. Contextual premise: Ann and Bob are planning a trip, and both are open to visiting Paul.
2. Assume Bob is cooperative at least insofar as he is forthcoming about where Paul lives.
3. Bob supplied less information than was required, seemingly contradicting 2 .

Conversational Implicature

Ann: What city does Paul live in?
Bob: Hmm . . . he lives in California.

Conversational implicature: Bob does not know which city Paul lives in.

1. Contextual premise: Ann and Bob are planning a trip, and both are open to visiting Paul.
2. Assume Bob is cooperative at least insofar as he is forthcoming about where Paul lives.
3. Bob supplied less information than was required, seemingly contradicting 2.
4. Assume Bob does not know which city Paul lives in.

Conversational Implicature

Ann: What city does Paul live in?
Bob: Hmm . . . he lives in California.

Conversational implicature: Bob does not know which city Paul lives in.

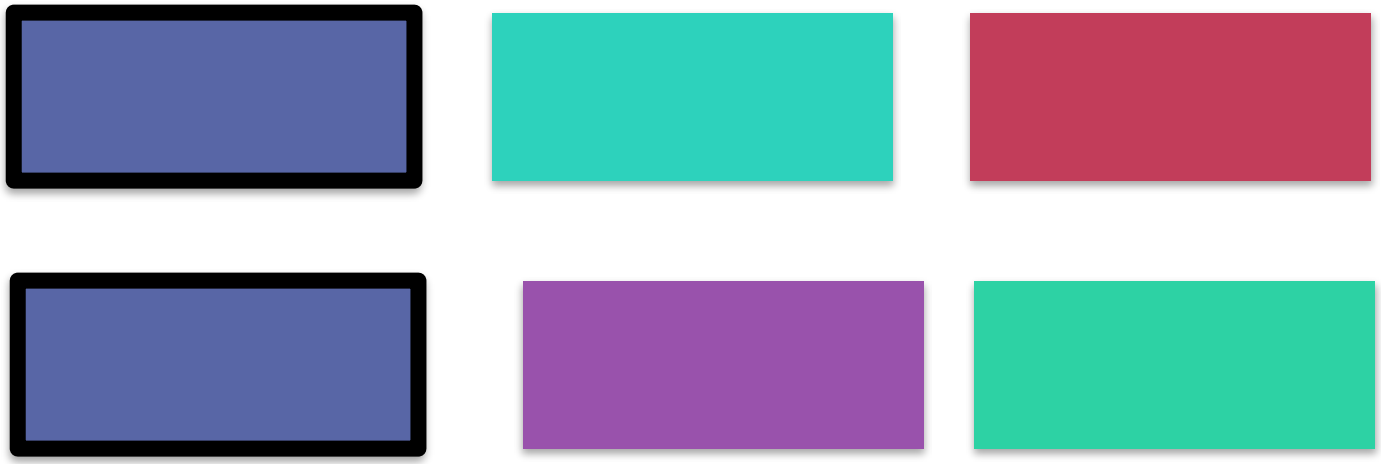
1. Contextual premise: Ann and Bob are planning a trip, and both are open to visiting Paul.
2. Assume Bob is cooperative at least insofar as he is forthcoming about where Paul lives.
3. Bob supplied less information than was required, seemingly contradicting 2.
4. Assume Bob does not know which city Paul lives in.
5. Then Bob's answer is optimal given his evidence.

Reference Games

- Simple form of *using language*
- Set of candidate referents R
- Encoding meaning
  - A speaker has an intent, which is a target referent $r \in R$
  - Speaker maps this intent r and context R to an utterance u
- Decoding meaning
  - A listener observes R and u
  - Listener *resolves* the reference u to $r' \in R$
- Success: $r = r'$

Reference Games

“Hat”



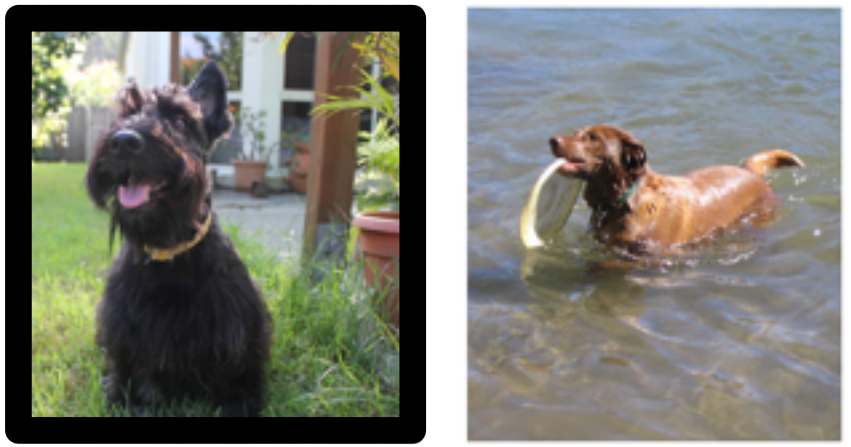
?

?

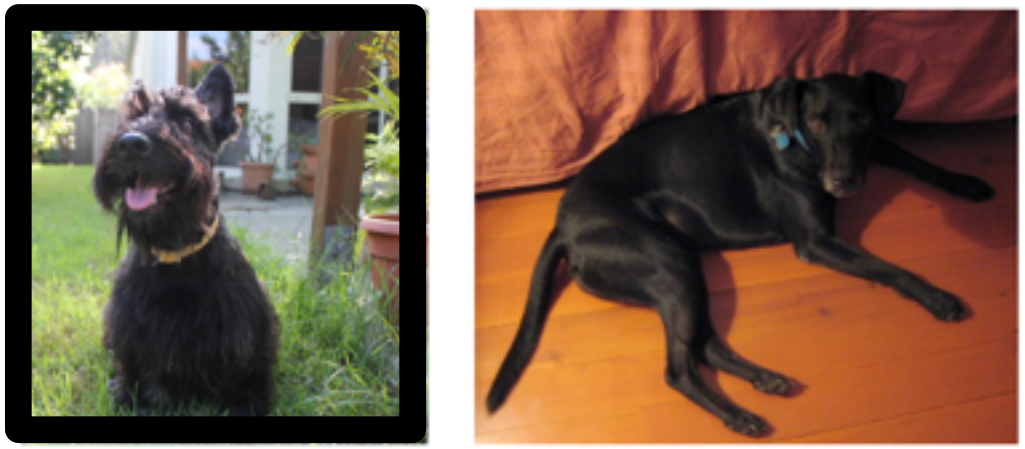
Monroe et al. 2017

Reference Games

“Ice skater”



?



?

Ji et al. 2022, Haber et al. 2019

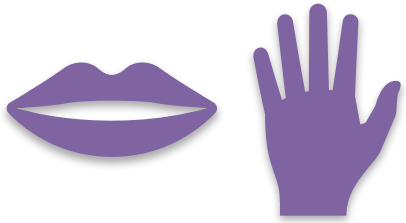
Demo!

PollEv.com/alanesuhr930

Speakers and Listeners

t

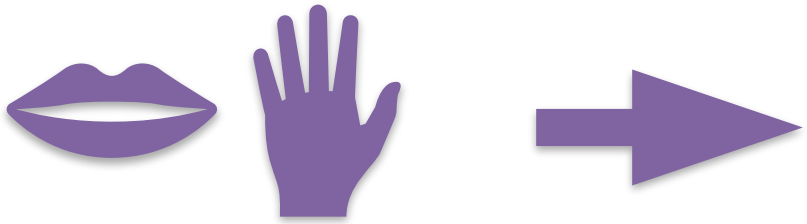
Speaker



Speakers and Listeners

t

Speaker



“glasses”

$$p_{\text{Literal}}^{\text{Speaker}}(\cdot \mid t)$$

	R1	R2	R3
hat		0	
glasses		1	
mustache		0	

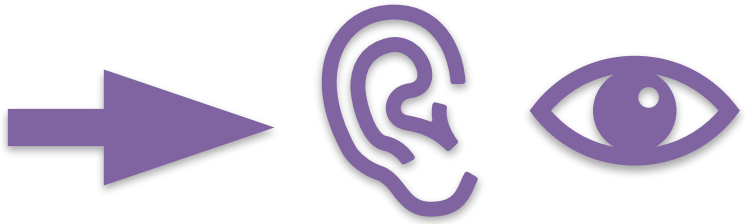
Speakers and Listeners

[[*"glasses"*]]

	R1	R2	R3
hat			
glasses	0	1	0
mustache			

$x =$ *"glasses"*

Listener

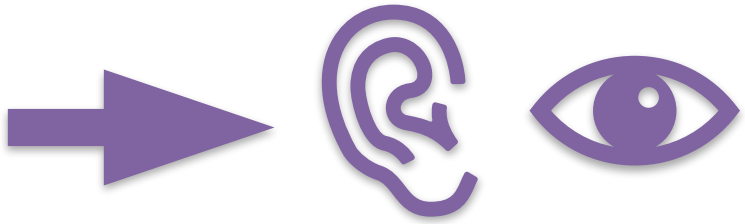


Speakers and Listeners

$$p^{\text{Listener}}_{\text{Literal}}(\cdot \mid x)$$

	R1	R2	R3
hat			
glasses	0	1	0
mustache			

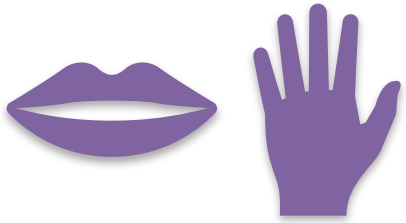
$x =$ *"glasses"*



Context-Dependence

t

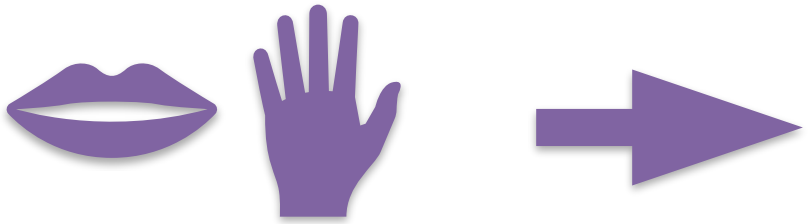
Speaker



Context-Dependence

t

Speaker



“glasses”

$$p^{\text{Speaker}}_{\text{Literal}}(\cdot \mid t)$$

	R1	R2	R3
hat		0	
glasses		1	
mustache		0	

Context-Dependence

[[*"glasses"*]]

	R1	R2	R3
hat			
glasses	0	1	1
mustache			

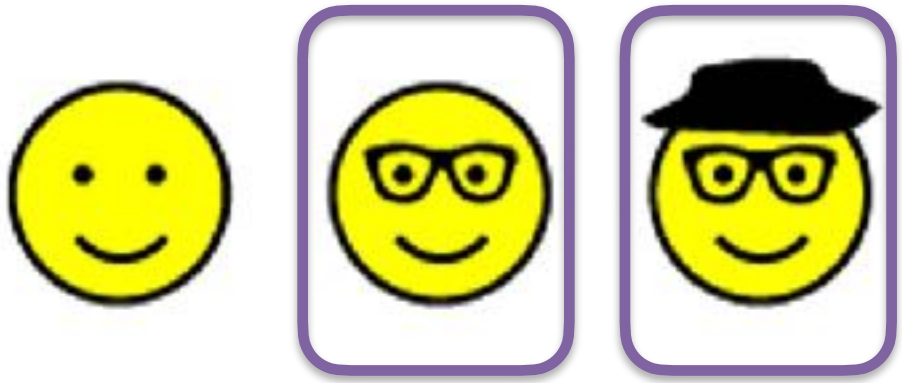
$x =$ *"glasses"*

Listener

Context-Dependence

$$p_{\text{Literal}}^{\text{Listener}}(\cdot \mid x)$$

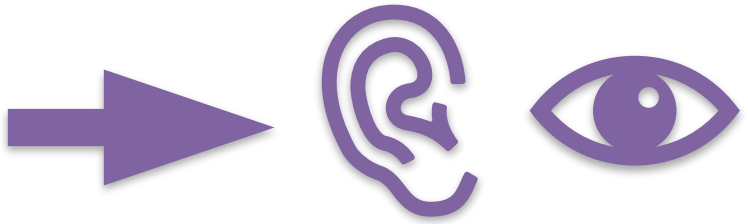
	R1	R2	R3
hat			
glasses	0	0.5	0.5
mustache			



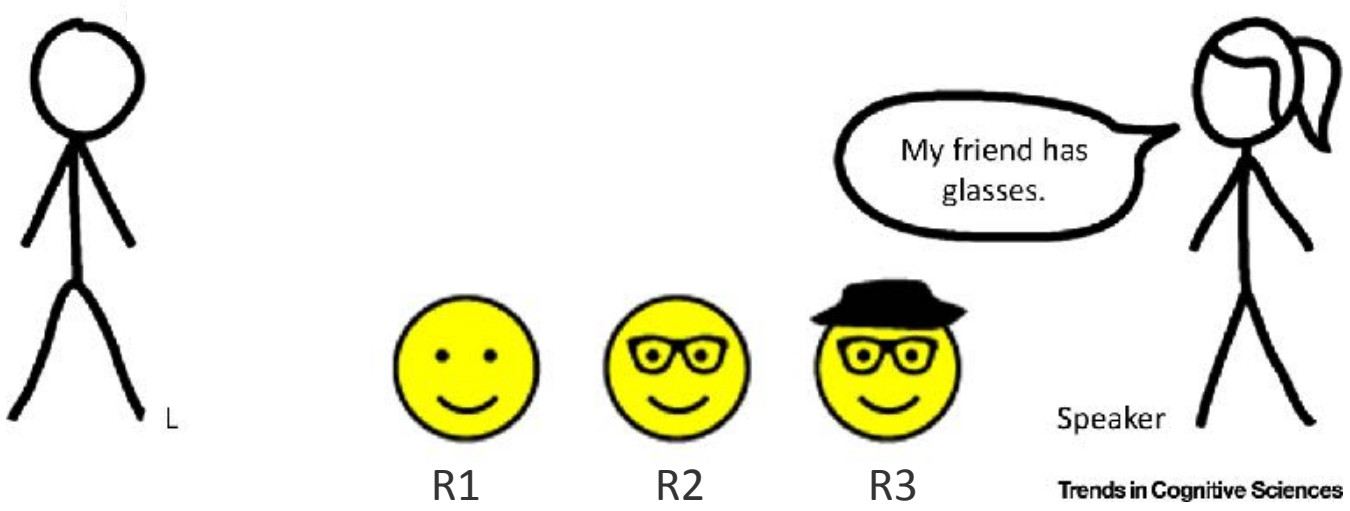
???

Listener

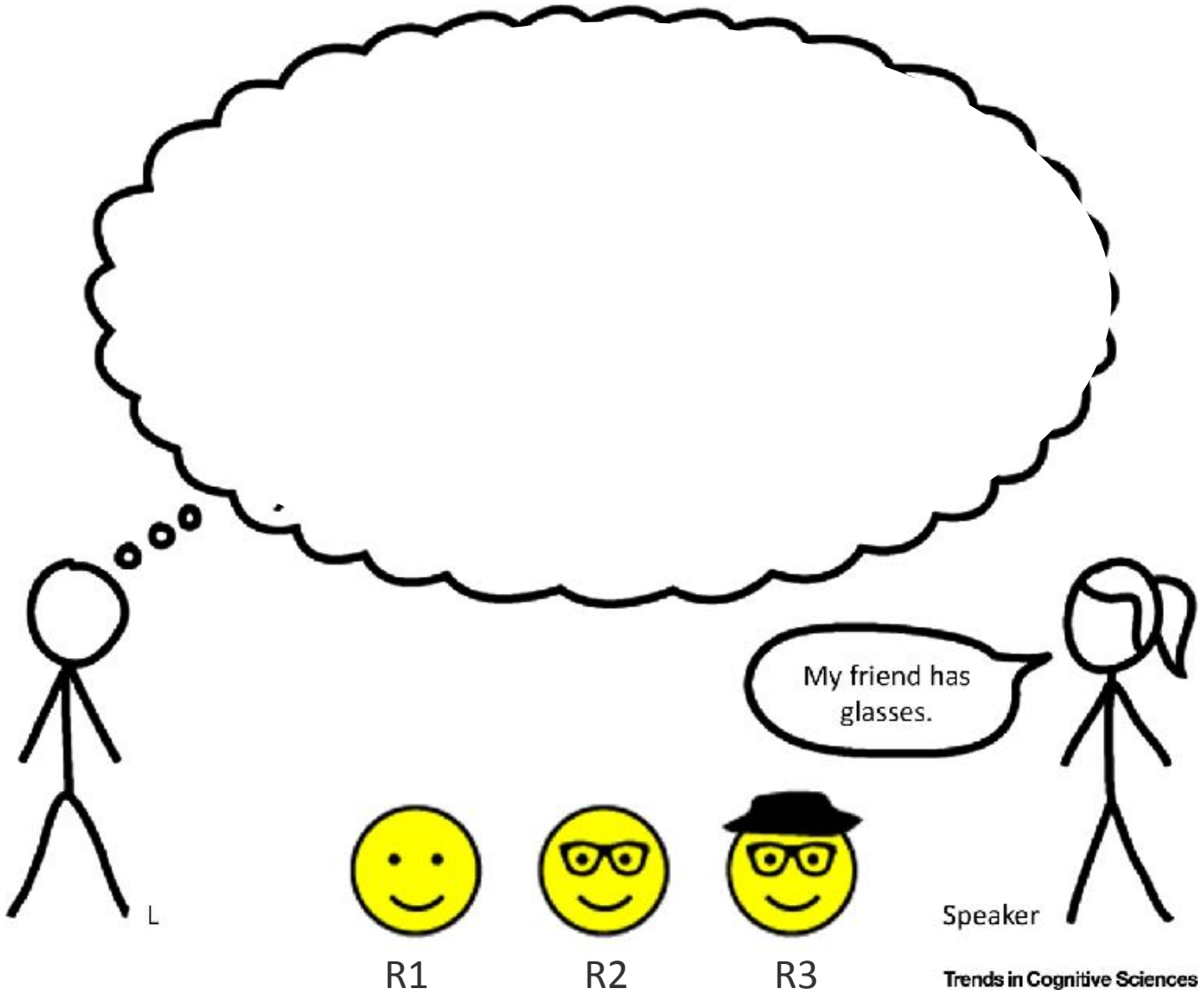
$x =$ *"glasses"*



Pragmatic Speakers and Listeners



Rational Speech Acts Model

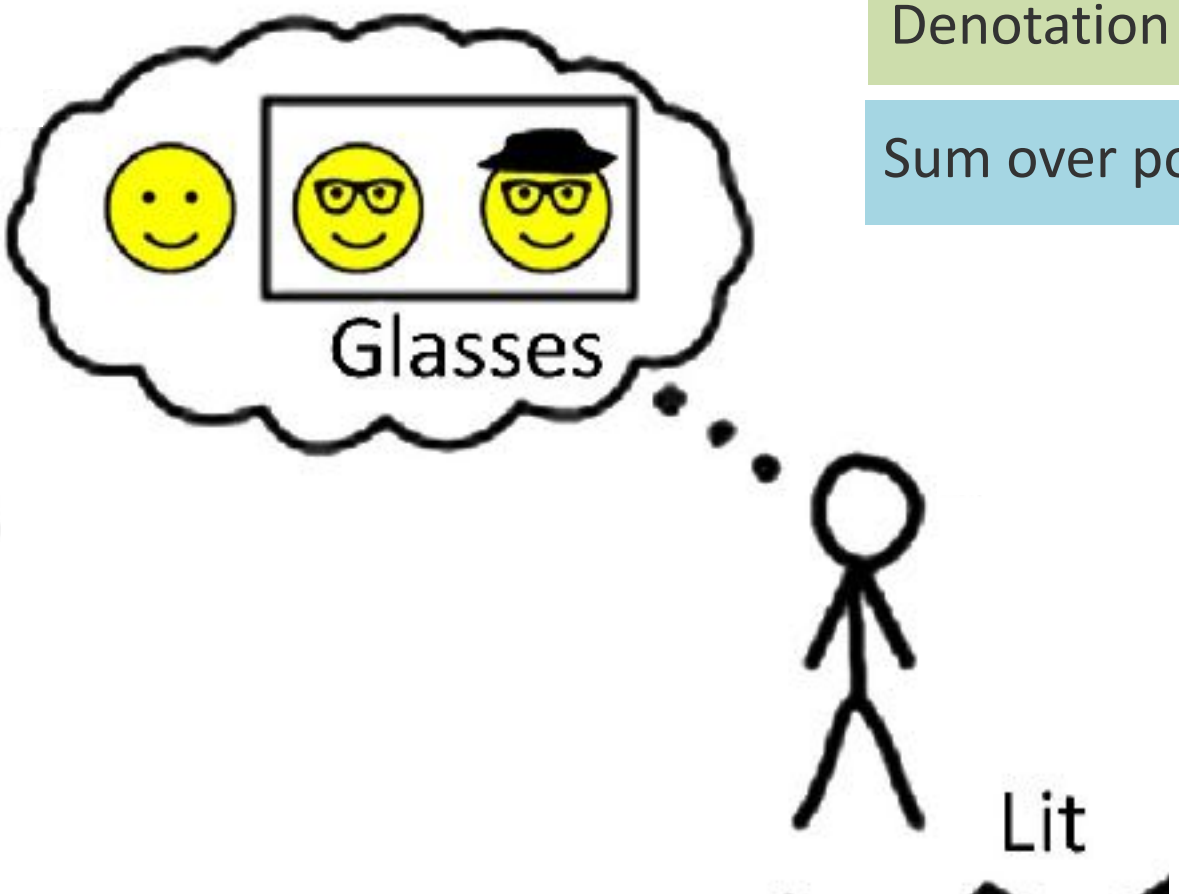


Rational Speech Acts Model

$$p_{\text{Literal}}^{\text{Listener}}(r \mid x) = \frac{[\![x]\!]_r}{\sum_{r' \in R} [\![x]\!]_{r'}}$$

Denotation of utterance

Sum over possible referents



Rational Speech Acts Model

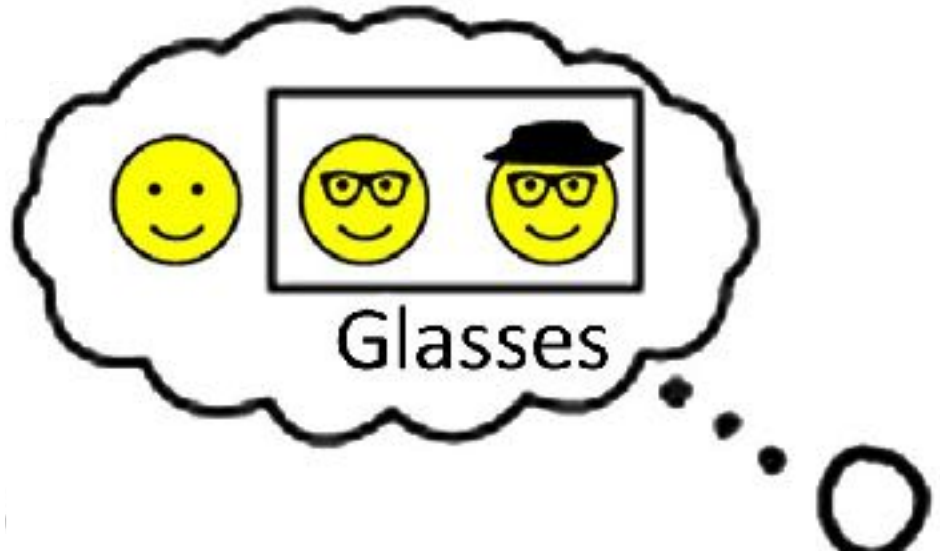
[[*"glasses"*]]

	R1	R2	R3
hat			
glasses			
mustache			

$$p_{\text{Literal}}^{\text{Listener}}(r \mid x) = \frac{[\![x]\!]_r}{\sum_{r' \in R} [\![x]\!]_{r'}}$$

Denotation of utterance

Sum over possible referents



Rational Speech Acts Model

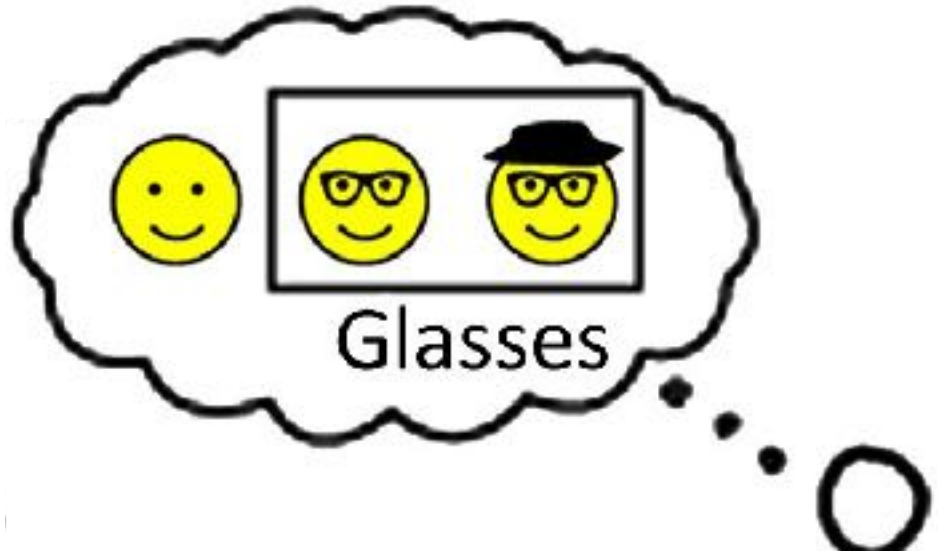
[[*"glasses"*]]

	R1	R2	R3
hat			
glasses	0	1	1
mustache			

$$p_{\text{Literal}}^{\text{Listener}}(r \mid x) = \frac{[\![x]\!]_r}{\sum_{r' \in R} [\![x]\!]_{r'}}$$

Denotation of utterance

Sum over possible referents



Rational Speech Acts Model

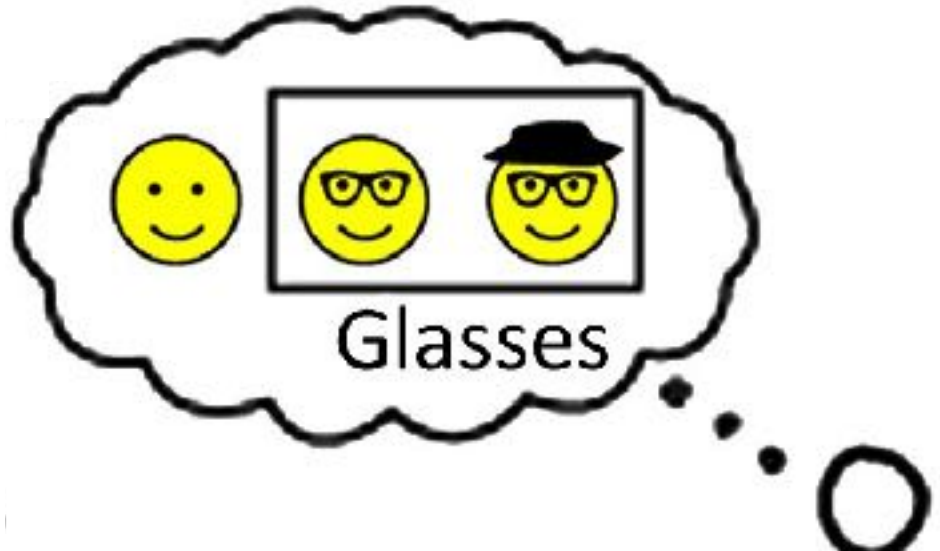
$$p^{\text{Listener}}_{\text{Literal}}(\cdot \mid x)$$

$$p^{\text{Listener}}_{\text{Literal}}(r \mid x) = \frac{[\![x]\!]_r}{\sum_{r' \in R} [\![x]\!]_{r'}}$$

Denotation of utterance

Sum over possible referents

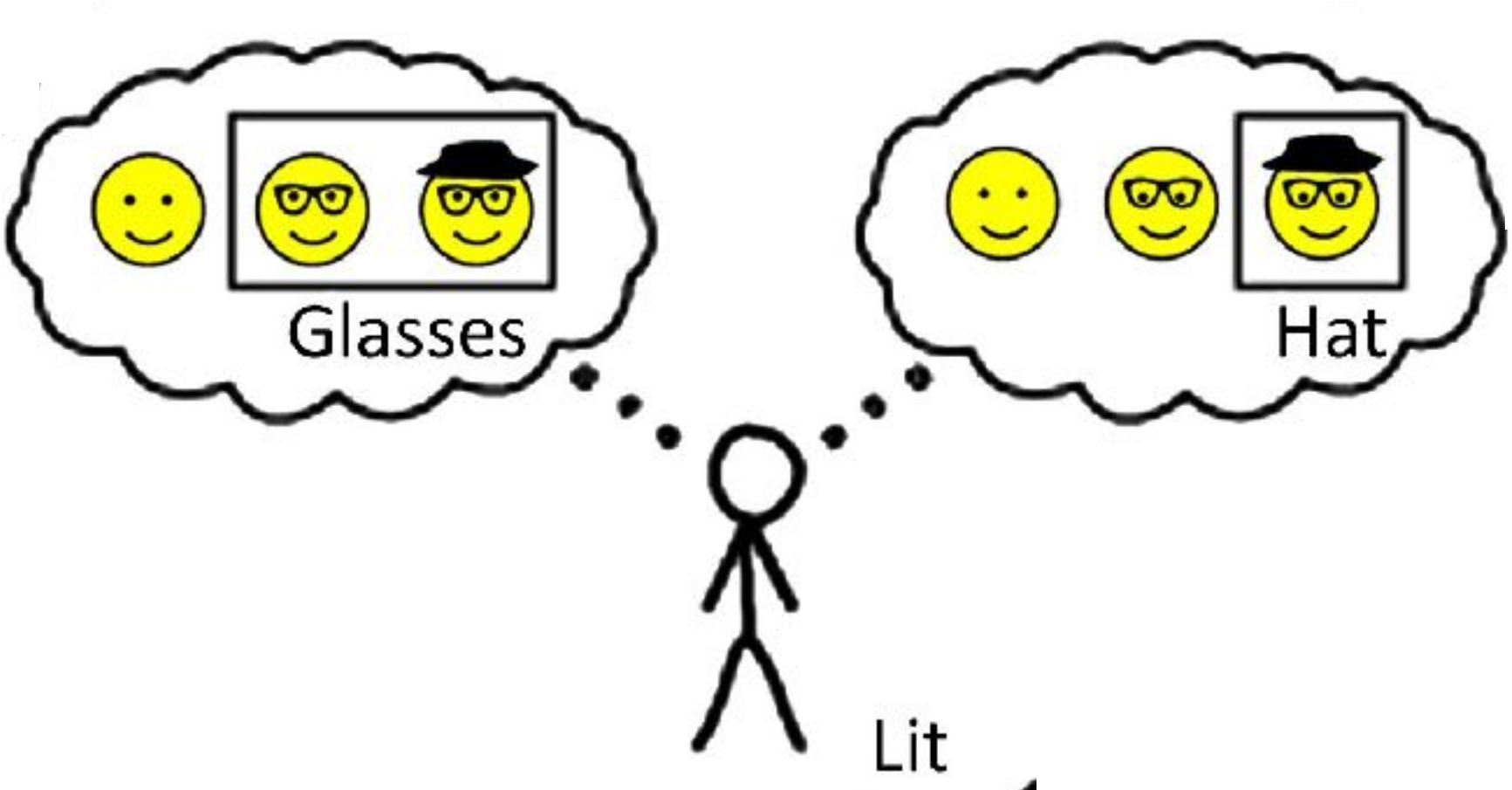
	R1	R2	R3
hat			
glasses	0	0.5	0.5
mustache			



Rational Speech Acts Model

$$p^{\text{Listener}}_{\text{Literal}}(\cdot \mid x) \qquad p^{\text{Listener}}_{\text{Literal}}(r \mid x) = \frac{[\![x]\!]_r}{\sum_{r' \in R} [\![x]\!]_{r'}}$$

	R1	R2	R3
hat	0	0	1
glasses	0	0.5	0.5
mustache	0	0	0



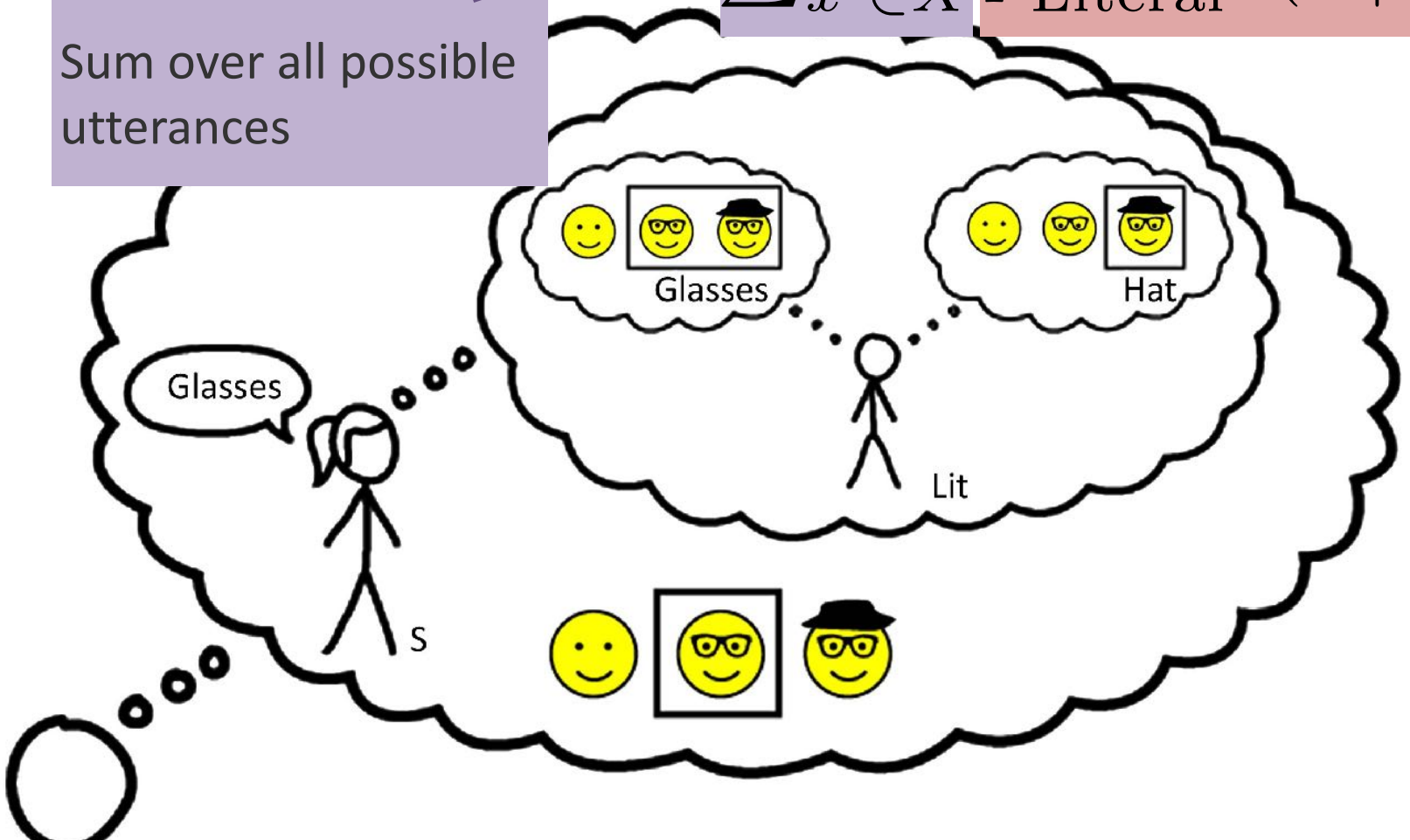
Rational Speech Acts Model

$$p^{\text{Listener}}_{\text{Literal}}(\cdot \mid x)$$

$$p^{\text{Speaker}}_{\text{Pragmatic}}(x \mid r) = \frac{p^{\text{Listener}}_{\text{Literal}}(r \mid x)}{\sum_{x' \in X} p^{\text{Listener}}_{\text{Literal}}(r \mid x')}$$

Sum over all possible utterances

	R1	R2	R3
hat	0	0	1
glasses	0	0.5	0.5
mustache	0	0	0

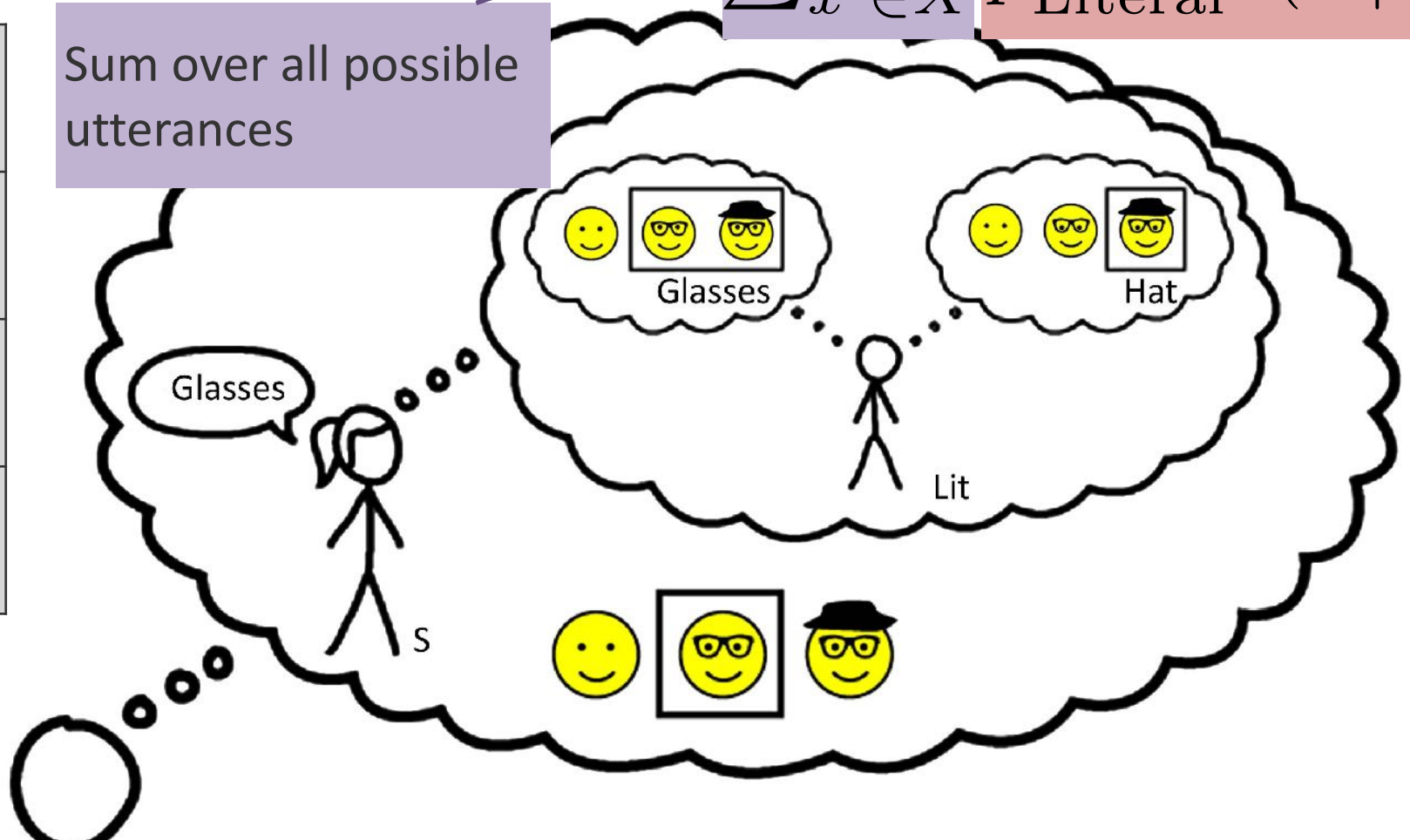


Rational Speech Acts Model

$$p^{\text{Speaker}}_{\text{Pragmatic}}(\cdot \mid r) \quad p^{\text{Speaker}}_{\text{Pragmatic}}(x \mid r) = \frac{p^{\text{Listener}}_{\text{Literal}}(r \mid x)}{\sum_{x' \in X} p^{\text{Listener}}_{\text{Literal}}(r \mid x')}$$

Sum over all possible utterances

	R1	R2	R3
hat	0	0	2/3
glasses	0	1	1/3
mustache	0	0	0



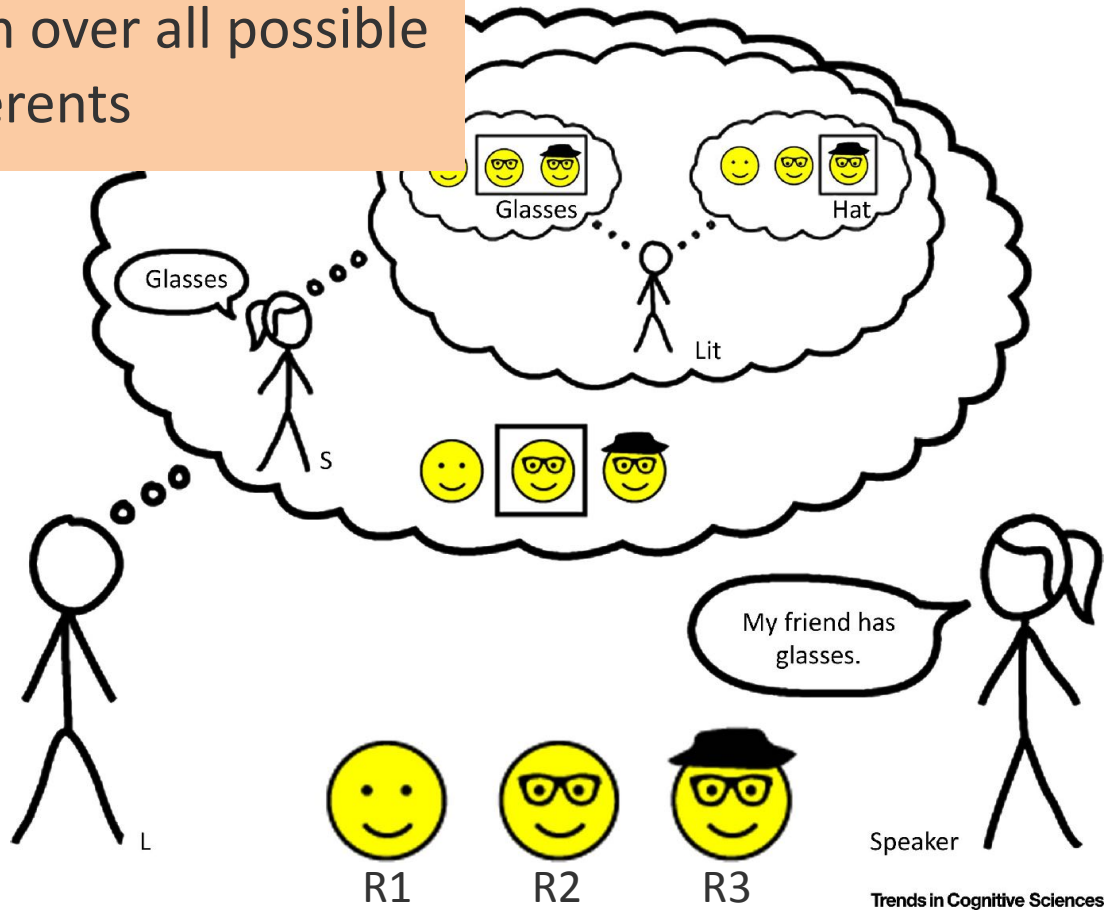
Rational Speech Acts Model

$$p^{\text{Speaker}}_{\text{Pragmatic}}(\cdot \mid r)$$

	R1	R2	R3
hat	0	0	2/3
glasses	0	1	1/3
mustache	0	0	0

$$p^{\text{Listener}}_{\text{Pragmatic}}(r \mid x) = \frac{p^{\text{Speaker}}_{\text{Pragmatic}}(x \mid r)}{\sum_{r' \in R} p^{\text{Speaker}}_{\text{Pragmatic}}(x \mid r')}$$

Sum over all possible referents

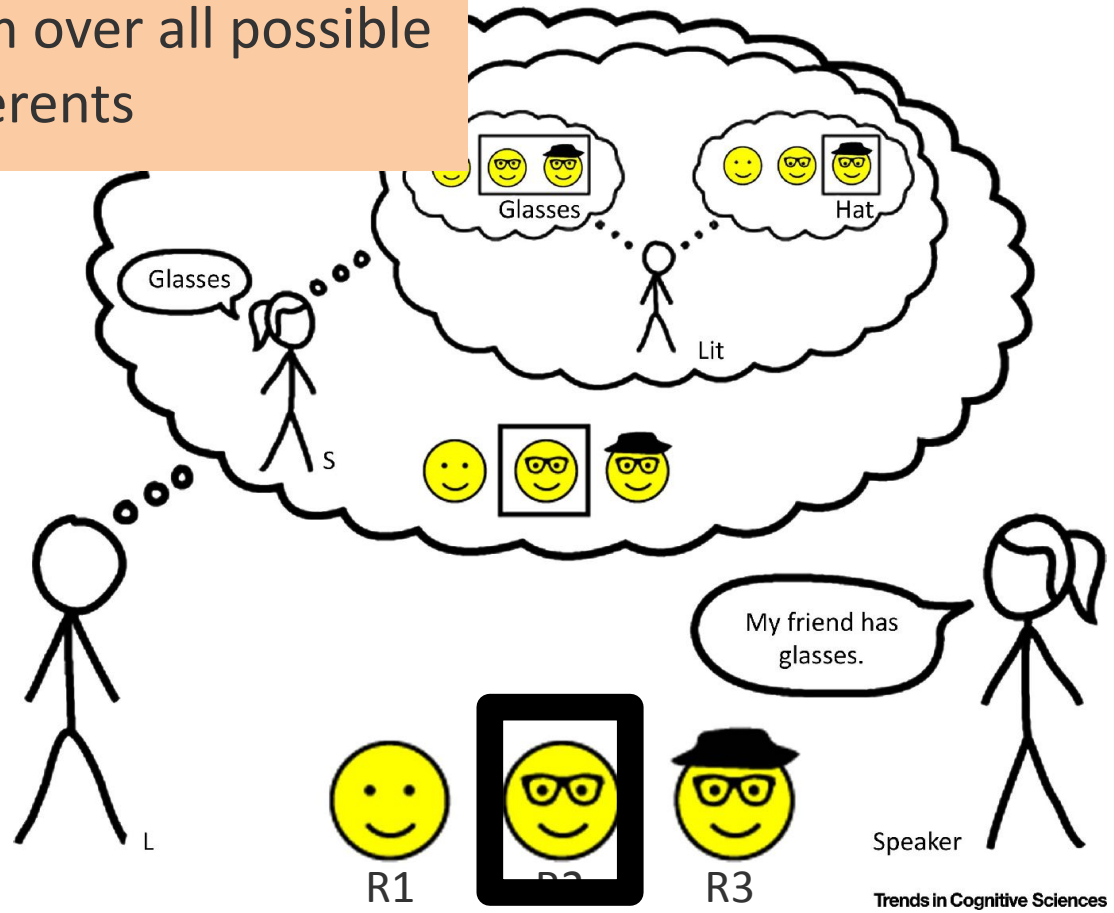


Rational Speech Acts Model

$$p^{\text{Listener}}_{\text{Pragmatic}}(\cdot \mid x) \qquad p^{\text{Listener}}_{\text{Pragmatic}}(r \mid x) = \frac{p^{\text{Speaker}}_{\text{Pragmatic}}(x \mid r)}{\sum_{r' \in R} p^{\text{Speaker}}_{\text{Pragmatic}}(x \mid r')}$$

Sum over all possible referents

	R1	R2	R3
hat	0	0	1
glasses	0	3/4	1/4
mustache	0	0	0



RSA: Review

- **Literal listener:** uses denotational semantics to map utterances to probability of referents

$$p_{\text{Literal}}^{\text{Listener}}(r \mid x) = \frac{[\![x]\!]_r}{\sum_{r' \in R} [\![x]\!]_{r'}}$$

RSA: Review

- **Literal listener:** uses denotational semantics to map utterances to probability of referents

$$p_{\text{Literal}}^{\text{Listener}}(r \mid x) = \frac{[\![x]\!]_r}{\sum_{r' \in R} [\![x]\!]_{r'}}$$

- **Pragmatic speaker:** re-normalizes probabilities over utterances given literal listener's interpretations

$$p_{\text{Pragmatic}}^{\text{Speaker}}(x \mid r) = \frac{p_{\text{Literal}}^{\text{Listener}}(r \mid x)}{\sum_{x' \in X} p_{\text{Literal}}^{\text{Listener}}(r \mid x')}$$

RSA: Review

- **Literal listener:** uses denotational semantics to map utterances to probability of referents

$$p_{\text{Literal}}^{\text{Listener}}(r \mid x) = \frac{[\![x]\!]_r}{\sum_{r' \in R} [\![x]\!]_{r'}}$$

- **Pragmatic speaker:** re-normalizes probabilities over utterances given literal listener's interpretations

$$p_{\text{Pragmatic}}^{\text{Speaker}}(x \mid r) = \frac{p_{\text{Literal}}^{\text{Listener}}(r \mid x)}{\sum_{x' \in X} p_{\text{Literal}}^{\text{Listener}}(r \mid x')}$$

- **Pragmatic listener:** takes into account alternative utterances the speaker *could* have used to refer to referents, but didn't

$$p_{\text{Pragmatic}}^{\text{Listener}}(r \mid x) = \frac{p_{\text{Pragmatic}}^{\text{Speaker}}(x \mid r)}{\sum_{r' \in R} p_{\text{Pragmatic}}^{\text{Speaker}}(x \mid r')}$$

RSA Variations

- Priors over referents

$$p_{\text{Literal}}^{\text{Listener}}(r \mid x) = \frac{[\![x]\!]_r \cdot P(r)}{\sum_{r' \in R} [\![x]\!]_{r'} \cdot P(r')}$$

RSA Variations

- Priors over referents

$$p^{\text{Listener}}_{\text{Literal}}(r \mid x) = \frac{[\![x]\!]_r \cdot P(r)}{\sum_{r' \in R} [\![x]\!]_{r'} \cdot P(r')}$$

- Utterance costs

$$p^{\text{Speaker}}_{\text{Pragmatic}}(x \mid r) = \frac{\exp(\log p^{\text{Listener}}_{\text{Literal}}(r \mid x) + C(x))}{\sum_{x \in X} \exp(\log p^{\text{Listener}}_{\text{Literal}}(r \mid x) + C(x))}$$

RSA Variations

- Priors over referents

$$p^{\text{Listener}}_{\text{Literal}}(r \mid x) = \frac{[\![x]\!]_r \cdot P(r)}{\sum_{r' \in R} [\![x]\!]_{r'} \cdot P(r')}$$

- Utterance costs

$$p^{\text{Speaker}}_{\text{Pragmatic}}(x \mid r) = \frac{\exp(\log p^{\text{Listener}}_{\text{Literal}}(r \mid x) + C(x))}{\sum_{x \in X} \exp(\log p^{\text{Listener}}_{\text{Literal}}(r \mid x) + C(x))}$$

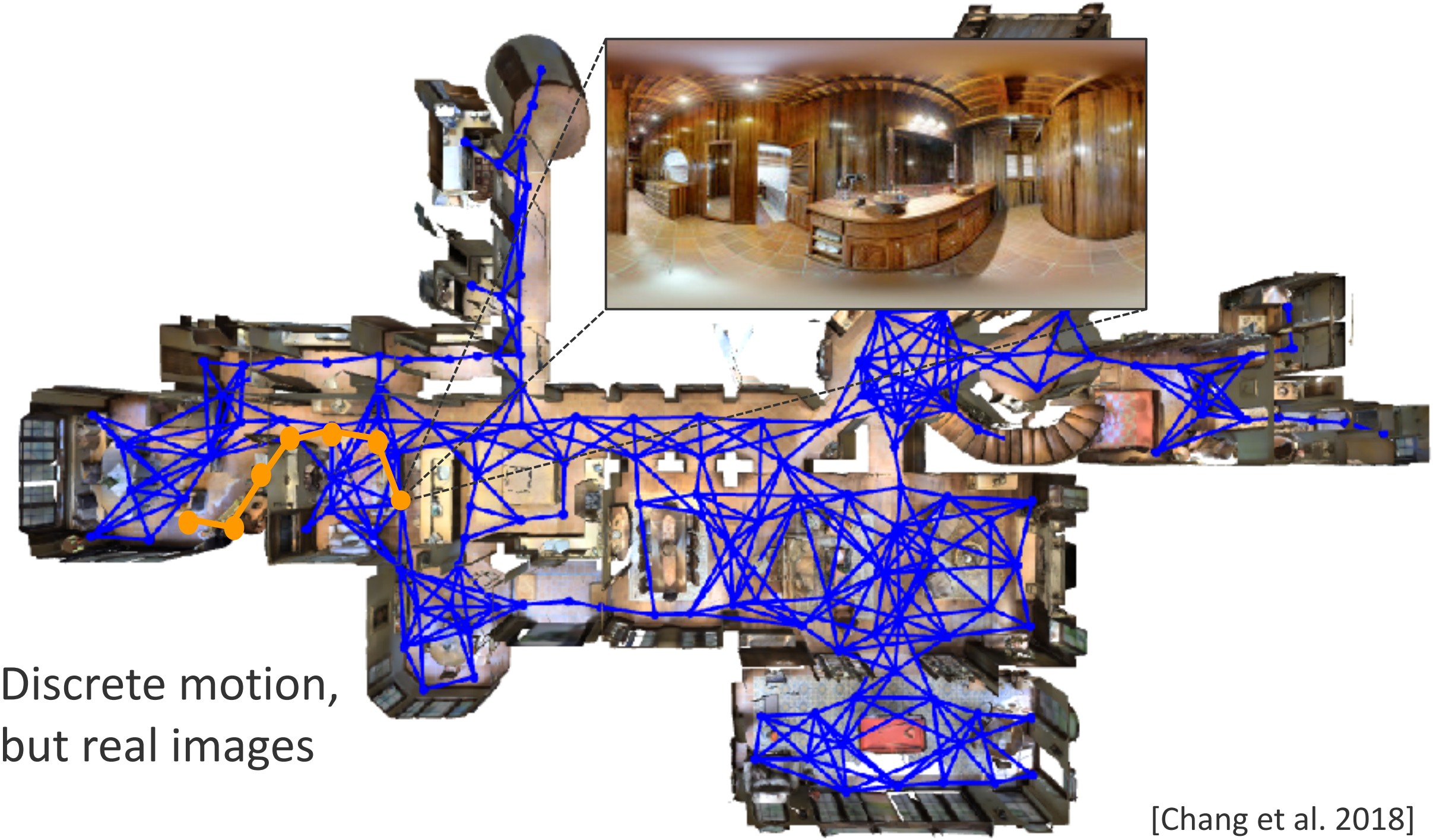
- Adjusting temperature of distributions

$$p^{\text{Speaker}}_{\text{Pragmatic}}(x \mid r) = \frac{\exp(\alpha \cdot (\log p^{\text{Listener}}_{\text{Literal}}(r \mid x) + C(x)))}{\sum_{x \in X} \exp(\alpha \cdot (\log p^{\text{Listener}}_{\text{Literal}}(r \mid x) + C(x)))}$$

Language Use Beyond Reference

Turn left and take a right at the table. Take a left at the painting and then take your first right. Wait next to the exercise equipment.

[*Vision-and-Language Navigation Task*. Anderson et al., 2018]



- Discrete motion, but real images

[Chang et al. 2018]

Instruction Following

Input instructions:

Go forward between the kitchen counters and then turn right into the living room. Walk forward onto the rug.

Output a route:

Instruction Generation

Input a route:

Output instructions:

Go forward between the kitchen counters and then turn right into the living room. Walk forward onto the rug.

Speakers and Listeners

Inputs

Outputs

Go forward between the kitchen counters…

Instruction *x*

Listener

$p^{\text{Listener}}_{\text{Literal}}(r \mid x)$

Route *r*

Route *r*

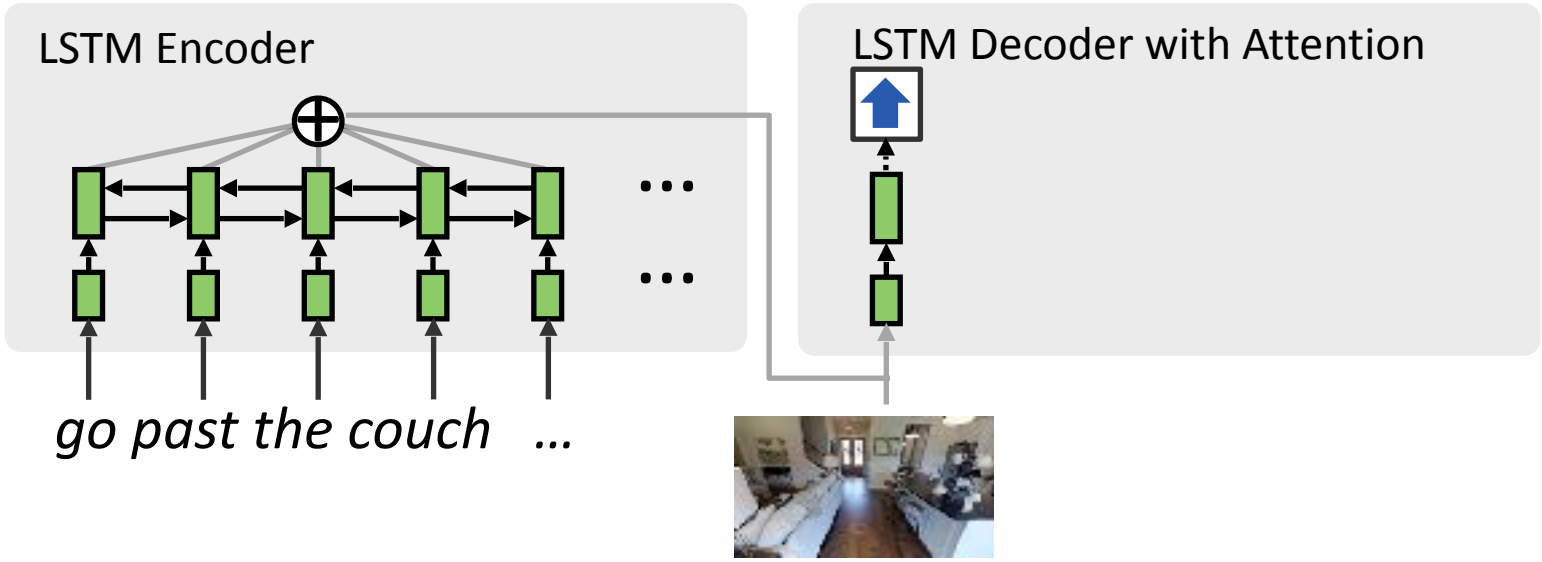
Speaker

$p^{\text{Speaker}}_{\text{Literal}}(x \mid r)$

Go forward between the kitchen counters…

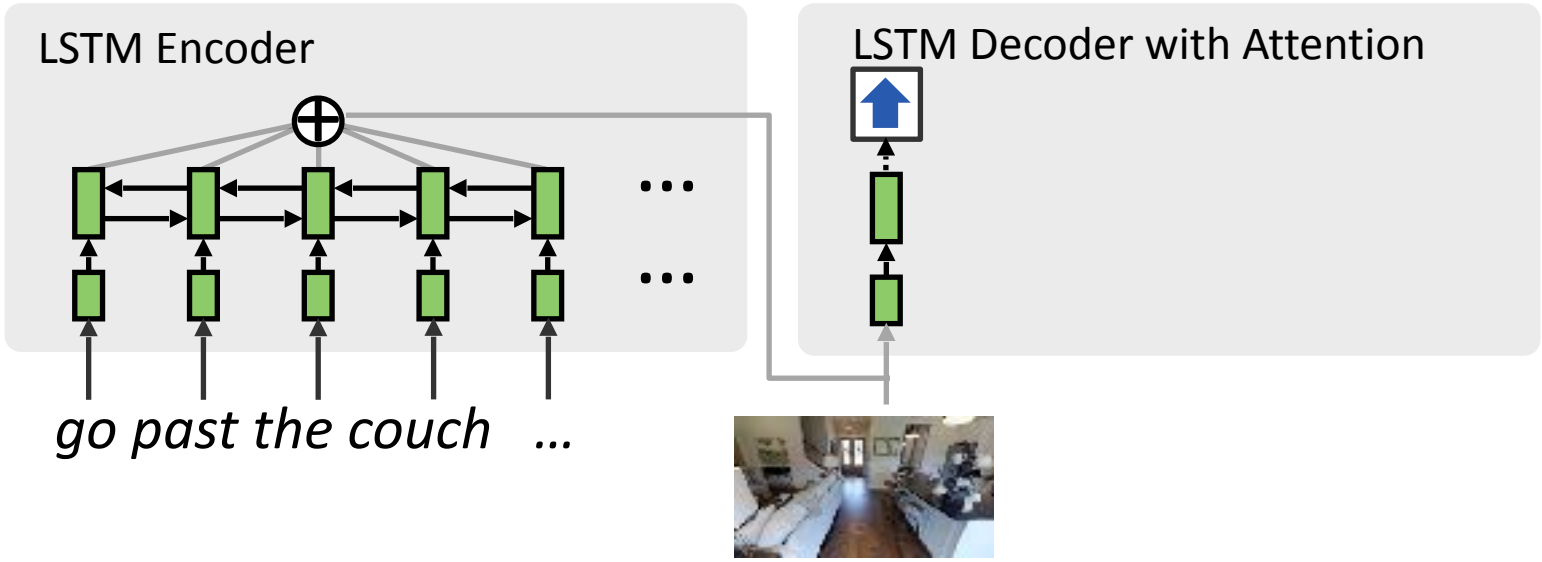
Instruction *x*

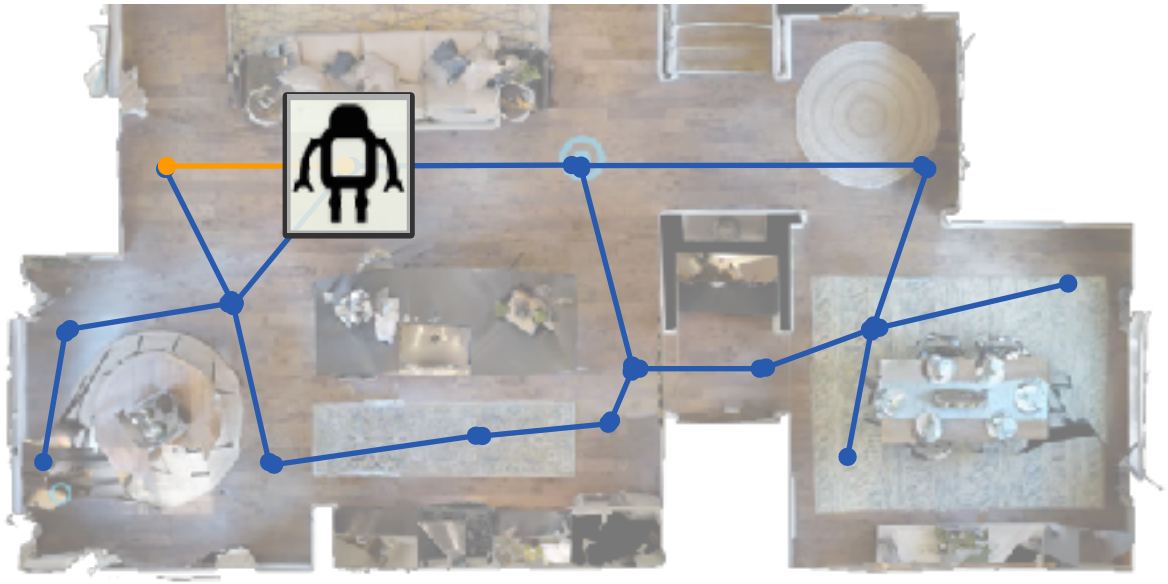
Literal Listener



[Anderson et al., 2018]

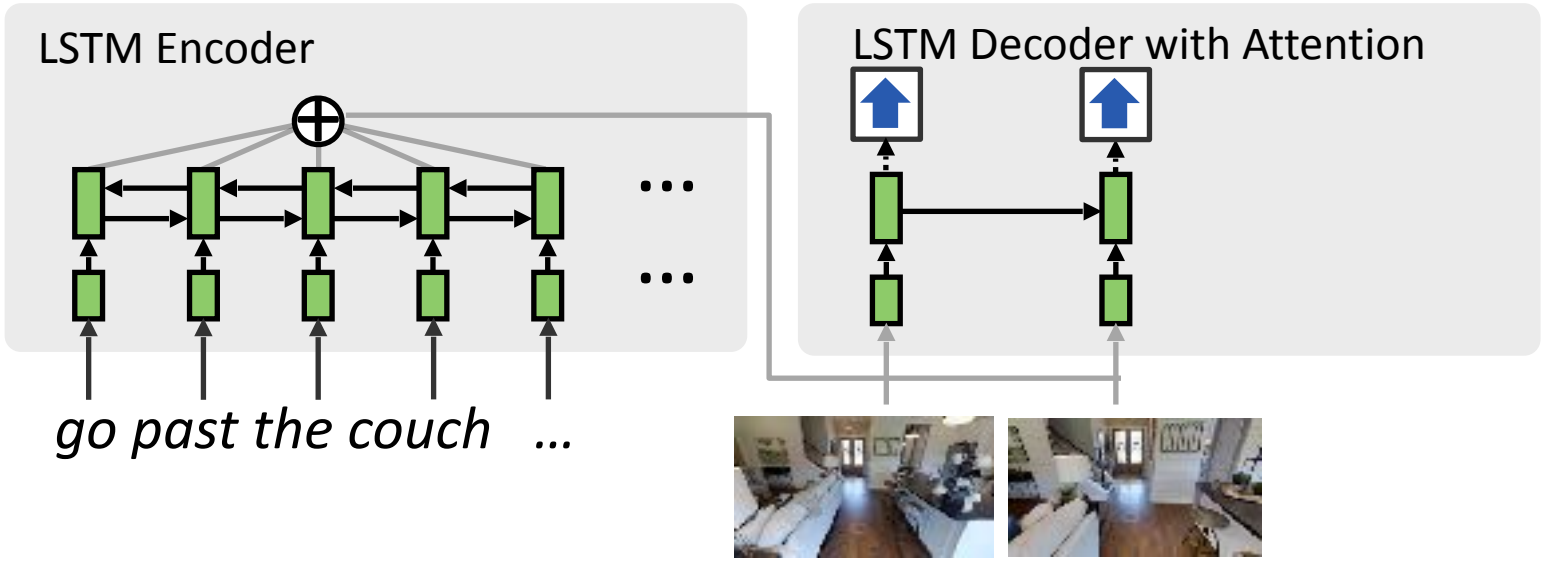
Literal Listener

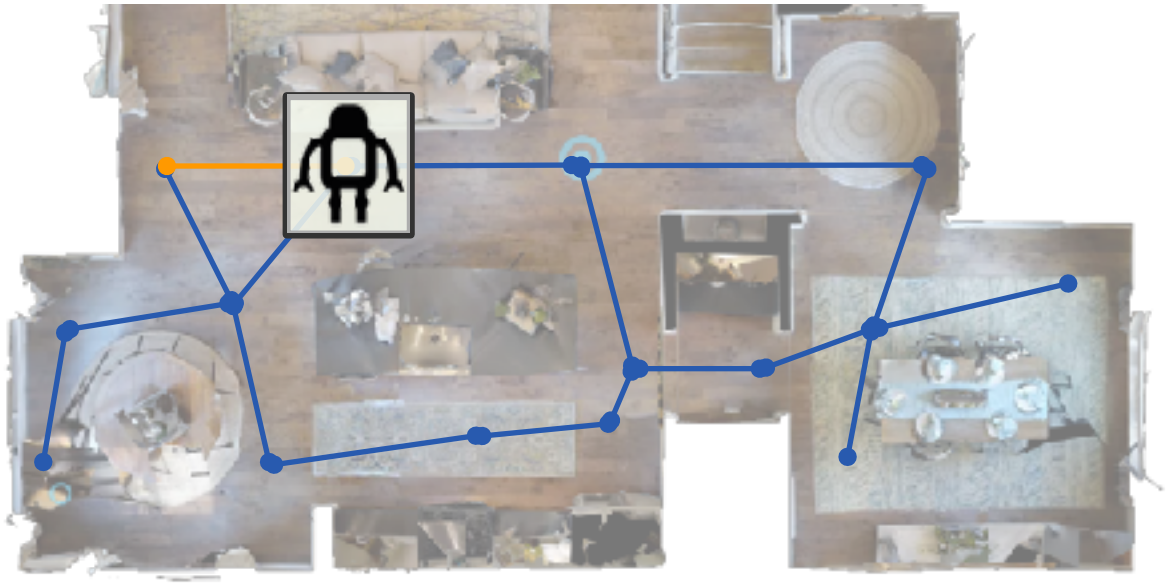




[Anderson et al., 2018]

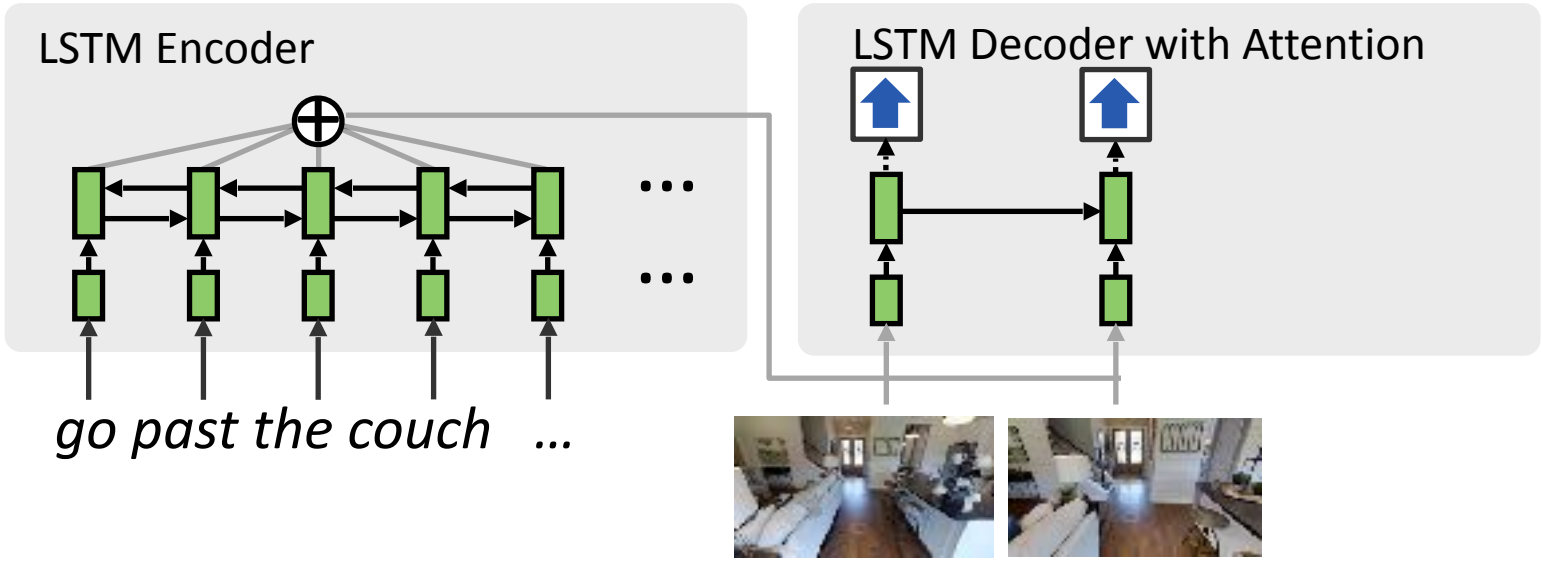
Literal Listener

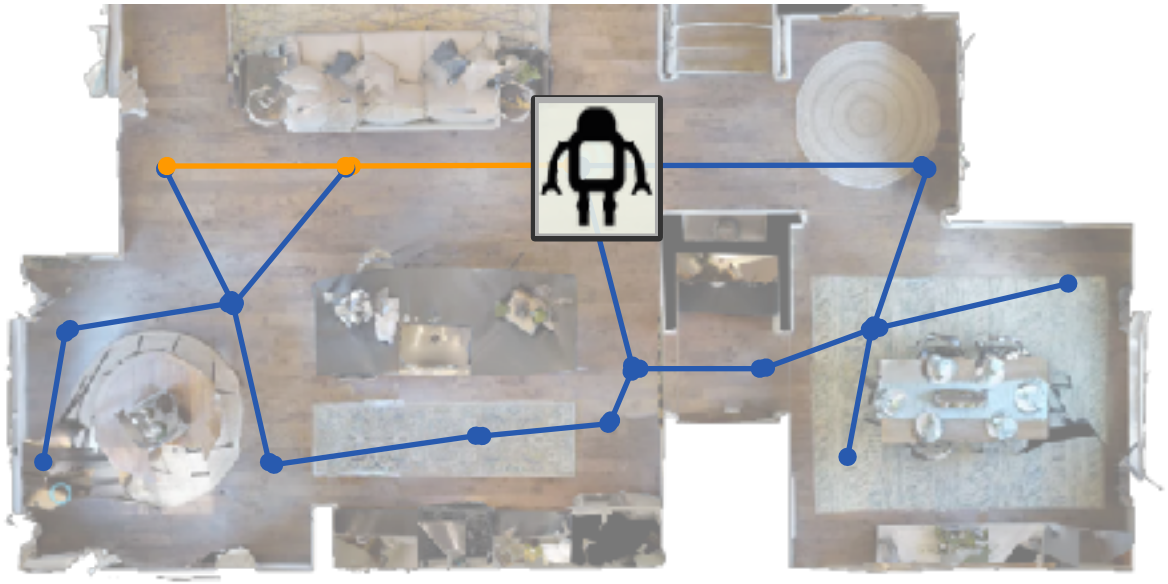




[Anderson et al., 2018]

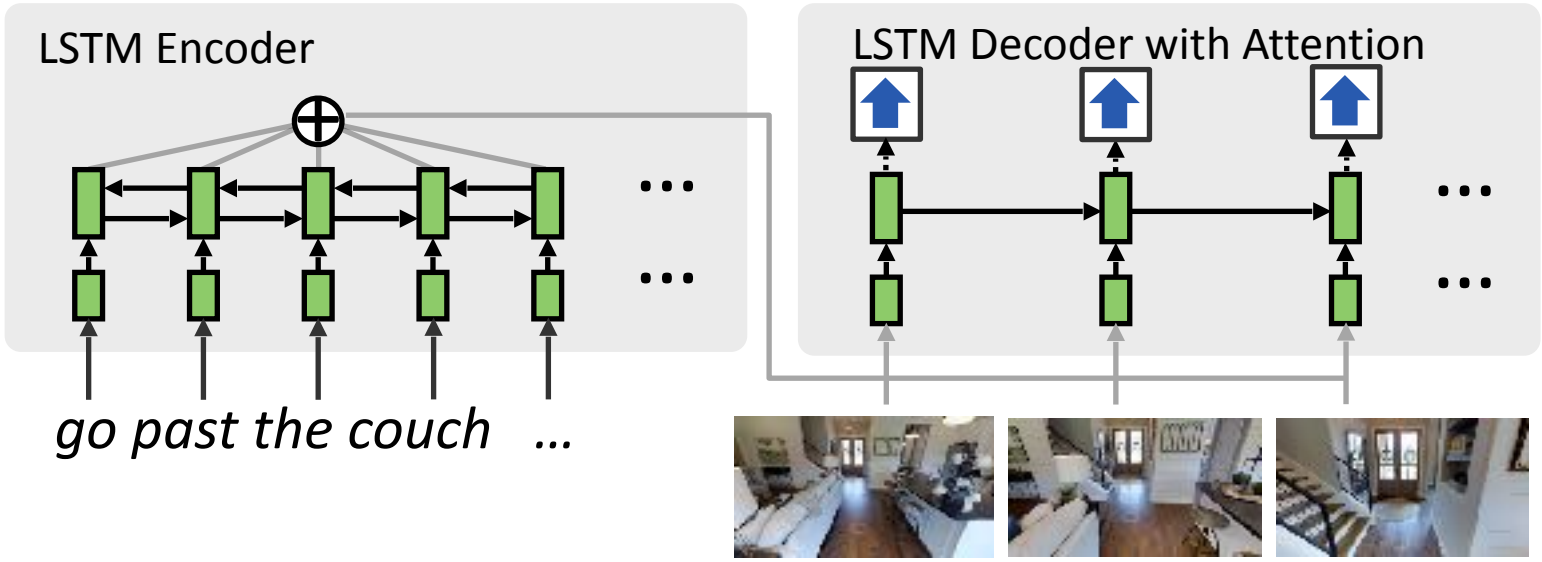
Literal Listener

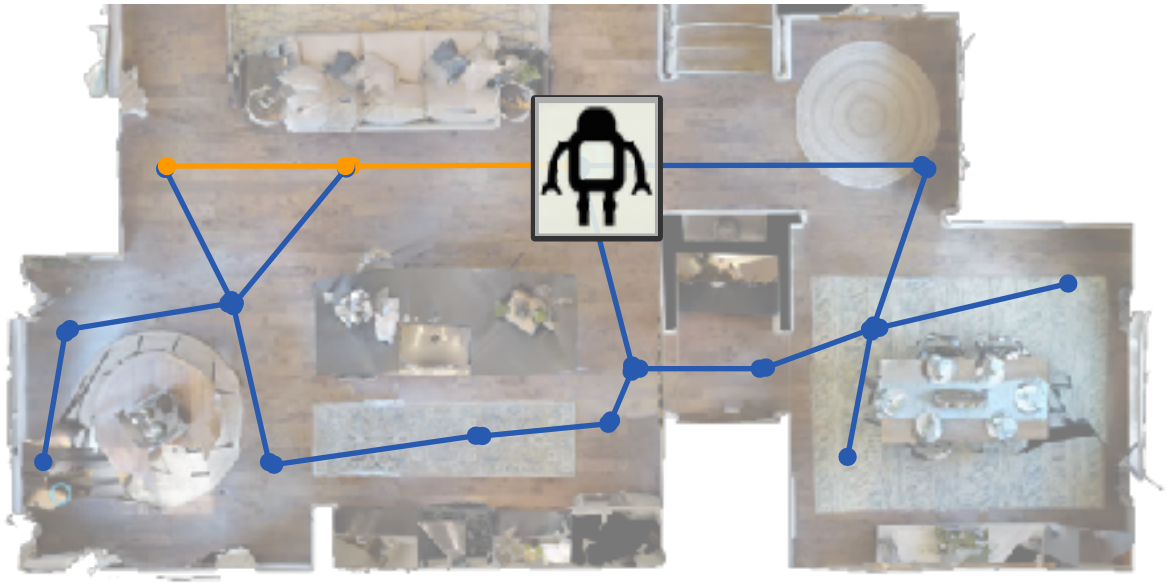




[Anderson et al., 2018]

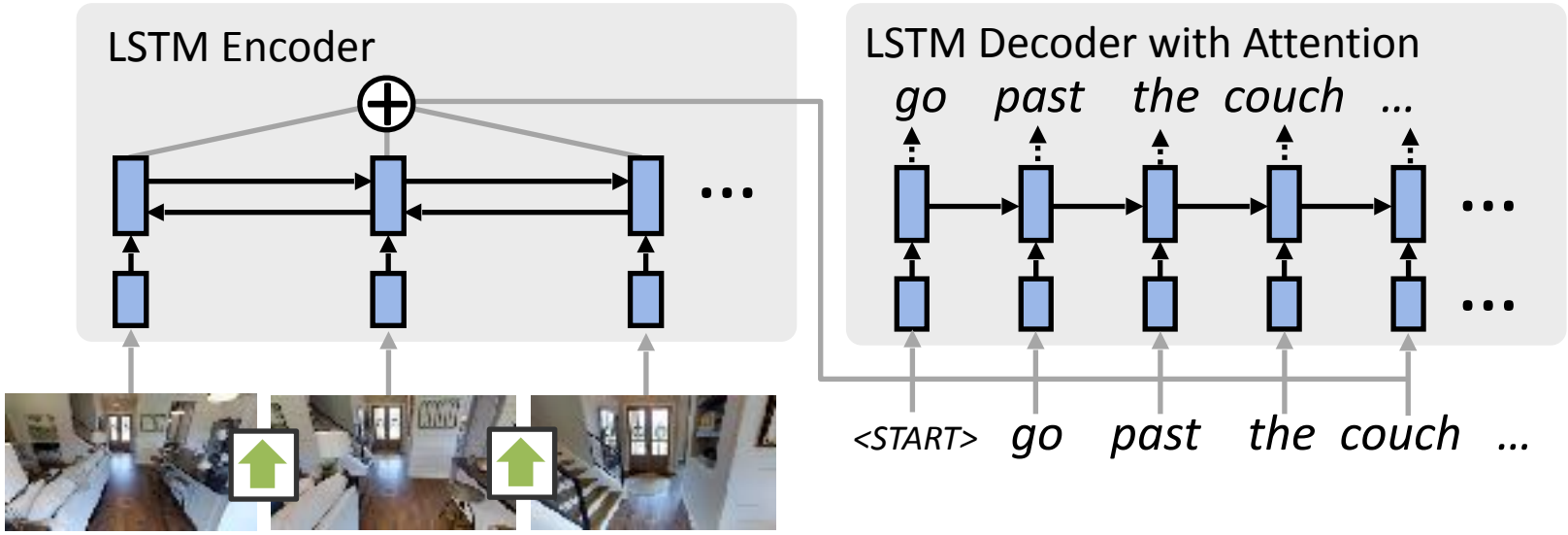
Literal Listener

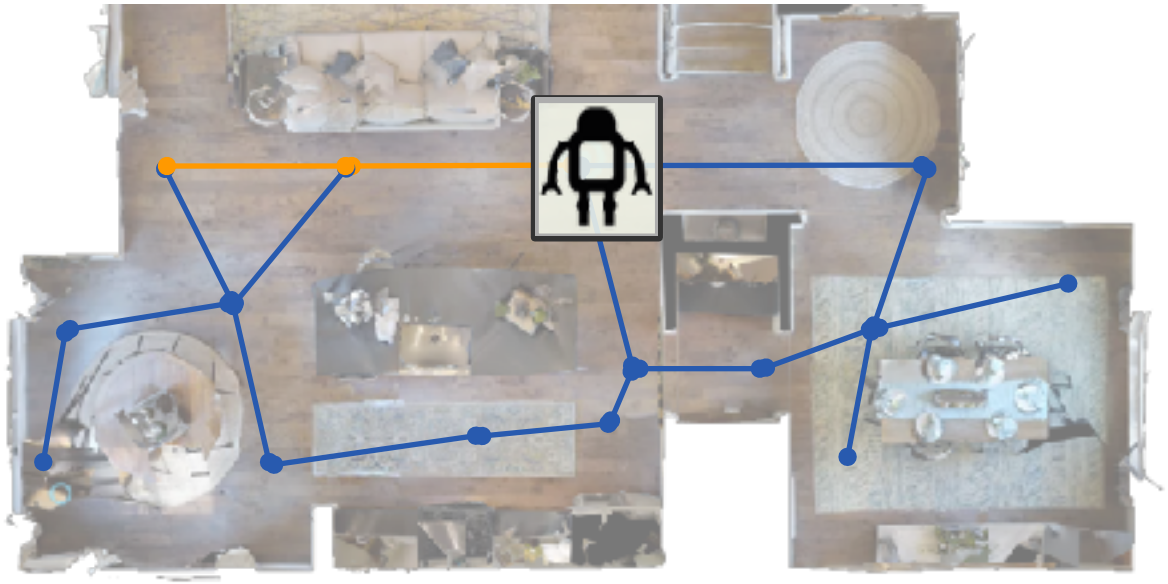




[Anderson et al., 2018]

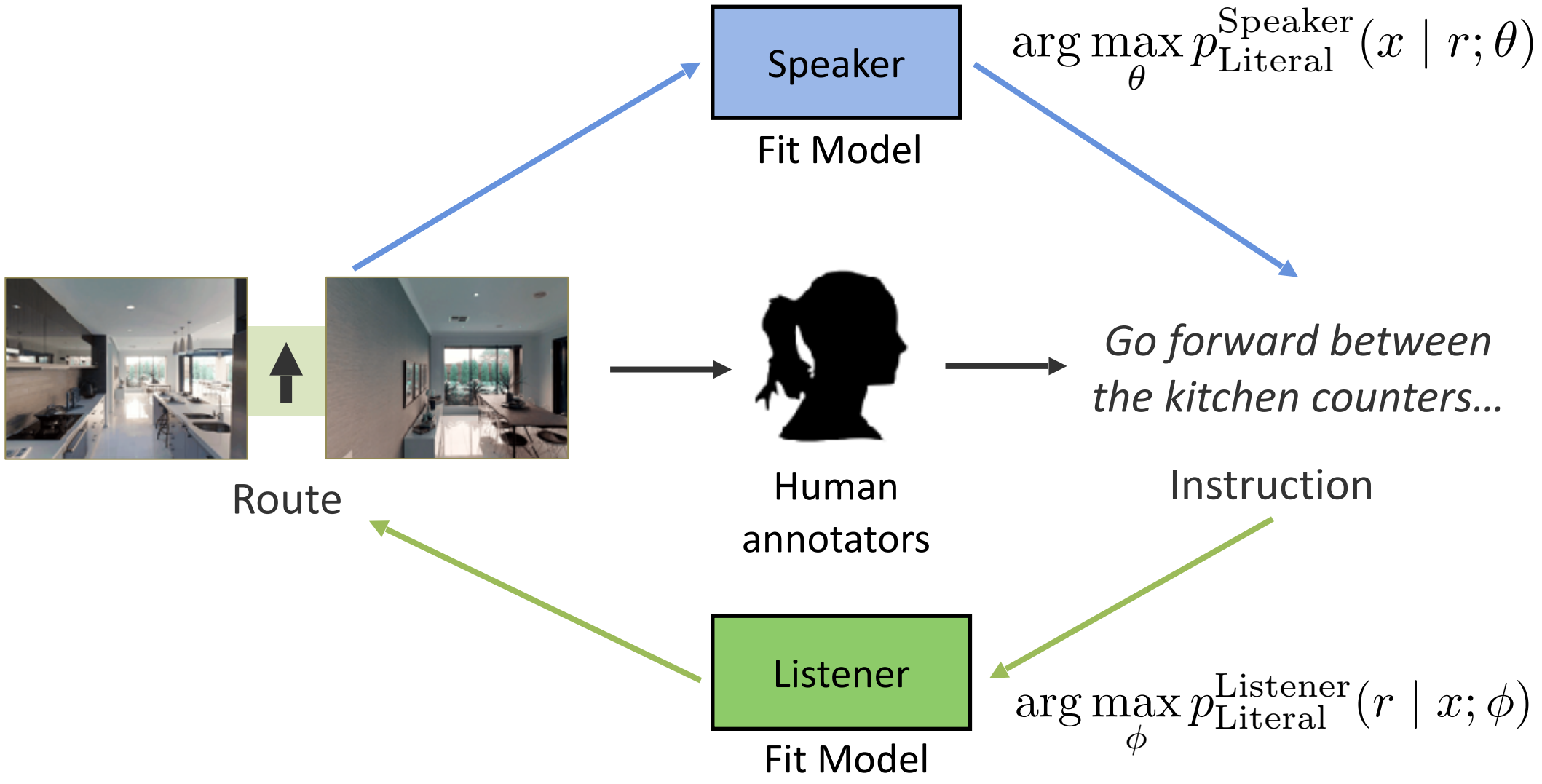
Literal Speaker





[Fried et al., 2018]

Training Literal Listener and Speaker



Pragmatic Instruction Generation

walk past the dining room table and chairs and take a right into the living room.

Speaker

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

Listener

Listener

Pragmatic Instruction Generation

walk past the dining room table and chairs and take a right into the living room.

Speaker

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

Listener

Listener

Pragmatic Instruction Generation

walk past the dining room table and chairs and take a right into the living room.

Speaker

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

Listener

0.4

$$p_{\text{Literal}}^{\text{Listener}}(r \mid x; \phi)$$

Listener

0.8