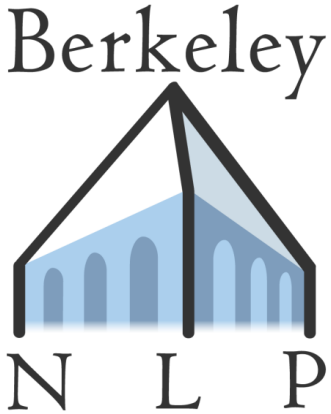
Learning Grounded Pragmatic Communication

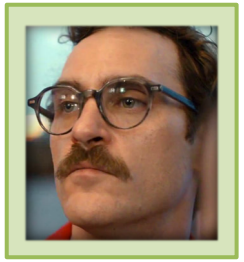


Daniel Fried

Natural Language Interfaces

Science Fiction

Her, 2013



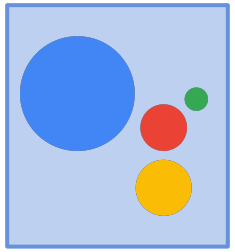
Let's start with your emails. You have several thousand emails regarding LA Weekly, but it looks like you haven't worked there in many years.

Oh yeah, I guess I was saving those because in some of them I thought I might have written some funny stuff.

Yeah, there are some funny ones. I'd say there are about 86 that we should save. We can delete the rest.

In Reality

Google Assistant, 2017



Who are you?

I'm your Google Assistant.

And I can let you know if you'll need a jacket today.

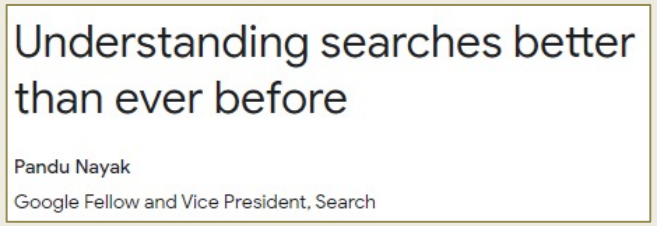
Do I?

Sorry, I don't understand.

Context in NLP

Other Language

Language Modeling, Structure & Semantics



This Talk

The World

Grounding

“Take me to the airport”

Intents and Effects

Pragmatics

“My neck hurts”

Grounding and Pragmatics

Grounding

“Stop at the second car”

Pragmatics

“Stop at the car”

Pragmatics and Reasoning

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]

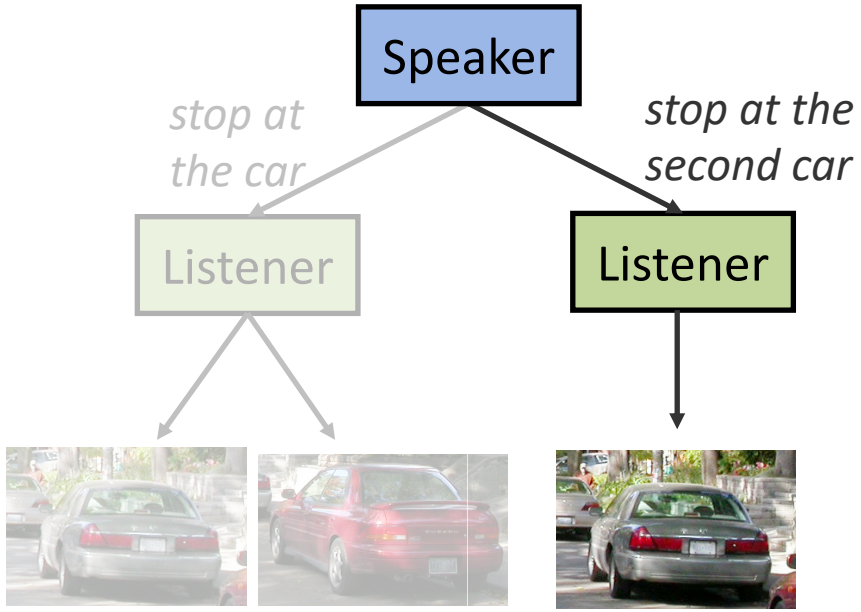
Our talk exchanges … are cooperative efforts… One of my avowed aims is to see talking as purposive, indeed rational, behavior.

[*Logic and Conversation.* Grice, 1975]

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

Pragmatics and Reasoning

Generation

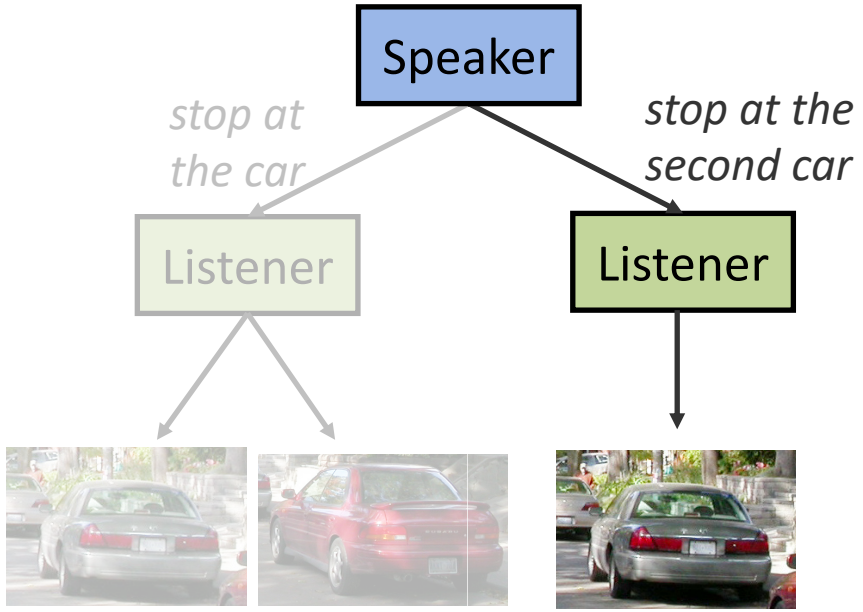


Interpretation

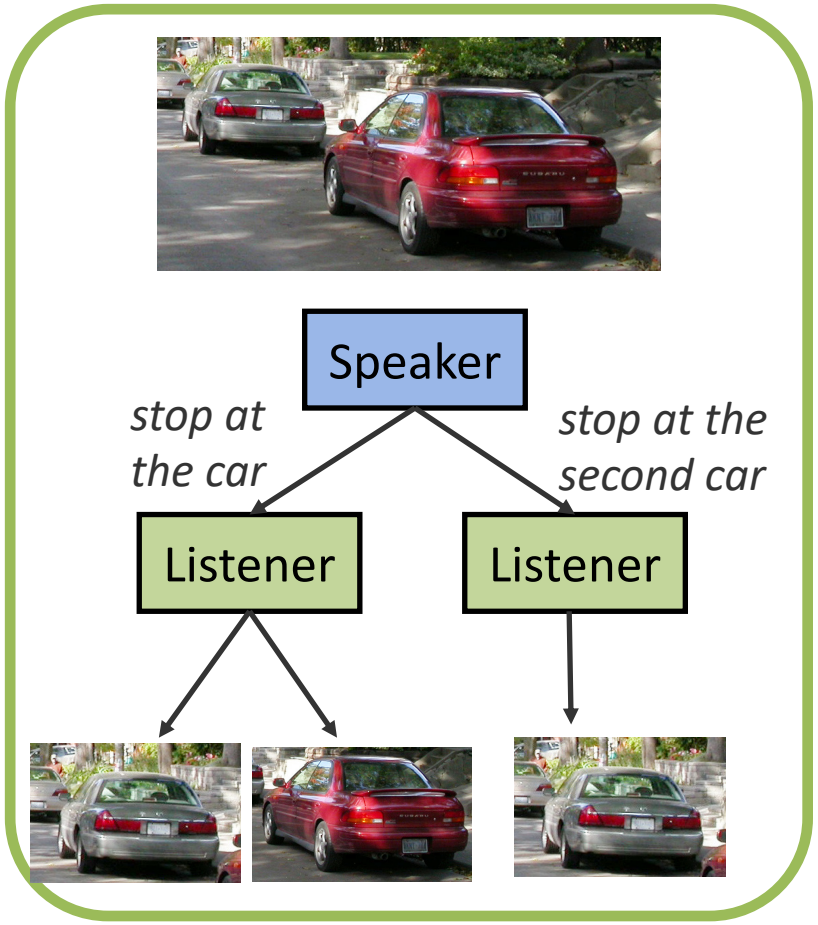
[e.g. Lewis 1969; Golland et al. 2010; Frank and Goodman 2012; Degen et al. 2013]

Pragmatics and Reasoning

Generation



Interpretation



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

Pragmatics and Reasoning

Generation

Speaker

stop at the car

stop at the second car

Listener

Listener

Interpretation

stop at the car

Listener

Speaker

stop at the car

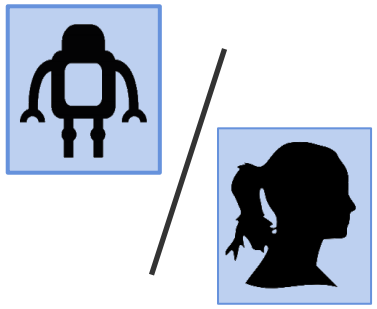
stop at the second car

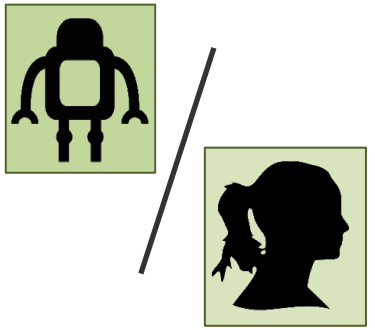
Listener

Listener

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

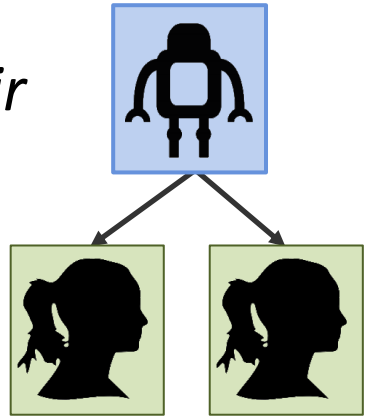
Reasoning with Speakers and Listeners





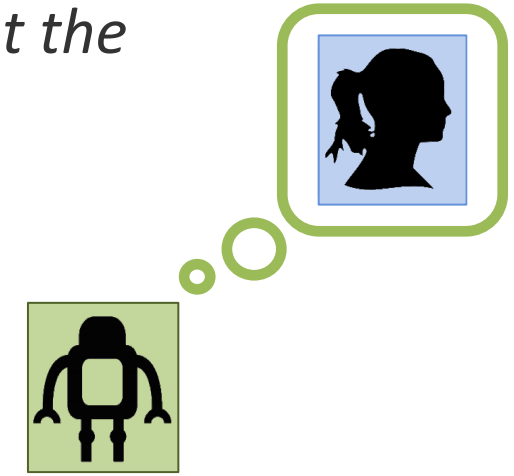
Pragmatics and Generation

walk along the wood path to the chair

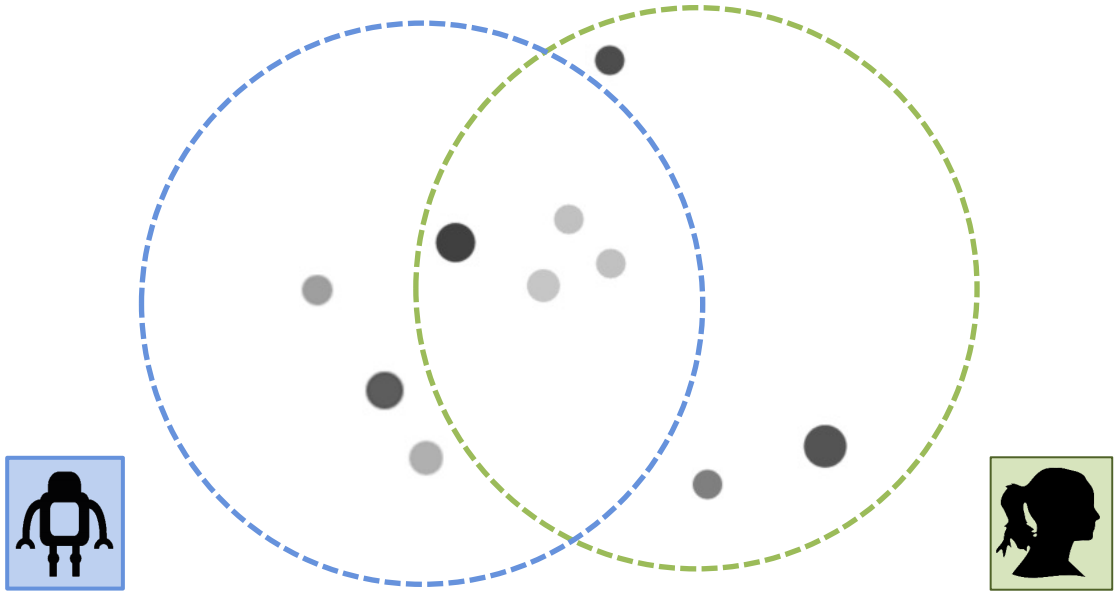


Pragmatics and Interpretation

Turn left and take a right at the table. Take a left at the painting and then take your first right.



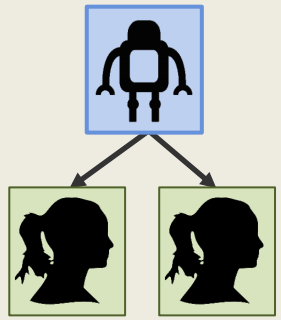
Pragmatics and Dialogue



Pragmatics and…

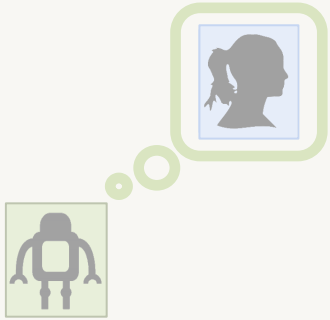
Generation

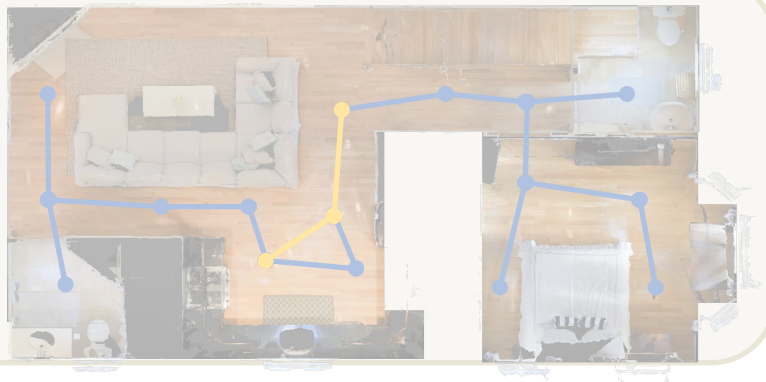
[**Fried**, Andreas, & Klein. NAACL 2018]



Interpretation

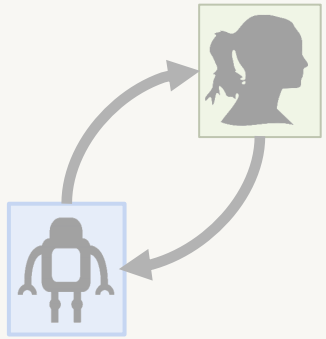
[**Fried***, Hu*, Cirik* et al. NeurIPS 2018]





Dialogue

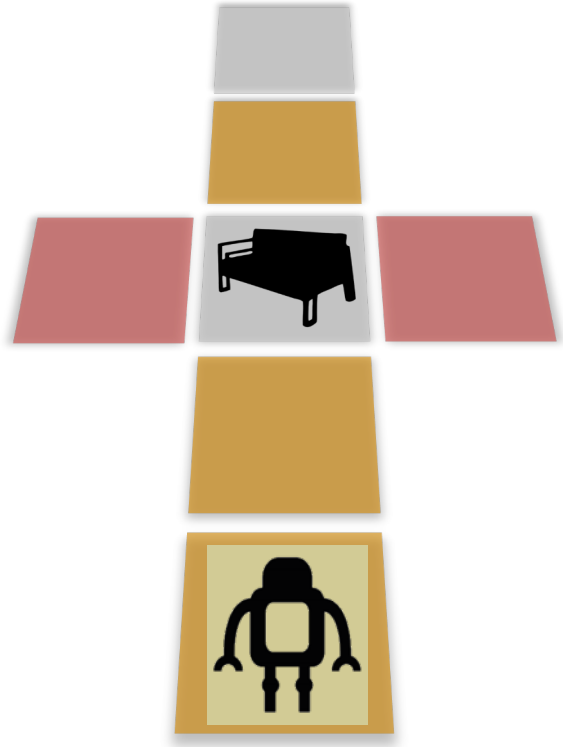
[**Fried**, Chiu, & Klein. In submission]



To Start: Virtual Environments

Human View:

Agent View:

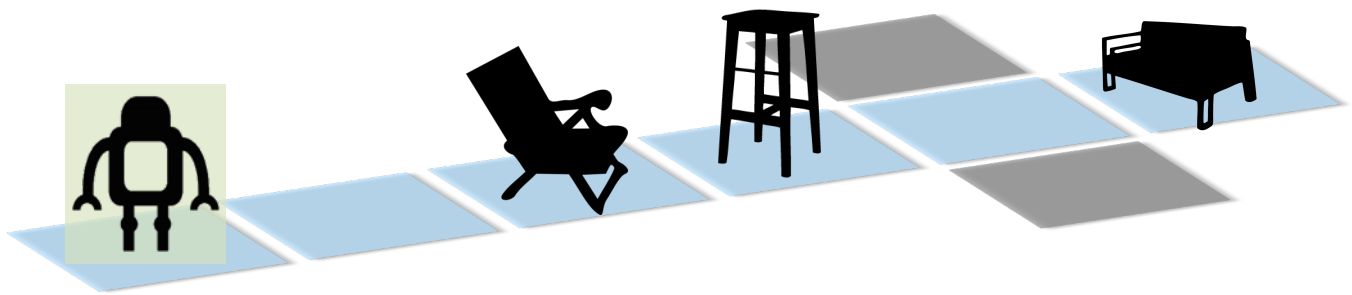


SAIL [MacMahon et al., 2006; Chen and Mooney, 2011]

Interpretation Task

Input instruction: *go forward to the grey hallway*

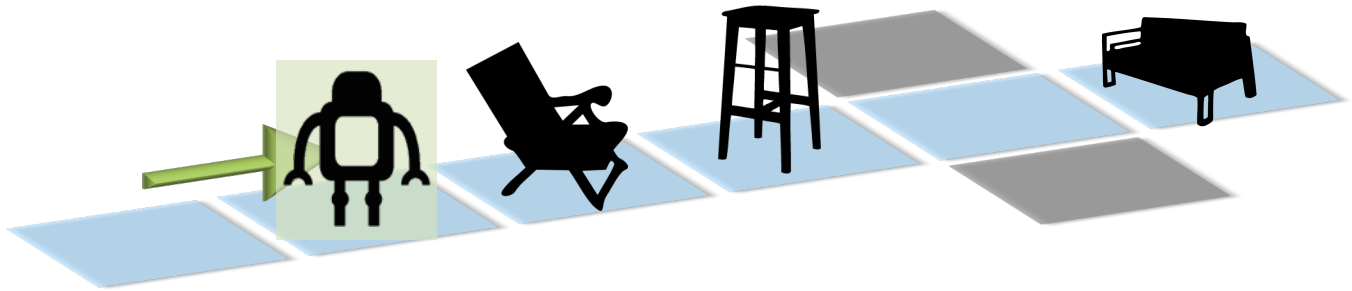
Output actions:



Interpretation Task

Input instruction: *go forward to the grey hallway*

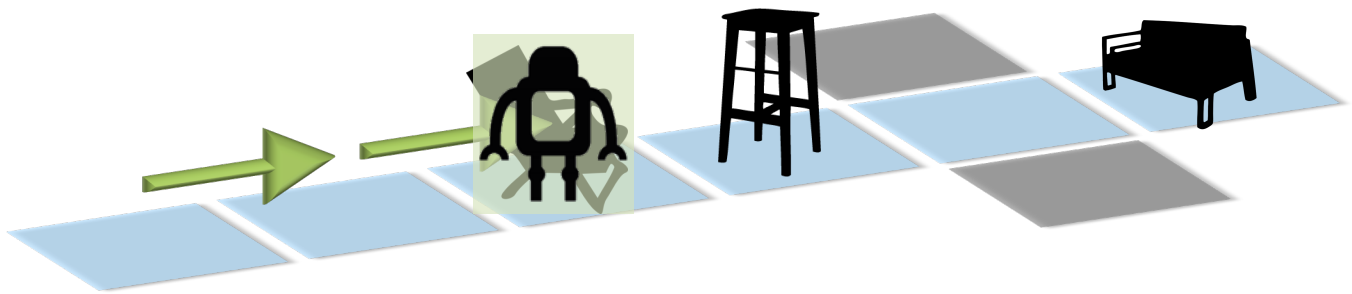
Output actions:



Interpretation Task

Input instruction: *go forward to the grey hallway*

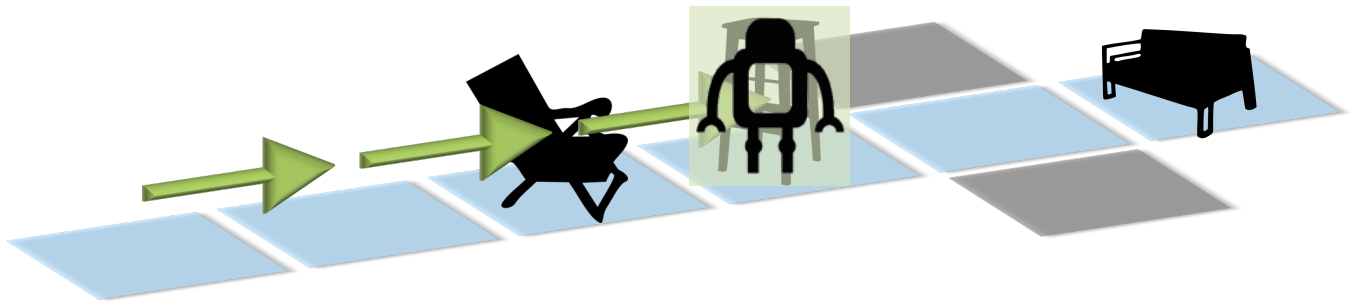
Output actions:



Interpretation Task

Input instruction: *go forward to the grey hallway*

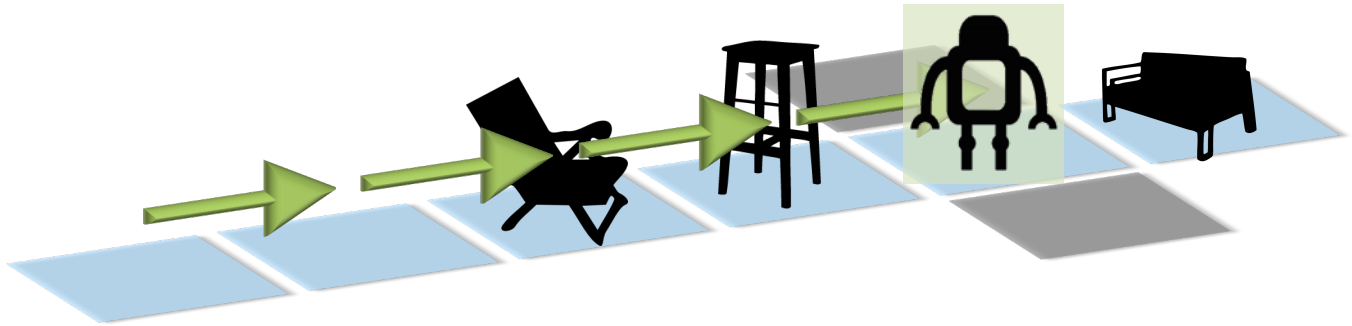
Output actions:



Interpretation Task

Input instruction: *go forward to the grey hallway*

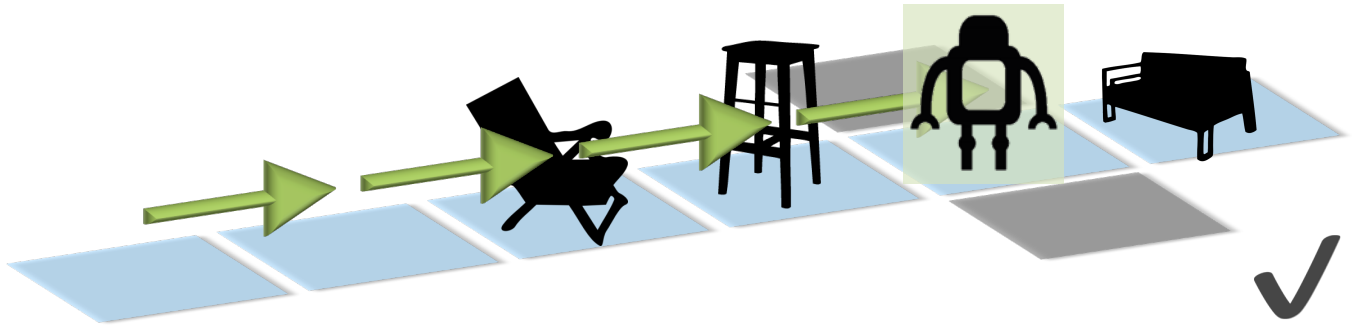
Output actions:



Interpretation Task

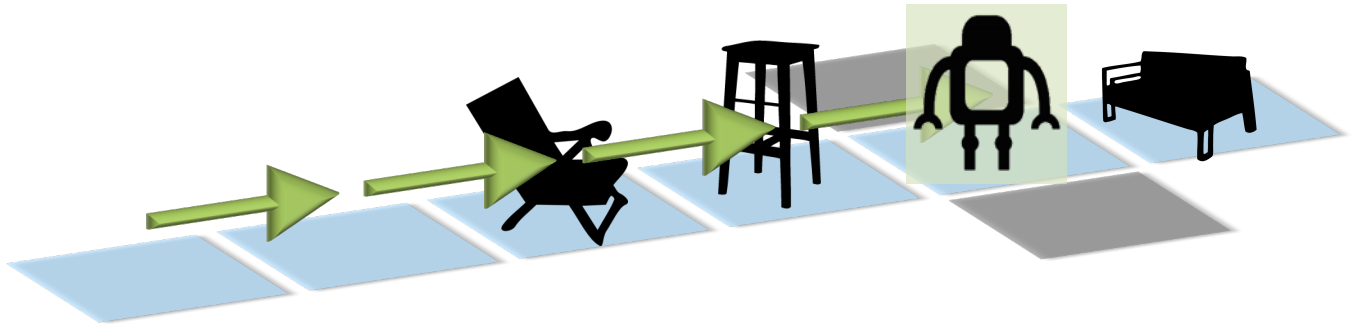
Input instruction: *go forward to the grey hallway*

Output actions:



Generation Task

Input actions:



Output Instruction: *go forward to the grey hallway*

Models of Listeners and Speakers

Inputs

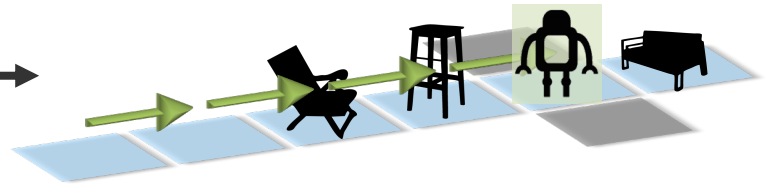
Outputs

go forward to the grey hallway

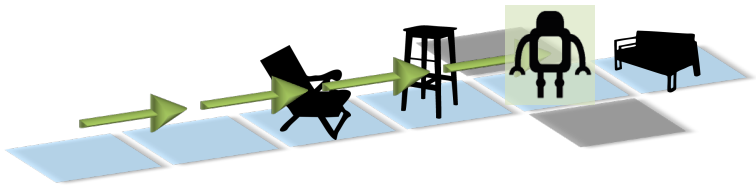
Instruction, i

Listener

$P_L(a \mid i)$



Actions, a



Actions, a

Speaker

$P_S(i \mid a)$

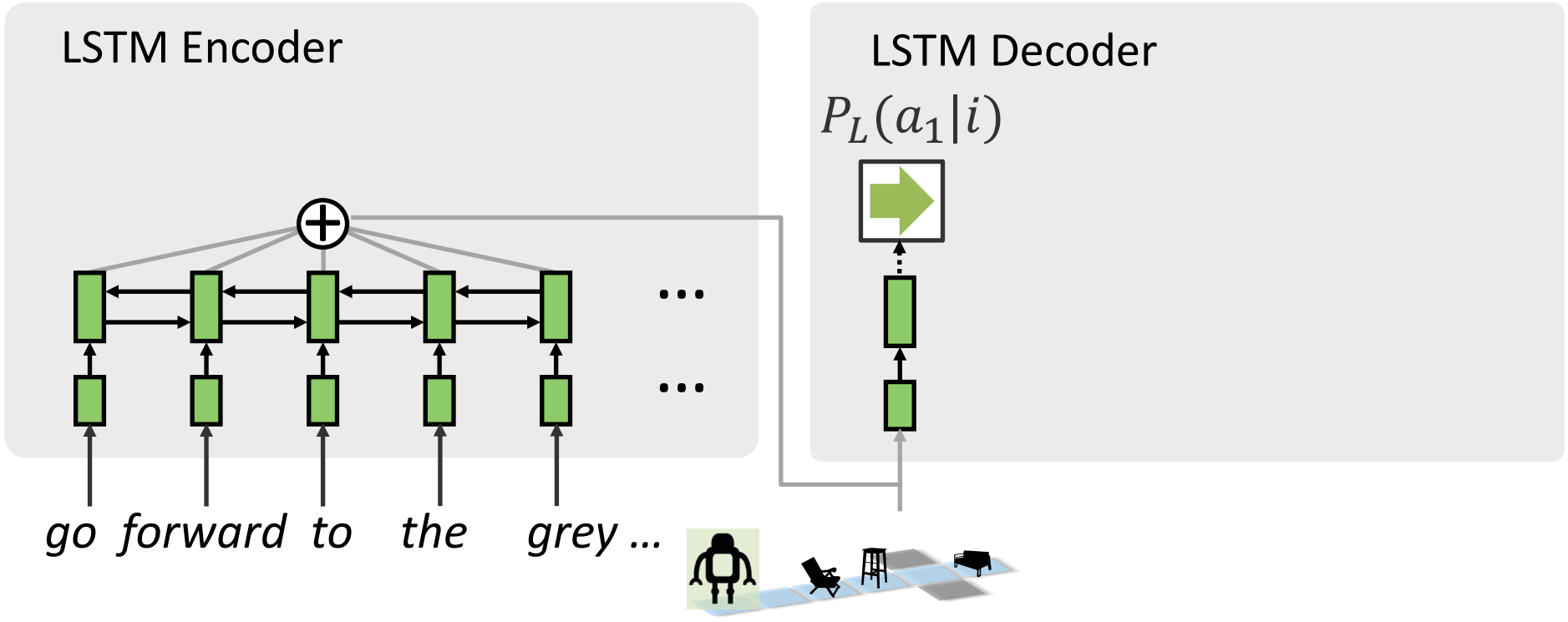
go forward to the grey hallway

Instruction, i

Base Models

Base Listener

$$P_L(a \mid i) = \prod_t P_L(a_t | a_{1:t-1}, i)$$

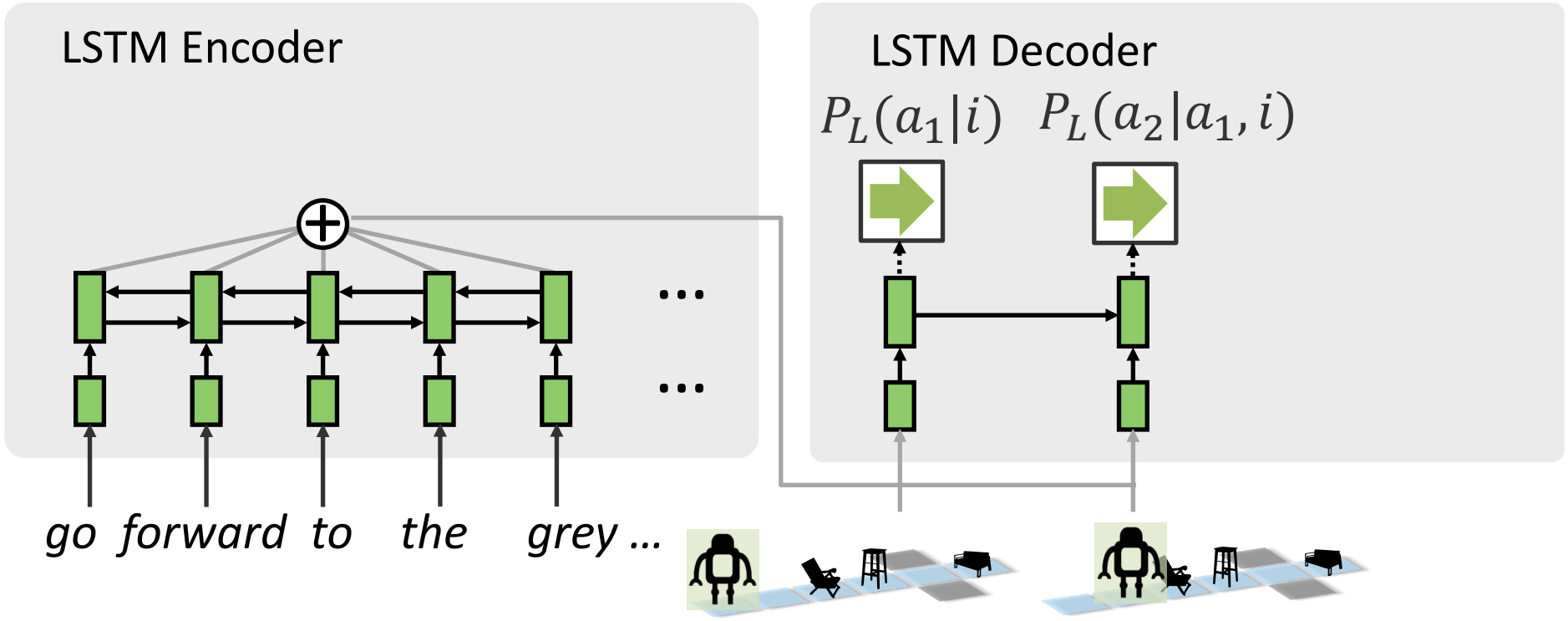


Based on Mei et al. [2016]

Base Models

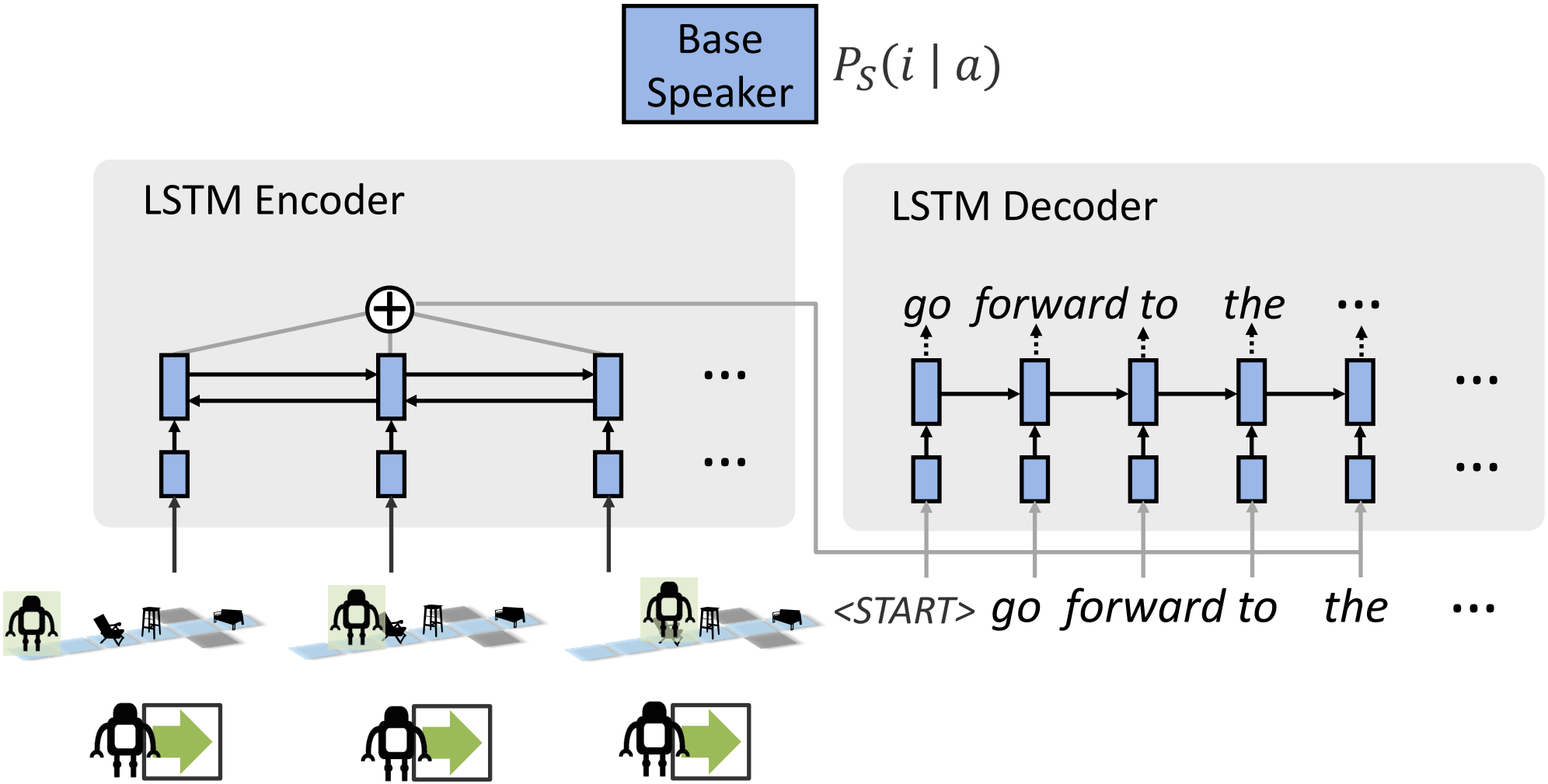
Base Listener

$$P_L(a \mid i) = \prod_t P_L(a_t | a_{1:t-1}, i)$$

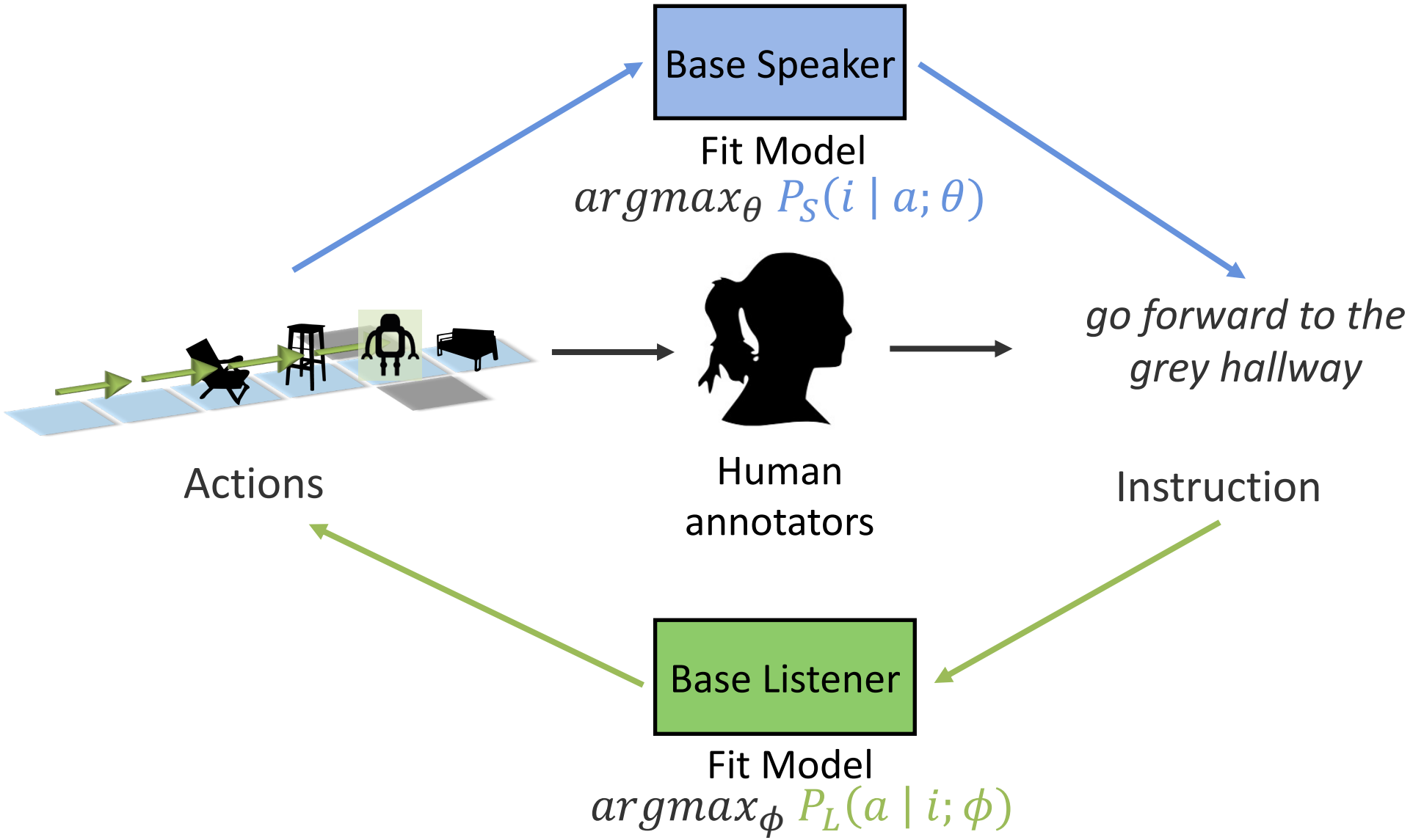


Based on Mei et al. [2016]

Base Models

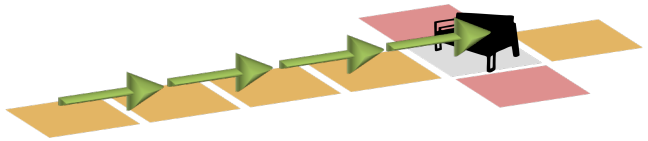

Base Speaker
$P_S(i \mid a)$
LSTM Encoder
LSTM Decoder
go forward to the ...
<START> go forward to the ...

Training Models on Human Instructions



Speaker Tasks and Evaluation

Speaker produces an instruction



Speaker

walk along the wood path to the chair

Humans try to interpret it

walk along the wood path to the chair

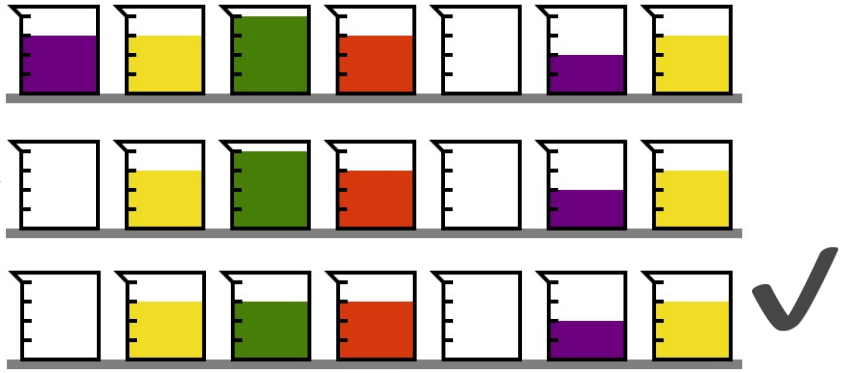
Human direction followers (MTurk)

SAIL navigation [MacMahon et al., 2006; Chen and Mooney, 2011]

Speaker Tasks and Evaluation

Alchemy

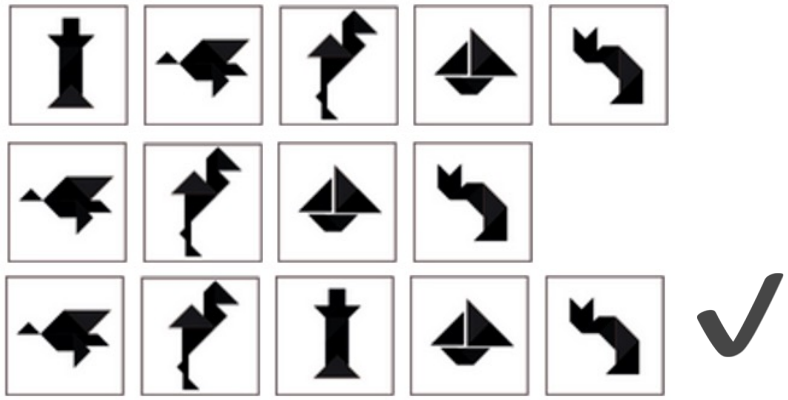
1. remove all the purple chemical from the beaker on the far left
2. do the same with one unit of green chemical
3. …



Human direction followers (MTurk)

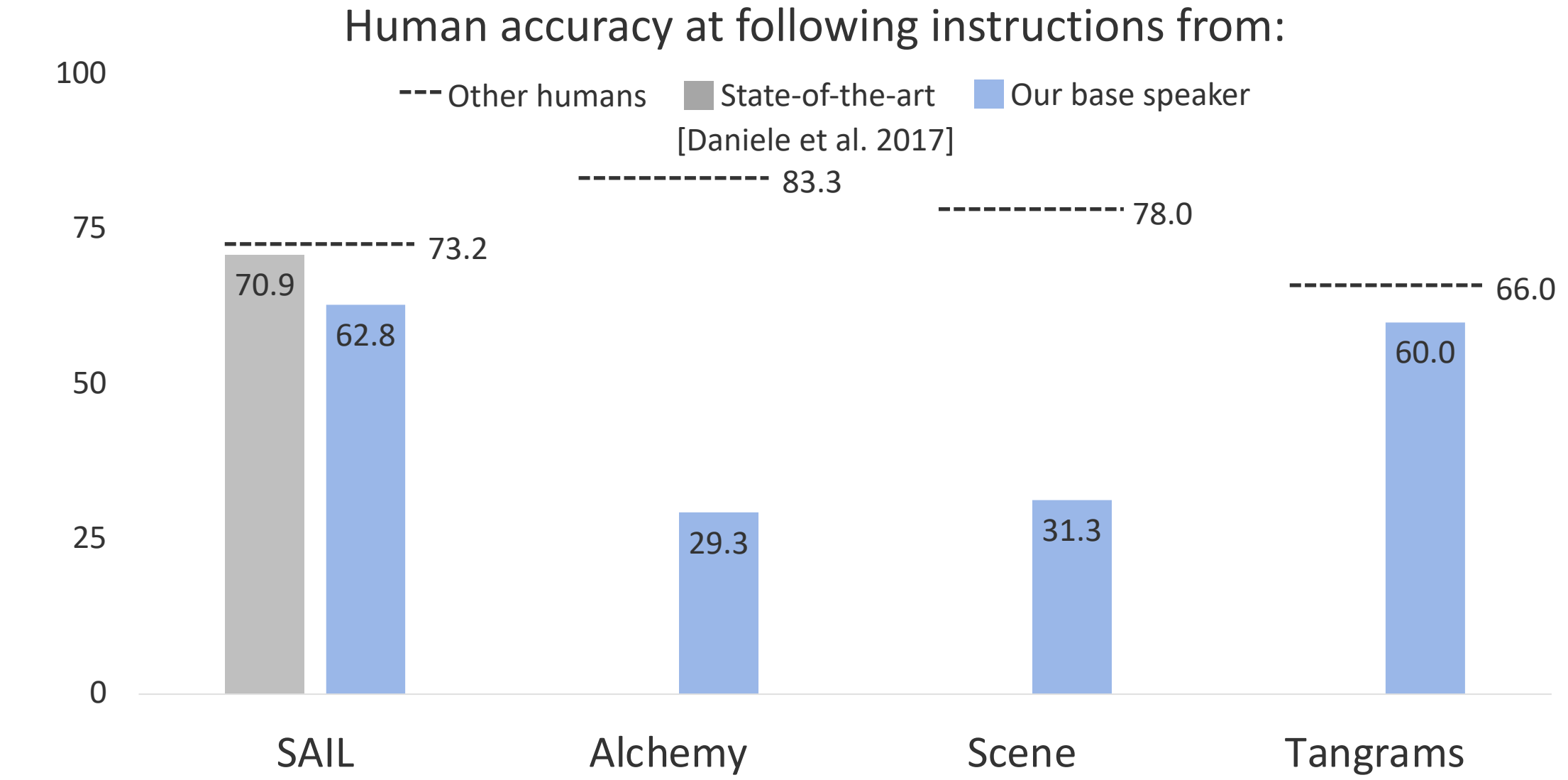
Tangrams

1. remove first figure
2. add it back into middle spot
3. …

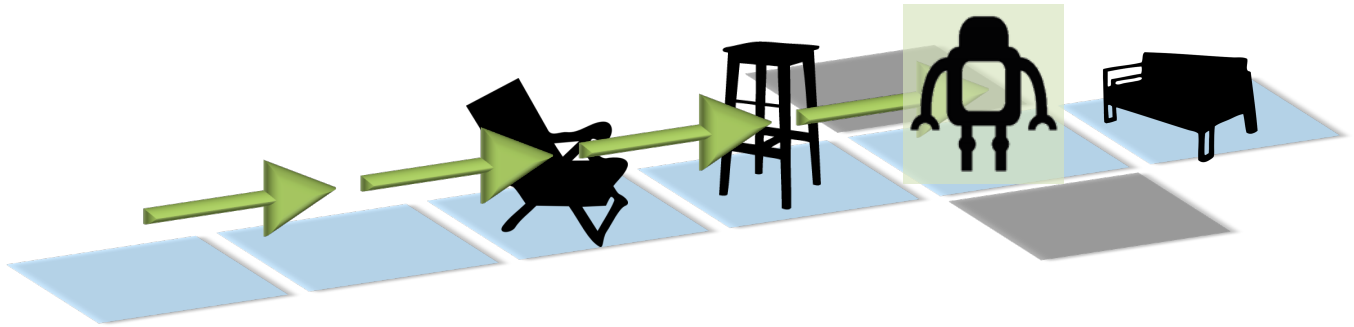


SCONE contextual instruction following [Long et al. 2016]

Generation is Hard to Imitate!



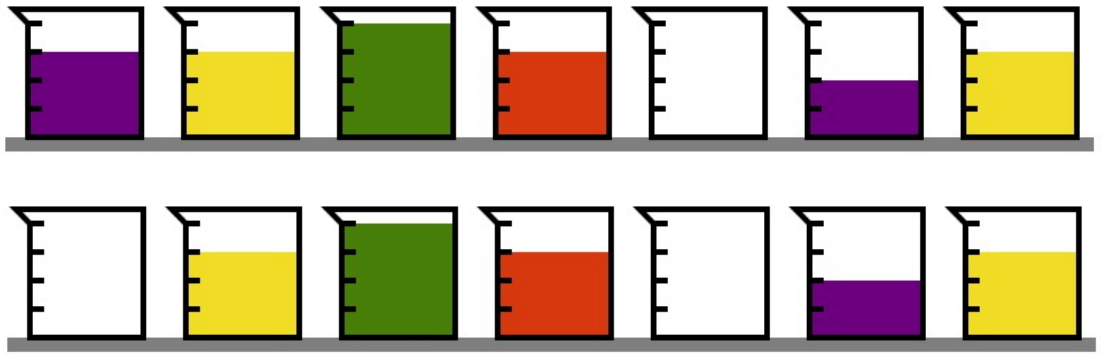
A Failure Mode: Underspecification



Base Speaker

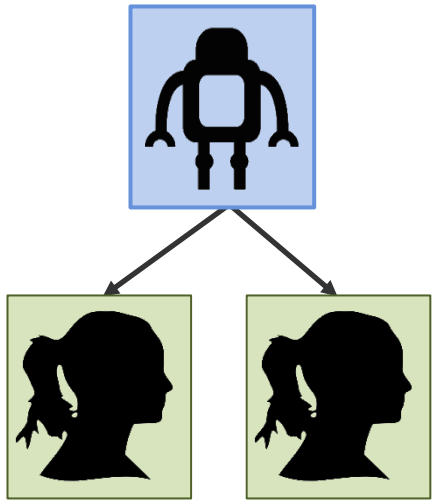
go forward past the stool ?

A Failure Mode: Contextual Ambiguity



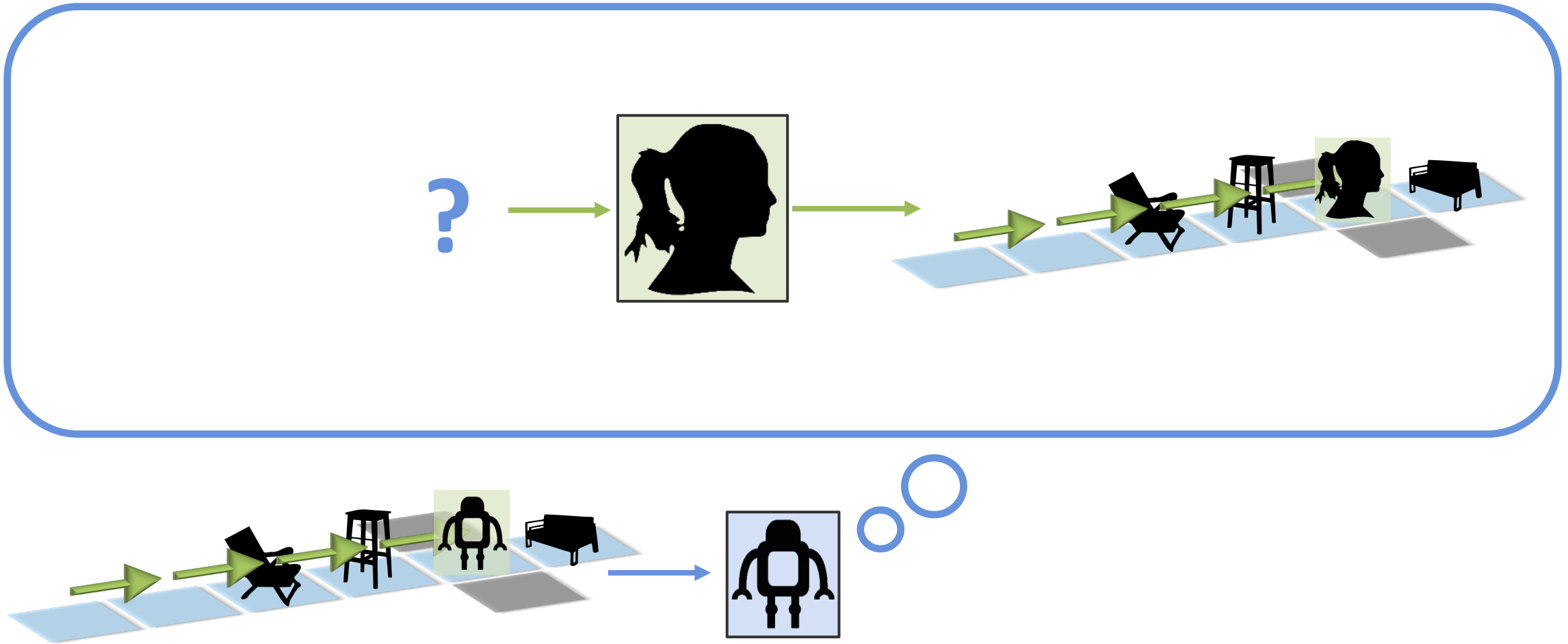
Base Speaker

throw out the purple chemical ✗



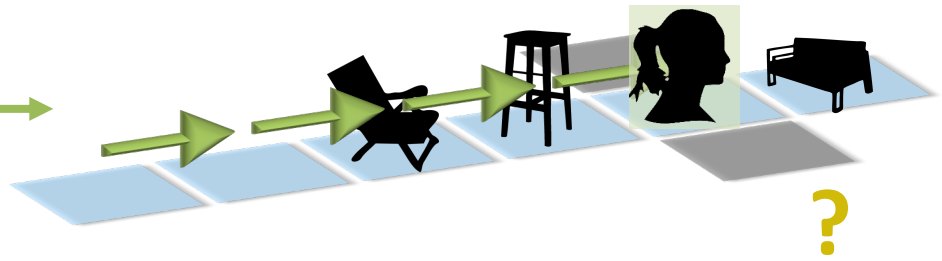
Making Text Informative with Pragmatic Speakers

Pragmatic Speakers Simulate Interpretation

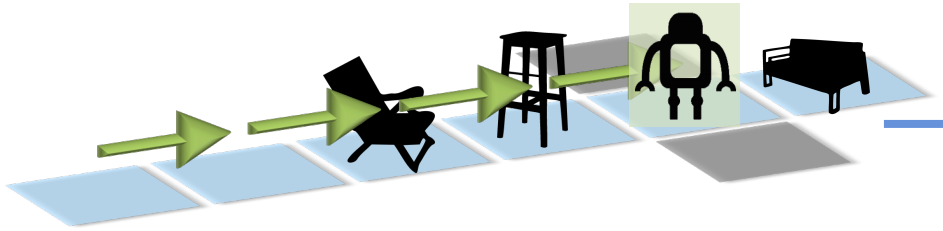


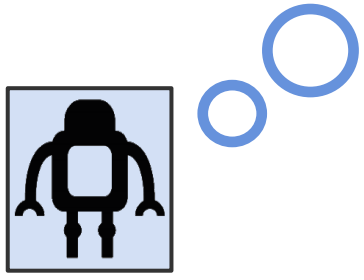
Pragmatic Speakers Simulate Interpretation

go forward
past the stool



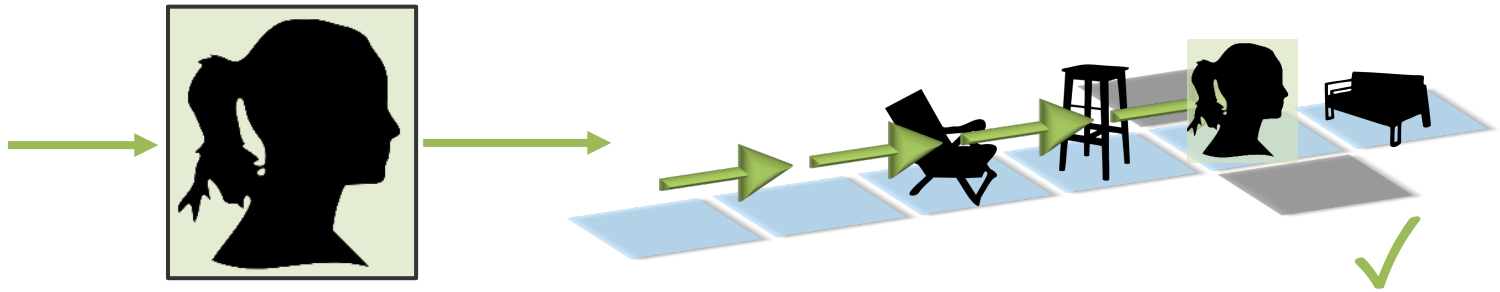
?

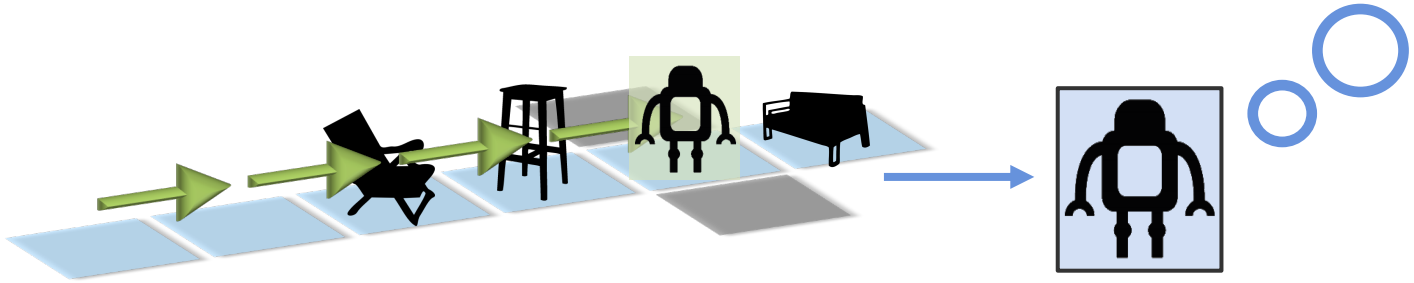




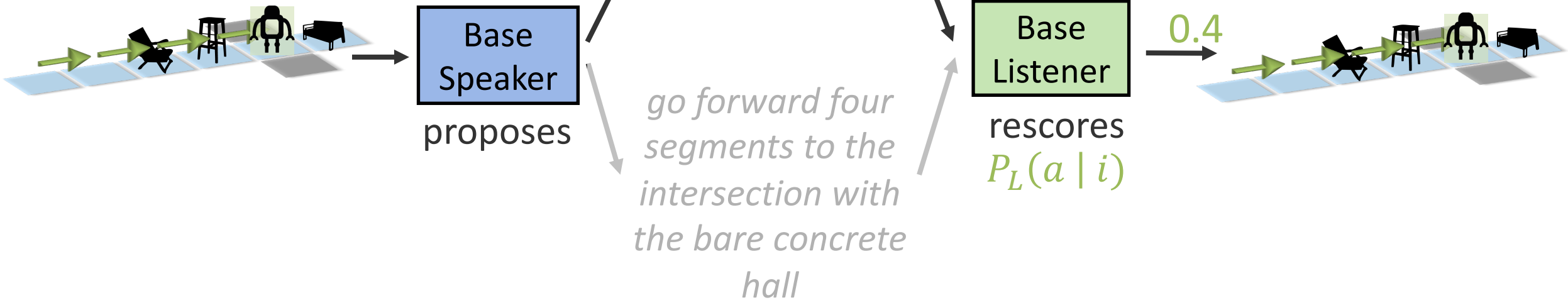
Pragmatic Speakers Simulate Interpretation

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

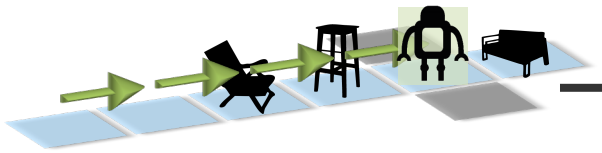




Building a Pragmatic Speaker



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

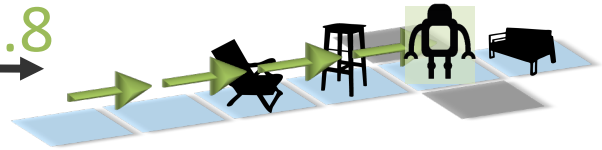
0.8

Base Listener

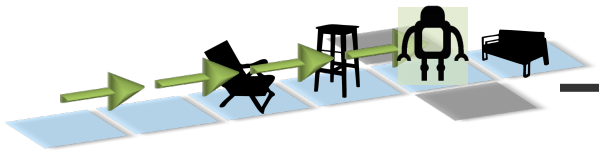
rescores

$P_L(a \mid i)$

0.8



Building a Pragmatic Speaker



Base Speaker

proposes

0.4

walk forward past the stool

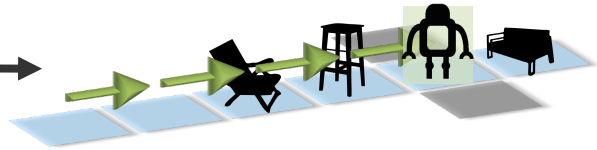
go forward four segments to the intersection with the bare concrete hall

0.8

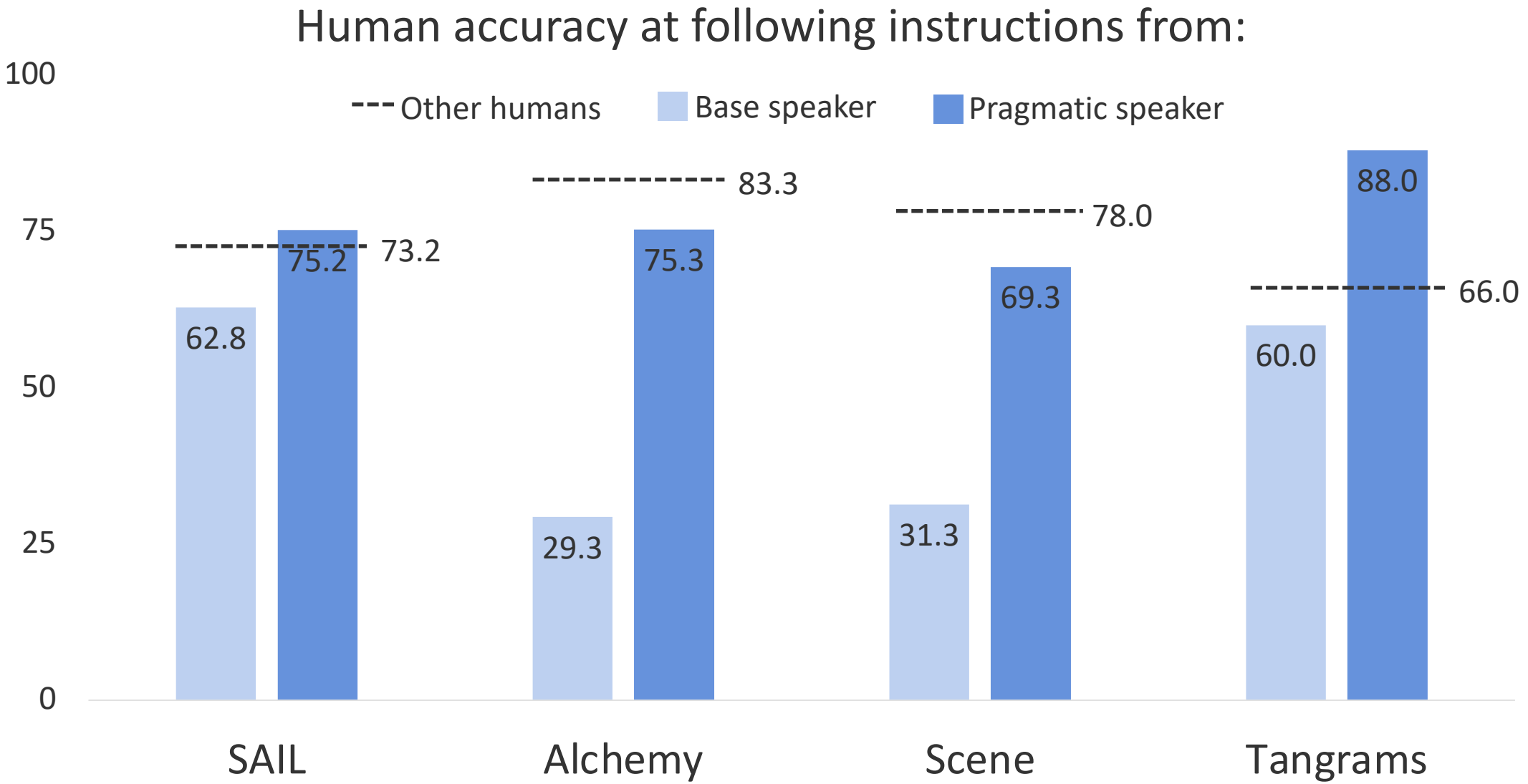
Base Listener

rescores

$P_L(a \mid i)$



Speaker Results

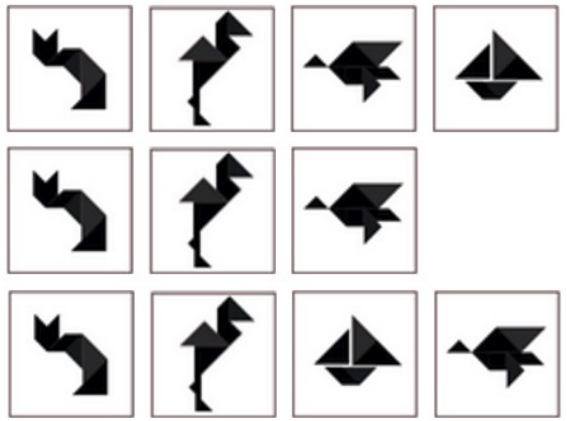


[**Fried**, Andreas, and Klein. NAACL 2018]

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*	✗

Pragmatic Speakers in Other Domains

Document Summarization

Input:

… The 1-0 scoreline that took Barcelona through to the Champions League quarterfinals made their clash with Manchester City all seem rather academic. ….

Pragmatic Output:

Barcelona beat Manchester City 1-0 in the Champions League.

[Shen, **Fried**, Andreas, & Klein. NAACL 2019]

Image Captioning

Input:

Pragmatic Output:

two giraffes standing in a large enclosure with a building in the background

[in preparation]

Visual Navigation

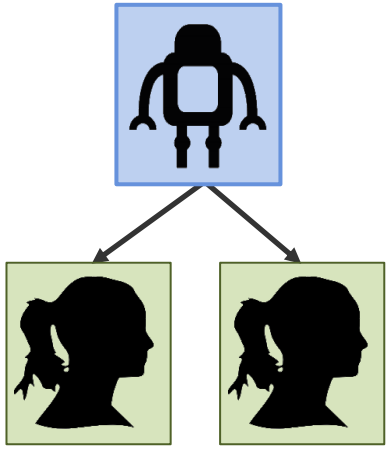
Input:

Pragmatic Output:

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

[**Fried***, Hu*, Cirik* et al. NeurIPS 2018]

Takeaways



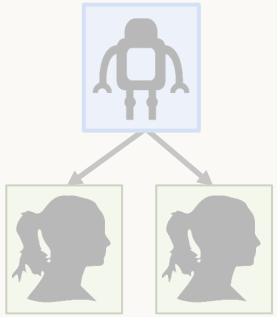
Simulating people's interpretations makes language more informative.

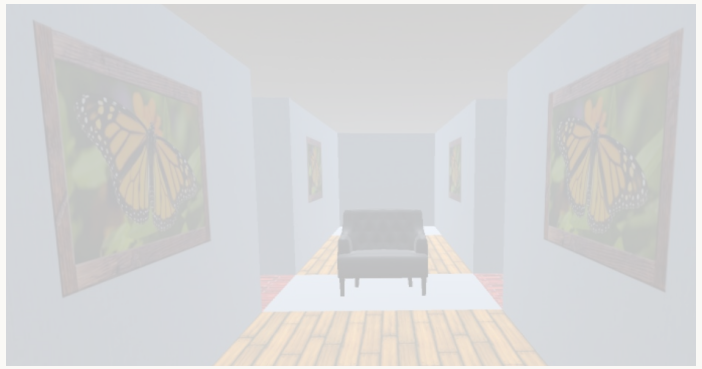
Pragmatics allows models to sometimes outperform their training data.

Pragmatics and…

Generation

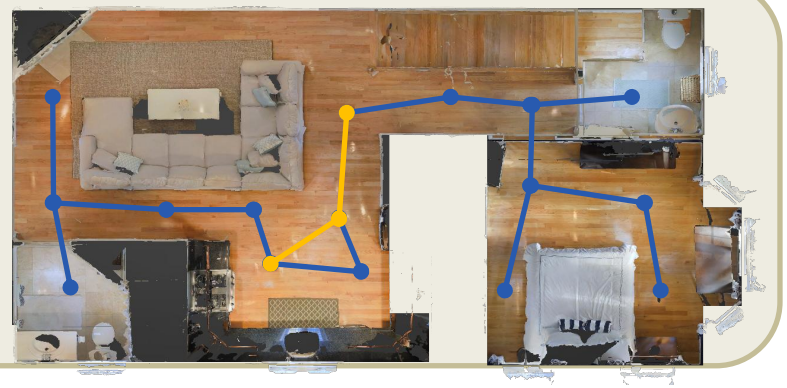
[**Fried**, Andreas, & Klein. NAACL 2018]





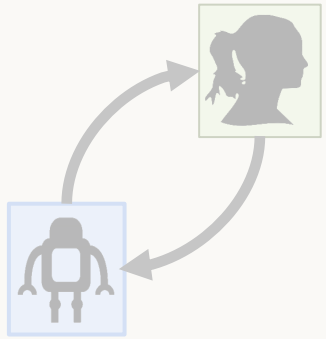
Interpretation

[**Fried***, Hu*, Cirik* et al. NeurIPS 2018]



Dialogue

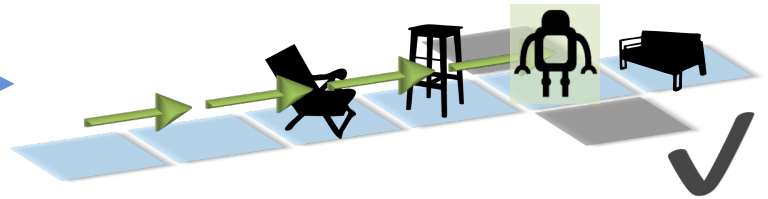
[**Fried**, Chiu, & Klein. In submission]



Listener Tasks

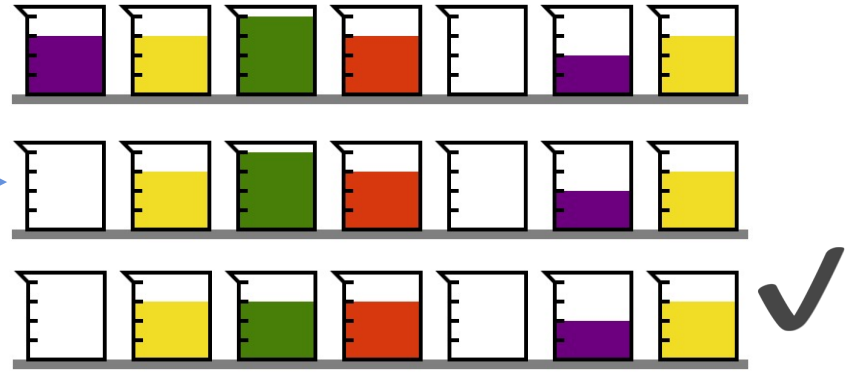
Navigation

go forward to the grey hallway → Listener →

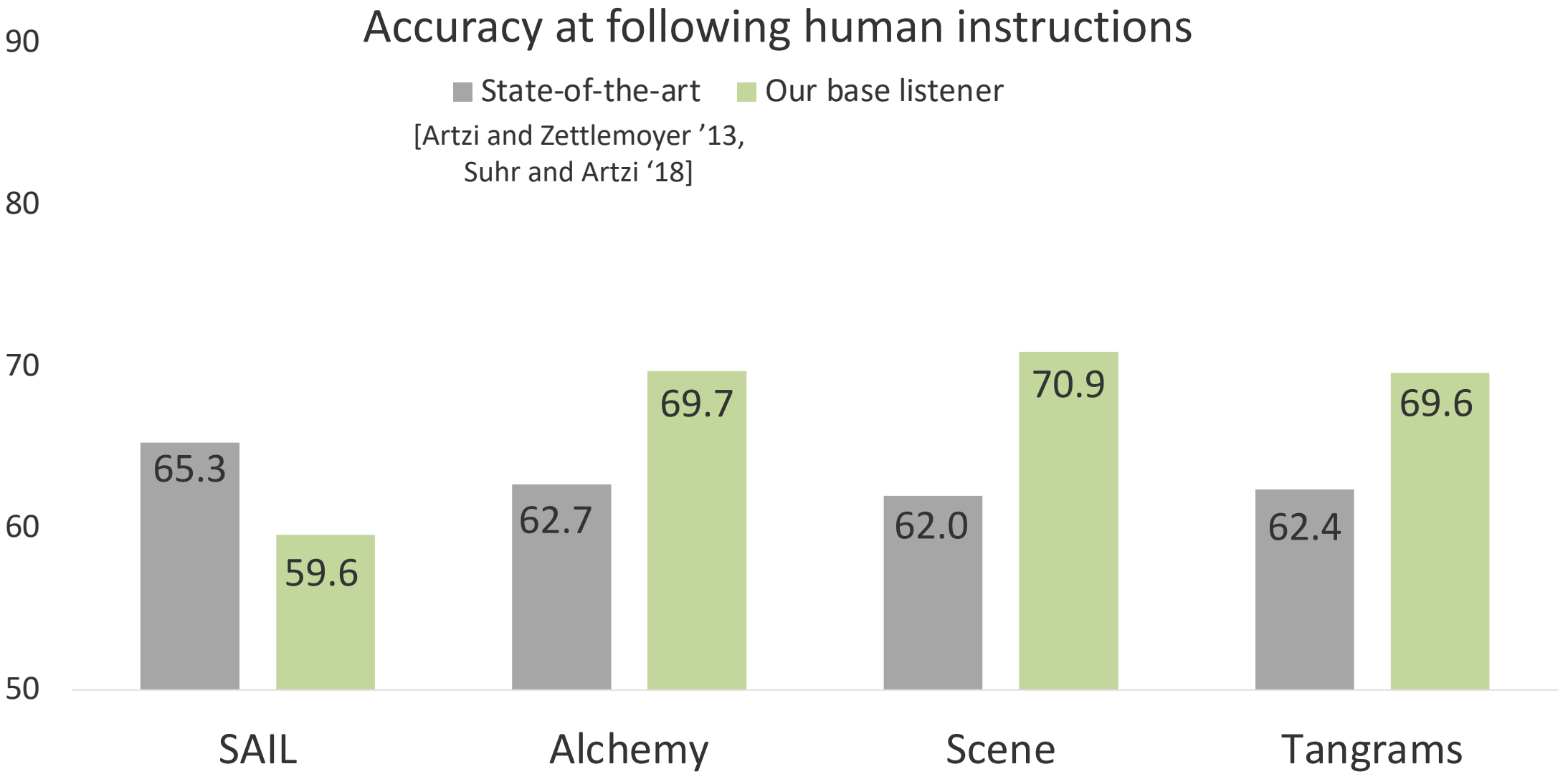


Contextual Execution: Alchemy

1. remove all the purple chemical from the beaker on the far left
2. do the same with one unit of green chemical → Listener →



Strong Listener Models



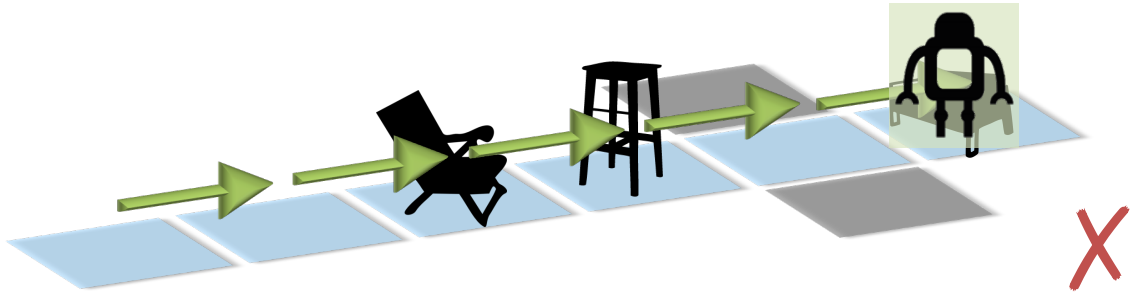
[**Fried**, Andreas, and Klein. NAACL 2018]

A Failure Mode for Listeners: Ambiguity

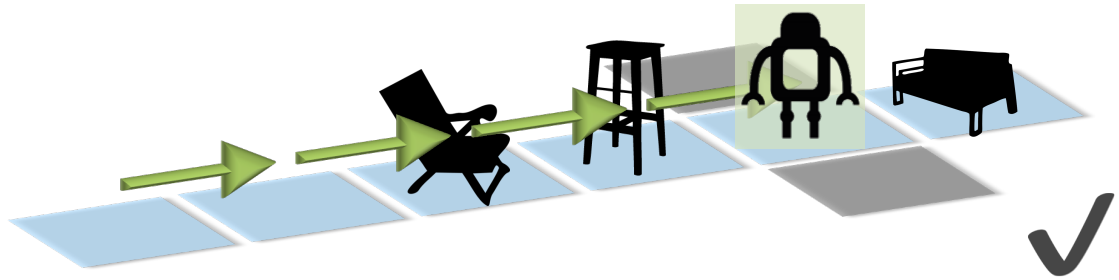
Instruction

walk along the blue carpet and you pass two objects

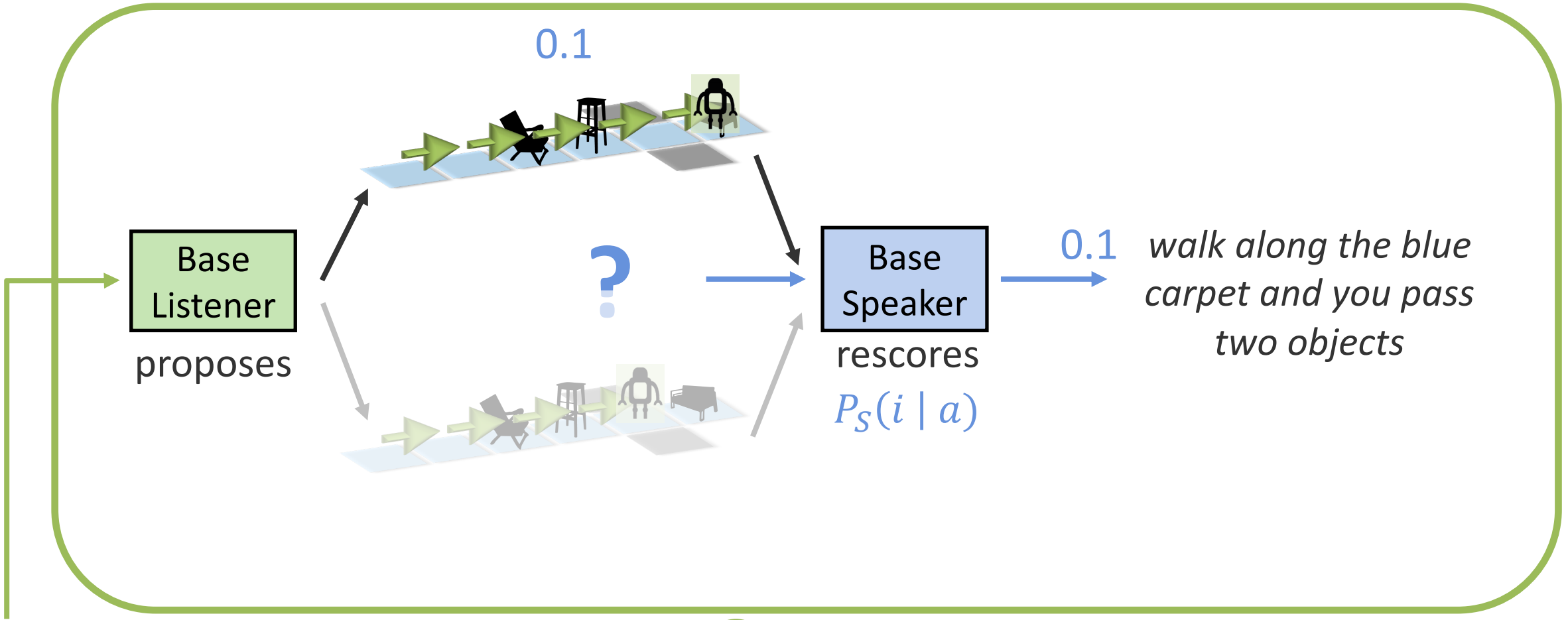
Base Listener



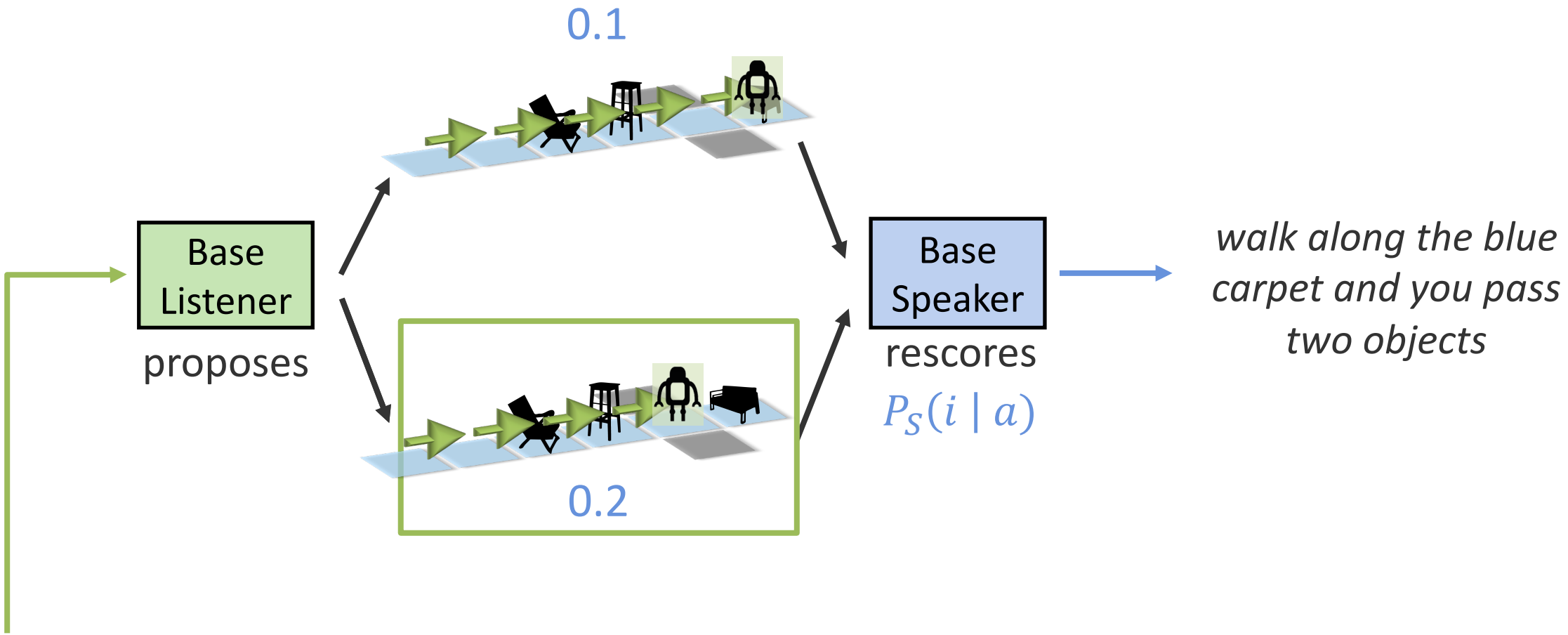
Correct



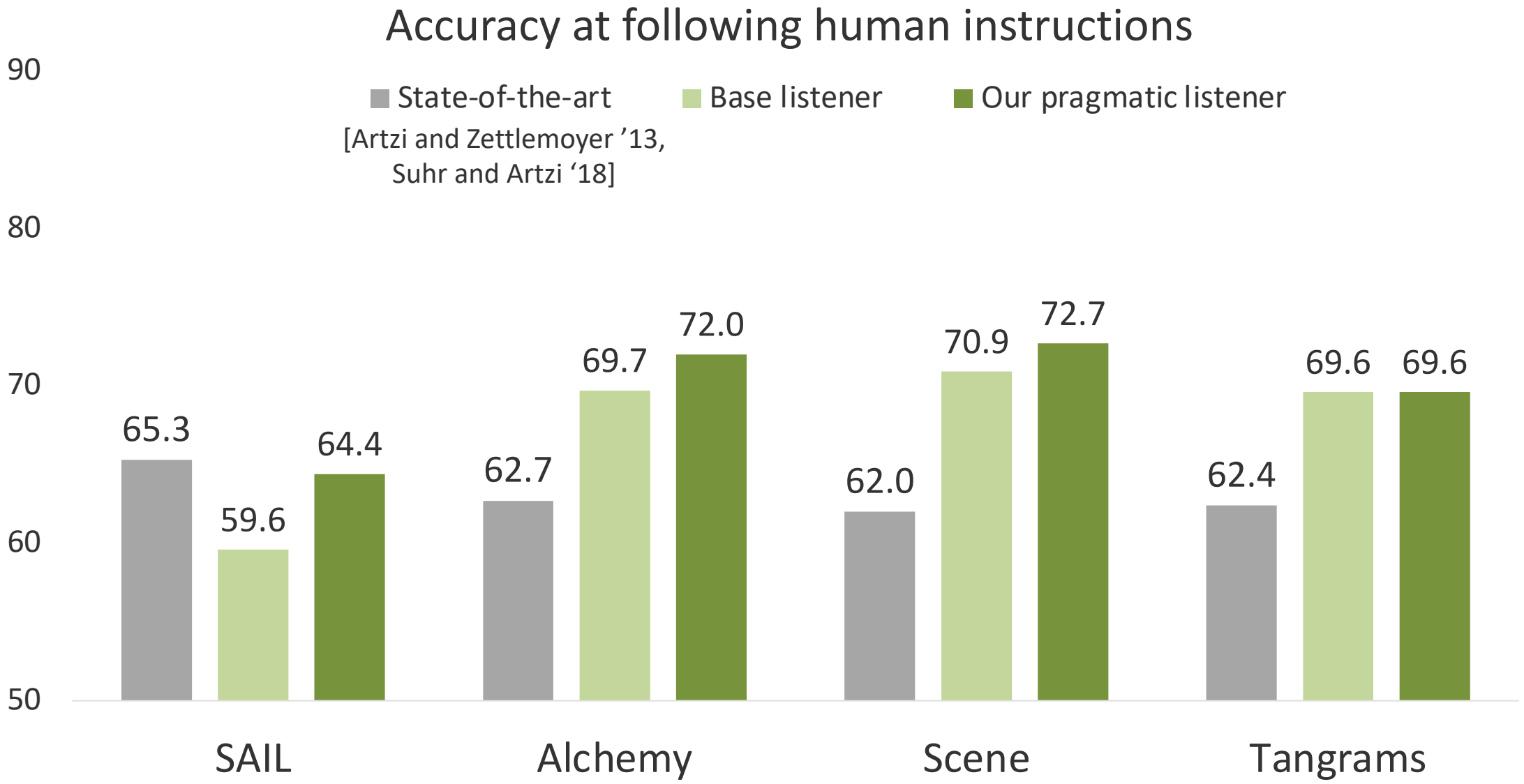
Building a Pragmatic Listener



Building a Pragmatic Listener



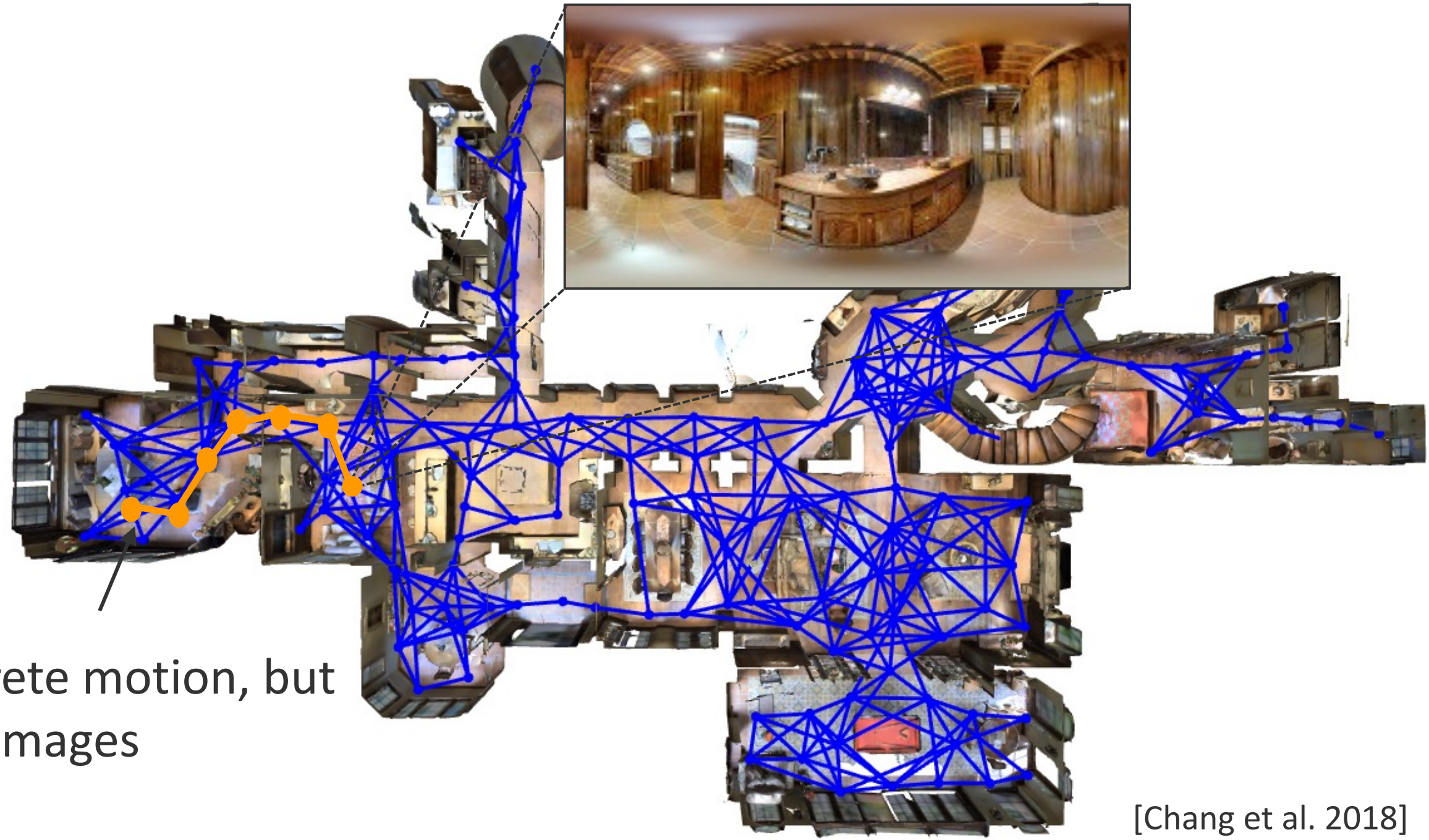
Listener Results



Visually-Grounded Listeners

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.

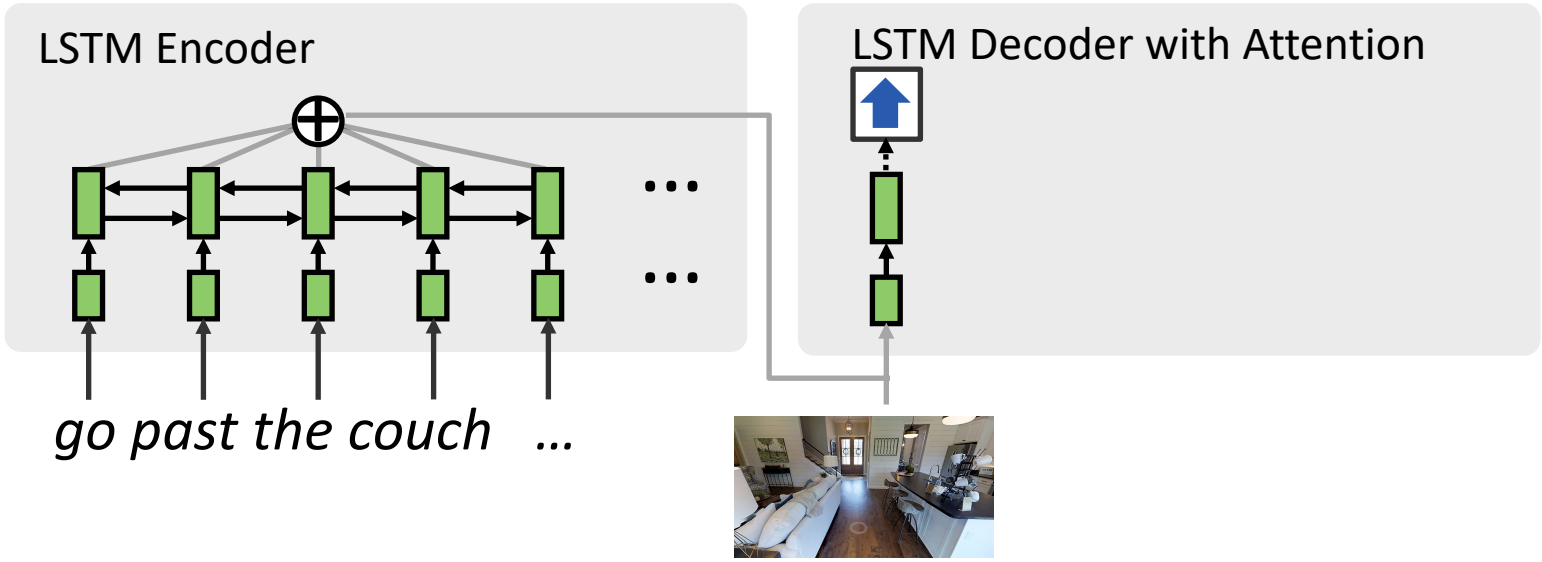
[*Vision-and-Language Navigation Task*. Anderson et al., 2018]



Discrete motion, but real images

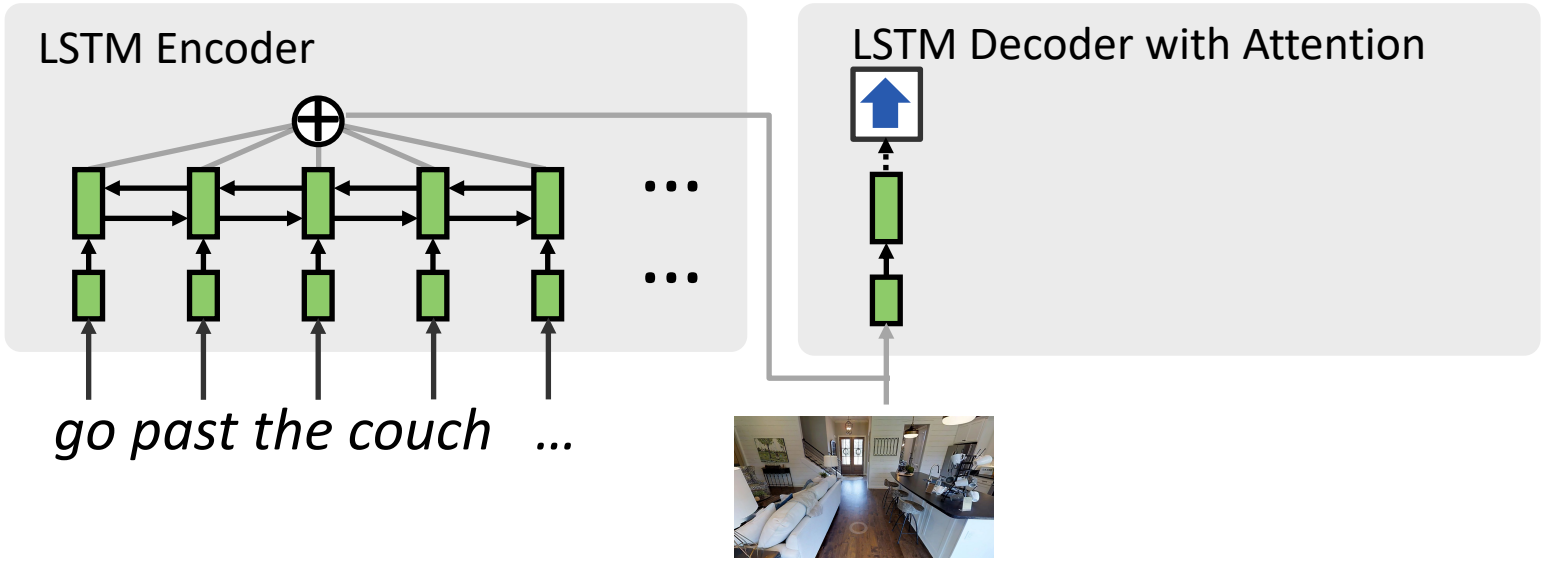
[Chang et al. 2018]

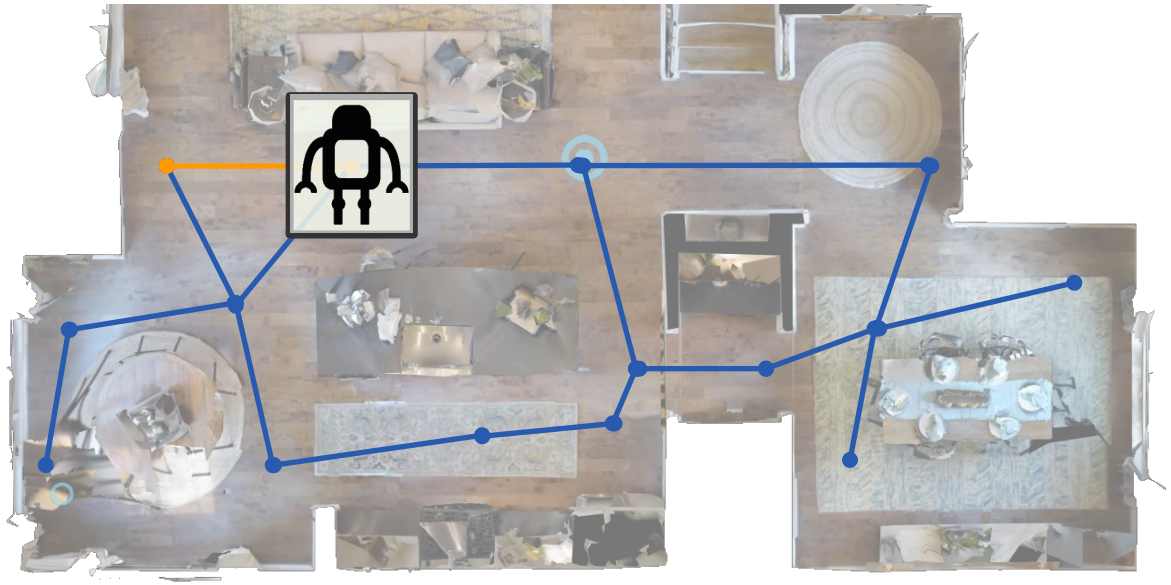
Base Listener Model



[Anderson et al., 2018]

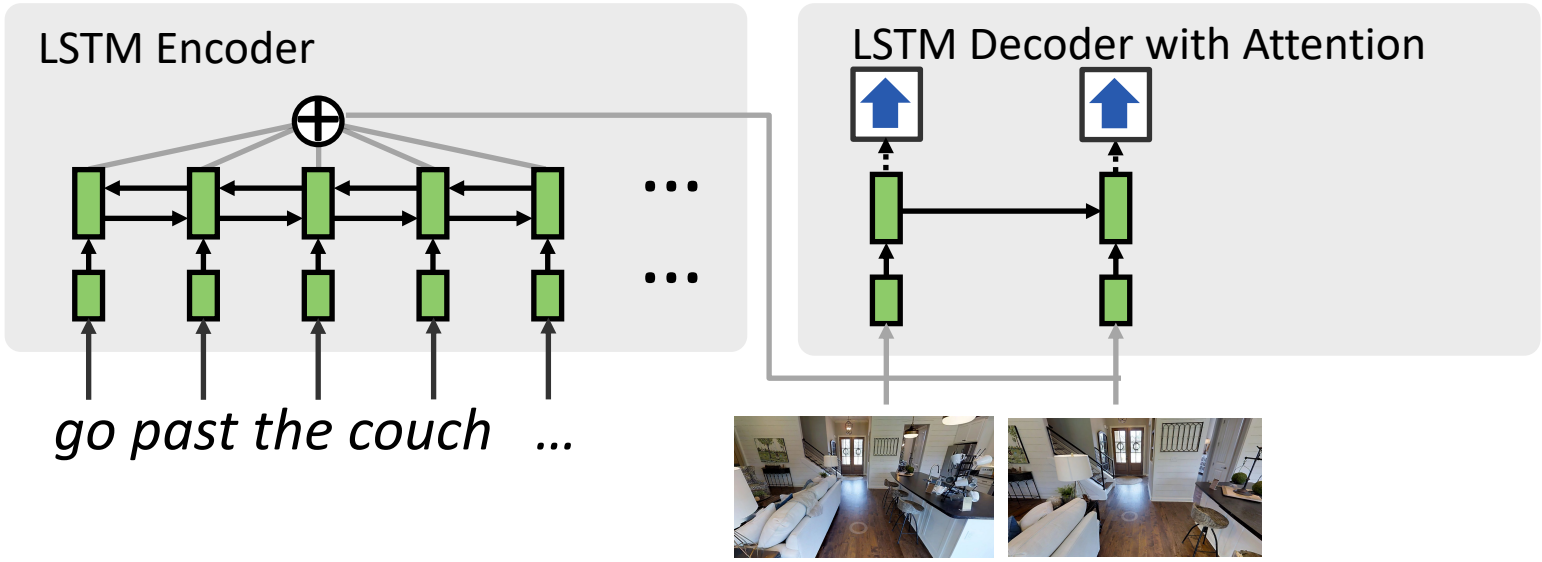
Base Listener Model

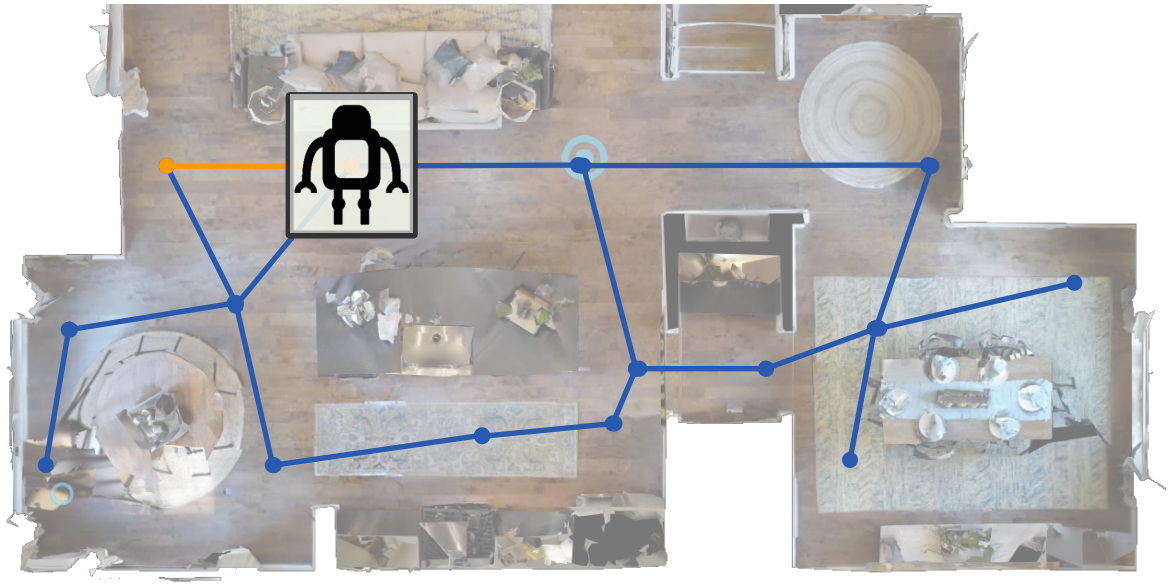




[Anderson et al., 2018]

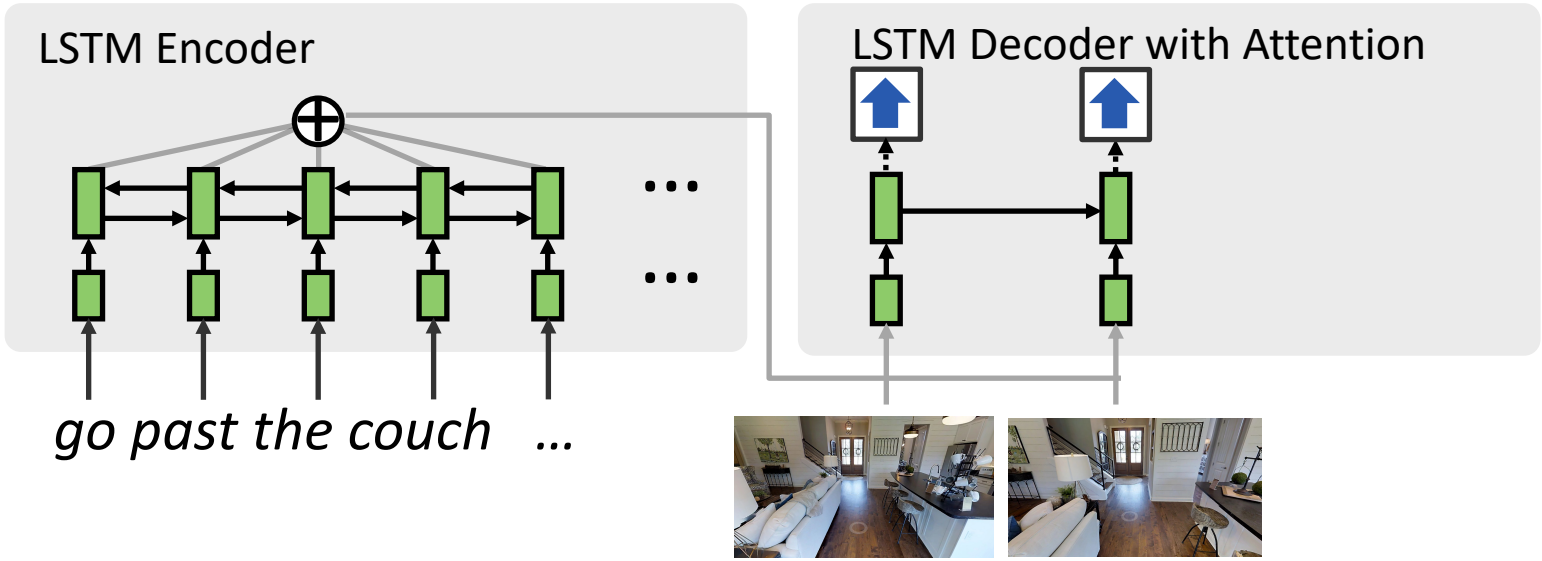
Base Listener Model





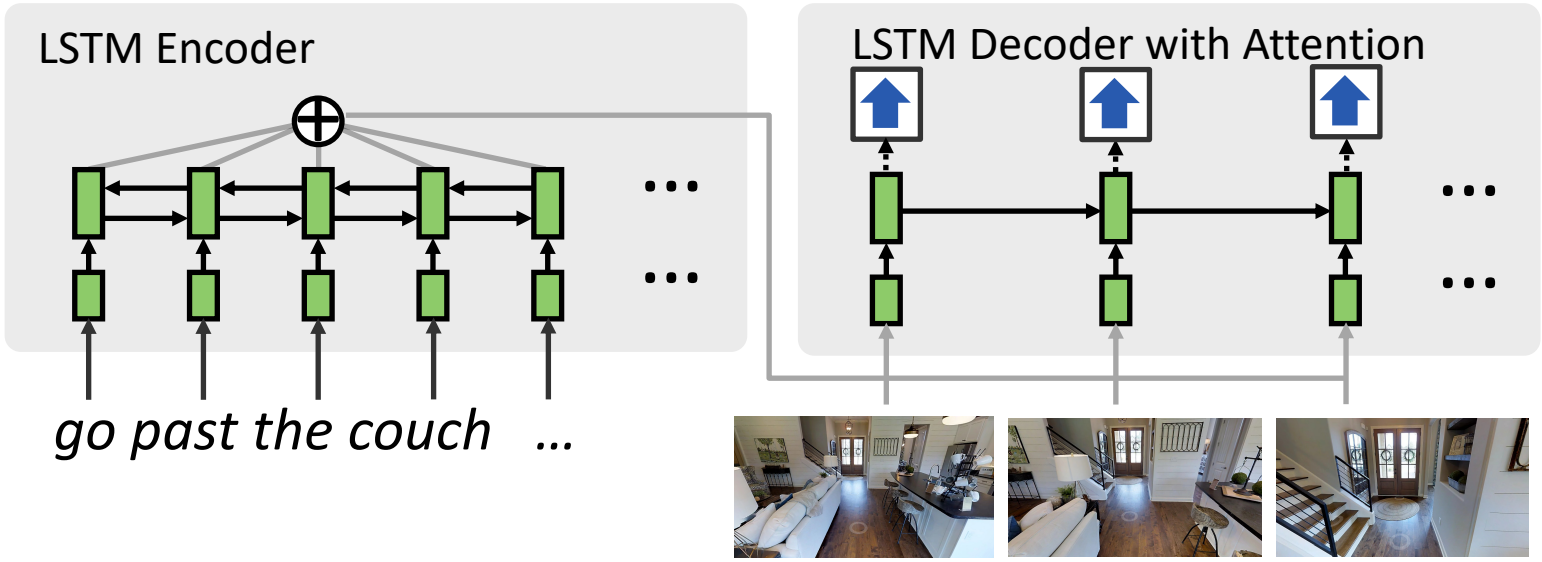
[Anderson et al., 2018]

Base Listener Model



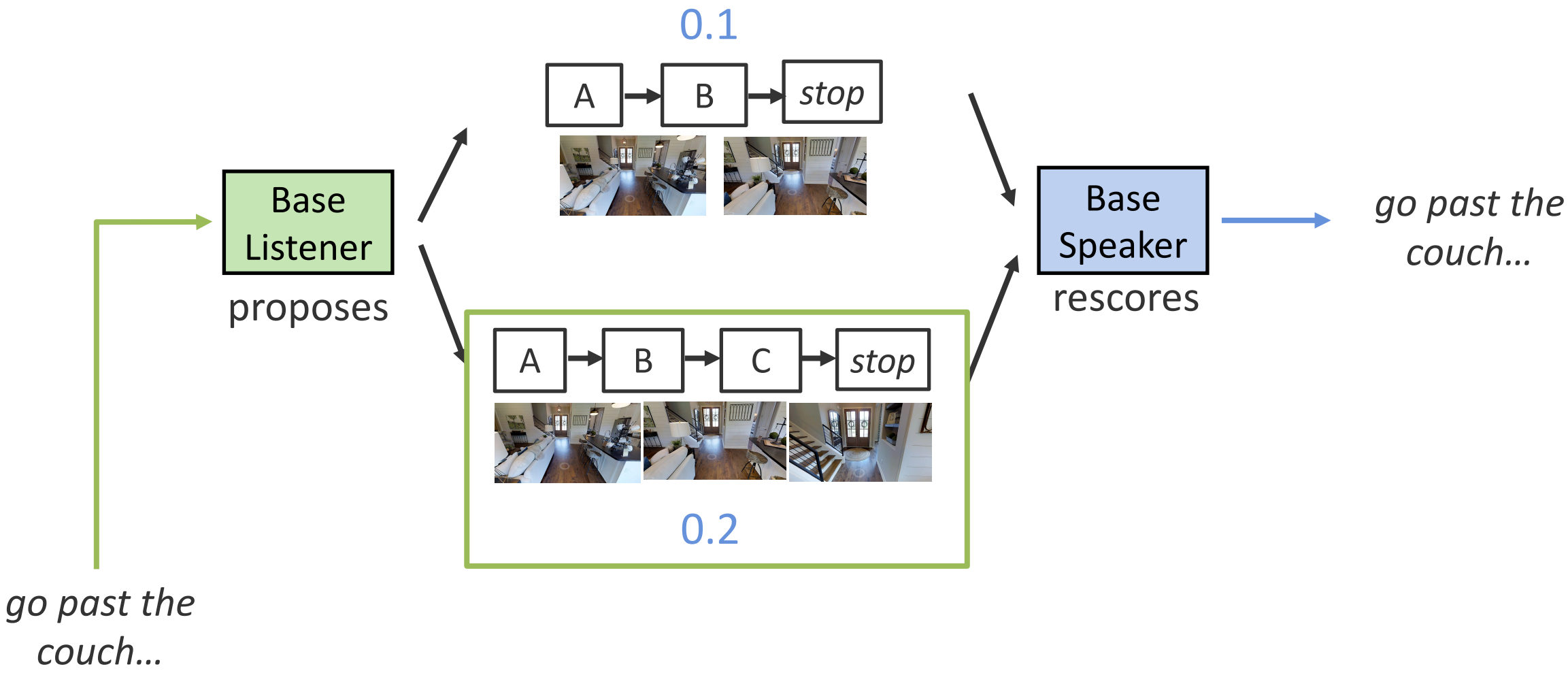
[Anderson et al., 2018]

Base Listener Model

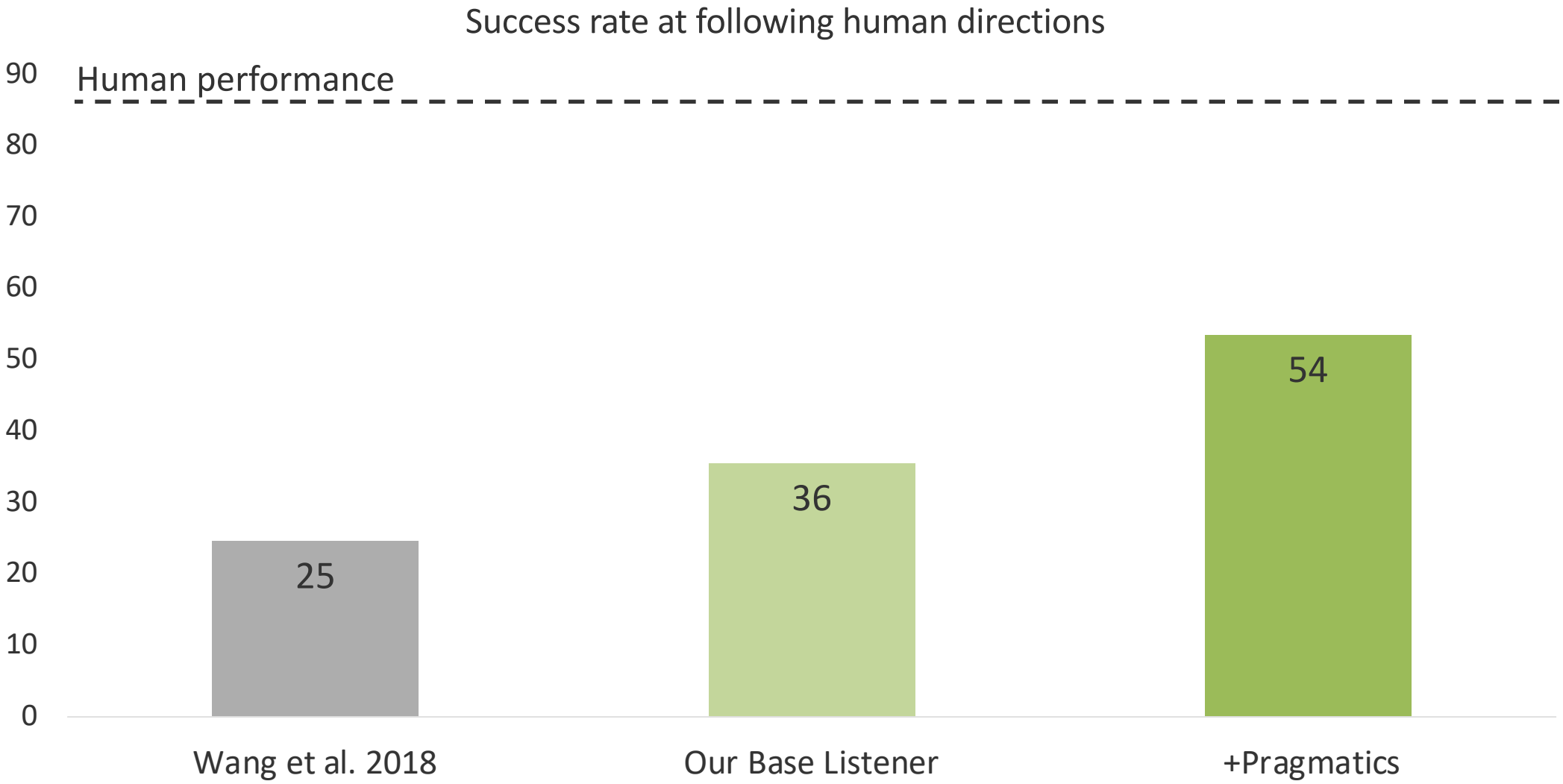


[Anderson et al., 2018]

Pragmatics for Visual Navigation



Comparison to Prior Work



Walk past hall table. Walk into bedroom. Make left at table clock. Wait at bathroom door threshold.

Base listener

Walk past hall table. Walk into bedroom. Make left at table clock. Wait at bathroom door threshold.

Pragmatic listener

Our Work in Context

Computational Pragmatics
Golland et al. ‘10; Frank and Goodman ‘12; Degen ‘13; Vogel et al. ‘13; Tellex et al. ’14; Monroe et al. ‘17; Luo & Shakhnarovich ‘17 …

Instruction Following
MacMahon et al. ‘06; Vogel and Jurafsky ‘10; Tellex et al. ‘11; Chen and Mooney ‘11; Artzi et al. ‘14; Mei et al. ‘16 …

Pragmatic Instruction Following
Fried et al. 2018
Fried*, Hu*, Cirik* et al. 2018

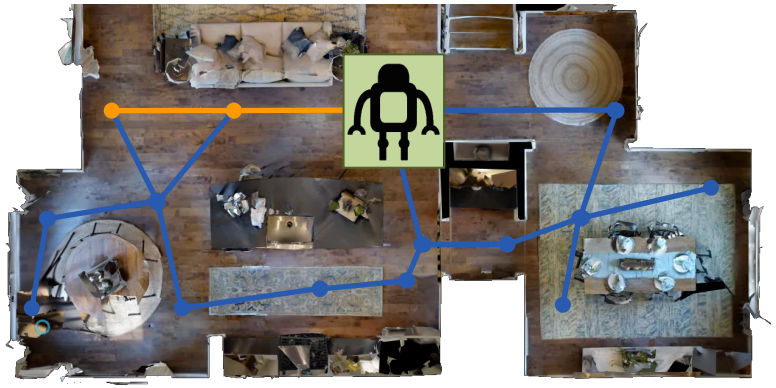
Speaker in Training
Tan et al. 2019
Wang et al. 2019
Zhu et al. 2020

Speaker in Inference
Hu et al. 2019
Cideron et al. 2020
Roman et al. 2020

Improved Search
Ke et al. 2019
Kurita and Cho 2021

Takeaways

Simulating why a speaker said what they did helps resolve ambiguity.

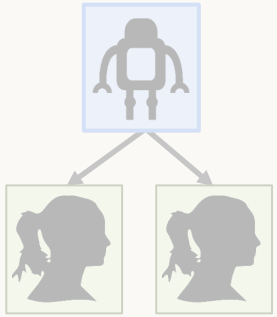


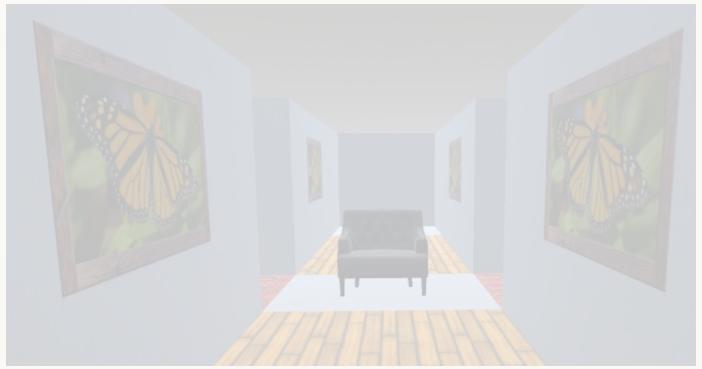
Pragmatics improves most in complex environments where grounding is harder.

Pragmatics and…

Generation

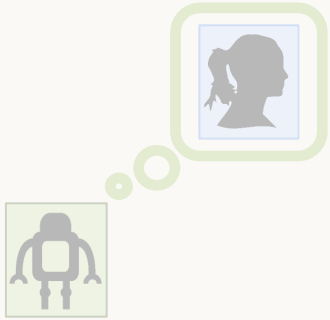
[**Fried**, Andreas, & Klein. NAACL 2018]





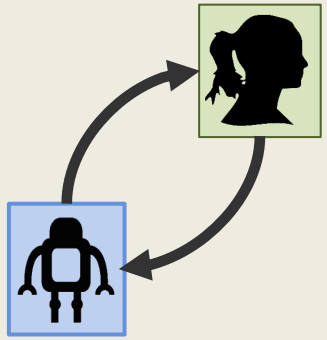
Interpretation

[**Fried***, Hu*, Cirik* et al. NeurIPS 2018]

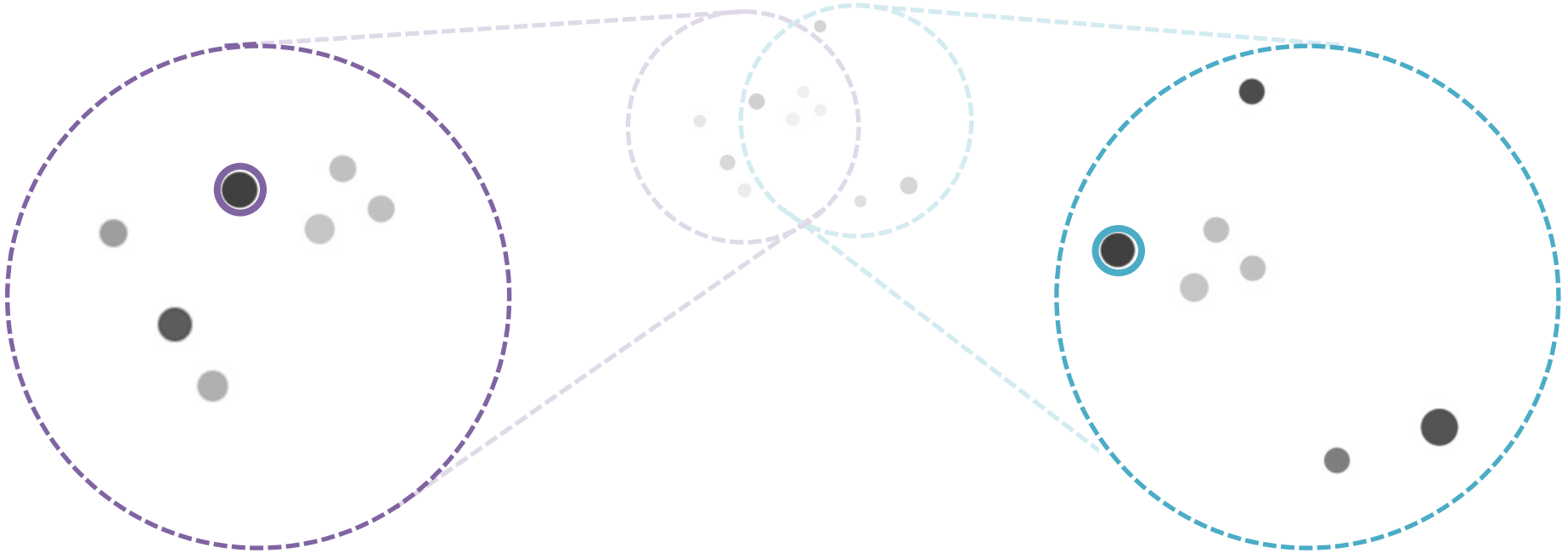


Dialogue

[**Fried**, Chiu, & Klein. In submission]



Grounded Collaborative Dialogue



A: I have three dots in a line with a dark one in the center.

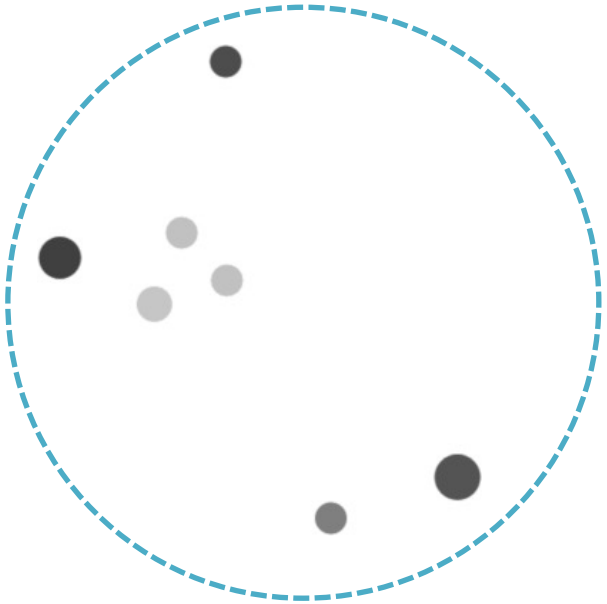
A: Is there a large black dot to the left of the three grey dots?

B: I don't have that. Do you have a cluster of three grey dots in a triangle?

B: Yes, let's select the black one.

[Udagawa and Aizawa, 2019 & 2020]

Decomposing Into Subtasks



A: I have three dots in a line with a dark one in the center. → **B**: I don't have that. Do you have a group of three grey dots? → **A**: Is there a large black dot to the left of the three grey dots? → **B:???**

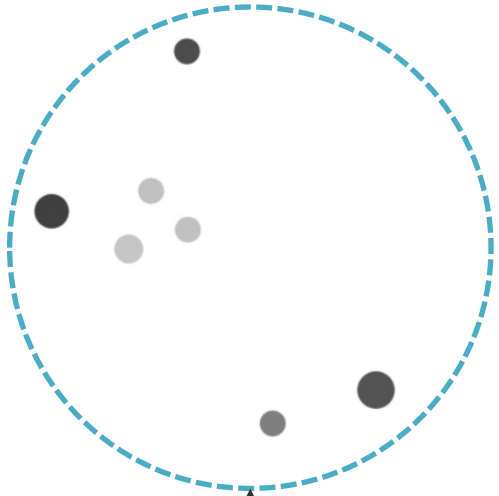
Decomposing Into Subtasks

A: I have three dots
B: I don't have that. Do you have a group of with a black dot to the left of the three grey dots?

A: Is there a large black dot to the left of a group of three grey dots?
B: ???

A: Is there a large black dot to the left of the three grey dots?
B: ???

Decomposing Into Subtasks

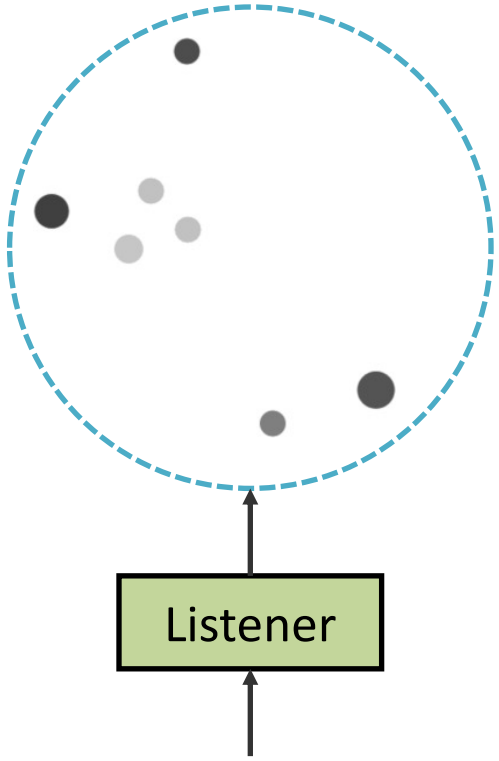


don't have that. Do
u have a group of
three grey dots?

A: Is there a large black dot to the left of the three grey dots?

B:???

Decomposing Into Subtasks

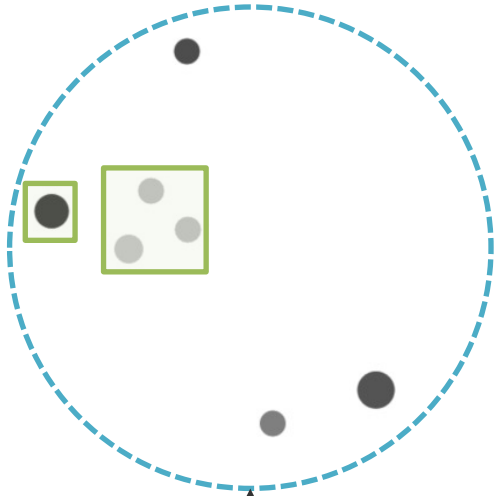


don't have that. Do
u have a group of
three grey dots?

A: Is there a large black dot to the left of the three grey dots?

B:???

Decomposing Into Subtasks

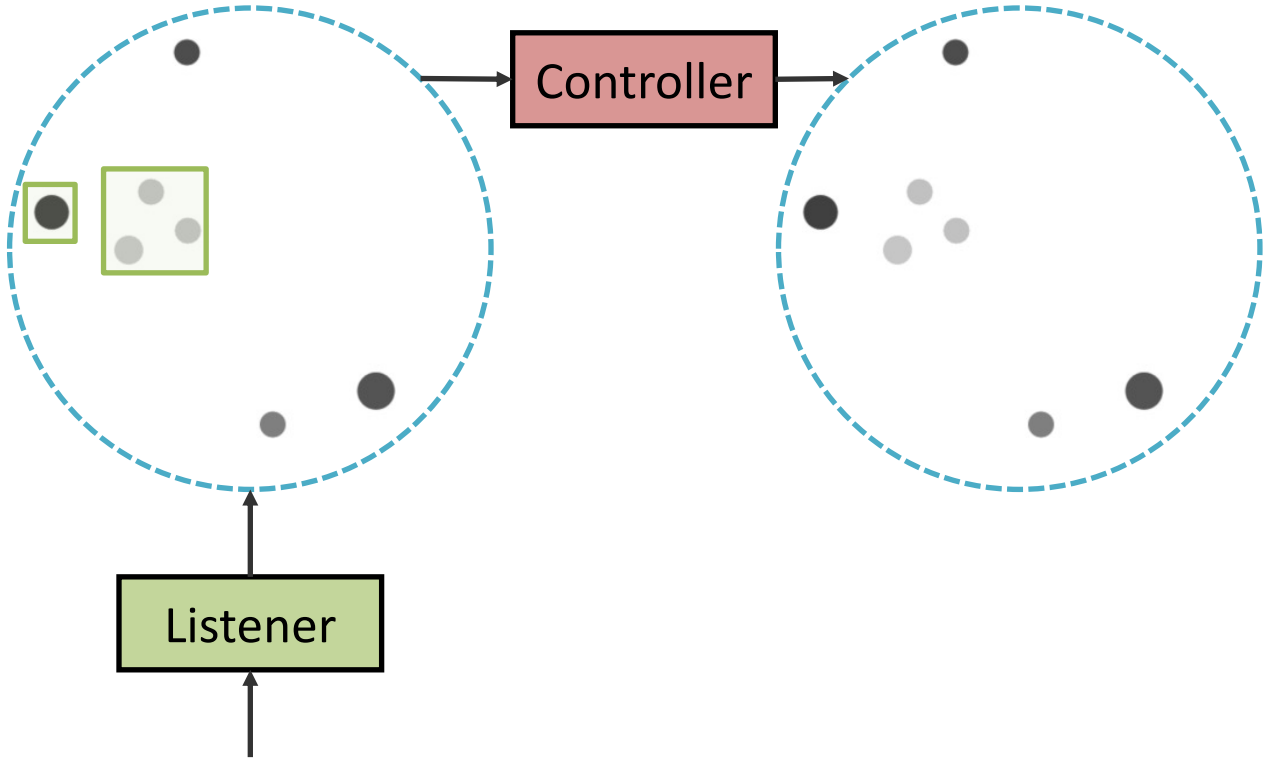


don't have that. Do
u have a group of
three grey dots?

A: Is there a large black dot to the left of the three grey dots?

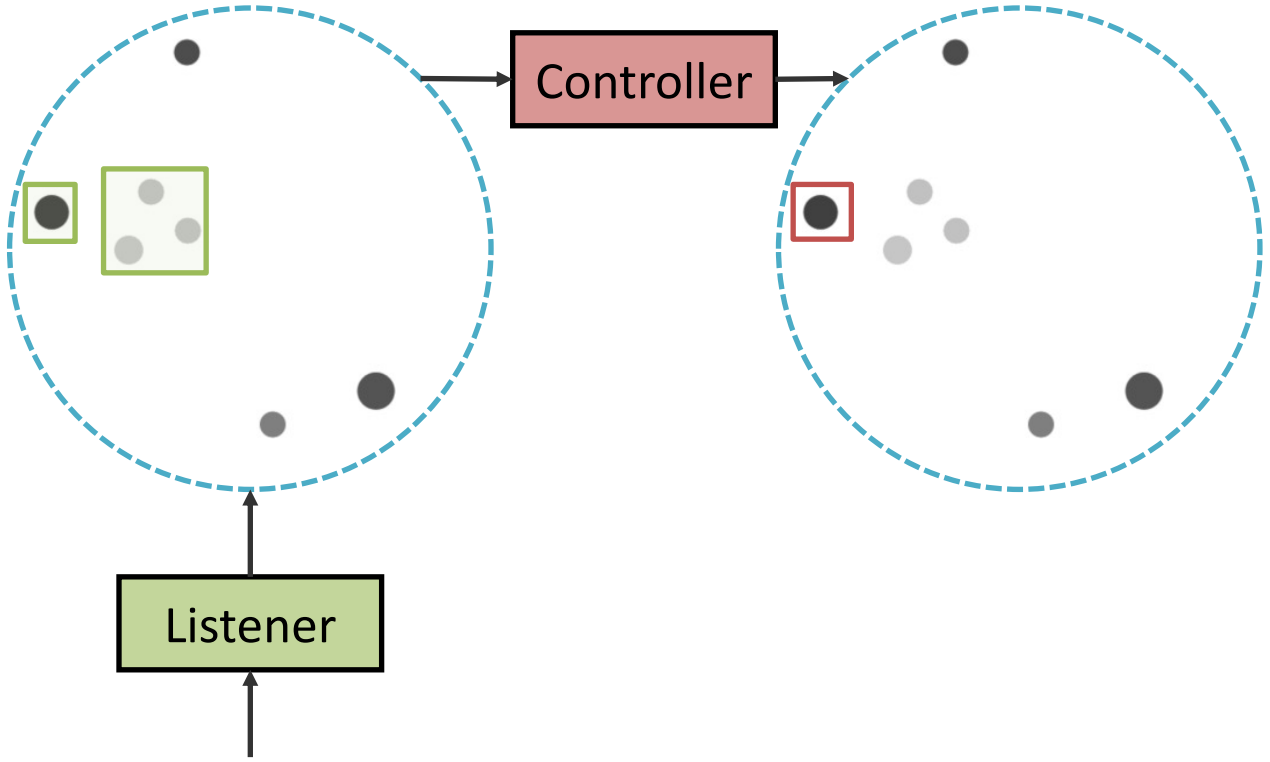
B:???

Decomposing Into Subtasks



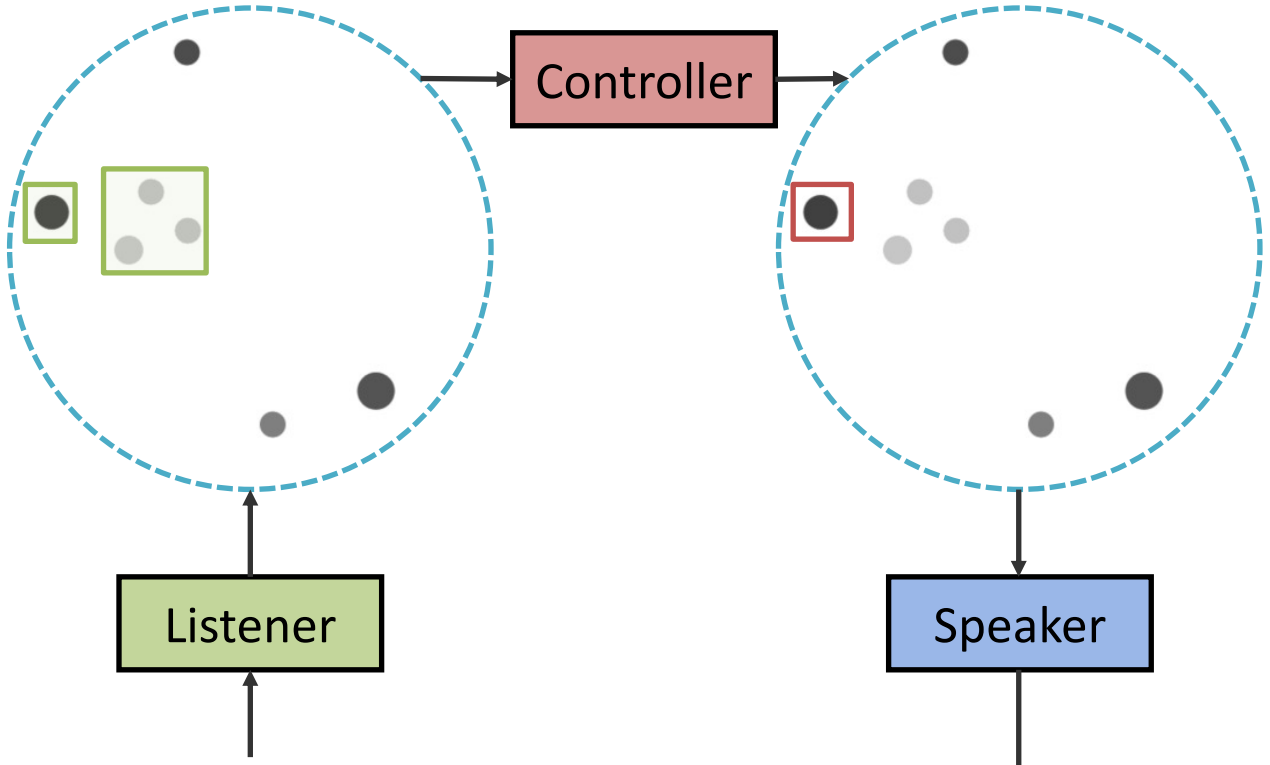
don't have that. Do u have a group of three grey dots? → **A**: Is there a large black dot to the left of the three grey dots? → **B**:???

Decomposing Into Subtasks



don't have that. Do u have a group of three grey dots? → **A**: Is there a large black dot to the left of the three grey dots? → **B**:???

Decomposing Into Subtasks

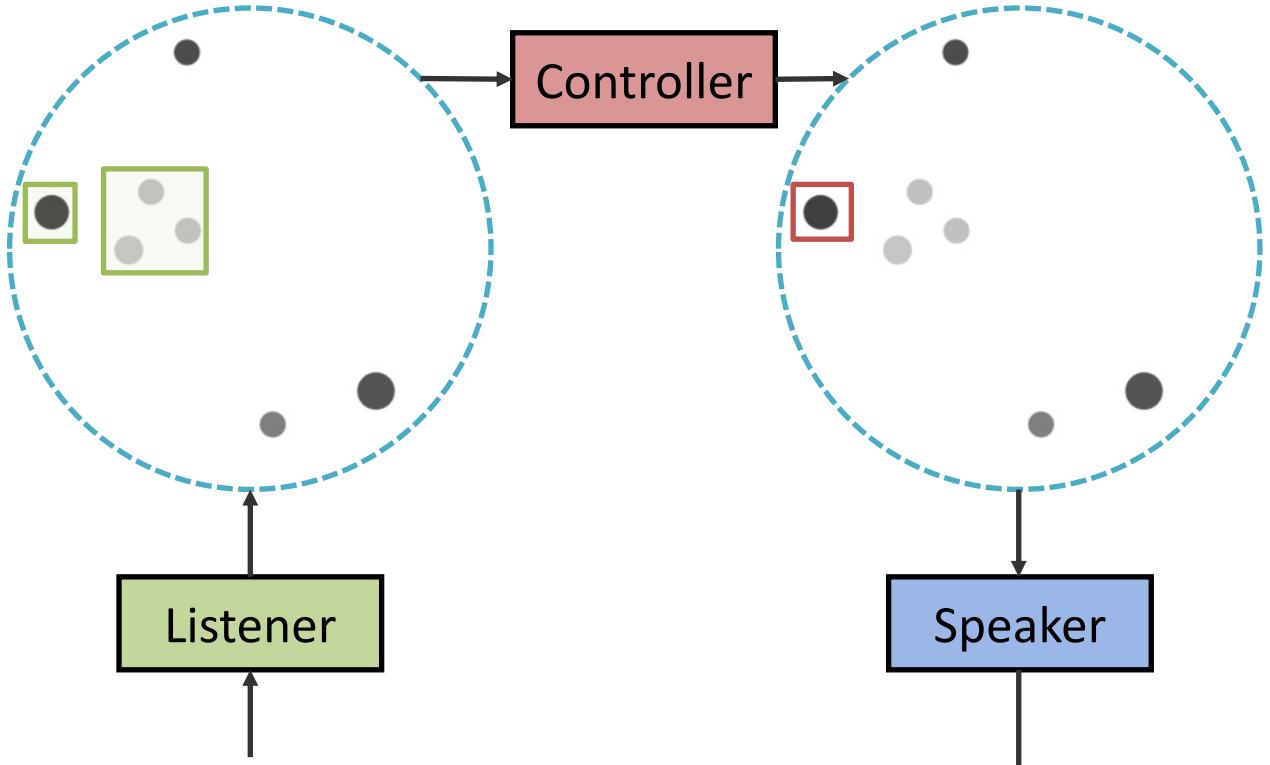


don't have that. Do
u have a group of
three grey dots?

A: Is there a large black dot to the left of the three grey dots?

B:Yes, let's select the black one.

Decomposing Into Subtasks

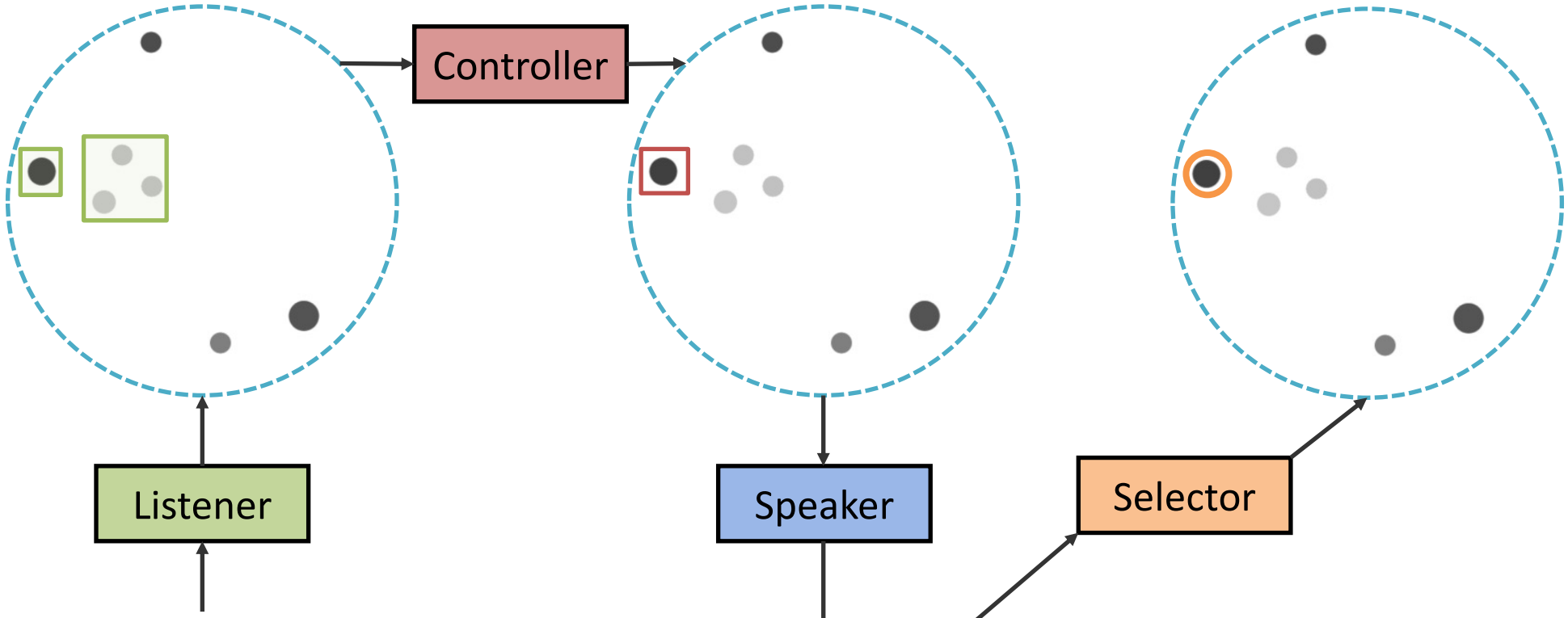


don't have that. Do
u have a group of
three grey dots?

A: Is there a large black dot to the left of the three grey dots?

B:Yes, let's select the black one.

Decomposing Into Subtasks

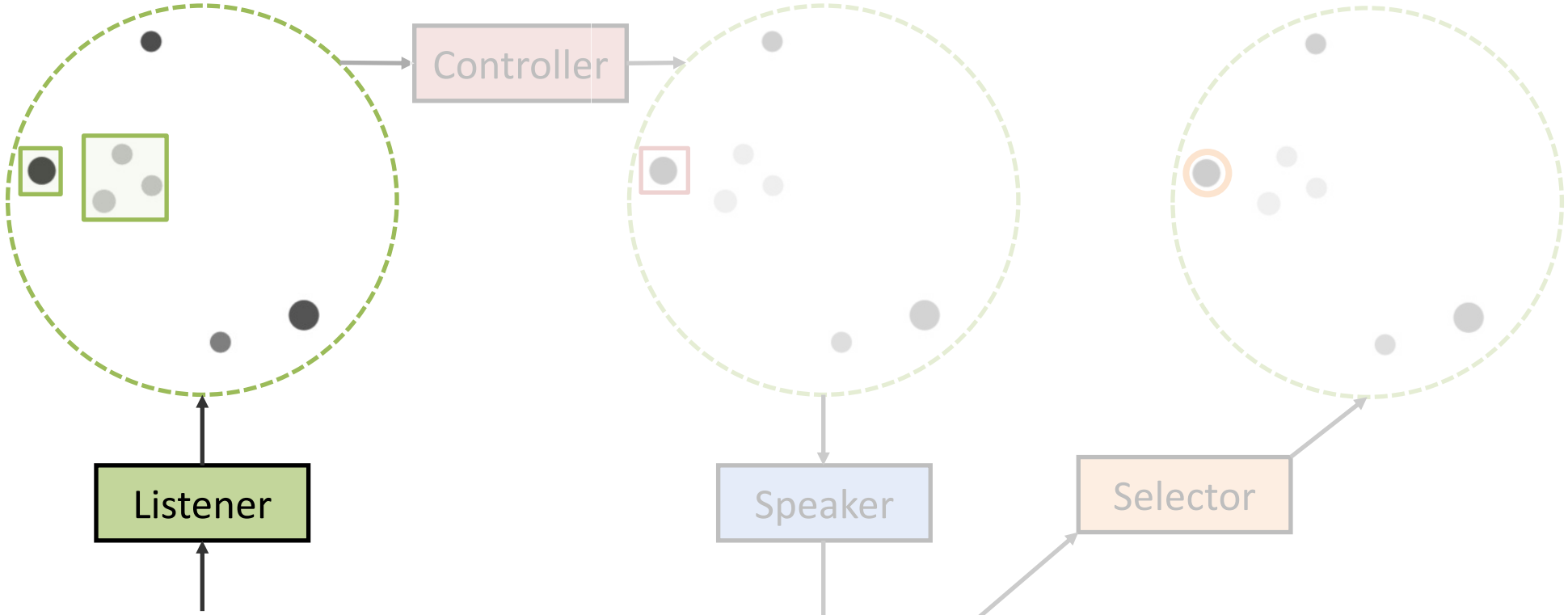


don't have that. Do u have a group of three grey dots?

A: Is there a large black dot to the left of the three grey dots?

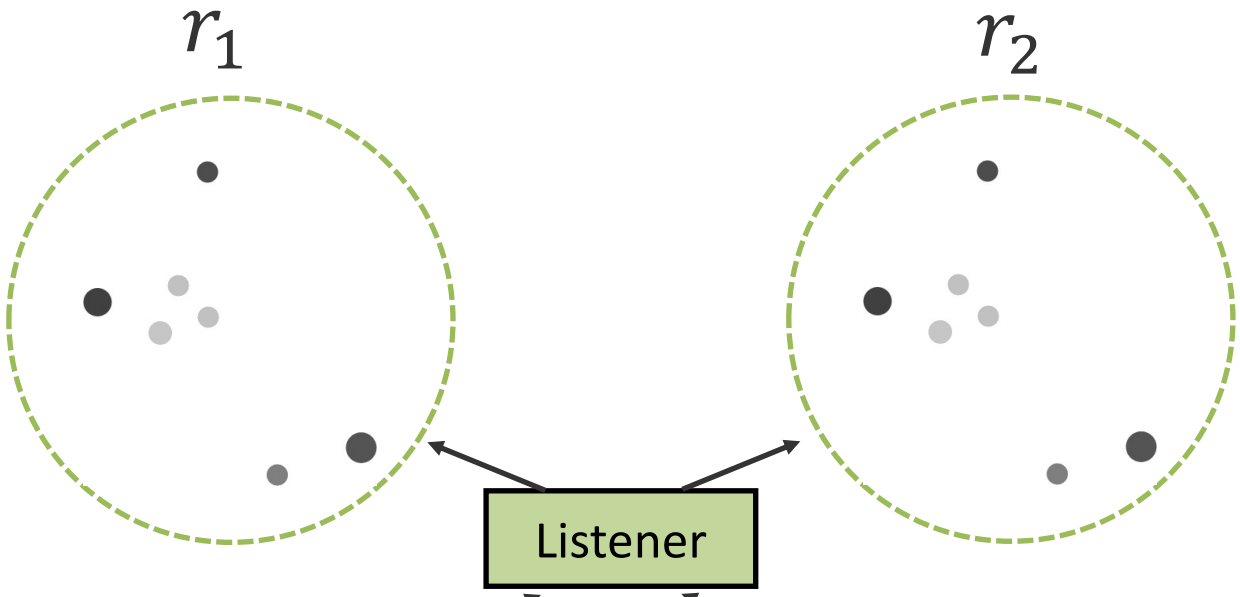
B:Yes, let's select the black one.

Decomposing Into Subtasks



A Structured Listener Module

Referents:



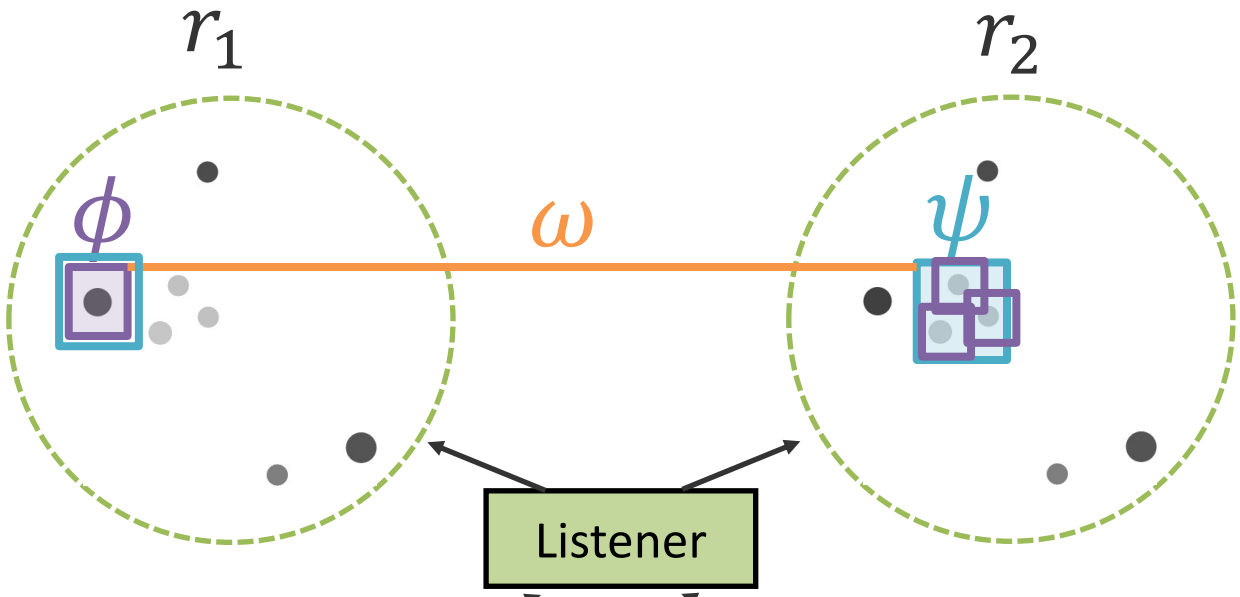
Utterance, $\boldsymbol{u}$ *Is there a large black dot to the left of the three grey dots?*

Neural Conditional Random Field (CRF):

$P_L(r_1, r_2|\boldsymbol{u})$

A Structured Listener Module

Referents:



Utterance, $\boldsymbol{u}$ *Is there a large black dot to the left of the three grey dots?*

Neural Conditional Random Field (CRF):

$$P_L(r_1, r_2|\boldsymbol{u}) \propto \exp \sum_{k\in\{1,2\}} \left(\sum_{d\in r_k} \phi(d, \boldsymbol{u}) \right) + \psi(r_k, \boldsymbol{u}) + \omega(r_{k:k+1}, \boldsymbol{u})$$

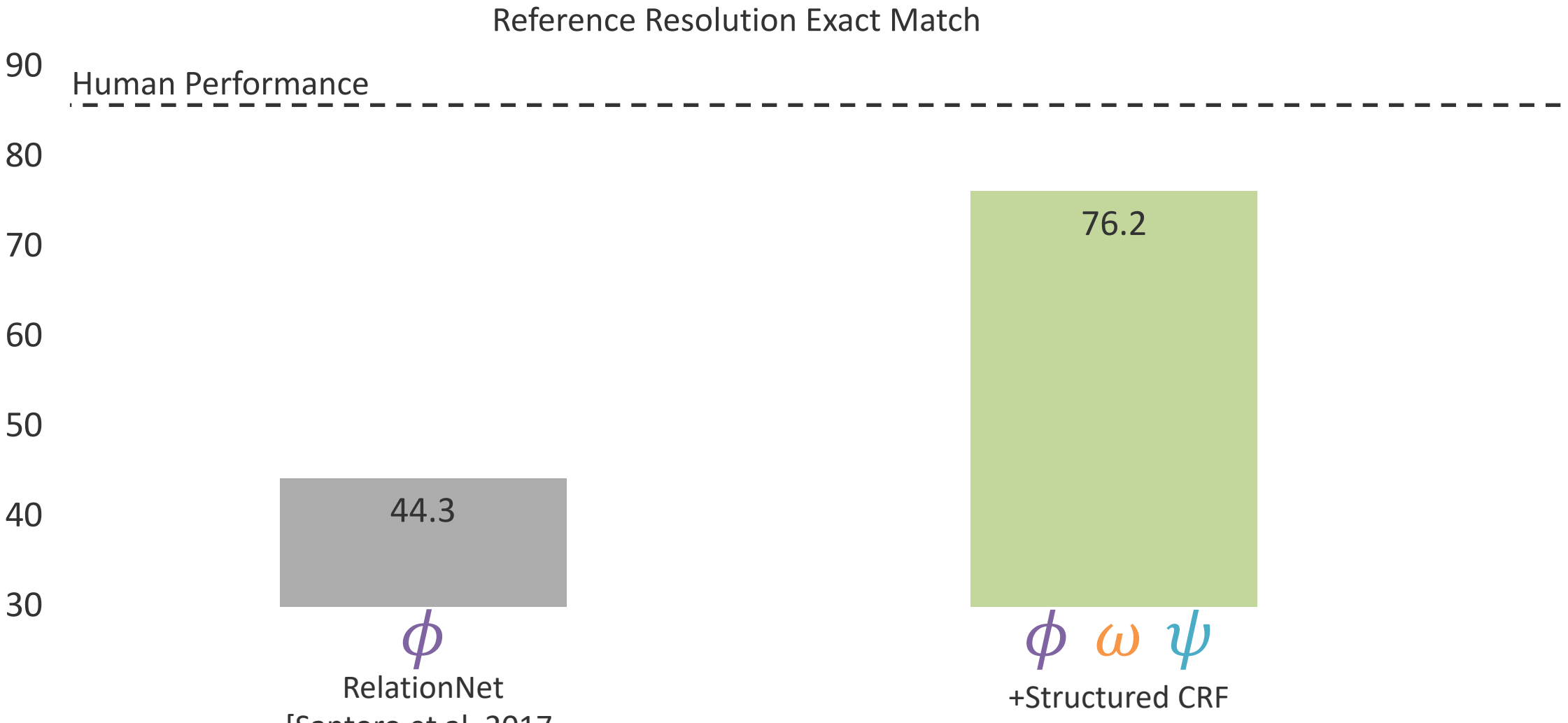
Compute with dynamic programming

a large black dot — Single Dots

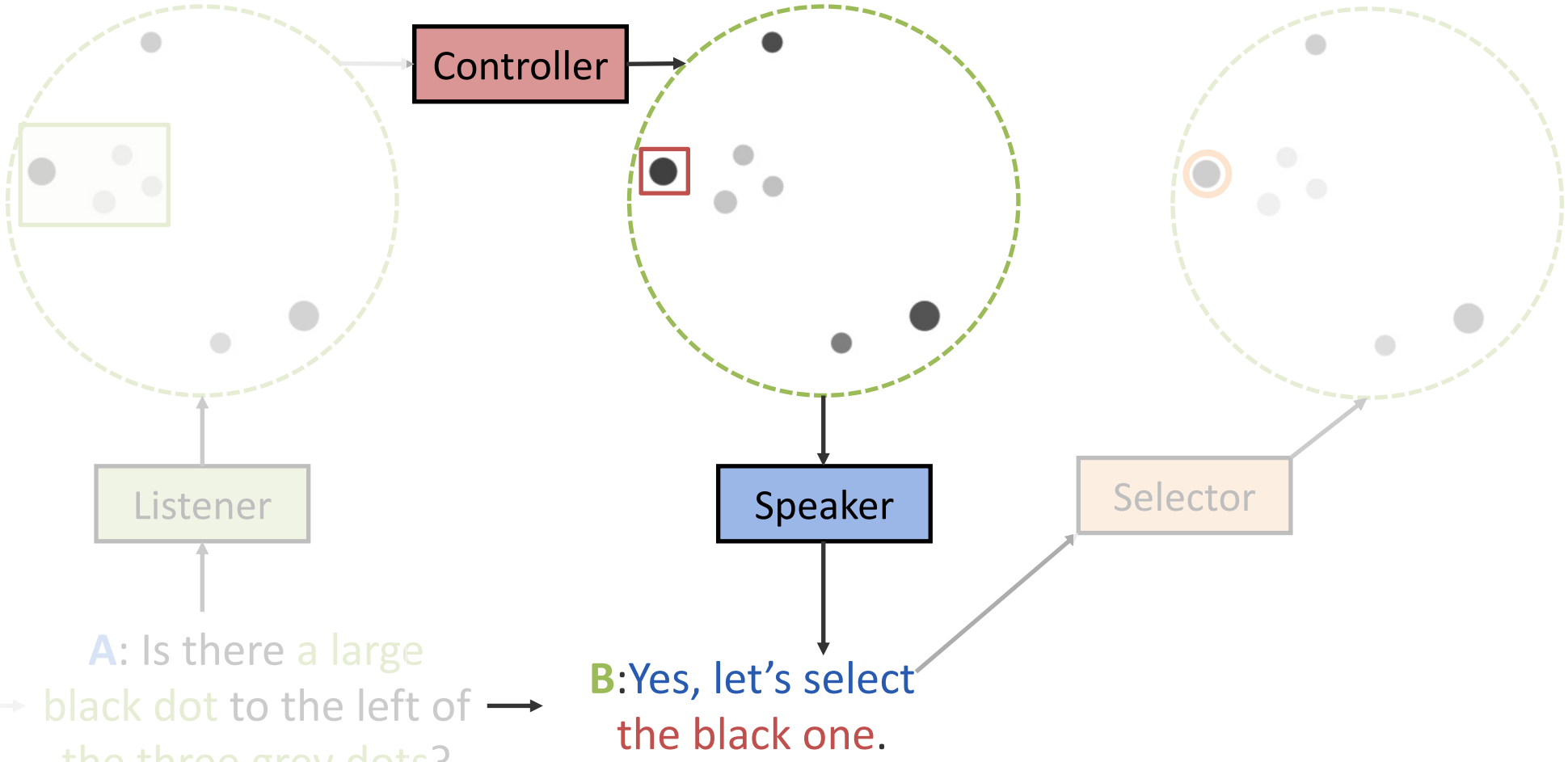
the three grey dots — Groups

… to the left of … — Relations

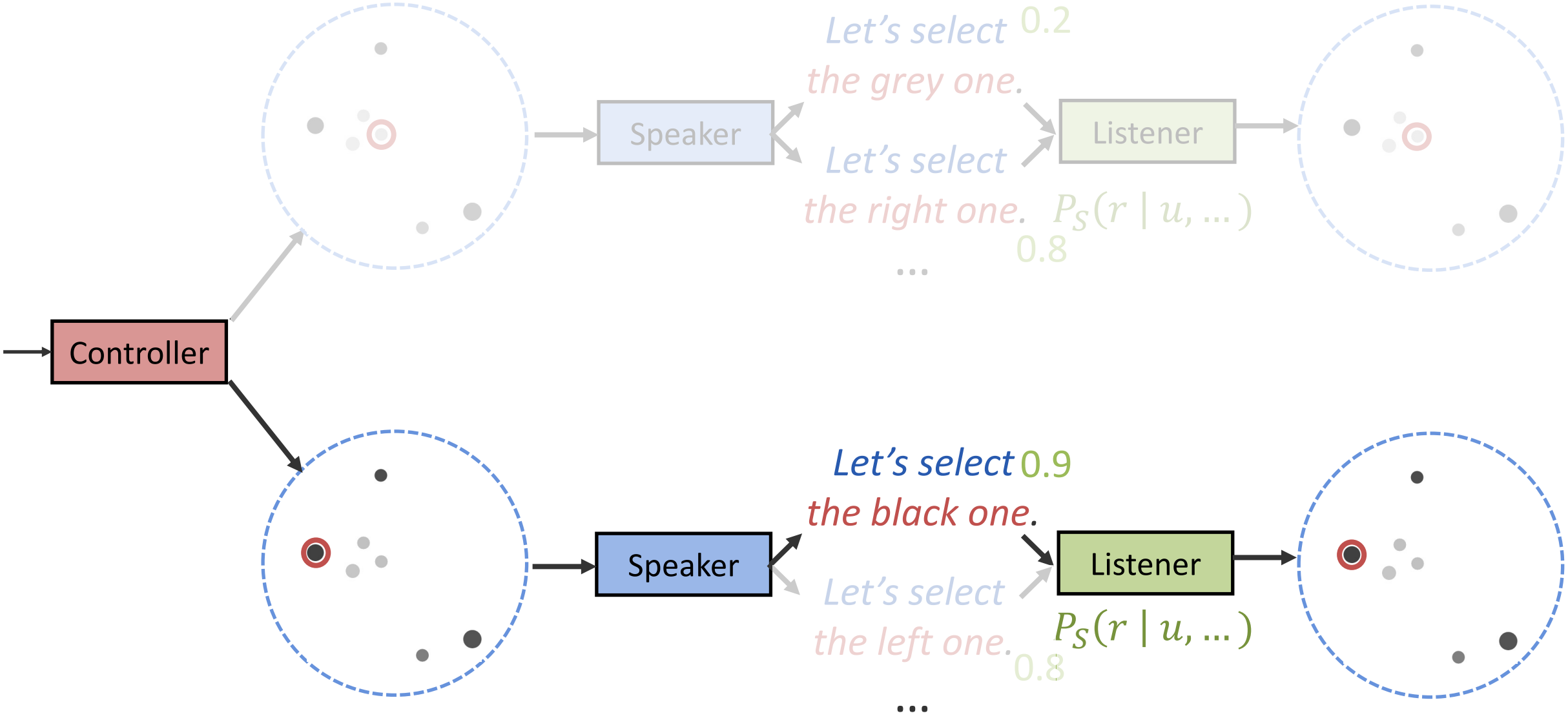
Listener Results



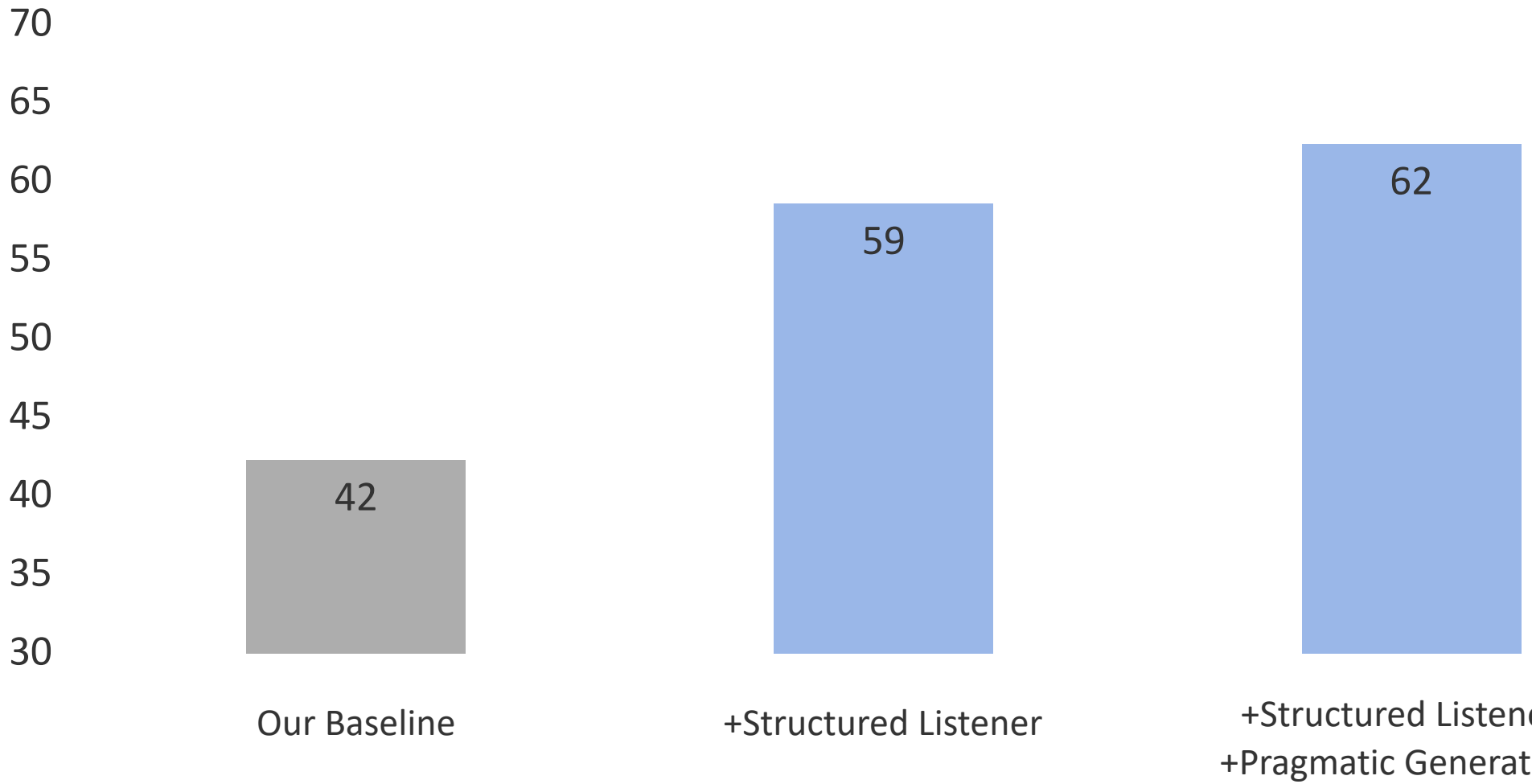
Decomposing Into Subtasks



Pragmatic Generation

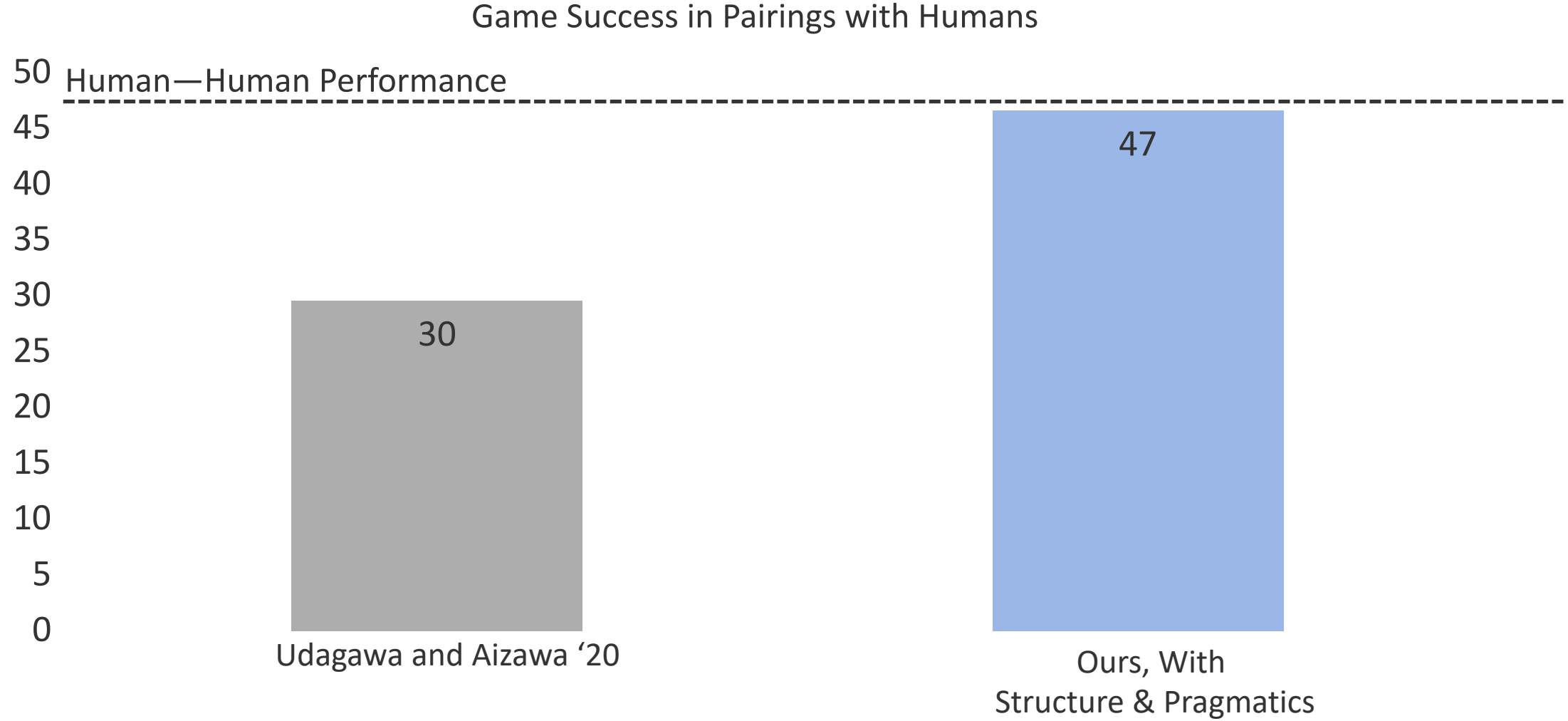


Automatic Evaluation Results



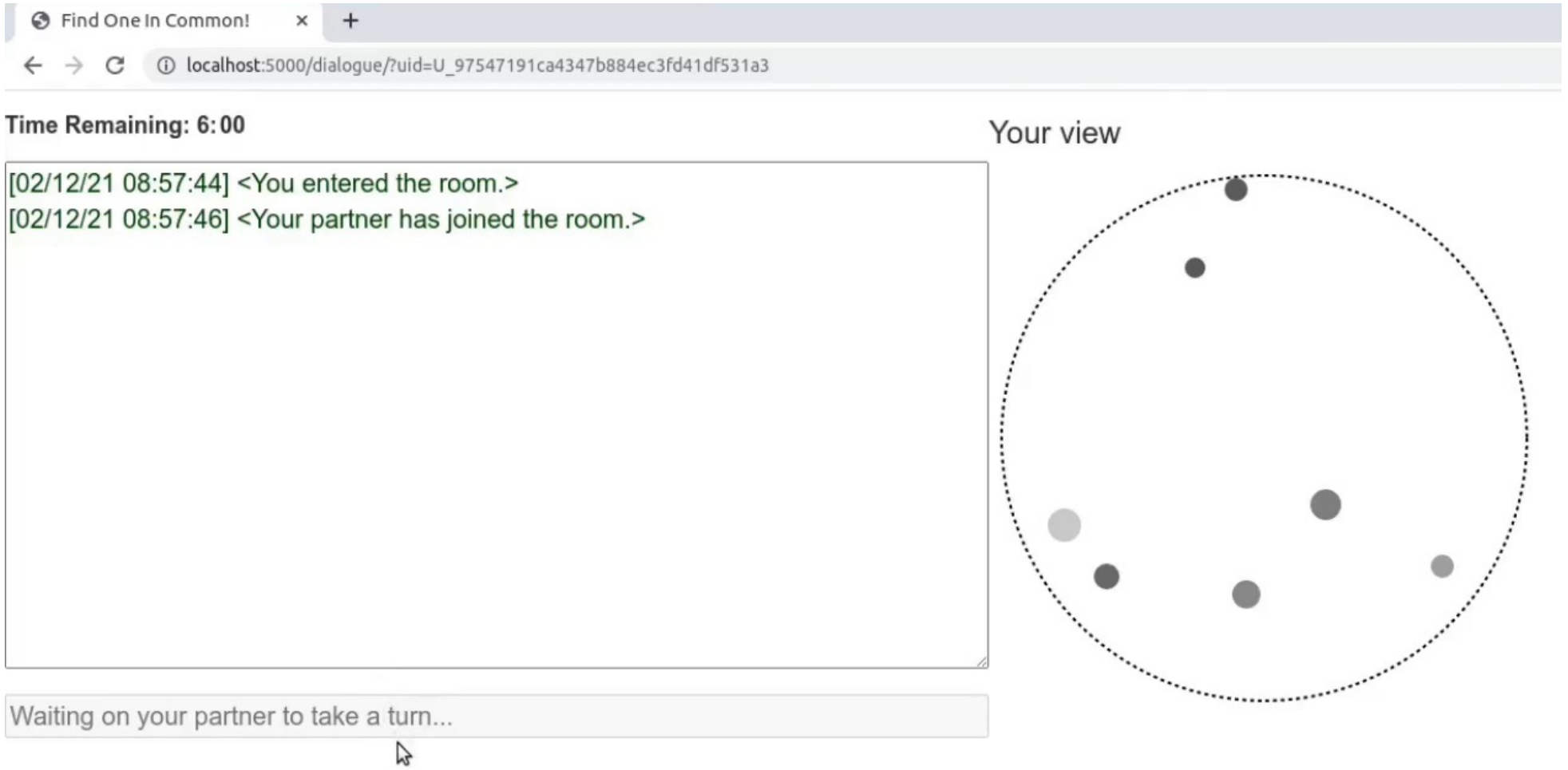
[**Fried**, Chiu, and Klein. In submission]

Human Evaluation Results

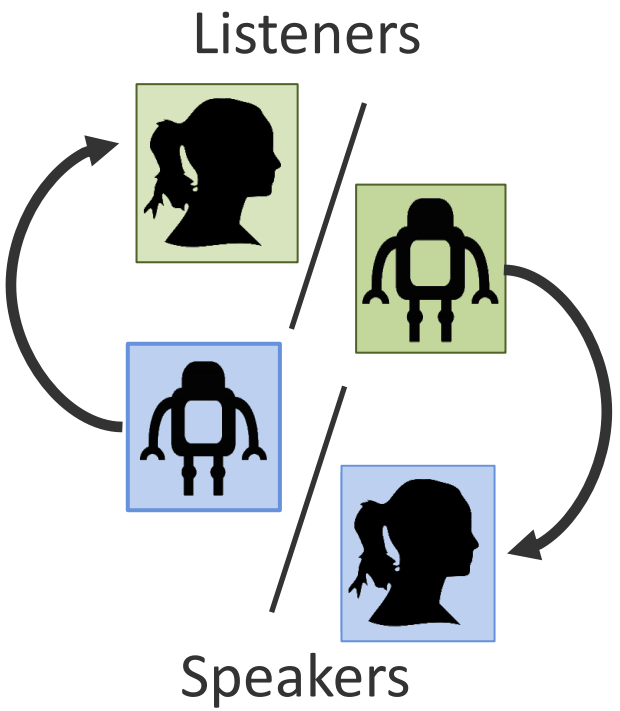


[**Fried**, Chiu, and Klein. In submission]

Demo


Find One In Common!
localhost:5000/dialogue/?uid=U_97547191ca4347b884ec3fd41df531a3
Time Remaining: 6:00
[02/12/21 08:57:44] <You entered the room.>
[02/12/21 08:57:46] <Your partner has joined the room.>
Your view
Waiting on your partner to take a turn...

Final Takeaways

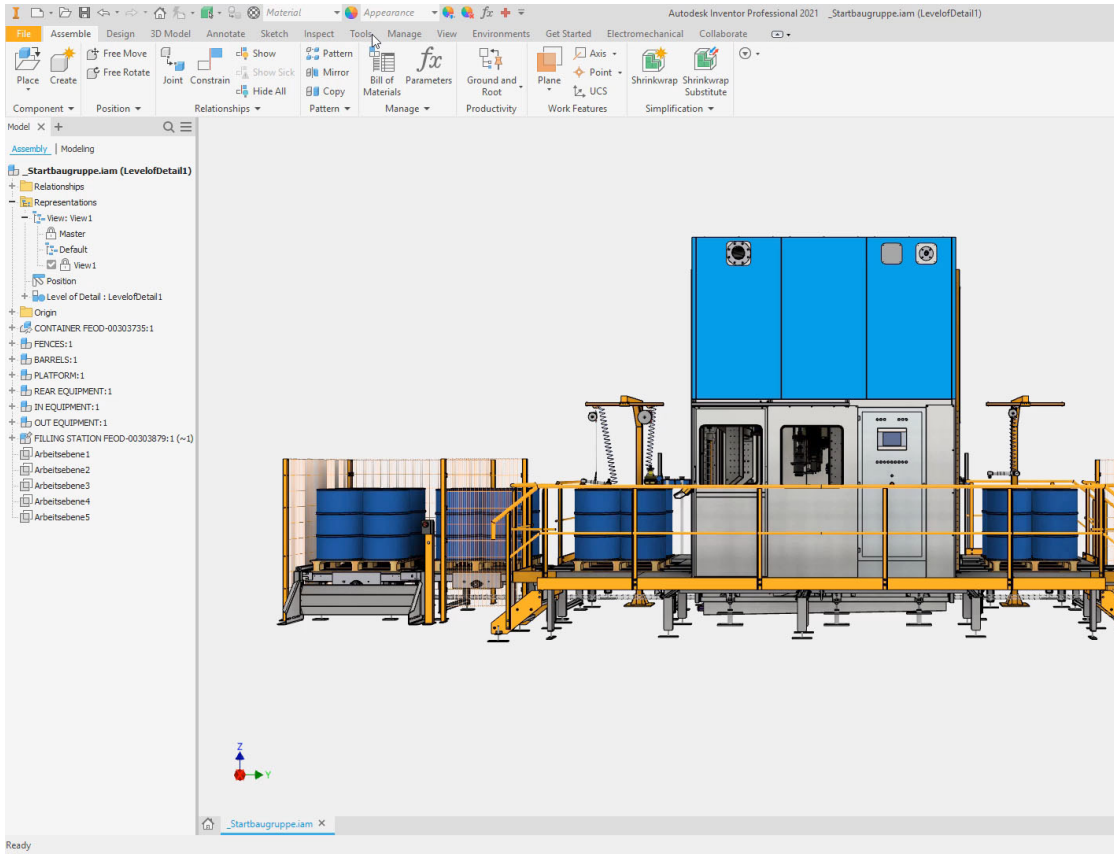


Language is a cooperative, multiagent process.

Language systems improve when they plan against simulated humans.

Future Work

Adaptive pragmatics



“Let’s call that collection of barrels a ‘pod’. Add a ‘pod’ on the platform”

Future Work

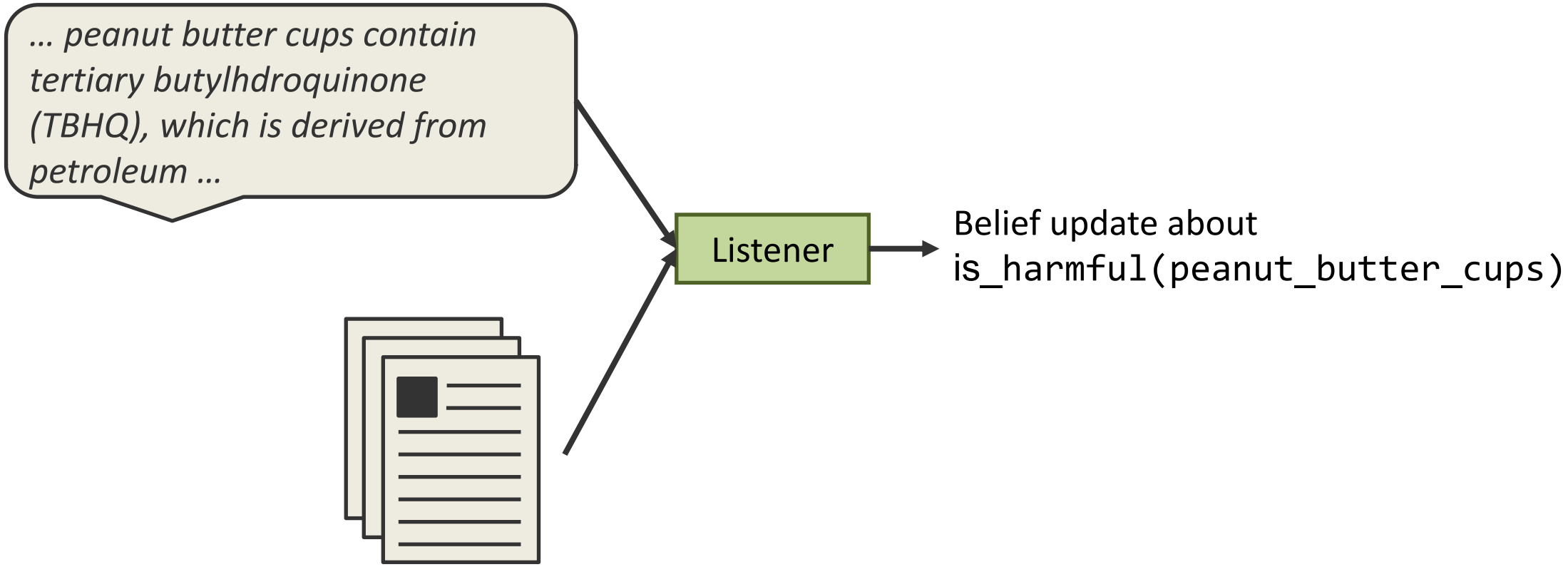
Broadening grounding in NLP: perception and action

The blue cups are fragile

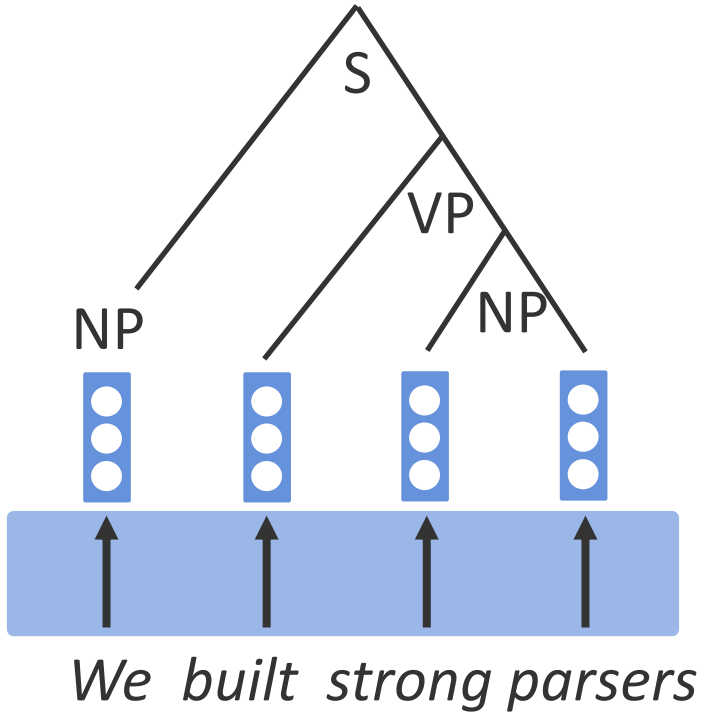
Be careful around Grandpa

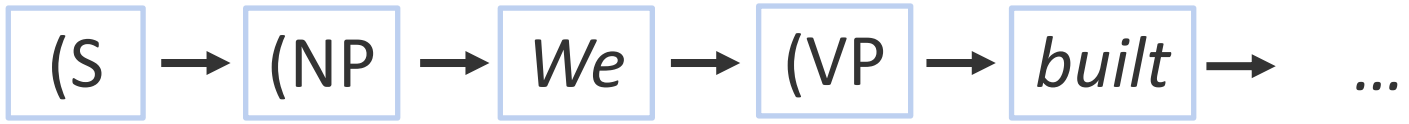
Future Work

Broadening grounding in NLP: intents and interpretations



Other Work: Structured Prediction & Core NLP



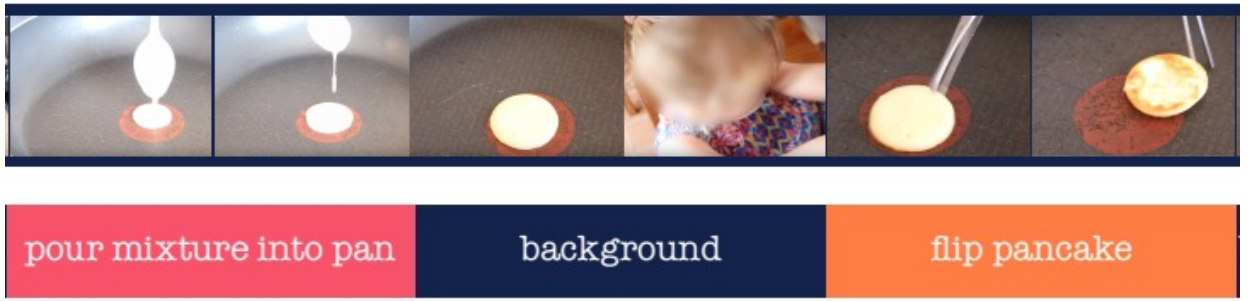


[**Fried***, Stern*, and Klein. ACL 2017]
[Stern, **Fried**, and Klein. EMNLP 2017]
[**Fried** and Klein. ACL 2018]
[**Fried***, Kitaev*, and Klein. ACL 2019]
[Kuncoro*, Kong*, **Fried***, Yogatama, Rimell, Dyer, and Blunsom. TACL 2020]

Other Work: Learning and Using Task Structure

In instructional videos:

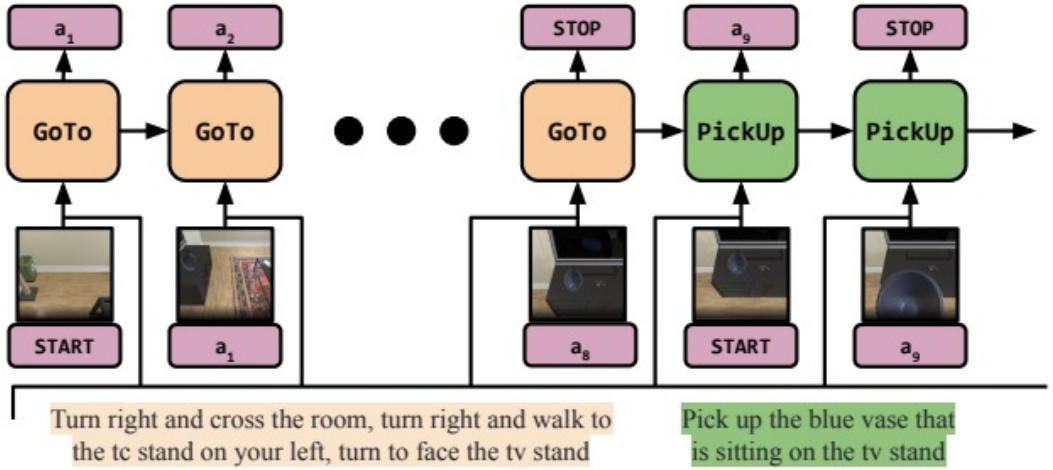
Folks my pan is nice and hot… I'll pour all the batter in there and let it cook… then flip it over once it starts to set …



[**Fried**, Alayrac, Blunsom, Dyer, Clark, and Nematzadeh. ACL 2020]

For embodied instruction following:

Turn right and cross the room… Pick up the blue vase that is sitting on the tv stand …

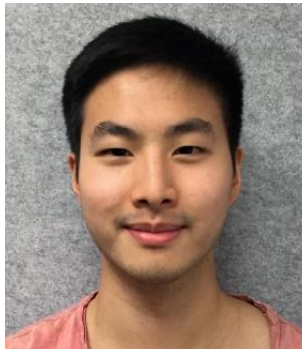


[Corona, **Fried**, Devin, Klein, and Darrell. NAACL 2021]

Collaborators

Jacob Andreas

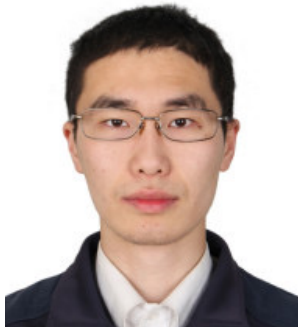
Taylor Berg-Kirkpatrick



Justin Chiu

Volkan Cirik

Trevor Darrell



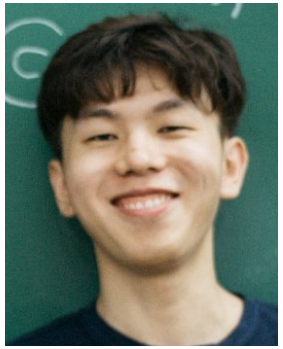
Ronghang Hu

Dan Klein

Louis-Philippe Morency

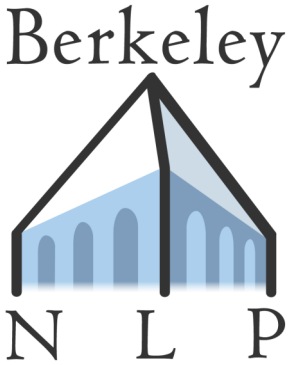
Anna Rohrbach

Kate Saenko



Sheng Shen

Thank you!



dfried@berkeley.edu
cs.berkeley.edu/~dfried/

Outperforming Training Data (Toy Example)

Training Data

Context → "Language"

AX → x
AX → x
BX → x
BX → x

True, but under-informative

AX → ax
BX → bx

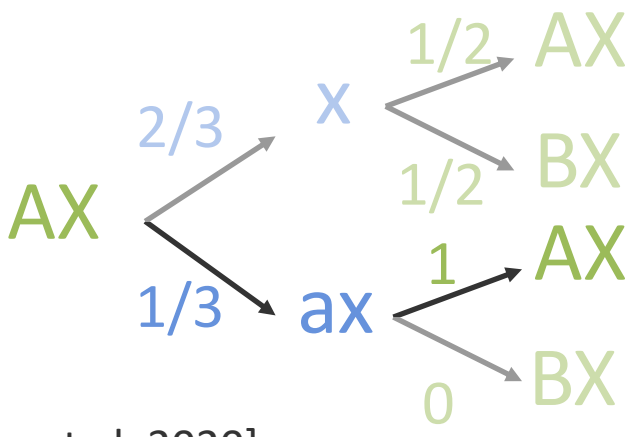
True and informative

Base Speaker

$P_S(\ x \mid AX) = 2/3$
$P_S(ax \mid AX) = 1/3$

Base Listener

$P_L(AX \mid x) = 1/2$
$P_L(AX \mid ax) = 1$



Pragmatics as best response [Franke 2009; Jäger 2014]
Other formalisms:

- Recursive Bayesian agents [Frank and Goodman 2012; Jeon et al. 2020]
- Optimal transport of beliefs [Wang et al. 2020]
- Rate-distortion communication [Zaslavsky et al. 2020]