

# Unified Pragmatic Models for Generating and Following Instructions



Daniel Fried, Jacob Andreas, and Dan Klein  
UC Berkeley



# Interpreting instructions

---

Instruction

*walk along the blue carpet and you pass two objects*

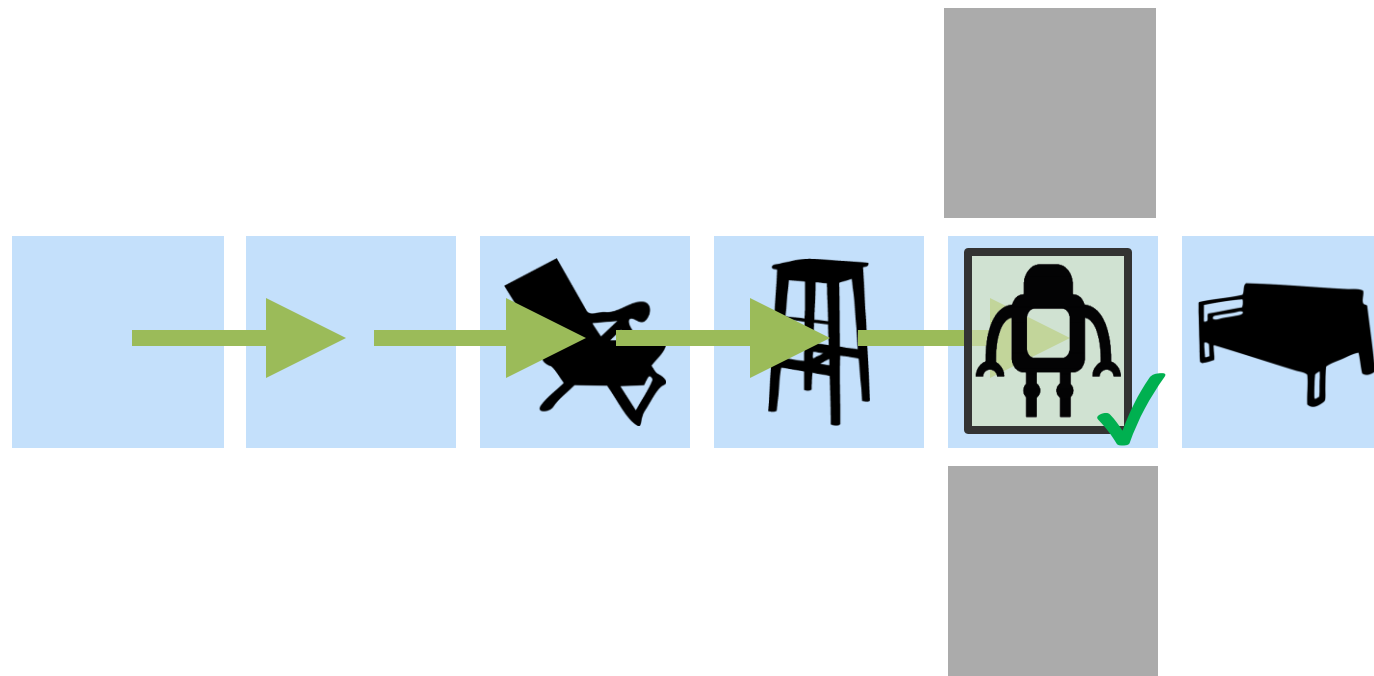




# Interpreting instructions

---

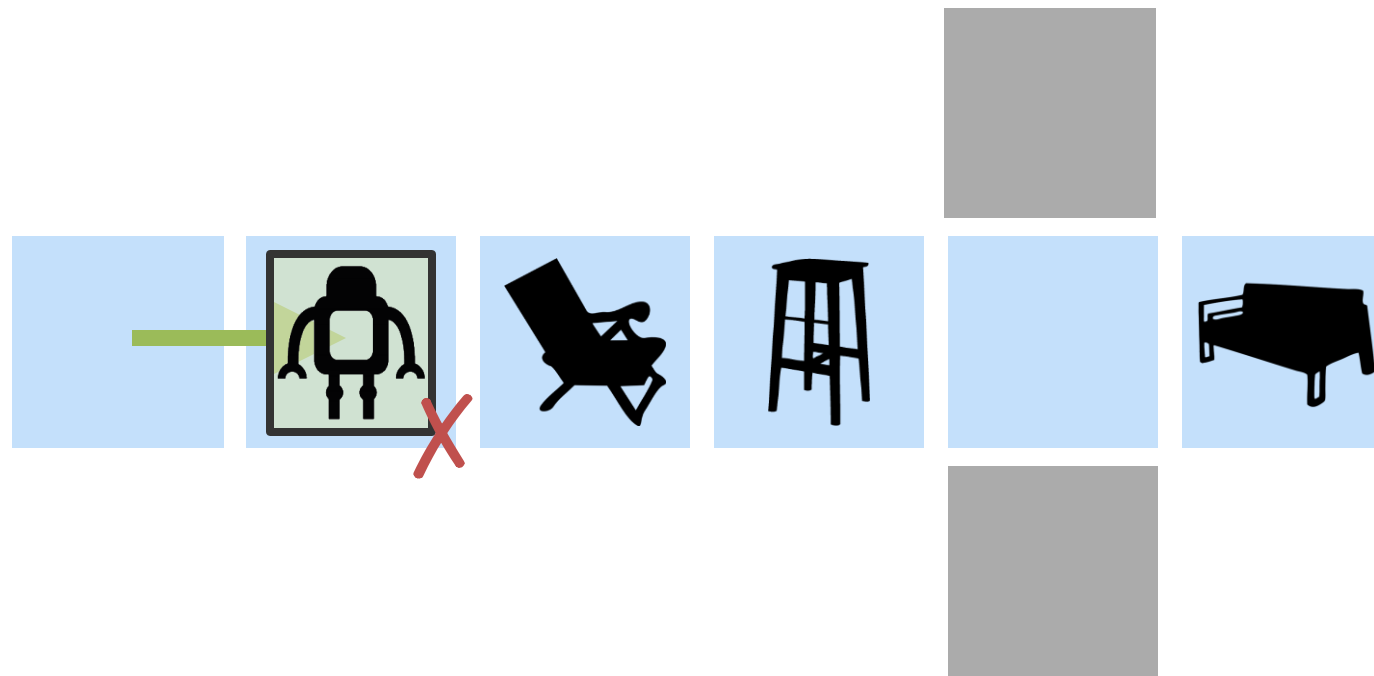
Instruction *walk along the blue carpet and you pass two objects*

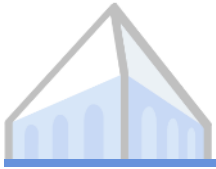




# Interpreting instructions

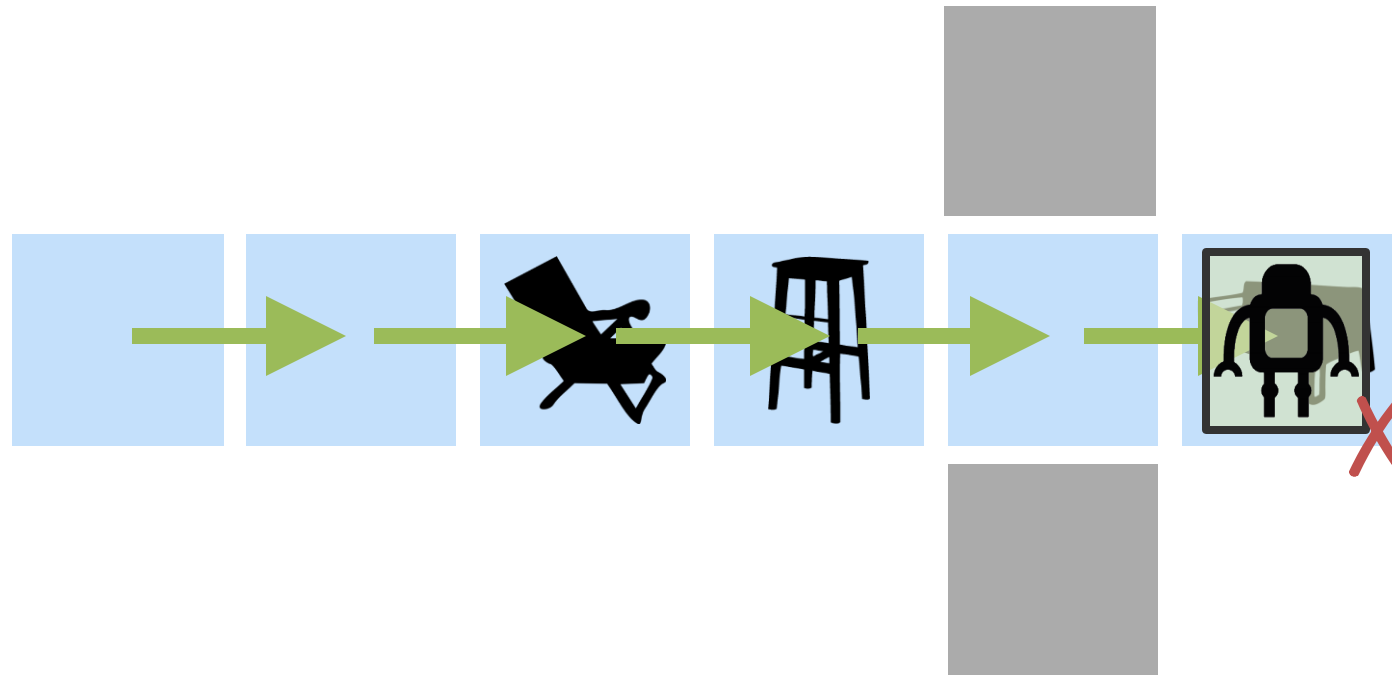
Instruction *walk along the blue carpet and you pass two objects*





# Interpreting instructions

Instruction *walk along the blue carpet and you pass two objects*

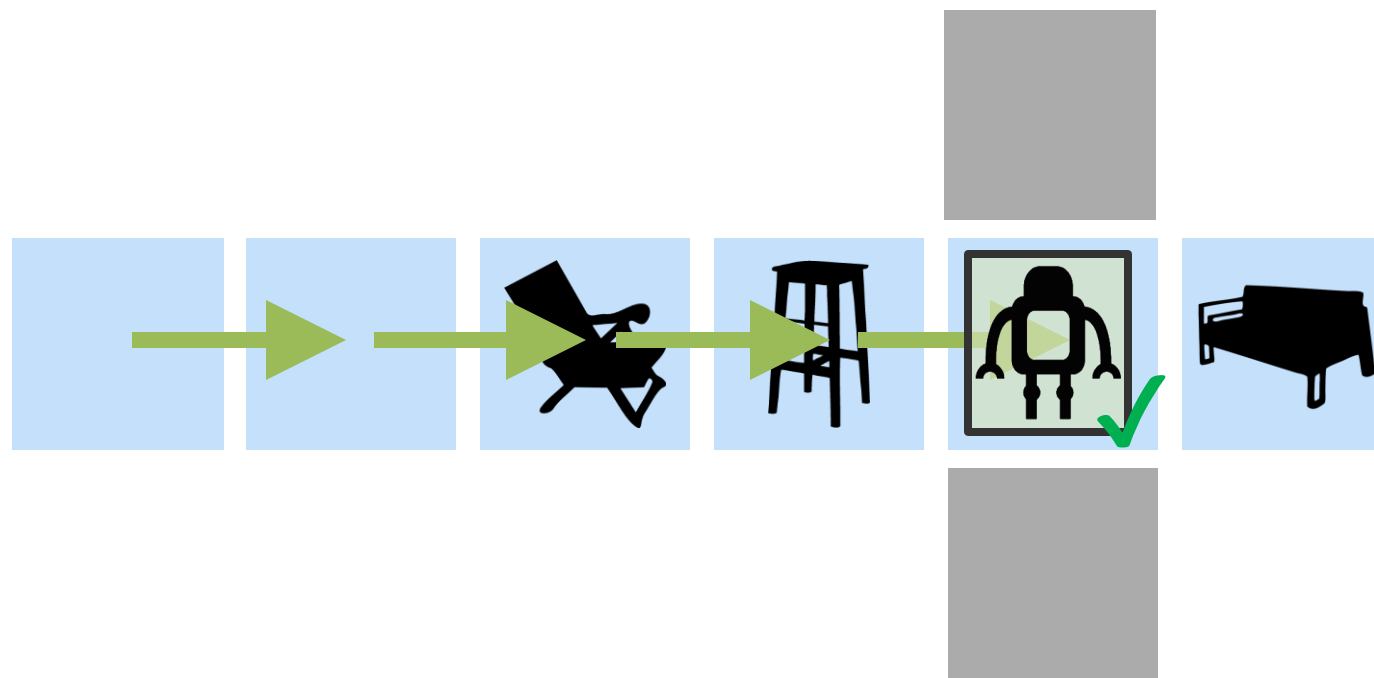




# Interpreting instructions

Instruction

*walk along the blue carpet and you pass two objects*



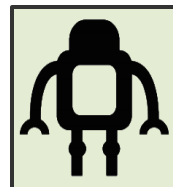


# Listener: reasoning about routes



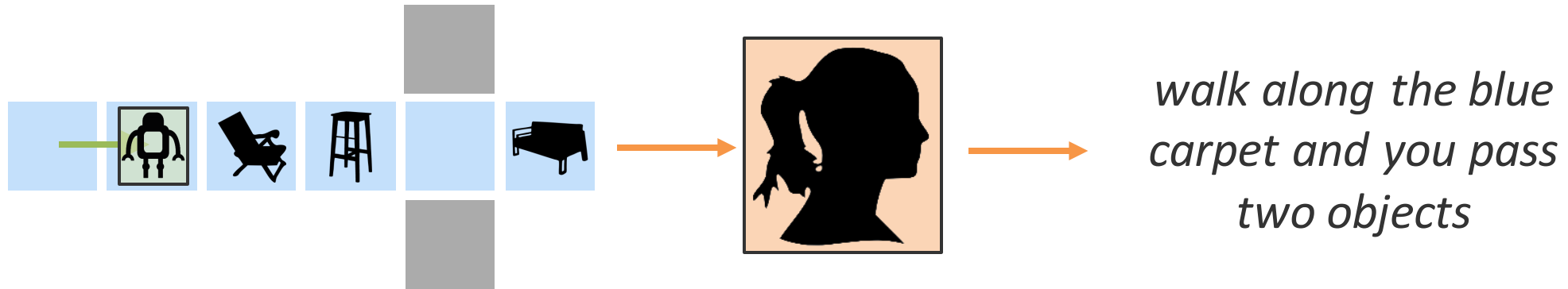
*walk along the blue  
carpet and you pass  
two objects*

*walk along the blue  
carpet and you pass  
two objects*

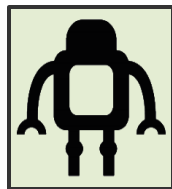




# Listener: reasoning about routes



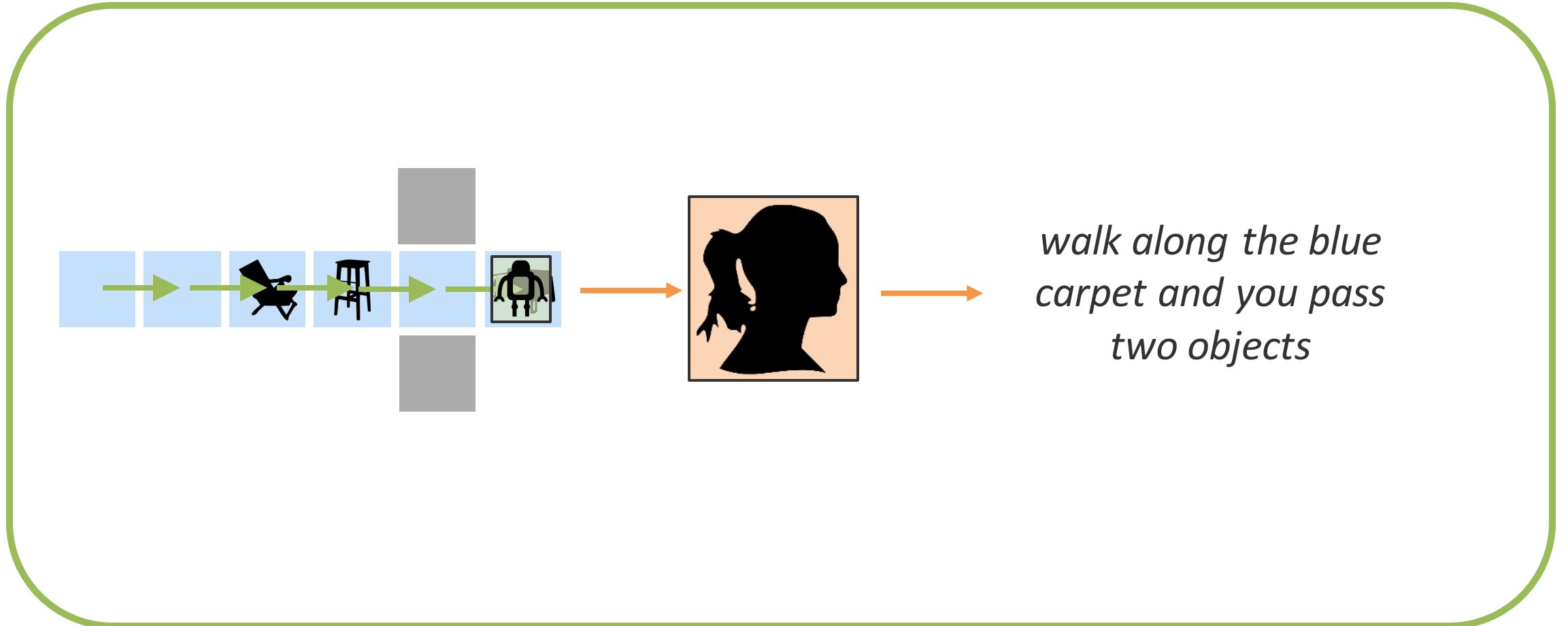
*walk along the blue carpet and you pass two objects*



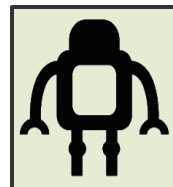




# Listener: reasoning about routes

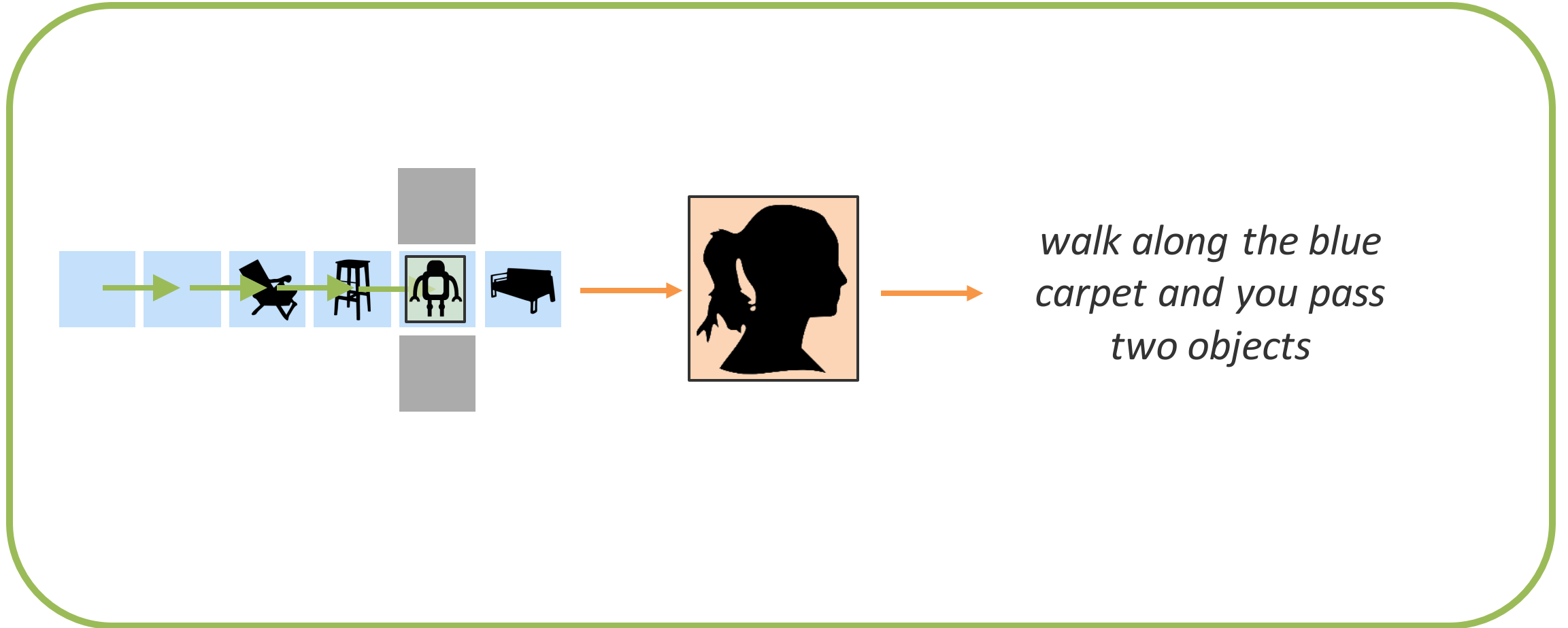


walk along the blue carpet and you pass two objects

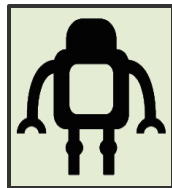




# Listener: reasoning about routes

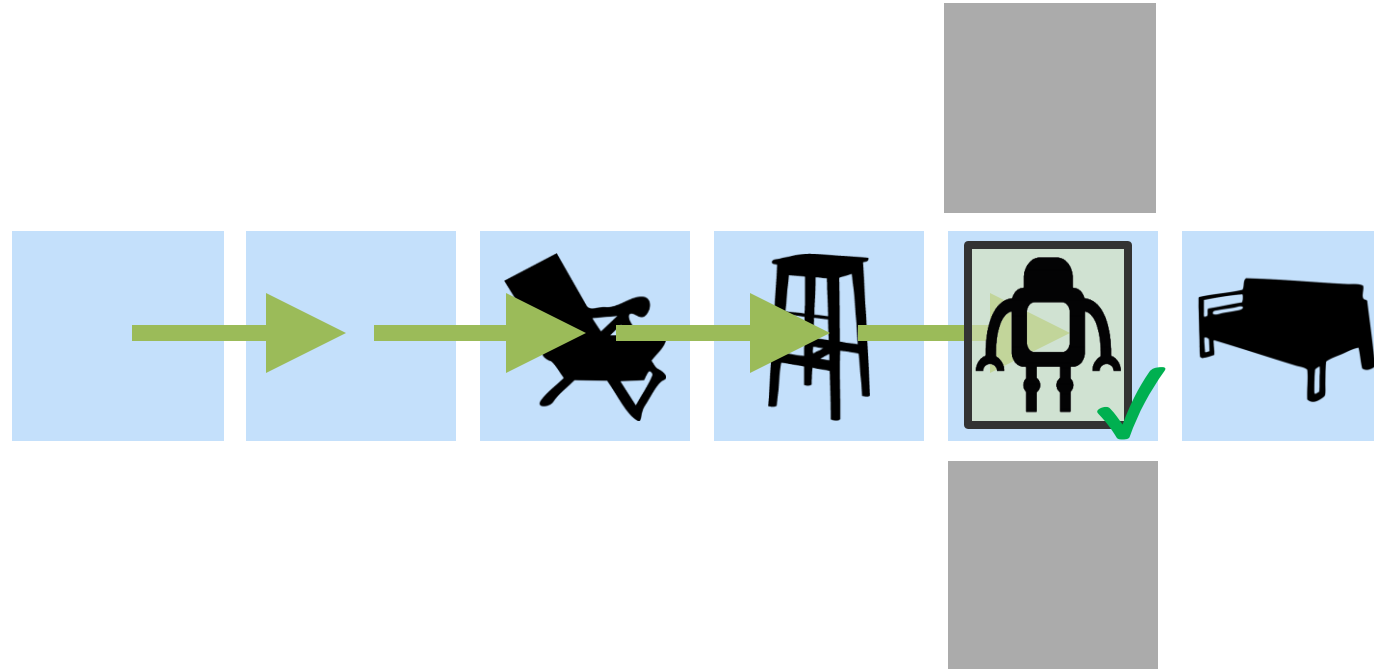


*walk along the blue carpet and you pass two objects*





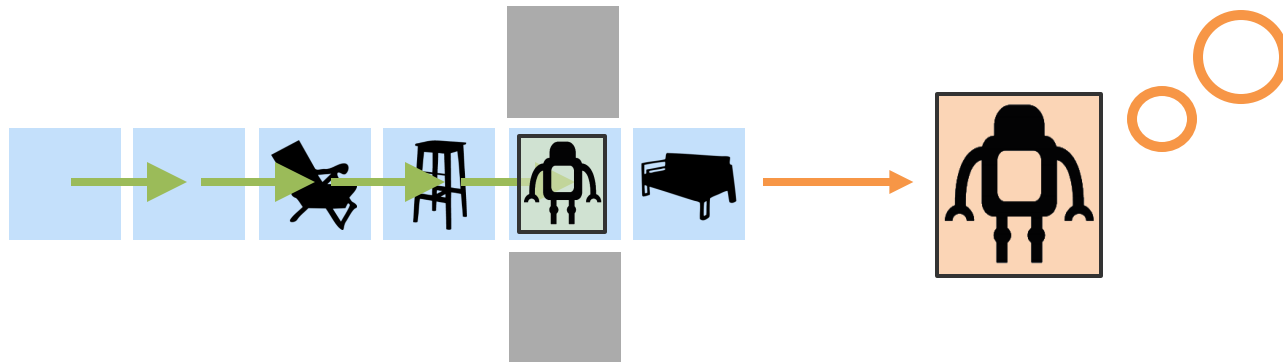
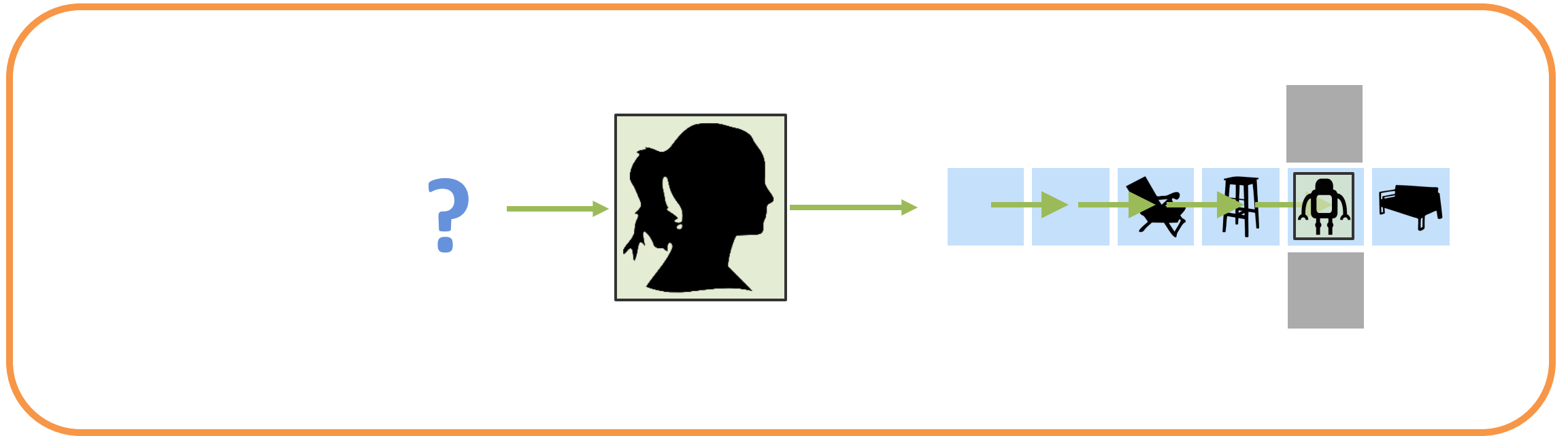
# Generating instructions



Generated Instruction: *go forward four segments to the intersection with the bare concrete hall*



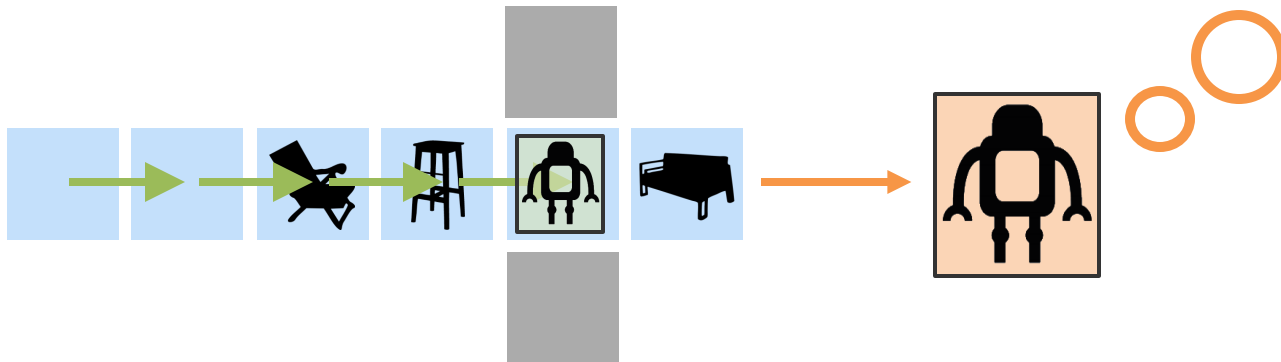
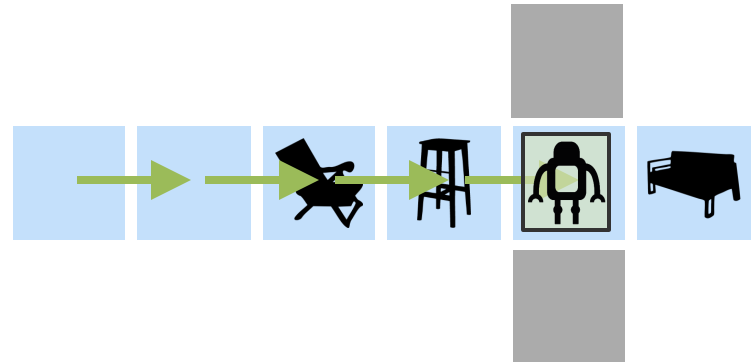
# Speaker: reasoning about interpretation





# Speaker: reasoning about interpretation

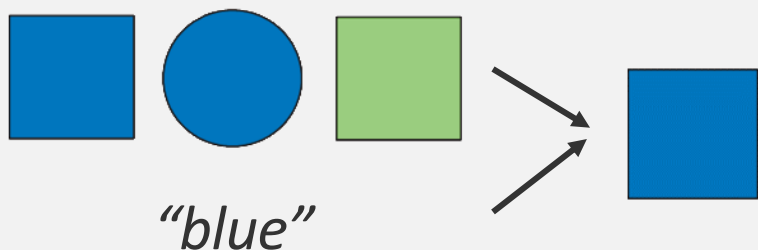
*go forward four  
segments to the  
intersection with the  
bare concrete hall*





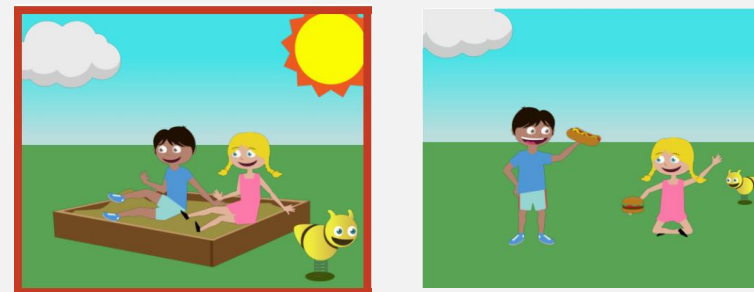
# Explicit pragmatic reasoning

## Interpretation



Frank and Goodman, 2012;  
Goodman and Stühlmüller, 2013;  
Wang et al., 2016

## Generation



*“The sun is in the sky”*

Golland et al., 2010; Monroe and  
Potts, 2015; Andreas and Klein, 2016;  
Mao et al., 2016; ...

## Both

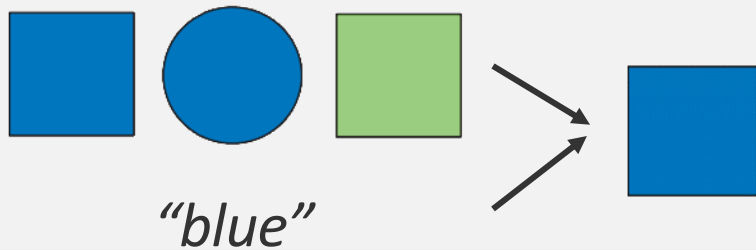
Monroe et al. 2017

Our work



# Explicit pragmatic reasoning

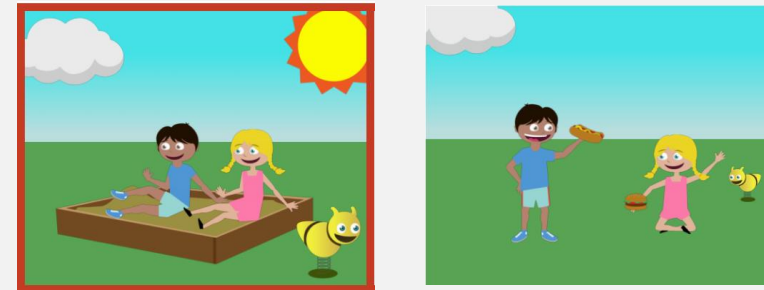
## Interpretation



Frank and Goodman, 2012;  
Goodman and Stühlmüller, 2013;

Wang et al., 2016

## Generation



*“The sun is in the sky”*

Golland et al., 2010; Monroe and  
Potts, 2015; Andreas and Klein, 2016;  
Mao et al., 2016; ...

Both

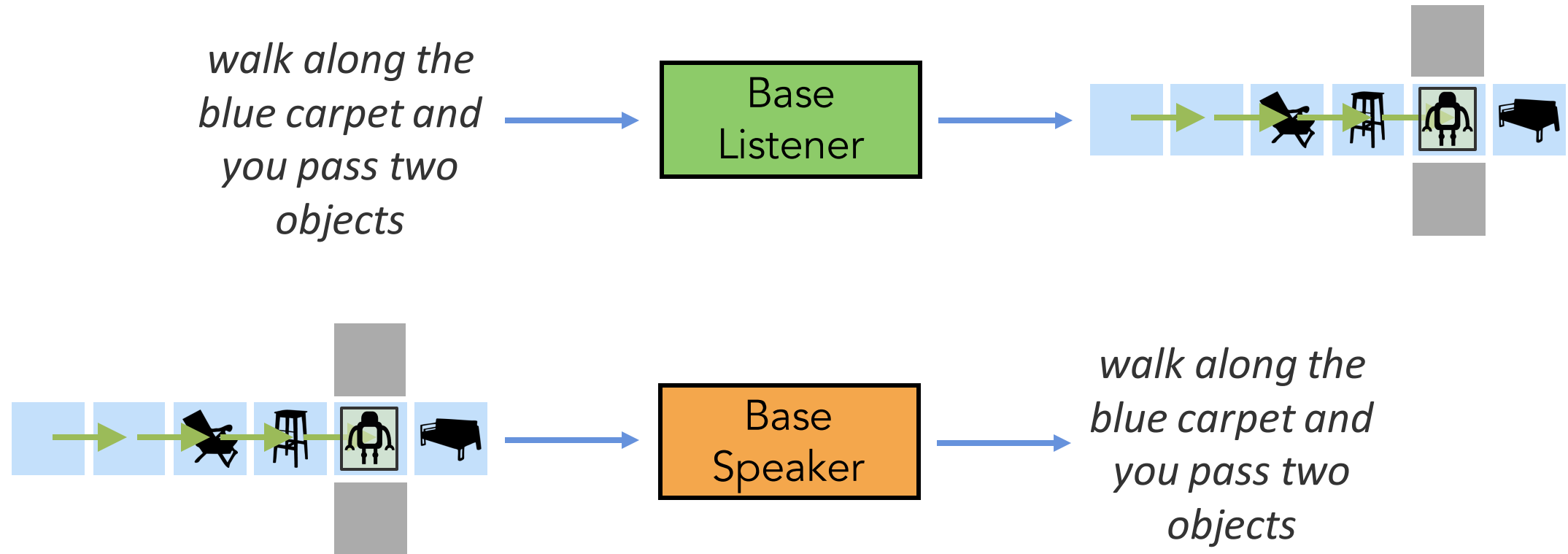
Monroe et al. 2017

Large, structured  
domains

Our work

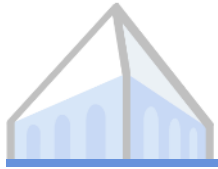


# Learned models of listeners and speakers



Both models trained independently on action sequences annotated with human instructions





# Building a pragmatic listener

Base  
Speaker

*walk along the blue  
carpet and you pass  
two objects*



Pragmatic  
Listener





# Building a pragmatic listener

?



Base  
Speaker



*walk along the blue  
carpet and you pass  
two objects*

*walk along the blue  
carpet and you pass  
two objects*

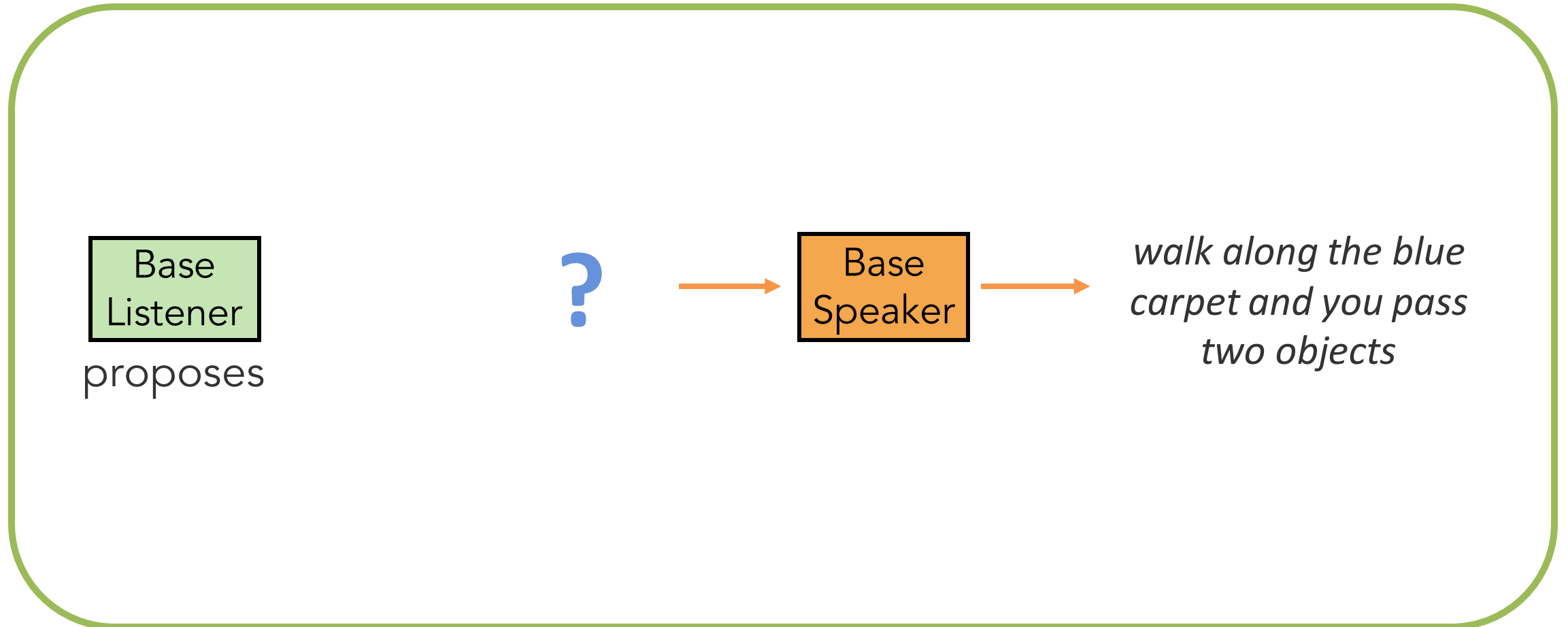


Pragmatic  
Listener





# Building a pragmatic listener

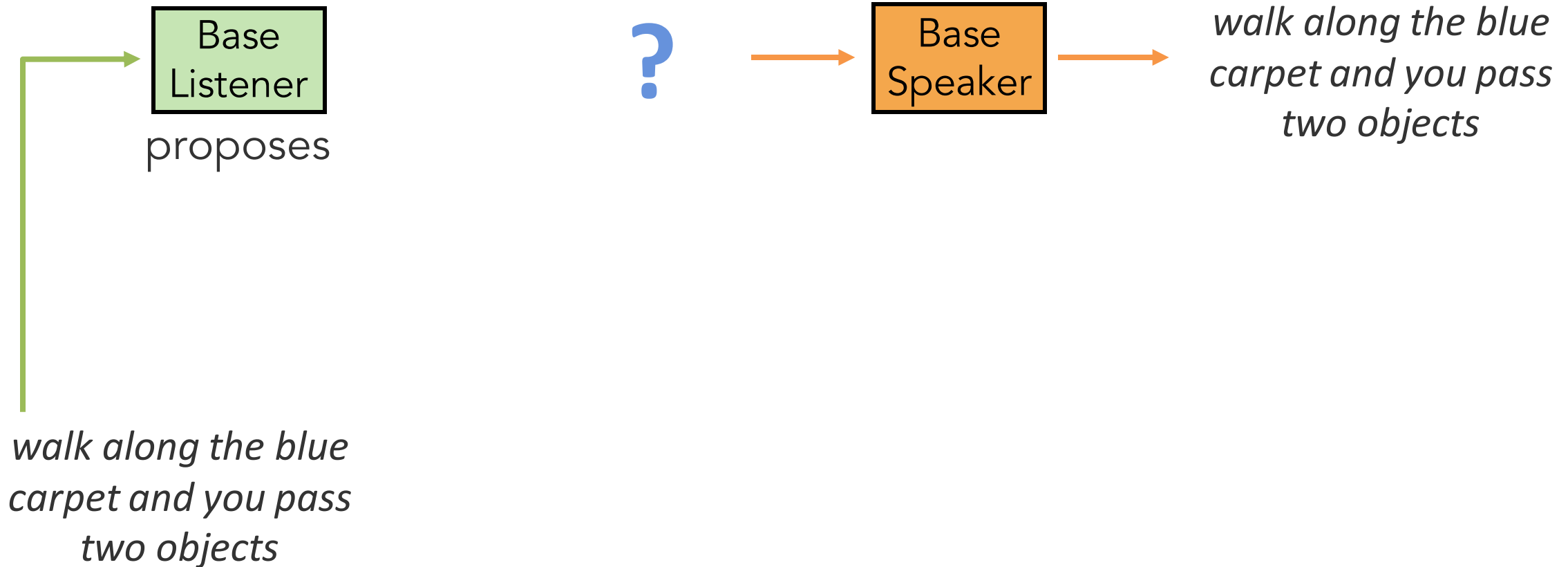


*walk along the blue carpet and you pass two objects*



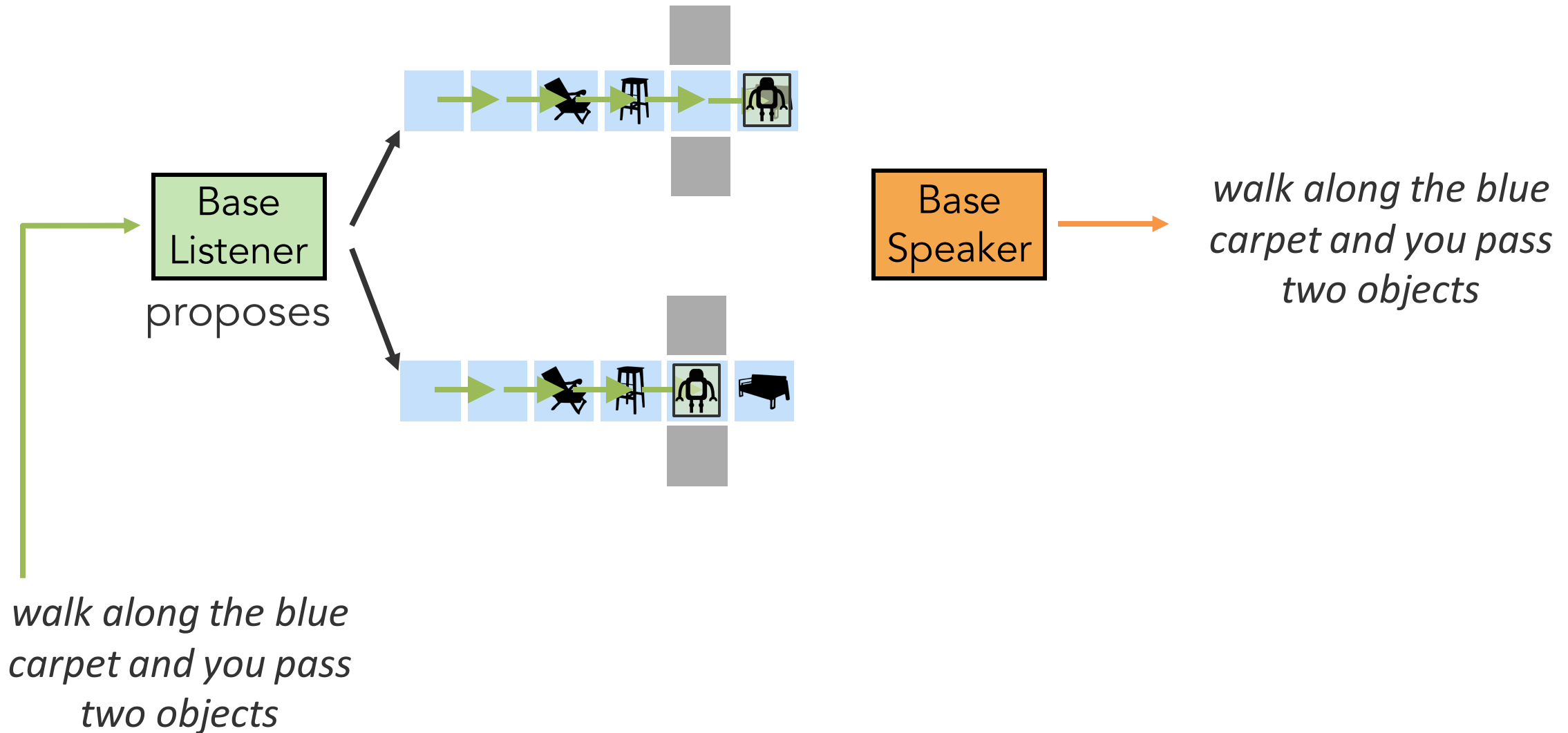


# Building a pragmatic listener



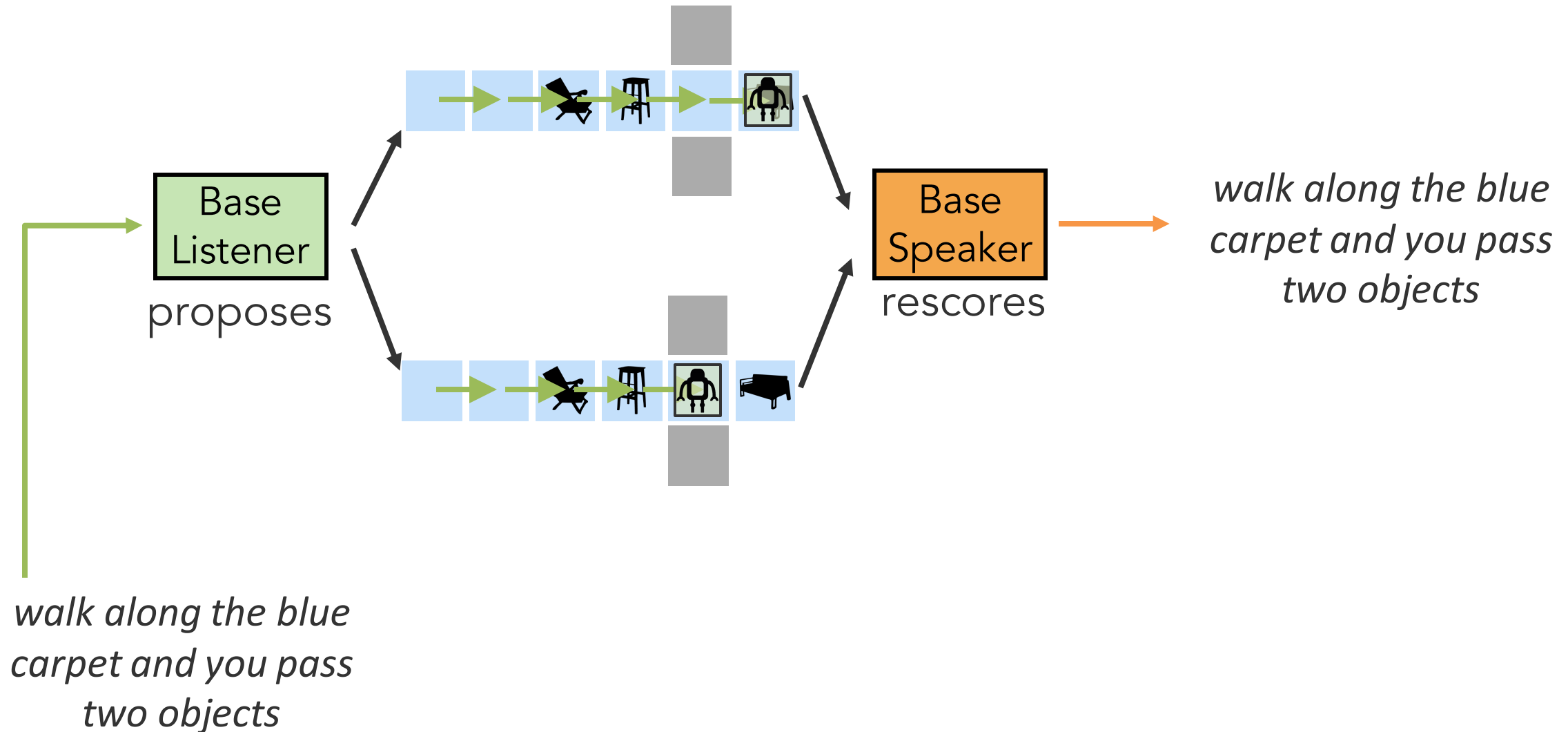


# Building a pragmatic listener



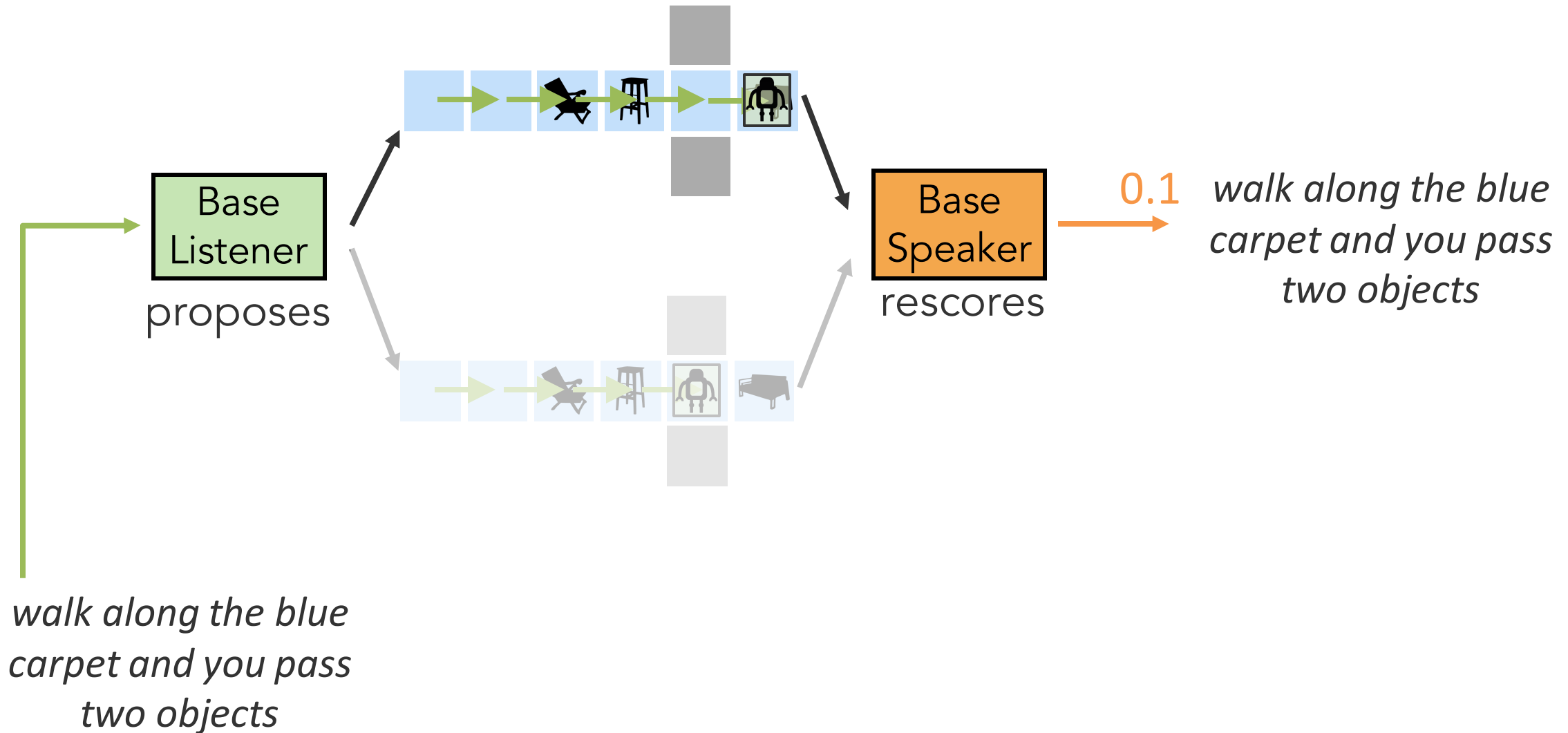


# Building a pragmatic listener



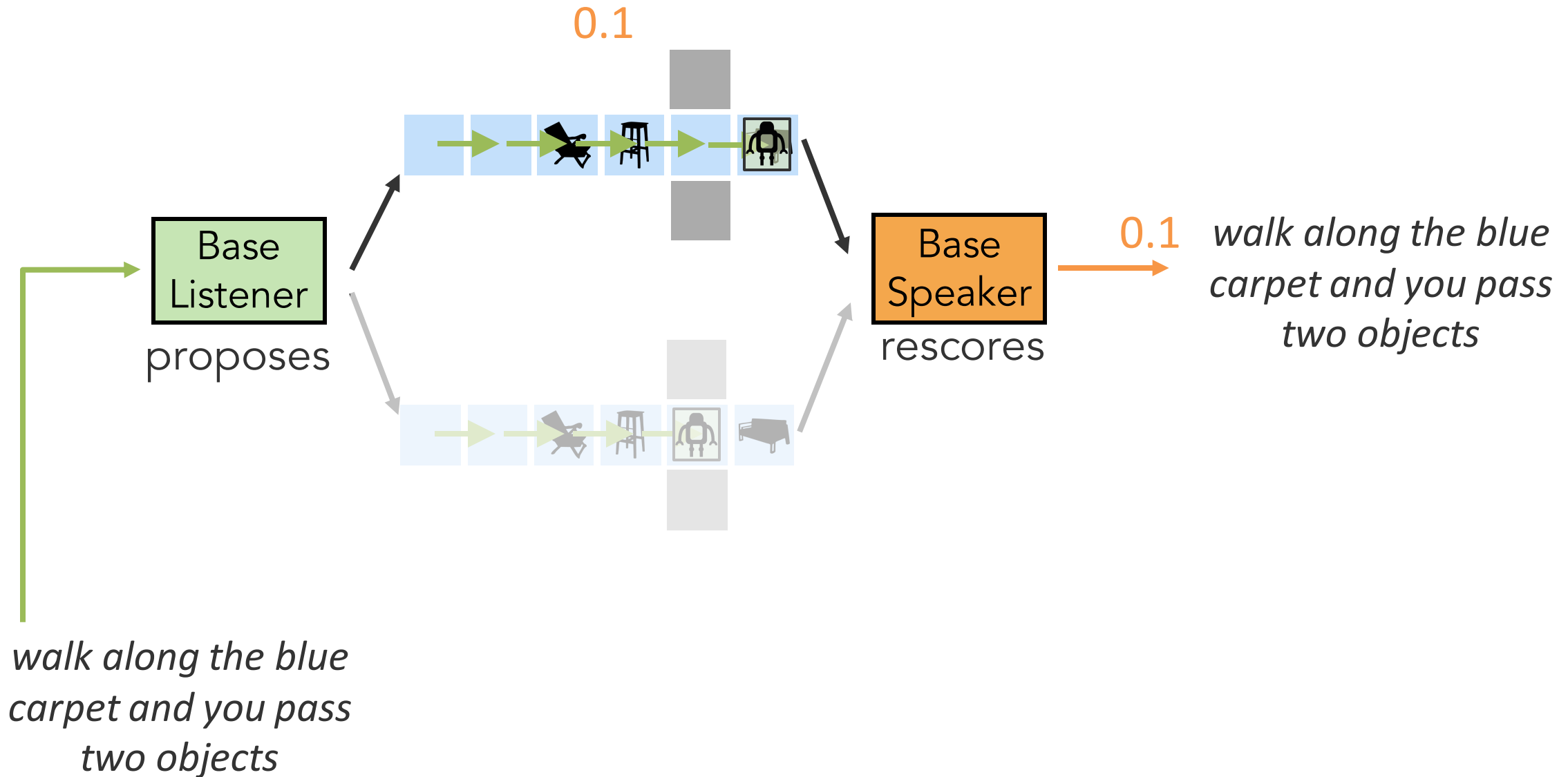


# Building a pragmatic listener





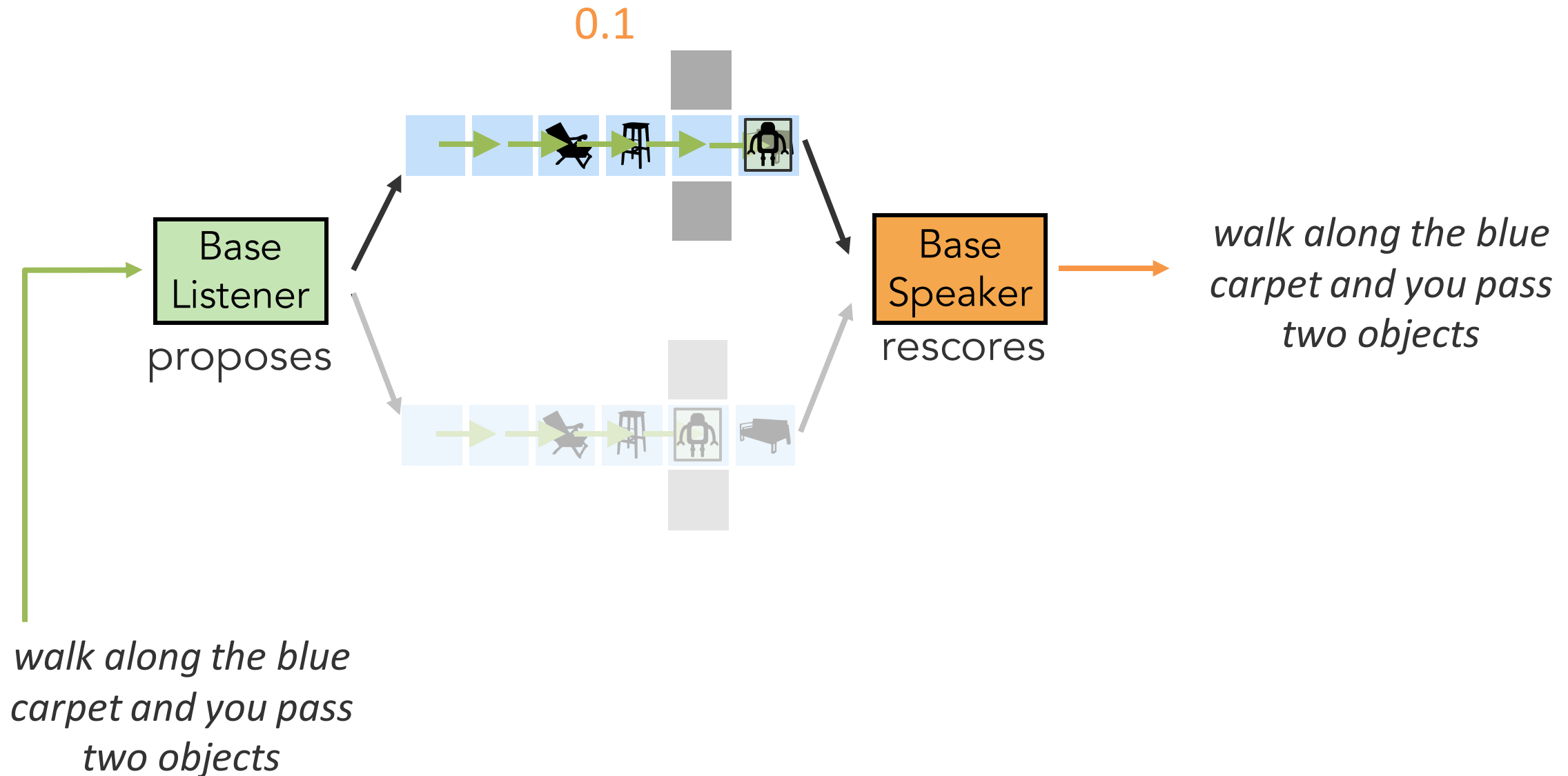
# Building a pragmatic listener





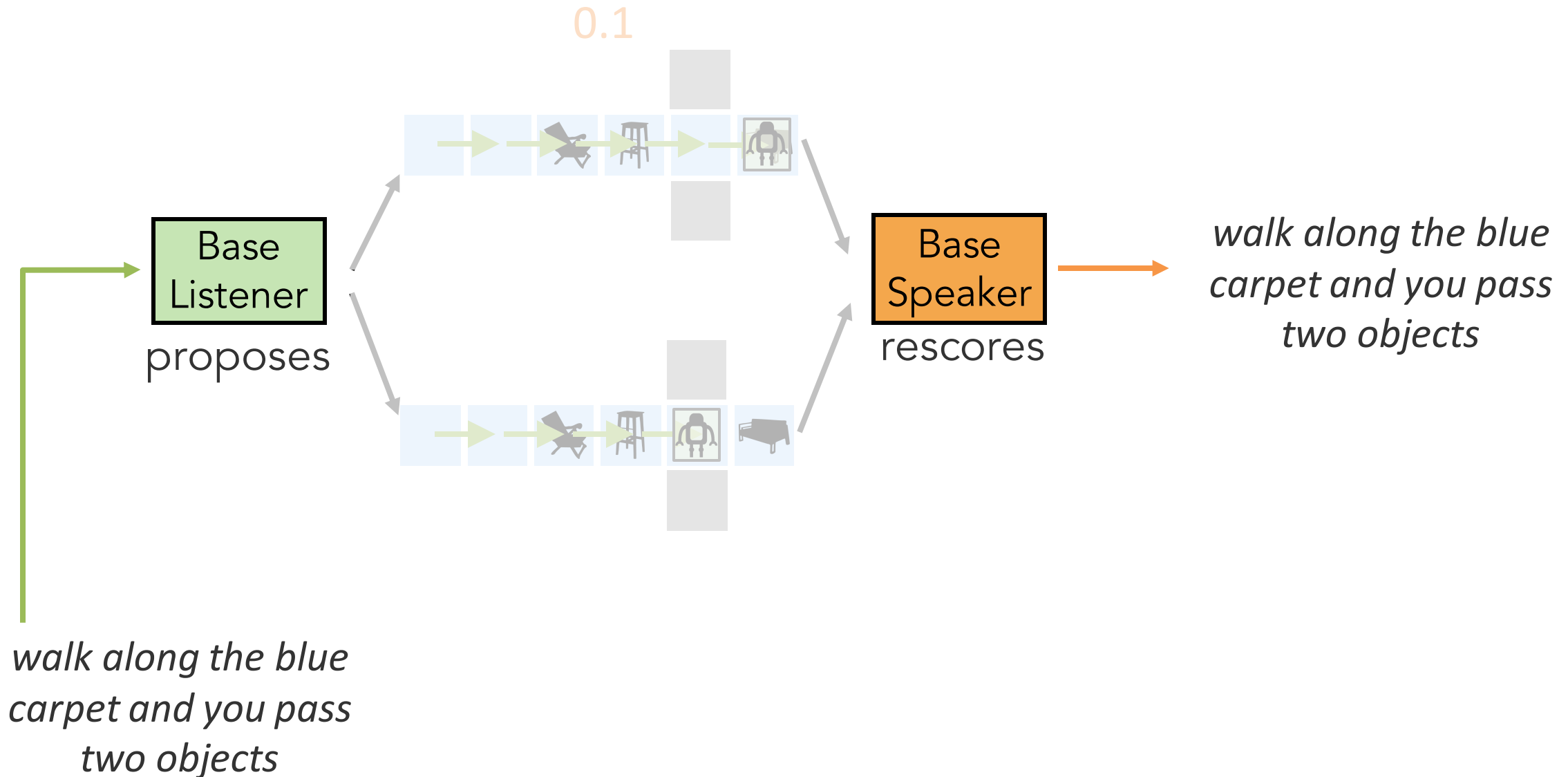


# Building a pragmatic listener



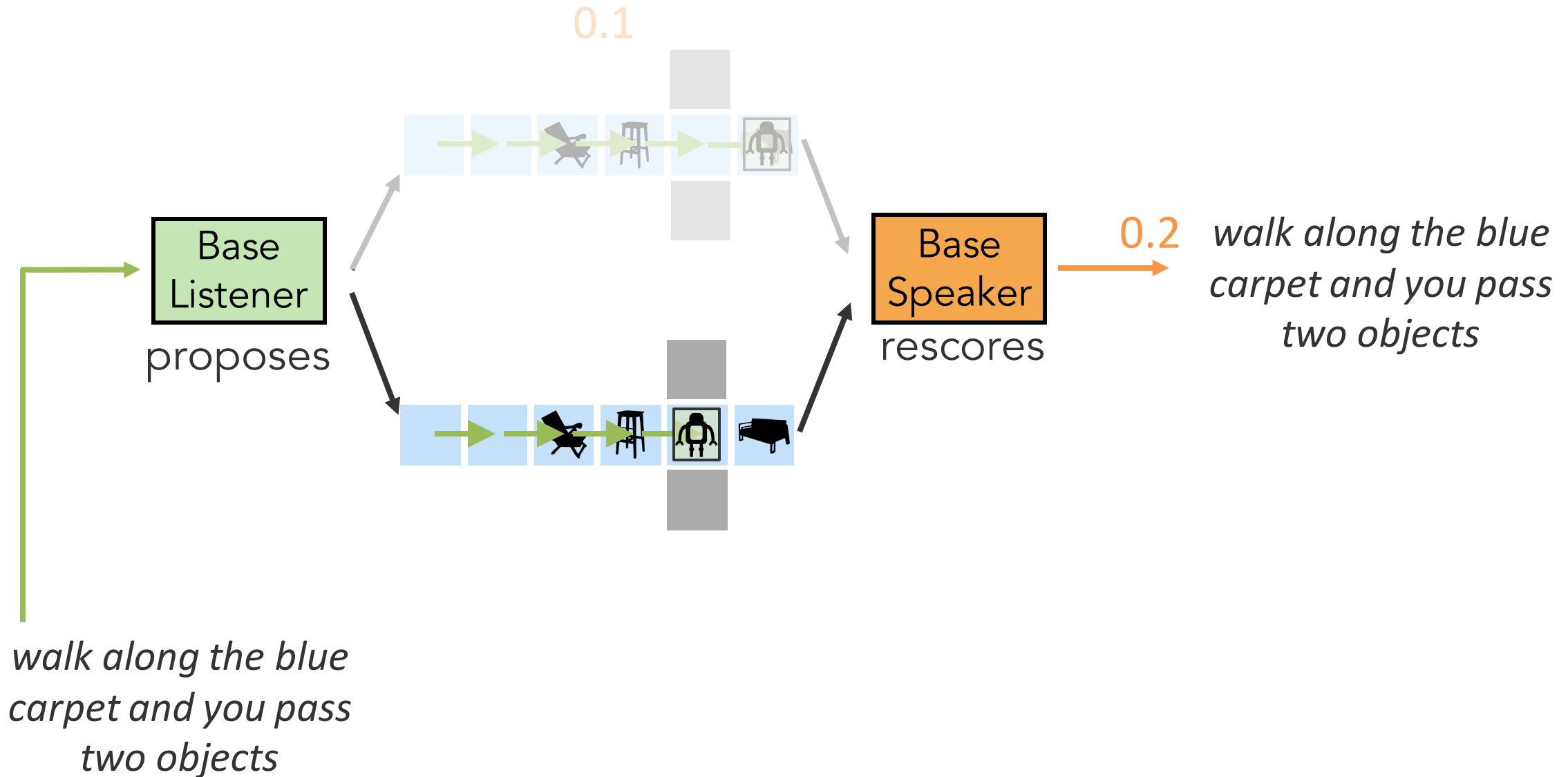


# Building a pragmatic listener



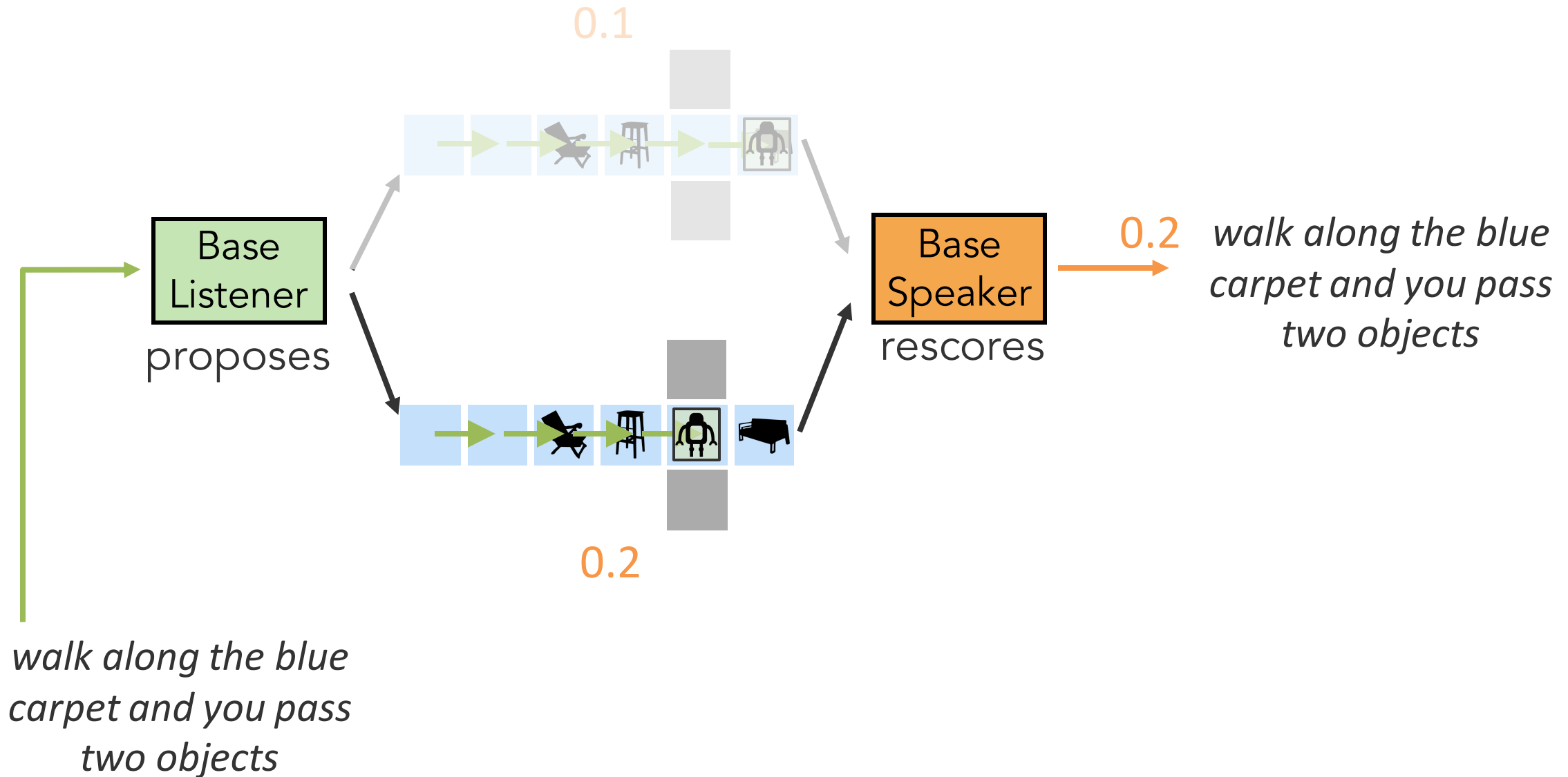


# Building a pragmatic listener



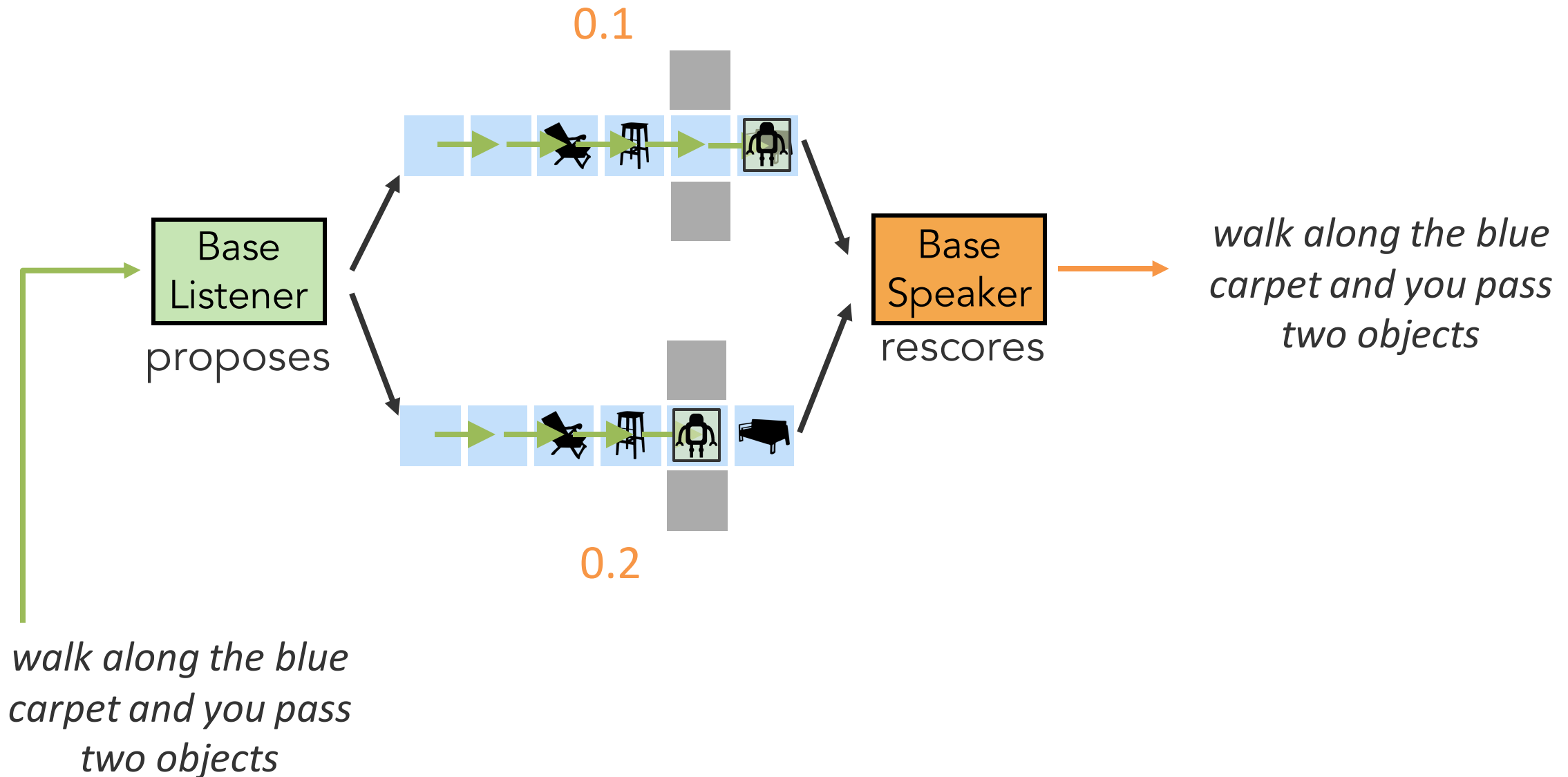


# Building a pragmatic listener



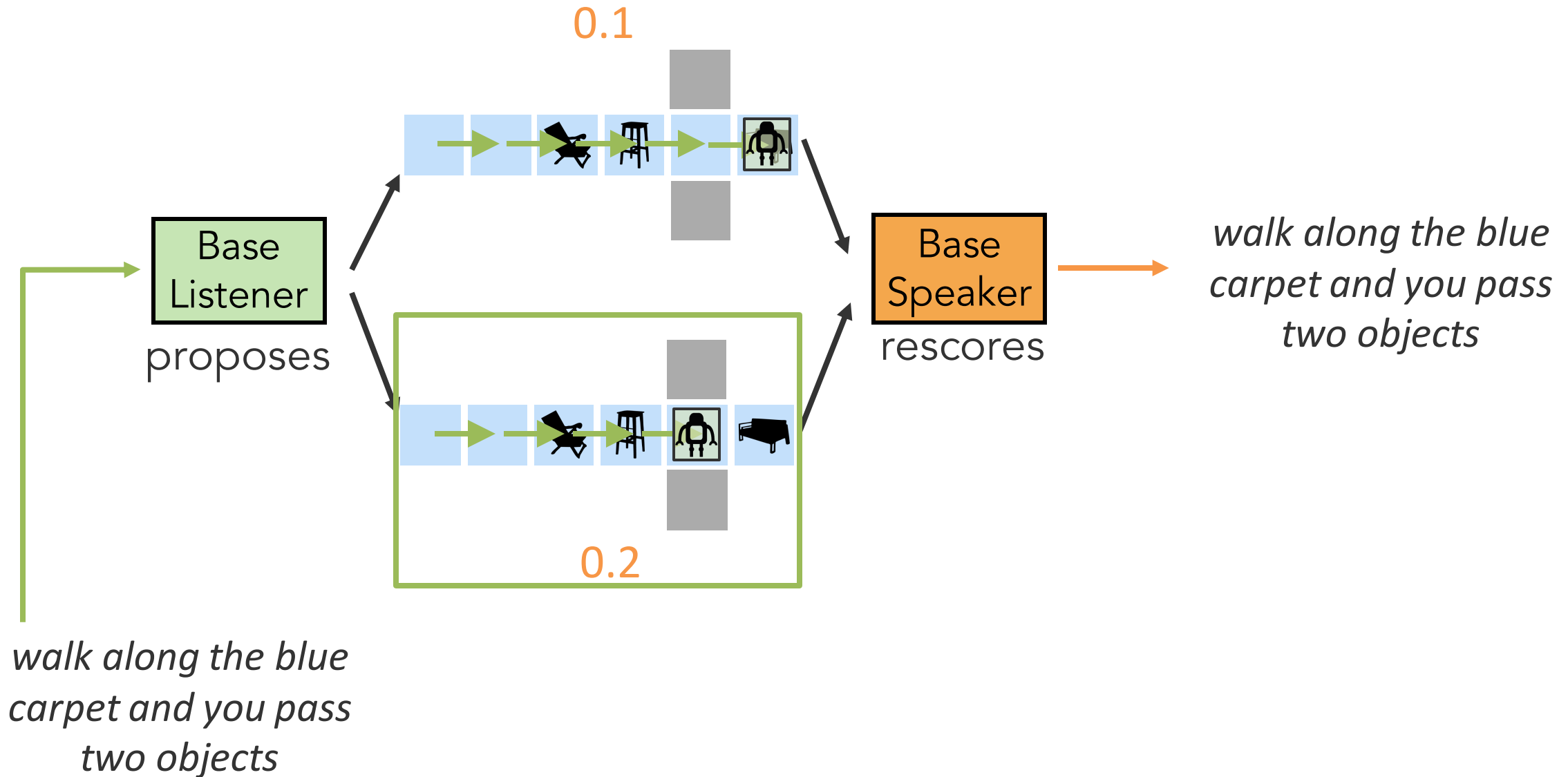


# Building a pragmatic listener





# Building a pragmatic listener

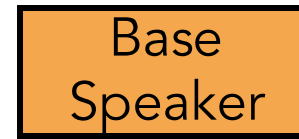




# Unified pragmatic inference

## Pragmatic Listener

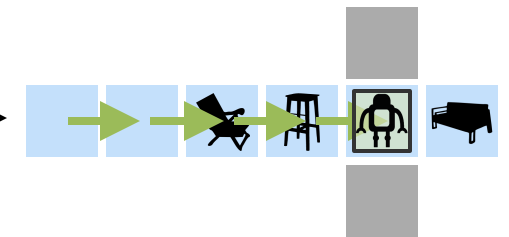
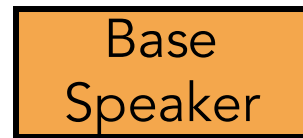
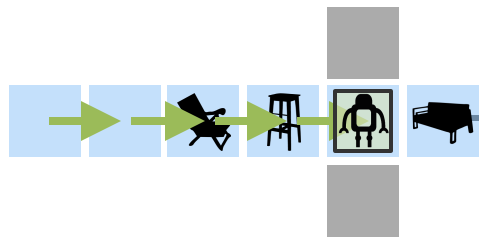
*walk along the blue carpet ...*



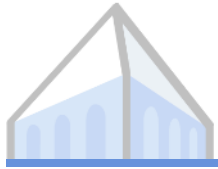
*walk along the blue carpet ...*

actions

## Pragmatic Speaker

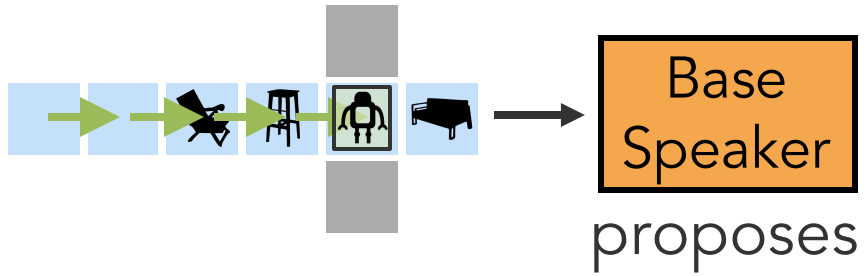


instructions

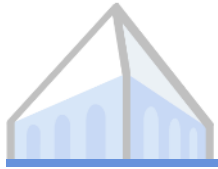


# Building a pragmatic speaker

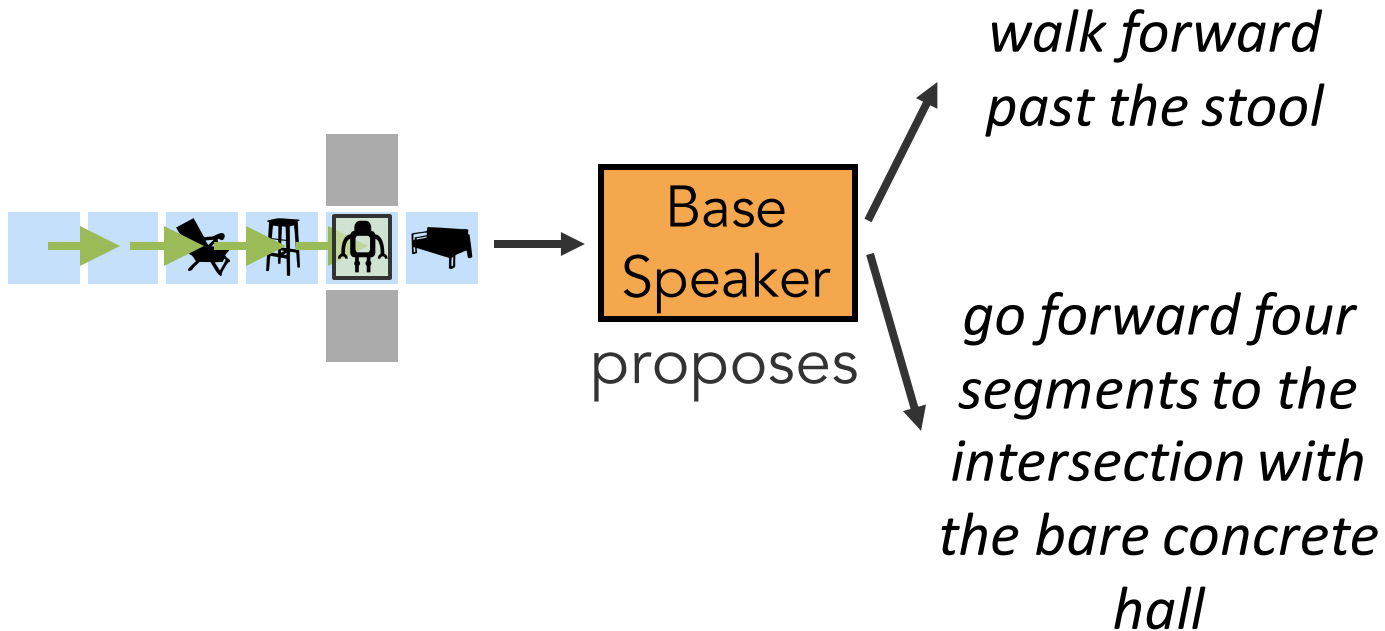
---





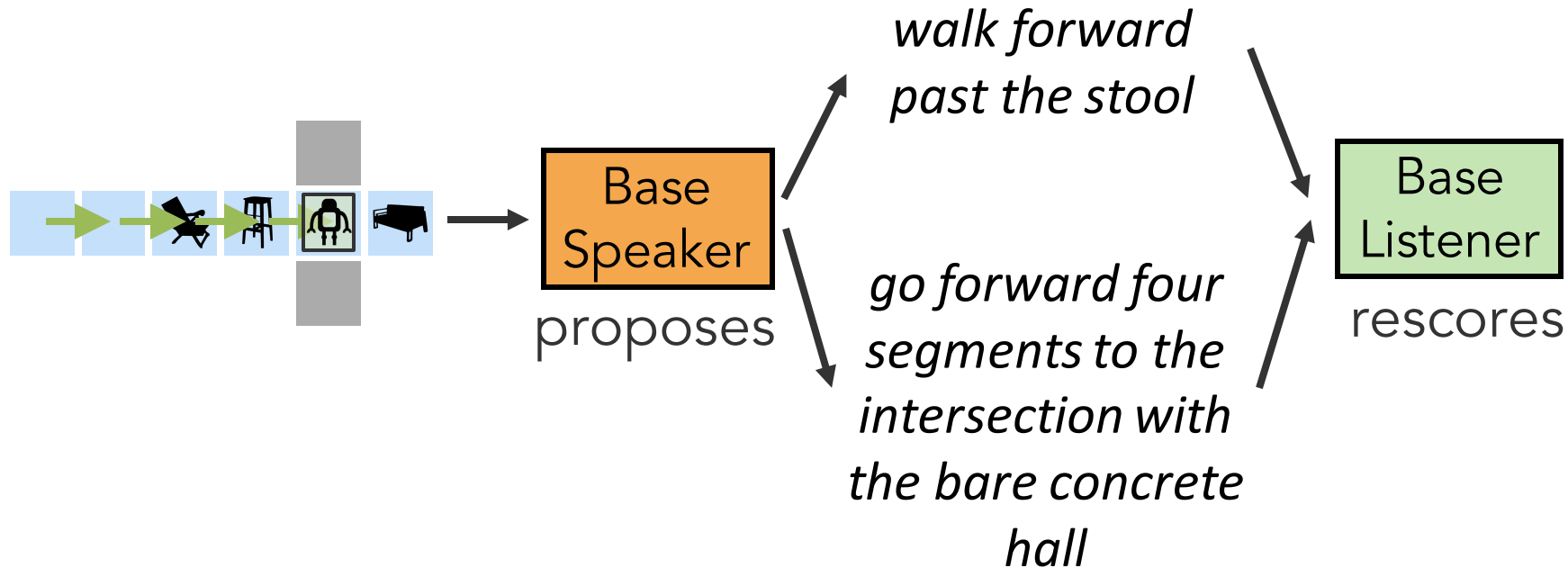


# Building a pragmatic speaker



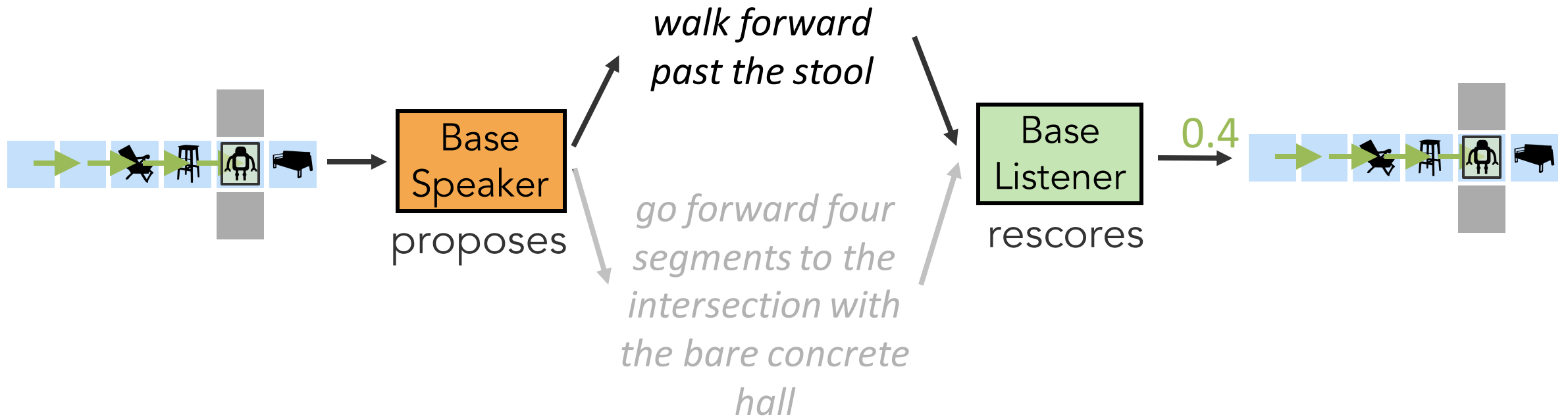


# Building a pragmatic speaker



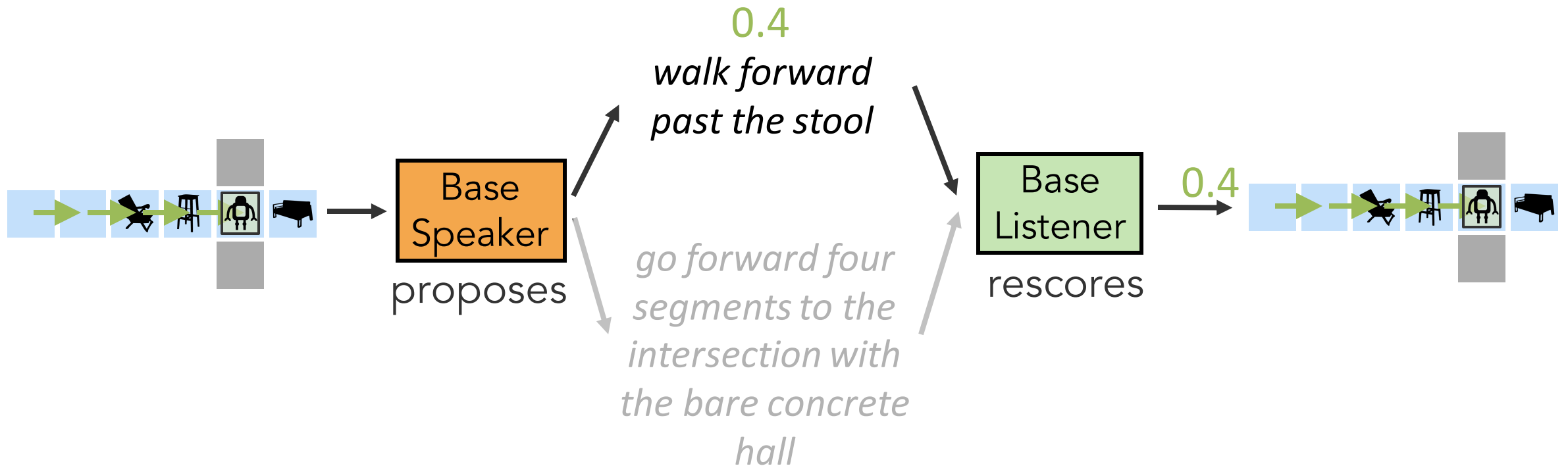


# Building a pragmatic speaker



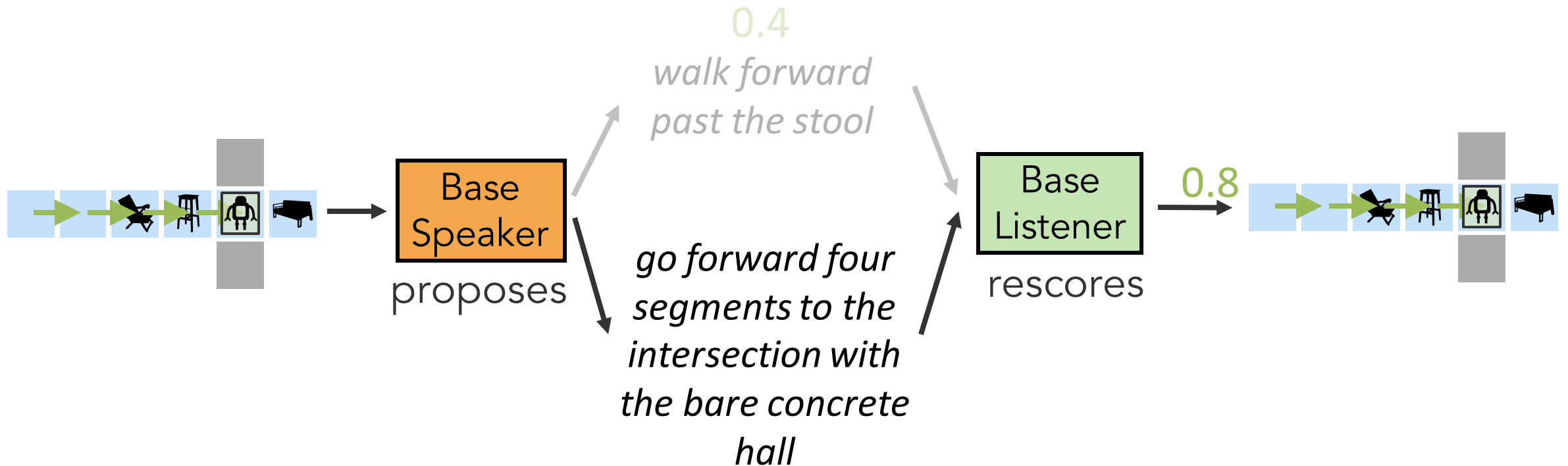


# Building a pragmatic speaker



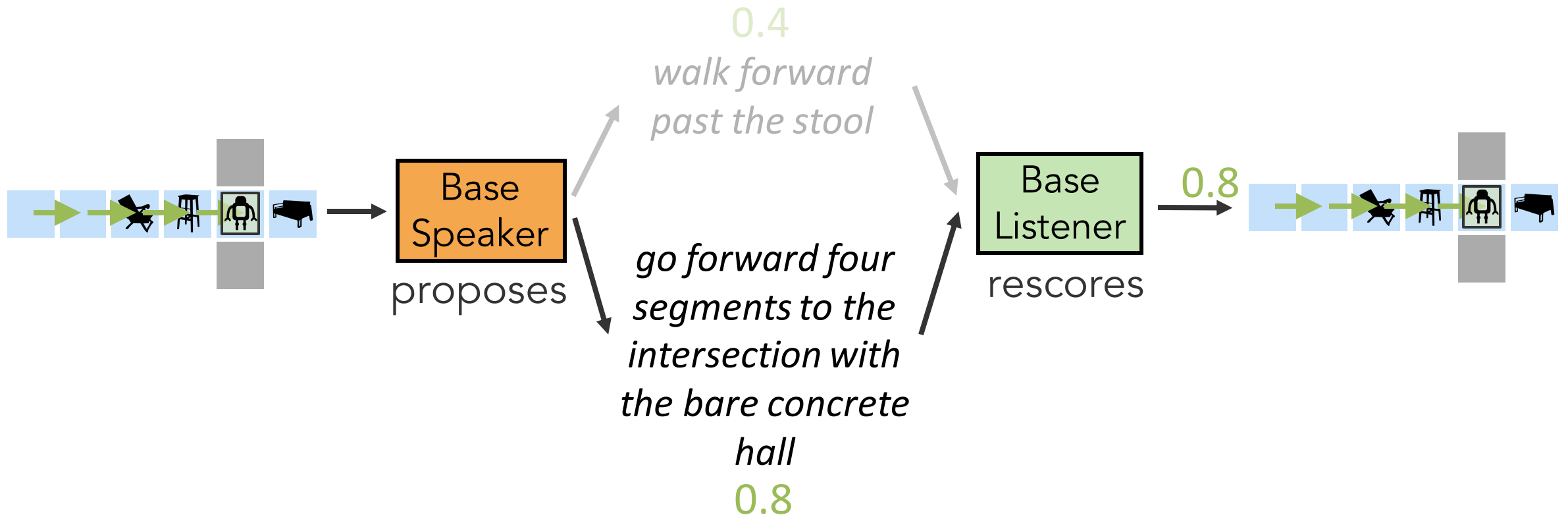


# Building a pragmatic speaker



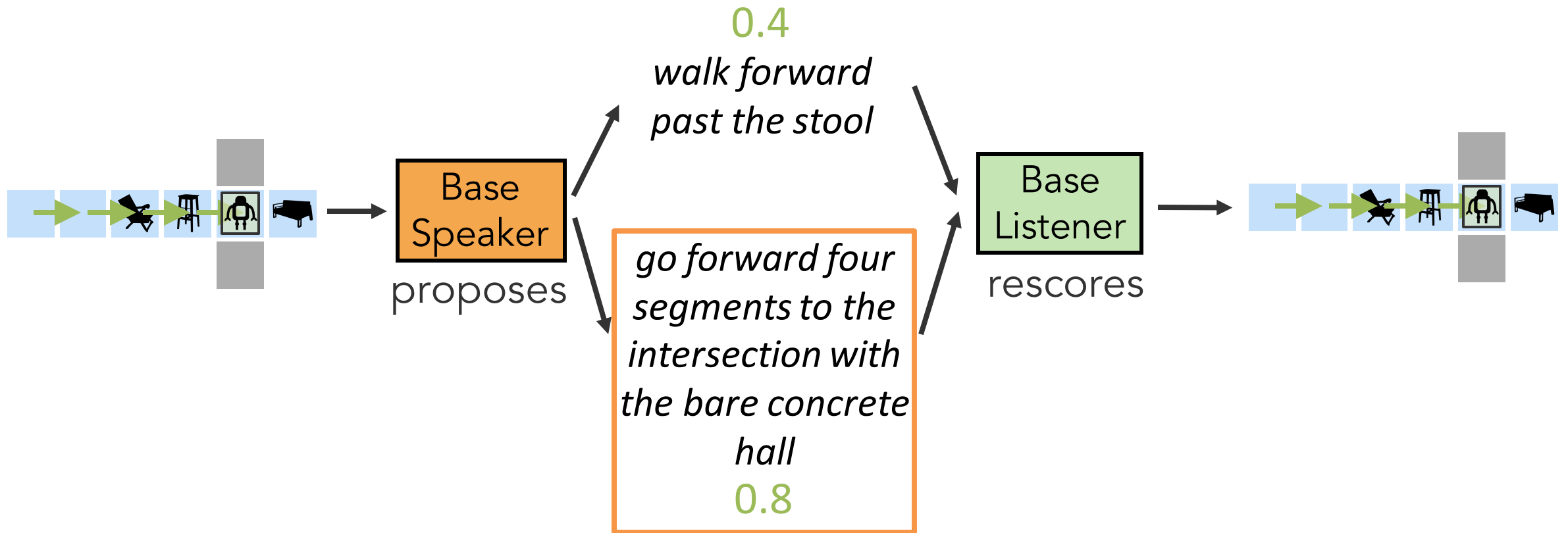


# Building a pragmatic speaker



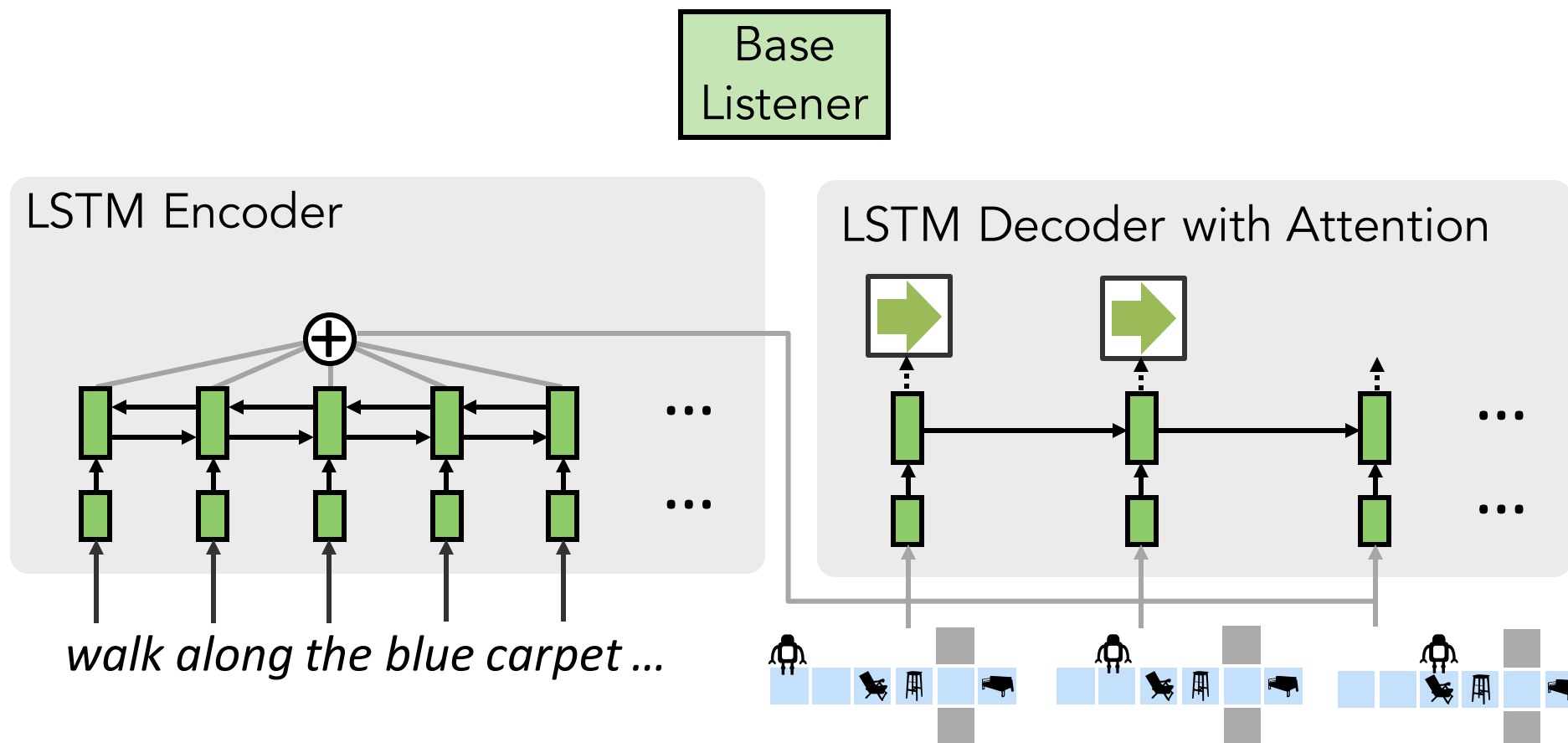


# Building a pragmatic speaker





# Base model implementations

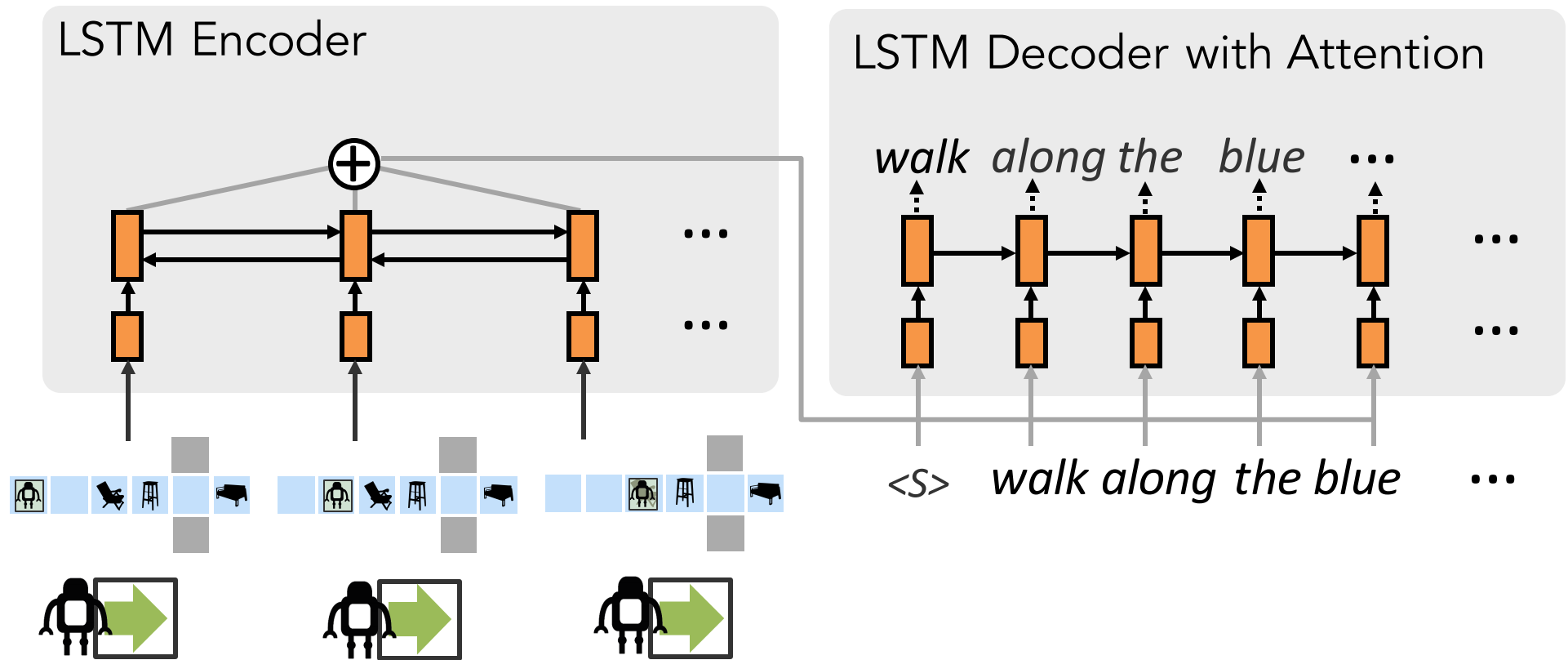






# Base model implementations

Base Speaker

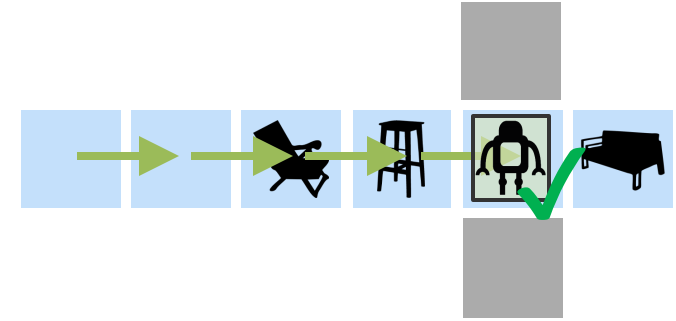




# Listener tasks

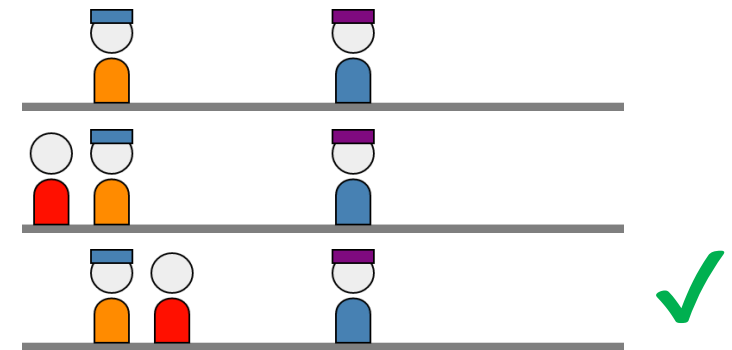
SAIL navigation [MacMahon et al., 2006; Chen and Mooney, 2011]

*walk along the  
blue carpet and  
you pass two  
objects*



Sequential Context-dependent Execution (SCONE) [Long et al. 2016]

*1. a red guy appears on  
the far left  
2. then to orange's  
other side*





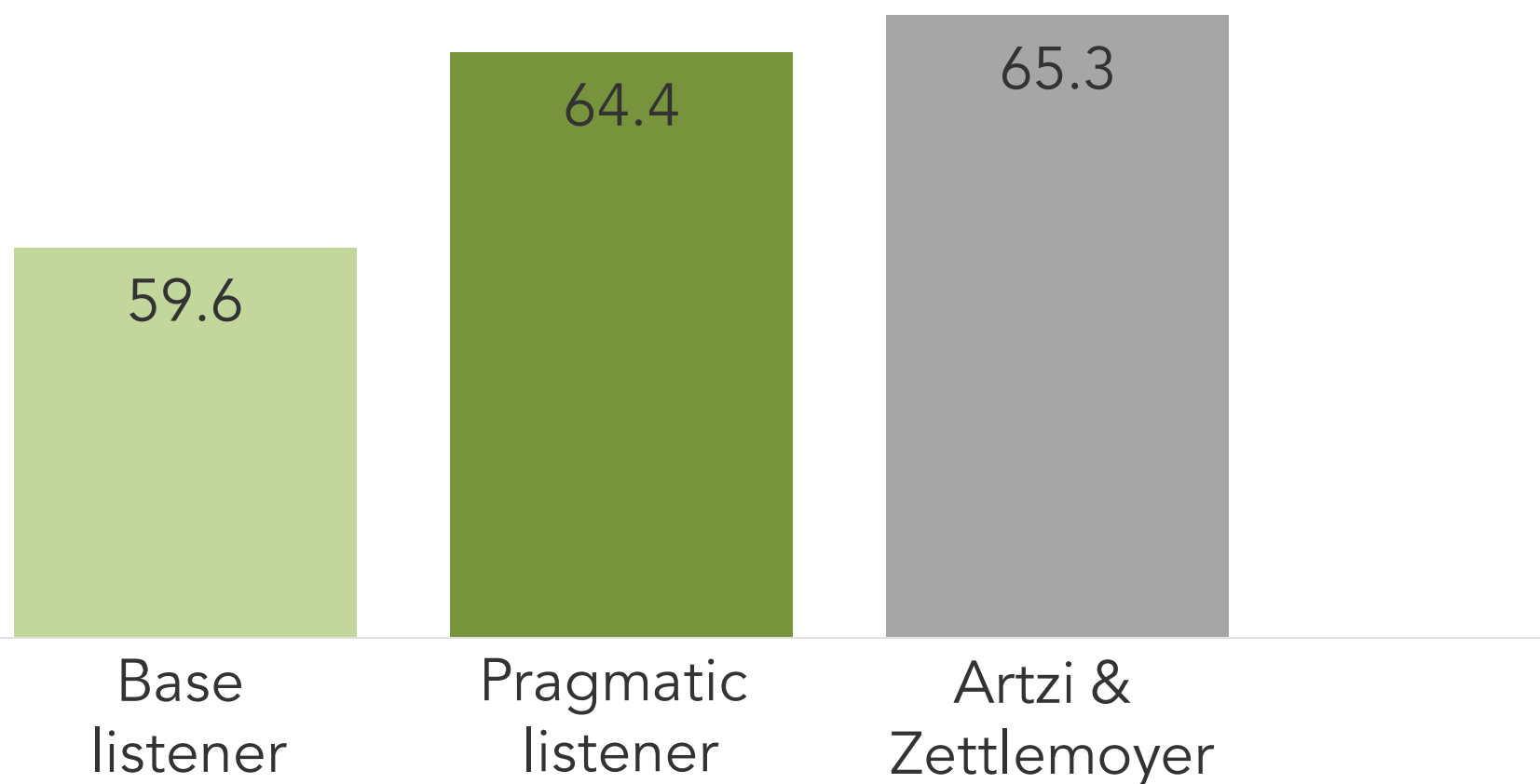
# Listener results, SAIL

Model accuracy at following human instructions

70

60

50



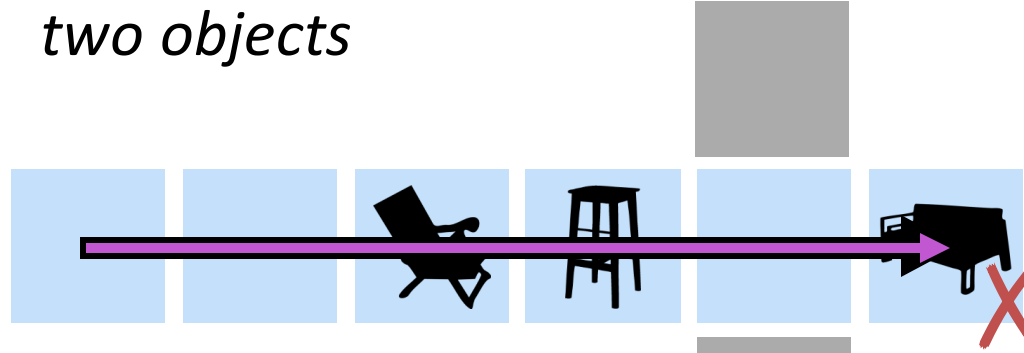


# Listener example, SAIL

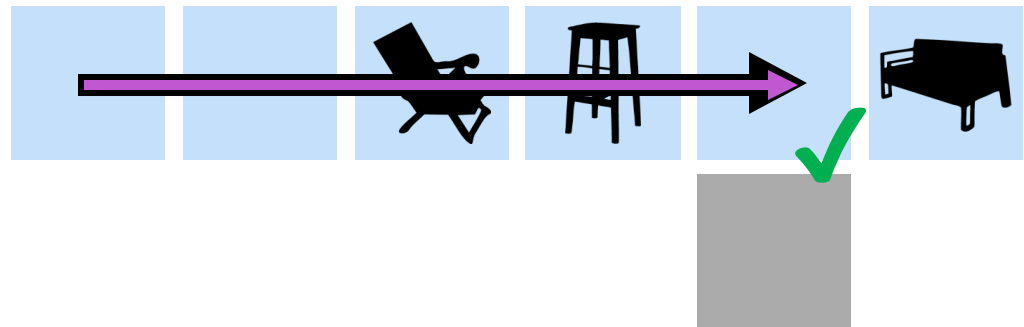
Instruction

*walk along the blue carpet and you pass two objects*

Base  
Listener



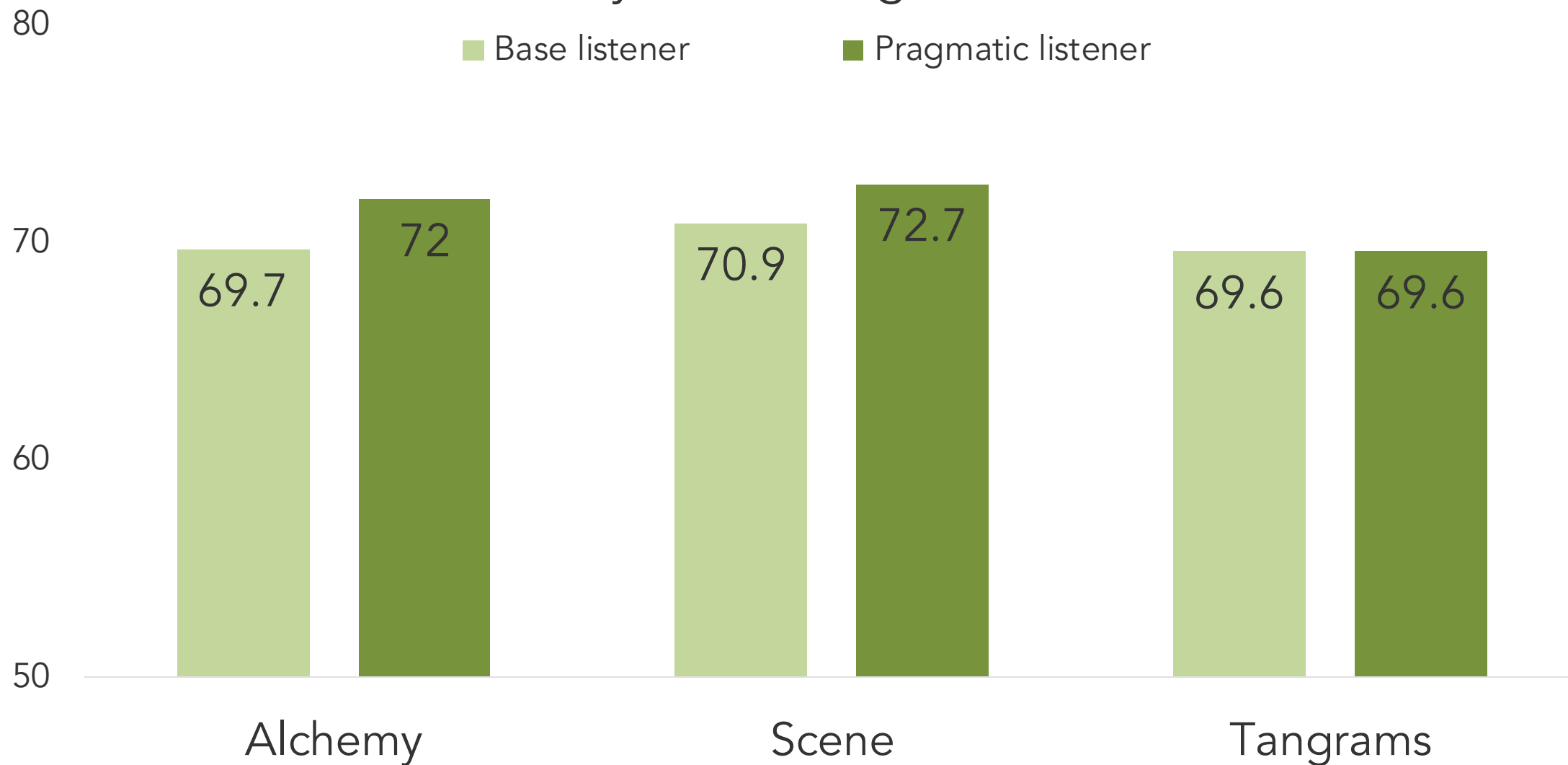
Pragmatic  
Listener





# Listener results, SCONE

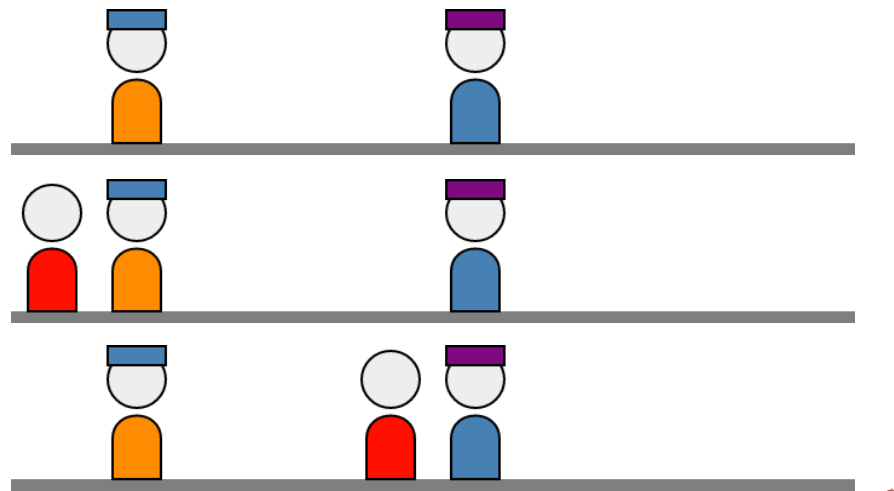
Model accuracy at following human instructions



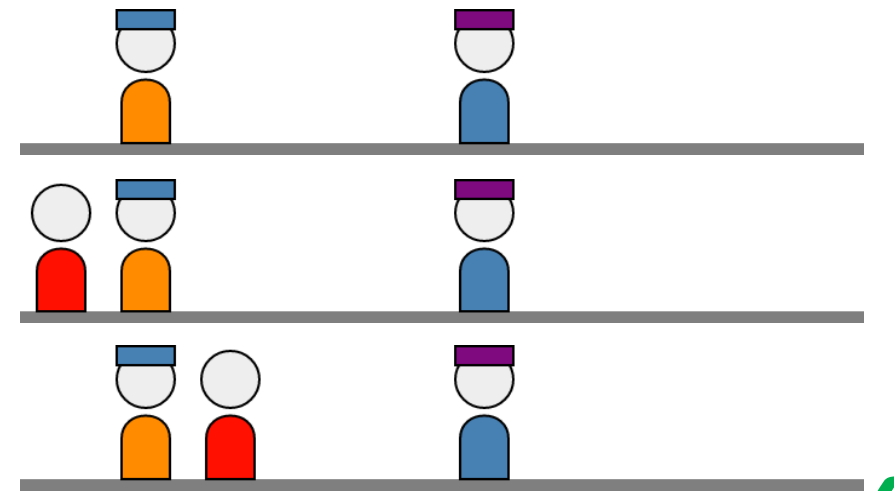


# Listener example, SCONE

Instruction *a red guy appears on the far left then to orange's other side*



Base Listener

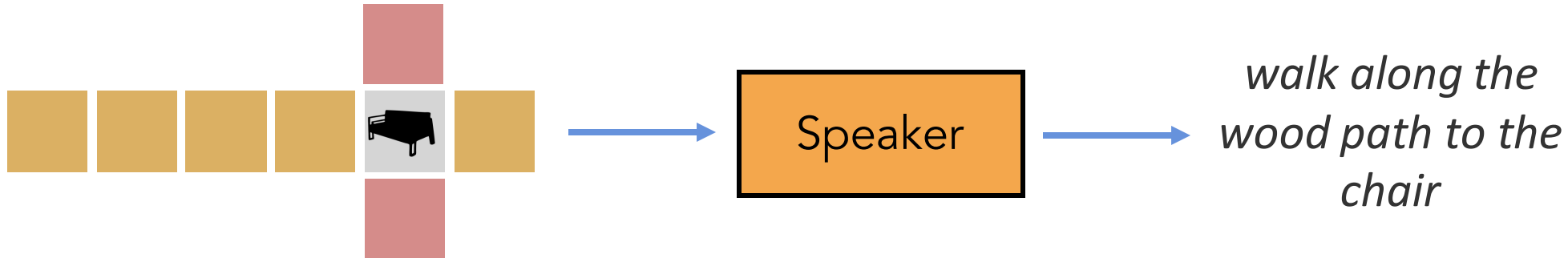


Pragmatic Listener



# Speaker task and evaluation

Speaker produces an instruction



Humans try to interpret it

*walk along the wood path to the chair*



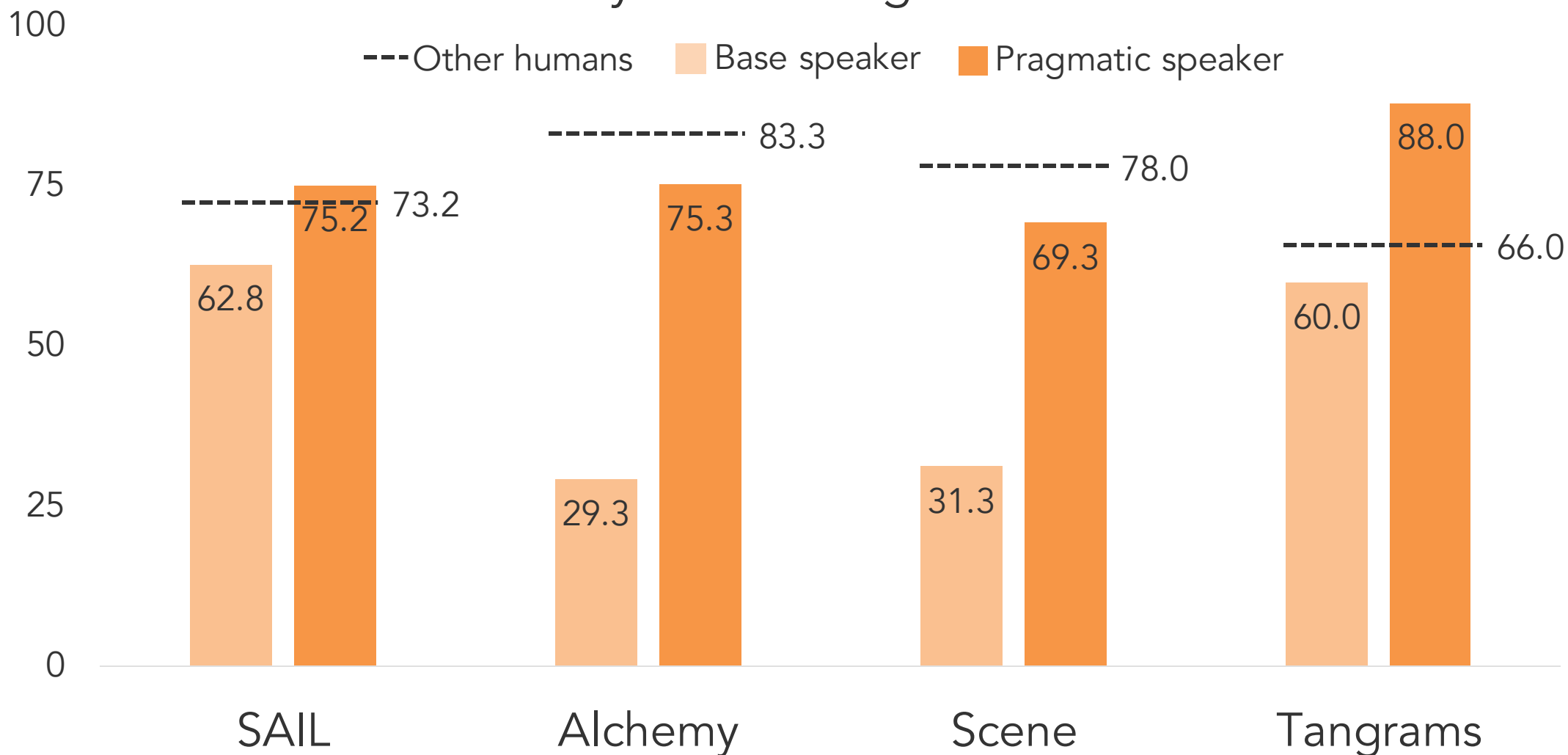
Human direction followers (MTurk)





# Speaker results

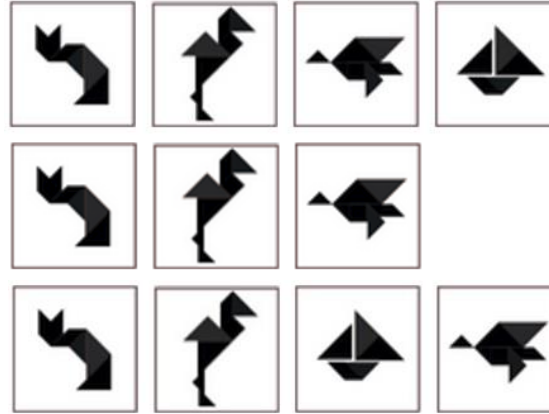
Human accuracy at following instructions from:







# Speaker example, SCONE



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*





# Real-world navigation

Matterport3D vision-and-language navigation dataset [Anderson et al., 2018]



## 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 .

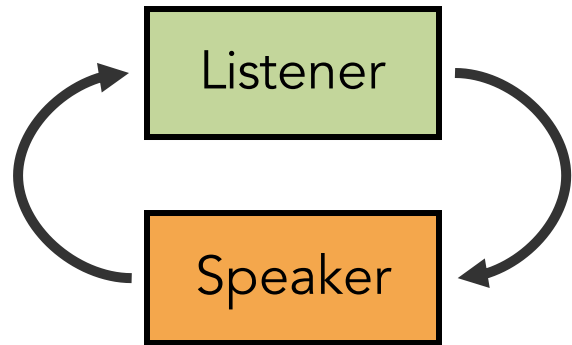
## pragmatic speaker:

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



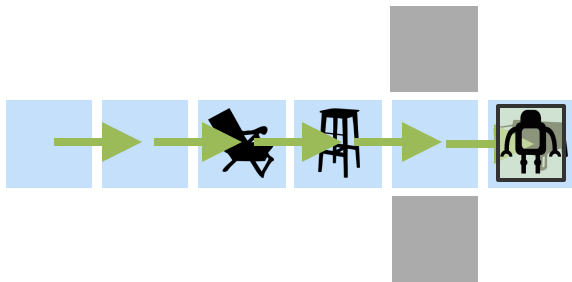
# Conclusions

---



Unified inference for sequential interpretation and generation

Reasoning counterfactually, and about likely interpretations



Pragmatics helps for complex tasks in structured domains

# Thanks!

<http://github.com/dpfried/pragmatic-instructions>