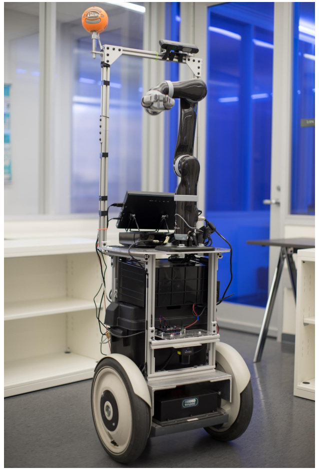
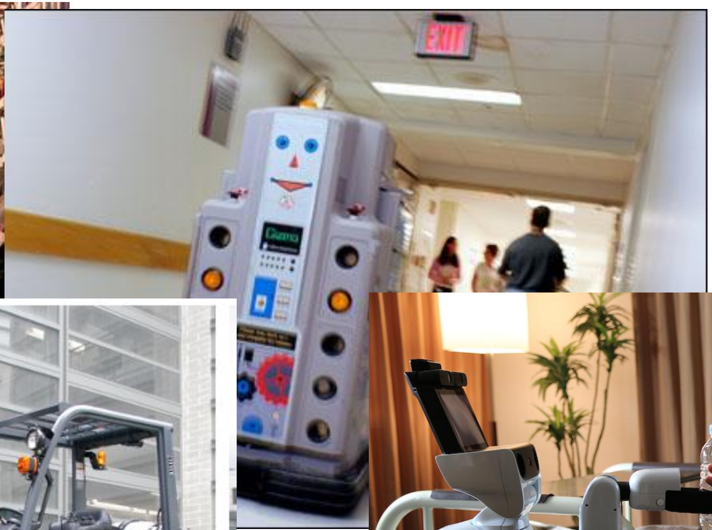
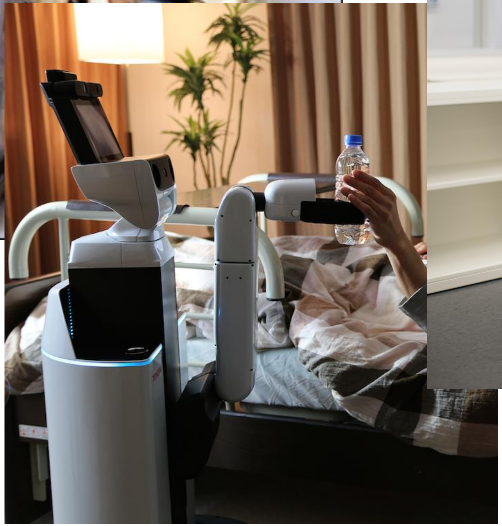
Continually Improving Grounded Natural Language Understanding through Human-Robot Dialog

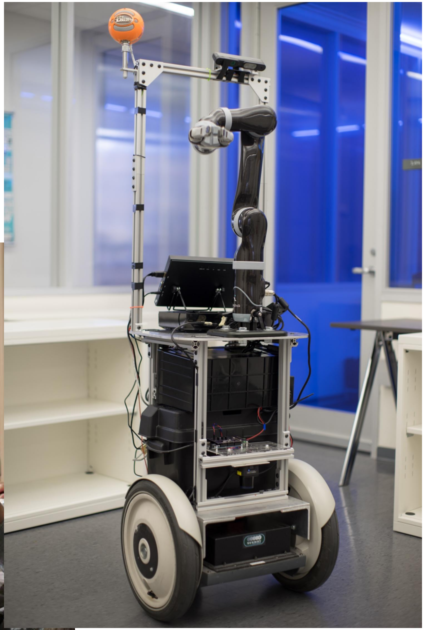
Jesse Thomason
University of Texas at Austin
Ph.D. Defense



Human-Robot Dialog



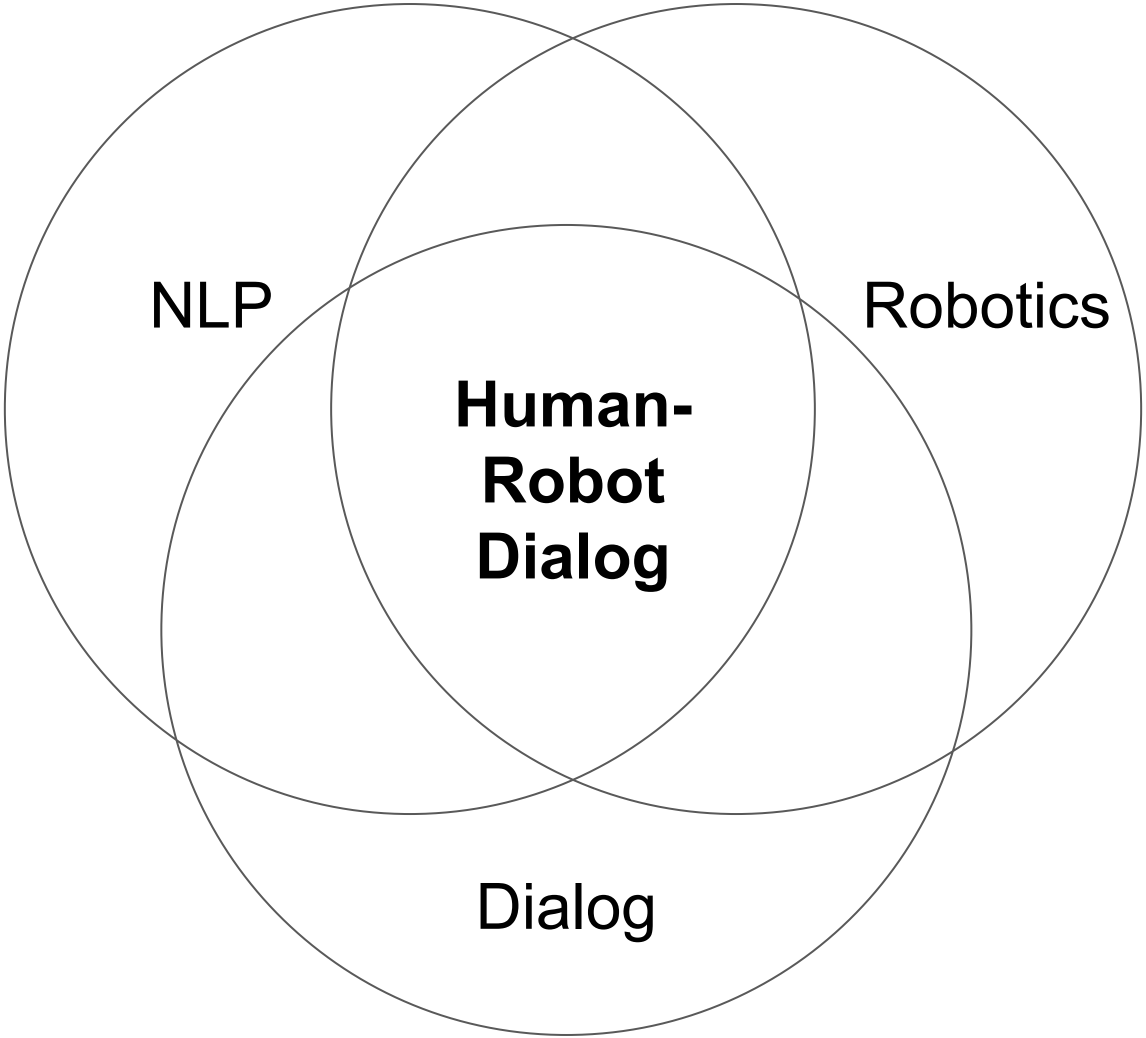


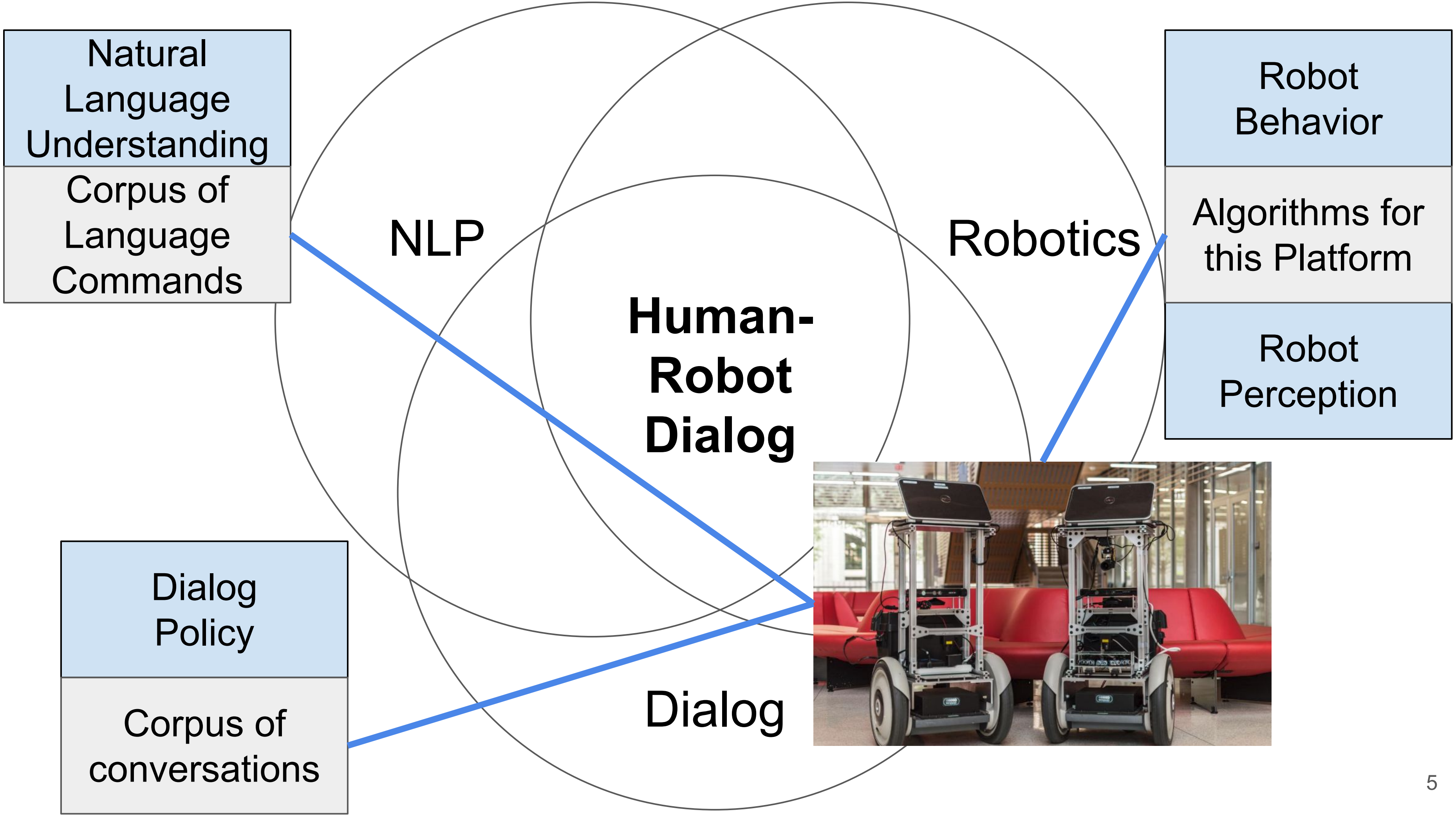


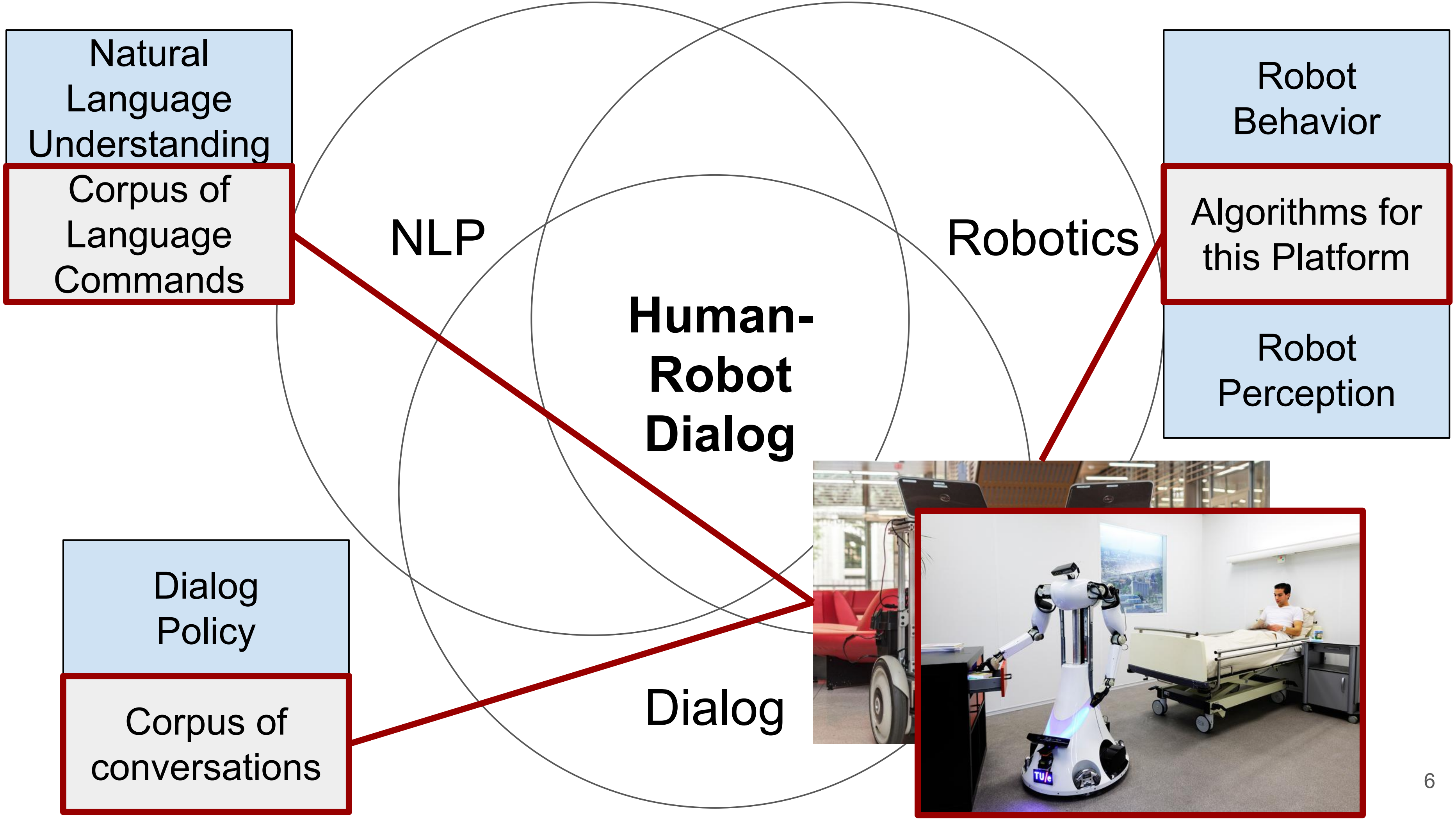
Human-Robot Dialog

“alert me if her heart rate decreases”
“bring me his chart”
“go and get the family”
“scalpel”

“text me when the speaker arrives”
“grab the empty, green bottle”
“lead him to alice’s office”
“get out of the way”

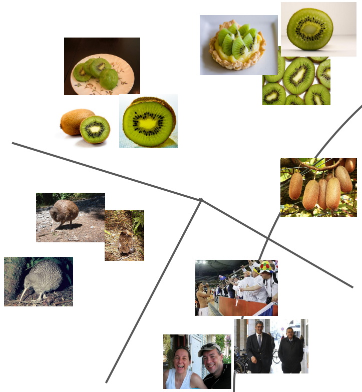

NLP
Robotics
Human-
Robot
Dialog
Dialog


Natural Language Understanding
Corpus of Language Commands
NLP
Robotics
Human-Robot Dialog
Dialog
Robot Behavior
Algorithms for this Platform
Robot Perception
Dialog Policy
Corpus of conversations


Natural Language Understanding
Corpus of Language Commands
NLP
Robotics
Robot Behavior
Algorithms for this Platform
Robot Perception
Human-Robot Dialog
Dialog Policy
Corpus of conversations
Dialog

Robot Dialog has Multiple Low-Resource Problems

- **My work**:
    - Develop algorithms for human-robot understanding that **overcome sparse training data.**
    - Use dialog to **correctly perform** user requests and **better understand** future requests.



NLP

Robotics

Dialog

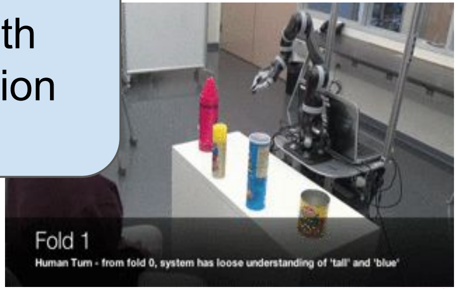
Polysemy Induction and Synonymy Detection (IJCAI'17)

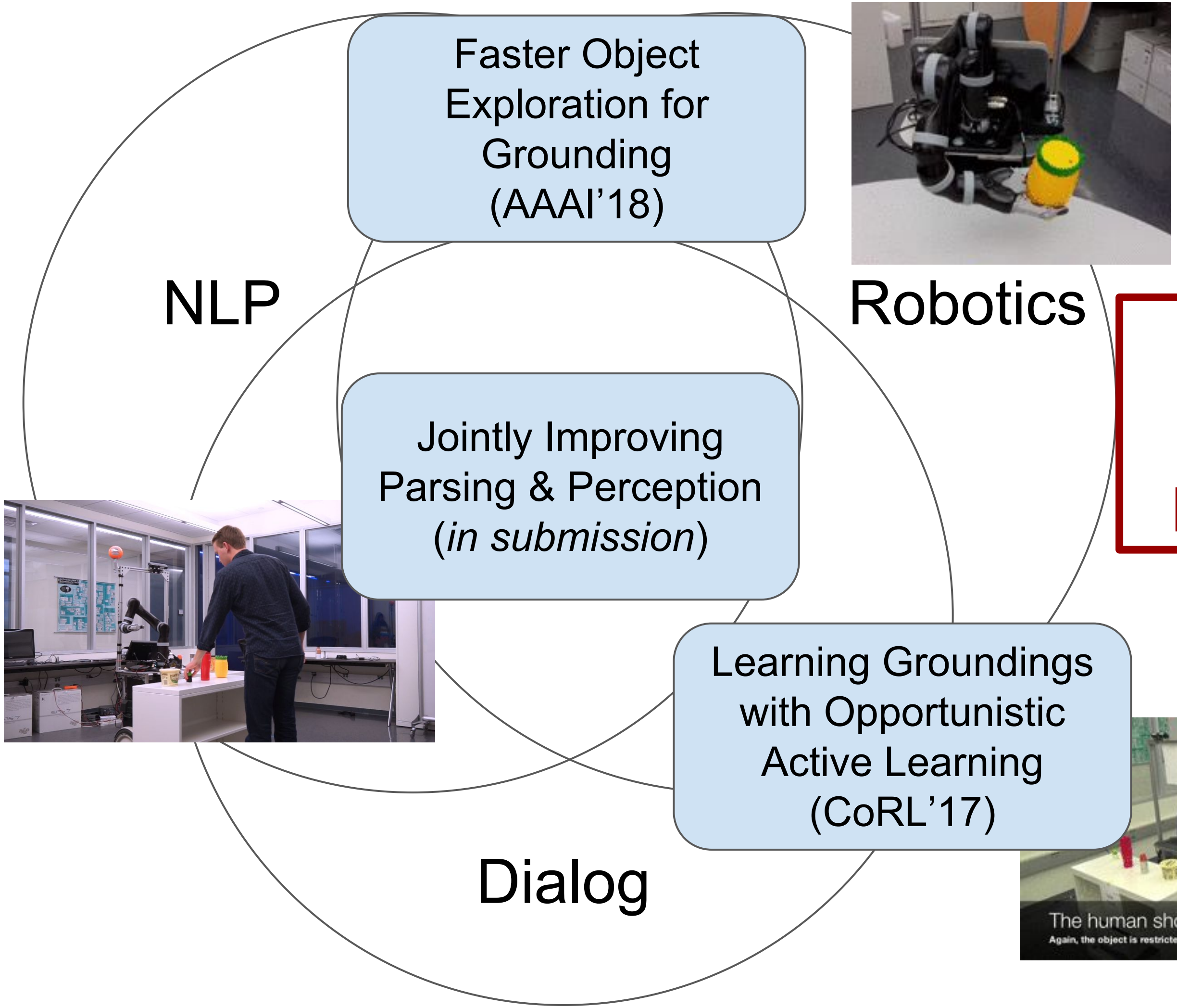
Human-Robot Dialog

Papers before proposal

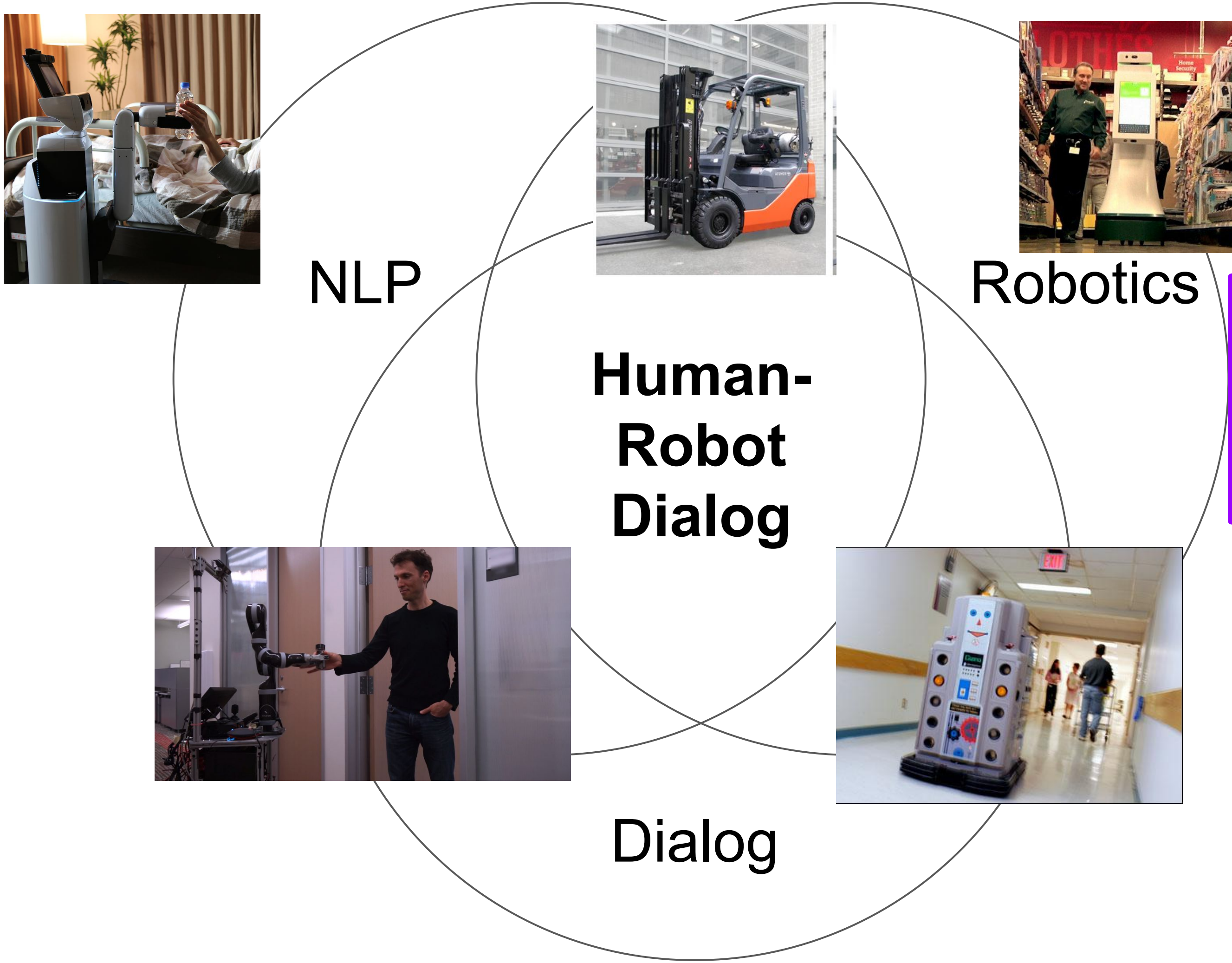
Improving Semantic Parsing through Dialog (IJCAI'15)

Learning Groundings with Human Interaction (IJCAI'16)



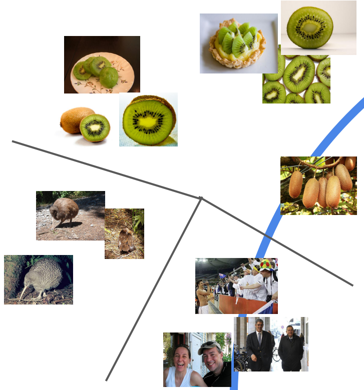


Papers since proposal



Next Directions

[Thomason et al., IJCAI'15]



NLP

Robotics

Dialog

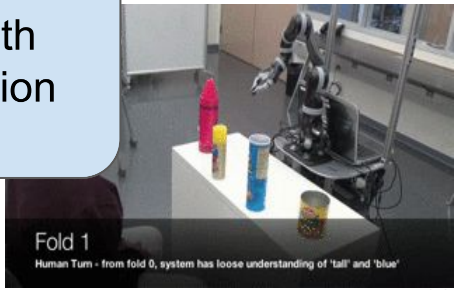
Polysemy Induction and Synonymy Detection (IJCAI'17)

Human-Robot Dialog

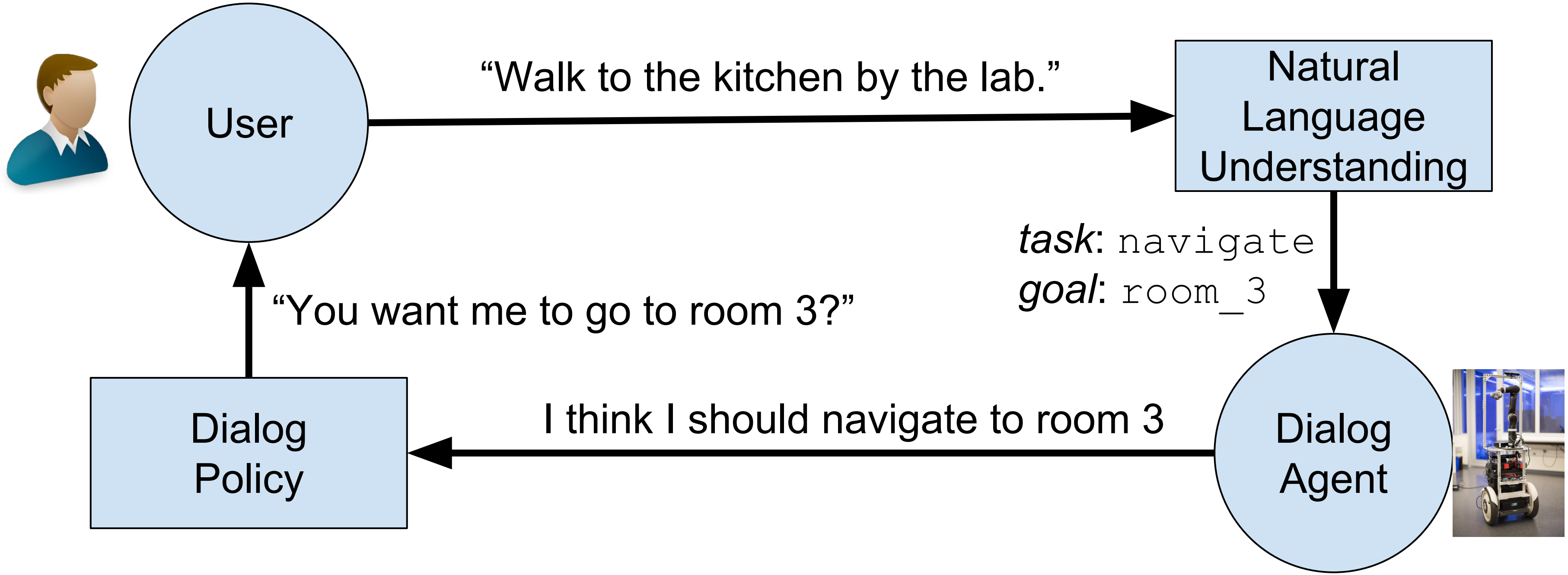
Improving Semantic Parsing through Dialog (IJCAI'15)

Learning Groundings with Human Interaction (IJCAI'16)

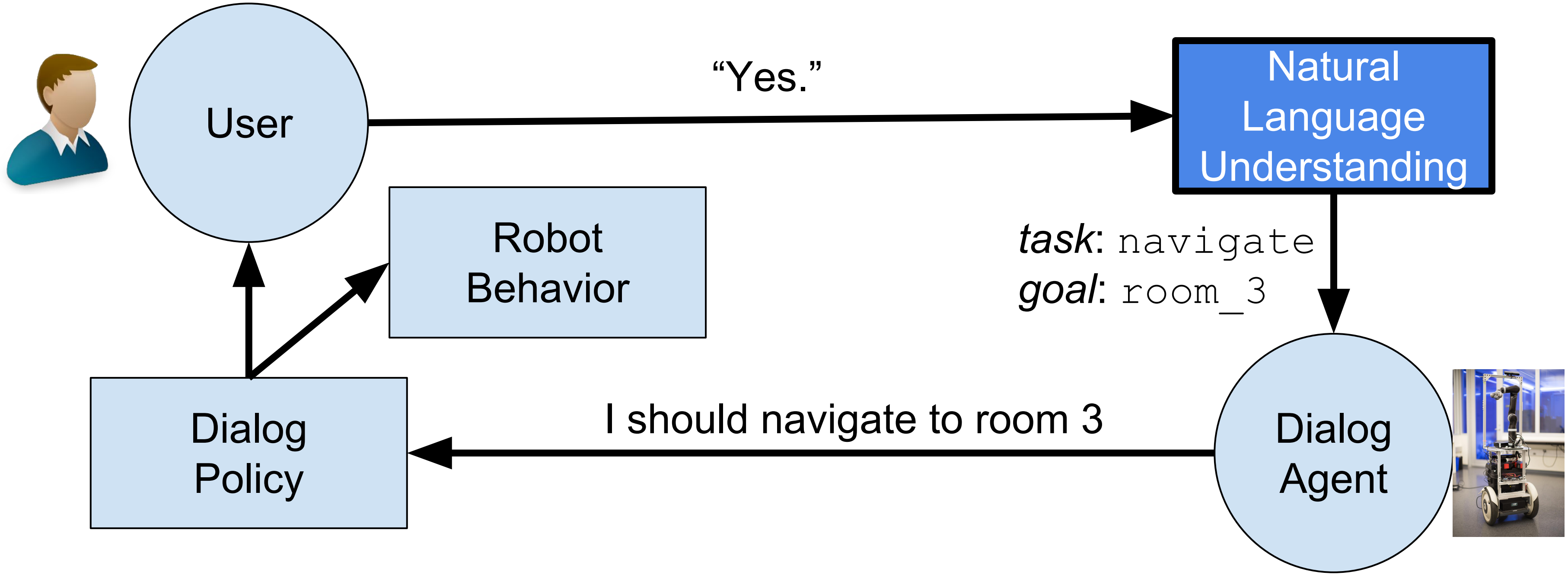
Papers before proposal



Dialog for Robots

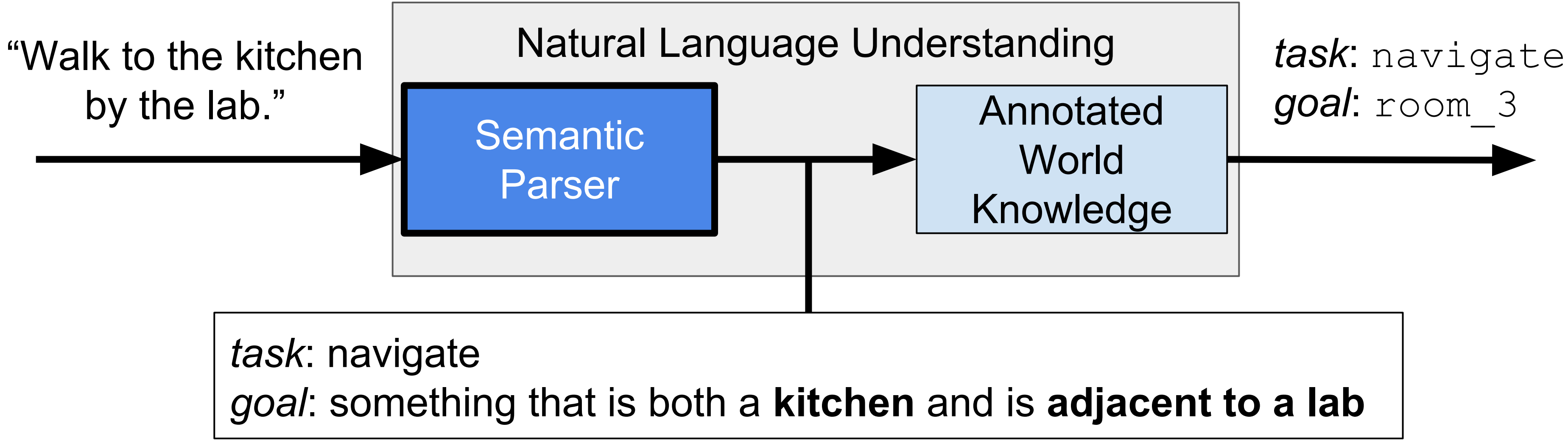


Dialog for Robots



[Thomason et al., IJCAI'15]

Natural Language Understanding



[Thomason et al., IJCAI'15]

Semantic Parser

- The parser can be initialized with low annotator effort.
  - In our experiments, we annotate **five** sentences.
  - Satisfies the low-resource constraints of human-robot dialog.
- But, **more training data** will improve performance.

[Thomason et al., IJCAI’15; Artzi and Zettlemoyer, EMNLP’11]

Inducing New Training Examples from Dialog

TASK TO COMPLETE

Dave Daniel wants the item in slot 5.

ROBOT	How can I help?
YOU	please bring the item in slot 5 to dave daniel
ROBOT	What should I bring to dave daniel?
YOU	a calander
ROBOT	I'm sorry, but I couldn't pinpoint what you meant by that.
ROBOT	What should I bring to dave daniel?
YOU	a day planner
ROBOT	I'm sorry, but I couldn't pinpoint what you meant by that.
ROBOT	What should I bring to dave daniel?
YOU	calendar
ROBOT	You want me to bring calendar to dave daniel?
YOU	yes
ROBOT	I thought so
ROBOT	Happy to help
YOU	

Final Task

Items available to robot:

Inducing New Training Examples from Dialog

TASK TO COMPLETE

Dave Daniel wants the item in slot 5.

ROBOT	How can I help?
YOU	please bring the item in slot 5 to dave daniel
ROBOT	What should I bring to dave daniel?
YOU	a calander
ROBOT	I'm sorry, but I couldn't pinpoint what you meant by that.
ROBOT	What should I bring to dave daniel?
YOU	a day planner
ROBOT	I'm sorry, but I couldn't pinpoint what you meant by that.
ROBOT	What should I bring to dave daniel?
YOU	calendar
ROBOT	You want me to bring calendar to dave daniel?
YOU	yes
ROBOT	I thought so
ROBOT	Happy to help
YOU	

Final Task

Items available to robot:

[Thomason et al., IJCAI'15]

Inducing New Training Examples from Dialog

TASK TO COMPLETE

Dave Daniel wants the item in slot 5.

ROBOT	How can I help?
YOU	please bring the item in slot 5 to dave daniel
ROBOT	What should I bring to dave daniel?
YOU	a calander
ROBOT	I'm sorry, but I couldn't pinpoint what you meant by that.
ROBOT	What should I bring to dave daniel?
YOU	a day planner
ROBOT	I'm sorry, but I couldn't pinpoint what you meant by that.
ROBOT	What should I bring to dave daniel?
YOU	calendar
ROBOT	You want me to bring calendar to dave daniel?
YOU	yes
ROBOT	I thought so
ROBOT	Happy to help
YOU	

Final Task

Items available to robot:

[Thomason et al., IJCAI’15]

Inducing New Training Examples from Dialog

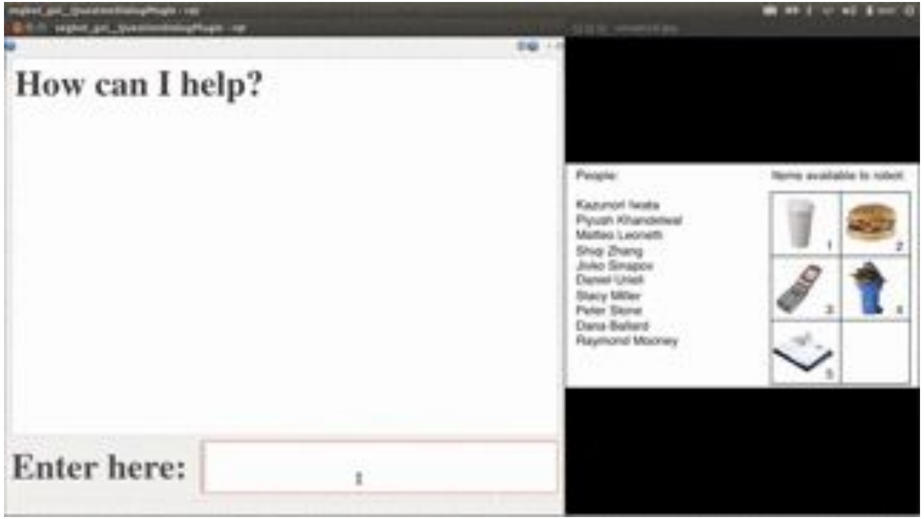
Induced Training Pairs

“please bring the item in slot five to dave daniel” `bring(calender, dave)`
“calander” `calendar`
“a day planner” `calendar`

Semantic Parser

[Thomason et al., IJCAI’15]

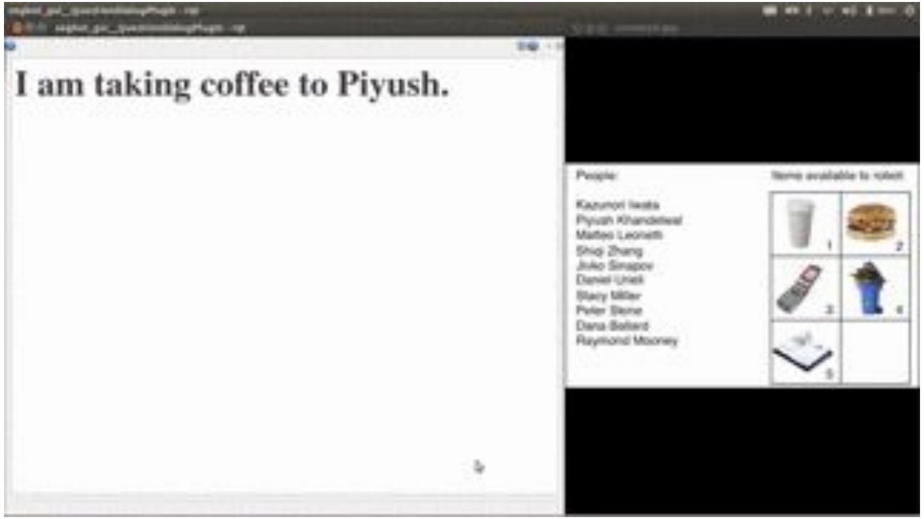
Demonstration



[Thomason et al., IJCAI'15]

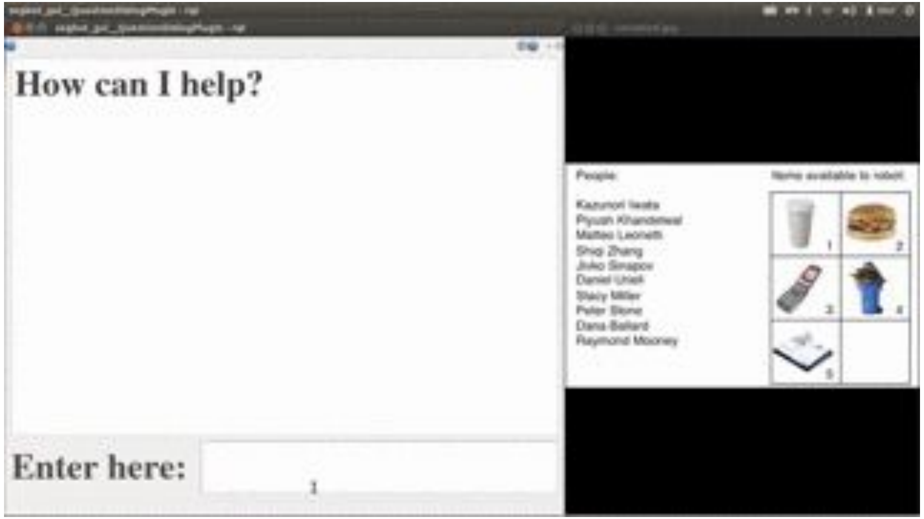
Demonstration

Demonstration



[Thomason et al., IJCAI’15]

Demonstration



[Thomason et al., IJCAI'15]

Dialogs that Clarify Meaning and Provide Supervision

Agent Belief (*task, goal, item, person*)	**Request**	**Question**
`(?, ?, ?, ?)`	*all*	"How can I help?" / "Can you reword your original request?"
`(navigate, ?, _, _)`	*goal*	"Where should I walk?"
`(deliver, _, ?, p)`	*item*	"What should I bring to `p`?"
`(navigate, r, _, _)`	*confirm*	"You want me to walk to `r`?"
...	...	...

[Thomason et al., IJCAI'15]

Dialogs that Clarify Meaning and Provide Supervision

Agent Belief (*task, goal, item, person*)	**Request**	**Question**
`(?, ?, ?, ?)`	*all*	"How can I help?" / "Can you reword your original request?"
`(navigate, ?, _, _)`	*goal*	"Where should I walk?"
`(deliver, _, ?, p)`	*item*	"What should I bring to `p`?"
`(navigate, r, _, _)`	*confirm*	"You want me to walk to `r`?"
...	...	...

[Thomason et al., IJCAI'15]

Dialogs that Clarify Meaning and Provide Supervision

TASK TO COMPLETE

Dave Daniel wants the item in slot 5.

ROBOT	How can I help?
YOU	please bring the item in slot 5 to dave daniel
ROBOT	What should I bring to dave daniel?
YOU	a calander
ROBOT	I'm sorry, but I couldn't pinpoint what you meant by that.
ROBOT	What should I bring to dave daniel?
YOU	a day planner
ROBOT	I'm sorry, but I couldn't pinpoint what you meant by that.
ROBOT	What should I bring to dave daniel?
YOU	calendar
ROBOT	You want me to bring calendar to dave daniel?
YOU	yes
ROBOT	I thought so
ROBOT	Happy to help
YOU	

Final Task

Expect whole command

Expect *item*

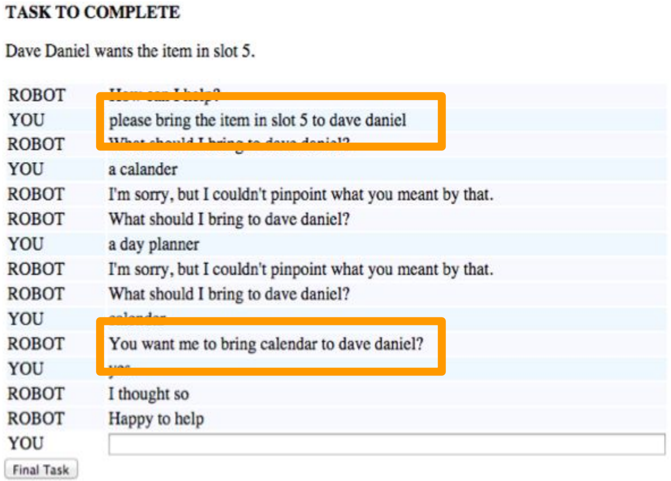
Expect *item*

Expect *item*

task: `deliver`
item: `calendar`
person: `dave_daniel`

Technical Contributions

- Design a dialog policy that allows us to **pair human language with latent meaning representations**.



- Improve semantic parsing **given very little initial in-domain data**.

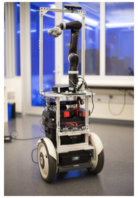
Experiments via Amazon Mechanical Turk

TASK TO COMPLETE

Dave Daniel wants the item in slot 5.

ROBOT	How can I help?
YOU	please bring the item in slot 5 to dave daniel
ROBOT	What should I bring to dave daniel?
YOU	a calander
ROBOT	I'm sorry, but I couldn't pinpoint what you meant by that.
ROBOT	What should I bring to dave daniel?
YOU	a day planner
ROBOT	I'm sorry, but I couldn't pinpoint what you meant by that.
ROBOT	What should I bring to dave daniel?
YOU	calendar
ROBOT	You want me to bring calendar to dave daniel?
YOU	yes
ROBOT	I thought so
ROBOT	Happy to help
YOU	

Final Task



x 50

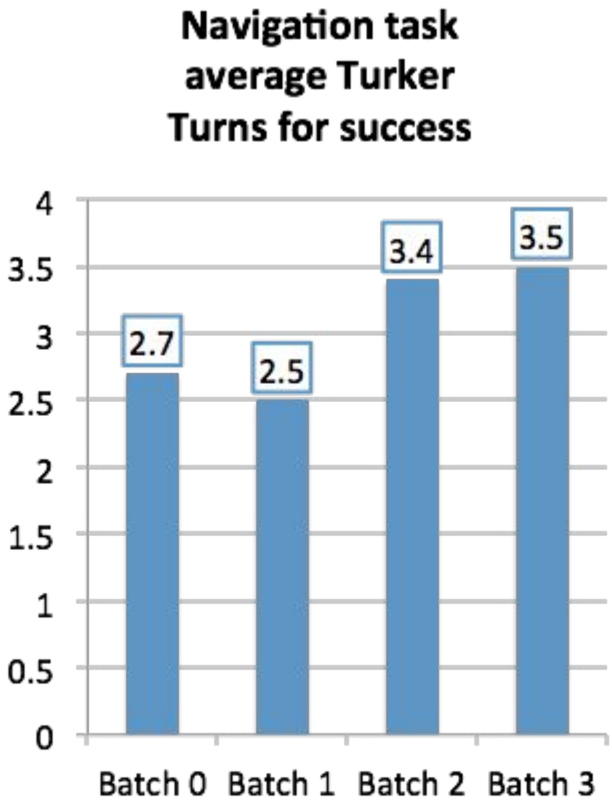
Induced Training Pairs

Semantic Parser

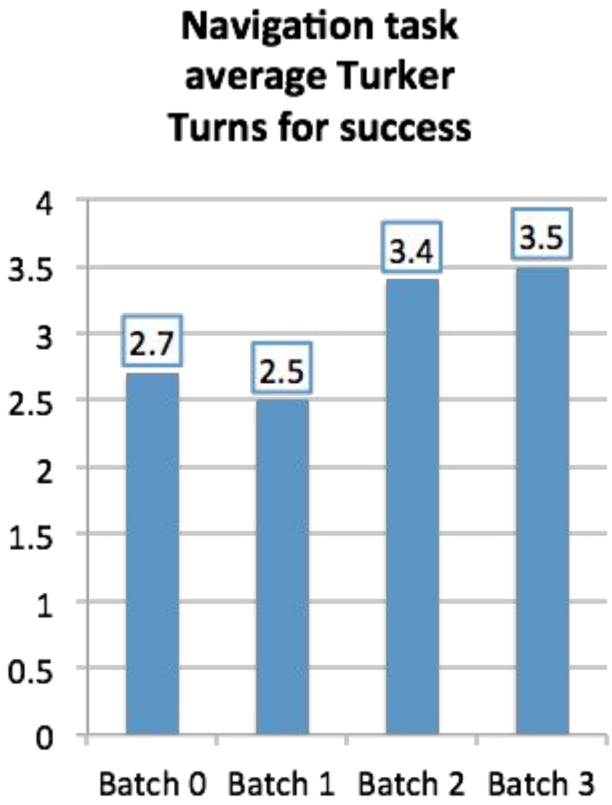
x 4

[Thomason et al., IJCAI'15]

Navigation Dialog Turns



Navigation Dialog Turns



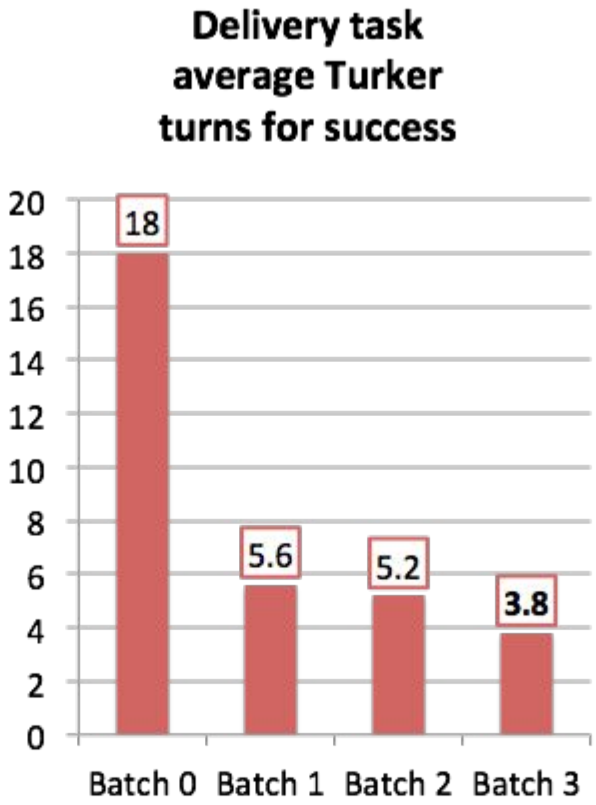
Robot: How can I help?
Human: go
…
Human: go to dave daniel's office

Induced Training Pairs

"go" `go(room_2)`
…

[Thomason et al., IJCAI'15]

Delivery Dialog Turns



- Statistically significant decrease.
- More arguments:

  harder to understand, so more to

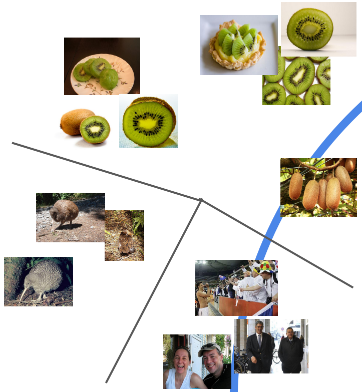
  gain from parser training.

Qualitative: One user wrote "the robot even fixed my typo when I mispelled calendar!"

Other Findings

- Users rate system more understanding and less frustrating.
- Results replicable on physical platform.

[Thomason et al., IJCAI’15]



Polysemy Induction and Synonymy Detection (IJCAI’17)

NLP

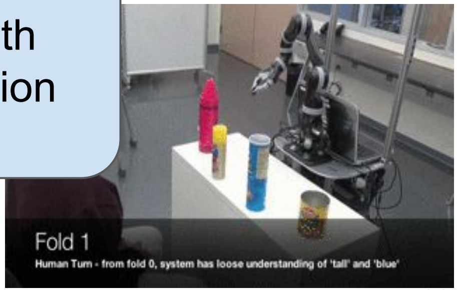
Robotics

Human-Robot Dialog

Papers before proposal

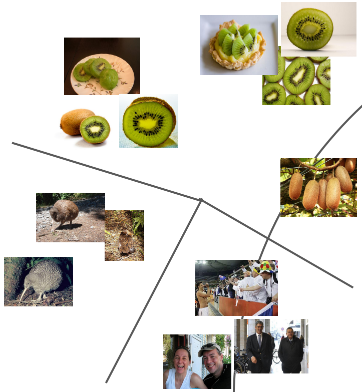
Improving Semantic Parsing through Dialog (IJCAI’15)

Learning Groundings with Human Interaction (IJCAI’16)



Dialog

[Thomason et al., IJCAI’16]



NLP

Polysemy Induction and Synonymy Detection (IJCAI’17)

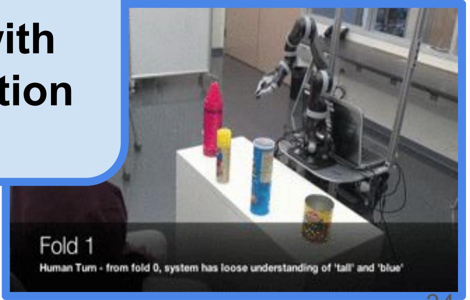
Robotics

Papers before proposal

Human-Robot Dialog

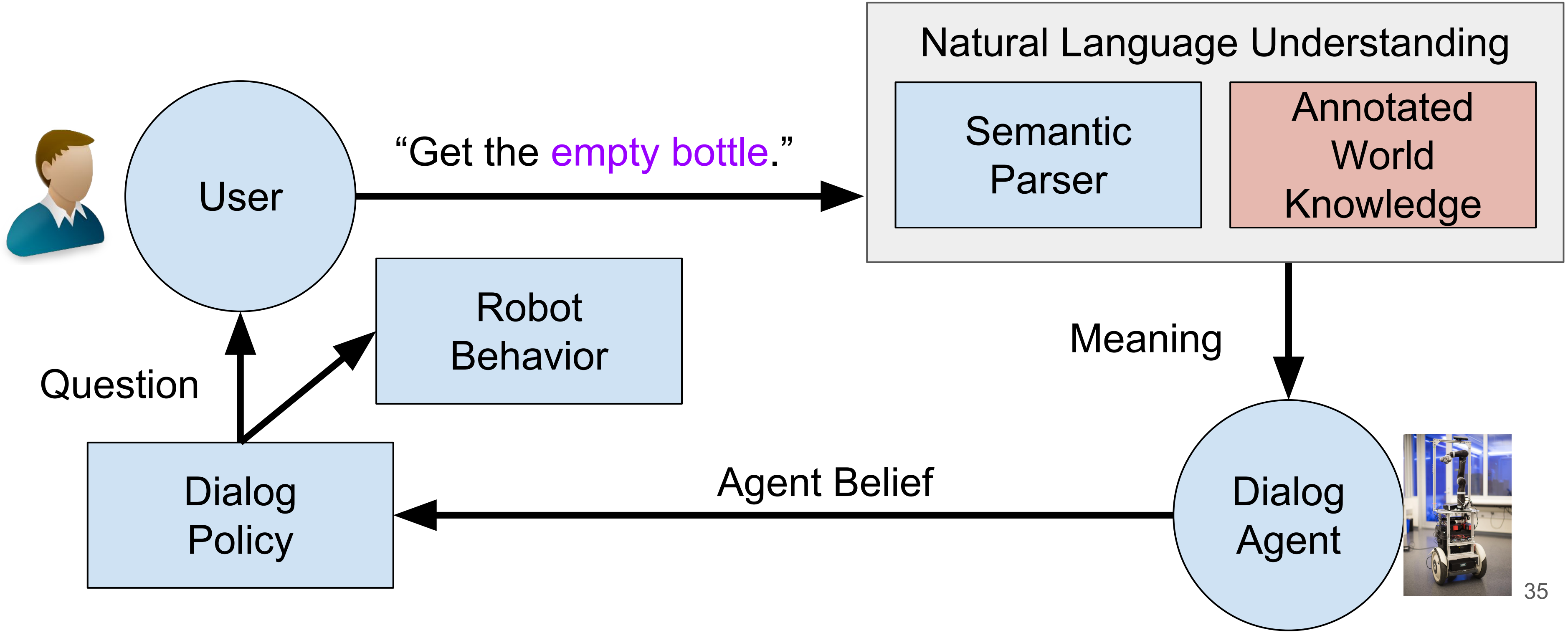
Improving Semantic Parsing through Dialog (IJCAI’15)

Learning Groundings with Human Interaction (IJCAI’16)

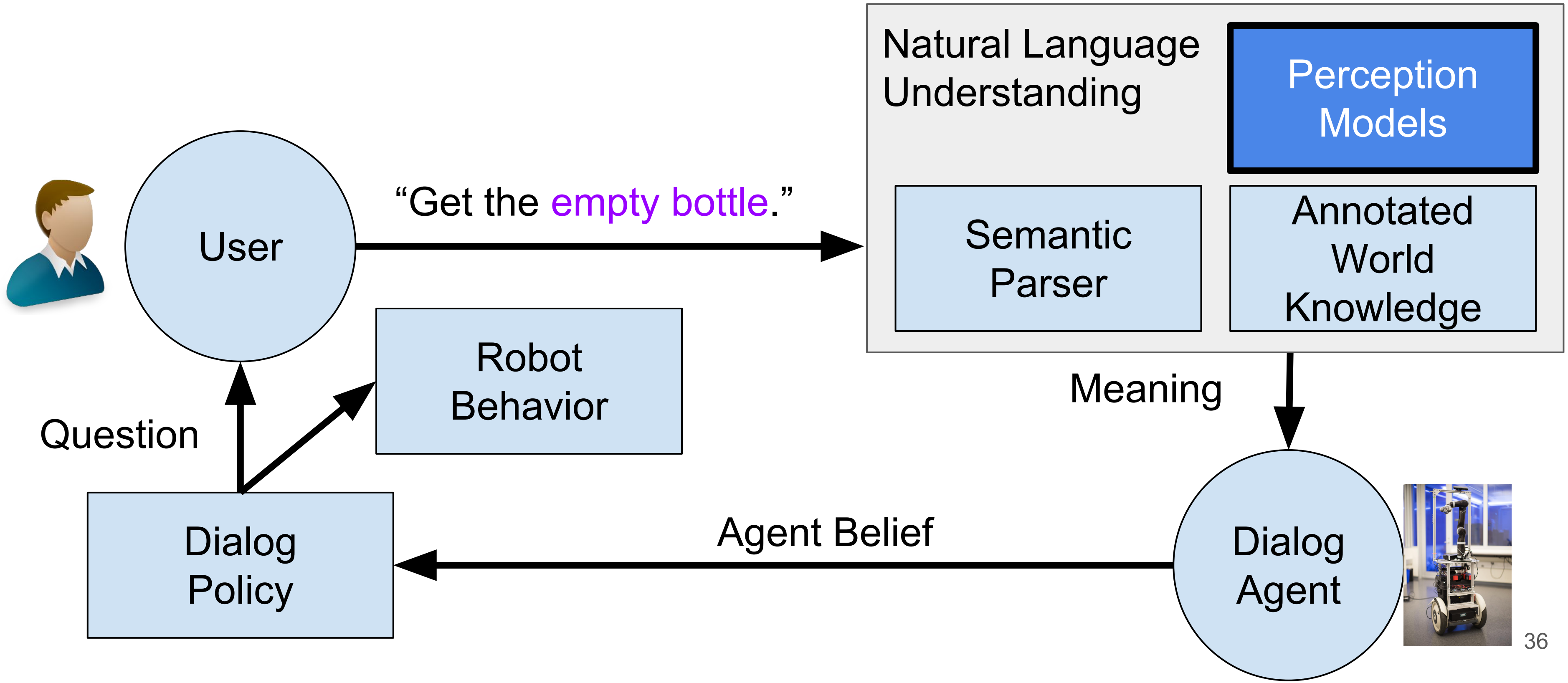


Dialog

We do not yet handle perception information



We need to perform *language grounding*



Language Grounding

`empty`?

Perception Models

yes

[Harnad, Physica D’90]

Language Grounding

- *Symbol grounding problem.*
- Historically use visual space.
- We use more than vision.

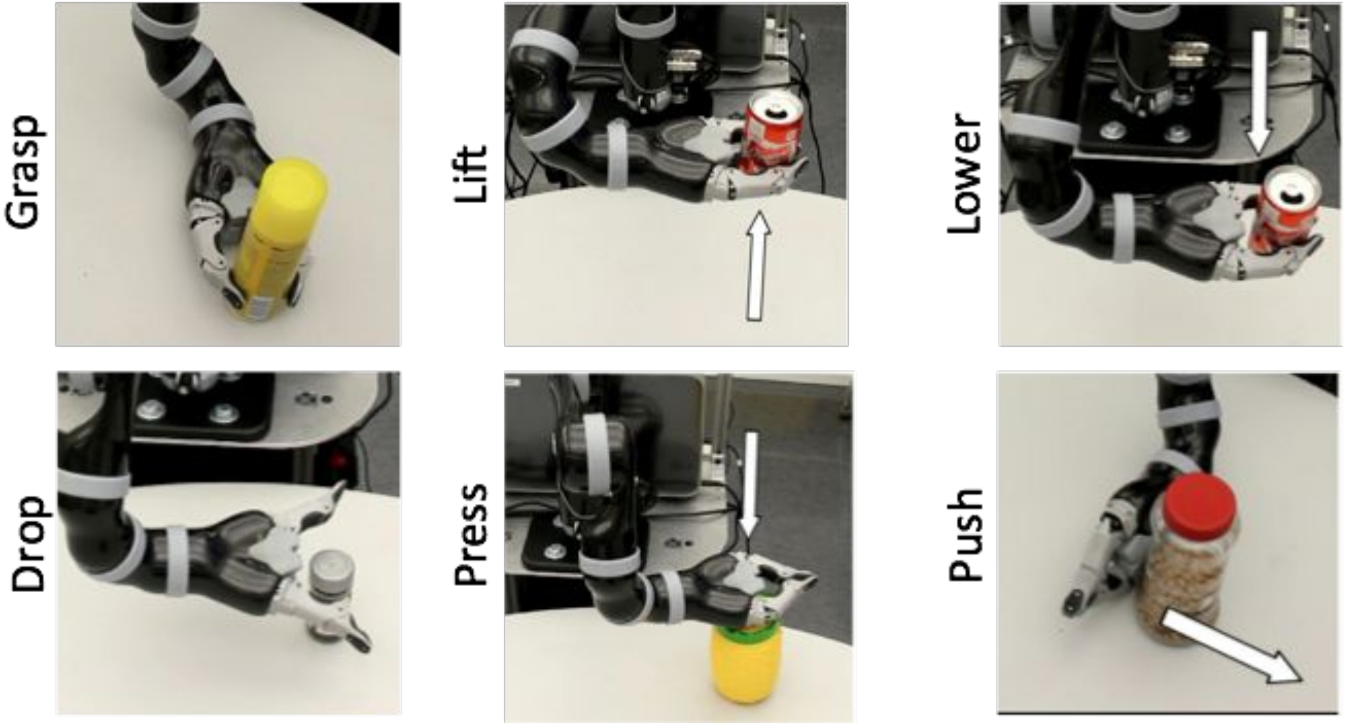
Language Grounding

Haptic sensors from arm give force information.

Audio signals from mic give sound information.

[Sinapov et al., IJCAI’16; Thomason et al., IJCAI’16; Simonyan and Zisserman, CoRR’14]

Perceptual Grounding



Look

color, ***shape***, and deep ***VGG*** features.

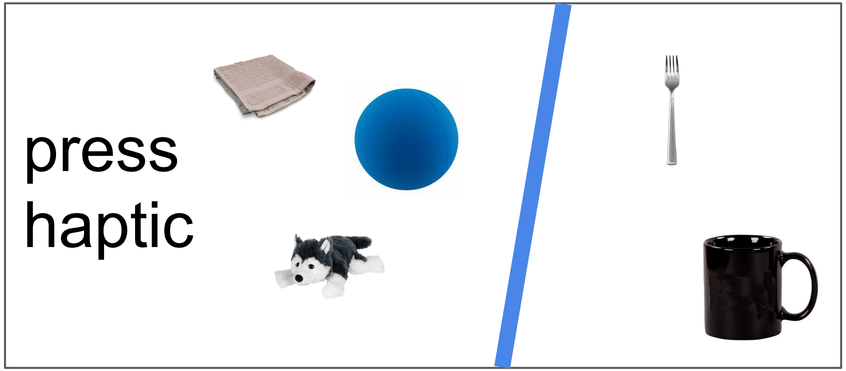
[Thomason et al., IJCAI’16]

Building Perceptual Classifiers

$G_{p,c}(o)$	SVM trained for predicate *p* and sensorimotor context *c* result on object *o*

p: squishy

c:



Few labeled examples, but SVMs can operate on this sparse data.

[Thomason et al., IJCAI'16]

Building Perceptual Classifiers

$\mathbf{G}_{p,c}(o)$	SVM trained for predicate *p* and sensorimotor context *c* result on object *o*

$$\boxed{d(p,o)} = sgn\left(\sum_{c \in C} w_{p,c}\mathbf{G}_{p,c}(o)\right)$$

Decision

[Thomason et al., IJCAI'16]

Building Perceptual Classifiers

$\mathbf{G}_{p,c}(o)$	SVM trained for predicate *p* and sensorimotor context *c* result on object *o*

$$\boxed{d(p,o)} = sgn\left(\sum_{\boxed{c \in C}} w_{p,c}\mathbf{G}_{p,c}(o)\right)$$

Decision

Sensorimotor Contexts

[Thomason et al., IJCAI'16]

Building Perceptual Classifiers

$\mathrm{G}_{p,c}(o)$	SVM trained for predicate *p* and sensorimotor context *c* result on object *o*

$$\boxed{d(p, o)} = sgn\left(\sum_{\boxed{c \in C}} w_{p,c} \boxed{\mathrm{G}_{p,c}(o)}\right)$$

Decision

Sensorimotor Contexts

Context SVM result

[Thomason et al., IJCAI'16]

Building Perceptual Classifiers

$\mathrm{G}_{p,c}(o)$	SVM trained for predicate *p* and sensorimotor context *c* result on object *o*

$$\boxed{d(p,o)} = sgn\left(\sum_{\boxed{c \in C}} \boxed{w_{p,c}}\boxed{\mathrm{G}_{p,c}(o)}\right)$$

Decision

Sensorimotor Contexts

Reliability Weight

Context SVM result

[Thomason et al., IJCAI'16]

Building Perceptual Classifiers

$\mathrm{G}_{p,c}(o)$	SVM trained for predicate *p* and sensorimotor context *c* result on object *o*

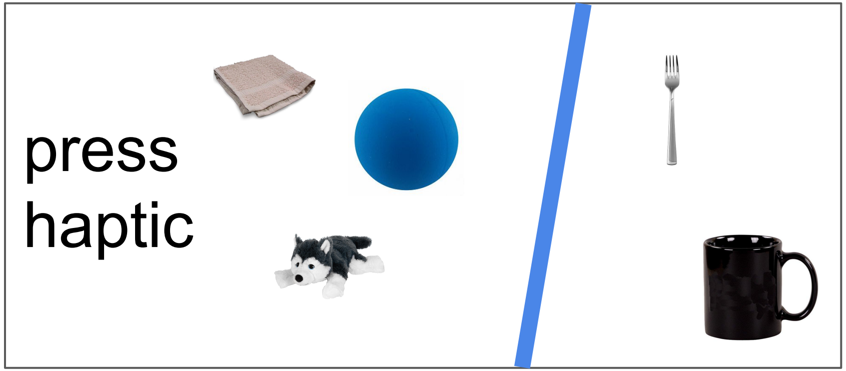
$$d(p,o) = sgn\left(\sum_{c\in C} \boxed{w_{p,c}} \mathrm{G}_{p,c}(o)\right)$$

Reliability weights estimated from xval

squishy	
sensorimotor context	$w_{p,c}$
press-haptics	0.5
grasp-haptics	0.3
...	...
look-VGG	0.01

Building Perceptual Classifiers

$G_{p,c}(o)$	SVM trained for predicate *p* and sensorimotor context *c* result on object *o*



Reliability weights estimated from xval

squishy	
sensorimotor context	$w_{p,c}$
press-haptics	0.5
grasp-haptics	0.3
...	...
look-VGG	0.01

[Thomason et al., IJCAI'16]

Building Perceptual Classifiers

$G_{p,c}(o)$	SVM trained for predicate *p* and sensorimotor context *c* result on object *o*

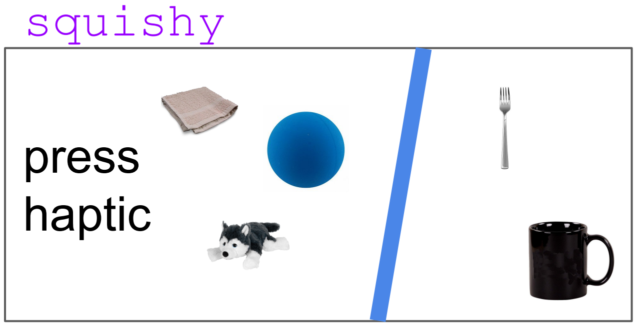
Reliability weights estimated from xval

squishy	
sensorimotor context	$w_{p,c}$
press-haptics	0.5
grasp-haptics	0.3
...	...
look-VGG	0.01

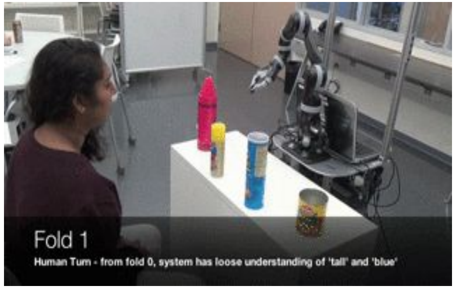
[Thomason et al., IJCAI'16]

Technical Contributions

- Ensemble SVMs over **multi-modal object features** to perform **language grounding**.



- Get language labels from natural **language game** with human users



[Thomason et al., IJCAI'16]

[Thomason et al., IJCAI'16]

Experiments Playing *I Spy*

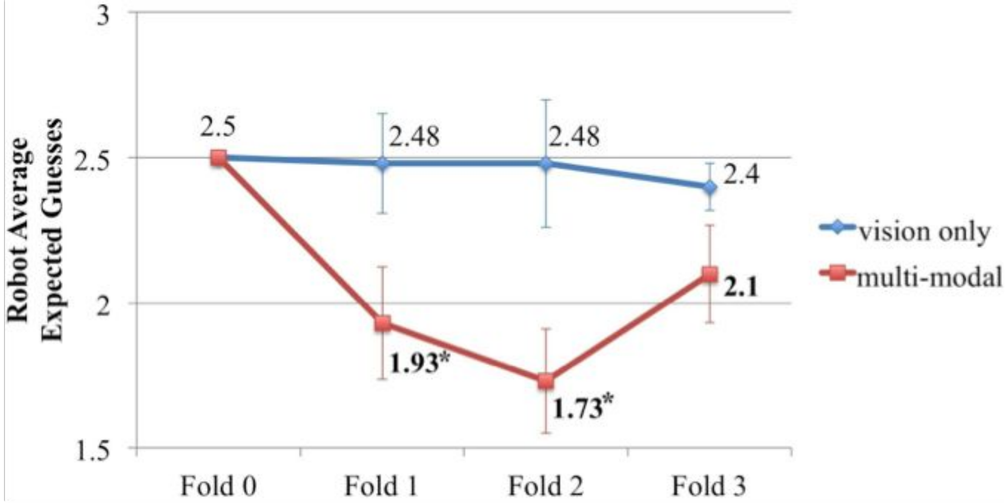
multi-modal

vs

vision only

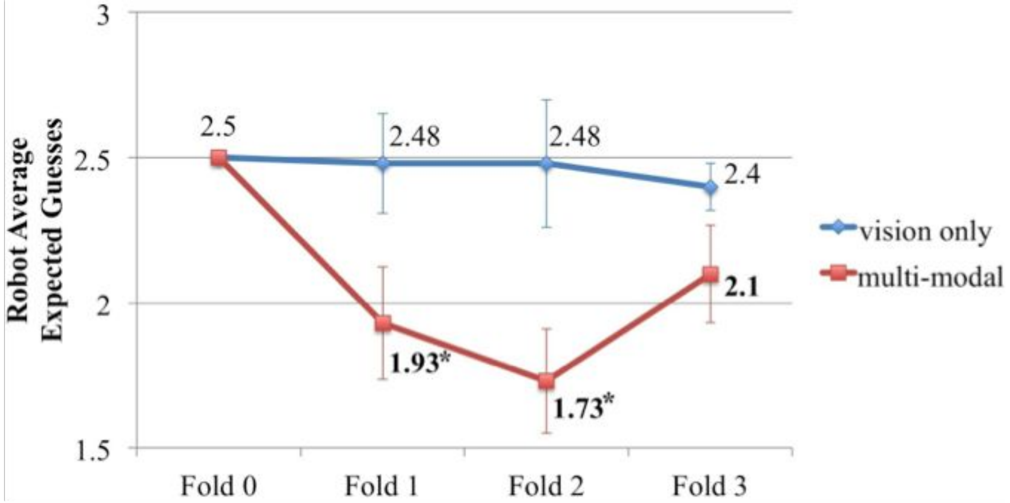
Experiments Playing I *Spy*

Four folds of objects for four rounds of training.



Bold: Lower than fold 0 average. *: Lower than vision only baseline

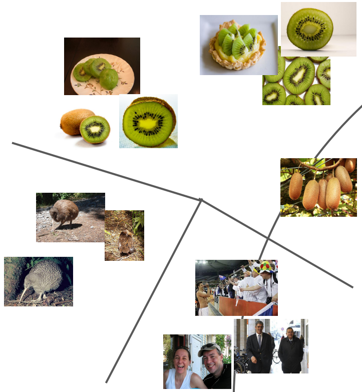
Problematic *I Spy* Object



Bold: Lower than fold 0 average. *: Lower than vision only baseline

Future: Be mindful of object *novelty* both for the learning algorithm and for human users.

[Thomason et al., IJCAI’16]



NLP

Robotics

Dialog

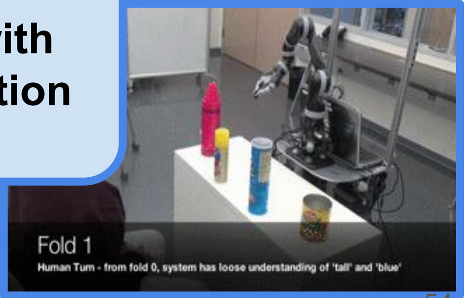
Polysemy Induction and Synonymy Detection (IJCAI’17)

Human-Robot Dialog

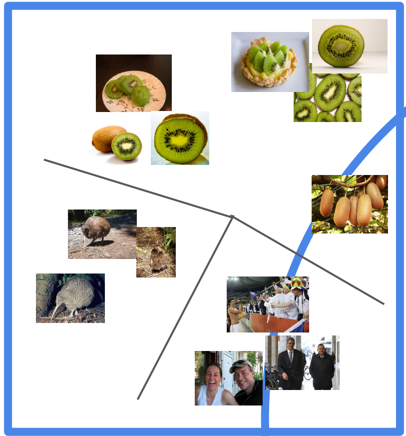
Papers before proposal

Improving Semantic Parsing through Dialog (IJCAI’15)

Learning Groundings with Human Interaction (IJCAI’16)



[Thomason et al., IJCAI'17]



Polysemy Induction and Synonymy Detection

NLP

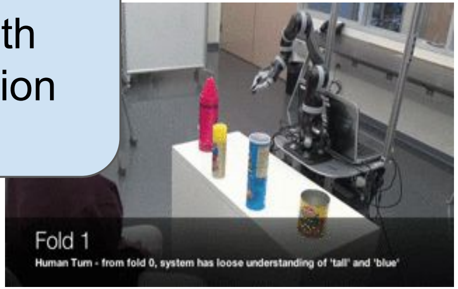
Robotics

Papers before proposal

Human-Robot Dialog

Improving Semantic Parsing through Dialog (IJCAI'15)

Learning Groundings with Human Interaction (IJCAI'16)



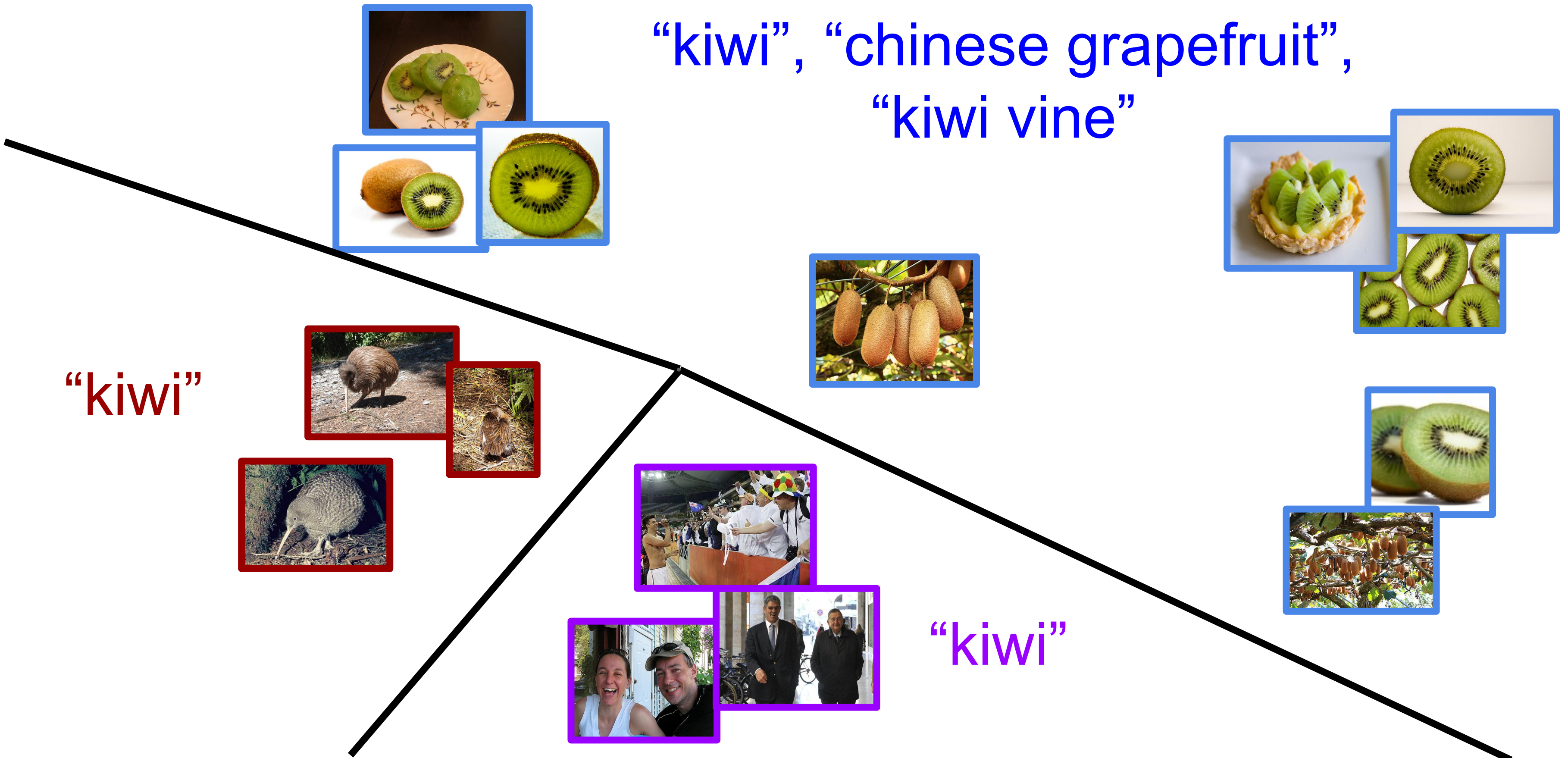
Dialog

[Thomason et al., IJCAI’17]

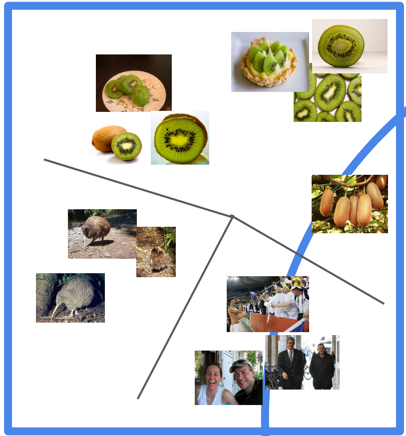
Unsupervised Word Synset Induction

[Thomason et al., IJCAI’17]

Unsupervised Word Synset Induction



[Thomason et al., IJCAI'17]



Polysemy Induction and Synonymy Detection

NLP

Robotics

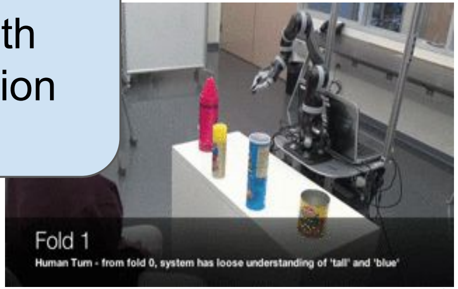
Papers before proposal

Human-Robot Dialog

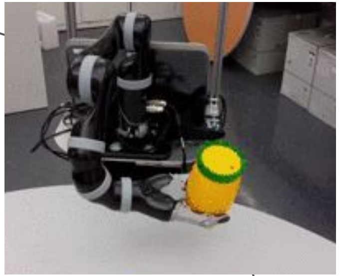
Improving Semantic Parsing through Dialog (IJCAI'15)

Learning Groundings with Human Interaction (IJCAI'16)

Dialog



Faster Object Exploration for Grounding (AAAI’18)



NLP

Robotics

Papers since proposal

Jointly Improving Parsing & Perception (*in submission*)

Learning Groundings with Opportunistic Active Learning (CoRL’17)

Dialog

[Thomason et al., AAAI'18]

Faster Object Exploration for Grounding (AAAI'18)

NLP

Robotics

Papers since proposal

Jointly Improving Parsing & Perception (*in submission*)

Learning Groundings with Opportunistic Active Learning (CoRL'17)

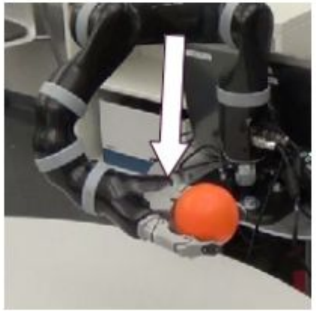
Dialog

[Thomason et al., AAAI'18]

Exploratory Behaviors

grasp (22s)

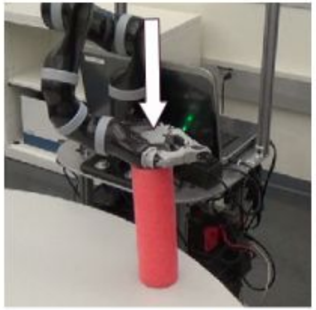
lift (11.1s)



lower (10.6s)

drop (9.8s)

push (22s)



press (22s)

+hold (5.7s)

+look (0.8s)

104s to explore an object once.

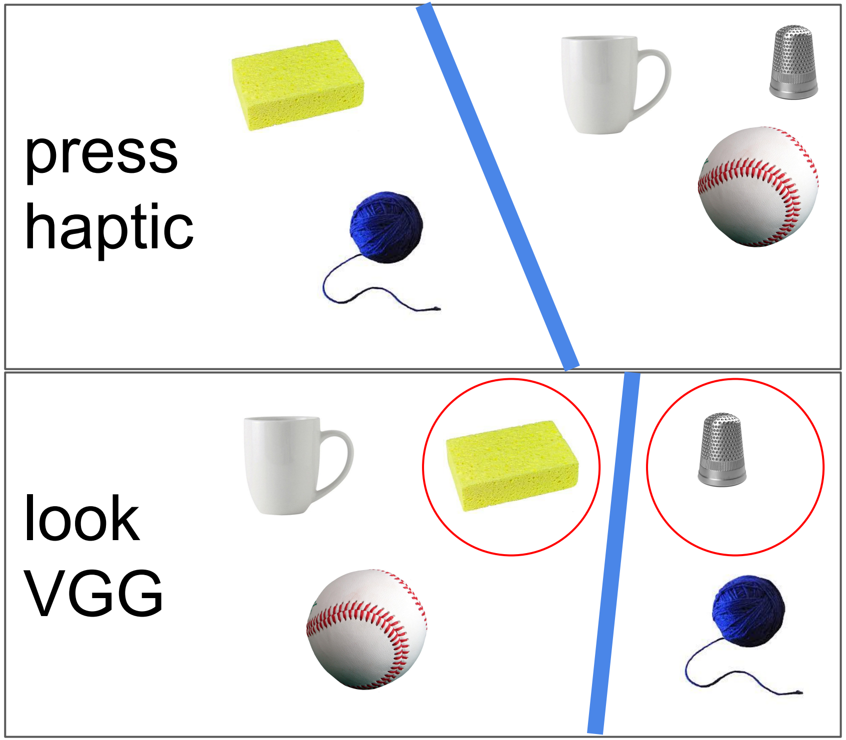
520s to explore an object five times.

4.5 **hours** to fully explore 32 objects.

[Thomason et al., AAAI’18]

Guiding Exploratory Behaviors

rigid:



squishy?

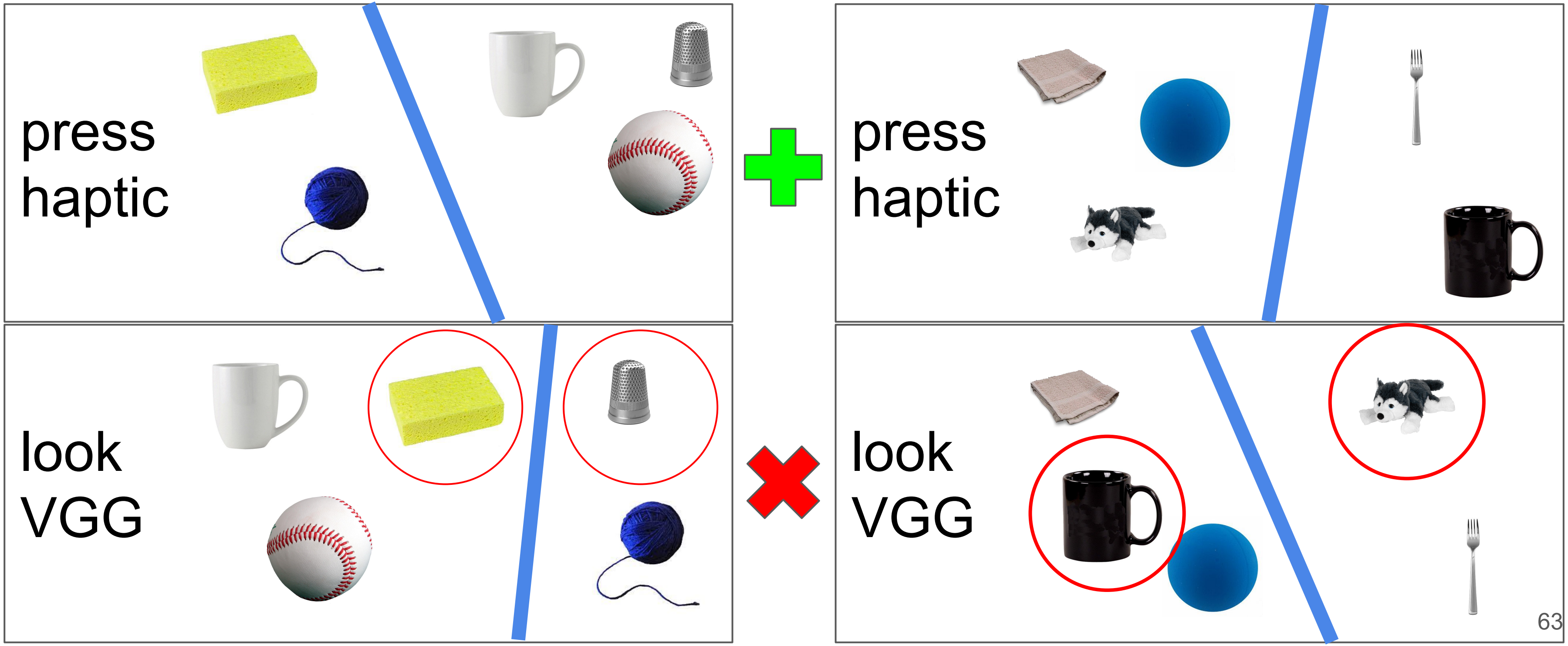
press?

look?

[Thomason et al., AAAI'18]

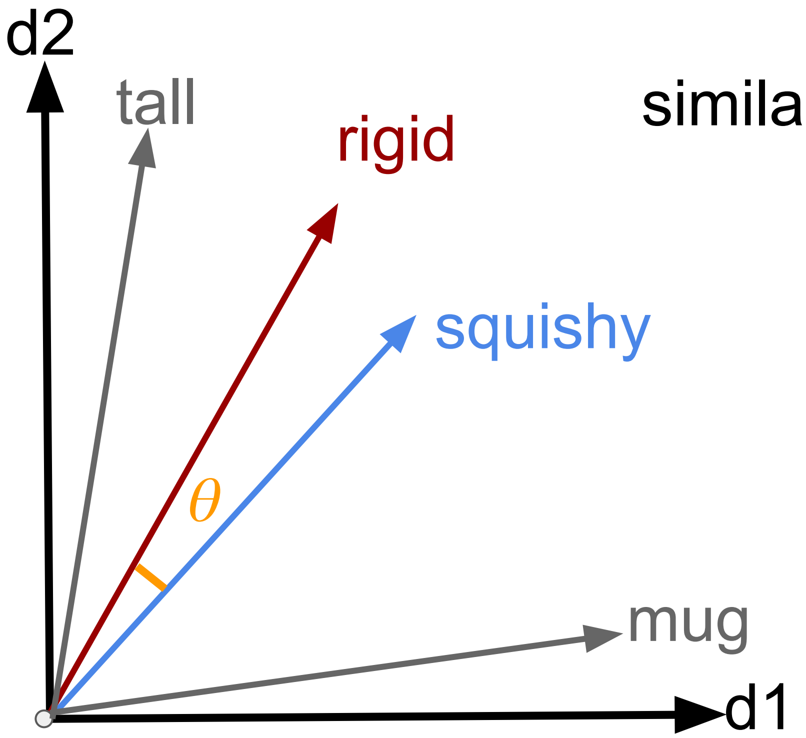
Guiding Exploratory Behaviors

rigid: squishy



[Thomason et al., AAAI'18;
Mikolov et al., NIPS'13]

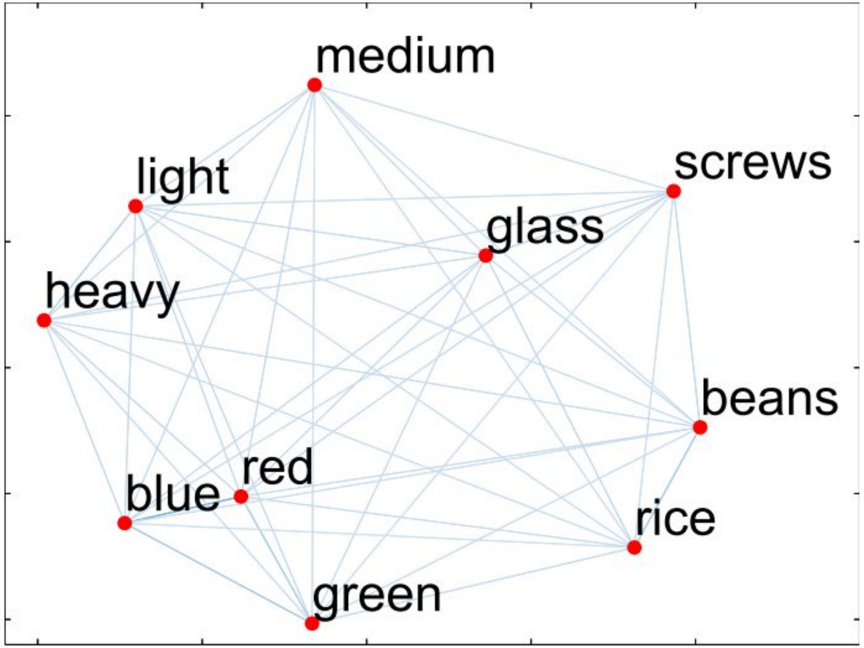
Guiding Exploratory Behaviors



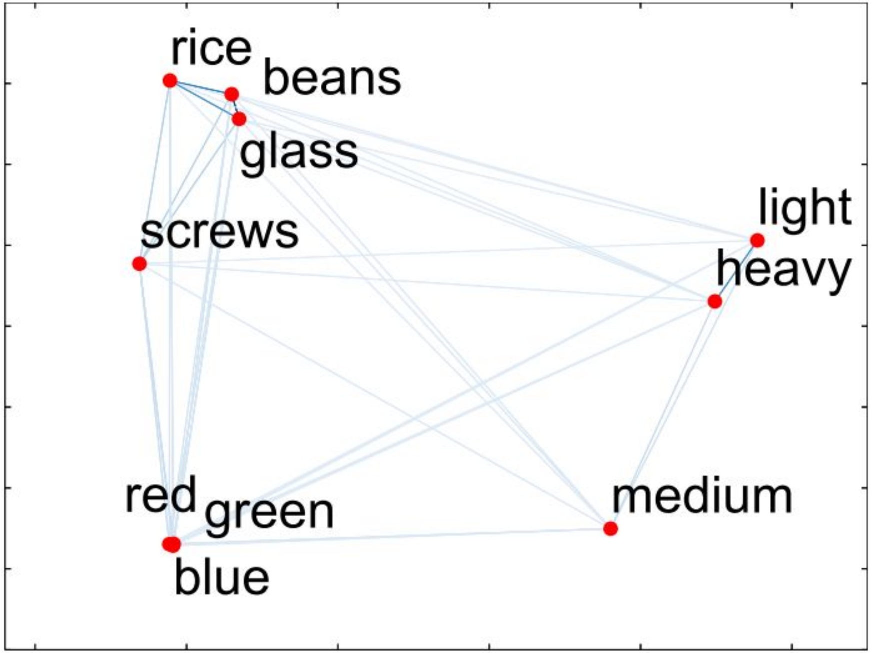
similarity(rigid, squishy) = $cos(\theta)$

[Thomason et al., AAAI’18]

Shared Structure: Embeddings and Features



2D-projection of word embeddings



2D-projection of behavior context features

[Thomason et al., AAAI'18]

Guiding Exploratory Behaviors using Embeddings

$$d(p,o) = sgn\left(\sum_{c \in C} \boxed{w_{p,c}} \mathrm{G}_{p,c}(o)\right)$$

$$\boxed{w_{q,c}} \approx \frac{1}{|P_q|} \sum_{\boxed{p \in P_q}} poscos(p,q) \boxed{w_{p,c}}$$

Surrogate reliability weights for new classifiers for *q*

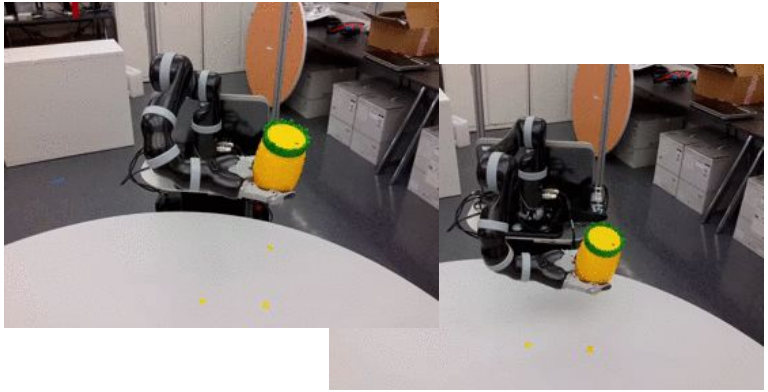
Nearest word-embedding predicates to *q*

Reliability weights for trained neighbor classifiers *p*

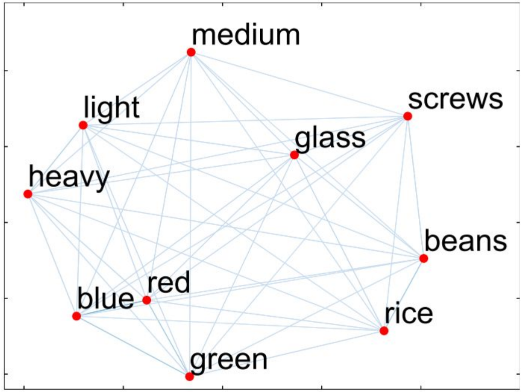
[Thomason et al., AAAI'18]

Technical Contributions

- Reduce exploration time when **learning a target new word**.

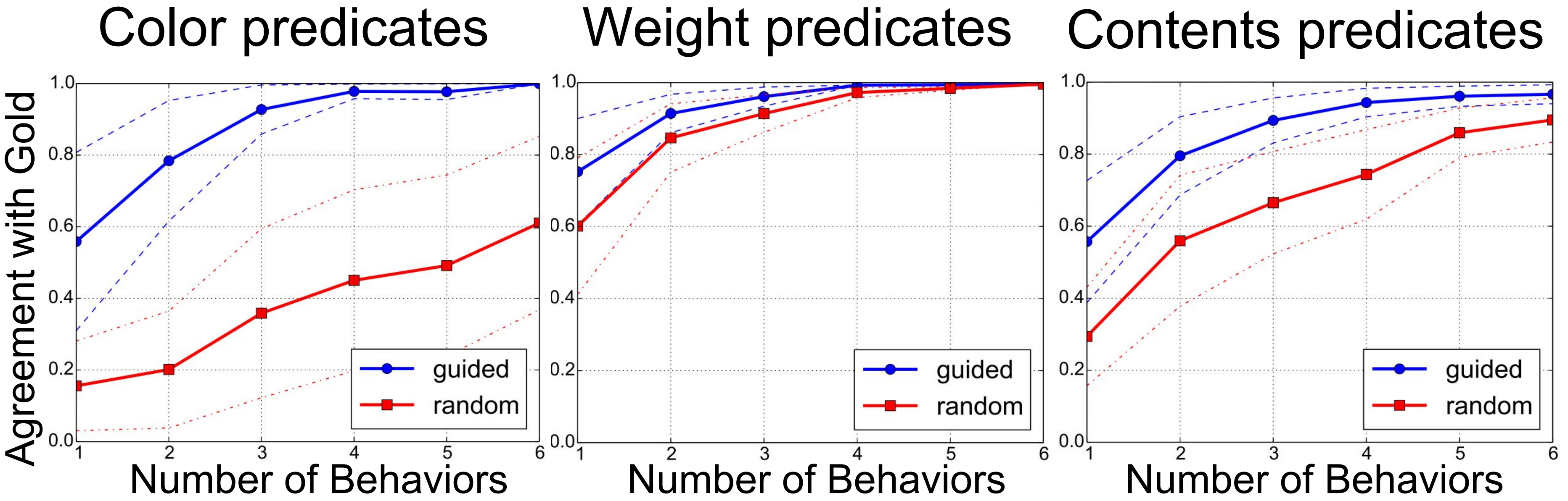


- Use word embeddings and human annotations to **guide behaviors.**



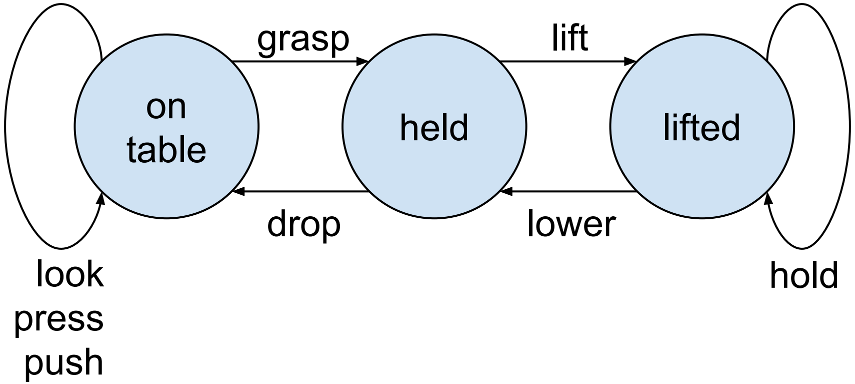
[Thomason et al., AAAI'18]

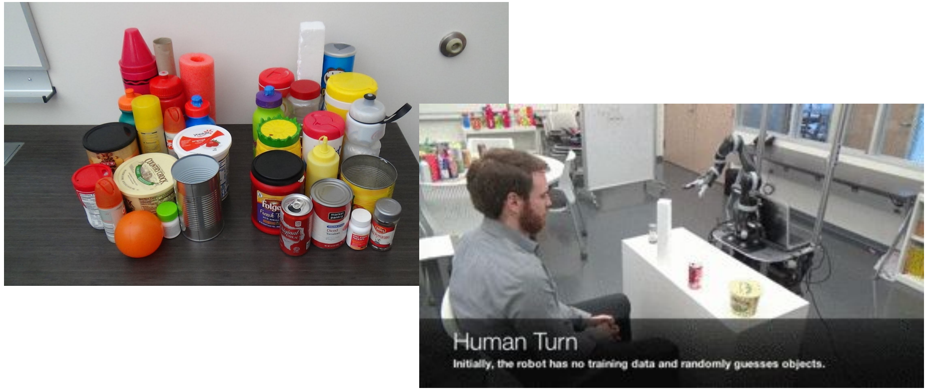
Results



(dotted lines show standard error)

Other Findings





- Human annotations help; “how would you tell if an object is *tall*?”
- Human annotations + word embeddings work better than either alone.

[Thomason et al., AAAI’18]

Faster Object Exploration for Grounding (AAAI’18)

NLP

Robotics

Papers since proposal

Jointly Improving Parsing & Perception (*in submission*)

Learning Groundings with Opportunistic Active Learning (CoRL’17)

Dialog

[Thomason et al., CoRL'17]

Faster Object Exploration for Grounding (AAAI'18)

NLP

Robotics

Papers since proposal

Jointly Improving Parsing & Perception (*in submisison*)

Learning Groundings with Opportunistic Active Learning

Dialog

Active Learning for Perceptual Questions

$$o_{\min}(p) = \operatorname{argmin}_{o \in O_{tr}}(\kappa(p, o))$$

The object for which the predicate classifier is least sure of the predicted label.

d(`bottle`, ) = -0.6

d(`bottle`, ) = 0.8

d(`bottle`, ) = 0.4

d(`bottle`, 

) = -0.2

[Thomason et al., CoRL'17]

Active Learning for Perceptual Questions

empty	
sensorimotor context	$\boldsymbol{w_{p,c}}$
lift-haptics	?
lift-audio	?
...	...
look-vgg	?

bottle	
sensorimotor context	$\boldsymbol{w_{p,c}}$
look-shape	0.6
look-vgg	0.5
...	...
lower-haptics	0.02

Active Learning for Perceptual Questions

$$prob(p) = \frac{1 - \kappa(p, o_{\min}(p))}{\sum_{q \in P \setminus \{p\}} 1 - \kappa(q, o_{\min}(q))}$$

Ask for a label with probability proportional to *un*confidence in least confident training object.

$$p \in \{q : q \in P \land \kappa(q, o_{\min}(q)) = 0\}$$

Ask for a positive label for any predicate we have insufficient data for.

Active Learning for Perceptual Questions

"Could you use the word bottle when describing this object?"

Ask for a label with probability proportional to *un*confidence in least confident training object.

"Can you show me something empty?"

Ask for a positive label for any predicate we have insufficient data for.

[Thomason et al., CoRL'17]

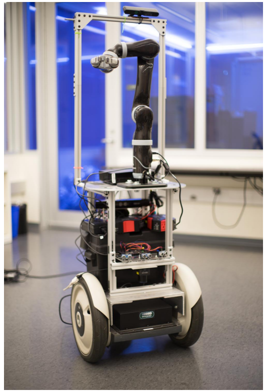
[Thomason et al., CoRL'17]

Technical Contributions

- Introduce an **opportunistic active learning** strategy for getting high-value labels.

- Show that *off-topic* questions **improve performance.**

"A full, yellow bottle."



"Would you describe this object as full?"

"Show me something red."

Experiments with Object Identification

"Would you describe this object as full?"

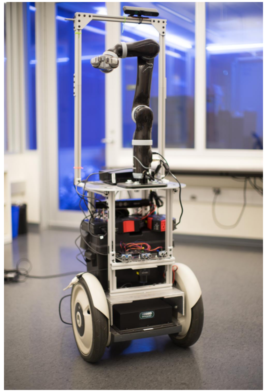
vs

"Show me something red."

Baseline Agent

Inquisitive Agent

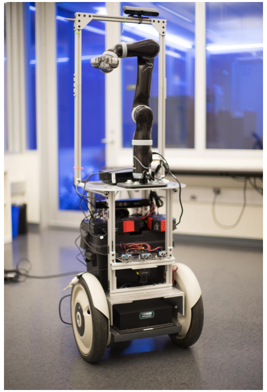
Results



“Would you describe this object as full?”

Baseline Agent

Rated less annoying.



“Show me something red.”

Inquisitive Agent

Correct object more often.

Rated better for real-world use.

[Thomason et al., CoRL’17]

Faster Object Exploration for Grounding (AAAI’18)

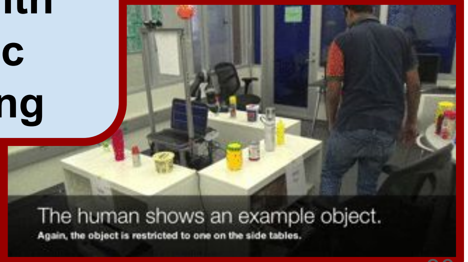
NLP

Robotics

Papers since proposal

Jointly Improving Parsing & Perception (*in submission*)

Learning Groundings with Opportunistic Active Learning



Dialog

[*in submission*]

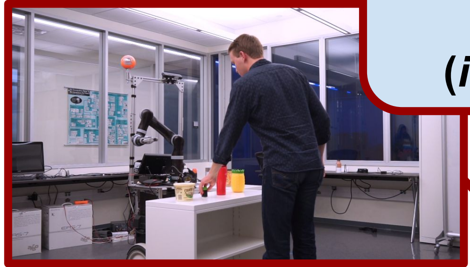
Faster Object Exploration for Grounding (AAAI'18)

NLP

Robotics

Papers since proposal

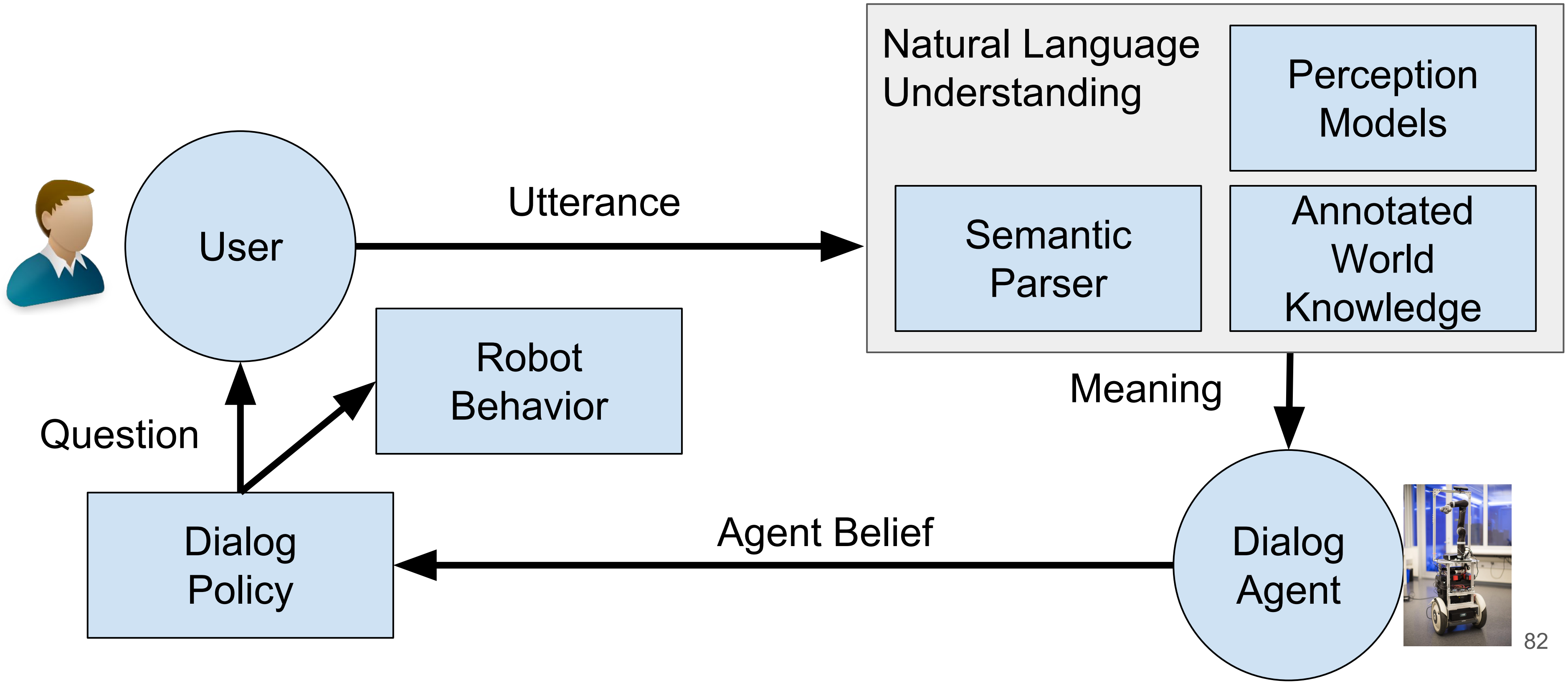
Jointly Improving Parsing & Perception (*in submission*)



Learning Groundings with Opportunistic Active Learning (CoRL'17)

Dialog

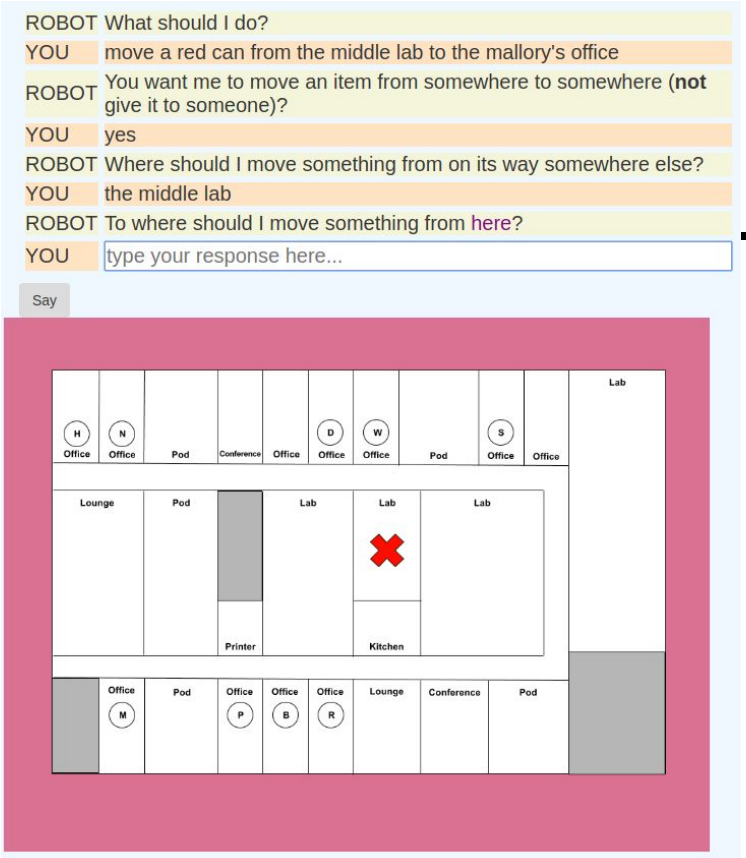
Human-Robot Dialog

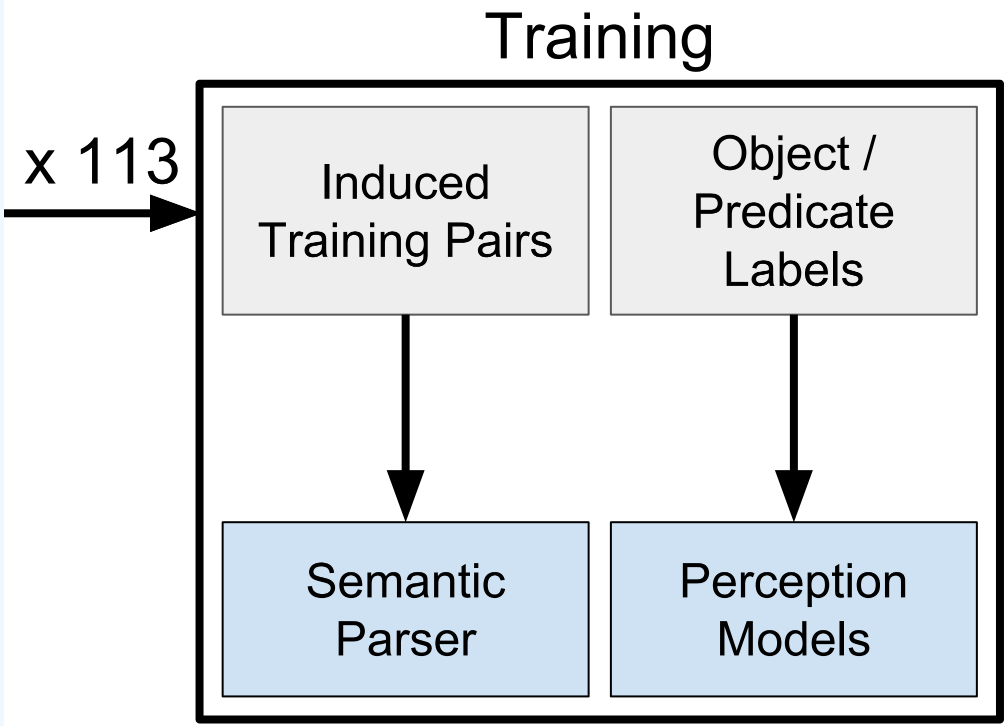


Jointly Improving Parsing and Perception

"Move a rattling container from lounge by the conference room to Bob's office."

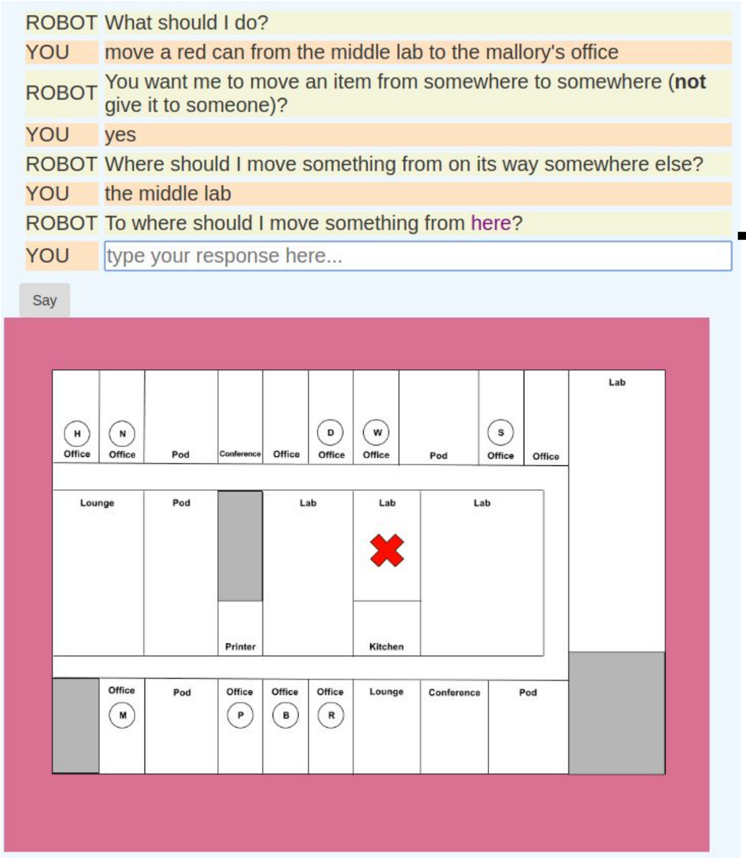
Experiments via Amazon Mechanical Turk

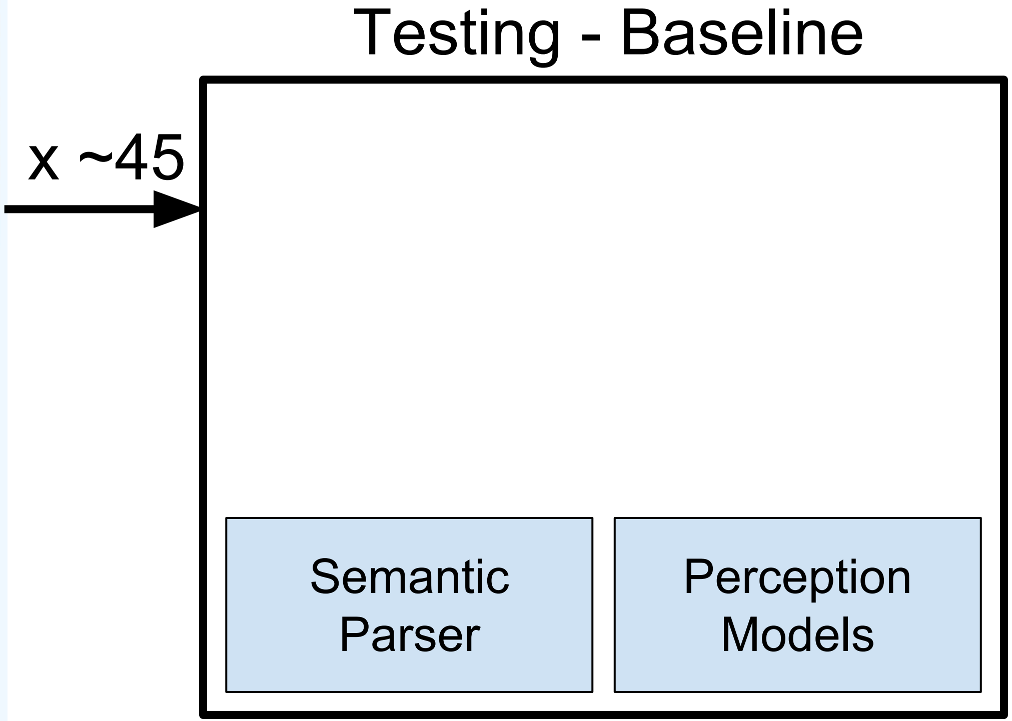




[*in submission*]

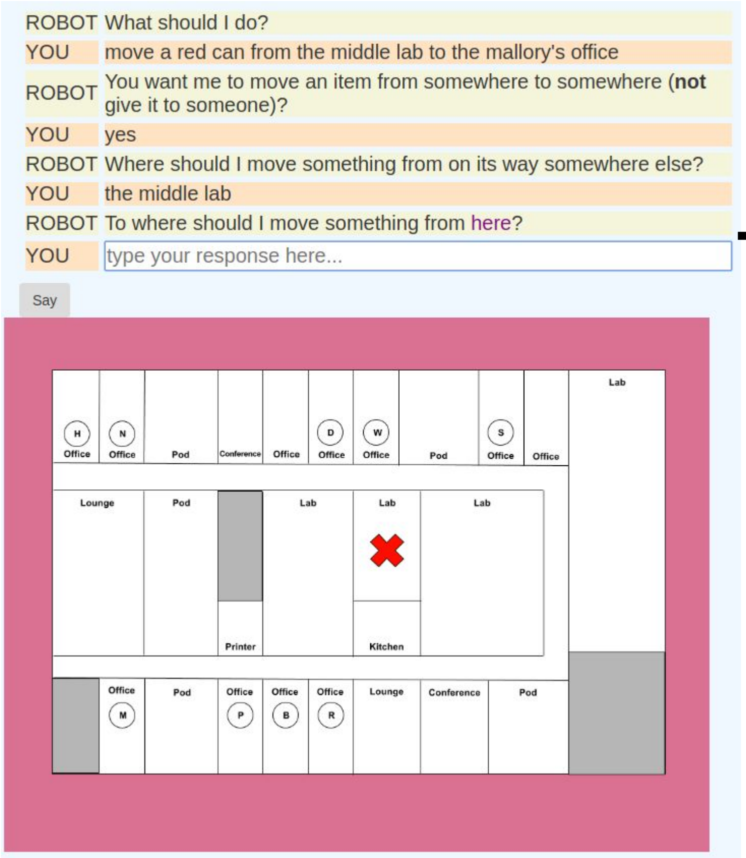
Experiments via Amazon Mechanical Turk



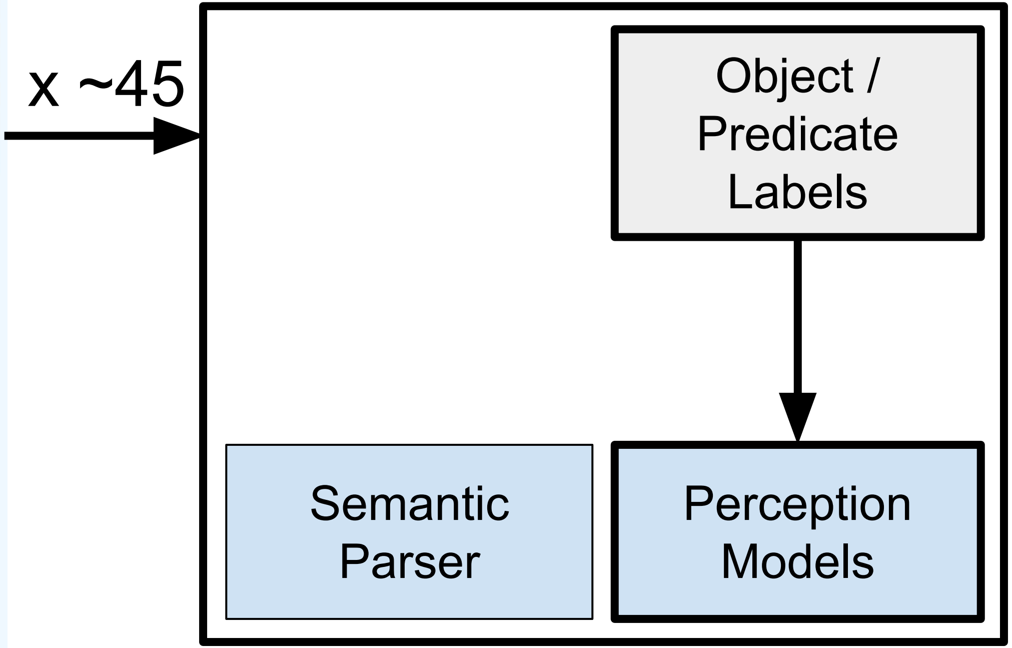


[in submission]

Experiments via Amazon Mechanical Turk

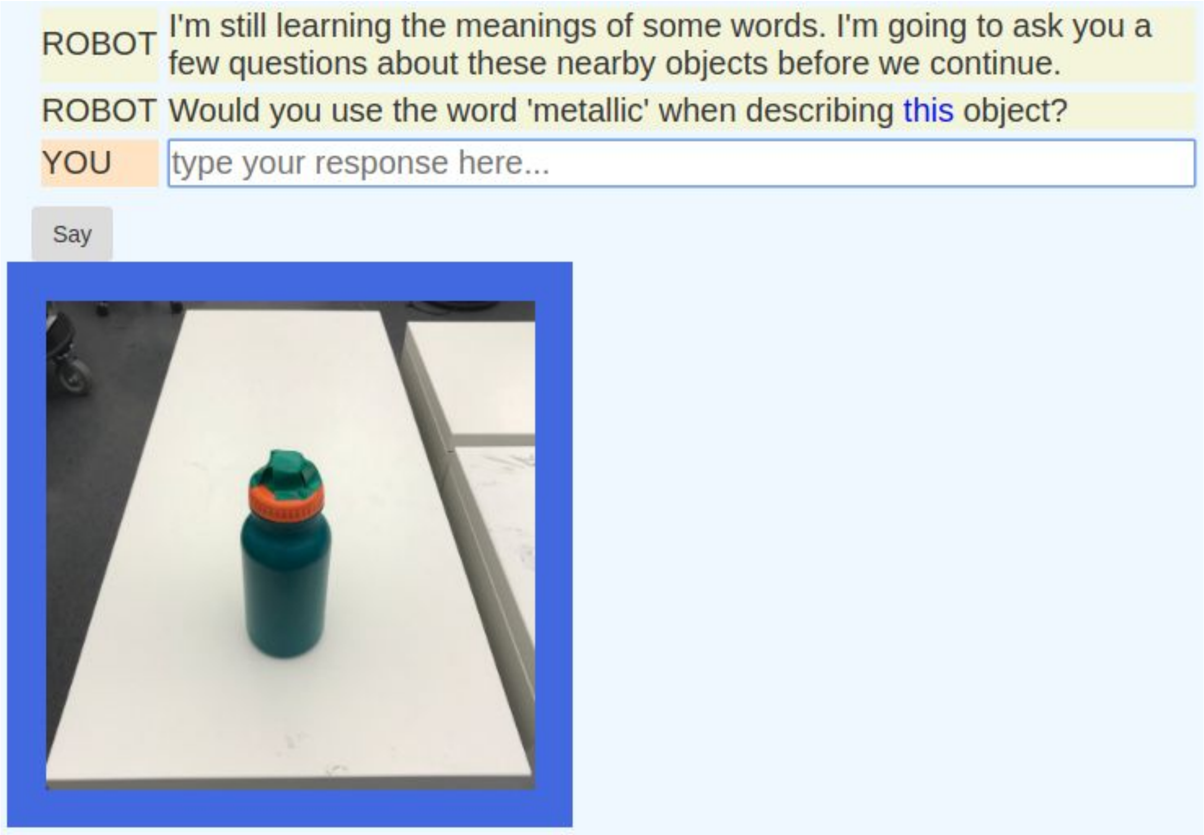


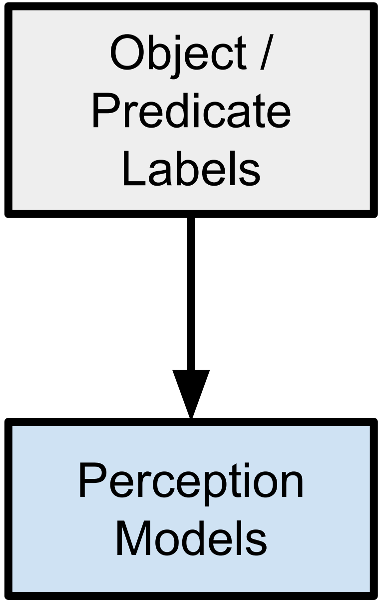
Testing - Perception



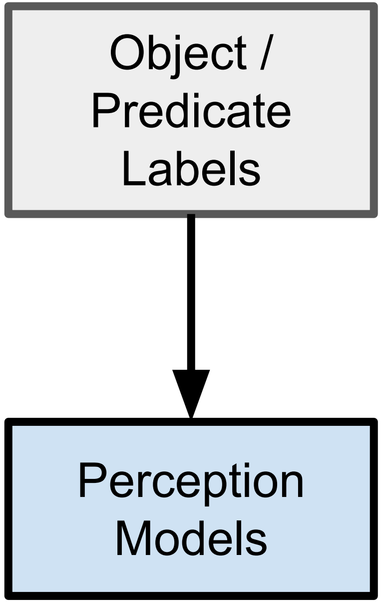
[*in submission*]

Getting Object/Predicate Labels in Dialog



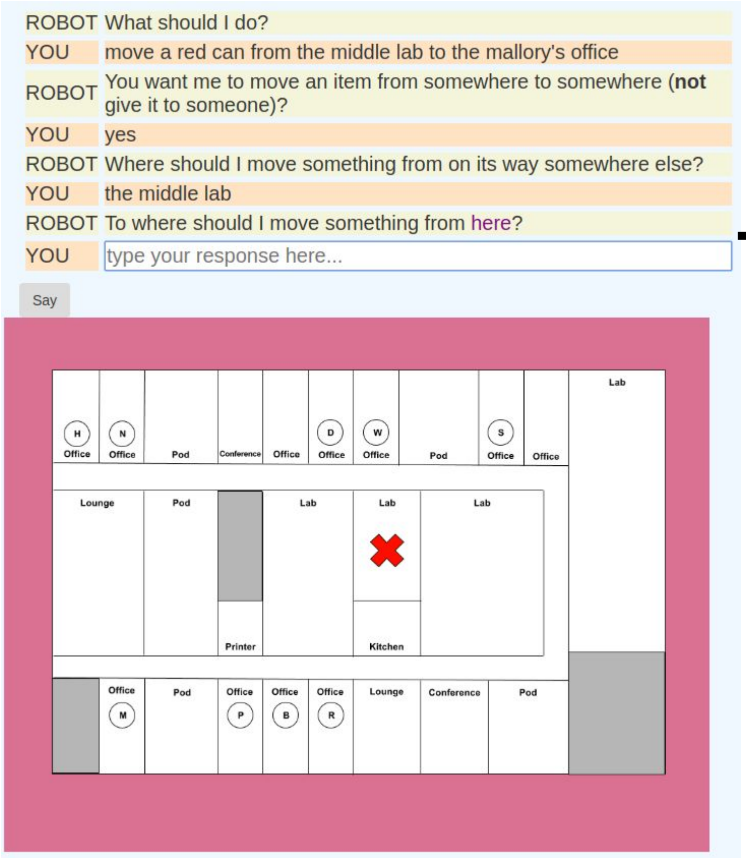


Getting Object/Predicate Labels in Dialog

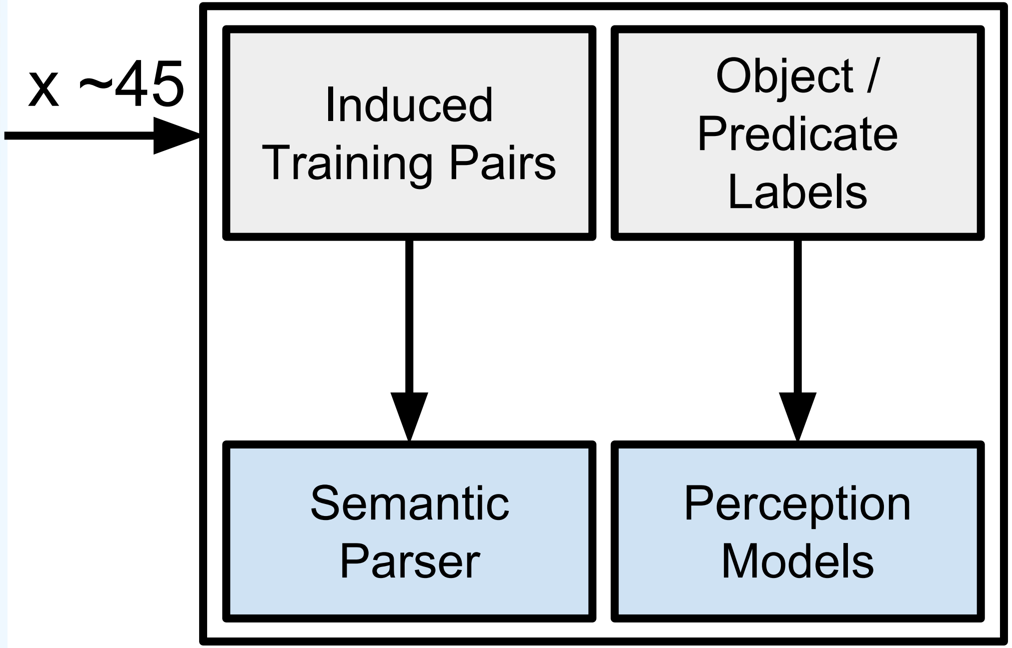


[*in submission*]

Experiments via Amazon Mechanical Turk

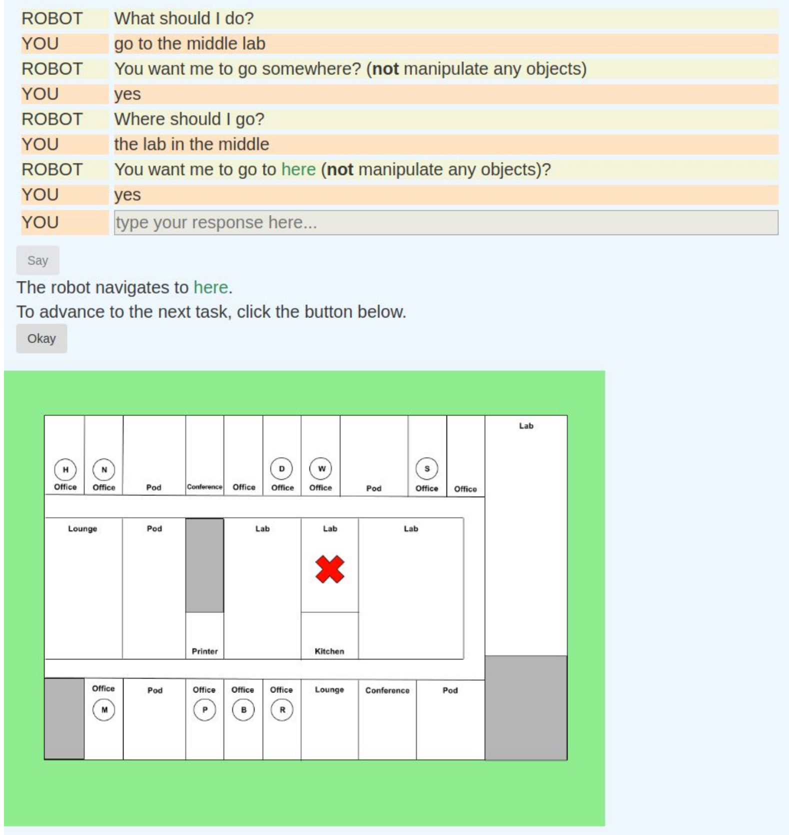


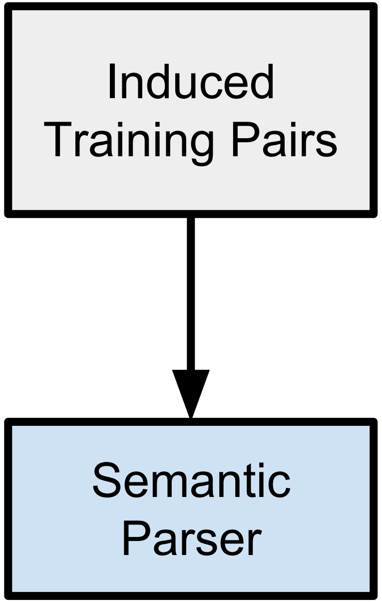
Testing - Parsing + Perception



[*in submission*]

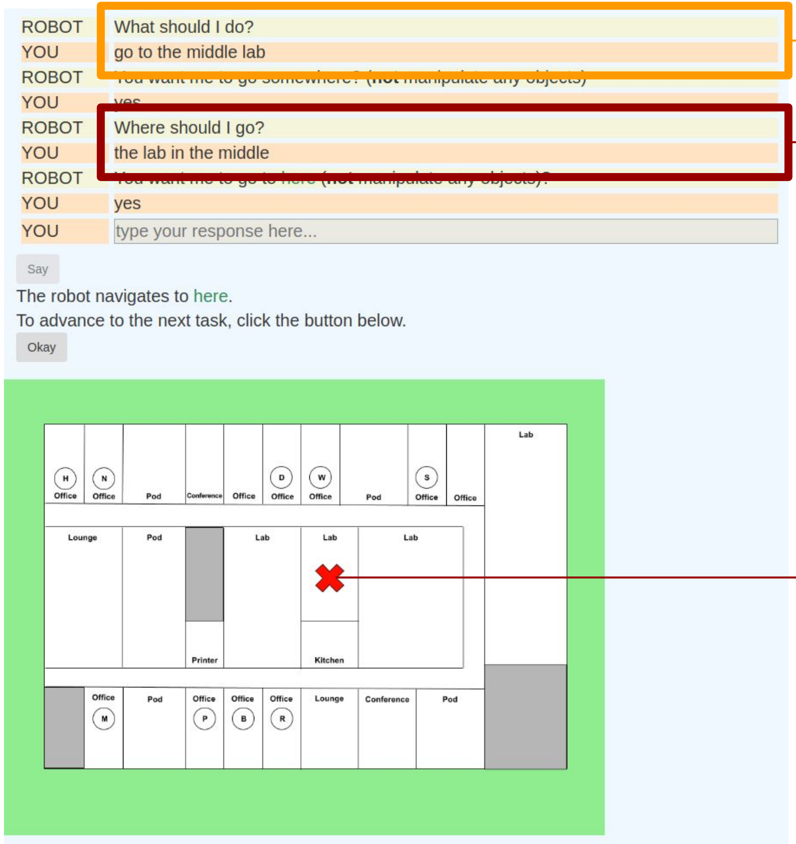
Inducing New Training Examples from Dialog





[*in submission*]

Inducing New Training Examples from Dialog



Expect whole command

Expect *goal*

task: `navigate`

goal: `room_3`

[*in submission*]

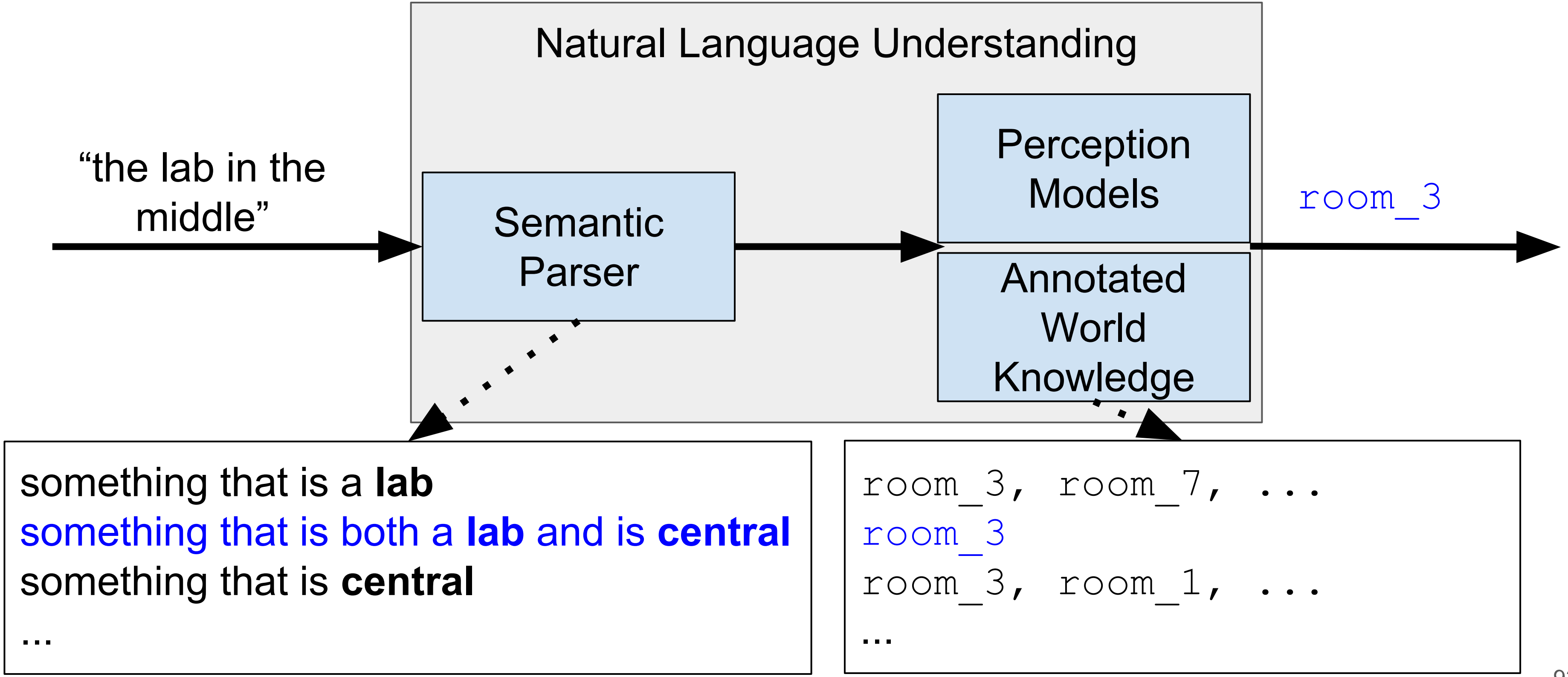
Inducing New Training Examples from Dialog

Induced Utterance/Denotation Pairs

"go to the middle lab" `navigate(room_3)`
"the lab in the middle" `room_3`

[*in submission*]

Natural Language Understanding



[*in submission*]

Inducing New Training Examples from Dialog

Induced Utterance/Denotation Pairs

"go to the middle lab" `navigate(room_3)`
"the lab in the middle" `room_3`

Semantic Parser

Perception Models

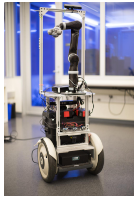
Annotated World Knowledge

Induced Parser Training Data

"go to the middle lab" `navigate(lab+central)`
"the lab in the middle" `lab+central`

[Mikolov et al., NIPS’13; *in submission*]

Using Embeddings for Out-of-Vocabulary Words



Induced Training Pairs

“deliver java to bob”

Word Embeddings

“deliver java to bob”

“deliver” -> “bring”
“java” -> “coffee”

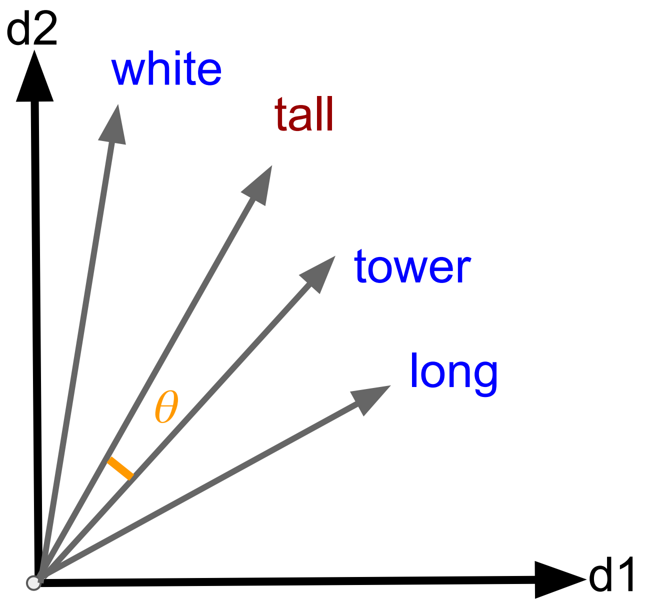
Semantic Parser

task: `deliver`
item: `coffee`
person: `bob`

[Mikolov et al., NIPS'13; *in submission*]

Using Embeddings to Find Perception Words

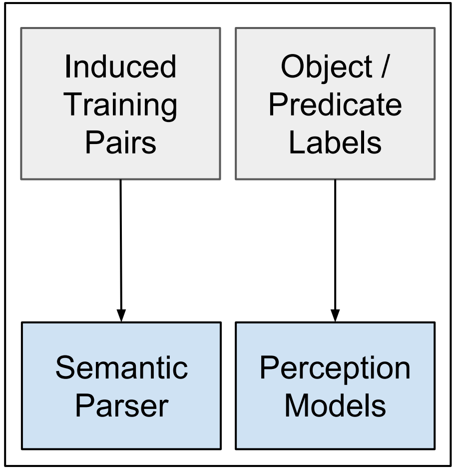
ROBOT	What should I deliver to this person?
YOU	a tall metal can
ROBOT	I haven't heard the word 'tall' before. Does it refer to properties of things, like a color, shape, or weight?
YOU	yes
ROBOT	Does 'tall' mean the same thing as 'tower'?
YOU	no
ROBOT	Does 'tall' mean the same thing as 'white'?
YOU	no
ROBOT	Does 'tall' mean the same thing as 'long'?
YOU	yes



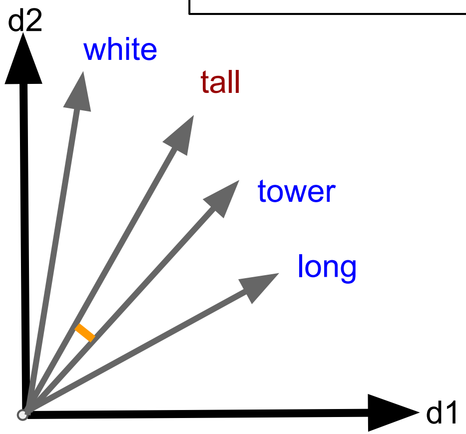
[*in submission*]

Technical Contributions

- Improve **both parsing and perception** from conversations.

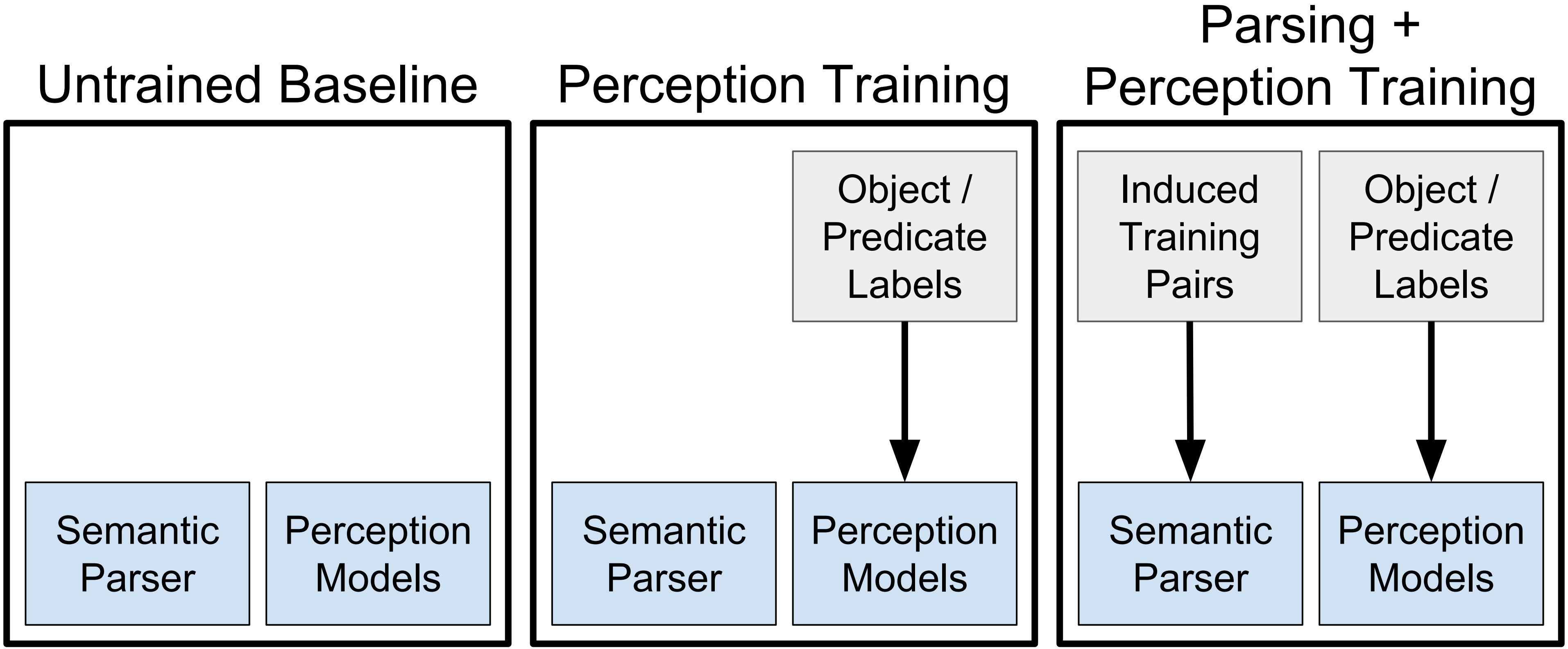


- Use word embeddings to **guide search for synonyms** and **novel perceptual predicates**.



[*in submission*]

Experiments via Amazon Mechanical Turk



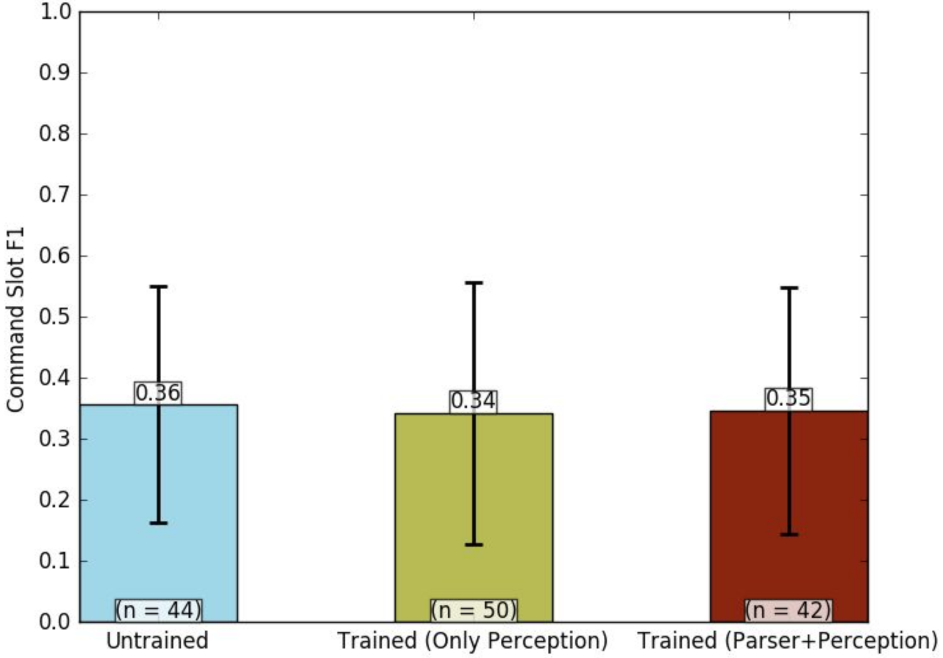
[*in submission*]

Metric - Semantic F1

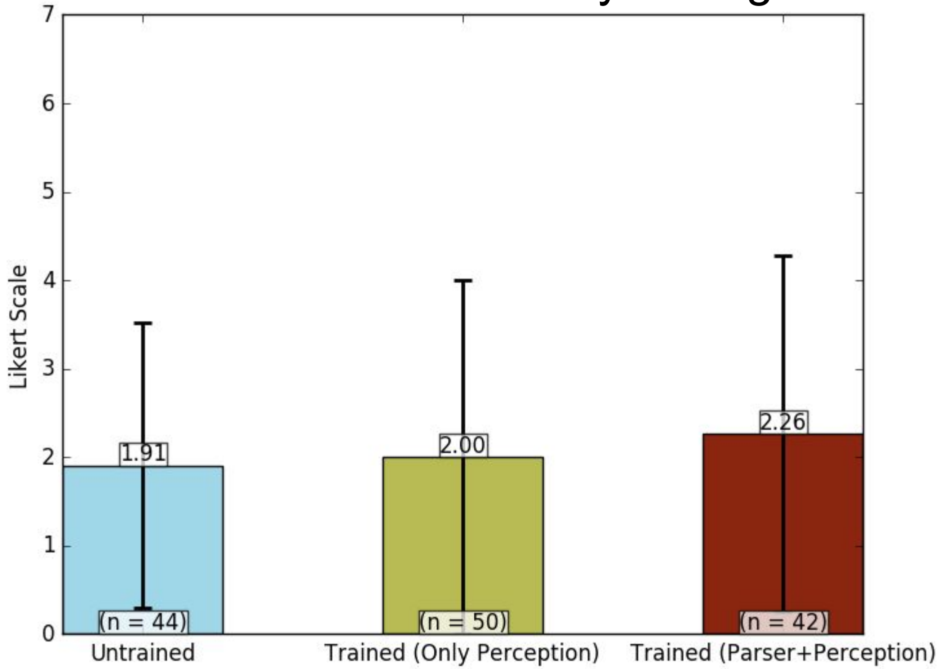
$$
\begin{aligned}
T_U =& \{(\text{action}, \text{deliver}), (\text{patient}, o_2), (\text{recipient}, p_1)\}, \\
T_G =& \{(\text{action}, \text{relocate}), (\text{patient}, o_2), (\text{source}, r_1), (\text{goal}, r_3)\}; \\
\text{precision}(T_U, T_G) =& \frac{|T_U \cap T_G|}{|T_U|} = \frac{1}{3}, \\
\text{recall}(T_U, T_G) =& \frac{|T_U \cap T_G|}{|T_G|} = \frac{1}{4}, \\
f(T_U, T_G) =& 2 \cdot \frac{\text{precision}(T_U, T_G) \cdot \text{recall}(T_U, T_G)}{\text{precision}(T_U, T_G) + \text{recall}(T_U, T_G)} = 0.286.
\end{aligned}
$$

Results - Navigation Task

Quantitative - Semantic F1



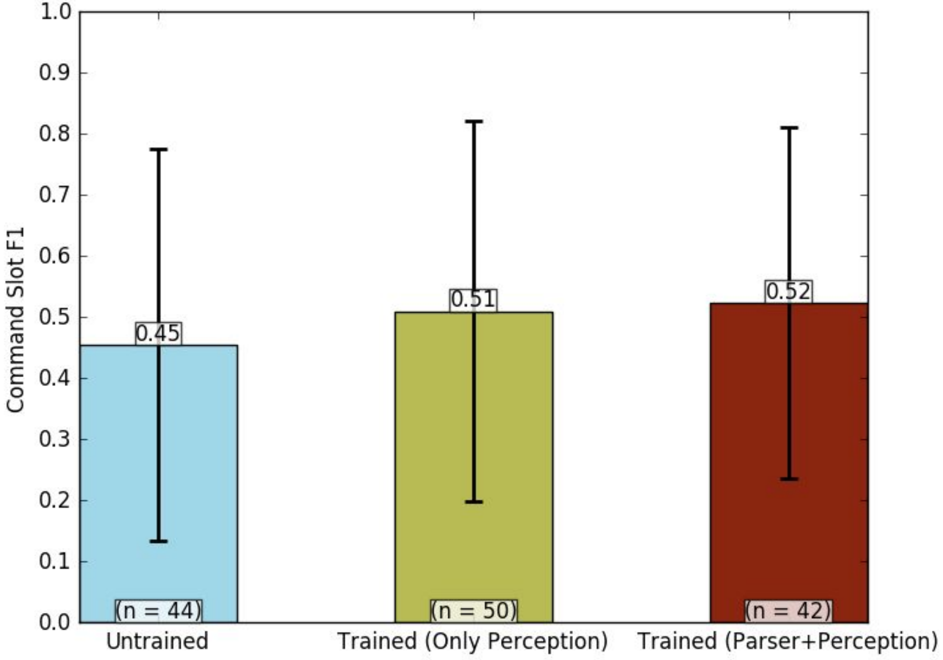
Qualitative - Usability Rating



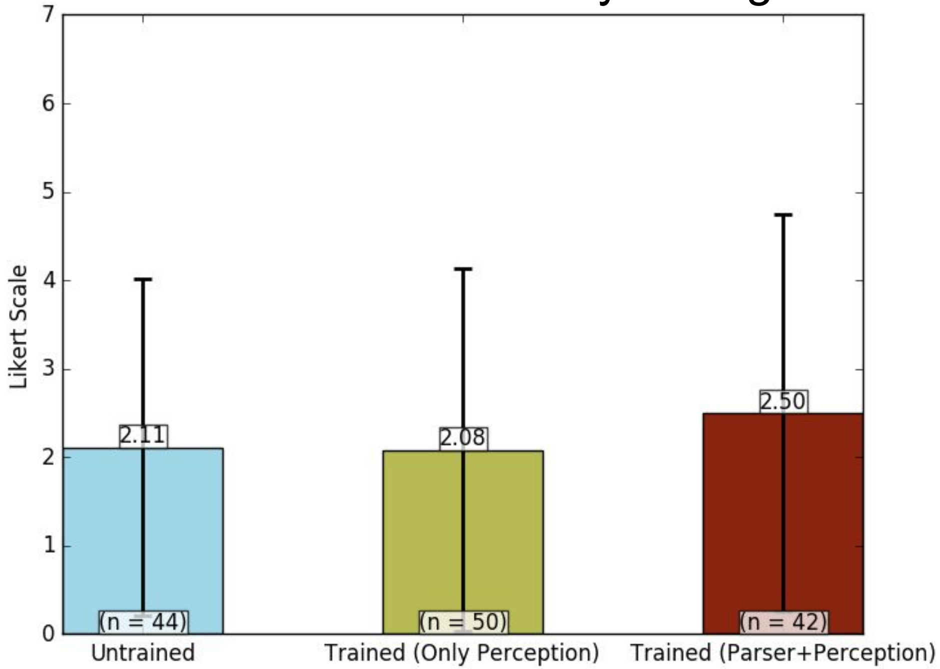
[*in submission*]

Results - Delivery Task

Quantitative - Semantic F1



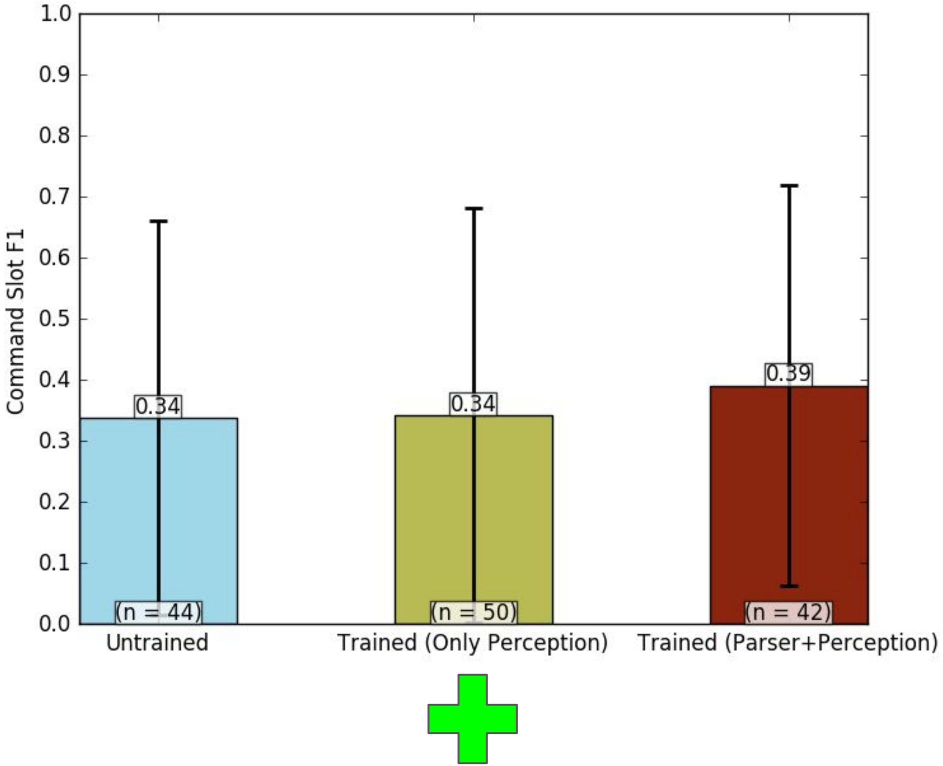
Qualitative - Usability Rating



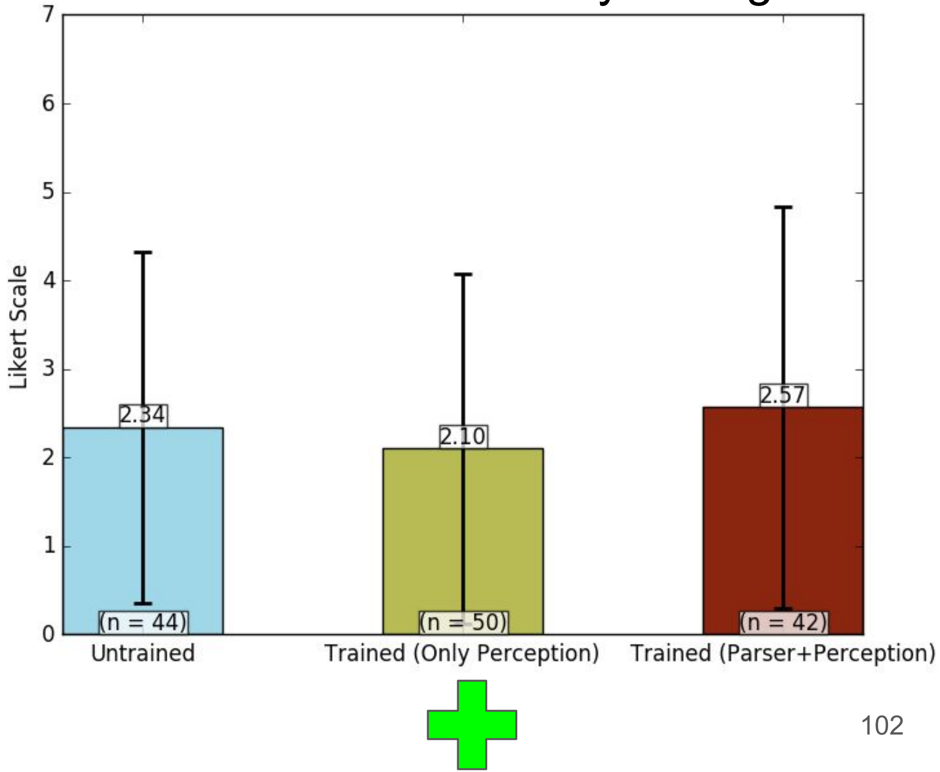
[*in submission*]

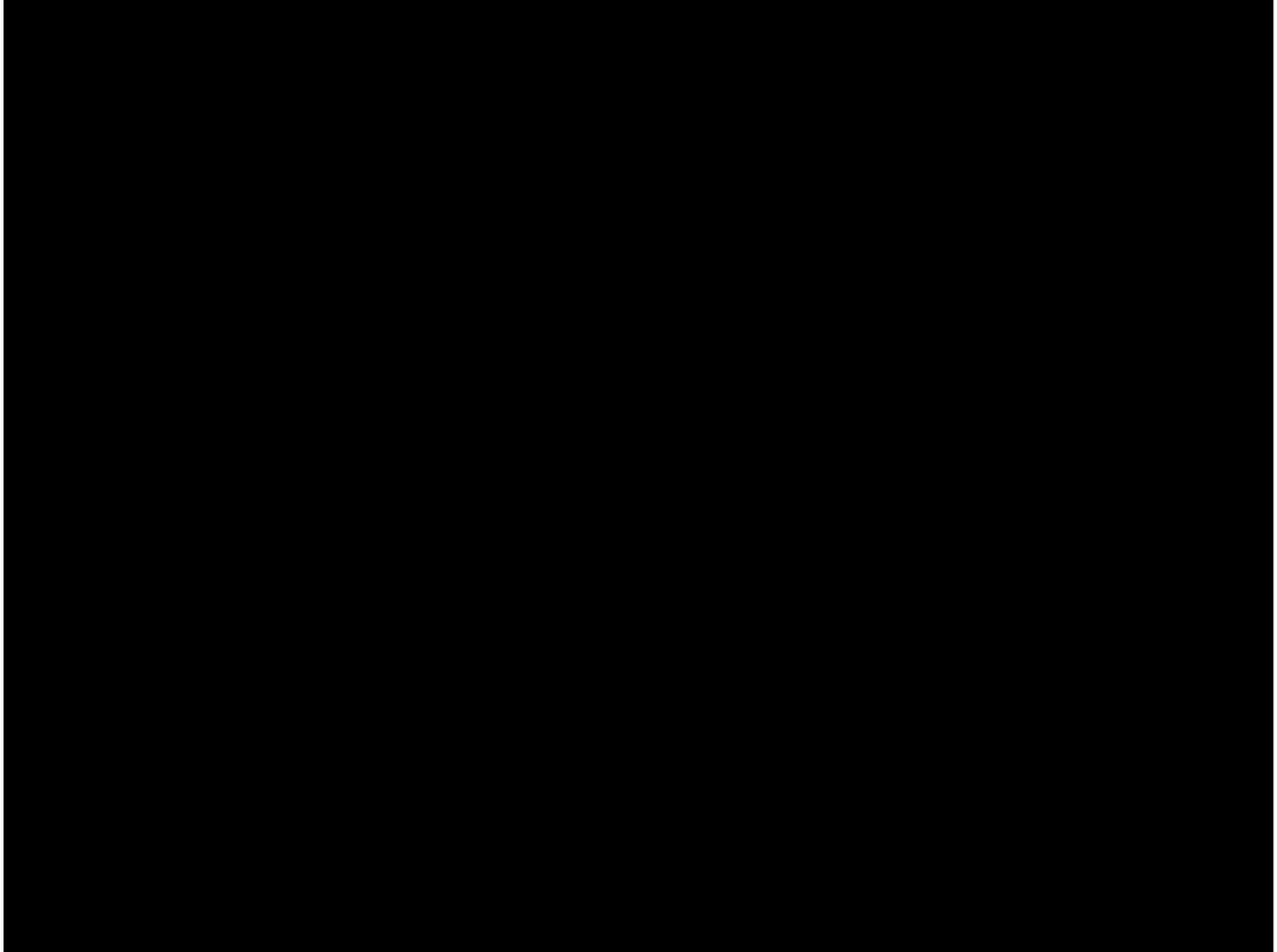
Results - Relocation Task

Quantitative - Semantic F1



Qualitative - Usability Rating





[*in submission*]

[*in submission*]

Faster Object Exploration for Grounding (AAAI'18)

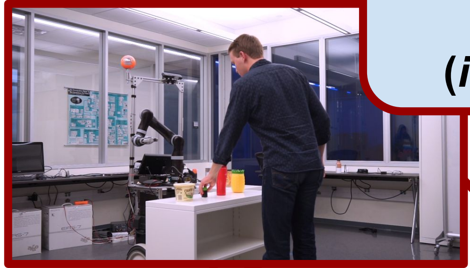
[*in submission*]

NLP

Robotics

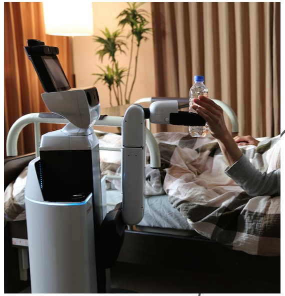
Papers since proposal

Jointly Improving Parsing & Perception (*in submission*)



Learning Groundings with Opportunistic Active Learning (CoRL'17)

Dialog



NLP

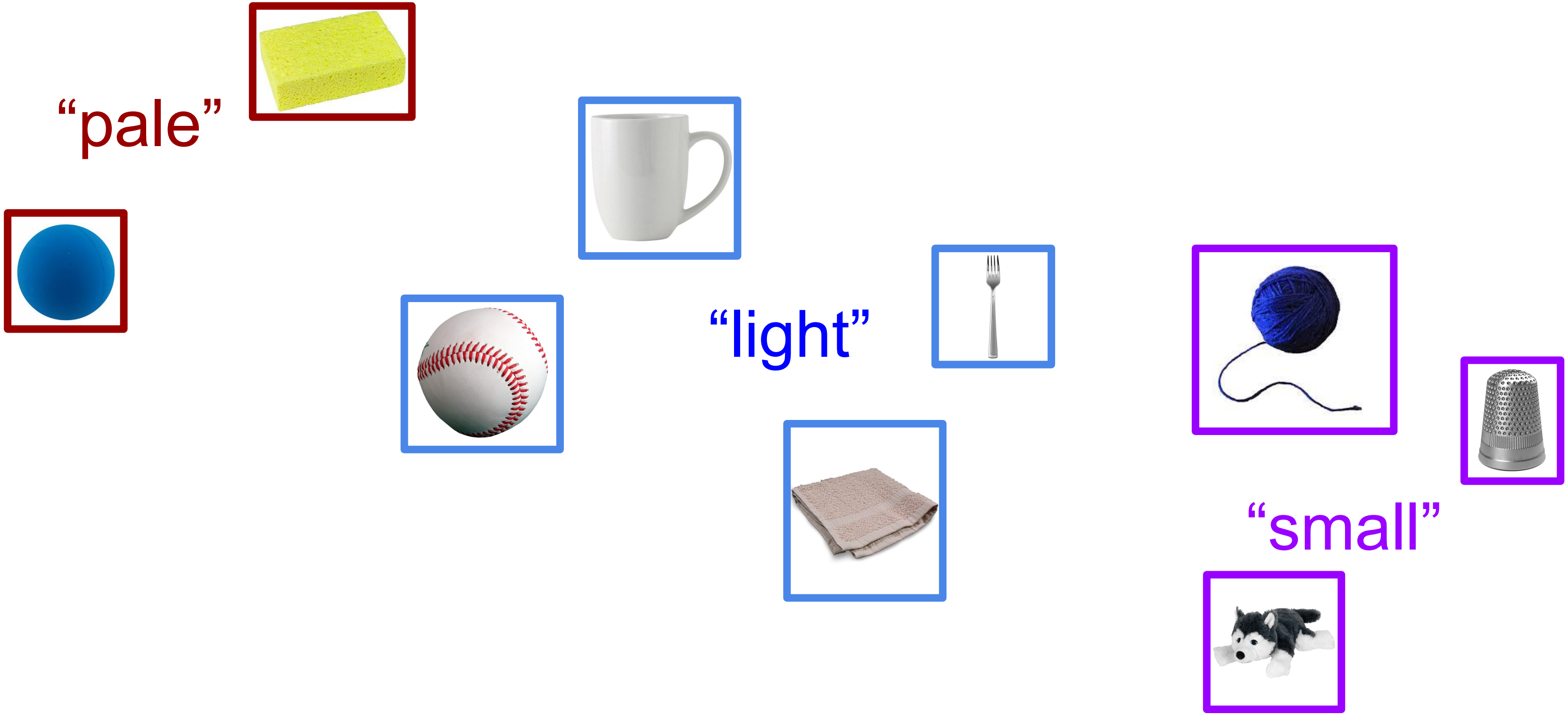
Robotics

Human-Robot Dialog

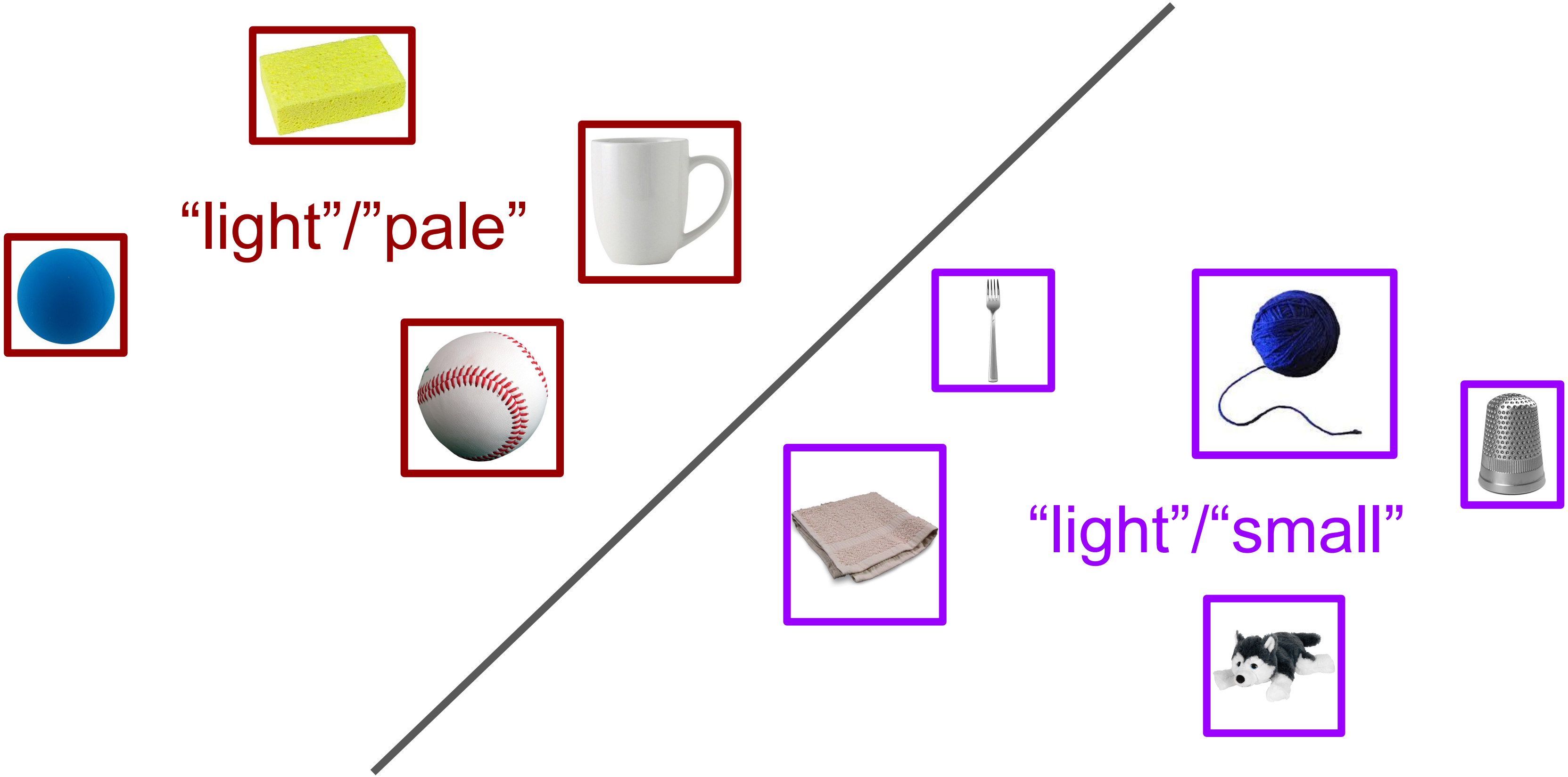
Next Directions

Dialog

Grounded Predicate Synset Induction



Grounded Predicate Synset Induction



Guided Exploration of New Objects

“Move a rattling container from the kitchen to bob’s office.”

`rattling`?

↓

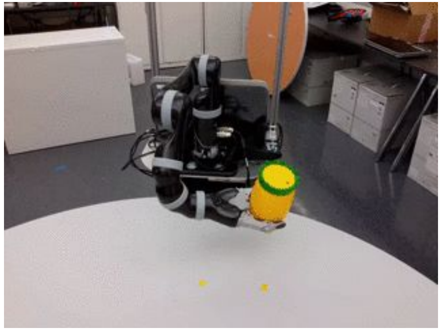
Guided Behavior(s)

↓

Perception Models

↓

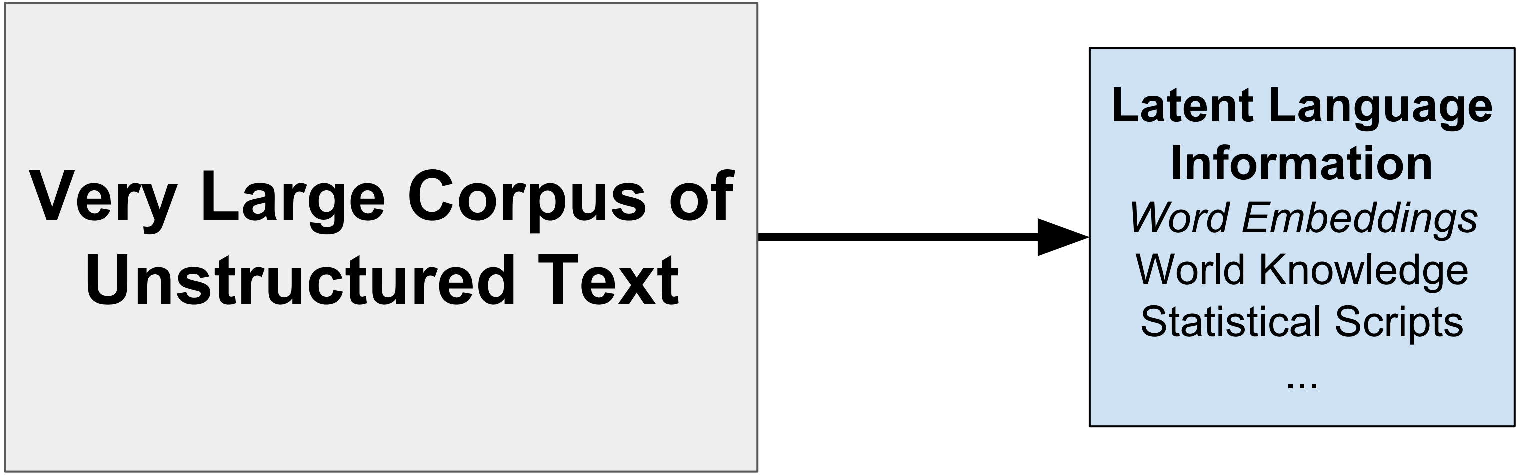
yes / no



Moving Forward

- The intersection of problems in human-robot dialog is **inherently low-resource**.
- Other parts of NLP, Robotics, and Dialog are not.
- We can **use big data and techniques** from these fields when solving problems in human-robot dialog.

Moving Forward - Using Big Data Where We Can



[Thomason et al., IJCAI'16; Simonyan and Zisserman, CoRR'14]

Moving Forward - Using Big Data Where We Can

Very Large Corpus of Training Examples
Crowd-sourced (ImageNet)

VGG Net

good features

`bottle`

[Burchfiel et al., RSS’17]

Moving Forward - Using Big Data Where We Can

Corpus of Object Representations from Exploratory Behaviors

→

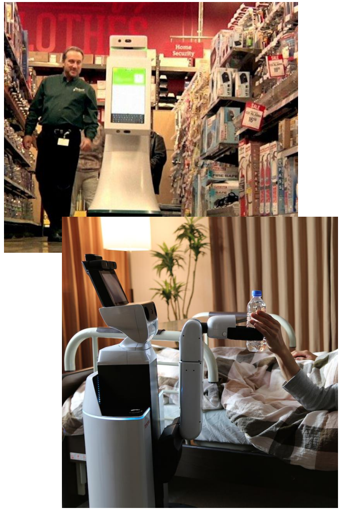
Latent Representations
Autoencoders
GANs
....

good features? →

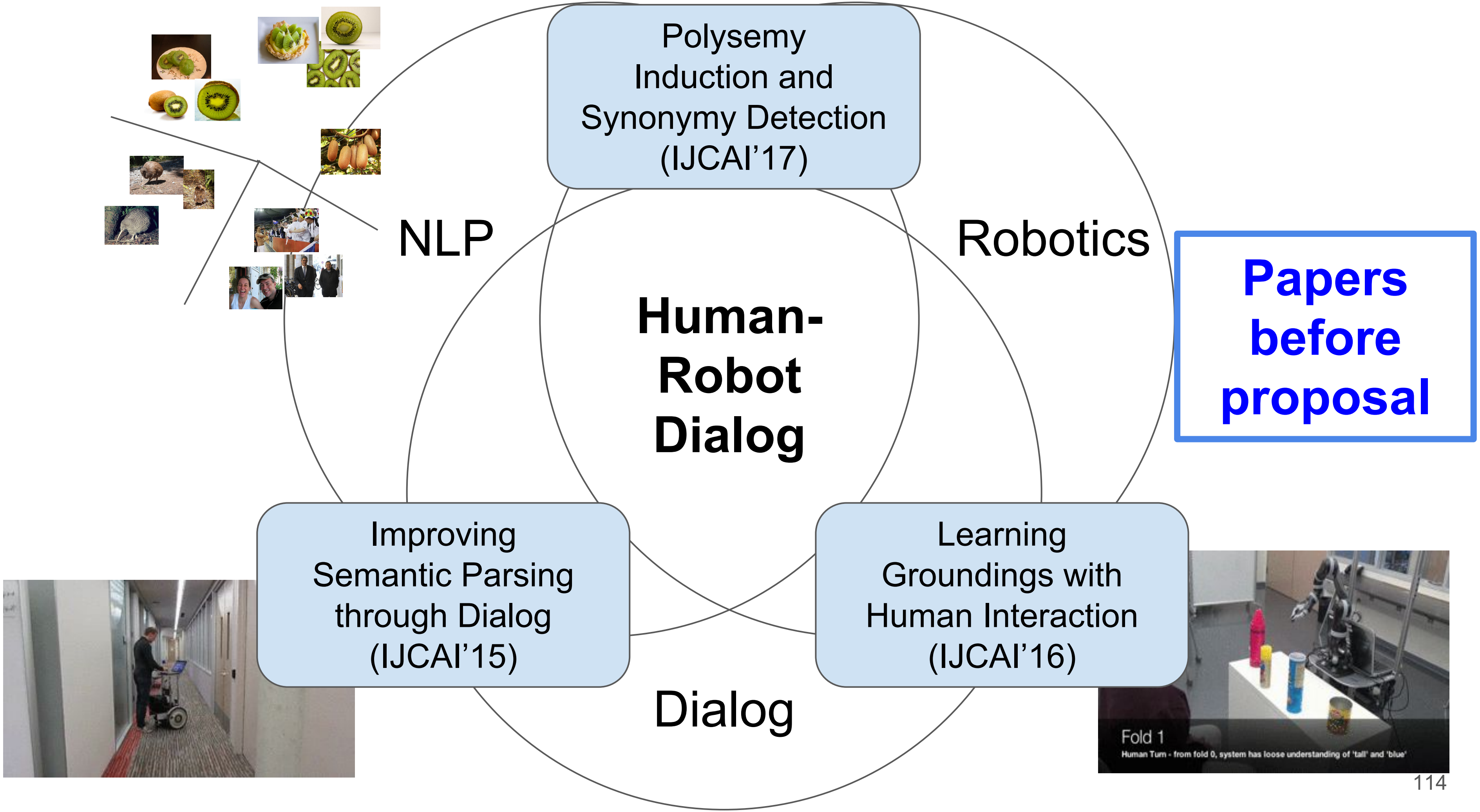
Moving Forward - Transfer Learning

Corpus of Human-Robot Dialogs

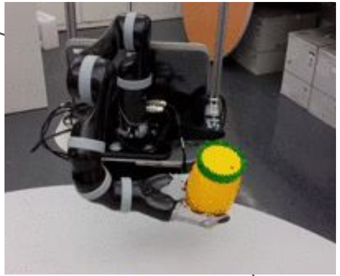
Similar domain shared commands
Sharing object representations



User ⟷ Robot → Robot Behavior


Polysemy Induction and Synonymy Detection (IJCAI’17)
NLP
Robotics
Human-Robot Dialog
Improving Semantic Parsing through Dialog (IJCAI’15)
Learning Groundings with Human Interaction (IJCAI’16)
Dialog
Papers before proposal
Fold 1
Human Turn - from fold 0, system has loose understanding of 'tall' and 'blue'

Faster Object Exploration for Grounding (AAAI'18)



NLP

Robotics

Papers since proposal

Jointly Improving Parsing & Perception (*in submission*)

Learning Groundings with Opportunistic Active Learning (CoRL'17)

Dialog

Acknowledgments

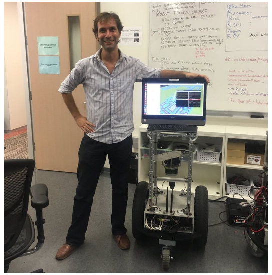
Ray
Mooney

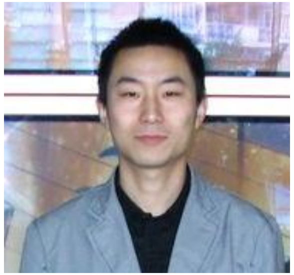
Peter
Stone

Scott
Niekum

Stefanie
Tellex

Acknowledgments


Jivko Sinapov


Shiqi Zhang


Aishwarya Padmakumar


Piyush Khandelwal


Rodolfo Corona


Harel Yedidsion

Justin Hart


Subhashini Venugopalan


Yuqian Jiang


Nick Walker

- *Jointly Improving Parsing and Perception for Natural Language Commands through Human-Robot Dialog*.
  Jesse Thomason, Aishwarya Padmakumar, Jivko Sinapov, Nick Walker, Harel Yedidsion, Justin Hart, Peter Stone, Raymond J. Mooney. (*in submission*)
- *Guiding Exploratory Behaviors for Multi-Modal Grounding of Linguistic Descriptions*.
  Jesse Thomason, Jivko Sinapov, Raymond J. Mooney, and Peter Stone. AAAI'18.
- *Improving Black-box Speech Recognition using Semantic Parsing*.
  Rodolfo Corona, **Jesse Thomason**, and Raymond J. Mooney. IJCNLP'17.
- *Opportunistic Active Learning for Grounding Natural Language Descriptions*.
  Jesse Thomason, Aishwarya Padmakumar, Jivko Sinapov, Justin Hart, Peter Stone, and Raymond J. Mooney. CoRL'17.
- *Multi-Modal Word Synset Induction*.
  Jesse Thomason and Raymond J. Mooney. IJCAI'17.
- *Integrated Learning of Dialog Strategies and Semantic Parsing*.
  Aishwarya Padmakumar, **Jesse Thomason**, Raymond J. Mooney. EACL'17.
- *BWIBots: A platform for bridging the gap between AI and human--robot interaction research*.
  Piyush Khandelwal, Shiqi Zhang, Jivko Sinapov, Matteo Leonetti, **Jesse Thomason**, Fangkai Yang, Ilaria Gori, Maxwell Svetlik, Priyanka Khante, Vladimir Lifschitz, J. K. Aggarwal, Raymond Mooney, and Peter Stone. IJRR'17.
- *Learning Multi-Modal Grounded Linguistic Semantics by Playing "I Spy"*.
  Jesse Thomason, Jivko Sinapov, Maxwell Svetlik, Peter Stone, and Raymond J. Mooney. IJCAI'16.
- *Learning to Interpret Natural Language Commands through Human-Robot Dialog*.
  Jesse Thomason, Shiqi Zhang, Raymond J. Mooney, and Peter Stone. IJCAI'15.

Graded Adjectives

- Think of gradation as a form of polysemy
- Semantic parser can use surrounding context
- Re-ranking of parses, as discussed, can help disambiguate

words

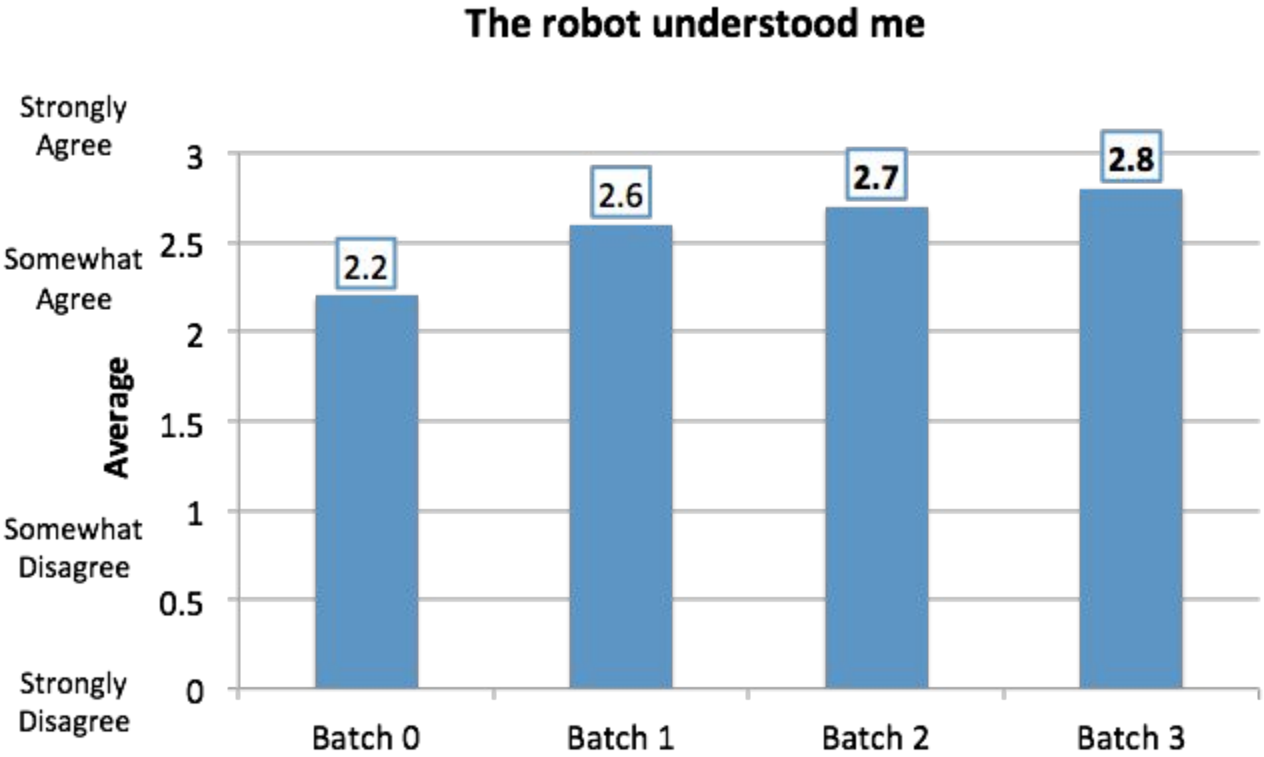
words

predicates

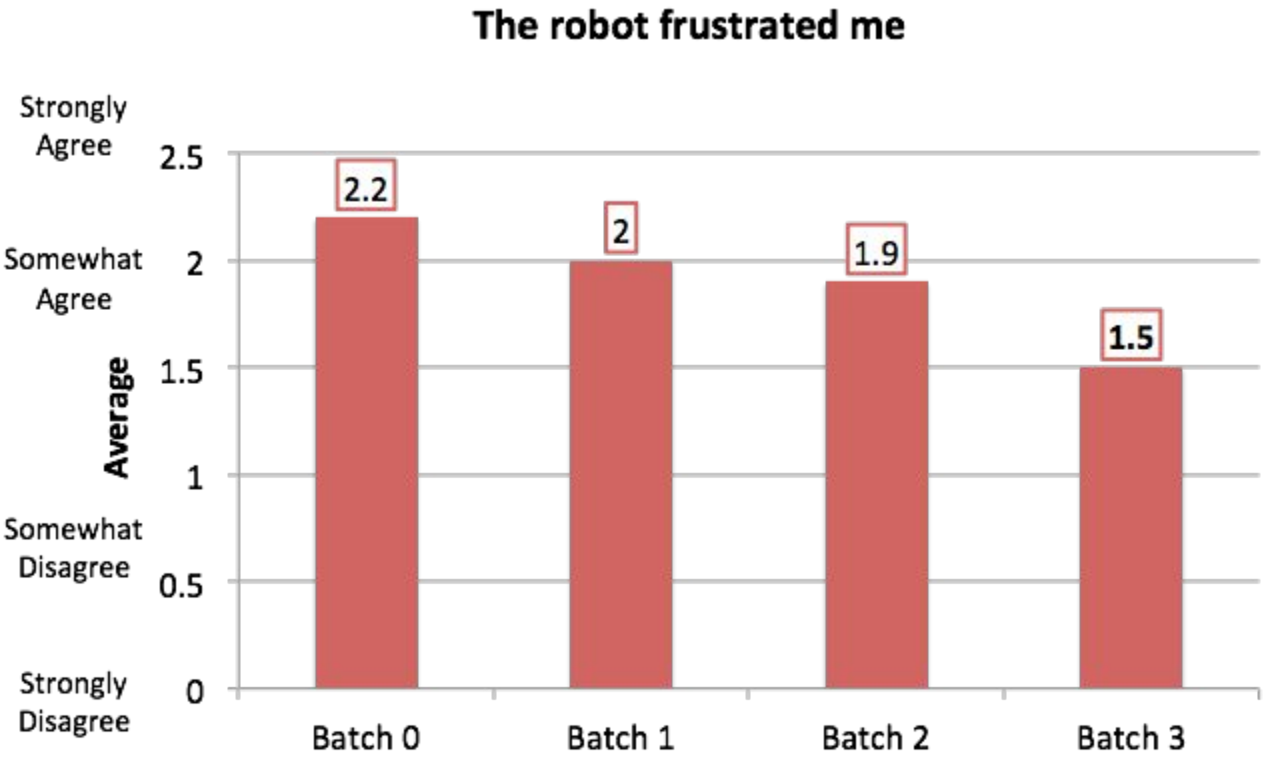
Comparative Adjectives

- E.g. “taller”, “heavier”; take two arguments: obj1, obj2
- Train classifier on the feature differences between obj1, obj2
- Can otherwise be handled with existing architecture
- Superlatives: majority winner object in pairwise comparative

Mechanical Turk Qualitative Results



Mechanical Turk Qualitative Results



Multi-modal Representation

[Thomason et al., IJCAI’17;
Deerwester et al., 1990;
Simonyan and Zisserman, CoRR’14]

- LSA embedding text features; VGG image features

Bat

“... most of the oldest known, definitely identified bat fossils were already very similar to modern microbats … ”

Bat

“... a baseball bat is divided into several regions …”

Bat

“... about 70% of bat species are insectivores … “

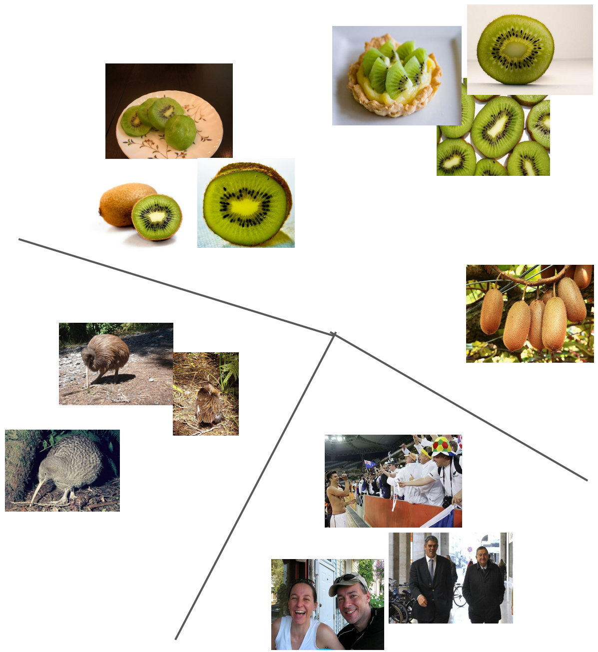
Bat

“... hickory has fallen into disfavor over its greater weight, which slows down bat speed … “

Technical Contributions

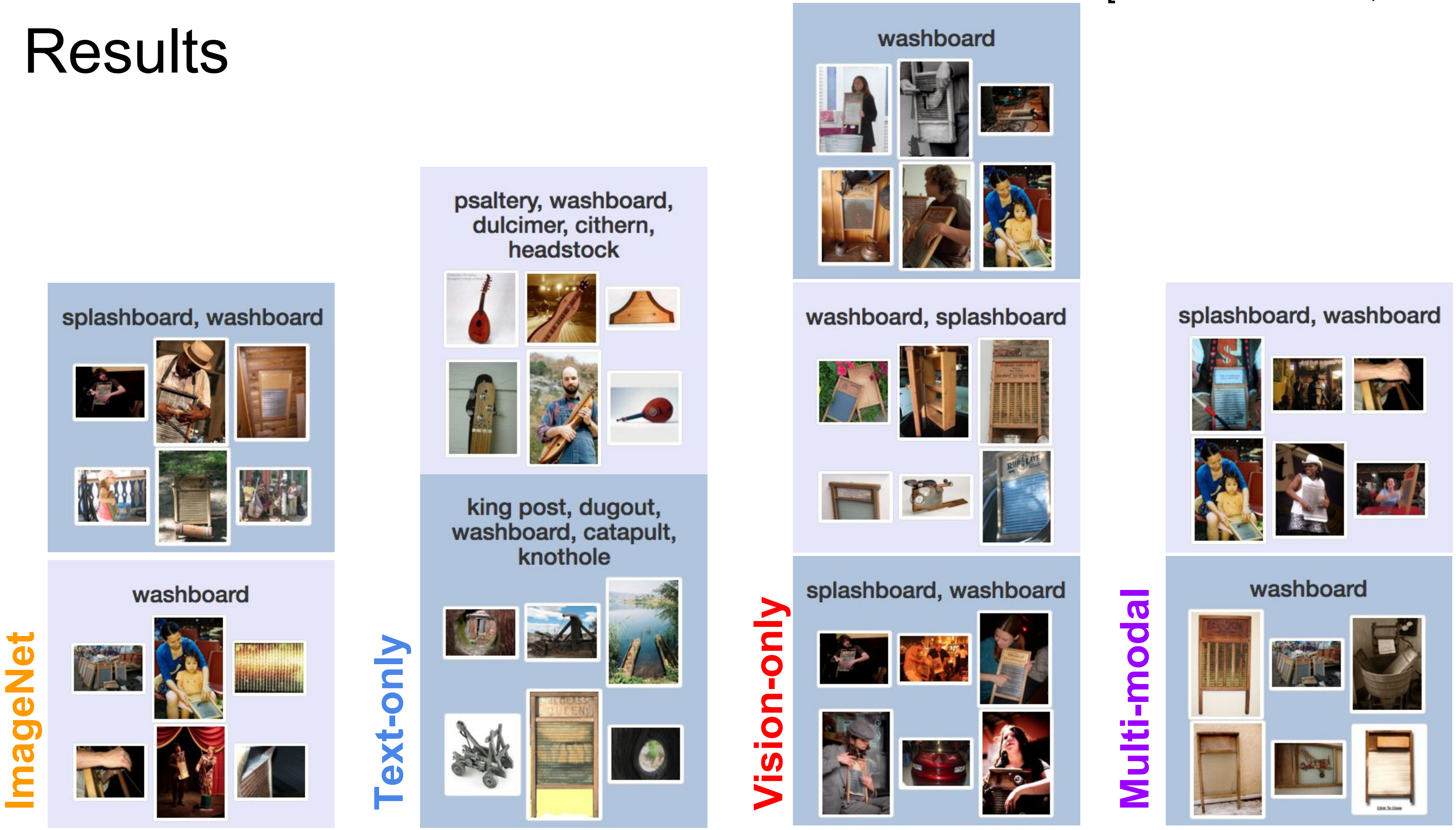
- Perform **unsupervised, multi-modal** sense induction and synonymy detection

- Create an ImageNet-like resource **without manual annotation**.

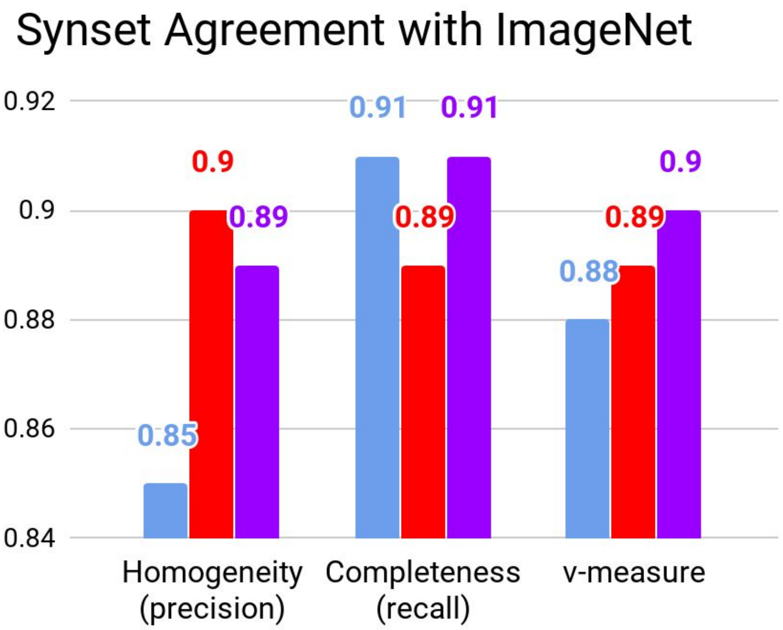


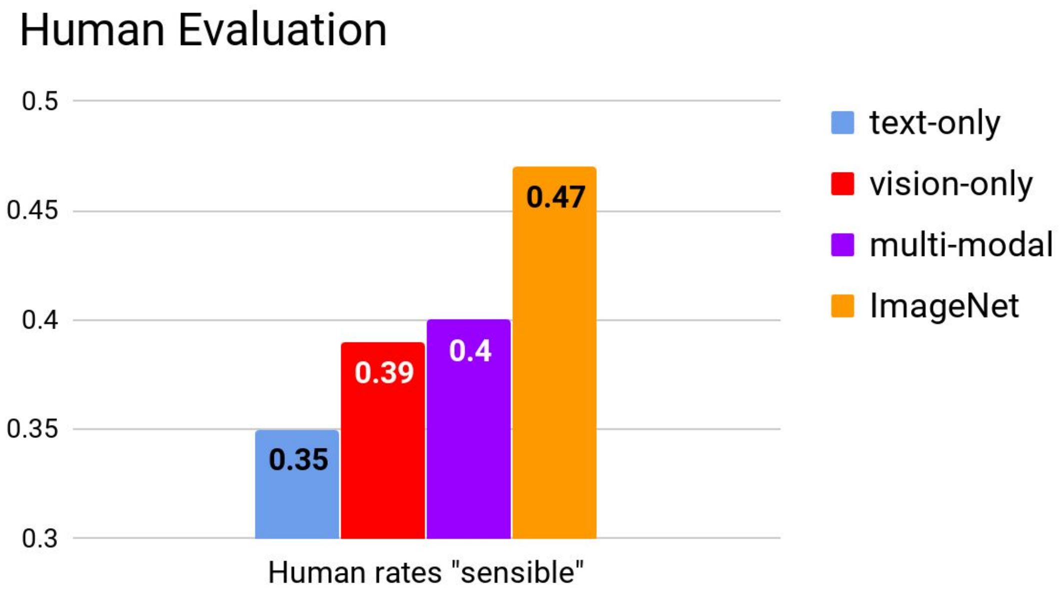
[Thomason et al., IJCAI'17]

Results

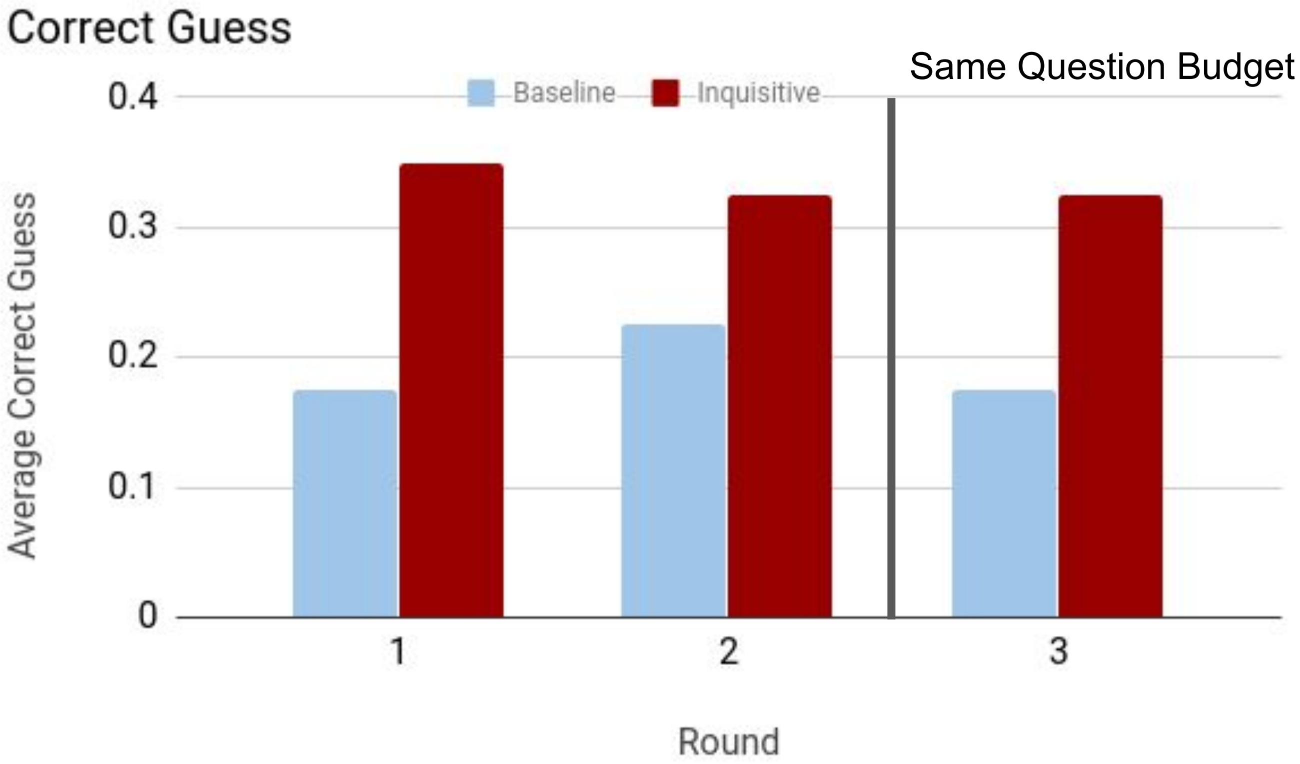


Results

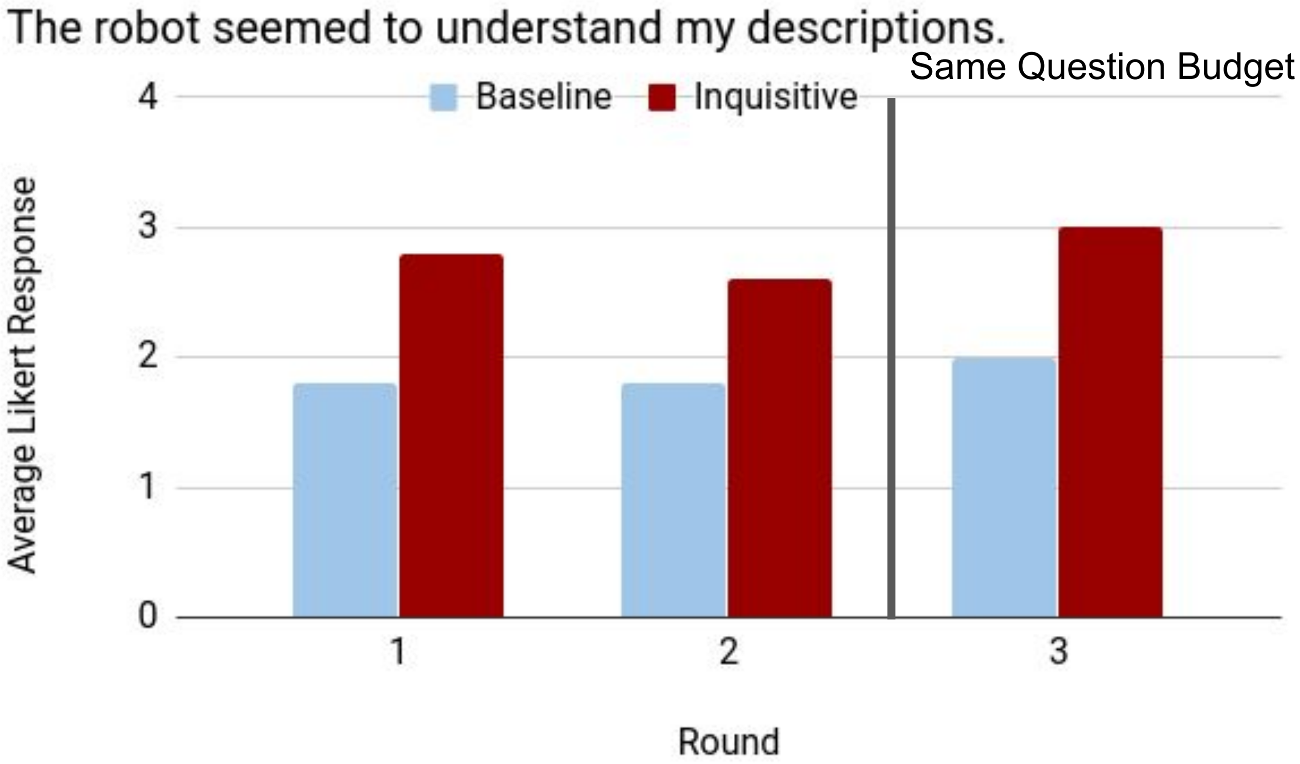

Synset Agreement with ImageNet
0.92
0.9
0.88
0.86
0.84
0.91
0.91
0.9
0.89
0.9
0.89
0.89
0.88
0.85
Homogeneity (precision)
Completeness (recall)
v-measure


Human Evaluation
0.5
0.45
0.4
0.35
0.3
0.47
0.4
0.39
0.35
Human rates "sensible"
text-only
vision-only
multi-modal
ImageNet

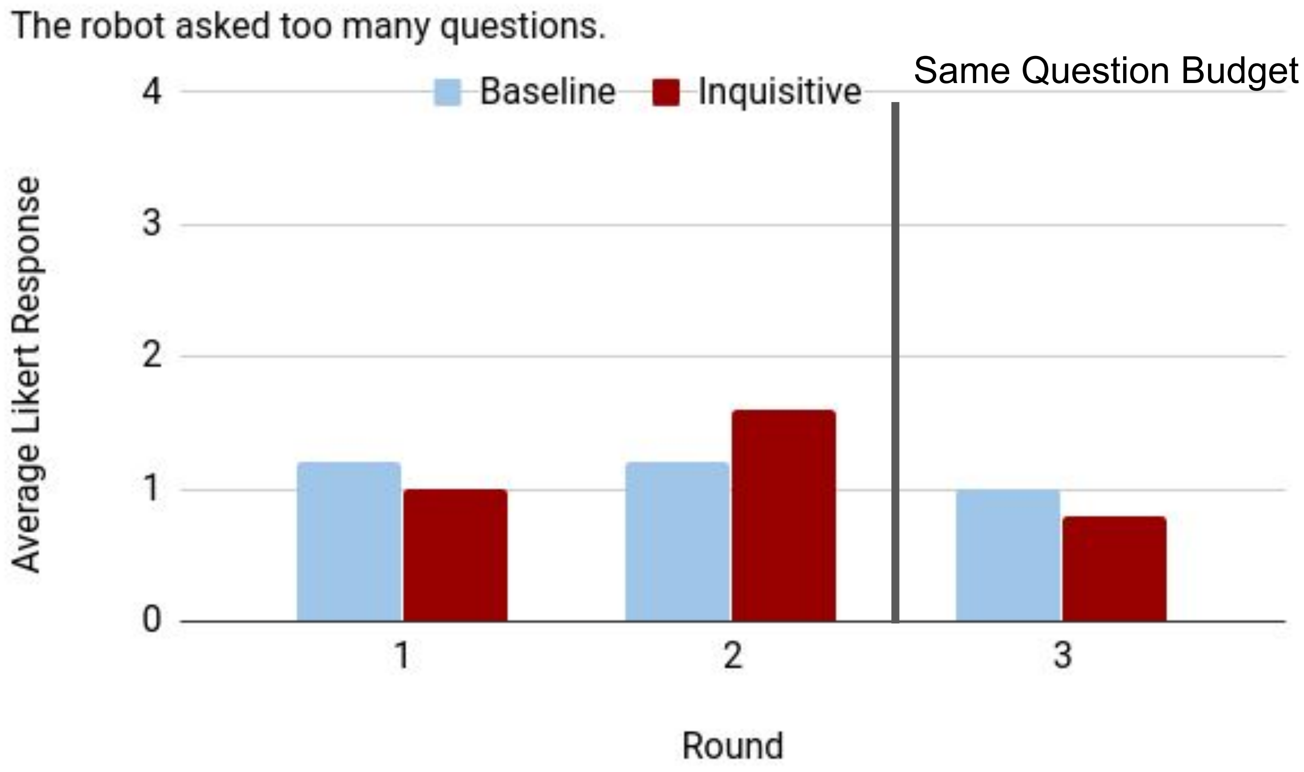
Results - Correct Object Selected



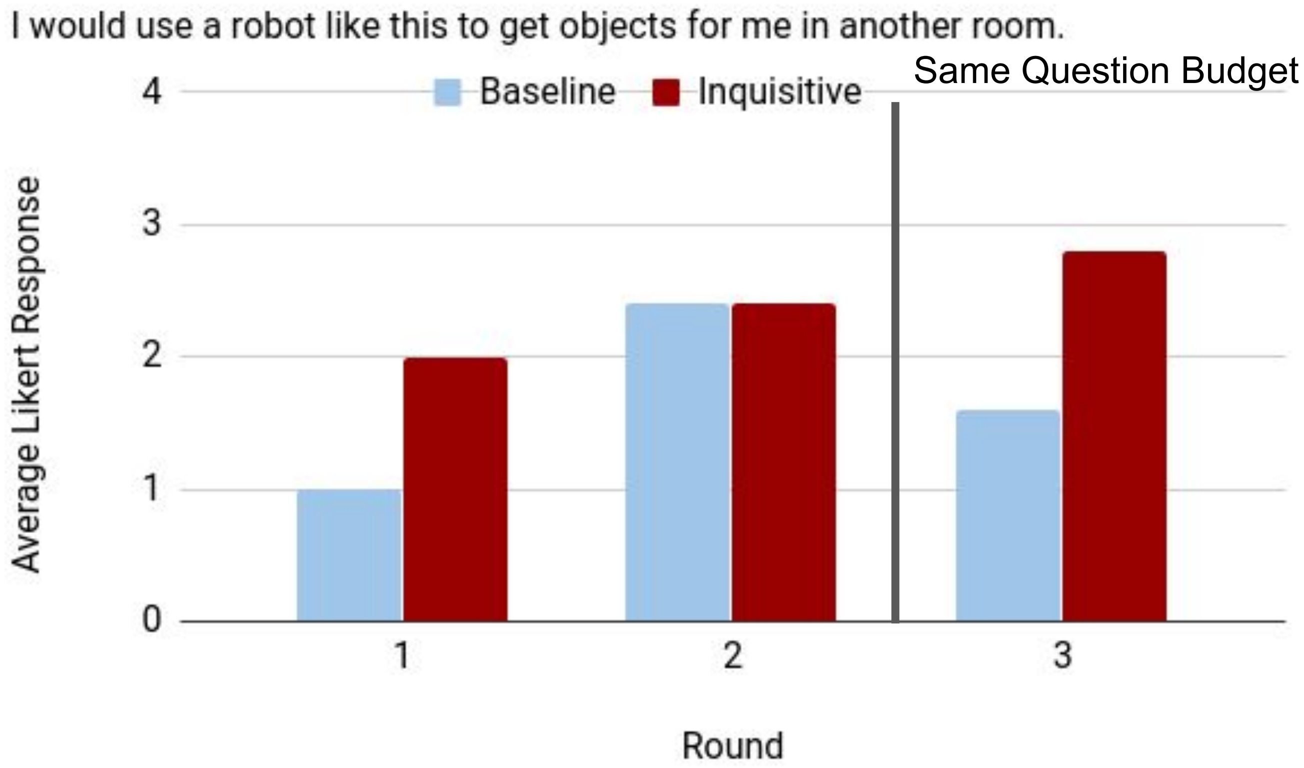
Results - Users Feeling Understood



Results - Users Annoyed



Results - Viable for Deployment



[Liang and Potts, Annual Review of Linguistics'15]

Learning from Denotations

- Given utterance-denotation pair, find a semantic form that is plausible for both

("rattling container", )

Learning from Denotations

- Use the parser to produce a beam of parses
- Use the grounder to find the denotations of those parses

("rattling container", )

Learning from Denotations

Semantic Parser

("rattling container" ,

the(λy.(rattling(y)))
the(λy.(rattling(y) $\wedge$ container(y)))
the(λy.(container(y)))
rattling $\wedge$ container
...

, )

Learning from Denotations

Semantic Parser

Grounding Modules

("rattling container" ,

$\text{the}(\lambda y.(\text{rattling}(y)))$

$\text{the}(\lambda y.(\text{rattling}(y) \wedge \text{container}(y)))$

$\text{the}(\lambda y.(\text{container}(y)))$

$\text{rattling} \wedge \text{container}$

...

, )

Learning from Denotations

("rattling container" , the(λy.(rattling(y) ∧ container(y))) , )

Learning from Denotations

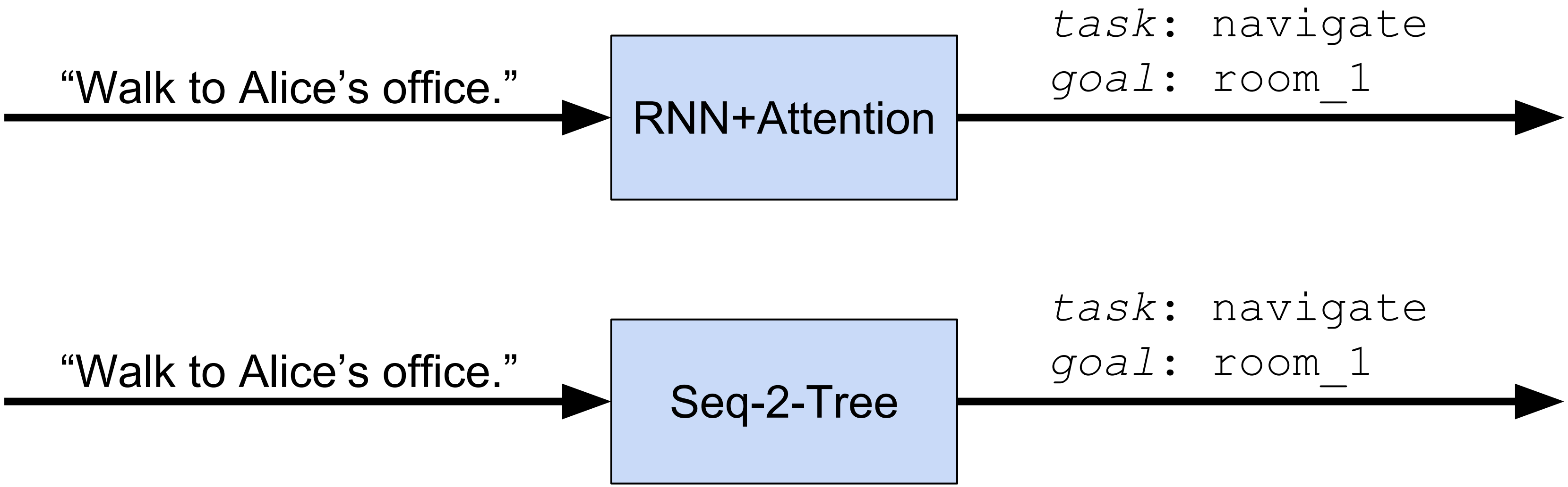
$$\left(\text{“rattling container”} , \ \text{the}(\lambda y.(\text{rattling}(y) \land \text{container}(y))) \right)$$

[*ongoing*]

[Jia, ACL’16; Dong, ACL’16]

Neural Parsing Methods

- Recurrent Neural Networks (RNNs) with Attention
- Sequence-to-Tree encoder-decoder networks

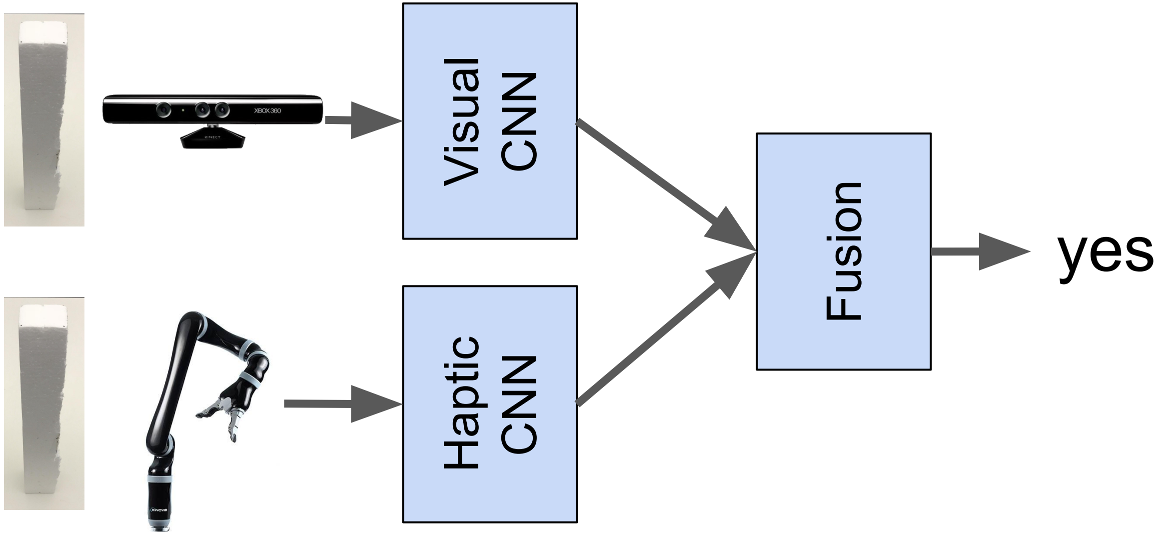


[Gao, ICRA’16]

Neural Perception Models

- Compress high-dimensional sensorimotor context information using Convolutional Neural Networks (CNNs)

textured?



[Das et al., CVPR'18]

Embodied Question Answering

- End-to-end deep model for joint parsing and perception