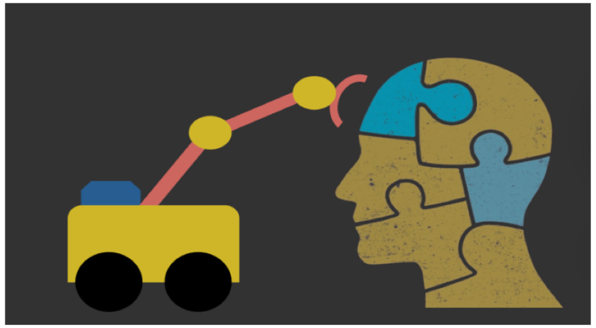
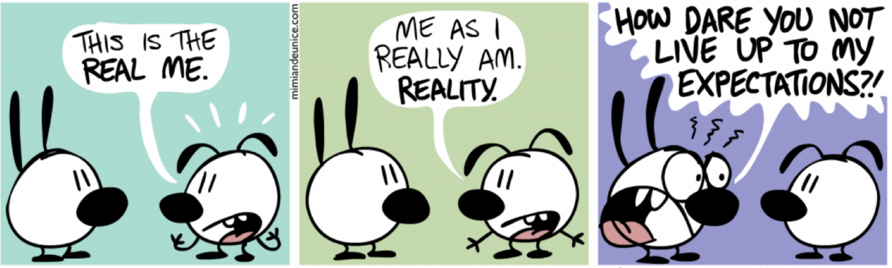
EAGER: Reconciling Model Discrepancies in Human-Robot Teams

Yu ("Tony") Zhang
Arizona State University, USA
yzhan442@asu.edu



Cooperative Robotic Systems

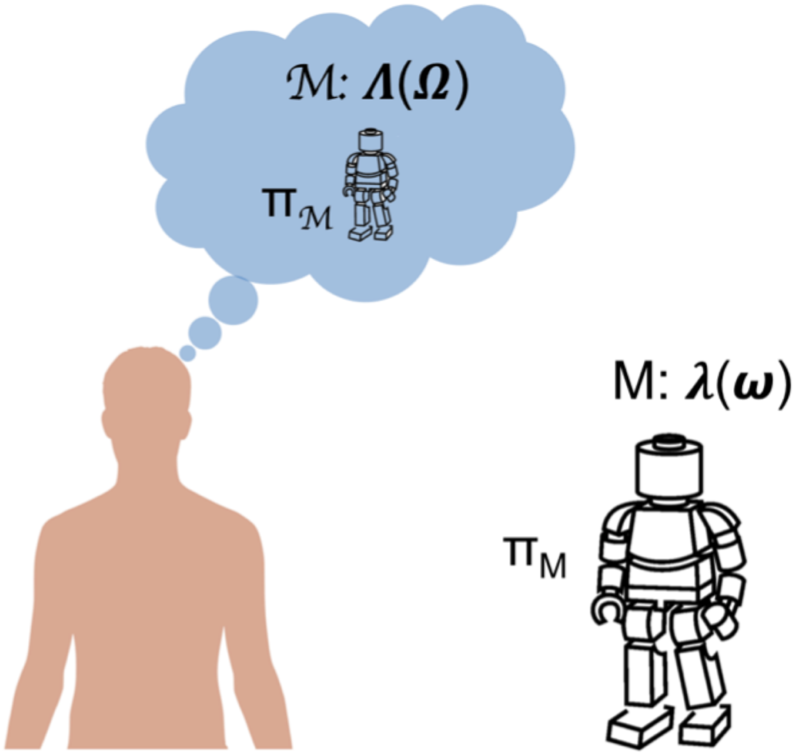
Motivation



from *mimiandeunice.com*

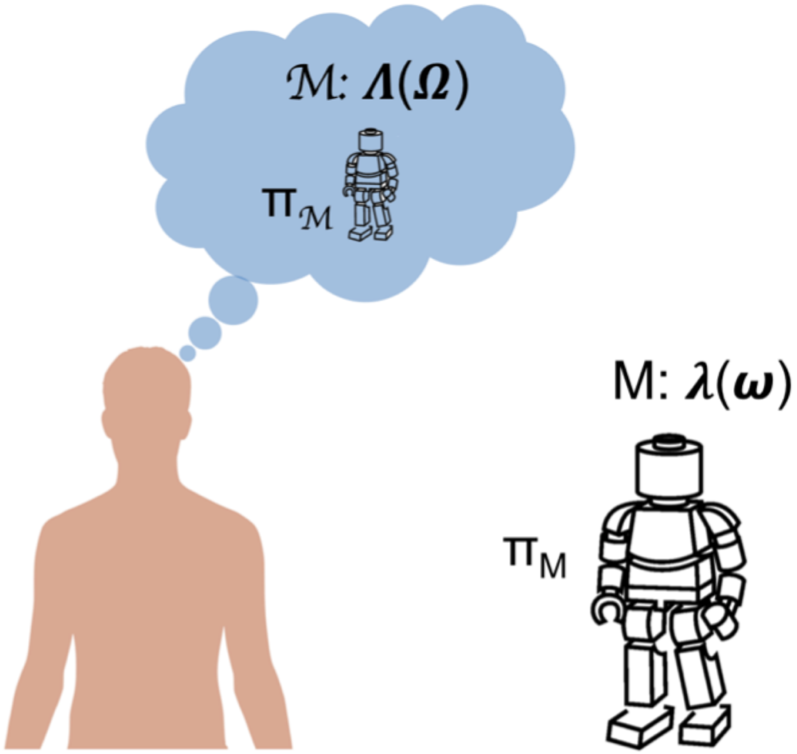
- Teammates have many conscious and subconscious ***expectations*** of others in terms of others' plans or behaviors
- The expected model (EM) and actual model (DM) may differ, leading to unmatched expectations, loss of situation awareness and trust

➢ This calls for general methods for ***model reconciliation***



Taxonomy of model reconciliation

Type	Align M with $\mathcal{M}$		Align $\mathcal{M}$ with M	
		Exp. AI	Exp. AI	Exp. AI
	Explicit	*Implicit*	*Explicit*	*Implicit*
Goal				
	Rwd Lrn [31, 33] Pln Rec [22, 73] Ins Und [17, 123]	IRL [5, 138]	Inv Grd [122] Int Prj [7, 56] Gol Aug [23]	Lgb Pln [34] Pln CoE [77] Com Act [75]
Behavior				
	Imi Lrn [107, 111] HuW Pln [21, 26]	NA	Crs Trn [94]	NA
Domain				
	Opn Wrd [120] Dom Lrn [135]	**RT 1, 2**	**RT 1, 2** Exu Gen* [53, 60]	Exp Inc [79] Com Act [75]
Cognition				
	Met Rsn [110] Com Rat [47]	DvP [92]	NA	**RT 3**



Prior work has focused mostly on model reconciliation with domain dynamics with other parts fixed or assumed.

Progress thus far

Model Reconciliation

Domain model reconciliation

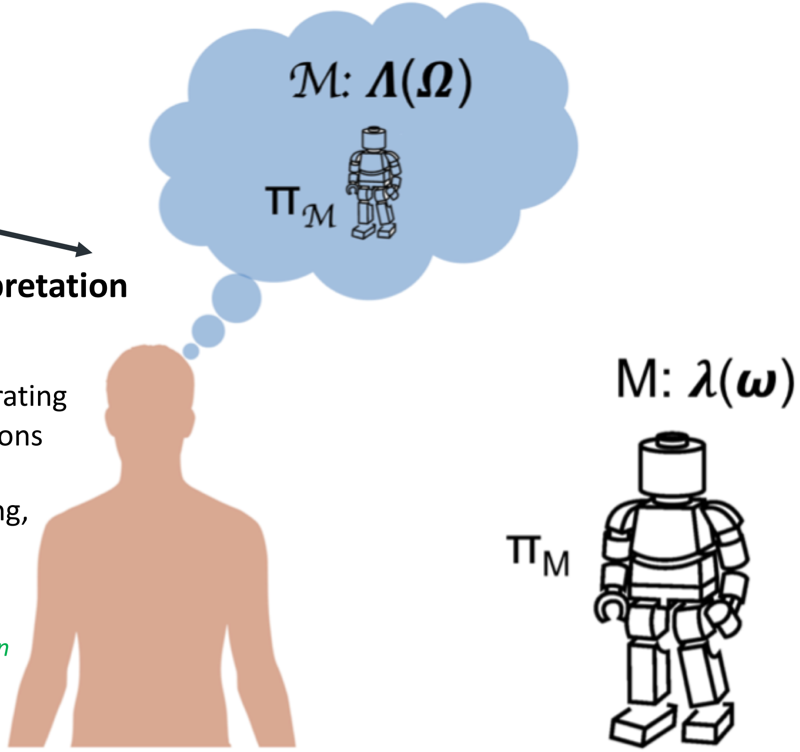
Cognitive model reconciliation

Goal reconciliation

implicit

Minimize content to convey

Reduce interpretation effort



Intellectual Merit

Explicability as Minimizing Distance from Expected Behavior, ***AAMAS 2019***

- *Approximate explicability measure using plan distance metrics for model reconciliation*

What Is It You Really Want of Me? Generalized Reward Learning with Biased Beliefs about Domain Dynamics, ***AAAI 2020 Oral***

- *Generalized reward learning with model reconciliation*

Online Explanation Generation for Human-Robot Teaming ***ICAPS WS 2019,*** Ready for *IROS*

- *Timing of model reconciliation for explanation generation*

Order Matters: Generating Progressive Explanations for Planning Tasks in Human-Robot Teaming, Under review

- *Ordering of model reconciliation for explanation generation*

Broader Impacts

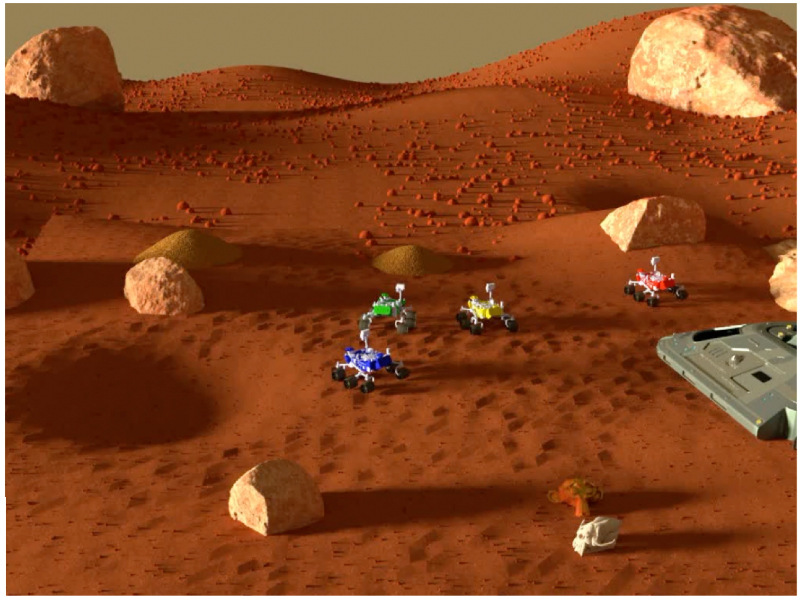
1) **Dissemination**: Keynote speaker at Intel, Chandler, AZ on "*Challenges in Cognitive Human-robot Teaming*"; organizing the Southwest Robotics Symposium
2) **K12 and high school**: Special Award Judge at Intel ISEF, 2016 and again in 2019, representing AAAI
3) **Undergraduate education**: Freshman E2
4) **Graduate education**: Incorporated research topics in Human-Aware Robotics class in Spring 2019
5) **Human-AI interaction**: model reconciliation method can push forward value alignment and safe AI

Awards:

Nominated for outstanding 1) master's mentor and 2) doctoral mentor

Student: 1) Outstanding Graduate Student award 2) Graduate fellowships 3) Summer schools

Reconciling cognitive models

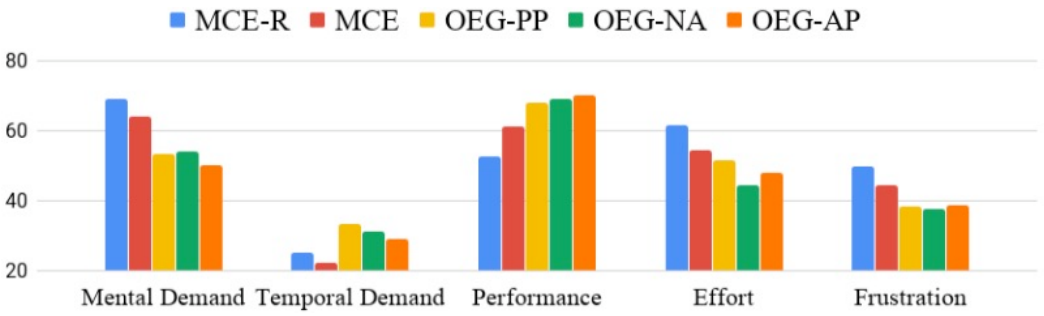


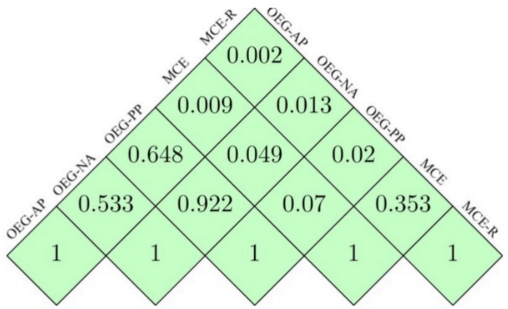
Timing for model reconciliation as explanation generation:

- ➢ **Breaking down a complex explanation into smaller parts and convey them in an online fashion, to save short-term or work memory**

	MCE-R	MCE	OEG-PP	OEG-NA	OEG-AP	Truth
Accuracy	0.746	0.804	**0.858**	**0.852**	**0.872**	
# Actions	8.789	7.263	**5.250**	**5.330**	**4.940**	2.0/30

- Identify long-term dependencies using planning methods
- Minimize information content to convey for model reconciliation
- Developed and evaluated three different online explanation generation methods





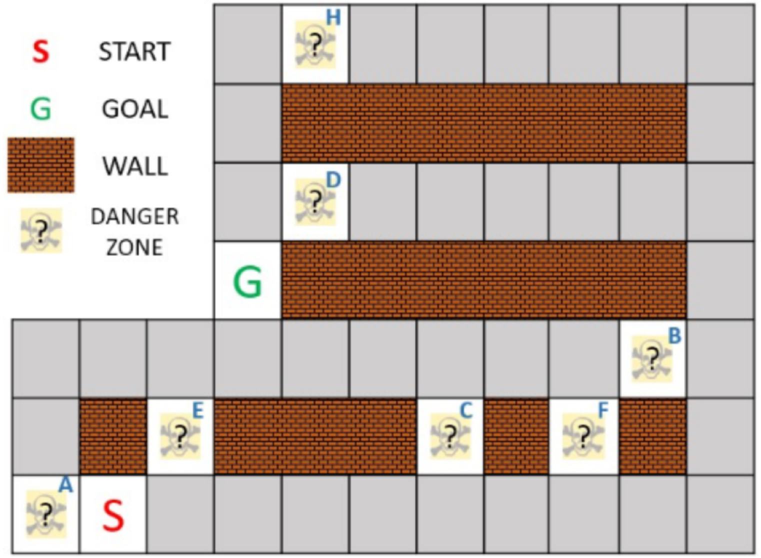
M. Zakershahrak, Z. Gong, N. Sadassivam, A. Hanni, Y. Zhang, *Online Explanation Generation for Human-Robot Teaming*, in ICAPS Workshop on Explainable AI Planning 2019.

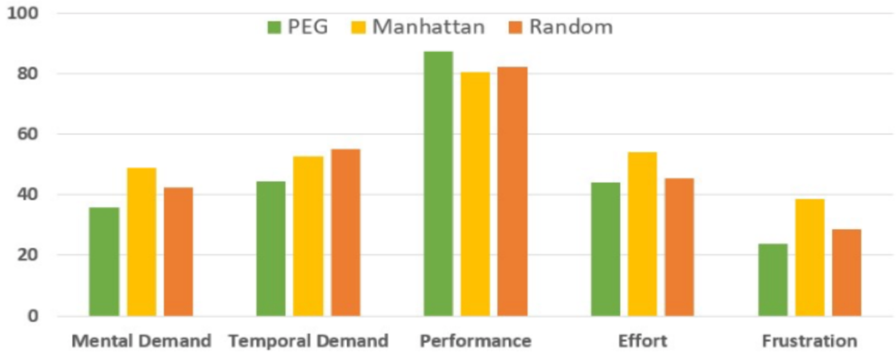
Reconciling cognitive models

Ordering for model reconciliation as explanation generation:

- ➢ **Identify the desirable order of conveying information to minimize interpretation effort. Order matters!**

- Developed learning method to associate features with ease of understanding in a sequential model
- Among these features, plan editing distance played an important role
- Evaluated explanation generation with only plan editing distance and results looked promising





M. Zakershahrak, R. Shashank, A. Sharma, Z. Gong, Y. Zhang, *Order Matters: Generating Progressive Explanations for Planning Tasks in Human-Robot Teaming*, under review

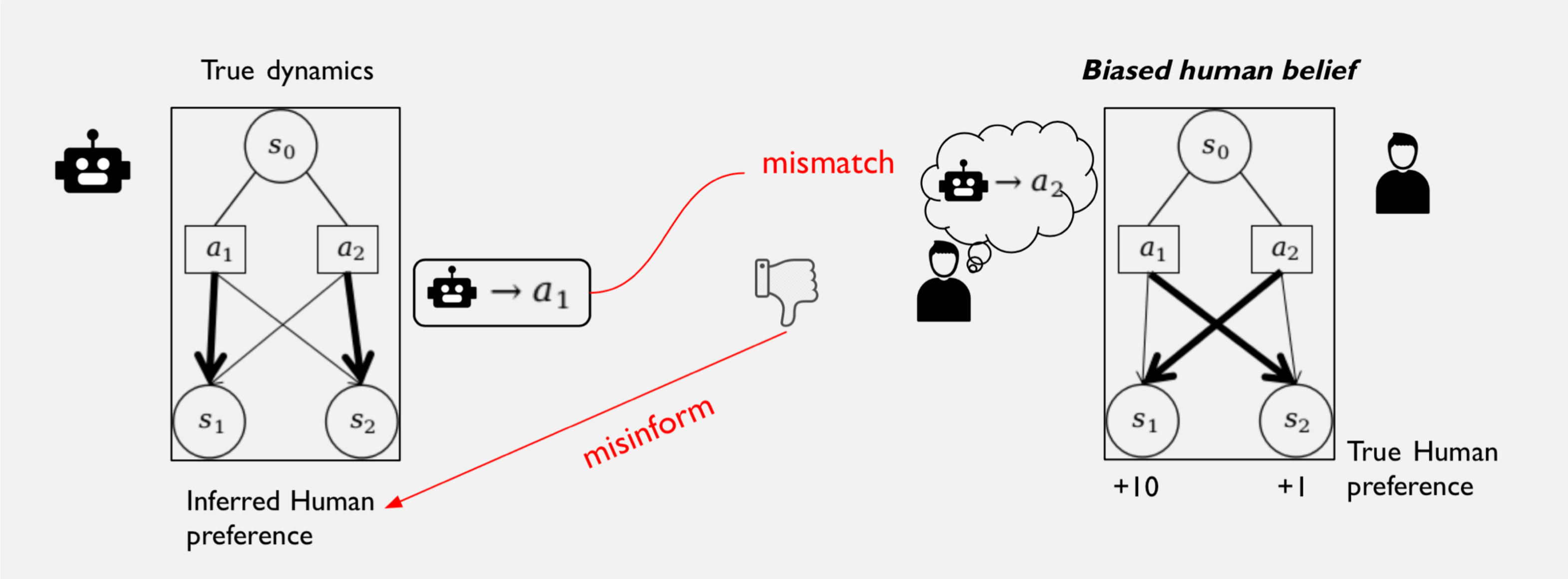
Domain model and goal reconciliation

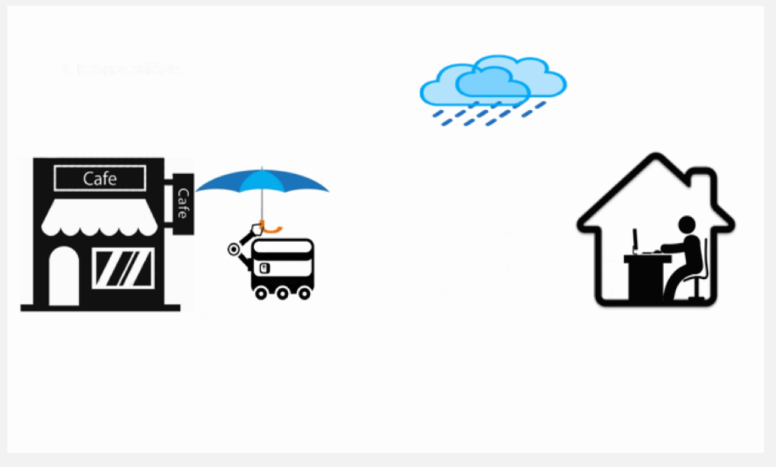
Problem settings:

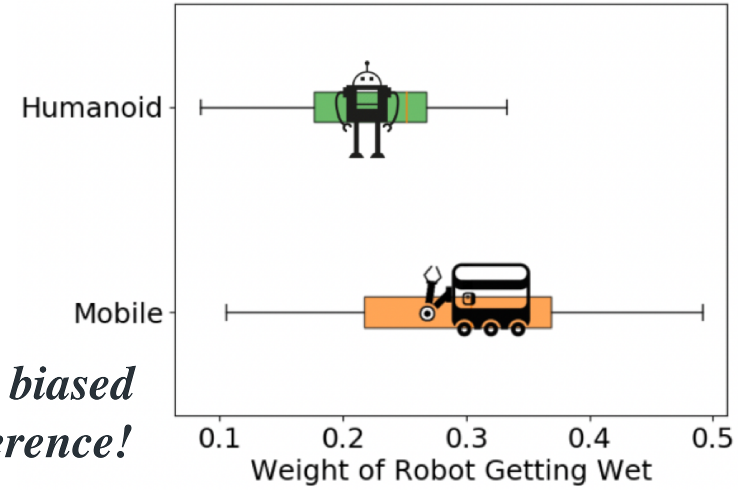
- (Non-expert) Human user observes the robot's behavior and provides feedbacks.
- Robot learns the human preferences from the feedbacks.

Z. Gong and Y. Zhang, *What Is It You Really Want of Me? Generalized Reward Learning with Biased Beliefs about Domain Dynamics, AAAI 2020*

Domain model and goal reconciliation







Reward learning without considering a biased belief learned the opposite human preference!

Z. Gong and Y. Zhang, *What Is It You Really Want of Me? Generalized Reward Learning with Biased Beliefs about Domain Dynamics, AAAI 2020*

Intellectual merit

- Recover the human's preferences or recover the domain dynamics
  - Inverse Reinforcement Learning and Reward Learning
    (Russell 1998, Ng et al. 2000, Abbeel et al. 2004, Ziebart et al. 2008, Daniel et al. 2014)
  - Dynamics learning given the human's reward function
    (Reddy et al. 2018; Unhelkar et al. 2019)
  - Simultaneous Estimation of Rewards and Dynamics
    (Herman et al. 2016)

- We generalize these learning methods to consider our problem settings:
  - ✓ The human may have a *biased hidden belief* about the robot's domain dynamics
  - ✓ The robot must infer about the human's preferences from their feedbacks only, subject to such biased beliefs

  [Note that, robot trajectories are not informative for learning the human biased belief.]

Z. Gong and Y. Zhang, *What Is It You Really Want of Me? Generalized Reward Learning with Biased Beliefs about Domain Dynamics, AAAI 2020*

Problem formulation

Given:

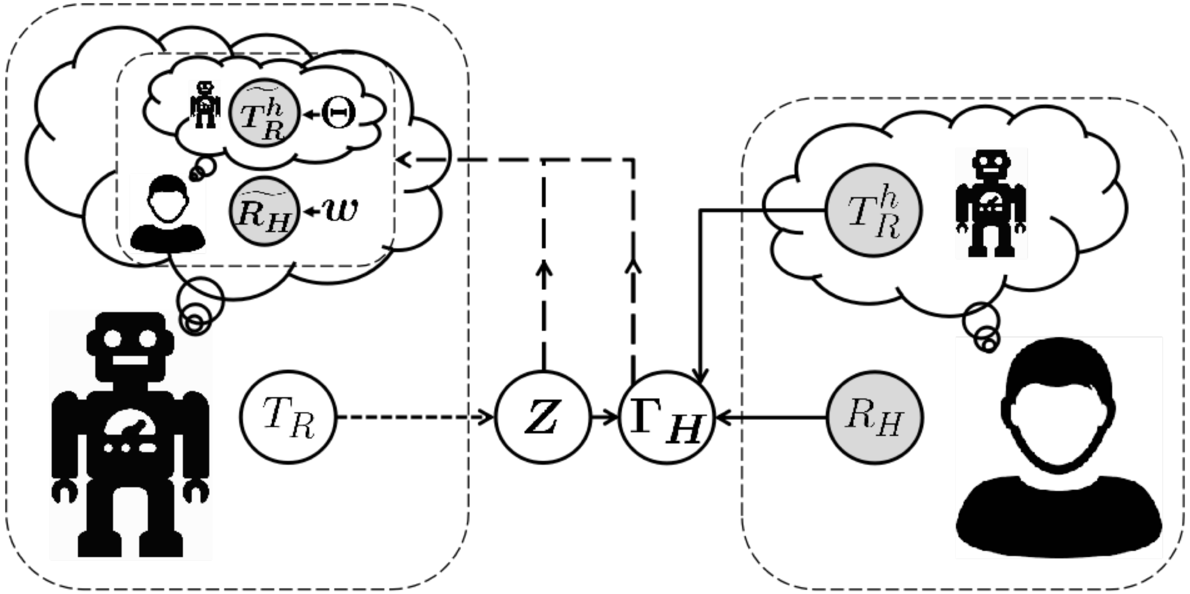
Robot's demonstrations $\boldsymbol{Z}$;

Human's ratings $\boldsymbol{\Gamma_H}$ for each instance in $\boldsymbol{Z}$.

To determine:

Human's true reward function $\boldsymbol{R_H}$;

Human's belief $\boldsymbol{T_R^h}$ about robot's domain dynamics.



In this work, our goal is to recover the **true reward function** and **biased beliefs** together.

Z. Gong and Y. Zhang, *What Is It You Really Want of Me? Generalized Reward Learning with Biased Beliefs about Domain Dynamics, AAAI 2020*

Problem solution

- Parameterization
  - Formulate the reward function R_H for a state s as follows:

$$R_H(s) = \boldsymbol{w} \cdot \Phi(s) \qquad \text{where} \quad \boldsymbol{w} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$$

  - Belief about robot domain dynamics are formulated using a set of Dirichlet distribution:

$$\boldsymbol{\Theta} = [\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_{|S|\times|A|}] \qquad \text{where} \quad \boldsymbol{\theta}_i \sim \mathrm{Dir}(\boldsymbol{\alpha}_i)$$

- Our goal is to learn the posterior distribution:

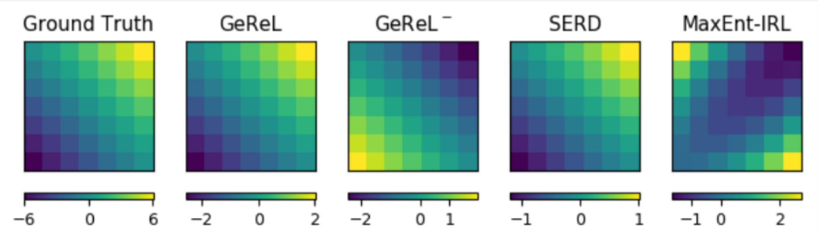
$$p(\boldsymbol{w}, \boldsymbol{\Theta} | \boldsymbol{\Gamma}_H, \boldsymbol{Z}) \text{ is approximated by } q(\boldsymbol{w}, \boldsymbol{\Theta} | \boldsymbol{\mu}, \boldsymbol{\mathcal{A}})$$

Variational posterior of the latent variables governed by $\boldsymbol{\mu}$ and $\boldsymbol{\mathcal{A}} = [\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_{|S|\times|A|}]$

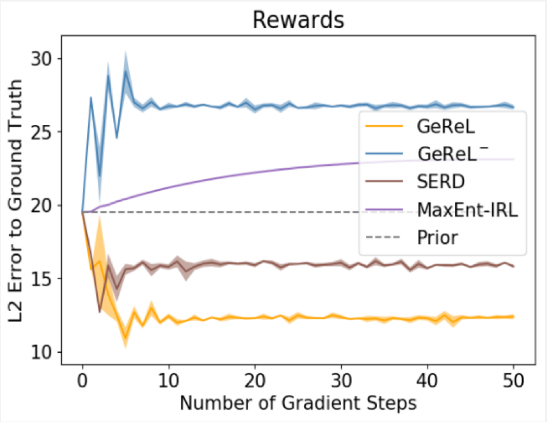
Z. Gong and Y. Zhang, *What Is It You Really Want of Me? Generalized Reward Learning with Biased Beliefs about Domain Dynamics, AAAI 2020*

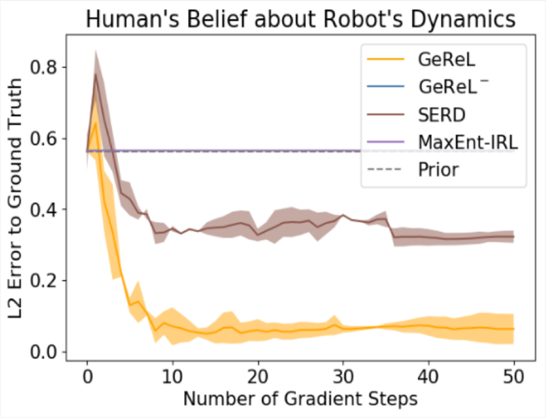
Results

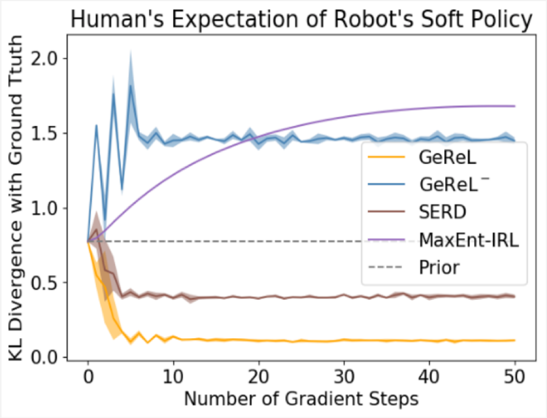
- **GeReL**: this work
- **GeReL$^-$**: basically uses GeReL without updating the domain dynamics.
- **Variant of SERD**: apply GeReL framework with soft Bellman equation.
- **MaxEnt-IRL**: demonstration generated using ground truth model.



Rewards learned by different approaches.



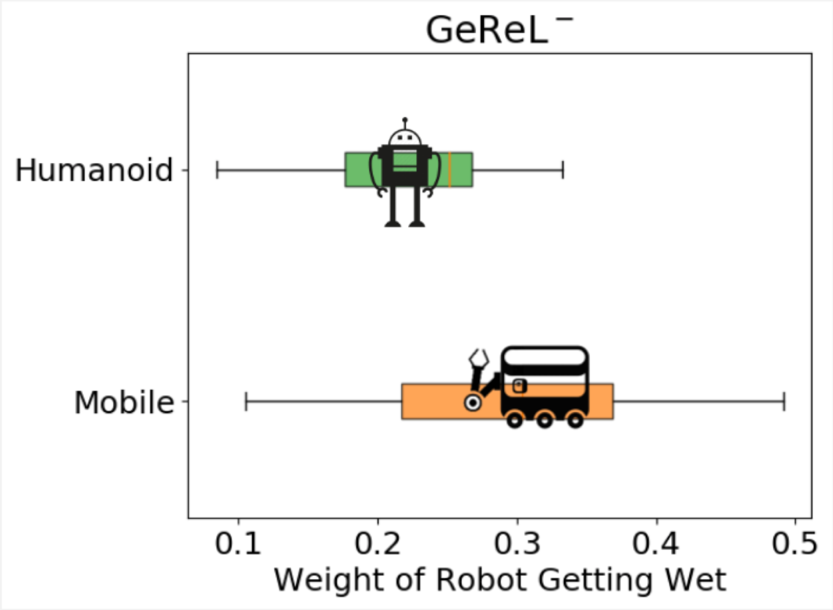


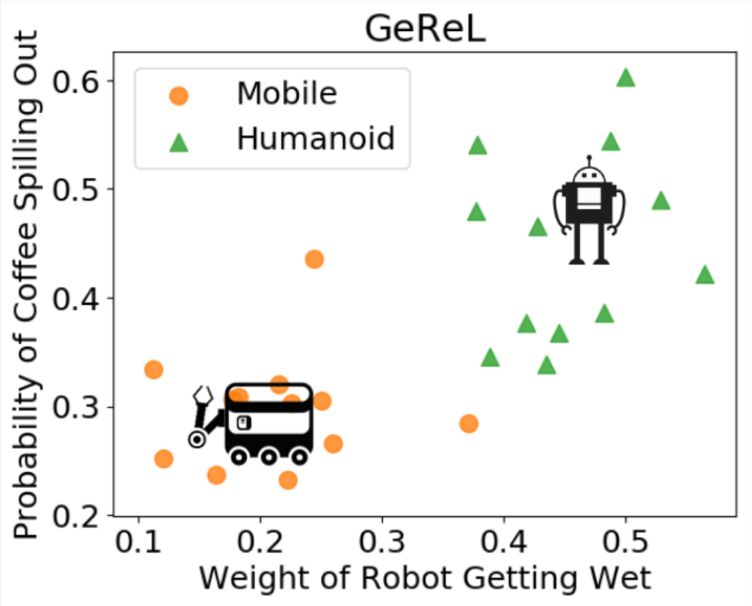


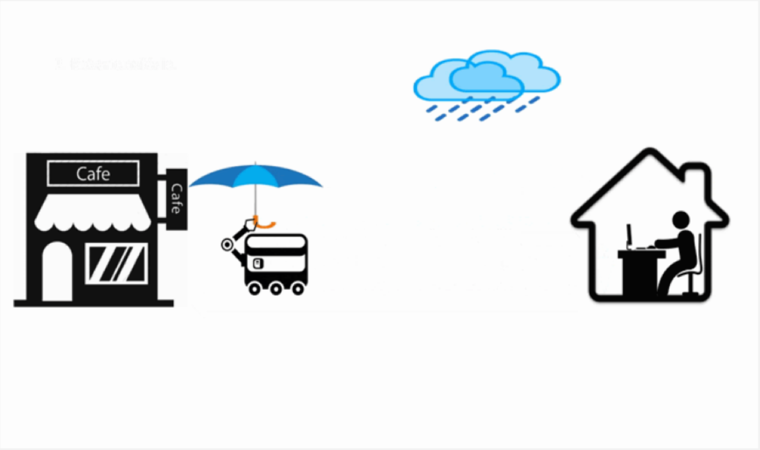
Z. Gong and Y. Zhang, *What Is It You Really Want of Me? Generalized Reward Learning with Biased Beliefs about Domain Dynamics, AAAI 2020*

Results

With our method







Z. Gong and Y. Zhang, *What Is It You Really Want of Me? Generalized Reward Learning with Biased Beliefs about Domain Dynamics, AAAI 2020*

Summary

Model Reconciliation

Domain model reconciliation

Goal reconciliation

Explicability as Minimizing Distance from Expected Behavior, ***AAMAS 2019***

- *Approximate explicability measure using plan distance metrics for model reconciliation*

What Is It You Really Want of Me? Generalized Reward Learning with Biased Beliefs about Domain Dynamics, ***AAAI 2020 Oral***

- *Generalized reward learning with model reconciliation*

Cognitive model reconciliation

implicit

Minimize content to convey

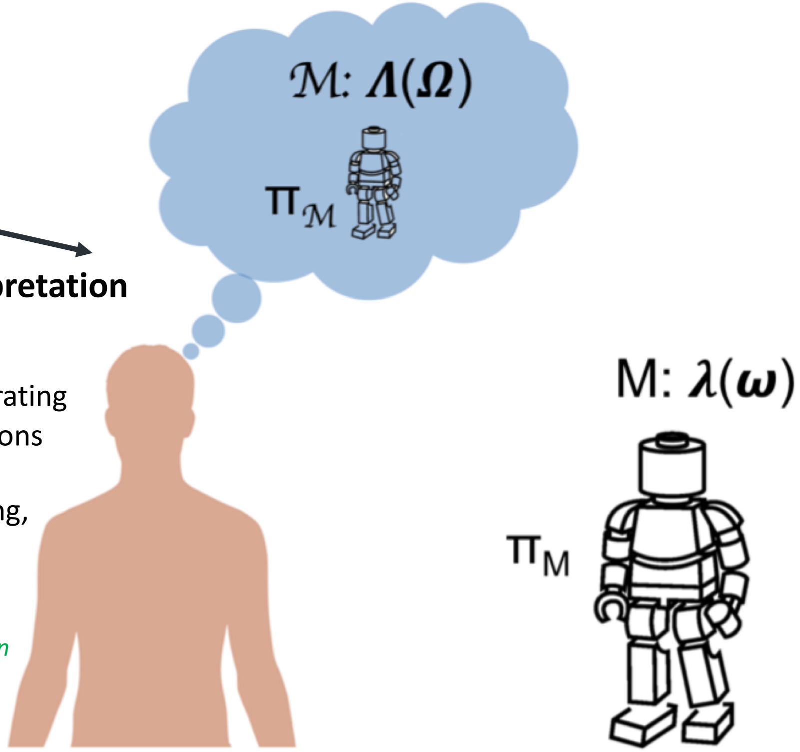
Online Explanation Generation for Human-Robot Teaming ***ICAPS WS 2019,*** Ready for *IROS*

- *Timing of model reconciliation for explanation generation*

Reduce interpretation effort

Order Matters: Generating Progressive Explanations for Planning Tasks in Human-Robot Teaming, Under review

- *Ordering of model reconciliation for explanation generation*



Future work

1. Long-term interaction with model reconciliation
2. Integration of model reconciliation methods
3. Investigate connections to value alignment and safe AI
4. Applications

Intellectual Merit

Broader Impacts

1) **Dissemination**: Keynote speaker at Intel, Chandler, AZ on "*Challenges in Cognitive Human-robot Teaming*"; organizing the Southwest Robotics Symposium
2) **K12 and high school**: Special Award Judge at Intel ISEF, 2016 and again in 2019, representing AAAI
3) **Undergraduate education**: Freshman E2
4) **Graduate education**: Incorporated research topics in Human-Aware Robotics class in Spring 2019
5) **Human-AI interaction**: model reconciliation method can push forward value alignment and safe AI

Awards:

Nominated for outstanding 1) master's mentor and 2) doctoral mentor

Student: 1) Outstanding Graduate Student award 2) Graduate fellowships 3) Summer schools