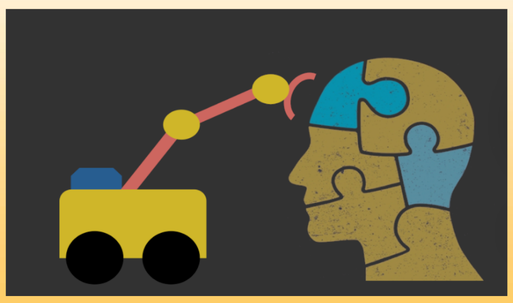
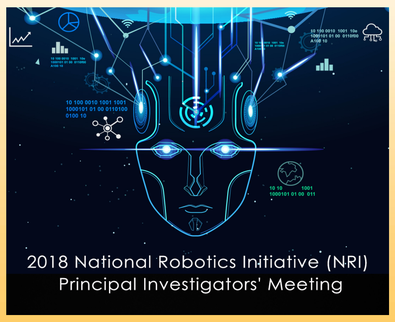
EAGER: Reconciling Model Discrepancies in Human-Robot Teams



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

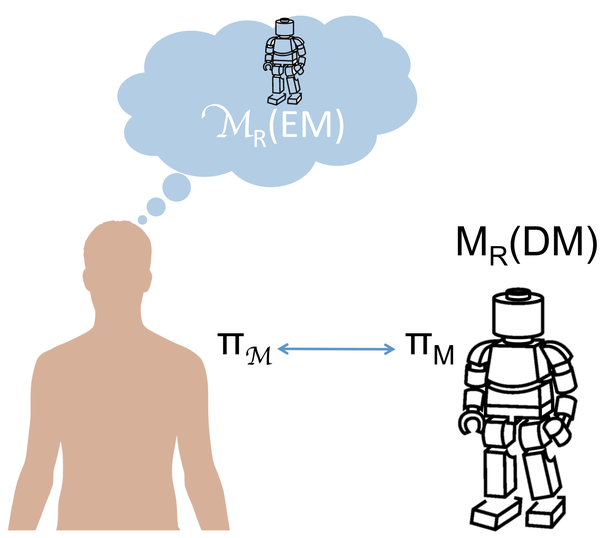


1. Motivation

- Teammates have many conscious and subconscious **expectations** of each other
- The expected (EM) and actual models (DM) may differ, leading to unmatched expectations, lost situation awareness and trust

➢ This is a problem of **model reconciliation**

A tragedy resulted as model differences:



2. Research Thrusts and Intellectual Merit

Two methods are proposed as **model reconciliation planning** (MRP) and **explanation generation** (MRE)

- **MRP**: biasing the robot's behavior to ***implicitly*** accommodate model differences – trade off between plan cost and explicability
- **MRE**: communicating to ***explicitly*** reduce model differences – search through the model space to find updates to EM that allow the expectation to be met

3. Model Reconciliation Planning (MRP)

- **Plan Explicability [1]**: ***learning*** the model of expectation (EM) from human labels; ***optimizing*** in terms of both plan cost and explicability (computed based on the learned model)

$$\underset{\pi_{M_R}}{\operatorname{argmin}}\, cost(\pi_{M_R}) + \alpha \cdot dist(\pi_{M_R}, \pi_{\mathcal{M}_R^*})$$

- **Interactive Plan Explicability [2]**: extending the model of expectation (EM) to an interactive setting; plan explicability is influenced by both human and robot actions in the context of a joint plan

4. Explanation Generation (MRE)

- **Multi-model Explanation [3]**: searching for explanations that satisfy the following properties: ***completeness***, ***conciseness***, and ***monotonicity***

E.g., $\mathcal{E}^{MCE} = \arg\min_{\mathcal{E}} |\Gamma(\widehat{\mathcal{M}}) \Delta \Gamma(\mathcal{M}^H)|$ minimally complete

- **Progressive Explanation [4]**: Considering not only the correctness of explanations for the explainee but also the ***cognition effort*** incurred for interpreting the explanation

```
Amy: Let's go to the outlet today.
Monica: My car is ready.
Amy: Great!
Monica: The rain will stop soon.
Amy: Wonderful!
Monica: By the way, today is a holiday
(shops closed).
Amy: You are telling me now!
Monica: Let us go to the central park!
Amy: ...
```

E.g., $\arg\min_{\langle \Delta(\widehat{M^H}, M^H) \rangle} \sum_{f_i \in \langle \Delta(\widehat{M^H}, M^H) \rangle} \rho_i$

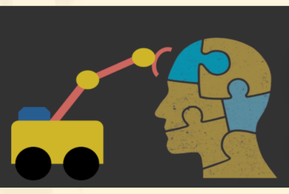
5. Summary of Current Progresses with References

1) Y. Zhang and M. Zakershahrak, *Progressive Explanation Generation for Human-robot Teaming*, under review

2) M. Zakershahrak, A. Sonawane, Z. Gong and Y. Zhang, *Interactive Plan Explicability in Human-Robot Teaming*, ROMAN, 2018

3) T. Chakraborti, S. Sreedharan, Y. Zhang, S. Kambhampati, *Plan Explanations as Model Reconciliation: Moving Beyond Explanation as Soliloquy*, IJCAI, 2017

4) Y. Zhang, S. Sreedharan, A. Kulkarni, T. Chakraborti, H. Zhuo, S. Kambhampati, *Plan Explicability and Predictability for Robot Task Planning*, ICRA, 2017



Cooperative Robotic Systems (CRS) Lab
COMPUTER SCIENCE AND ENGINEERING, ARIZONA STATE UNIVERSITY

ASU