Effect of Adapting to Human Preferences on Trust in Human-Robot Teaming

SHREYAS BHAT, JOSEPH B. LYONS, CONG SHI, X. JESSIE YANG

INDUSTRIAL & OPERATIONS ENGINEERING
UNIVERSITY OF MICHIGAN

UNIVERSITY OF MIAMI
MIAMI HERBERT BUSINESS SCHOOL

AFRL
THE AIR FORCE RESEARCH LABORATORY

Introduction

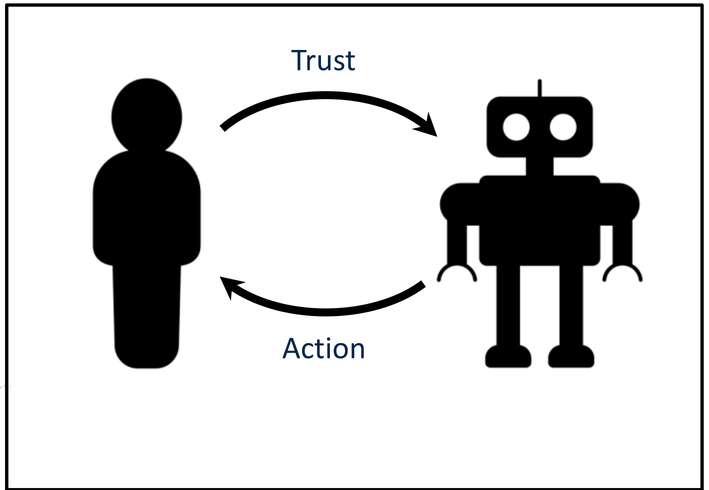
Trust is a key factor to facilitate effective collaboration [1]

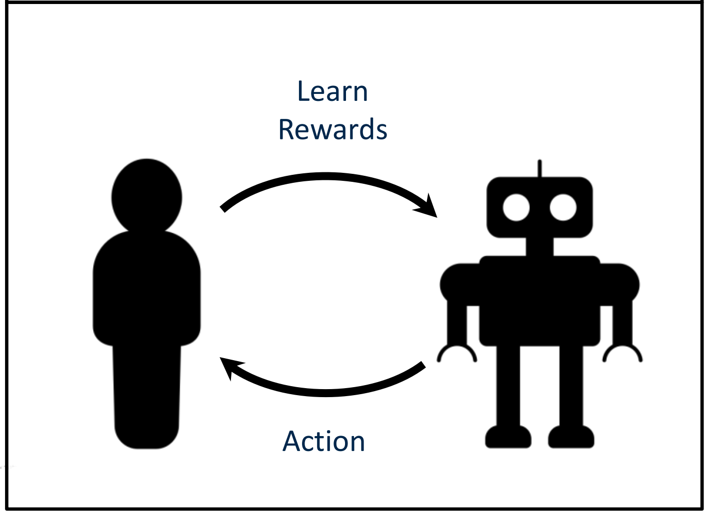
Trust has been used to drive the decision-making of robots in human-robot teams [2, 3]

- However, most prior research fixes reward functions for the human and the robot that the team tries to maximize

Value alignment [4, 5] refers to the field of study trying to match the “values” of the robot to that of the human

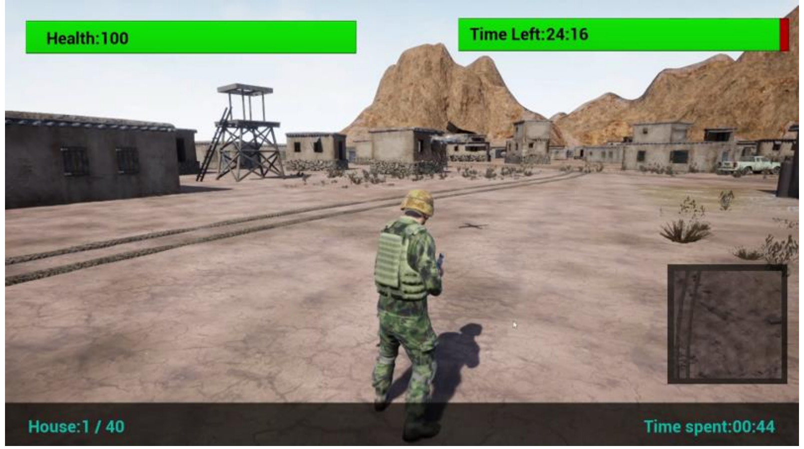
- This is done by learning reward functions for the human through their behavior
- The effect of value alignment on trust is an unexplored area

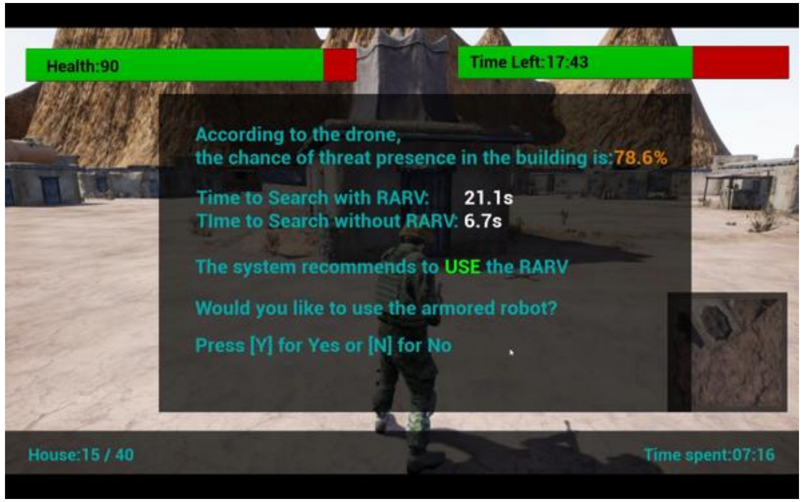




Task scenario

- Human-robot team searches through a town for potential threats (armed gunmen)
- At each site i, a drone scans and reports the level of threat $\hat{d}_i$
- The robot knows some prior information about threat in any site d_i
- The robot recommends whether
  - the human should breach the site directly
  - or they should deploy an armored robot for protection



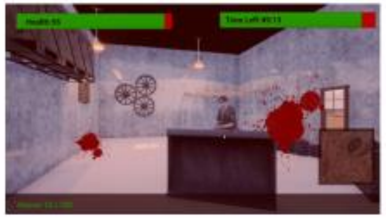


Task scenario

- The human chooses an action and observes the outcome of the action
- The human then reports their level of trust $\hat{t}_i$ on the recommendations
- The team then moves to the next site
- Their goal is to:
  - Minimize damage to the soldier
  - Finish the mission as quickly as possible

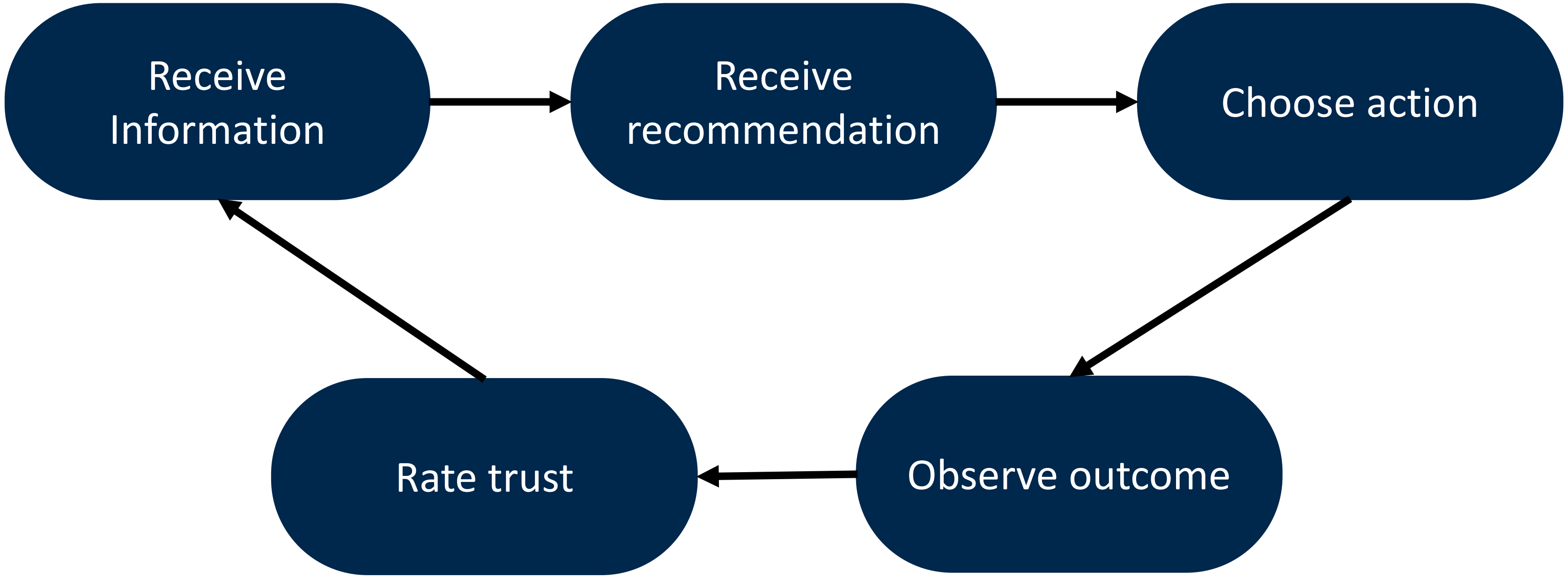

(a) No Threat, RARV Not Used


(b) No Threat, RARV Used


(c) Threat, RARV Not Used


(d) Threat, RARV Used

Task flow


Receive Information
Receive recommendation
Choose action
Observe outcome
Rate trust

Problem Formulation

- We formulate the interaction as a trust-aware Markov Decision Process (trust-aware MDP)

- A trust-aware MDP consists of:
  - States
  - Actions
  - Reward function
  - Transition function
  - Human behavior model

Trust-Aware MDP

- States: $$t_i \sim Beta(\alpha_i, \beta_i)$$
- Actions: $$a_i^h, a_i^r \in \{0, 1\}$$
- Reward function: $$R_i^r = -w_h^r h(D, a) - w_h^r c(a)$$ $$R_i^h = -w_h^h h(D, a) - w_h^h c(a)$$
- Transition function [6]: $$\alpha_i = \alpha_{i-1} + p_i w^s$$ $$\beta_i = \beta_{i-1} + (1 - p_i) w^f$$

$$\boxed{P_j = \begin{cases} 1, & \text{if } R_j^h(a_j^r) \geq R_j^h(1 - a_j^r), \\ 0, & \text{otherwise.} \end{cases}}$$

Trust-Aware MDP

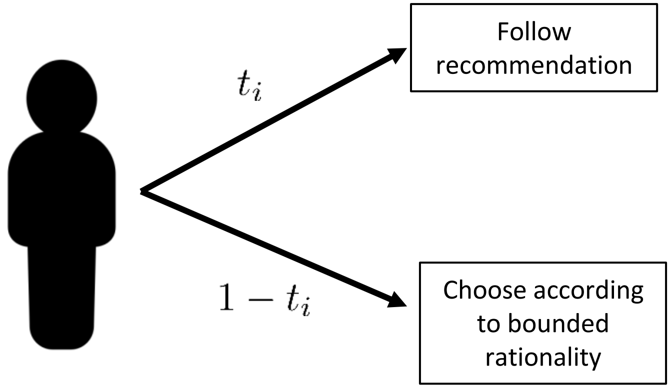
- Human behavior model:

$$P(a_i^h = a | a_i^r = a) = t_i + (1 - t_i) q_a,$$
$$P(a_i^h = 1 - a | a_i^r = a) = (1 - t_i)(1 - q_a).$$

$$\boxed{q_a = \frac{\exp(\kappa E[R_i^h(a)])}{\sum_{a' \in \{0,1\}} \exp(\kappa E[R_i^h(a')])}}$$

We call this the bounded-rationality-disuse model of human behavior

- The idea is that the human will accept and follow the recommendation with a probability equal to their current level of trust
- If they do not accept the recommendation, they choose an action based on the bounded rationality model



Bayesian Inverse Reinforcement Learning

- We use Bayesian Inverse Reinforcement Learning to learn personalized reward weights for each human during interaction

$$b_{i+1}(w_h^h = w) \propto \begin{cases} P(a_i^h = a_i^r | a_i^r) b_i(w), & \text{if } a_i^h = a_i^r, \\ P(a_i^h = 1 - a_i^r | a_i^r) b_i(w), & \text{otherwise.} \end{cases}$$

- This is done by maintaining and updating a distribution $b(w)$ on the possible reward weights w_h^h associated with losing health

- We compute the reward weight w_c^h associated with losing health as

$$w_c^h := 1 - w_h^h$$

Interaction Strategies (Conditions)

Non-learner: Assumes that the human shares the robot's reward function

Non-adaptive-learner: Learns personalized reward functions for each human. It only uses these for performance estimation and behavior prediction. It still optimizes its original reward function

Adaptive-learner: Learns personalized reward functions for each human and adopts it as its own

Experiment

Participants

- 12 students from the University of Michigan
- Age: 21.9 ± 2.4 years

Measures

- Subjective trust
- Behavioral trust

Design

- Within-subjects
- Counterbalanced

Results

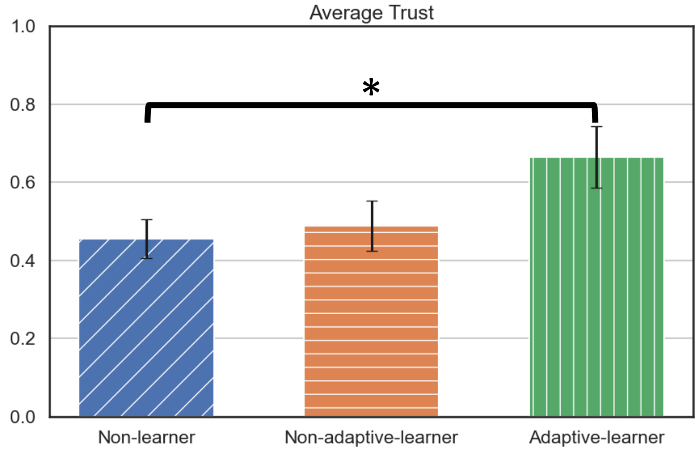
- Repeated measures ANOVA were performed for comparing the three strategies in various trust measures

- Bonferroni adjustment was used in post-hoc pairwise comparisons

Results (Contd.)

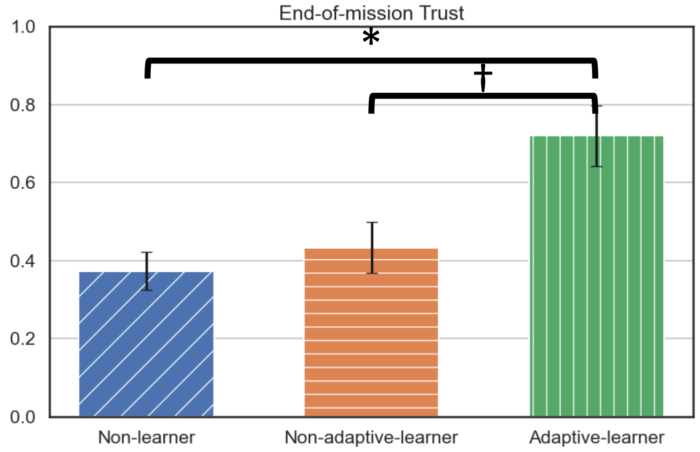
Average Trust Over Interaction

- $F(2, 22) = 4.968, p = 0.019$
- Adaptive-learner > Non-learner



End-of-mission Trust

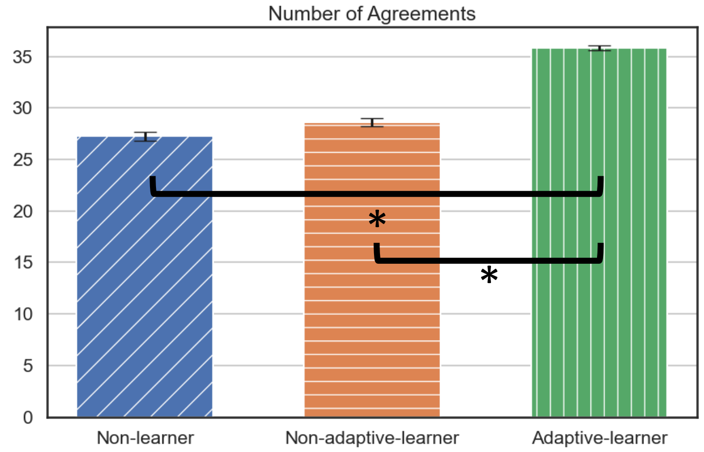
- $F(2, 22) = 7.455, p = 0.003$
- Adaptive-learner > Non-learner
- Adaptive-learner > Non-adaptive-learner



Results (Contd.)

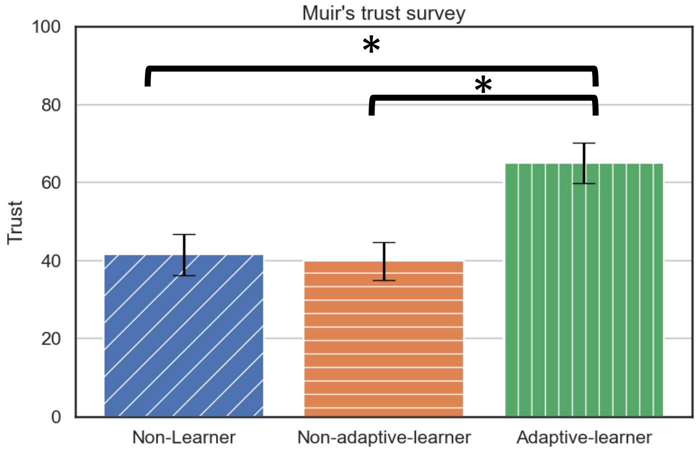
Number of Agreements

- $F(2,22) = 13.732, p < 0.001$
- Adaptive-learner > Non-learner
- Adaptive-learner > Non-adaptive-learner



Subjective Trust (Muir's Survey)

- $F(2,22) = 11.962, p < 0.001$
- Adaptive-learner > Non-learner
- Adaptive-learner > Non-adaptive-learner



Summary and Conclusions

Summary

- We demonstrated the use of Bayesian Inverse Reinforcement Learning for aligning robot's values to human preferences
- We compared human trust between three interaction strategies for the robot

Conclusions

- Aligning to the human's preferences increases subjective as well as behavioral trust

Future Work

- We started the Bayesian IRL algorithm with a uniform prior. Comparisons between the strategies when starting with a data-driven prior will be a good next step
- Comparing performance between the three interaction strategies

References

1. P. A. Hancock, T. T. Kessler, A. D. Kaplan, J. C. Brill, and J. L. Szalma, “Evolving Trust in Robots: Specification Through Sequential and Comparative Meta-Analyses,” Human Factors, vol. 63, no. 7, pp. 1196–1229, 2020.

2. S. Bhat, J. B. Lyons, C. Shi and X. J. Yang, "Clustering Trust Dynamics in a Human-Robot Sequential Decision-Making Task," in IEEE Robotics and Automation Letters, vol. 7, no. 4, pp. 8815-8822, Oct. 2022, doi: 10.1109/LRA.2022.3188902.

3. Min Chen, Stefanos Nikolaidis, Harold Soh, David Hsu, and Siddhartha Srinivasa. 2020. Trust-Aware Decision Making for Human-Robot Collaboration: Model Learning and Planning. J. Hum.-Robot Interact. 9, 2, Article 9 (June 2020), 23 pages. https://doi.org/10.1145/3359616

4. Luyao Yuan et al., In situ bidirectional human-robot value alignment. Sci. Robot.7,eabm4183(2022).DOI:10.1126/scirobotics.abm4183

5. Fisac, J.F. et al. (2020). Pragmatic-Pedagogic Value Alignment. In: Amato, N., Hager, G., Thomas, S., Torres-Torriti, M. (eds) Robotics Research. Springer Proceedings in Advanced Robotics, vol 10. Springer, Cham. https://doi.org/10.1007/978-3-030-28619-4_7

6. Guo, Y., Yang, X.J. Modeling and Predicting Trust Dynamics in Human–Robot Teaming: A Bayesian Inference Approach. Int J of Soc Robotics 13, 1899–1909 (2021). https://doi.org/10.1007/s12369-020-00703-3

7. Guo, Y., Yang, J.X., & Shi, C. (2023). TIP: A Trust Inference and Propagation Model in Multi-Human Multi-Robot Teams. Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction.

Thank You

SHREYAS BHAT

(shreyasb@umich.edu)

Research funded by: