Effects of Proactive Explanations by Robots on Human-Robot Trust

Lixiao Zhu and Tom Williams| MIRRORLab | Colorado School of Mines

Background

- Appropriate levels of trust must be established between teammates to achieve high levels of team performance [1].

- The transparency of robots' internal beliefs, desires, and intentions can help establish an appropriate level of human-robot trust [2,3].

- Verbal communicative behaviors, such as explanation generation, can provide robot transparency [3,4].

Proactive Explanatory Behavior

- Previous studies on robot explanation generation have mainly focused only on reactive explanations: post-hoc explanations generated in response to a request from a human teammate to explain a previous behavior.

- In contrast, little work has explored proactive explanations: explanations that are generated before an action is performed.

- Proactive Announcement vs Proactive Explanation.

Human-Subject Experiment

- Research goal: to understand the fundamental relationship between human-robot trust of robots and proactive explanatory behaviors.

- Two hypotheses:

  - *H1: Robots that generate proactive explanatory behavior will be more trusted than robots that do not.*

  - *H2: Robots that generate proactive explanations will build greater human-robot trust than those that perform proactive announcements.*

- Participants collaboratively engaged with robots in a novel resource management task.

Sponsors

This work was funded in part by:

Early Career Award
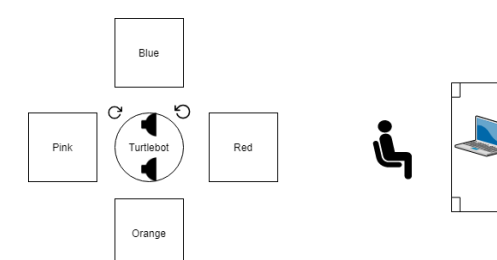
Results

- Robots that generated proactive explanations were more trusted than those that generated no proactive explanatory behaviors.

- More data must be collected before concluding precisely what effect is had by proactive announcements.
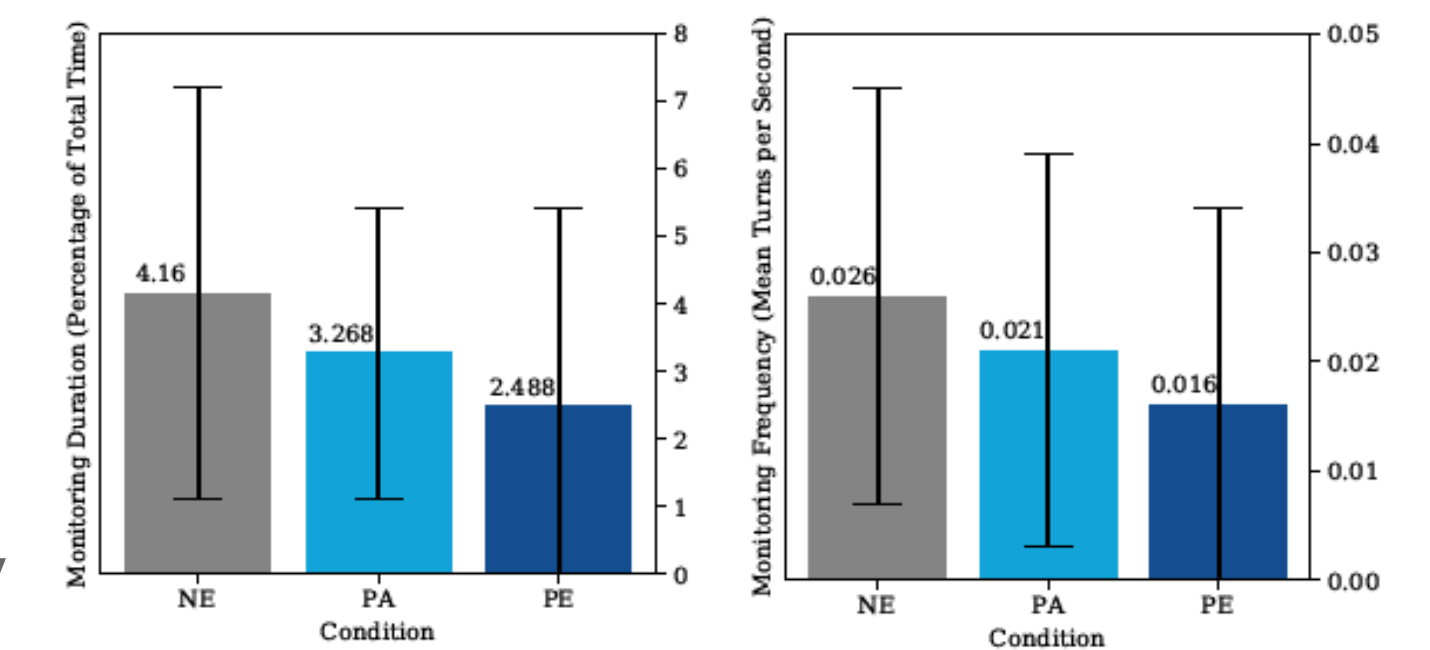
Future Work

- To collect sufficient data to confirm or refute the inconclusive findings presented in this paper.

- To identify the optimal policies for navigating the tradeoff between trust and workload that is presented during explanation generation.

- To interrogate the new research questions that have been identified regarding theory of mind and the fundamental nature of human-robot trust.

Reference

- [1] Billings, D., Schaefer, K., Llorens, N., Hancock, P.: What is trust? defining the construct across domains. In: Proc. Conf. of the American Psych. Assoc. (2012)

- [2] Helldin, T.: Transparency for Future Semi-Automated Systems : Effects of transparency on operator performance, workload and trust. Ph.D. thesis, University of Skövde (2014)

- [3] Mercado, J.E., Rupp, M.A., Chen, J.Y.C., Barnes, M.J., Barber, D., Procci, K.: Intelligent agent transparency in human–agent teaming for multi-uxv management. Human Factors58(3), 401–415 (2016)

- [4] McManus, T., Holtzman, Y., Lazarus, H., Anderberg, J., Ucok, O.: Transparency, communication and mindfulness. Journal of Management Development (2006)

CONTACT

Lixiao Zhu

lizhu@alumni.mines.edu

Tom Williams

twilliams@mines.edu

mirrorlab.mines.edu

MIRRORLab
Mines Interactive Robotics Research

CS@Mines

DEPARTMENT OF
COMPUTER SCIENCE