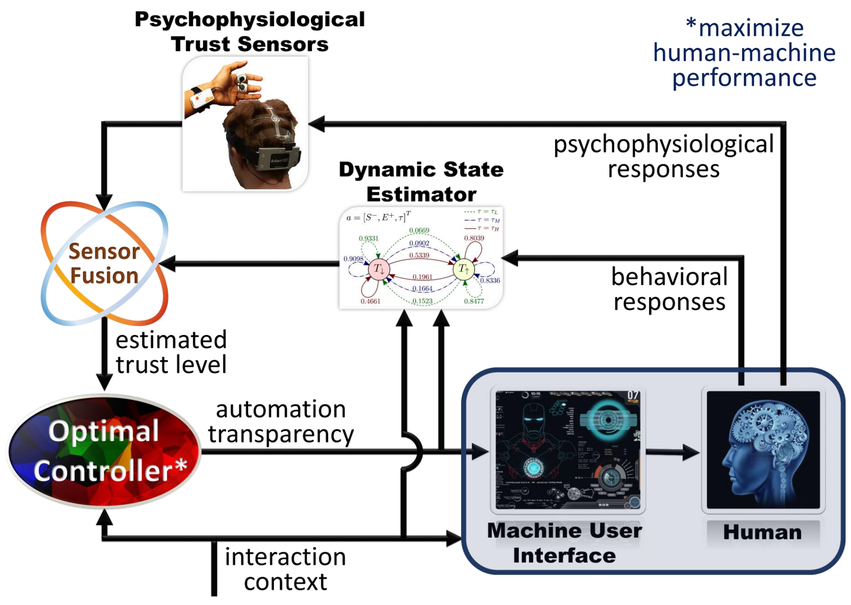
Reimagining Human-Machine Interactions Through Trust-Based Feedback

Research Assistant: Kumar Akash (Ph.D. Student), kakash@purdue.edu
Undergraduate Researcher: Griffon McMahon, gmcmahon@purdue.edu
Principal Investigators: Dr. Neera Jain and Dr. Tahira Reid
Sponsor: National Science Foundation & Bilsland Dissertation Fellowship

Problem Statement

- Automation loses its utility if humans override such systems due to a fundamental lack of trust [1]
- Accidents may occur due to mistrust [2]
- Trust should be appropriately calibrated to avoid disuse or misuse of automation [3]

We aim to design intelligent machines that can respond to changes in human trust in real-time.



Feedback approach to dynamically vary machine user interface based on human trust

[1] B. M. Muir, "Trust between humans and machines, and the design of decision aids," International Journal of Man-Machine Studies, vol. 27, no. 5-6, pp. 527 – 539, 1987.
[2] M. Richtel and C. Dougherty, "Googles driverless cars run into problem: Cars with drivers," New York Times, vol. 1, 2015.
[3] J. Lee and K. A. See, "Trust in Automation: Designing for Appropriate Reliance," Human Factors, vol. 46, no. 1, pp. 50–80, 2004.

Methodology

Human Subject Studies
Collection of time-domain data of human **psychophysiological & behavioral responses**

System Identification and Machine Learning

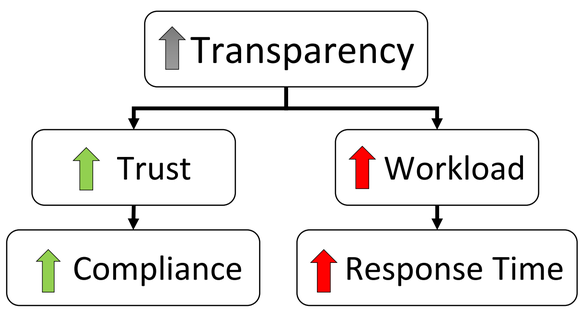
Plant Model
Variations in machine UI to human behavioral responses

Sensor Model
Psychophysiological responses to trust level

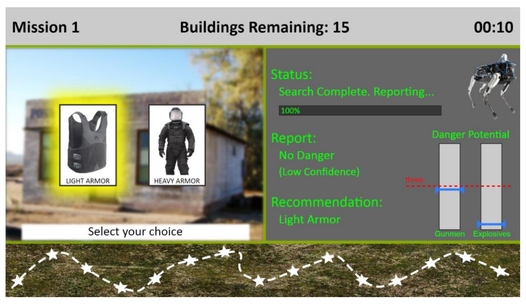
Control Algorithm Design
Optimal controller to maximize Human-Machine Interaction performance

Background

- Automation transparency: **amount and utility of information** to assist operator's comprehension about an automation's intent and reasoning
- Automation transparency affects human trust and workload



Human Subject Study



- Reconnaissance mission with a **robotic assistant** (transparency varied)
- Search 15 buildings quickly
- Participants decide to send light-armored or heavy-armored squad to search building based on the robot's recommendation

Data Collection

- 203 participants' behavioral data used for model estimation
- 48 in-lab participants recruited for psychophysiological data collection
- 80 participants' behavioral data used to validate the control algorithm

PURDUE MECHANICAL ENGINEERING

Results

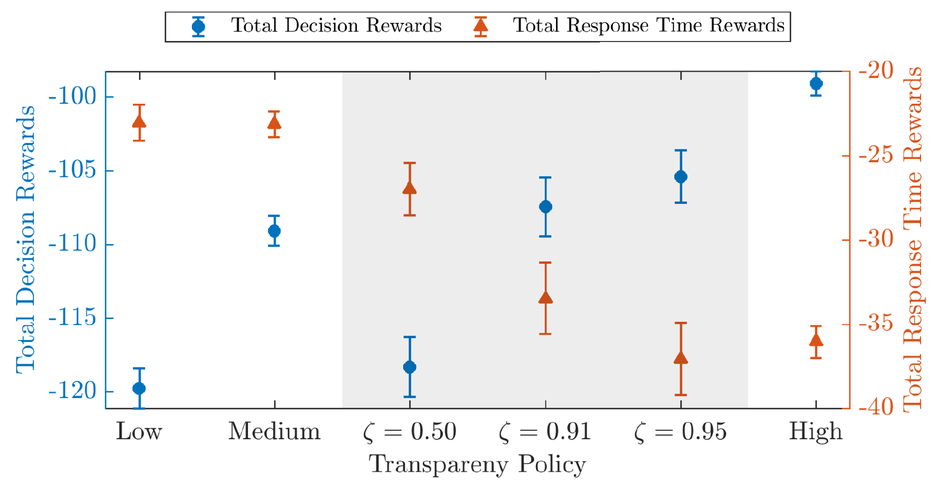
Plant Model [4]

- Trust and workload modeled as a Partially Observable Markov Decision Process (**POMDP**)
- Baum-Welch algorithm used to estimate models with collected data
  - **Independent** trust and workload models
  - **Coupled** trust-workload model

Control Algorithm [4]

Control objective is to balance the **trust-workload tradeoff** based on:

- decision rewards: to **safely** complete the task
- response time rewards: to achieve **time-efficiency**

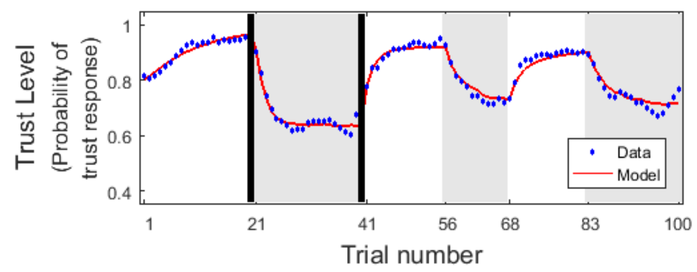


- Control policies can dynamically vary transparency to address the tradeoff (shown in gray region) whereas a fixed transparency (low, medium, or high) throughout the interaction cannot

Psychophysiological Sensor Model [5] [6]

Real-time sensing of trust using psychophysiological sensors needed to **augment trust estimate** when behavioral data is limited or unavailable

- Two-stage feature selection (ReliefF and SFFS) conducted from exhaustive set of time & frequency domain features of **EEG** & **GSR**
- A quadratic discriminant analysis classifier used for probabilistic classification of trust level
- An average accuracy of **78.55%** obtained for sensing categorical trust



Reliable trials (white)
100% accuracy

Faulty trials (gray)
50% accuracy

[4] K. Akash, G. McMahon, T. Reid, and N. Jain, Human Trust-based Feedback Control: Dynamically varying automation transparency to optimize human-machine interactions." IEEE Control Systems Magazine, July 2019. (Under Review)
[5] K. Akash, W.-L. Hu, N. Jain, and T. Reid, "A Classification Model for Sensing Human Trust in Machines Using EEG and GSR." ACM Transactions on Interactive Intelligent Systems. (Accepted)
[6] W. Hu, K. Akash, N. Jain, and T. Reid, "Real-Time Sensing of Trust in Human-Machine Interactions." 1st IFAC Conference on Cyber-Physical & Human-Systems, Florianopolis, Brazil, December 7-9, 2016.

Next Steps

Sensor Fusion Design

- Combining trust estimation from dynamic state estimator and psychophysiological response data to improve overall trust estimate

Generalization to other Contexts

- Extending the framework to other autonomous contexts such as **automated driving**, **industry automation, smart buildings**, etc.

Potential Impact of Research

- More **reliable** and **efficient** operation of safety-critical systems
- Extensions to other research domains interested in dynamic models of human trust (e.g. sociology, e-commerce)

Acknowledgements

This material is based upon work supported by the National Science Foundation under Award No. 1548616. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

PURDUE The Graduate School ADVANCE TO A HIGHER DEGREE
Bilsland Dissertation Fellowship

RAY W. HERRICK LABORATORIES