A Computational Model of Coupled Human Trust and Self-Confidence Dynamics

Research Assistants: Madeleine Yuh
Principal Investigator(s): Dr. Neera Jain
Contact email: *myuh@purdue.edu, neerajain@purdue.edu*

Project Objective

- Increasing complexity of human-automation interactions necessitates automation that is responsive to human cognitive behavior to avoid its misuse, disuse, and abuse [1].
- Trust and self-confidence cognitive states affect the human's decision to rely on automation [2] but most existing model frameworks consider only trust.
- The ability to estimate and predict cognitive states enables autonomous systems to aptly respond and adapt to humans for better task performance.
- BUT a mathematical model of human cognitive state evolution is required

Objective: Develop a probabilistic dynamic model for real-time estimation and prediction of human trust and self-confidence behavior

[1] R. Parasuraman and V. Riley, "Humans and Automation: Use, Misuse, Disuse, Abuse," *Hum Factors*, vol. 39, no. 2, pp. 230–253, Jun. 1997
[2] J. D. Lee and N. Moray, "Trust, self-confidence, and operators' adaptation to automation," International Journal of Human-Computer Studies, vol. 40, no. 1, pp. 153–184, 1994.

Methodology

Trust and self-confidence modeled using Partially Observable Markov Decision Process (POMDP)
Coupling exists between

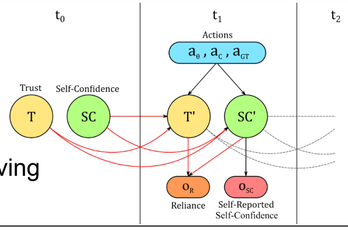
- Trust and Self-confidence states across consecutive time steps
- Reliance observation and trust and self-confidence states.

States	Trust and Self-Confidence
Actions	Collisions, Game Time, Θ Level
Observations	Reliance, Self-Reported Self-Confidence

Transition Probability	$\tau(s'|s,a) = \tau(s_T'|s_T, s_{S_C}, a)\tau(s_{S_C}'|s_T, s_{S_C}, a)$
Emission Probability	$\varepsilon(o|s) = \varepsilon(o_R|s_T, s_{SC})\ \varepsilon(o_{SC}|s_{SC})$

Therefore, coupling is applied to

- ***formulation of the transition probabilities*** (probability of transitioning to the current state given the previous states and actions)
- ***emission probabilities*** (probability of observing the emitted observation given current state).



Acknowledgements

This material is based upon work supported by the National Science Foundation under Award No.183690. Any opinions, findings, and material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Methodology cont.

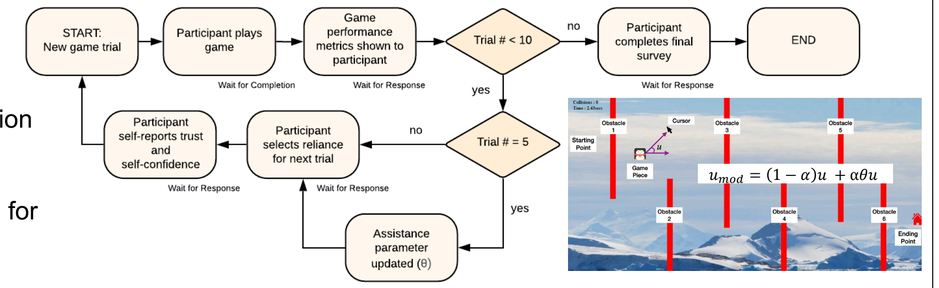
Human Subject Study:

- Obstacle avoidance game with automation assistance
- Θ acts as a proxy of the input of the automation that scales the human's input u while playing the game
- Participants enable $(\alpha = 1)$ or disable$(\alpha = 0)$ automation assistance prior to each trial

$$\boldsymbol{u_{mod} = (1-\alpha)u + \alpha\theta u}$$

Data Collection: 340 participants' behavioral data used for model parameter estimation

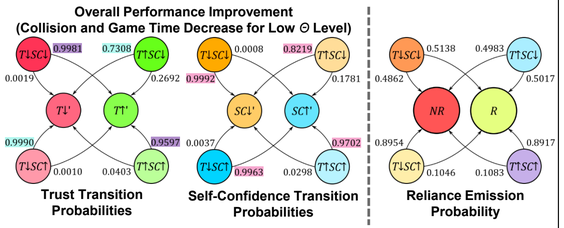
System Identification and Machine Learning:
Trained POMDP model using behavioral response data and a genetic algorithm



Results

Transition Probabilities:

- **Demonstration of Attribution Theory:** depending on performance and input from the automation assistance, participants attribute their successes and failures to either themselves or the automation.
- For overall performance improvement case shown, SC level likely to be maintained.
  - T↑' → Performance attributed to automation.
  - T↓' → Performance attributed to human.

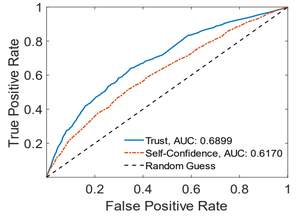


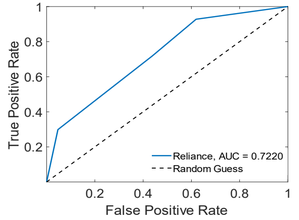
Reliance Emission Probabilities:

- SC↑ → Trust is proportional to likelihood to rely
- SC↓ → Reliance likelihood almost equally distributed

Model Performance:

- Receiver operating characteristic (ROC) curves are generated for both the cognitive states and reliance observation.
- Higher AUCs correspond to better model classification performance.

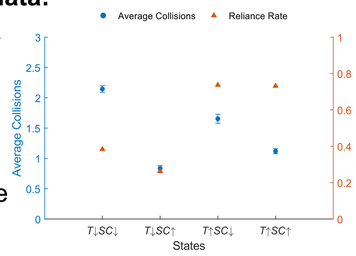




Future Work and Potential Impact of Research

From self reported data:

- T↓SC↓ and T↑SC↓ correspond to poor performance.
- Combinations of T-SC states result in different reliance behaviors



Closing the Loop: Design feedback control algorithms to calibrate the human's self-confidence and trust. Behavioral trends from self-reported data can be incorporated in future algorithm design.
Generalize the model to other contexts: Adapt POMDP framework to other HAI contexts and incorporate other control algorithms in place of Θ.

PURDUE UNIVERSITY | The Ray W. Herrick Laboratories