Grey-box Modeling of Human Cognitive State Dynamics in Automated Driving Contexts

Research Assistants: Sibibalan Jeevanandam (sjeevana@purdue.edu), Xipeng Wang, Tyler Hsieh | Principal Investigator: Dr. Neera Jain (neerajain@purdue.edu)

PURDUE UNIVERSITY®

Ray W. Herrick Laboratories

Overview

During safety-critical applications of autonomous (or semi-autonomous) systems in highly stochastic environments involving human interaction, human complacency may cause misuse of automation, sometimes leading to fatal accidents. While efforts have been made to understand cognitive factors (such as trust in the automation) responsible for human behavior during automated driving [1], comparatively **less research has been done to characterize the underlying dynamics of these cognitive factors.** Knowledge of the evolution of these factors in real time can be used to vary automation factors such as transparency to reduce unsafe operation of automation.

In this work, we aim to identify a grey-box model rooted in the state-space of cognitive factors affecting human decision making during conditionally automated driving. **This work represents the first effort in modeling cognitive states in an experiment that is not event-based; the human engages continuously with the automation.**

- We present our human subjects experiment to elicit measurable changes in these states, with constraints informed by human factors research.
- We collect heterogeneous measurements of the human's behavior and physiology.
- We identify the underlying dynamic characteristics of the human's cognitive states using system identification techniques.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Award No. 2145827. Any opinions, findings, and material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

[1] S. Jeevanandam, M. Williamson Tabango, X. Wang, and N. Jain, "A Novel Experiment Design for Studying Multiple Cognitive Factors in Conditionally Automated Driving Contexts" in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Phoenix, AZ, September 9-13, 2024.

[2] W. -L. Hu, K. Akash, T. Reid and N. Jain, "Computational Modeling of the Dynamics of Human Trust During Human–Machine Interactions," in *IEEE Transactions on Human-Machine Systems*, vol. 49, no. 6, pp. 485-497, Dec. 2019.

Human Subject Experiment

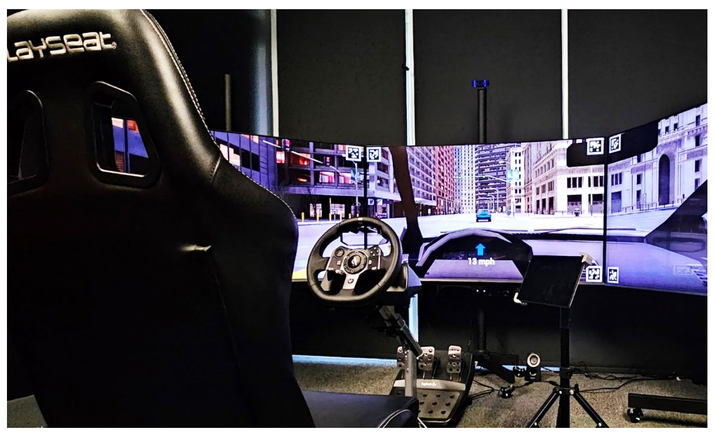

Figure 1: Driving simulator setup

Figure 2: Ego-vehicle approaching a construction zone

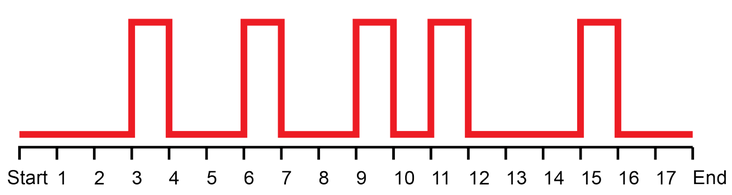

Figure 3: Binary signal for task complexity; changes occur at intersections (numbered)

Experiment Objective: To perturb cognitive states of interest by varying task complexity in a medium fidelity driving simulator (Figure 1).

- Task complexity is varied as a binary signal (Figure 3).
- Low complexity is city driving with low traffic; high complexity is navigating through construction zones with workers (Figure 2).
- Automation is 100% reliable.
- Each participant receives the same order of task complexity changes to enable studying individual-specific differences (e.g., automation bias) and characterizing uncertainty in different measurements.
- Participants' response is captured via behavioral (reliance on automation, eye-gaze), physiological (heart rate, functional Near Infrared Spectroscopy (fNIRS), galvanic skin response), and subjective (pre- and post-experiment questionnaires) measures.
- The in-person study was approved by the Institutional Review Board at Purdue University. Participants were compensated at $20/hr.
- Data was collected for 12 participants.

Polar H10 Heart Rate Monitor

NIRSport2 by NIRx

Neon by Pupil Labs

Shimmer 3+ GSR Unit

Figure 4: Sensor suite

Data Pre-Processing

- Participants are equally divided into training and testing participants by random selection.
- Reliance on automation (a binary signal) is converted to a continuous-valued metric with range 0-100 (Distrust) by defining it as the likelihood of a participant not relying on automation (based on [2]) for the set of training and testing participants separately.
- The likelihood of high task complexity (TC) is similarly computed for both sets.

$$Distrust(t_k) = \frac{100}{|P|}\sum_{i\in P}(1 - Reliance^i(t_k)), \qquad TC(t_k) = \frac{100}{|P|}\sum_{i\in P} TC^i(t_k)$$

t_k : Discrete-time instants
P : Participants in the training/testing sets
$|P|$: Cardinality of P

ARX Model

$$\boldsymbol{Distrust(t_k) = a\, Distrust(t_{k-1}) + b\, TC(t_k) + e(t_k)}$$
$$\boldsymbol{Trust(t_k) = 100 - Distrust(t_k)}$$

- An AutoRegressive with eXogenous inputs (ARX) model of **order 1** is identified using the training set.

$$\boldsymbol{a = 0.9907, b = 0.0083}$$

- The identified model predicts Trust for the test population well (Figure 5).
- The population model is also used to predict Reliance on automation at a participant level ($Reliance^i(t_k)$).
- Prediction performance of the identified model varies significantly between participants in the illustrative examples (Figures 6 and 7).

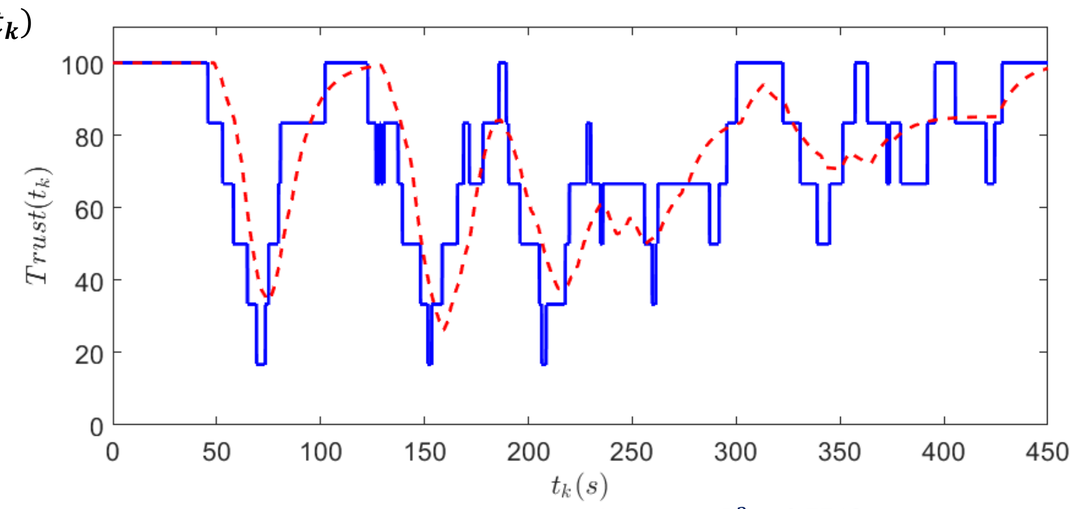

Figure 5: Test data vs model output; $R^2 = 0.7742$

Participant 9 chose to not rely on the automation in all construction zones, (similar to mean population behavior); the identified model predicts this trend well.

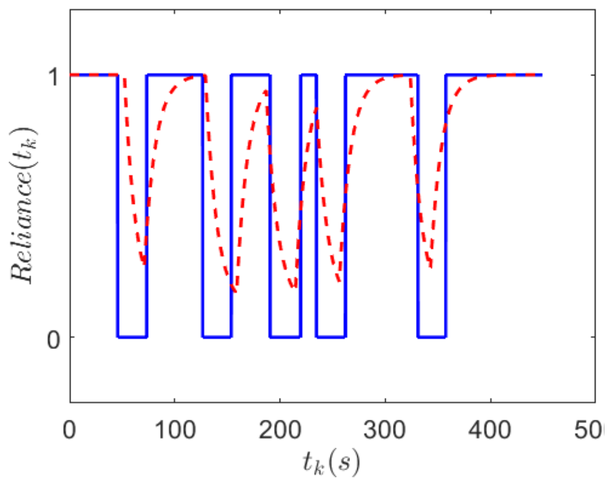

Figure 6: Test data vs model output (P9); $R^2 = 0.5099$

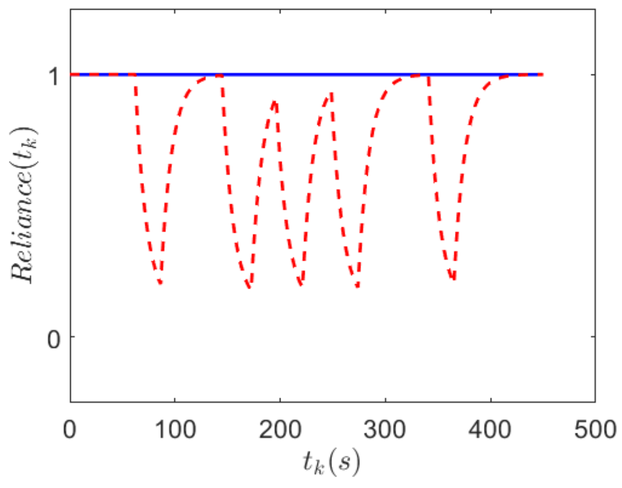

Figure 7: Test data vs model output (P11)

Participant 11 constantly relied on the automation even in construction zones unlike majority of the population; the population model is not suited to predict responses like this.

Future Work

- Characterize variance in behavior ($e(t_k)$) using Reproducing Kernel Hilbert Space (RKHS) methods, and identify dynamic models with well-informed assumptions for noise.
- Identify models to predict workload using heart-rate, galvanic skin response, and fNIRS signals.

CONCLUSIONS

- We perturbed trust in the automation by designing an experiment with changes in task complexity via presence of construction zones.
- A first-order ARX model captures the population trust dynamics in an automated driving context.
- The population model performs well at a participant level for some participants but not others.
- Individual-specific factors (such as automation bias) need to be characterized to group similar participants.