JAIN RESEARCH LAB

Do You Trust It? A Set-Based Approach to Estimating Human Trust and Risk Perception During Automated Driving

Research Assistant: Sibibalan Jeevanandam (sjeevana@purdue.edu) | 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 uncertain environments involving human interaction, such as automated driving, human complacency may cause misuse of automation, sometimes leading to fatal accidents. While efforts have been made to understand cognitive states—such as trust—responsible for human behavior during automated driving [1], comparatively **less research has been done to estimate these states in real time.** Estimates of these states could be leveraged by the automation to respond and adapt to the human to improve safety and performance outcomes.

In this work, we leverage a hybrid modeling framework [2] to build a set-valued state estimator for human cognitive factors, such as trust (T) and risk perception (R). Conventional state estimation methods, such as Kalman filtering, use a probabilistic characterization of uncertainty which poses a challenge when using human data with highly intermittent observations. Instead, we use a set-based approach to handle information obtained using quantized measurements (cognitive factors intermittently self-reported by the user), and a binary behavioral measurement (human reliance on the automation). **This work represents the first effort in estimating cognitive states in an experiment that is not event- or trial-based; the human engages continuously with the automation.**

Human Subject Experiment Description

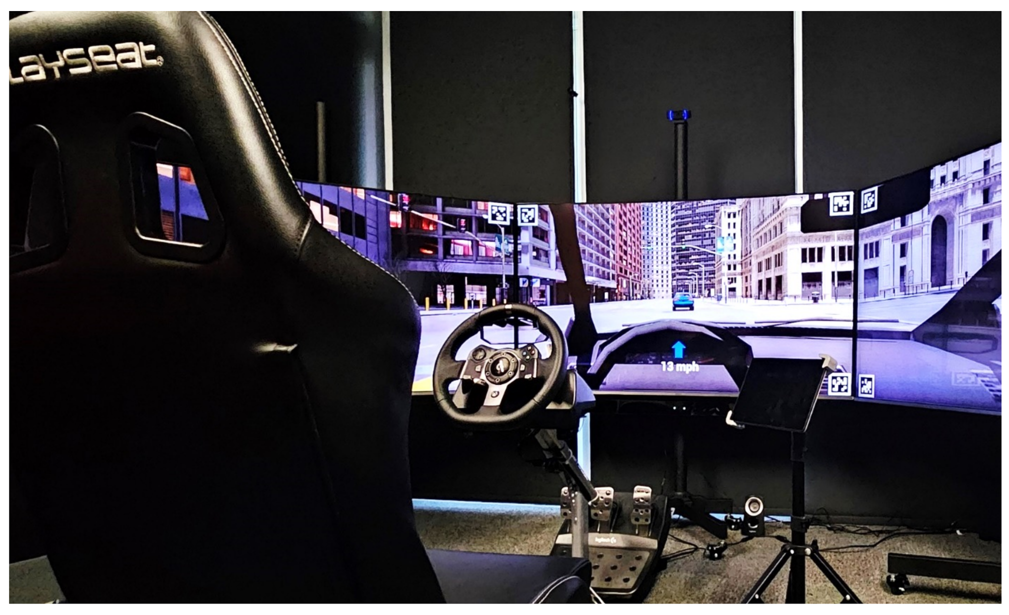


***Figure 1**: Driving simulator setup*

***Figure 3**: Ego-vehicle approaching a construction zone*

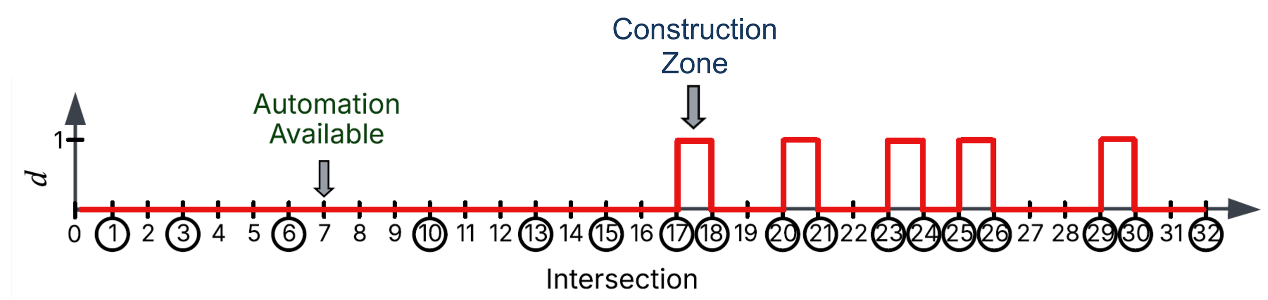


***Figure 2:** Binary signal for task complexity*

Experiment Objective: To perturb the human driver's cognitive states (x) and consequently their reliance (q) on the automation by varying task complexity (d) in a medium fidelity driving simulator (**Figure 1**).

- Task complexity is varied as a binary signal (**Figure 2**).
- Low complexity is city driving with low traffic; high complexity is navigating through construction zones with workers (**Figure 3**).

- Participants navigate through a pre-defined route in an ego-vehicle with SAE Level 3 automation during a single, continuous drive.
- When available, participants are free to engage or disengage the automation at their discretion.
- Self-reports (y) for cognitive states are solicited at intersections (circled in Figure 2) on a scale from 0 to 100, in increments of 5.

Hybrid Dynamical Model

- In prior work [2], we modeled the evolution of the cognitive states ($x = [T\ R]^T \in \mathbb{R}^2$) and reliance ($q \in \{0,1\}$) during changes in task complexity (d) using a hybrid dynamic model, given by

$$x(k+1) = Ax(k) + Bd(k) + c + w, \forall k = 0, \cdots, N$$

$$q(k) = \begin{cases} 1, & \text{if } x \in S_1 \\ 0, & \text{if } x \in S_0 \end{cases}$$

Decision Regions

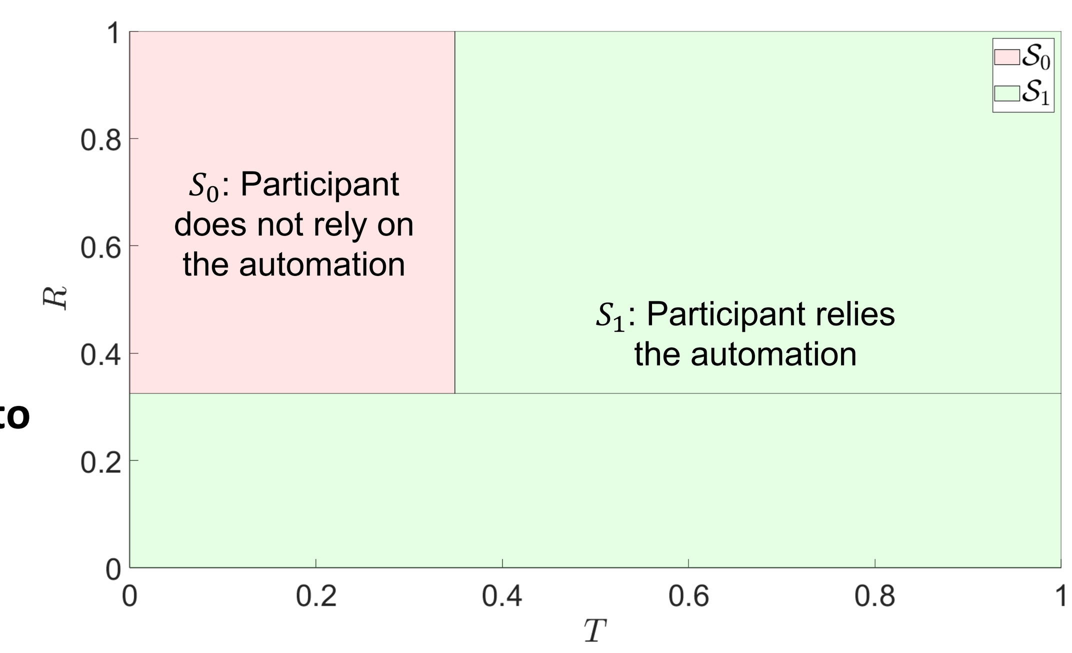


***Figure 4:** Decision regions in P003's cognitive state space, identified using a decision tree*

- **The proposed model structure is low-dimensional and can be used to identify participant-level models, capturing individual-specific behaviors**
- Self-reports (y) are quantized, intermittent measurements of the cognitive states, such that

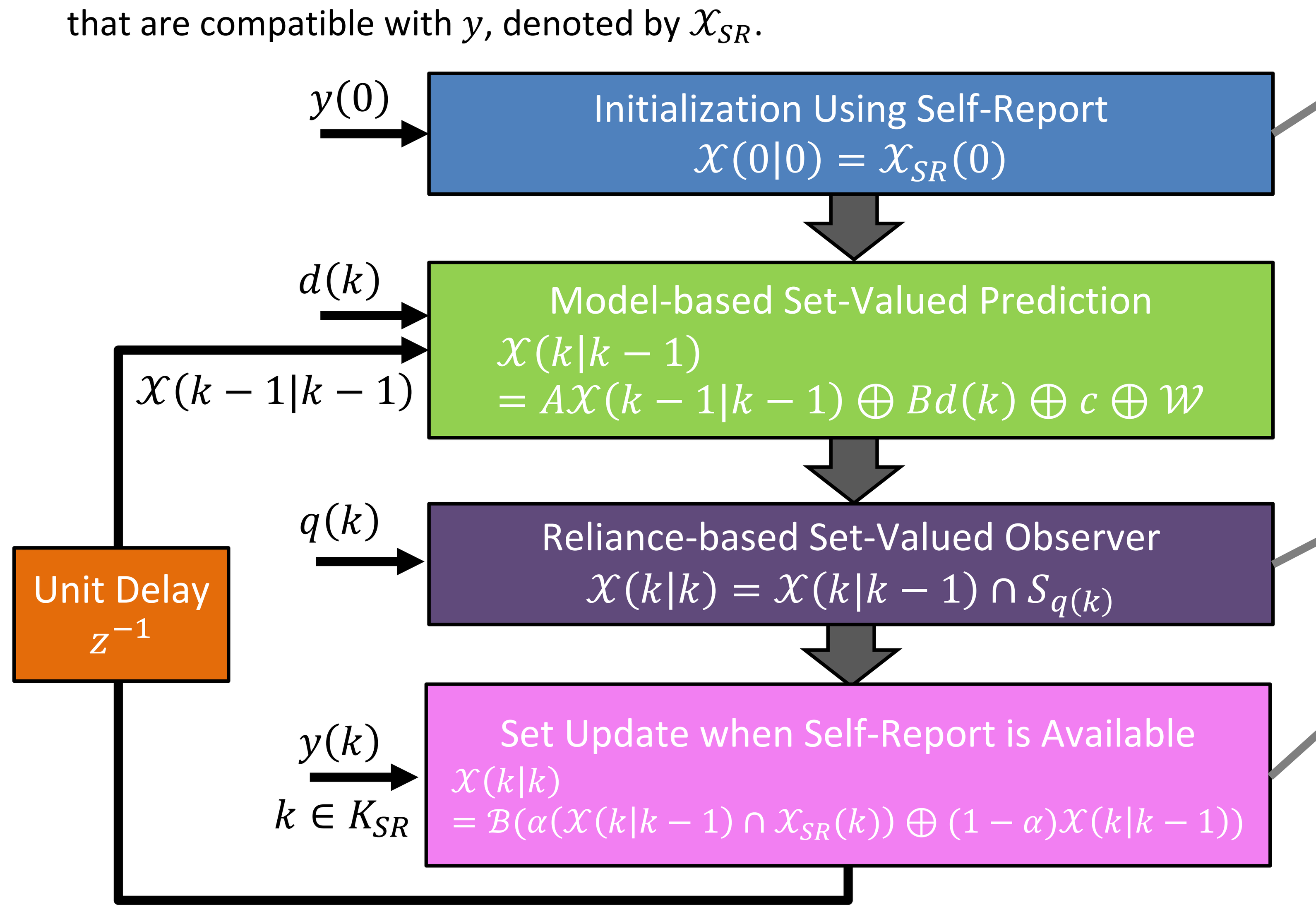
$$z(k) = x(k) + v, k \in K_{SR}$$

$$y_i(k) = \begin{cases} 0, & \text{if } z_i(k) < -\frac{\Delta}{2} \\ Y_j, & \text{if } z_i(k) \in \left[Y_j - \frac{\Delta}{2}, Y_j + \frac{\Delta}{2}\right] \\ 100, & \text{if } z_i(k) > 100 + \frac{\Delta}{2} \end{cases}$$

Δ: Quantization step size
$i \in \{1,2\}$: Denotes the i^{th} entry of y
Y_j: Possible values of y_i $(0, 5, \cdots, 100)$
K_{SR}: Set of k for which self-reports are available
$\mathcal{B}(\mathcal{X})$: Axis-aligned bounding box of $\mathcal{X}$

Set-Valued State Estimator

- We assume bounded process and measurement noise, i.e., $|w| < \delta^w, |v| < \delta^v$.
- Given a self-report (SR) y, the feasible measurement set is the set of all states x that are compatible with y, denoted by $\mathcal{X}_{SR}$.



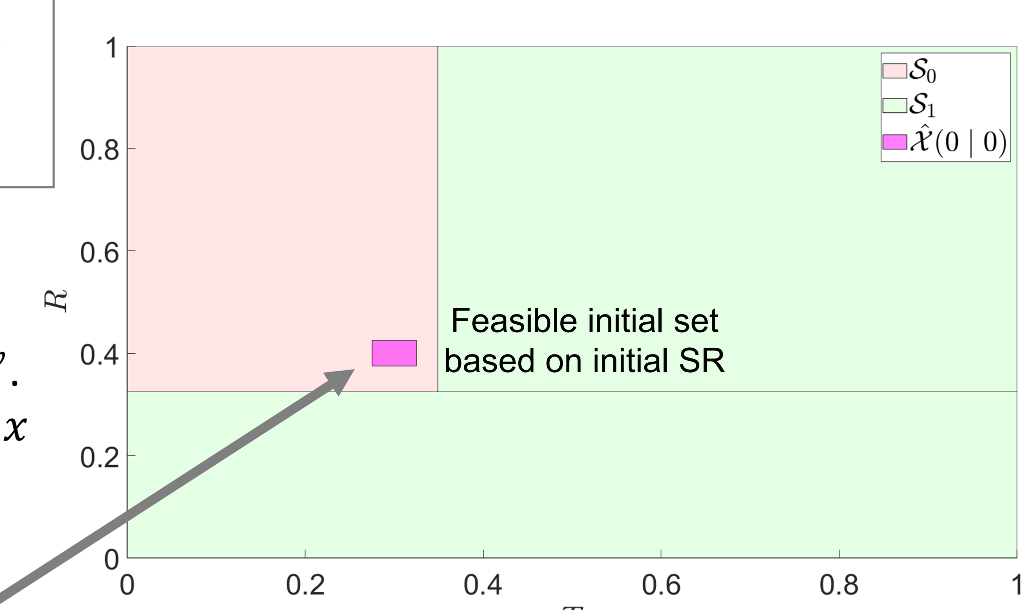


***Figure 5:** Set-valued estimator is initialized using a self-report*

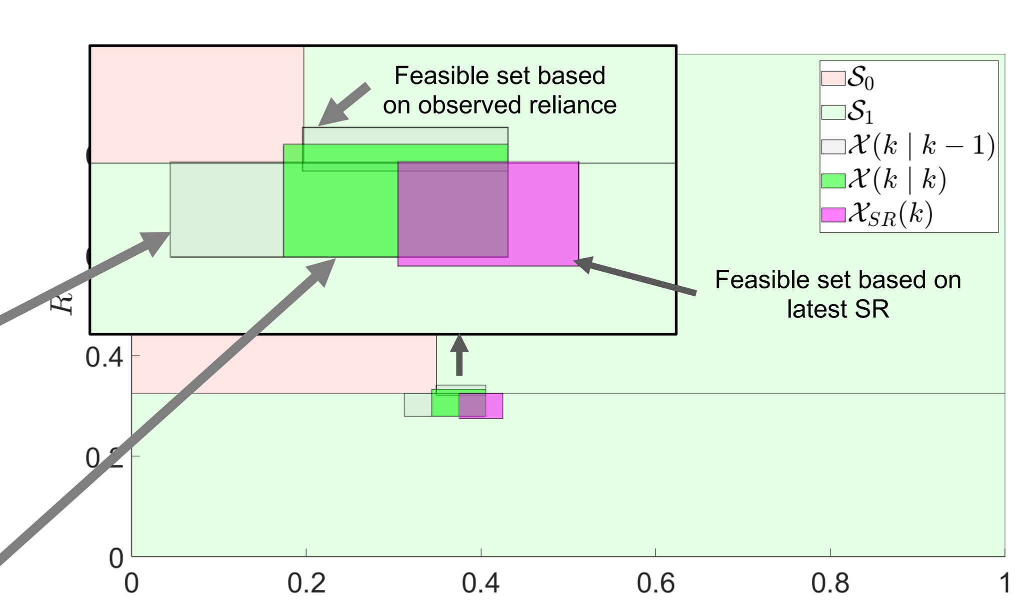


***Figure 6:** Estimate is updated using self-reports when available, and observed reliance on automation*

Future Work

- Devise a systematic method to identify the process and measurement noise bounds (δ^w, δ^v) as well as the update gain (α)
- Evaluate the empirical quality of the estimator by quantifying the uncertainty (using set size), and consistency with measured states

CONCLUSIONS

- We designed a human cognitive state estimator to track latent cognitive states such as trust in the automation and perception of risk in real-time during conditionally automated (SAE Level 3) driving.
- We used a set-based approach to handle quantized (self-reported cognitive states) as well as binary (reliance) measurements.
- We demonstrated the framework on participant data collected from an in-person human study.

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] S. Jeevanandam, and N. Jain, "A hybrid dynamic model for predicting human cognition and reliance during automated driving," in *2025 IEEE 28th International Conference on Intelligent Transportation Systems (ITSC)*, Nov. 2025.