Using Trust in Automation to Enhance Driver-(Semi)Autonomous Vehicle Interaction and Improve Team Performance

Hebert Azevedo-Sa[a], Xi Jessie Yang[a], Lionel P. Robert Jr.[a], Dawn M. Tilbury[a]

[a]_University of Michigan-Ann Arbor_

1. Introduction

Trust in robots has been gathering attention from multiple directions [1, 2, 3, 4, 5, 6, 7], as it has a special relevance in the theoretical descriptions of human-robot interactions [8, 9]. It is essential for reaching high acceptance and usage rates of robotic technologies in society [10], as well as for enabling effective human-robot teaming [5]. Researchers have been trying to model the development of trust in robots to improve the overall "rapport" between humans and robots. In particular, researchers sought to characterize factors that induce changes in drivers' trust in automated driving systems (ADSs)—the systems that provide the autonomous capabilities of (semi)autonomous vehicles—over their interactions [11].

Trust in automation [12, 13] is a popular concept among researchers, and evolved from the theory behind interpersonal trust development [14, 15]. Miscalibration of trust in automation is a common issue that jeopardizes the effectiveness of automation use [16]. It happens when a user's trust levels are not appropriateate to the capabilities of the automation being used. Users can be: _undertrusting_ the automation—when they do not use the functionalities that the machine can perform correctly because of a "lack of trust"; or _overtrusting_ the automation—when, due to an "excess of trust", they use the machine in situations where its capabilities are not adequate [17, 18, 19]. The influence of different risk types (internal, from the ADS itself, and external, from the environment) [20] and of situational awareness in trust development has been investigated [21]. Different risk types were found to influence trust development and to moderate the impacts of trust on non-driving-related task (NDRT) performance and on trusting behaviors. Moreover, when the ADS provides high quality information, drivers are able to increase their trust in the automation and achieve better performance on NDRTs.

The main objective of this work is to examine driver's trust development in the ADS. We aim to model how risk factors (e.g.: false alarms and misses from the ADS) and the short term interactions associated with these risk factors influence the dynamics of drivers' trust in the ADS. The driving context facilitates the instrumentation to measure trusting behaviors, such as drivers' eye movements and usage time of the automated features. Our findings indicate that a reliable characterization of drivers' trusting behaviors and a consequent estimation of trust levels is possible. We expect that these techniques will permit the design of ADSs able to adapt their behaviors to attempt to adjust driver's trust levels. This capability could avoid under- and overtrusting, which could harm their safety or their performance.

2. Methodology

We seek to characterize trust dynamics in short term interactions, i.e., during the different events that occur when drivers operate vehicles featuring ADSs. In other words, we aim to describe how trust in the ADS changes immediately after the ADS meets (or fails to meet) drivers' expectations. For this we have quantified drivers' trust in the ADS at specific time instances and used linear mixed effects models [22] to establish a state-space model for trust dynamics (1),

$$T(t+1) = \mathbf{A}T(t) + \mathbf{B} \begin{bmatrix} L(t) \\ M(t) \\ F(t) \end{bmatrix} + u(t) \,; \quad \begin{bmatrix} \varphi(t) \\ \pi(t) \\ \upsilon(t) \end{bmatrix} = \mathbf{C}T(t) + w(t) \,, \tag{1}$$

where T represents drivers' self-reported trust; $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ represent the state-space model matrices; The time variable t indexes an interaction event, characterized by the occurrence of true alarms (when the ADS is able to identify an obstacle on the road), false alarms (when the ADS issues an alarm but there is no

obstacle on the road) and misses (when the ADS is unable to identify an obstacle on the road). These events are respectively represented by $L(t)$, $F(t)$ and $M(t)$, which are inputs of the model. φ, π and υ represent, respectively, the measures of visual focus (measured from eye-tracking), NDRT performance and automation usage (observations); $u(t)$ and $w(t)$ represent random zero-mean noises intended to capture drivers' individual characteristics and biases. With this model, we develop a trust estimator able to provide real-time estimates of trust $T(t)$.

For data acquisition, we use an autonomous vehicle simulator in a user experiment where we expose the participants to different behaviors from the ADS. The participants experience two trials, i.e., in straight and curvy roads. Within each of the two trials there are 12 interaction events. We request the participants to self-report their trust in the ADS and measure the real time variables that represent their trusting behaviors. We use Muir's trust questionnaire [17] and scale the self-reported scores to match the interval from 1 to 100 points. Figure 1 shows our experimental setup.
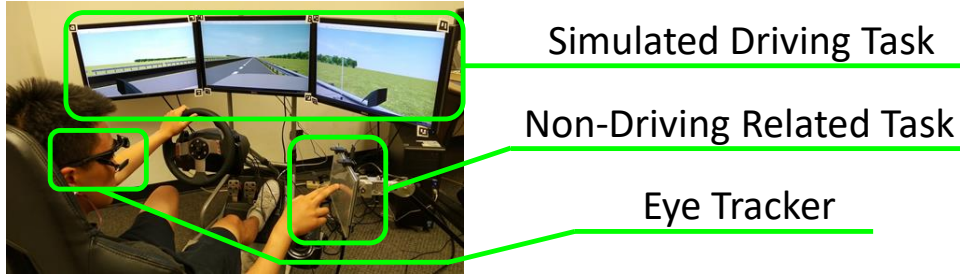

Figure 1: The setup for user experiments. Drivers interact with a simulated autonomous vehicle while performing a NDRT and reporting their trust in the system.

3. Results

We used the data acquired from 80 participants to compute the parameters for the trust dynamics model (1). Table 1 presents the state-space model matrices. With the identified parameters we designed the trust estimator, which is based on a linear Kalman filter. Figure 2 presents the trust estimation results for a randomly chosen participant, bringing together their two trials and the different estimation results for the 12 events within each trial.
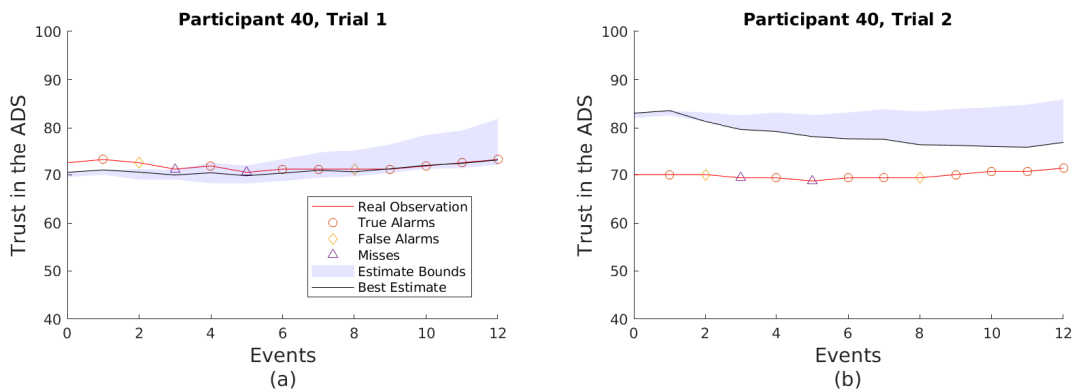

Figure 2: Curves representing the self-reported trust levels and the calculated estimates of trust for the 40th participant of the experiment, over 12 events. The events could be true alarms, false alarms, or misses. The blue shaded bands indicate the approximate bounds for the estimates.

The trust estimate bounds represented in Figure 2 are approximations obtained with the juxtaposition of 100 simulated estimates. The estimates vary with the noise introduced in the Kalman filter. This variability is

Table 1: State-space model parameters for trust in the ADS dynamics

Parameter	Value Estimate	S.E.M
A	$[1.00]$	$[0.25]$
B	$\begin{bmatrix} 0.224 & -0.670 & -0.798 \end{bmatrix}$	$\begin{bmatrix} 0.079 & 0.084 & 0.083 \end{bmatrix}$
C	$\begin{bmatrix} 7.01 \times 10^{-3} \\ 4.23 \times 10^{-3} \\ 9.20 \times 10^{-3} \end{bmatrix}$	$\begin{bmatrix} 3.6 \times 10^{-4} \\ 1.3 \times 10^{-4} \\ 1.0 \times 10^{-4} \end{bmatrix}$
σ_u^2	0.26	–
$\boldsymbol{\Sigma}_w = \begin{bmatrix} \sigma_\varphi^2 & 0 & 0 \\ 0 & \sigma_\pi^2 & 0 \\ 0 & 0 & \sigma_v^2 \end{bmatrix}$	$\begin{bmatrix} 0.18 & 0 & 0 \\ 0 & 0.07 & 0 \\ 0 & 0 & 0.06 \end{bmatrix}$	–

Note: S.E.M = Standard Error of the Mean.

due to the uncertainty represented by the random noise parameters u and w. The best estimate and the real observation curves are also represented in the plots.

Our results show that the proposed method can successfully track the self-reported trust levels from the observation variables (φ, π and v). However, comparing the curves from Figures 2(a) and 2(b), it is notable that the success of the estimation process highly depends on a precise initial estimate of trust in the ADS. In Figure 2(b), the difference between curves got reduced over the interactions, but the total of 12 events was not sufficient for the trust estimator to track the observed self-reported trust levels. Two causes contribute for this: first, the variation of drivers' trust in the ADS is a slow process, which can be noted by the flatness of the red curves in Figure 2; second, the noise terms (u and w) have large variances, as they intentionally capture the individual differences of behaviors from each participant. These large variances slow down the convergence of the curves, as the Kalman filter structure has a conservative behavior when the observations are contaminated by noise. This limitation might be addressed by establishing a longer term and individualized training process for the trust estimator, capable of providing more precise initial trust estimates. Furthermore, this process might also be used for tuning the parameters of the model, which would improve the accuracy of the estimates.

4. Conclusion

In this work we explored how trust develops over interactions between drivers and ADSs, and how the reliability of the ADS affects this trust development. We quantified trust and introduced an estimation method to describe and track trust dynamically in the short term interactions represented by true alarms, false alarms and misses. This method might be used for the design of ADSs that could adapt their behaviors to avoid trust miscalibration and improve team effectiveness, focusing on increasing safety and performance in the driving context.

References

[1] M. Chen, S. Nikolaidis, H. Soh, D. Hsu, S. Srinivasa, *Planning with trust for human-robot collaboration, in: Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction, ACM, Chicago, IL, 2018, pp. 307–315.*

[2] M. Körber, E. Baseler, K. Bengler, *Introduction matters: Manipulating trust in automation and reliance in automated driving, Applied Ergonomics 39 (1) (2018) 9–21. doi:10.1016/j.apergo.2017.07.006.*

[3] T. B. Sheridan, *Extending Three Existing Models to Analysis of Trust in Automation: Signal Detection, Statistical Parameter Estimation, and Model-Based Control, Human Factors: The Journal of the Human Factors and Ergonomics Society 42 (3) (2019) 001872081982995. doi:10.1177/0018720819829951.*
URL http://journals.sagepub.com/doi/10.1177/0018720819829951

[4] T. Kundinger, P. Wintersberger, A. Riener, *(Over)Trust in Automated Driving, in: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, ACM, Glasgow, UK, 2019, pp. 1–6. doi:10.1145/3290607.3312869.*
URL https://doi.org/10.1145/3290607.3312869

[5] P. Robinette, M. Novitzky, C. Fitzgerald, M. R. Benjamin, H. Schmidt, *Exploring human-robot trust during teaming in a real-world testbed, in: 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), IEEE, Daegu, Korea., 2019, pp. 592–593.*

[6] J.-G. Lee, J. Gu, D.-H. Shin, *Trust in unmanned driving system, in: Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts, ACM, Portland, Oregon, USA., 2015, pp. 7–8.*

[7] Y. Xie, I. P. Bodala, D. C. Ong, D. Hsu, H. Soh, Robot capability and intention in trust-based decisions across tasks, in: 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), IEEE, 2019, pp. 39–47.

[8] A. Steinfeld, T. Fong, D. Kaber, M. Lewis, J. Scholtz, A. Schultz, M. Goodrich, Common metrics for human-robot interaction, in: Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction, ACM, Salt Lake City, Utah, USA., 2006, pp. 33–40.

[9] D. R. Billings, K. E. Schaefer, J. Y. Chen, P. A. Hancock, Human-robot interaction: developing trust in robots, in: Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction, ACM, Boston, Massachusetts, USA., 2012, pp. 109–110.

[10] J. S. Metcalfe, A. R. Marathe, B. Haynes, V. J. Paul, G. M. Gremillion, K. Drnec, C. Atwater, J. R. Estepp, J. R. Lukos, E. C. Carter, W. D. Nothwang, Building a framework to manage trust in automation, in: Micro- and Nanotechnology Sensors, Systems, and Applications IX, Vol. 10194, SPIE, Bellingham, WA, 2017, p. 101941U. doi:10.1117/12.2264245.

[11] H. Zhao, H. Azevedo Sá, C. Esterwood, X. J. Yang, L. Robert, D. Tilbury, Error type, risk, performance and trust: Investigating the impacts of false alarms and misses on trust and performance, in: In Proceedings of the Ground Vehicle Systems Engineering and Technology Symposium (GVSETS 2019), NDIA, Novi, MI, 2019, pp. 1–8.

[12] B. M. Muir, Trust between humans and machines, and the design of decision aids, International Journal of Man-Machine Studies 27 (5-6) (1987) 527–539. doi:10.1016/S0020-7373(87)80013-5.
URL https://www.sciencedirect.com/science/article/pii/S0020737387800135

[13] J. D. Lee, N. Moray, Trust, self-confidence, and operators' adaptation to automation, International Journal of Human-Computer Studies 40 (1) (1994) 153–184. doi:10.1006/IJHC.1994.1007.
URL https://www.sciencedirect.com/science/article/pii/S107158198471007X?via%3Dihub

[14] B. Barber, The logic and limits of trust, Vol. 96, Rutgers University Press, New Brunswick, NJ, 1983.

[15] J. K. Rempel, J. G. Holmes, M. P. Zanna, Trust in Close Relationships, Tech. Rep. 1, APA (1985).
URL https://pdfs.semanticscholar.org/4727/fcf320e6f8c3a8bbd9d7bac22708825f48ad.pdf

[16] A. B. Christensen, C. R. Dam, C. Rasle, J. E. Bauer, R. A. Mohamed, L. C. Jensen, Reducing overtrust in failing robotic systems, in: 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), IEEE, 2019, pp. 542–543.

[17] B. M. Muir, N. Moray, Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation, Ergonomics 39 (3) (1996) 429–460. doi:10.1080/00140139608964474.
URL http://www.tandfonline.com/doi/abs/10.1080/00140139608964474

[18] K. Hoff, M. Bashir, A theoretical model for trust in automated systems, in: CHI '13 Extended Abstracts on Human Factors in Computing Systems on - CHI EA '13, ACM Press, New York, New York, USA, 2013, p. 115. doi:10.1145/2468356.2468378.
URL http://dl.acm.org/citation.cfm?doid=2468356.2468378

[19] W.-L. Hu, K. Akash, T. Reid, N. Jain, Computational Modeling of the Dynamics of Human Trust During Human-Machine Interactions, IEEE Transactions on Human-Machine Systems 1 (1) (2018) 1–13. doi:10.1109/THMS.2018.2874188.
URL https://ieeexplore.ieee.org/document/8502860/

[20] L. Petersen, H. Zhao, D. Tilbury, X. J. Yang, L. Robert, et al., The influence of risk on driver's trust in semi-autonomous driving, in: In Proceedings of the Ground Vehicle Systems Engineering and Technology Symposium (GVSETS 2018), NDIA, Novi, MI, 2019, pp. 1–7.

[21] L. Petersen, D. Tilbury, X. J. Yang, L. Robert, Effects of augmented situational awareness on driver trust in semi-autonomous vehicle operation, Proceedings of the 9th Ground Vehicle Systems Engineering and Technology Symposium 1 (1) (2017) 1–8.

[22] H. Swaminathan, H. J. Rogers, Estimation procedures for hierarchical linear models, Multilevel modeling of educational data (2008) 469–519.