Prospectus

Measuring Real-time Trust in an Autonomous Driving Simulator
(Technical Topic)

Analysis of the Role of Trust in the Acceptance of Driverless Cars
(STS Topic)

By

Hill Duggan

31 October, 2019

Technical Project Team Members: Jack McManus

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

Signed: Hill Duggan

Approved: _____ Date 11/20/2019

Approved: _____ Date _____

Introduction

The rapidly growing popularity of advanced driver assistance systems (ADAS) technology signals society's foray into the realm of autonomous driving, with the logical endgame being the widespread availability of fully self-driving cars. The implications of a driverless society are profound: "drastic reduction of accidents, deaths and injuries; giving access to mobility to minorities and disabled people; boosting the economy; a reduction in traffic congestion and pollution" (Stewart, Musa, & Croce, 2019). 94% of serious crashes are due to human error, a statistic that further illustrates the life-saving potential of automated driving systems (National Highway Traffic Safety Administration, 2019). Today, it is hardly rare to see cars on the road that make use of ADAS technologies like automatic lane keeping, blind spot detection, or automatic emergency braking; however, when it comes to fully automated cars, the public is more tentative. A Cox Automotive study (2018) showed that of those polled, only 16% would feel comfortable allowing an autonomous vehicle drive them without the option to take over. To better understand both the causes and implications of this problem of public trust, I have been working with a research team to design and conduct experiments on human test subjects, with the goal of measuring human trust in the context of an autonomous driving simulator.

Years of research have been poured into making autonomous driving feasible, safe, and ultimately ubiquitous. However, while the continuation of research and technical improvement is paramount, one of the primary obstacles to overleap is social: garnering public trust of autonomous vehicles. By understanding the role and importance of human trust as a factor in this issue, researchers and developers can engender a reputation of safety surrounding automated vehicles; only then will the public be receptive to adopting change on such a large scale.

In order to understand the role of human trust in the prospect of societal absorption of autonomous vehicles, the technical and social factors are inextricably linked, and thus equally important to consider. It is critical to view the development of self-driving cars as a part of a larger lattice, where both technical and human factors are meshed together. Moreover, it is important to not only measure human trust of self-driving cars, but to understand its implications. Below I explain a technical solution to improve the collection of real-time trust data in a driving simulator. Additionally, using Actor Network Theory as a framework, I assess the multidimensional problem of public acceptance of autonomous vehicles.

Technical Problem

Given the high cost and potential danger of real-life testing of autonomous driving systems, it is often more practically feasible to perform research within the context of some sort of driving simulator. Simulators afford researchers the ability to devise their own driving algorithms, construct their own road networks, and repeatedly test variations of specific scenarios that may be hazardous or expensive to test in real life (Silver, 2018). Current ADS research frequently involves placing human test subjects in an autonomous driving simulator, running them through a virtualized scenario, and collecting sensory data for analysis. A subset of ongoing research utilizes a semi-autonomous simulator, allowing the test subjects to switch between manual and autonomous driving modes based on their trust of the system to handle incidents programmed into the driving tracks.

The goal of many studies is to observe the effects of driving simulator features on user trust. In the context of technical research, trust requires a specific definition, which we've accepted to be the subject's delegation of responsibility for actions to the autonomous system, and willingness to accept the potential risk and uncertainty. However, even in studies that

specifically aim to investigate the role of trust in a driving system, relevant data is only collected in the form of pre and post-experiment surveys (Koo et al., 2014). Moreover, an important metric that current researchers fail to keep track of is a real-time user-reported indicator of trust of the driving simulator. There are extensive benefits to considering real-time trust in conjunction with pre and post-experiment trust, namely the ability to directly observe *what* specific factors in an autonomous driving simulator positively or negatively impact a user's trust, and to what degree they affect it.

The goal of my research has been to design and conduct human subject experiments in a driving simulator to study trust and perceived safety of an autonomous driving simulator. Prior to experimentation, we used survey data to form a partially observable Markov decision process (POMDP) with human trust as a latent variable. The trust-POMDP model provides a principled approach for the system to infer the trust of a human subject through interaction, reason about the effect of its own actions on human behaviors, and choose actions that maximize user trust over the long term (Chen, Nikolaidis, Soh, Hsu, Srinivasa, 2018). We constructed a road network based on the POMDP model. The system brings the subject from a fixed starting point to a fixed finish point, making different routing decisions based on whether the subject trusted the autonomous algorithm to handle certain incidents (e.g. a pedestrian crossing or an obstacle in the road), or felt it was necessary to switch from autonomous mode to manual mode. While in autonomous mode, the participant is instructed to rate his/her current trust of the system on a scale from 1 to 7 (1 being strong distrust, 7 being full trust.) This is done in real time using two buttons on the steering wheel, the left being to decrement the trust score and the right being to increment. The conductors of the experiment take notes on any fluctuations in trust and switches to or from manual mode.

STS Problem

In March of 2018, 49 year-old Elain Herzberg died after being struck by an Uber self-driving car that failed to perform an emergency stop; a US federal investigation later determined that the car's emergency stop system was disabled entirely. Following the highly publicized disaster, the percentage of Americans who reported they would be too afraid to ride in a self-driving car rose from 63% to 73% (Hawkins, 2018). A case like this creates a major setback for the emerging market of autonomous driving systems, leading many people to question their safety and reliability; as of now, it is an open question whether or not self-driving cars will be successfully integrated into society.

Autonomous vehicles are one of the most potentially disruptive innovations on this generation's technological horizon (Cox, 2017). Many newer car models already have aspects of automation as design features, such as automatic lane-keeping, automatic braking, and blind-spot detection. However, scientists agree that there are many essential hurdles to overleap before fully driverless cars become a market mainstay. The 2018 Uber crash highlights one of current autonomous vehicles' shortcomings: pedestrian handling. For autonomous vehicles to be significantly safer than manual vehicles there is also a necessity for improvements in human driver handling, car-to-car connectivity, and mitigation of car-hacking (Moldrich & Woollaston, 2018). Swift progress is being made in these functional issues, and scholars are correct to acknowledge the importance of addressing them. However, one of the largest obstacles in the way of autonomous vehicles isn't functional, and it goes underemphasized by scholars in the field: human trust.

Recognizing the importance of the acceptance of target consumers is integral to the success of any emerging technology (Taherdoost, 2017). If scientists continue to over-centralize

the purely technical deficiencies autonomous vehicles, they may fail to understand the critical role consumer trust plays in their prospective transition to consumer availability. Using Actor-Network Theory as a lens through which to view this problem, I argue that it is not only the technological limitations of autonomous vehicles, but also public trust that lies in the way of the aforementioned transition. As a framework, Actor Network Theory "attempts to 'open the black box' of science and technology by tracing the complex relationships that exist between governments, technologies, knowledge, texts, money, and people (Cressman, 2009)." Autonomous vehicles exist as a technological network with human and non-human actors. The network's impetus, or builder, is the subset of the automotive industry that is currently working to develop and roll out self-driving cars; Tesla and Uber are examples of companies within this actor. There are many other actor-networks within this broader web, including relevant legislature, cultural ideals, and self-driving cars themselves. However, the actor-network that I argue holds the most power is the potential consumers. The stabilization of the autonomous driving system network is contingent on public trust and acceptance; thus, it is critical to analyze the power-dynamic between the consumers and the researchers and producers.

Conclusion

This technical report will delineate a new method to measure and analyze real-time trust in an autonomous driving simulator. This will allow researchers to see with specificity what factors causes users to trust or distrust self-driving cars. In the STS report, I employ Actor Network Theory to examine the role of trust emerging autonomous driving system network, especially as it pertains to the fluctuating power dynamic between self-driving car researchers and public consumers.

The findings of the technical report will aid in the problem of measuring the fickle variable of human trust. Once trust has been measured, the results of the STS report offer a perspective with which to understand it. In doing so, we can elucidate the grander role public trust plays in the stabilization of not just the autonomous driving industry, but also any burgeoning technological network.

Word Count: 1564

References

Autonomous Vehicle Awareness Rising, Acceptance Declining. (2018, August 16). Retrieved October 18, 2019, from https://www.coxautoinc.com/news/evolution-of-mobility-study-autonomous-vehicles/.

Chen, M., Hsu, D., Nikolaidis, S., Srinivasa, S., & Soh, H. (2018). Planning with Trust for Human-Robot Collaboration. *HRI '18 Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, 307–315. Retrieved from https://humanrobotinteraction.org/2018/

Cox, L. (2017, June 14). Autonomous Vehicles - The most disruptive innovation of a generation. Retrieved October 20, 2019, from https://disruptionhub.com/autonomous-vehicles-disruptive-innovation-generation/.

Cressman, D. (n.d.). A Brief Overview of Actor-Network Theory: Punctualization, Heterogeneous Engineering & Translation.

Hawkins, A. J. (2018, May 22). Self-driving car crashes put a dent in consumer trust, poll says. Retrieved October 20, 2019, from https://www.theverge.com/2018/5/22/17380374/self-driving-car-crash-consumer-trust-poll-aaa.

Koo, J., Kwac, J., Ju, W., Steinert, M., Leifer, L., & Nass, C. (2014). Why did my car just do that? Explaining semi-autonomous driving actions to improve driver understanding, trust, and performance. *International Journal for Interactive Design and Manufacturing*. doi: 10.1007/s12008-014-0227-2

Lynberg, M. (n.d.). Automated Vehicles for Safety. Retrieved October 18, 2019, from https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety.

Silver, D. (2018, November 1). Simulation Becomes Increasingly Important For Self-Driving Cars. Retrieved October 19, 2019, from https://www.forbes.com/sites/davidsilver/2018/11/01/simulation-becomes-increasingly-important-for-self-driving-cars/#2c05bdd05583.

Stewart, L., Musa, M., & Croce, N. (2019, August 12). Look no hands: self-driving vehicles' public trust problem. Retrieved October 18, 2019, from https://www.weforum.org/agenda/2019/08/self-driving-vehicles-public-trust/.

Taherdoost, H. (2017). Elsevier Procedia Manufacturing Volume 22, 2018, Pages 960-967 Procedia Manufacturing A review of technology acceptance and adoption models and theories. *International Conference Interdisciplinarity in Engineering*, 960–967. Retrieved from https://www.sciencedirect.com/

Woollaston, V., & Moldrich, C. (2018, October 18). Driverless cars of the future: How far away are we from autonomous cars? Retrieved from https://www.alphr.com/cars/1001329/driverless-cars-of-the-future-how-far-away-are-we-from-autonomous-cars.