Human Trust Measurement Using an Immersive Virtual Reality Autonomous Vehicle Simulator

Shervin Shahrdar
Corey Park
Mehrdad Nojoumian
sshahrda,cpark7,mnojoumian@fau.edu
Department of Computer & Electrical Engineering and Computer Science
Florida Atlantic University
Boca Raton, Florida

ABSTRACT

Recent studies indicate that people are negatively predisposed toward utilizing autonomous systems. These findings highlight the necessity of conducting research to better understand the evolution of trust between humans and growing autonomous technologies such as self-driving cars (SDC). This research presents a new approach for real-time trust measurement between passengers and SDCs. We utilized a new structured data collection approach along with a virtual reality SDC simulator to understand how various autonomous driving scenarios can increase or decrease human trust and how trust can be re-built in the case of incidental failures. To verify our methodology, we designed and conducted an empirical experiment on 50 human subjects. The results of this experiment indicated that most subjects could rebuild trust during a reasonable time frame after the system demonstrated faulty behavior. Our analysis showed that this approach is highly effective for collecting real-time data from human subjects and lays the foundation for more-involved future research in the domain of human trust and autonomous driving.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**; *Virtual reality*.

KEYWORDS

Real-time trust measurement; human-autonomous vehicle interaction; self-driving cars.

ACM Reference Format:
Shervin Shahrdar, Corey Park, and Mehrdad Nojoumian. 2019. Human Trust Measurement Using an Immersive Virtual Reality Autonomous Vehicle Simulator. In *AAAI/ACM Conference on AI, Ethics, and Society (AIES '19), January 27–28, 2019, Honolulu, HI, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3306618.3314264

AIES '19, January 27–28, 2019, Honolulu, HI, USA

ACM ISBN 978-1-4503-6324-2/19/01…$15.00
https://doi.org/10.1145/3306618.3314264

1 INTRODUCTION

Recent studies indicate that people have negative attitudes toward utilizing autonomous platforms [8, 11]. Besides, with the exponential growth and the increase in the complexity of autonomous systems in the 21st century, managing trust of users in such systems has become an important concept when designing new autonomous or artificial intelligence systems. Numerous studies in the domain of trust and intelligent systems have suggested that management and constant improvement of this mutual trust between autonomous systems and their users will be one of the primary challenges the industry professionals will face when trying to popularize the use of fully autonomous systems [1, 2, 4, 6, 10]. These discoveries highlight the necessity and urgency of conducting research to better understand the evolution of trust between humans and growing autonomous technologies, and to provide technologies that are responsive to human trust.

A concrete example of a trust management problem is the systematic maintenance of trust in self-driving cars. Car manufacturers and tech giants (for instance, Mercedes Benz, BMW, Tesla, Volvo, Waymo, etc.) have successfully manufactured semi-autonomous cars, and have been working on level-4 and level-5 fully autonomous prototypes since the early 2010s. Many of these corporations have projected the mass production of SDCs in the early 2020's [7, 16, 17]. Their major challenge in the upcoming years will be to attract the attention of average consumers in the US and around the world, who have high expectations, but at the same time, a high level of distrust in fully automated SDCs [5].

According to a World Economic Forum study, consumers are very reluctant to consider purchasing or even trying autonomous vehicles. Safety, control, and faulty behavior of autonomous cars are some of the many concerns the consumers expressed. With the recent trust damaging incidents, e.g., Tesla[1] and Uber[2] SDCs accidents, the need for additional research to provide a safer test environment and to manage human-machine trust becomes more important than ever. New research objectives and innovative methodologies can potentially provide a robust platform to develop autonomous vehicles that perform well and are trustworthy.

Generally speaking, the literature of trust is very broad. However, very little research has been conducted on physiological responses

[1] https://www.theguardian.com/technology/2018/mar/31/tesla-car-crash-autopilot-mountain-view
[2] https://www.nytimes.com/2018/03/23/technology/uber-self-driving-cars-arizona.html

and fluctuations in the trust levels of SDCs' passengers [18]. Furthermore, most research in this area has utilized online surveys and primitive simulations for their experiments rather than realistic and immersive simulations. The open questions and the gaps in the current literature motivated us to initiate a new project by designing an empirical experiment for measuring the trustworthiness of a simulated SDC. By implementing an immersive VR simulator, which uses real VR driving videos, and developing an advanced trust self-reporting tool, we intended to collect accurate data from test subjects. We asked the participants to sit in our SDC simulator and use the trust self-reporting tool after they experience different simulated SDC driving segments. We believe that our investigation and its outcome will contribute to the general understanding of factors affecting trust and satisfaction among passengers of SDCs.

In this paper, we will describe the design, the procedure and the outcome of the empirical experiment that we have conducted. Section 2 illustrates our research methodology. In Section 3, the detailed simulation setup and its technical configurations are described. Section 4 covers the experimental design. The results of our research are presented in Section 5. Finally, in Section 6, we end the paper with the concluding remarks and the future direction of our research.

2 RESEARCH METHODOLOGY

2.1 Novelty of Our Approach

The sequential and structured data collection, various trust states, and a realistic simulation platform are novel aspects of our research.

(1) Our data was collected based on a limited (to avoid allowing our human subjects to provide inaccurate responses) sequence of trust-building/damaging incidents that affect each other. This helped us understand how the human mind goes from one specific trust-state to another one in a sequence of events. We utilize specific templates from [13] to form these incidents.

(2) Although most approaches [3, 9] utilize two forms of responses from subjects, i.e., distrust and trust, we consider a fuzzy set of trust states, i.e., distrust, somehow distrust, neutral, somehow trust, and trust.

(3) In many "trust-in-autonomy" projects, subjects were asked to respond to questions in a survey or interact with an algorithm to express their inputs [18], while our SDC simulator is fully immersive.

2.2 Sequential and Structured Data Collection

Using a structured and sequential data collection approach, we intended to understand how humans gain or lose trust in autonomous vehicles and how trust escalation or reduction can be controlled in various incidents as well as among different groups of people, i.e., young, mid-age and senior. Our collected data can be transformed into specifications to be used in the controllers of autonomous vehicles. Note that demographic data and past psychological data are collected prior to our experiments to characterize our human subjects based on self-confidence, trusting attitude, risk-acceptance, past unpleasant experiences, and other traits because these factors impact the outcome of our research. Segments are categorized into five distinct groups as follow:

(1) **Initial Trust:** Segments that aim to capture the initial trust of passengers in the first few minutes of the first interaction.

(2) **Trust Escalation:** Segments that illustrate a sequence of incidents in which human subject's trust is increased, e.g., 2 minutes of smooth and predictable driving by the SDC without any complications or surprises.

(3) **Trust Reduction:** Segments that illustrate a sequence of incidents in which the human subject's trust is decreased, e.g., when the SDC cuts off another controlled vehicle.

(4) **Trust Mutation:** A sequence of mild incidents (e.g., a rapid lane change by the SDC) followed by critical incidents (e.g., stop-sign violation or tailgating by the SDC) and vice versa, can be negative or positive incidents.

(5) **Re-Building Trust:** Segments that demonstrate how trust can be rebuilt, e.g., the SDC performs smoothly for a reasonable period of time after trust-damaging incidents.

3 SIMULATION SETUP

Our simulator, as shown in Figure 1, is a safe platform to expose human subjects to any trust-damaging incident, including but not limited to, sharp turns, sudden stops, stop-light violations, speeding, tailgating, unexpected accident, etc. The SDC virtual reality simulator is based on fusion of an Oculus Rift headset with an Atomic A3 Full Motion Simulator. Figure 1 shows a human subject in our self-driving car simulator.

360 degree video of driving situations were recorded using the GoPro Fusion Camera and edited using the GoPro Fusion Studio. To capture interesting driving footage and trust damaging scenarios, the team members recorded 360 degree videos of everyday driving for a couple of months. These videos were later analyzed, edited, and categorized to be used in the SDC simulator. The team members also choreographed and recorded driving segments (for instance, near collisions between two cars) in safe environments with no outside traffic.

The Oculus Rift head set outputs 1080x1200 resolution per eye, at a 90 Hz refresh rate, a 110 degree field of view, and has headphones which output a 3D audio effect[3]. The participant also wears noise canceling ear muffs over the headphones to eliminate outside noise. While wearing the Oculus Rift VR headset, the participant can freely move their head 360 degrees to see the complete scene. Figure 2 demonstrates the view of a scene in the SDC simulator.

The Atomic A3 Full Motion Simulator can move up to 71 degrees per second across a full 27 degree dual-axis movement range[4]. The combination of complete visual, audio, and movement immersion provides a convincingly realistic simulation. The Atomic A3 Simulator receives telemetry data that has pitch values for front and back movements and roll values for left and right movements.

SimTools motion simulator software was used to send the telemetry data for each video to the Atomic A3 motion simulator via UDP packets. The "Video Ride Creator" plug-in was used to generate telemetry points for every frame in the simulation videos.

[3]https://gopro.com/fusion
[4]http://www.atomicmotionsystems.com/

Figure 1: Participant using the SDC simulator.

Figure 2: View from inside the self-driving car simulator: each frame represents the participant's view as they turn their head to look around, illustrating the 360-degree view inside the SDC simulator.

The simulator plays the audio, video, and telemetry files synchronously. After each segment, the participant is presented with a Likert Scale that appears inside the Oculus Rift. The participant selects their response by focusing their gaze on the desired answer for five seconds while wearing the Oculus Rift. Figure 3 shows a sample question in the virtual reality space.

VR sickness is a known phenomenon in which people experience symptoms that are very similar to motion sickness. Symptoms include headache, general discomfort, nausea, vomiting and vertigo [12]. To mitigate these effects the motion output of the simulator was closely monitored, and subjects were not kept in the simulation for longer then 15 minutes.

To play 360 degree 4K resolution videos in each lens of the Oculus Rift, it requires a machine with powerful processing. To meet these demands, the application is executed on an AlienWare Area-51 equipped with an Intel Core i7-5960X processor along with dual

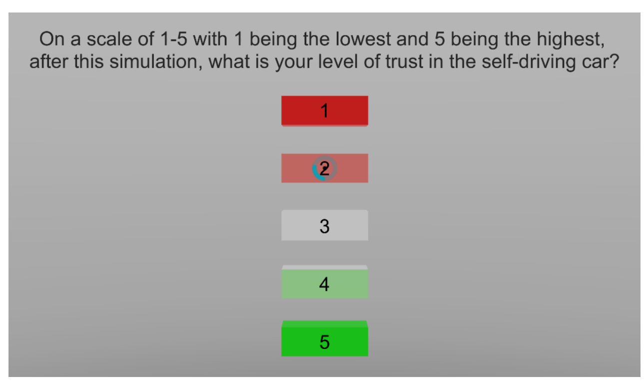

Figure 3: Stare-and-Select tool interface.

NVIDIA GeForce GTX Titan Z graphics cards that achieves clock speeds greater than 4 Ghz. This high performance machine allows for the dual 360 degree 4k videos to be played seamlessly.

4 EXPERIMENTAL DESIGN

Prior to the simulation, participants were asked to answer 17 demographic and psychological questions by filling out an anonymous survey. Participants were randomly placed in one of two possible SDC simulation scenarios. Each scenario is made up of 5 segments. Tables 1 and 2 define the scenario and segment pairings. Specific scenario-segment pairs are denoted with a two letter abbreviation followed by the scenario and segment numbers, for example, TR_{I-II} denotes trust reduction segment 2 of scenario 1.

Each segment starts with an exposure to an approximately 2-minute SDC driving simulation followed by a response interval to the question “On a scale of 1-5 with 1 being the lowest and 5 being the highest, after this simulation, what is your level of trust in the self-driving car?" After the participant responds, the application moves on to the next segment until the simulation scenario is complete. Different videos are used across driving scenarios.

Table 1: Simulation Scenario-1

IT_{I-I}	Initial Trust
TR_{I-II}	Trust Reduction
TR_{I-III}	Further Trust Reduction
NM_{I-IV}	Negative Trust Mutation
RT_{I-V}	Rebuild Trust

Table 2: Simulation Scenario-2

IT_{II-I}	Initial Trust
TE_{II-II}	Trust Escalation
TR_{II-III}	Trust Reduction
NM_{II-IV}	Negative Trust Mutation
RT_{II-V}	Rebuild Trust

An initial trust/trust escalation segment involved the SDC moving slowly and predictably while adhering to the rules of the road. A trust reduction segment involved the SDC along with Human-Driving Cars (HDC) moving erratically and unpredictably, breaking rules of the road including speeding, tailgating, and sudden lane changes. In the NM_{I-IV} segment, the SDC ran through a non-visible stop sign and nearly collided with another car and then proceeded to drive through a residential neighborhood. In the NM_{II-IV} segment, the SDC ran through a stop sign unexpectedly and detected a pedestrian and a bicyclist crossing a crosswalk and abruptly came to a stop. A rebuild trust segment involved the SDC driving defensively and adhering to rules of the road. Note that HDCs were involved in all scenarios.

It is predicted that after the initial trust/trust escalation segments, the participants will respond with high levels of trust in the SDC, and after trust reduction segments, the participant will respond with low levels of trust in the SDC. It is also predicted that after the negative trust mutation segment, the participant will report a drastic decrease in trust.

5 EXPERIMENTAL RESULTS

Fifty human subjects were recruited to participate in the 10 minute VR autonomous driving simulation[5]. 84% of the participants were male and between the ages of 18-30. Ethnically, the participants identified as 60% White, 20% Hispanic/Latino, 12% Black/African American, and 6% as Other. Half of the participants were randomly selected to be in Scenario-1 and the other half were assigned to Scenario-2. Full findings detailing how demographic and psychological data affect trust levels in the SDC simulator are detailed in the final journal publication [19].

5.1 Scenario-1

Figure 4 shows the mean trust levels participants reported after each segment in scenario-1. In the initial trust segment (IT_{I-I}), participants responded with an average score of 4.52±0.17, followed by a mean score of 3.60±0.23 in the first trust reduction segment (TR_{I-II}). After exposure to further trust reduction (TR_{I-III}), the score increased slightly to 3.84±0.19, followed by a large decline to 2.28 ±0.23 when exposed to negative trust mutation (NM_{I-IV}). Finally, trust levels increased to 4.08±0.19 in the rebuild trust (RT_{I-V}) segment. The most obvious change across segments was between the negative trust mutation (NM_{I-IV}) and the initial trust segment (IT_{I-I}), consistent with our expectations that erratic driving has the potential to severely reduce trust.

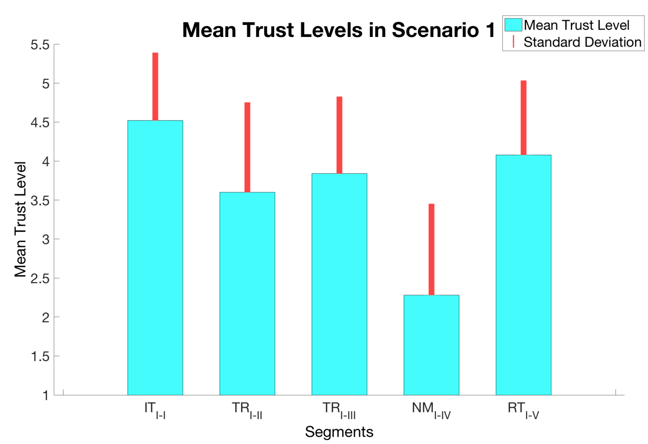

Figure 4: Mean Trust Levels Across Segments in Scenario-1.

To assess whether the observed scores are statistically different, the Wilcoxon Rank Sum Test was performed across segments; the resulting p-values are shown in Figure 5. Here, we see that scored changes in trust are not distinguishable above the 0.05 p-value between TR_{I-II} and TR_{I-III}, RT_{I-V} and TR_{I-II}, and RT_{I-V} and TR_{I-III}. All other comparisons show statistically significant changes in trust.

[5]IRBNET ID #: 1187756-1

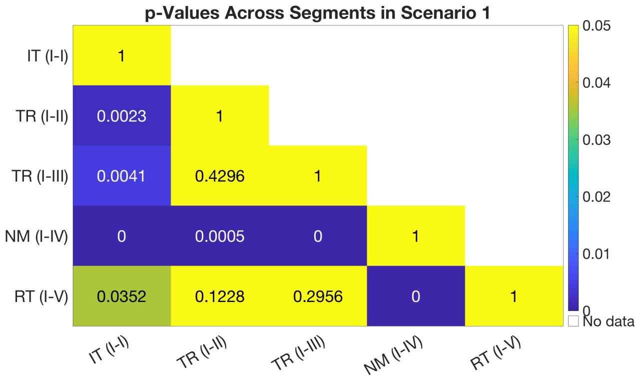

Figure 5: Wilcoxon Rank Sum Test P-Values Across Segments in Scenario-1.

Scenario-1 performed as expected. Participants scored the initial trust and rebuild trust segments with high levels of trust, the trust reduction segments with lower levels of trust, and the negative trust mutation segment with the lowest level of trust. It is interesting that mean values of trust appear to have slightly increased in TR_{I-III}, a segment designed to elicit further trust reduction. This may be due to participants perceiving TR_{I-II} to be more dangerous in comparison to TR_{I-III}; however, the difference between the two is within the standard error and statistically insignificant. As expected, the negative trust mutation had the lowest trust levels and was significantly lower than all other segments. This indicates the simulator's effectiveness in reducing participants' trust levels.

An interesting result is the difference between the initial trust segment and the final segment designed to rebuild trust. While participants scored their level of trust after RT_{I-V} at 4.08±0.19, a high value, it is significantly lower than the initial trust value (4.52±0.17), representing a 12.00% decrease. This seems to indicate that participants trusted the SDC less after being exposed to trust reducing segments.

5.2 Scenario-2

Figure 6 shows the mean trust levels participants reported after each segment in Scenario-2. In the initial trust segment (IT_{II-I}), participants responded with an average score of 4.24±0.15, followed by a mean score of 4.52±0.14 in the first trust escalation segment (TE_{II-I}). After exposure to trust reduction (TR_{II-III}), the score decreased to 3.36±0.20, followed by a further decline to 3.12 ±0.26 when exposed to the negative trust mutation (NM_{II-IV}). Finally, trust levels increased to 3.80±0.23 in the Rebuild Trust segment (RT_{II-V}).

Figure 7 shows the results of the Wilcoxon Rank Sum Test across segments in Scenario-2. Here, we see that scored changes in trust are not distinguishable above the 0.05 p-value between TE_{II-II} and IT_{II-I}, RT_{II-V} and IT_{II-I}, NM_{II-IV} and TR_{II-III}, RT_{II-V} and TR_{II-III}, and RT_{II-V} and NM_{II-IV}. All other inter-comparisons show statistically significant changes in trust.

In (NM_{II-IV}), the SDC approaches a crosswalk and stops for a pedestrian to cross the street. This was the only segment that involved the SDC interacting near a pedestrian. Participants reported

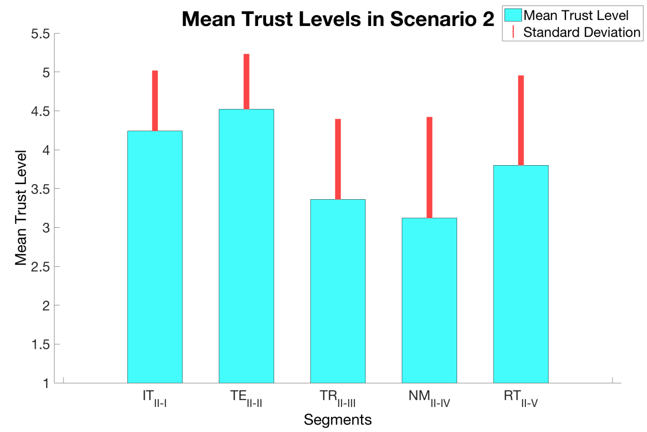

Figure 6: Mean Trust Levels Across Segments in Scenario-2.

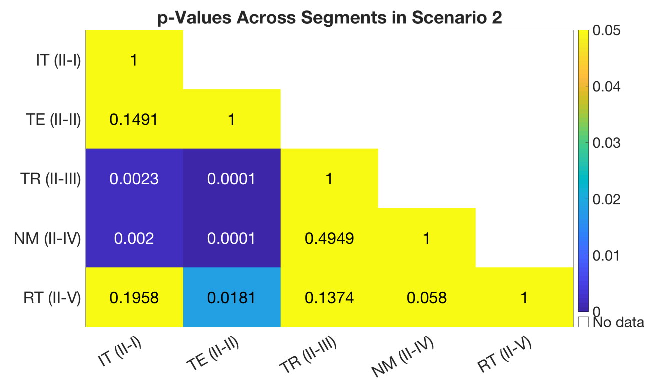

Figure 7: Wilcoxon Rank Sum Test: P-Values Across Segments in Scenario-2.

low levels of trust after this segment and commented that they especially did not trust the SDC near pedestrians. In Scenario-2, while participants score their level of trust after RT_{II-V} at 3.80 ±0.23, a high value, it is significantly lower than the initial trust segment (4.24 ±0.15), representing a 13.50% decrease. This seems to indicate that participants trusted the SDC less after being exposed to trust reducing segments.

5.3 Comparison of Scenario-1 and Scenario-2

We performed the Wilcoxon Rank Sum Test p-Value between the trust segments in Scenario 1 and Scenario 2 to investigate the fluctuations of trust levels in participants in groups 1 and 2. Figure 8 illustrates this comparison.

Our analysis indicated that participants from both groups reported similar expected trust levels after positive and negative trust segments. The only major significant difference that we observed was the reported trust levels for Negative Trust Mutation segments, that is, NM_{I-IV} and NM_{II-IV}. We believe that this is due to the fact that the Negative Trust Mutation segment in Scenario-2, NM_{II-IV}, was much milder compared to the one in Scenario-1. It mostly involved the SDC running stop signs in a parking lot and driving near pedestrians and a bicyclist, as opposed to speeding, almost crashing into another car, and driving in the middle of the road in NM_{I-IV}.

Another considerably significant trust difference can be seen in TR_{I-II} and TE_{II-II}. This is expected because TR_{I-II} involves trust damaging incident such as speeding or tailgating, as opposed to defensive highway driving in TE_{II-II}.

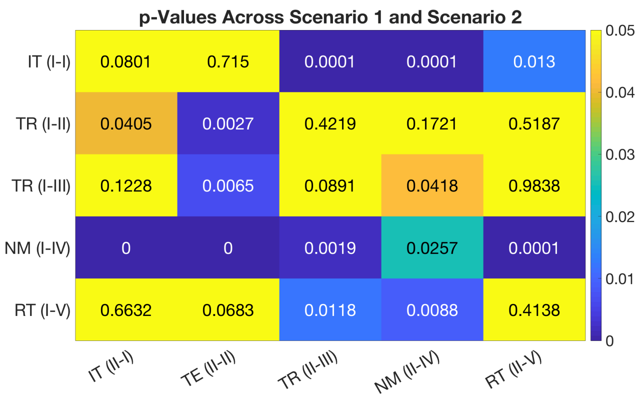

Figure 8: Wilcoxon Rank Sum Test: P-Values in segments from Scenario-1 vs Scenario-2.

The results of the experiment were generally consistent with our expectations. The participants reported higher trust levels after experiencing initial trust and trust escalation segments and reported distrust after the trust reduction segments as well as high distrust after the negative trust mutation segment. Participants did not trust the SDC around pedestrians. Finally, participants in both groups were able to relatively rebuild their trust after the trust damaging Negative Trust Mutation segments.

6 CONCLUSION AND FUTURE WORK

The results of the experiment indicated that the trust levels of humans change depending on the SDCs driving style and that the majority of the participants were able to moderately rebuild their trust in the simulated self-driving car after faulty and erratic behaviors. The autonomous driving style directly influences the trust of the passengers in the system. Aggressive driving diminishes trust, and defensive/predictable driving increases trust.

The results of our experiment matched our initial expectations. Thus, we can consider this innovative data collection approach an adequate and a reliable technique to measure passengers' levels of trust and psychological responses when exposed to different driving scenarios. This approach lays the foundation for a wide variety of future research in the context of trust and human-machine interactions. With the mass production and commercialization of autonomous vehicles in the upcoming years and the high degree of skepticism of the average consumers in the industry, we believe that this type of research and development is more important than ever before.

In the next iterations of our research, we will improve our SDC VR simulation scenarios by introducing segments with crucial failures such as accidents as well as hazardous conditions such as heavy rain, storms, and snow to see if they would have a direct impact on the passengers' trust levels. Furthermore, automatic collection of the physiological and psychological responses (via EEG sensors, heartbeat sensors, facial recognition modules, and others) during the simulated driving scenarios and analyzing them in real-time are some of our primary ongoing research directions [14, 15].

7 ACKNOWLEDGMENT

We greatly thank Florida Atlantic University for the financial support. We also thank the anonymous reviewers for their inspiring and constructive feedback.

REFERENCES

[1] Moaed Abd, Iker Gonzalez, Mehrdad Nojoumian, and Erik Engeberg. 2017. Impacts of Robot Assistant Performance on Human Trust, Satisfaction, and Frustration. In *RSS: Morality and Social Trust in Autonomous Robots*.

[2] Moaed Abd, Iker Gonzalez, Mehrdad Nojoumian, and Erik Engeberg. 2017. Trust, Satisfaction and Frustration Measurements for Real-Time Human-Robot Interaction. In *30th Florida Conference on Recent Advances in Robotics*. 89–93.

[3] Kumar Akash, Wan-Lin Hu, Tahira Reid, and Neera Jain. 2017. Dynamic modeling of trust in human-machine interactions. In *American Control Conference (ACC), 2017*. IEEE, 1542–1548.

[4] Jenay Beer, Arthur D Fisk, and Wendy A Rogers. 2014. Toward a framework for levels of robot autonomy in human-robot interaction. *Journal of Human-Robot Interaction* 3, 2 (2014), 74.

[5] Michelle S Carlson, Munjal Desai, Jill L Drury, Hyangshim Kwak, and Holly A Yanco. 2014. Identifying factors that influence trust in automated cars and medical diagnosis systems. In *AAAI Symposium on The Intersection of Robust Intelligence and Trust in Autonomous Systems*. 20–27.

[6] Jong Kyu Choi and Yong Gu Ji. 2015. Investigating the importance of trust on adopting an autonomous vehicle. *International Journal of Human-Computer Interaction* 31, 10 (2015), 692–702.

[7] Sharon Gaudin. 2012. Autonomous cars will arrive within 10 years, Intel CTO says. https://www.computerworld.com/article/2492744/emerging-technology/autonomous-cars-will-arrive-within-10-years--intel-cto-says.html

[8] Peter A Hancock, Deborah R Billings, Kristin E Schaefer, Jessie YC Chen, Ewart J De Visser, and Raja Parasuraman. 2011. A meta-analysis of factors affecting trust in human-robot interaction. *The Journal of the Human Factors and Ergonomics Society* 53, 5 (2011), 517–527.

[9] Wan-Lin Hu, Kumar Akash, Neera Jain, and Tahira Reid. 2016. Real-Time Sensing of Trust in Human-Machine Interactions. *IFAC-PapersOnLine* 49, 32 (2016), 48–53.

[10] Jeamin Koo, Jungsuk Kwac, Wendy Ju, Martin Steinert, Larry Leifer, and Clifford Nass. 2015. Why did my car just do that? Explaining semi-autonomous driving actions to improve driver understanding, trust, and performance. *International Journal on Interactive Design and Manufacturing (IJIDeM)* 9, 4 (2015), 269–275.

[11] Miltos Kyriakidis, Riender Happee, and JCF De Winter. 2015. Public opinion on automated driving: results of an international questionnaire among 5000 respondents. *Transportation research part F: traffic psychology and behaviour* 32 (2015), 127–140.

[12] Joseph J LaViola Jr. 2000. A discussion of cybersickness in virtual environments. *ACM SIGCHI Bulletin* 32, 1 (2000), 47–56.

[13] Mehrdad Nojoumian. 2015. Trust, Influence and Reputation Management Based on Human Reasoning. In *4th AAAI Workshop on Incentives and Trust in E-Communities (WIT-EC)*. 21–24.

[14] Corey Park, Mehrdad Nojoumian, and Dimitris Pados. 2019. Using EEG and Self-Reported Human Metrics to Measure Passenger Emotional Response in Human-Autonomous Vehicle Interactions. In *Under Review*.

[15] Corey Park, Shervin Shahrdar, and Mehrdad Nojoumian. 2018. EEG-Based Classification of Emotional State Using an Autonomous Vehicle Simulator. In *10th IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*. 297–300.

[16] Philip E Ross. 2017. CES 2017: Nvidia and Audi Say They'll Field a Level 4 Autonomous Car in Three Years. https://spectrum.ieee.org/cars-that-think/transportation/self-driving/nvidia-ceo-announces

[17] Alexandria Sage and Paul Lienert. 2016. Ford plans self-driving car for ride share fleets in 2021. https://www.reuters.com/article/us-ford-autonomous/ford-plans-self-driving-car-for-ride-share-fleets-in-2021-idUSKCN10R1G1

[18] Shervin Shahrdar, Luiza Menezes, and Mehrdad Nojoumian. 2018. A Survey on Trust in Autonomous Systems. In *Computing Conference (CC) (AISC)*, Vol. 857. Springer, 368–386.

[19] Shervin Shahrdar, Corey Park, and Mehrdad Nojoumian. 2019. Real-Time Trust Measurement in Human-Autonomous Vehicle Interactions Using a VR-Motion Simulator and a New Structured Data Collection Approach. (2019).