NEW STRUCTURED DATA COLLECTION APPROACH FOR REAL-TIME TRUST MEASUREMENT IN HUMAN-AUTONOMOUS VEHICLE INTERACTIONS

by

Shervin Shahrdar

A Thesis Submitted to the Faculty of

The College of Engineering & Computer Science

In Partial Fulfillment of the Requirements for the Degree of

Master of Science

Florida Atlantic University

Boca Raton, FL

May 2018

NEW STRUCTURED DATA COLLECTION APPROACH FOR REAL-TIME TRUST MEASUREMENT IN HUMAN-AUTONOMOUS VEHICLE INTERACTIONS

by

Shervin Shahrdar

This thesis was prepared under the direction of the candidate's thesis advisor, Dr. Mehrdad Nojoumian, Department of Engineering & Computer Science, and has been approved by the members of his supervisory committee. It was submitted to the faculty of the College of Engineering & Computer Science and was accepted in partial fulfillment of the requirements for the degree of Master of Science.

SUPERVISORY COMMITTEE:

Mehrdad Nojoumian, Ph.D.
Thesis Advisor

Imadeldin Mahgoub, Ph.D.

Dimitrios Pados, Ph.D.

Nurgun Erdol, Ph.D.
Chair, Department of Engineering & Computer Science

Stella N. Batalama, Ph.D.
Dean, The College of Engineering & Computer Science

Khaled Sobhan, Ph.D.
Dean, Graduate College

April 27, 2018
Date

ACKNOWLEDGEMENTS

I would like to express my sincerest gratitude to my graduate advisor, Dr. Nojoumian for guiding and advising me throughout my graduate studies here at Florida Atlantic University. Dr. Nojoumian helped expand my knowledge in many different areas such as Human-Robot Interaction (HRI), autonomous systems, cryptography, and blockchain technologies. He also steered me in the right direction when I faced difficult research challenges.

I would like to also thank the experts, Dr. Imadeldin Mahgoub and Dr. Dimitrios Pados, who agreed to review my research as my thesis committee members. Also, I would like to thank all my professors here at FAU. Thanks to them, I was able to undoubtedly learn and digest practical and theoretical computer science concepts during my studies.

I would also like to thank my research partner Corey Park for collaborating with me on various projects and coming up with excellent ideas.

Special thanks goes to Ahsan Sanaullah for assisting me with developing various software modules related to this research.

Finally, I would like to express sincere gratitude to my parents, family, and friends for providing unlimited support and motivation throughout my years of study and research. This accomplishment would not have been possible without them.

ABSTRACT

Author:	Shervin Shahrdar
Title:	New Structured Data Collection Approach for Real-Time Trust Measurement In Human-Autonomous Vehicle Interactions
Institution:	Florida Atlantic University
Thesis Advisor:	Dr. Mehrdad Nojoumian
Degree:	Master of Science
Year:	2018

Most of 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 therefore 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 (VR) 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 timeframe after the system demonstrated faulty behavior. Furthermore, we discovered that the cultural background and past trust-related experiences of the subjects affect how they lose or regain their trust in SDCs. Our analysis showed that this model 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.

I wholeheartedly dedicate this thesis to my parents,

to Dr. Nojoumian,

to all my professors,

and my friends and colleagues

NEW STRUCTURED DATA COLLECTION APPROACH FOR REAL-TIME TRUST MEASUREMENT IN HUMAN-AUTONOMOUS VEHICLE INTERACTIONS

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

The rapid growth of technology has resulted in the automation of many daily tasks humans had to perform themselves in the past. From industrial robots used in factories to autopilot systems and self-driving cars, automation continues to aid humans with repetitive, tedious, and monotonous tasks. Every day, automation introduces newer and more sophisticated concepts and systems in different areas of our lives including, but not limited to, our homes, the daily commute, the workplace, and the military.

As automated systems evolve and their levels of complexity advance over time, their presence in various daily activities in humans lives leads to the development of new notions in human-autonomy interactions, for example, trust, satisfaction, and frustration, to name a few. Studies have indicated that one of the primary challenges in successfully integrating advanced autonomous systems and artificial intelligence technologies in humans lives will be the management and development of mutual trust [9].

The possible misuses and abuses that humans would bring into automation technologies is another prominent issue in this domain. According to [10], users can become overly dependent on an automation technology, attempt to use functions that are out of the scope of the system, or not monitor the system adequately. These are due to trust or distrust in these technologies. This highlights the importance of educating those who intend to purchase or use an autonomous technology, mainly with respect to the proper use of these technologies.

Furthermore, the usage of information that is given to users regarding an autonomous system is also important. In [10], human subjects were given information sources regarding the usage of an autonomous system with faulty behavior. This study revealed that when more errors occur, the participants will not use the information provided to them due to a lack of trust. Therefore, a proper trust management strategy would help the users utilize such information even in the presence of errors.

Finally, a fair amount of consideration should be given to the safety of autonomous systems. Indeed, users tend to be unaware of functions that an autonomous vehicle can perform, often due to the over-complexity that appears throughout the system. The authors in [11] state that safety would increase with simplification and possible training through a simulated interface between autonomous systems and drivers.

In this thesis, we study how the passengers of a simulated self driving car respond to various modes of autonomous driving. By implementing a realistic VR SDC simulator as well as an advanced, easy to use trust self reporting tool, we aimed to understand how people gain or lose trust in the simulated SDC, and what factors affect the fluctuations in their trust levels.

This thesis is structured as follows: In Chapter 2, a survey of major studies in the domain of trust and autonomous systems is presented. Chapter 3 describes our simulation setup in detail, the equipments we used, and the challenges we faced when developing our simulation. Chapter 4 presents a new approach for measuring and managing trust in autonomous vehicles by utilizing VR technology, our experimental approach, followed by our experiment and its results. Finally, Chapter 5 contains the limitations in the current literature, new ideas, and the future direction of the research discussed in this thesis.

1.1 OUR MOTIVATION

Recent studies indicate that people generally have negative attitudes toward utilizing autonomous platforms such as carriers with autopilot modes; according to recent findings by researchers at Chapman University[1], Americans expressed the highest levels of fear about man-made disasters followed by fears about technology. Furthermore, with the exponential

[1] http://www.chapman.edu/wilkinson/research-centers/babbie-center/survey-american-fears.aspx

growth and increase in the complexity of autonomous systems in the 21st century, managing the trust of users in such systems has become an important concept when designing new AI and autonomous systems. Numerous studies in the domain of trust and AI have suggested that the management and constant improvement of this mutual trust between autonomous systems and their users will be one of the primary challenges faced by industry professionals when attempting to popularize the use of fully autonomous systems [12, 13, 9]. These interesting discoveries highlight the necessity and urgency of conducting research to better understand the evolution of trust between humans and developing autonomous technologies and to provide technologies that are responsive to human trust.

A concrete example of the trust management problem that we mentioned previously would be the systematic maintenance of trust in SDCs. Car manufacturers and tech giants (e.g., Tesla, Mercedes Benz, GM, Volvo, Waymo, and Intel) 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 [14, 15, 16]. 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 not only high expectations but also a high level of distrust in fully automated SDCs [17].

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 the autonomous car are some of the many concerns the consumers expressed (see Figure 1.1 for the detailed break down of the consumer concerns in this study). These are valid consumer concerns. With the recent trust damaging, fatal car accidents involving SDCs by Tesla and Uber [18] [19], the need for additional research to provide a safer test environment and managing 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.

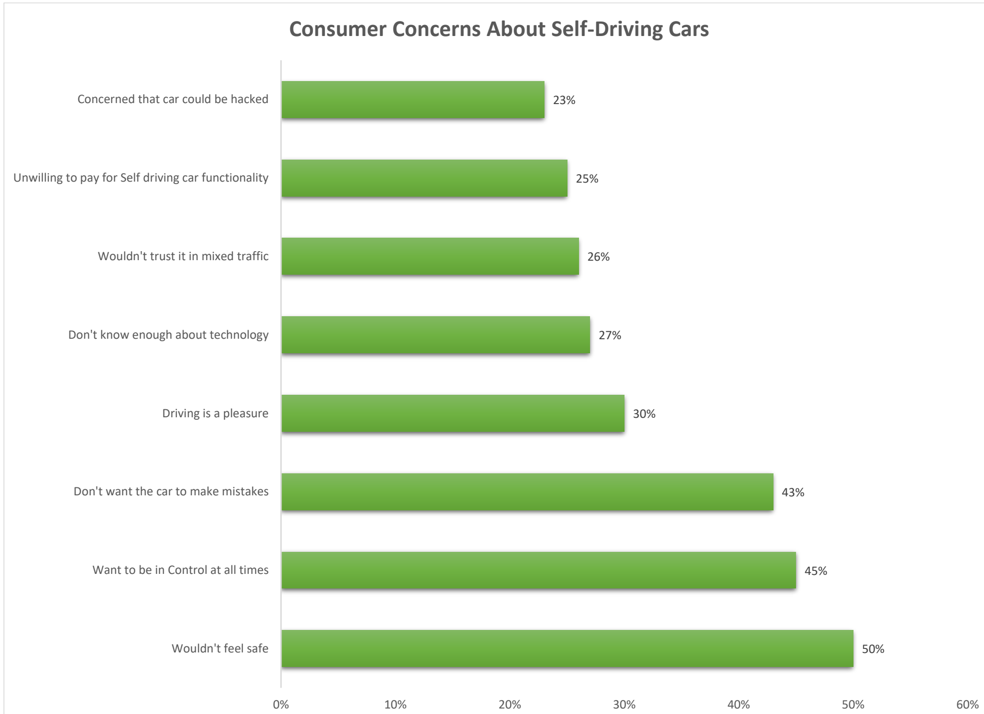

Figure 1.1: Top Reasons for Consumer Distrust in SDCs [1]

In the domain of trust, AI, and autonomous systems, the literature related to trust is very broad. Very little research has been conducted on physiological responses and fluctuations in the trust levels of passengers of SDCs [20]. Furthermore, most research in this area has utilized online surveys and primitive simulations for their experiments rather than realistic, immersive simulations (See Chapter 2 for a brief review of literature in the domain of SDCs). [2]. The open questions and the gaps in the current literature motivated us to attempt to contribute to filling these gaps by designing an empirical experiment to measure the trustworthiness of a simulated SDC. By implementing an immersive VR simulation, which uses real VR driving videos, and utilizing an advanced trust self-reporting software that we developed, we intended to collect accurate data from test subjects. We asked the par-

[2]In this phase of our research we did not use any devices to measure physiological responses.

ticipants to use our trust self-reporting tool after they experience different simulated SDC driving segments (see Section 4.1). We believe that our investigation and its outcome will contribute to the general understanding of factors affecting trust and satisfaction among passengers of SDCs.

1.2 NOVELTY OF OUR APPROACH

To the best of our knowledge, the following items are novel aspects of our research methodology and approach:

- *Sequential and Structured Data Collection*: 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 (see Figure 4.1) to form these incidents.

- *Various Trust States*: Although most approaches [21, 22] utilize two forms of responses from subjects (distrust and trust), we consider a fuzzy set of trust states (distrust, somehow distrust, neutral, somehow trust, and trust).

- *Realistic Simulation Platforms*: In many "trust-in-autonomy" projects (except the robotic ones), subjects were asked to respond to questions in a survey or interact with an algorithm to express their inputs. To the best of our knowledge, our SDC simulator is the first platform that is utilized in a similar project. In fact, it is a very safe test bed to expose our human subjects to inconvenient trust-damaging incidents.

Note that our intention here is to construct computational (or quantitative) models of trust by experimental methods. We intend to understand principles of trust reduction and escalation in real-time human-autonomy interactions. In the later phases of our research, we will define our models' specifications and construct computational models that are understandable to controllers.

1.3 HIGHLIGHTS OF OUR NEW DISCOVERIES

In this research, we discovered a few interesting things after conducting our experiment on the levels of trust between humans and the different driving styles of SDC, and performing data analysis on the demographic data that we collected prior to the experiment. We list below some of the highlights of our findings (see Chapter 4 for a detailed report of these findings).

1. Females gained or lost their trust in the simulated SDC with a slower pace compared to males, who lost or gained their trust with a faster pace.

2. Subjects from Hispanic/Latino backgrounds gained or lost their trust with a slower pace compared to others, who lost and gained their trust with a faster pace.

3. Subjects who expressed their trust can be easily rebuilt, re-gained their trust with a faster pace versus subjects who mentioned otherwise.

4. Subjects with recent bad trust-related experiences gained or lost their trust with a faster pace whereas, subjects who had bad trust-related experiences long time ago, gained or lost their trust with a slower pace.

5. Subjects with a high-level of trust in SDC technology lost their trust with a slower pace and regained their trust with a faster pace whereas, subjects with a low-level of trust in this technology, lost their trust with a faster pace and regained their trust with a slower pace.

1.4 BASIC DEFINITIONS

1.4.1 Autonomous Systems

The definition of an autonomous system continuously changes [23]. The Merriam-Webster dictionary currently defines *autonomy* as “the quality or state of being self-governing; especially, the right of self-government.” The concept of autonomy has existed for thousands of years in many different areas including philosophy, sociology, and politics. It is worth mentioning that the second part of the autonomous term, i.e. *nomos*, means *law* in Greek. Therefore, an autonomous system is an independent entity that creates its own laws [24]. A more specific definition would be a machine that is capable of performing tasks by itself and without explicit human control [25].

Although automation was introduced to human civilization many years ago and has been widely utilized since the industrial revolution [26], autonomous systems and industrial robots are relatively new technologies that were introduced in recent decades. In this thesis, our primary focus will be trust in SDCs.

1.4.2 Levels of Autonomous Driving

According to SAE Internationals standard J3016 [27], the levels of autonomy in SDCs are broken down into six categories:

1. **Level 0** - No Automation: The human user manually controls the system 100% of the time. Most of the manufactured vehicles today are in this category.

2. **Level 1** - Driver Assistance: In certain situations such as changing lanes and adjusting the vehicles speed, the system can assist the driver. The driver still drives the car manually at all times.

3. **Level 2** - Partial Automation: The car can autonomously engage the breaks, steer, or accelerate in certain scenarios, but human involvement in tasks such as detecting hazards or traffic signals is still be required

4. **Level 3** - Conditional Automation: The car can manage most of the driving with minimal human intervention. Human attention would still be required if a system failure occurs.

5. **Level 4** - High Automation: Most of the driving will be done autonomously. However, depending on certain geographic areas, minimal human intervention is sometimes required.

6. **Level 5** - Full Automation: All driving autonomous at this level. The only human intervention required would be entering the destination into the system.

Level	Autonomy
0	No Automation
1	Driver Assistance
2	Partial Automation
3	Conditional Automation
4	High Automation
5	Full Automation

Table 1.1: Levels of Autonomous Driving

1.4.3 Trust and its Measurement

Trust is a term that has many different definitions in various contexts such as psychology, sociology, economics, and computer science. Currently, there is no uniform definition of trust [28]. Prior research indicates that there are over 300 definitions in various research areas, and in the context of human-autonomy interactions, there are too many definitions and notions of trust. These notions include the measurement of trust [23], computational models of trust [29], and human-inspired models of trust [30], to name a few. The formal definition of computational trust between two entities can be defined as follows [31]:

Definition 1 *Let* $T_i^j(t)$ *be the* trust value *assigned by a group of players* P_j *to a specific player* P_i *in period* t*. Let* $T_i(t) : \mathbb{N} \mapsto \mathbb{R}$ *be the trust function that illustrates how trustworthy*

P_i is.

$$T_i(t) = \frac{1}{n-1}\sum_{j\neq i}^{n} T_i^j(t),$$

where $-1 \leq T_i(t) \leq +1$ defines the upper and lower bounds of the trust value, and $T_i(0) = 0$ determines the initial trust value, which is assigned when an interaction starts.

Indeed, this formal definition illustrates that trust can be quantified and measured. Accordingly, actions or behaviors can be modified. The goal in technological systems is to have a quantifiable model of trust so the system can be responsive to human trust.

The Oxford dictionary defines the notion of trust as "a firm belief in the reliability, truth, or ability of someone or something [32]." If we consider this simple definition in our context, the definition of trust would be: a strong human belief in the reliability, truth, or ability of an autonomous system. Some examples of trust in autonomous systems include trust in robots, machines, SDCs, autonomous airplanes, and software agents.

CHAPTER 2

LITERATURE REVIEW TRUST IN AUTONOMOUS SYSTEMS

In this chapter, we review the related studies in four categories in order of publication date. The classification includes trust in SDCs, trust in human-robot or human-machine interaction, and trust in autopilot systems. Currently, the level of trust between humans and autonomous systems is known to be very low. This results in humans not allowing autonomous systems to be completely in charge, especially in dangerous situations.

2.1 TRUST IN SDCS

Uggirala et al. [33] analyzed trust in a situation in which users were given information about the capacity of the autonomous vehicle. This study aimed to decrease uncertainty and optimize system performance by making the users knowledgeable about the functions that the autonomous car can perform. The participants completed training to become familiar with the system. Subsequently, they had to judge whether the vehicle would be capable of efficiently completing certain functions given reference lines. This study concluded that when users are knowledgeable about the system, their trust in it increases.

The authors in [34] conducted a study related to the ability of SDCs to function in snow conditions. In this study, 59 drivers were chosen to sit in an autonomous simulator cockpit. One group of drivers were given information about the risks and uncertainties of the SDC when driving in heavy snow conditions, whereas a second group of drivers were not told anything about the SDCs capabilities. This experiment indicated that the first group of drivers who were knowledgeable about the risks and uncertainties did not trust the SDC, and they preferred to override the system so they could drive the car manually. However, the second group did not override the autonomous system to drive the car manually and had more trust in the system.

In [35], Howard focused on factors affecting trust in SDCs. The author examined average consumers attitudes towards SDCs. This research discovered that most consumers have positive feelings toward the ease of use that comes with SDCs. In the context of fully autonomous vehicles, users would not have to feel frustrated when driving in heavy traffic or finding parking in busy areas. One can imagine that at some time in the future, commuters will be able to take naps or watch movies while the car drives them to wherever they desire. The author also discovered that most individuals have concerns regarding the cost, liability, and potential loss of control in SDCs. Income and gender are other variables that affect consumers attitudes towards SDCs. For example, subjects with higher levels of income were more concerned about liability, but subjects with lower levels of income were more concerned about the loss of control.

Carlson et al. [17] conducted a statistical analysis in the domain of autonomous vehicles and autonomous diagnostic systems. They created an online survey and asked human subjects about various scenarios related to SDCs and the use of IBM Watson in critical medical situations (e.g., to determine types of cancer). It was discovered that most test subjects had concerns regarding the past performance of the car, reliability, errors, software/hardware failures, and the liability of the car manufacturer. Similarly, it was discovered that top factors that affect trust during the use of IBM Watson in critical medical situations are accuracy and past performance. The result of this study indicated that, regardless of the domain, most people tend to prioritize safety, efficiency, and failure rates when deciding to trust an autonomous system.

Kyriakidis et al. [36] created an international questionnaire related to the general publics opinion of automated driving. The questions were related to concerns, acceptance, and willingness to purchase an SDC. Most of the 5000 participants from 109 countries agreed that fully automated cars have the potential to be very popular among consumers by 2050. However, most were also concerned about safety, malicious activities/hacking, and legal issues related to autonomous vehicles. The authors also found that most of the educated subjects had more income and lived in developed countries. This subset of participants was uncomfortable with the idea of SDCs transmitting data to external sources. They were concerned about malicious use of the transmitted data.

The authors in [37] explored the possibilities of developing trust in SDCs using techniques that are currently available. This study argued that SDCs aim to make our lives easier and reduce the number of accidents. However, in many situations such as unpredictable hazards and intense weather situations, a human driver's reactions are superior and needed. Testing was discovered to be a critical aspect of determining if the car is trustworthy enough to be on the road or has the potential to develop trustworthiness overtime. SDCs utilize machine learning and image processing techniques to provide functions like detecting pedestrians or stop signs. It was argued that many people require very high accuracy in the functions of their SDCs (close to 100%), but machine learning algorithms cannot produce such accurate results.

Butakov and Ioannou [38] suggested that users levels of comfort and trust will increase if the design and dynamics of cars autopilot systems are closer to those in regular vehicles. In this study, the authors presented a methodology that allows the custom modification of autopilot modes such as adaptive cruise control and automatic lane change systems based on individual preferences.

Payre et al. [39] conducted an experiment and analyzed how alternating levels of trust would affect a driver's reaction time. The main objective of this experiment was to measure *manual control recovery* (MCR) times when emergency situations arise in fully autonomous cars. These cars are certified and eligible to be used by drivers with a standard driver license. Nonetheless, the drivers are still accountable for their vehicles, including regaining MCR in an emergency and remaining in the driver's seat with their seat-belt buckled at all times. This study demonstrated that higher trust levels in fully autonomous cars resulted in slower reaction times, which can create a hazardous environment for drivers. This experiment can help companies become more aware of problems that over-trust can cause.

Akash et al. [21] presented a gray-box modeling approach that has the capability of capturing different variations of human behavior related to human-machine trust. In an experiment involving 581 human subjects, the authors utilized a computer simulation platform by which human subjects were shown an obstacle detection system (as used in SDCs) and were asked if they trust or distrust the image processing algorithm used in the system. The study discovered that human trust significantly decreases in faulty scenarios. An observation of this study was that the amount of trust in the system slightly increases after around 8 to 10 trials when a negative experience has already lowered their trust. Additionally, this study investigated the effects of national origins, culture, and gender on the level of trust. The results indicated that Americans usually have less trust in autonomous vehicles compared to people from other nationalities such as Mexicans and Indians. This finding matches previous discoveries in this domain. The overall conclusion of this study suggests that a perfect autonomous system should be able to collect data (e.g., psychological factors and demographic information) from its users and use this data to maintain and improve trust.

Finally, Daziano et al. [40] investigated consumers willingness to purchase SDCs by conducting an online experiment. After analyzing data collected from over 1260 human subjects around the world, it was estimated that the average household tends to pay about \$$3,500$ more for partial automation and \$$4,900$ for full automation when purchasing a new vehicle. This research also found that consumers preferences regarding different levels of automation (i.e. low, semi, full) are highly variable. A significant portion of participants even preferred to pay more than \$$10,000$ extra for fully automated vehicles. Based on information from this study, we can detect a pattern in consumers' behavior which suggests that public interest in SDCs is increasing rapidly. We believe that this public interest will spike in the near future as issues such as trust between humans and SDCs as well as the reliability of autonomous technologies are resolved. A summary of this section's results is shown in Table 2.1.

Reference	Summary	Approach	Concentration
[33]	When users are knowledgeable about an autonomous system, their trust in the system increases	Simulation	Training & education
[34]	Drivers who are knowledgeable about risks, don't trust SDCs in snow conditions	Simulation	Risks and uncertainties
[35]	Consumers have positive feelings toward the ease of use that comes with SDCs	Survey	Fender and income & trust
[17]	Errors, software and hardware failures will affect trust	Survey	Safety, efficiency, Malfunction & trust
[36]	Users are concerned about safety, hacking, and legal issues	Survey	The future of SDCs
[37]	Unpredictable hazards are still an issue that needs to be resolved	N/A	Safety
[38]	Level of trust increases if the design and dynamics of SDC are closer to what they are in regular vehicles	Simulation	Autopilot modes
[39]	Over-trust is an issue and can potentially cause hazardous situations	Simulation	Training & education
[21]	A model to capture the dynamic variations of human trust	Survey	Dynamic trust
[40]	Consumers are willing to pay significantly more for autonomous features	Survey	Consumer behavior

Table 2.1: Trust in SDCs

2.2 TRUST IN HUMAN-MACHINE INTERACTION

Muir [41] provided an analysis of trust in human-machine interaction from a psychological perspective. The author illustrated how trust evolves when *decision support systems* (DSS) are used. DSS are computer programs that assist an individual or an organization when making decisions. These decisions include ranking important documents, buying and selling stocks quickly, choosing a target market, and many other important decisions [42]. Since DSS have a significant impact on some critical decisions, the user's trust in such systems becomes a crucial factor when designing decision support systems. The author initially analyzed the psychological trust models among humans and then developed a human-machine trust model. In this model, the concept of *trust calibration* is introduced in which the user has the responsibility of calibrating their trust based on the reliability of the decision support system. The author suggested that there are certain factors and design goals that should be considered to better calibrate trust when designing decision support systems. These factors are as follows:

1. *Aiming to improve the user's perception of the trustworthiness of decision support systems*: This would require the user to understand how a DSS works and become familiar with the predictability of the system's decisions. The author recommended that this can be achieved by putting each user on a trial period of using a DSS to improve the user's perception. A candidate solution would be the use of a simulation environment to allow the user to freely explore the DSS without any fear or concern related to wrong or dangerous decisions.

2. *Modifying the decision support systems criterion of trustworthiness*: To achieve this goal, the DSS has to demonstrate a history of efficiency and good performance. It was recommended that the users have access to statistical data such as information related to the systems performance.

3. *Continuous identification and fixing the causes of poor trust calibration*: To improve the level of trust between humans and machines, the system (or developers) should detect low levels of trust and fix the problem. This study indicated that some of the main causes of low levels of trust might be due to users incorrect expectations. Thus, users calibration training for this issue is crucial.

Another fascinating study, which examined the role of trust in decision support systems and autonomous aids, was conducted by Madhaven and Wiegmann [43]. Since the role of DSS and autonomous aids in making critical decisions has increased significantly over time, this paper proposed a framework by which human trust in autonomous aids can be increased over time. The proposed trust framework utilizes psychological traits that affect trust among humans. It uses these traits to provide a set of instructions for the DSS so that human trust is increased. The outcome of this research contributed to the identification of several important psychological factors such as favoritism (human vs. robot partners) and subjective bias in users, which affect the human-robot trust relationship. These variables are critical for the development of decision support systems as well as autonomous systems that interact with humans.

Factory automation is another line of research in this domain. Lee and Moray [44] executed an experiment to characterize variation in an operator's trust during an interaction with a semi-automatic pasteurization plant. The authors investigated the relationship between changes in the operators' control strategies and trust. In the same line of research and based on [45], Muir [46] conducted two experiments to test influential variables in human-machine trust and provided an experimental analysis of the theoretical trust model proposed a few years earlier. The results of these experiments indicated that the perceived competence of an autonomous system relates to the amount of trust a user may have in the system. For example, if a user detected that the system might be incapable of doing its job, i.e. incompetence, they would manually take control of the system, and as a result, their trust would drastically decrease. Another finding of this study indicated that the amount of monitoring the autonomous system requires will decrease if the level of trust in the system increases. The author suggested that the findings in this research could be used by industry professionals to determine which properties of autonomous systems could have vulnerabilities that might display incompetence and lower human trust. By doing so and predicting the patterns of human trust, they would be able to increase the overall effectiveness of the autonomous system.

Dassonville et al. [47] investigated the issue of trust within the context of a teleoperation system, which is a type of system in which human operators control a machine or system from a distance. The authors conducted experiments on the role of self-confidence in human-machine interactions. A teleoperation system is composed of three components as follows:

1. *Master universe*: The master universe is the environment in which the operator re-

sides. An example would be a military drone operator sitting in a container in the middle of a desert and controlling a strike drone somewhere far away in a combat zone.

2. *Slave universe*: Similarly, the slave universe is the environment in which the machinery or the system operates through the operator's commands. This universe is composed of hardware and a group of sensors.

3. *Space between the master universe and the slave universe*: The space between the master universe and the slave universe contains data transmission (e.g., the internet), fast computers, and decision control systems.

In this study, a simulated experiment was performed by having an operator to use a joystick (master universe) that was connected to a computer (space in between) to control a cursor on the screen (slave universe). This experiment was conducted on two student populations of literature and scientific studies. The study discovered that the first population appeared more self-confident in operating the machine; however, both groups had similar levels of trust in the system.

In [48], Moray and Idnagaekri analyzed how trust in autonomous systems leads to a lower level of human supervision. They discovered that as subjects become more reliant on the system, a decrease in constant monitoring occurs. This study also scrutinized a situation in which participants overly trusted a system, which could lead to malpractice in certain situations.

The authors in [49] conducted a study based on the effects of continuous and discrete malfunctions within autonomous systems. The study had two parts. The first section tested human participants based on continuous and discrete faults separately, whereas the second part intertwined two malfunction types. These experiments found a significant decrease in trust after five continuous failures. However, there was no significant reduction in trust after one discrete malfunction. The results of this study demonstrated how trust was dissipated and how users relied on autonomous systems based on previous faults.

Dzindolet et al. [50] performed an experiment to improve trust in autonomous systems. This study involved participants detecting a soldier camouflaged in an area. They had the option of manually guessing whether the soldier was there or obtaining assistance from an automated aid. The first study measured human trust in the system before any interaction with the system. The results indicated that the operator would trust the system that had higher approval ratings and fewer errors. The second study compared the number of mistakes the user made to the number of errors produced by the system. To accomplish this, two separate groups were selected. One group had a system that made twice as many errors as the user, and the other system made half as many errors as the user. The result demonstrated that those who had more errors were more inclined to stick to their decision.

Finally, Merritt [51] examined the importance of considering differences in human behaviors in the context of human-automation interaction. The author conducted an empirical study by providing an experiment related to X-ray screening. Subjects were asked to use simulation software to detect dangerous items such as weapons in luggage. They were given the options of scanning the x-ray image manually and flagging it if they spot anything suspicious or having a fictional autonomous system examine the image and report any issues. This study found that individual differences among subjects affect the value of trust in autonomous systems, even if the characteristics of the autonomous system are constant. This study suggests that future researchers should consider human characteristics when designing experiments for trust analysis and measurements. The summary of this section's results is shown in Table 2.2.

2.3 TRUST IN HUMAN-ROBOT INTERACTION

Murphy et al. [52] investigated the use of autonomous rescue robots in combat situations as well as cases where victims were unable to be reached, for example, victims trapped in earthquake rubble. In this study, the purpose of the rescue mission was to find the victims,

Reference	Summary	Approach	Concentration
[41]	User should calibrate their trust based on the reliability of DSS	Analytical	DSSs
[43]	Introduced framework for trust improvemnt	Framework	DSSs
[44]	Changes in operators' trust	Simulation	Control strategies
[46]	Manual override of the system by distrustful users	Simulation	Incompetence of autonomous systems
[47]	Role of self-confidence and trust in teleoperation systems	Simulation	Trust and teleoperation systems
[48]	Theoretical trust models were explored	Meta-Analysis	Human-machine interaction
[49]	Demonstrated the changes in trust based on malfunctions	Simulation	Malfunctions and their impact on trust
[50]	Trust is a critical factor in automation reliance decisions	Survey	Automation reliance
[51]	Different people have different trust levels in autonomous system	Simulation	User perceptions of trust

Table 2.2: Trust in Human-Machine Interaction

check for vital signs, and help the victims until they were rescued. The study discovered that the success of these robots simply depended on the victims' trust. Specifically, it was crucial that the victims allowed the robot to help and collect data for optimal recovery and assistance, which could be achieved if a high level of trust was established.

Human trust also depends on the failure rate of the autonomous systems. For instance, a study of commercially available ruggedized robots operating under field conditions showed a *mean-time-between-failures* (MTBF) of 12.26 hours and an availability rate of 37% [53]. This finding indicates that if the robotic systems reduce their failure rates, their reliability will increase, and subsequently, users confidence in their performance will increase. It is apparent that, due to recent advancements in technology, the mean-time-between-failures has been decreased in autonomous systems even outside of robotics.

Parasuraman and Miller [54] investigated the concept of trust and etiquette in the domain of *human-robot interaction* (HRI). Given that respect and etiquette significantly affect the level of trust in many human-to-human social interaction scenarios, the authors argued that these factors also impact humans perception of autonomous robots. In this study, etiquette is described as a set of prescribed and proscribed behaviors that permits meaning and intent to be ascribed to actions. This study also conducted an experiment related to the role of etiquette in HRI. Human subjects used flight simulator software called *Multi-Attribute Task* (MAT) and communicated with the autonomous system using different communication styles such as interrupting the user and being impatient. The empirical evidence obtained in this experiment showed that both etiquette and reliability affect humans trust of autonomous robots.

Stormont [55] showed a low level of trust in HRI. The author investigated the factors that affect trust between humans and robotic systems. He discovered that one reason for a low level of confidence in autonomous systems is their low level of reliability. The author also discovered that unpredictability is another factor affecting trust between humans and autonomous systems. He argued that, in various hazardous circumstances such as battlefields (as shown in Figure 2.1) and rescue missions, the unpredictability of robots becomes a significant problem for human supervisors. Although the autonomous nature of robots and their capacity for quick decision-making are considered positive traits, a problem arises when the life or death of humans depends on the choices of a robot. Indeed, questions such as "Should life or death decisions be made by an autonomous system?" have been the focus of many researchers. The same study executed a simulation of robots assisting firefighters in a hazardous fire situation. The simulation showed that although the firefighters did not initially trust these robots, their reliance on and trust in the robots increased as the mission progressed and the firefighters became tired. As a result, they finally allowed the robots to extinguish the fire.

The authors in [56] investigated how culture and appearance might have an impact on trust. The team sampled participants from China, Germany, and Korea to analyze different cultural backgrounds. The participants were asked to interact with a robot that knew the culture of each participants country. The results were scaled based on likeability, engagement, trust, and satisfaction. The outcomes demonstrated that each participant perceived

Figure 2.1: Autonomous and Semi-Autonomous Aobots Used in Battlefields [2].

the robot differently, thus showing the need for increased concentration in different areas. Producers may use this information to create robots which are more unique to specific regions and cultures to improve trust and demand within particular communities.

Hancock et al. [3] provided a comprehensive analysis of factors affecting trust in human-robot interaction. This study classified factors affecting trust in HRI into three different categories (i.e. human, robot, and environment), as shown in Figure 2.2. Human-related factors include training, expertise, situational awareness, and demographic information. Similarly, robot-related factors are behavior, dependability, reliability, level of automation, failure rates, false alarms, transparency, and attribute-based factors such as location, personality, adaptability, robot type, and anthropomorphism (having human traits). Finally, environmental factors include teamwork, culture, communication, shared mental models, task type, task complexity, and multi-tasking. This paper discovered that robot performance has the highest impact on human trust.

Yagoda and Gillan [57] proposed a new mechanism for measuring the value of trust in the context of HRI. This measurement was based on multiple factors such as team configuration, team processes, the context, the task, and the system. The proposed trust measuring mechanism was developed using two experiments. The results of these two studies were combined to create a new HRI trust measuring tool.

Penders et al. [58] investigated HRI in "no-visibility" conditions, which means the

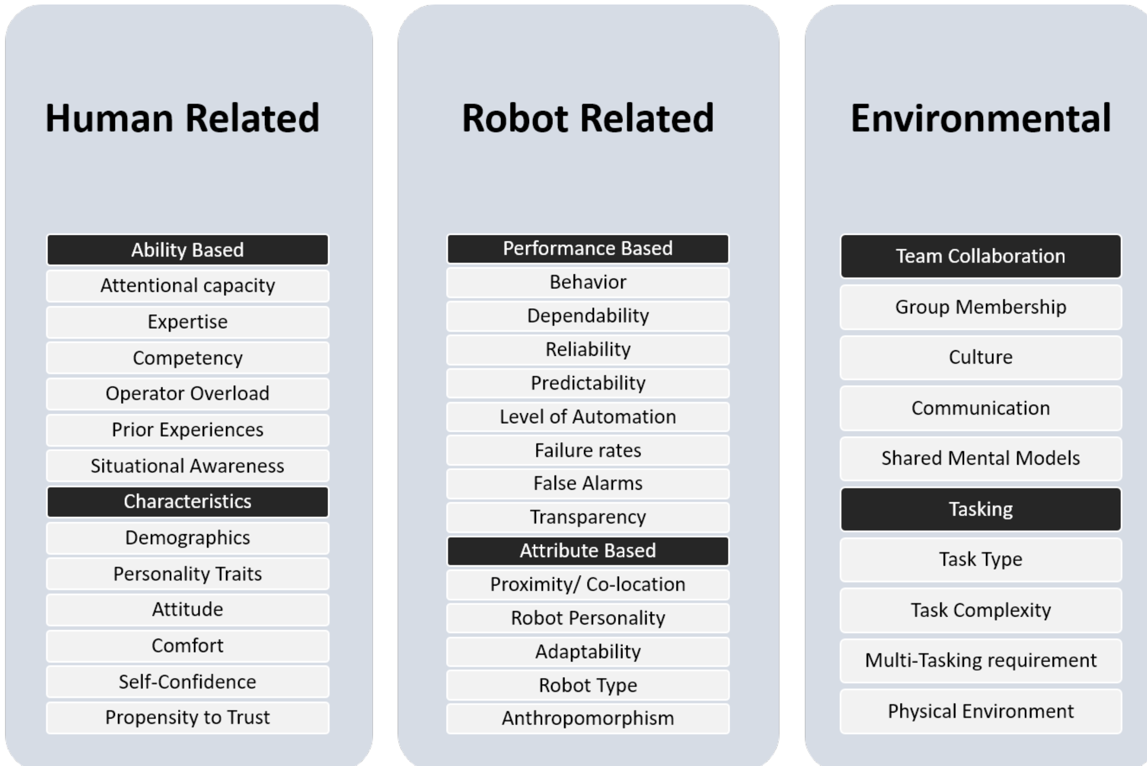

Figure 2.2: Trust Factors Identified by Hancock [3].

human subjects might be visually impaired or blind. Therefore, they would have to trust the robot completely. This study analyzed interactions between visually impaired people and their guide-dogs and examined the variables that could be utilized in the design and behavior of robots to improve human trust. These variables include human dominance, cooperation over time, and accountability. It is worth mentioning that Castelfranchi and Falcone [59] investigated how control negatively affects trust. Their research discovered that human trust will decrease if a participant is forced to take control of an autonomous system. We do believe that this is a prominent issue that should be considered in no-visibility conditions.

Wang et al. [60] investigated human-robot trust in the context of underwater semi-autonomous robots. To increase the performance of a submarine robot, the operator's trust in the submarine robots capabilities must be established and sustained. This study proposed a trust model that mainly deals with recording the robot's past performance, the human operators performance, and the fault rates of humans and robots. A semi-autonomous robot known as YSI EcoMapper AUV was tested in this study. The authors show the effectiveness of this trust model through a simulation-based approach.

An essential aspect of improving human trust with respect to robotic systems is the real-time measurement of human subjects emotional responses. If artificial intelligence technologies can learn and interpret the emotional states of humans, they will be able to adjust themselves accordingly to be responsive to these emotional states. Hu et al. [22] provided a trust sensor model that utilizes psychophysiological measurements of human subjects. The objective of this project was to determine whether humans psychological factors captured through sensors such as *Electroencephalography* (EEG) and *Galvanic Skin Response* (GSR) can be used to manage trust in the context of human-robot interaction. After a series of experimental studies, statistical analyses, and classification, this study concluded that psychophysiological measurements could be used to measure trust in human participants. However, the mean accuracy of this claim was only 71.57%, so this method cannot be used for all humans. The authors believe that human subjects' demographic information should be considered in any experimental study.

Finally, due to technological advancements in recent years, care-giving robots are becoming popular and common in our lives. As a result, trust in these robotic systems is a prominent issue from the consumers perspective. In a recent study [61, 62], the authors measured levels of trust, satisfaction, and frustration in the context of care-giving robots by designing several experiments in which human subjects interacted with a Baxter robot through a sequence of trust-building or trust-damaging incidents. In this study, various scenarios were tested, for example, delivering an object with different speeds or accidentally dropping the object. The authors discovered that the assistant robots performance affected the human participants level of trust, satisfaction, and frustration when a sequence of structured incidents is considered. The summary of this section's results is shown in Table 2.3.

Reference	Summary	Approach	Concentration
[52]	Success of the rescue robots depends on their reputation	Survey	Rescue robots
[53]	Reduction in failure rates leads to increase in trust	Simulation	Improving trust by reducing failure rates
[54]	Etiquette affects human trust as well as the reliability of autonomous robots	Simulation	Etiquette & trust
[55]	Unpredictability of robots affects trust	Simulation	Military and hazardous environments
[56]	The appearance of a robot affects human perception and trust	Robot Experiment	Trust and cultural differences
[3]	The performance of robots has the highest impact on trust in the context of human-robot interaction	Survey	Human-robot interaction
[57]	Created a trust measuring tool for human-robot interaction	Survey	Trust measurement
[58]	To enhance trust in human-robot interaction, a number of design choices need to be made	Live Experiment	Improving trust in human-robot interaction
[59]	Control on autonomous systems increases or decreases trust depending on the circumstances	Cognitive analysis	Trust and artificial intelligence
[60]	If humans trust a robot, its performance will increase	Robot experiment	Trust and semi-autonomous robots
[22]	Psychophysiological measurements can be used to measure human trust	Survey	Psychophysiological measurements & trust
[61][62]	Performance of a robot directly affects trust, satisfaction and frustration	Robot Experiment	Trust, satisfaction, and frustration

Table 2.3: Trust in Human-Robot Interaction

2.4 TRUST IN AUTOPILOT SYSTEMS

de Vries et al. [63] showed how planned routes in manual and automatic modes affected trust. This experiment involved instructing a group of participants to plan a route and then choose to complete it manually and automatically ten times each. The study proved that automatic failures had more negative impacts on trust compared to manual failures. Participants were more likely to forgive themselves for the error they had committed than the failures that happened during the automatic mode. The results demonstrated a bias towards participants trust in the manual mode as opposed to the automatic mode.

In [64], the authors conducted research on human trust regarding air traffic management systems. They provided guidelines and strategies for improving trust in autopilot systems over time. They argued that air traffic control operators currently utilize many automated and semi-automated computer tools, and more usage of fully automated systems is anticipated in the near future. Thus, the operators will have to trust components of autonomous (or fully autonomous) air traffic management systems, for example, radar systems and communication tools. The procedure to address the trust issue can be formed through multiple development phases as follows:

1. Assign experienced air traffic controllers to develop a system.

2. Provide high-quality simulations.

3. Provide training for the controllers.

4. Schedule a transition period for the controllers.

5. Keep the old technology in case failures occur.

Jiang et al. [65] discovered that there is a direct correlation between specific types of errors that occur and operators' trust in the autonomous system. In this study, the participants were monitored for one week. The first day was based solely on training them to recognize errors and functions of the system. During the remainder of the trial, the team examined how participants felt about false alarms given by high-risk systems. The results demonstrated that there was a significantly greater decrease in trust towards systems that consistently output false alarms.

In an experimental study [66], the authors investigated the methods by which human subjects were able to judge the performance of complex autonomous systems. To ensure accuracy in the investigation, the participants completed training in measuring the accuracy and performance of airplanes. The participants were then asked to analyze an airplane's performance and rate it as friendly or hostile based on the measured speed, altitude, range, and time in the air. The post-training results were relatively accurate, and demonstrated that the judgments became more accurate when participants learned what they should be looking for in complex autonomous systems.

Finally, Winter et al. [67] investigated humans distrust of autonomous airplanes. In this study, participants were asked if they prefer to be on a commercial airplane with two pilots (i.e. a pilot and a co-pilot), an airplane with a pilot in the cockpit and a co-pilot working remotely, or an airplane with both pilots controlling the aircraft remotely. The authors found that the human subjects would have a high degree of discomfort if they were on a fully autonomous commercial airplane with both pilots only overseeing the movements and controlling the airplane remotely. They also mentioned that the subjects would have a high degree of distrust when only one pilot was in the cockpit. This study also discovered that humans trust in autonomous airplanes is related to their culture. For example, the test subjects from India felt more comfortable if they were on a fully autonomous aircraft than subjects from the United States. A summary of this section's results is shown in Table 2.4.

Reference	Summary	Approach	Concentration
[63]	Failures caused by system have more negative impacts on trust	Simulation	Trust in automatic and manual failures
[64]	Provides a procedure for improving trust in air traffic management systems	Survey	Trust in air traffic management systems
[65]	Greater decrease in trust when systems continuously outputted false alarms	Empirical Experiments	False alarms & trust
[66]	Participants analyzed airplanes and rated them as friendly or hostile	Simulation	trust management in aircraft autopilot features
[67]	Culture directly affects trust	survey	Trust in autopilot systems

Table 2.4: Trust in Autopilot Systems

2.5 CONCLUSION

In this chapter, we reviewed the existing literature on trust in autonomous systems. We investigated technical papers that examined trust between humans and SDCs, robots, machines, and autopilot systems. The reviewed studies provide new discoveries as well as recommendations to manage and improve trust between humans and autonomous systems. The literature on trust, however, is still very broad and does not address particular issues that are currently present such as how all the theoretical trust frameworks can be transformed into models that are understandable to machines. Further research discoveries and innovative methodologies can potentially provide a reliable platform to develop trustworthy SDCs that perform well.

CHAPTER 3
SIMULATION SETUP AND CHALLENGES

In this chapter, we will go over the design of our simulation setup and its technical configurations in detail. Section 3.1 covers simulation setup and the equipments used to develop this VR simulator. Section 3.2 discusses the realistic nature of our simulator, our video capturing mechanism, and the use of real VR driving videos. Section 3.3 goes over the programming languages and the technology used to develop our trust self reporting tool. In Section 3.4 we will go over some of the challenges we faced during the development of this simulator, and how we managed to resolve them, and we will go over the high level architecture of the VR SDC simulator and the development phases we executed to implement each component of the SDC VR simulator.

3.1 SIMULATION SETUP

Our simulator is a safe platform to expose human subjects to any inconvenient trust-damaging incidents, including but not limited to sharp turns, sudden stops, stop-light violations, speeding, tailgating, unexpected accidents, and others. The SDC VR simulator is based on the fusion of an Oculus Rift VR headset with an Atomic A3 Full Motion Simulator. Figure 3.1 shows a participant sitting in the simulator.

3.1.1 Atomic A3 motion simulator

Atomic A3 motion simulator was utilized to create this simulation. This motion simulator is designed by Atomic Motion Systems, a UK based company, and manufactured and distributed in the US by Talon Simulations [4]. The Atomic A3 Full Motion Simulator is

Figure 3.1: Participant Using the SDC Simulator.

capable of moving up to 71 degrees per second across a full 27 degree dual-axis movement range [68]. The combination of complete VR visual effects, 3D audio, and movements provides a convincingly realistic simulation. Hence, it is used for realistic simulations such as NASA's Hybrid Reality Work [69] [70]. Figure 3.2 displays the Atomic A3 motion simulator with one of the participants experiencing once of the trust segments.

For motion, The Atomic A3 Simulator receives telemetry data via UDP packets. Each UDP packet contains pitch are roll values in degrees. Pitch values are used for front and back movements. Similarly, roll values for left and right movements (e.g., negative value for pitch indicates backwards motion).

3.1.2 Oculus Rift VR Headset

Oculus Rift VR headset was used in combination with the Atomic A3 motion simulator to create the SDC Motion Simulator. Developed and manufactured by Oculus VR, which is a sub-division of Facebook Inc. [71], the Oculus Rift VR headset outputs 1080x1200 resolution graphics per eye at a 90 Hz refresh rate and a 110-degree field of view; it also

Figure 3.2: Atomic A3 Motion Simulator [4] and Alienware Area 51

has headphones which output 3D audio effects. Oculus Rift is mainly used for gaming and entertainment, but recently, it has been gaining momentum in many different applications such as education, media, virtual casinos, etc. [72]. See Figure 3.3 for a picture of Oculus Rift.

Figure 3.3: Oculus Rift VR Headset & Oculus LED Sensor [5]

For input, Oculus Rift takes in 6DOF (3-axis rotational tracking and 3-axis positional tracking) input data from an IR LED sensor sensor that tracks the user's head movements [5]. This allows the VR "feel" when the user moves their head around in the simulation. See Figure 3.3 for a picture of Oculus Rif's LED sensor.

3.1.3 Alienware Area 51

To run a VR simulation with realistic graphics, a high end machine with a top notch quality graphics card and processor is needed. In order to address this problem, we used an Alienware Area 51 desktop computer. This $10,000 machine is equipped with an Intel Core i7-6950X (10 cores) processor along with dual NVIDIA GeForce GTX Titan Z (12 GB GDDR5X each) graphics cards. This high end machine with superior computing power allowed us to run our Simulation with no problems. See Figure 3.2 for a picture of Alienware Area 51.

3.1.4 Advanced Noise Canceling earmuffs

To increase the immersion of the SDC simulation, and reduce the exterior noise when the simulation is running (e.g, the noises coming from the mechanical movements of the simulator), a pair of ClearArmor 141001 Shooters Hearing Protection Safety Ear Muffs were used to cancel any kind of a distracting noise. These earmuffs have a noise cancellation rating of 31 decibels. See Figure 3.4 for an illustration of ClearArmor earmuffs.

Figure 3.4: ClearArmor Earmuffs [6]

3.2 REALISTIC VR ENVIRONMENT

To provide realistic simulation with high degree of immersion, 360-degree videos of driving various driving styles were recorded using a GoPro Fusion Camera and edited using GoPro Fusion Studio software. The videos were exported from Fusion Studio at 4k resolution as MP4s along with 360-degree MP3 audio files [8].

3.2.1 Gopro Fusion

Gopro Fusion is one of the latest cameras that allows 360 video capture with a very high quality (4k and 5k). Released in November 2017, Gopro Fusion features two lenses; one in the front, and one in the back of the camera and records two videos simultaneously, which are later stitched together using Gopro's video stiching software, Gopro Fusion Studio. After using 360fly 4k, and experiencing low VR quality (see 3.4), we decided to purchase a Gopro Fusion, which drastically improved the video quality. Figure 3.5 illustraces the Gopro Fusion camera.

Figure 3.5: 360Fly 4K Action Sports Camera [7] and Gopro Fusion [8]

3.2.2 Capturing Trust Damaging Footage

One of the team members volunteered to mount a Gopro Fusion camera in his car using the Gopro's Goosneck mount (See Figure 3.8). To capture interesting driving footage and trust damaging scenarios, the team member recorded 360 degree videos of his commute to and from work everyday for the duration of a month. These videos were later analyzed, edited, and categorized to be used in the SDC simulation.

3.3 SOFTWARE

A variety of software and programming languages were utilized for different to develop different components of the SDC simulator such as scripting, automation, video editing.

3.3.1 Unity Game Engine & C# Scripting

The SDC simulator application was built in the Unity Game Engine using C#. Unity is one of the most powerful game engines in the video game industry. It is also used for production of VR simulations, since it is compatible with the Oculus Rift headset. The compiled C# executable is responsible for synchronously playing the video, audio and sending the telemetry UDP packets to the simulator. To maintain a clean and maintainable architecture, Our experienced software development team coded the simulation based on the best practices of Object Oriented Design patterns and SOLID principles.

Furthermore, the SDC simulator software is responsible for creating structured directories with log files and spreadsheets for the data collected from the test subjects after every test run.

3.3.2 Simtools

SimTools motion simulator software was used to send the telemetry data for each video to Atomic A3 motion simulator via UDP packets. By using a plug-in called "Video Ride Creator", we were able to create telemetry points for every frame in the simulation videos. Another plug-in, "Video Ride Player" was used to rapidly send the telemetry points as UDP packets to Atomic A3's motion simulator Software, Symphinity, which recieved the UDP packets and moved the chair [73].

3.3.3 Powershell and Batch

PowerShell was used to automate multiple tasks in the SDC motion simulator. We wrote multiple PowerShell scripts combined with Windows batch scripts to synchronize the activities of the Simtools software and the Unity executable, so that when the videos are being played in VR, the UDP packets would get sent to the chair automatically with no human intervention.

3.3.4 Stare-and-Select Trust Reporting Tool

We developed a software component called "Stare-and-Select" (SAS). SAS was used as our main data collection tool in this phase of our experiment. After each simulated autonomous driving segment was completed, the participants were presented with a Likert scale that appeared inside the VR environment. The participants were able to select their response by just focusing their gaze on the desired answer for five seconds while wearing the Oculus Rift. We believe that this data collection approach is novel because it is quick, it does not distract the users, and it is pleasant to work with. Figure 3.6 shows a sample question in the VR environment.

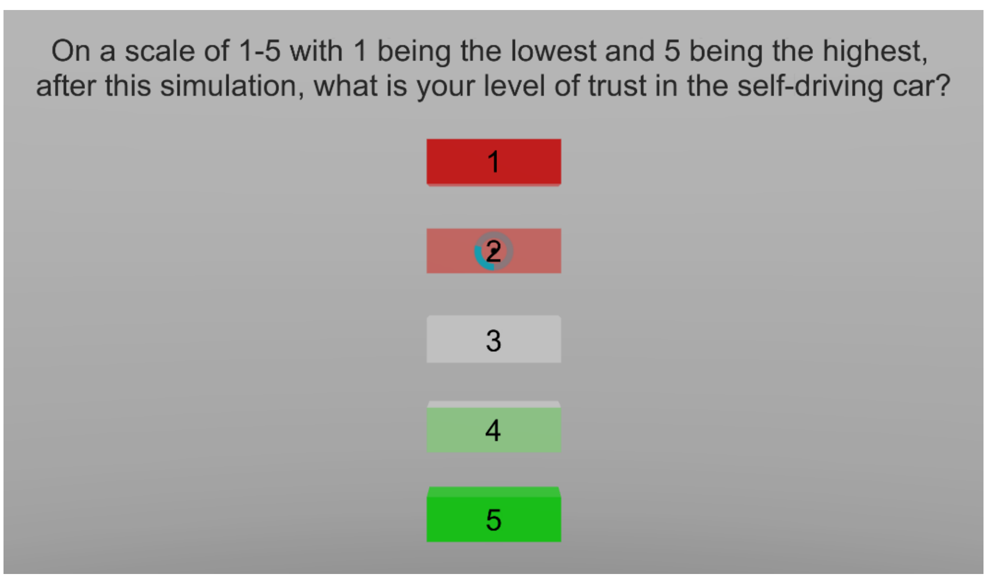

Figure 3.6: SAS Trust Reporting Tool

3.4 CHALLENGES

We encountered several challenges during the development of the SDC VR simulator and SAS. Some of the obstacles we faced include Low VR video quality, Telemetry Meta-Data optimization, video shakiness, and VR sickness.

3.4.1 VR Video Quality

360 videos are known to have lower quality when viewed in VR headsets. This is due to the inability of many 360 degree cameras to capture footage with very high resolutions and low quality video stitching algorithms [74].

Initially, 360Fly 4K action sports camera was used to capture driving footage, but after testing on Oculus Rift, we discovered that the video quality is very poor. To increase the immersive nature of the SDC simulation, we had to wait until November 2017 for Gopro Fusion to be released into the market, which proved to capture higher quality driving footage. Figure 3.5 shows 360Fly 4K Action Sports camera.

3.4.2 Obtaining Telemetry Meta-Data

Obtaining and optimizing the telemetry data for the Atomic A3 motion simulator was the most difficult challenge that we faced when developing the SDC VR Simulator. Initially, we used smart phone apps like MotionLogger [75] and SensorData [76] to capture the pitch and roll telemetry values of the simulated autonomous driving scenarios. However, after deploying and testing on the Atomic A3 motion simulator, we noticed the raw data is insufficient for the simulation because the movements were not completely in sync with the movements in the VR videos. For example, when the car was making a right turn, there was a massive delay or even no motion detected by the sensor. This was due to the fact that the sensor resided inside the car, as well as the fact that the car did not do anything too erratic such as drifting. To Address this problem, we had to manually go through each telemtery file and manually adjust the pitch and roll values to ensure that they are in sync with the car movements. Figure 3.7 is a flowchart that visualizes the high level view of the process and the development cycle of the SDC VR simulator.

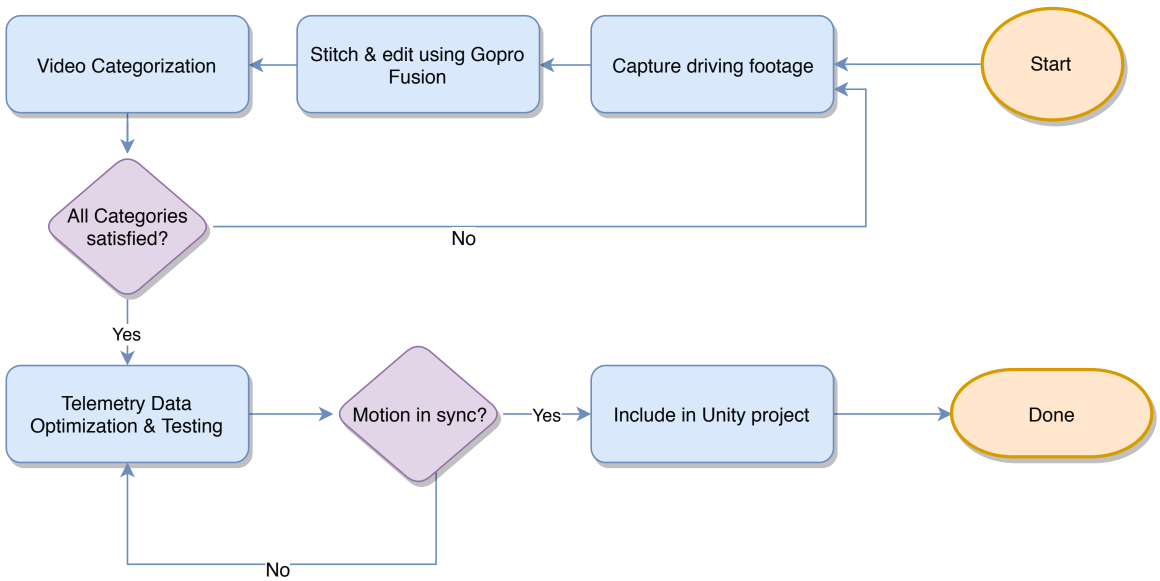

Figure 3.7: SDC VR Simulator Development Phases

3.4.3 Shakiness

Another major challenge when developing the SDC VR Simulator was the shakiness in the videos. In the initial phases of the development, we mounted our camera on a bicycle helmet, and had our driver wear the bicycle helmet when driving and capturing footage. This approach was problematic. Whenever the driver looked around (e.g., to check the blind spots and side view mirrors), it disrupted the VR video due to the shakiness.

To attempt to address this issue, we purchased a “Goose Neck” mount for our Gopro camera, and attempted to attach it to the sun visor of the car upside down. We noticed a slight improvement, but still detected shakiness on bumpy roads and sharp turns. After several trials and errors, we attached the Goose Neck mount to the moon-roof of the car upside down, and positioned the Gopro Fusion in front of the steering wheel. Figure 3.8 shows the position of the Gopro Fusion Camera.

Figure 3.8: Gopro Fusion Mounted onto The Car's Sun-roof

3.4.4 VR Sickness

VR sickness is a known phenomenon in which people experience symptoms that are very similar to motion sickness symptoms when they watch or interact with VR content. Some of the symptoms include headache, general discomfort, nausea, vomiting and vertigo [77]. In the initial development phases of the SDC motion simulator, our early test runs indicated that excessive motion in the simulated autonomous driving videos causes VR sickness. To address this problem, we lowered the motion output of the Atomic A3 motion simulator to 5%. Furthermore, in our IRB application we stated that the eligible test subjects cannot be pregnant, have history of nausea, vertigo, or motion sickness.

CHAPTER 4

REAL-TIME TRUST MEASUREMENT IN HUMAN-SDC INTERACTIONS

In this chapter, we will describe the procedure and outcome of the empirical experiment that we have conducted to investigate the level of trust, distrust, and satisfaction of passengers of SDCs. Section 4.1 illustrates our research methodology. Section 4.2 covers the experimental design. The results of our research and our demographic data analysis are presented in Section 4.3. Finally, in Section 4.4, we end the paper with concluding remarks and the future direction of our research.

4.1 HUMAN-INSPIRED TRUST MODELING: RESEARCH METHODOLOGY

In our methodology, experiments are executed to measure initial trust, trust escalation, and trust reduction in human-autonomy interactions. Our goal is to understand how humans gain or lose trust while interacting with technological systems such as SDCs. We intend to understand humans' perception of trustworthy and untrustworthy actions. Specifically, we performed experiments and examined interactions between human subjects and our test bed under the lens of trust in a controlled setting. We intended to understand (a) how we can establish trust between humans and autonomous vehicles, (b) how we can sustain trust over time, and (c) how we can rebuild trust when incidental failures occur. From the outcomes of our experiments, we derived humans' perception of trustworthy/untrustworthy actions.

4.1.1 Our Sequential and Structured Data Collection Approach

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. Our data collection scenarios must have a specific structure and capture certain information. As shown below, scenarios are categorized into five distinct groups. Note that demographic data and past psychological data are collected prior to our experiments to characterize our human subjects regarding self-confidence, trusting attitude, risk-acceptance, past unpleasant experiences, and other traits because these factors impact the outcome of our research.

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

2. **Trust Escalation:** Scenarios 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:** Scenarios that illustrate a sequence of incidents in which the human subject's trust is decreased, for example, when the SDC cuts off another controlled vehicle.

4. **Trust Mutation:** Second and third scenarios can be 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.

5. **Re-Building Trust:** Scenarios that demonstrate how trust can be rebuilt; for example, the SDC performs smoothly for a reasonable period of time after trust-damaging incidents.

Consider the following sample templates for five consecutive incidents in a 10-minute interaction, as shown in Figure 4.1. For instance, the initial trust measurement occurs within the first few minutes, followed by a trust reduction incident, a further trust reduction incident, and a negative trust mutation followed by rebuilding trust in the last few minutes. Every 2 minutes, the human subject has an opportunity to express whether the interaction is reliable by selecting one of the following options through the VR headset (by staring at the option for five seconds): "Red/No, Light Red/Maybe Not, White/Neutral, Light Green/Maybe Yes, Green/Yes." [1] Note that we considered two out of eight structured templates in the present work.

Template-1	Template-2	Template-3	Template-4	Template-5	Template-6	Template-7	Template-8
Initial Trust	Initial Trust	Initial Trust	Initial Trust	Initial Trust	Initial Trust	Initial Trust	Initial Trust
Trust Reduction	Trust Reduction	Trust Escalation	Trust Escalation	Trust Escalation	Trust Escalation	Trust Reduction	Trust Reduction
Further Trust Reduction	Further Trust Reduction	Further Trust Escalation	Further Trust Escalation	Trust Reduction	Trust Reduction	Trust Escalation	Trust Escalation
Trust Mutation (negative)	Trust Mutation (positive)	Trust Mutation (positive)	Trust Mutation (negative)	Trust Mutation (positive)	Trust Mutation (negative)	Trust Mutation (positive)	Trust Mutation (negative)
Rebuild Trust	Rebuild Trust	Strengthen Trust	Rebuild Trust	Rebuild Trust	Rebuild Trust	Rebuild Trust	Rebuild Trust

Figure 4.1: Structured and Sequential Data Collection: Eight Templates For Various Scenarios.

4.2 EXPERIMENTAL DESIGN

Prior to the simulation, participants were asked to answer 16 demographic and psychological questions by filling out an anonymous survey using Google Forms. Then, participants were randomly placed in one of two possible SDC simulation scenarios. Each scenario is made up of 5 segments. Specific scenario-segment pairs are denoted with a two letter abbreviation followed by the scenario and segment numbers, for example $\mathrm{TR}_{\mathrm{I-II}}$ denotes trust reduction segment II of scenario I. Each segment starts with an exposure to an approximately 2 minute SDC driving simulation followed by a response interval to the question

[1]The colors serve as a visual guide for participants to indicate low trust versus high trust

"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. Tables 4.1 and 4.2 define the scenario and segment pairings.

$\mathrm{IT}_{\mathrm{I-I}}$	Initial Trust
$\mathrm{TR}_{\mathrm{I-II}}$	Trust Reduction
$\mathrm{TR}_{\mathrm{I-III}}$	Further Trust Reduction
$\mathrm{NM}_{\mathrm{I-IV}}$	Negative Trust Mutation
$\mathrm{RT}_{\mathrm{I-V}}$	Rebuild Trust

Table 4.1: Simulation Scenario 1 Based on Template-1

$\mathrm{IT}_{\mathrm{II-I}}$	Initial Trust
$\mathrm{TE}_{\mathrm{II-II}}$	Trust Escalation
$\mathrm{TR}_{\mathrm{II-III}}$	Trust Reduction
$\mathrm{NM}_{\mathrm{II-IV}}$	Negative Trust Mutation
$\mathrm{RT}_{\mathrm{II-V}}$	Rebuild Trust

Table 4.2: Simulation Scenario 2 Based on Template-6

An initial trust/trust escalation section involves the SDC moving slowly and predictably while adhering to the rules of the road. A trust reduction section involves the SDC moving erratically and unpredictably, breaking rules of the road including speeding, tailgating, and sudden lane changes. In the negative trust mutation section, the SDC runs through a stop sign and nearly collides with another car and then proceeds to drive erratically through a residential neighborhood. In the second negative trust mutation segment, the SDC runs through multiple stop signs and detects a pedestrian and a bicyclist crossing a cross walk and comes to a stop. See Figure 4.2 for the scene of the pedestrian crossing the street.

Figure 4.2: Segment NM_{II-IV}: SDC Detects a Pedestrian, But Does Not Come To a Complete Stop

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. An objective of this study is to determine if trust can be rebuilt after the participant experiences a negative trust mutation or general decrease in trust. We predict that exposure to more trust building interactions will help rebuild trust after a trust reducing event. Since participants in Scenario 2 have more time to recover from their trust reduction experience with more exposure to positive trust experiences, we expect them to report higher levels of trust in the SDC simulator than participants in Scenario 1. We are also interested in investigating the sequential impacts of our trust-damaging or trust-building scenarios due to the structured data collection approach that we utilize. For the detail breakdown of trust segments used in scenario 1 and 2, see Figure 4.3

Segment	Summary
IT	Highway driving, defensive driving, using turn signals
TR	0-60 acceleration, frequent lane changes, speeding, not allowing other vehicles to pass
NM	Running stop signs, almost crashing in to other cars, speeding in residential areas, driving in the middle of the road, driving near pedestrians
RT	Defensive driving, using turn signals, driving in traffic jams

Table 4.3: Detailed Breakdown of Negative and Positive Trust Segments in both Scenarios

4.3 RESULTS

Fifty human subjects were recruited to participate in the 10 minute VR autonomous driving simulation [2] . Most of the participants were college students aged 18-30 years and recruited on or near Florida Atlantic University's campus. Half of the participants were randomly selected to be in Scenario 1, and the other half were assigned to Scenario 2.

4.3.1 Demographics & Psychological Questions

As we mentioned previously, before the experiment, we asked each participant sixteen demographics, psychological, and trust questions through an anonymous Google Forms survey. The reason we collected this data prior to the actual experiment was because we wanted to see if there are correlations between the background of the participant, their negative and positive past experiences and their reported trust levels in the simulated SDC. See section 4.3.5 for the correlations found in the actual trust levels.

Figure 4.3 includes pie charts to visualize the combined participant responses for ques-

[2] IRBNET ID #: 1187756-1

tions 1-7 from the anynomous survey we conducted prior to the experiment.

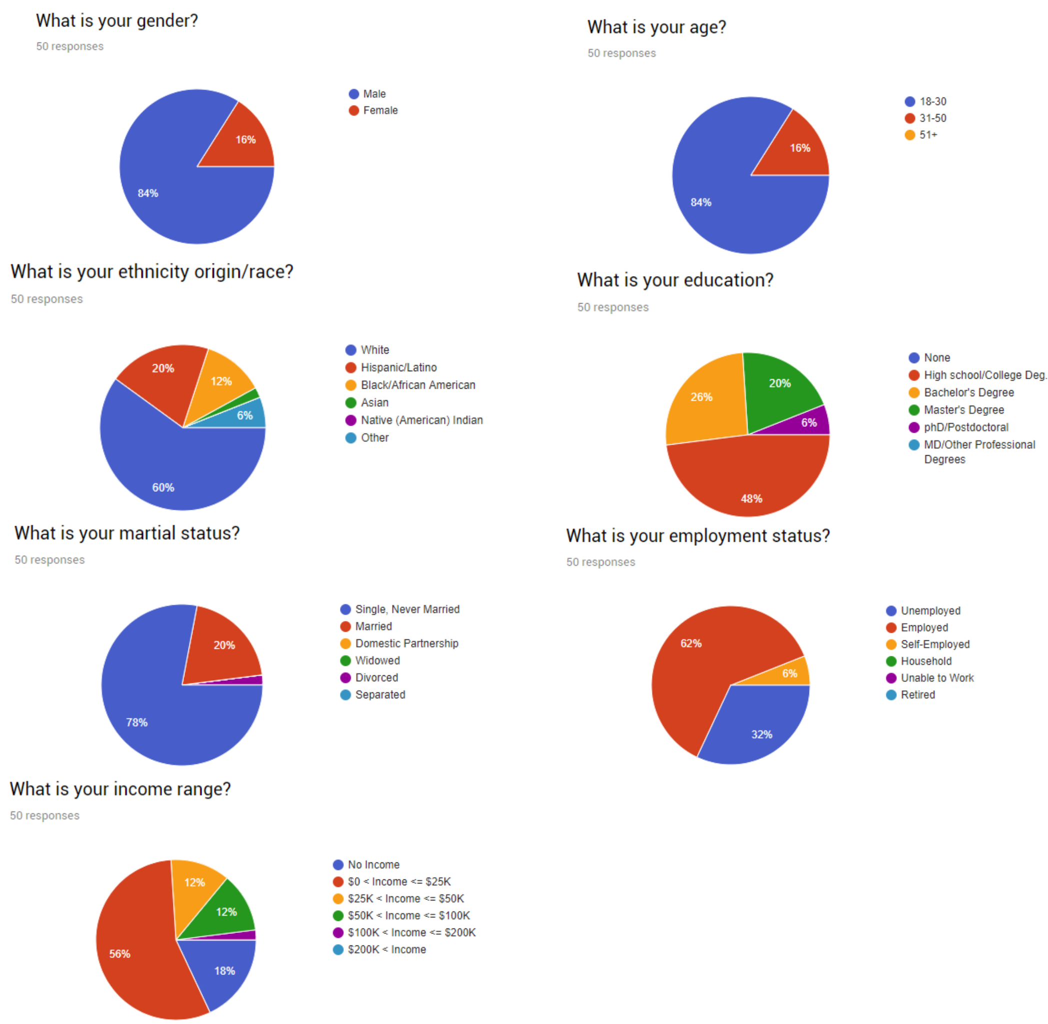

Figure 4.3: Survey Questions 1-7

Figure 4.4 shows the results of combined participant responses for questions 8-11.

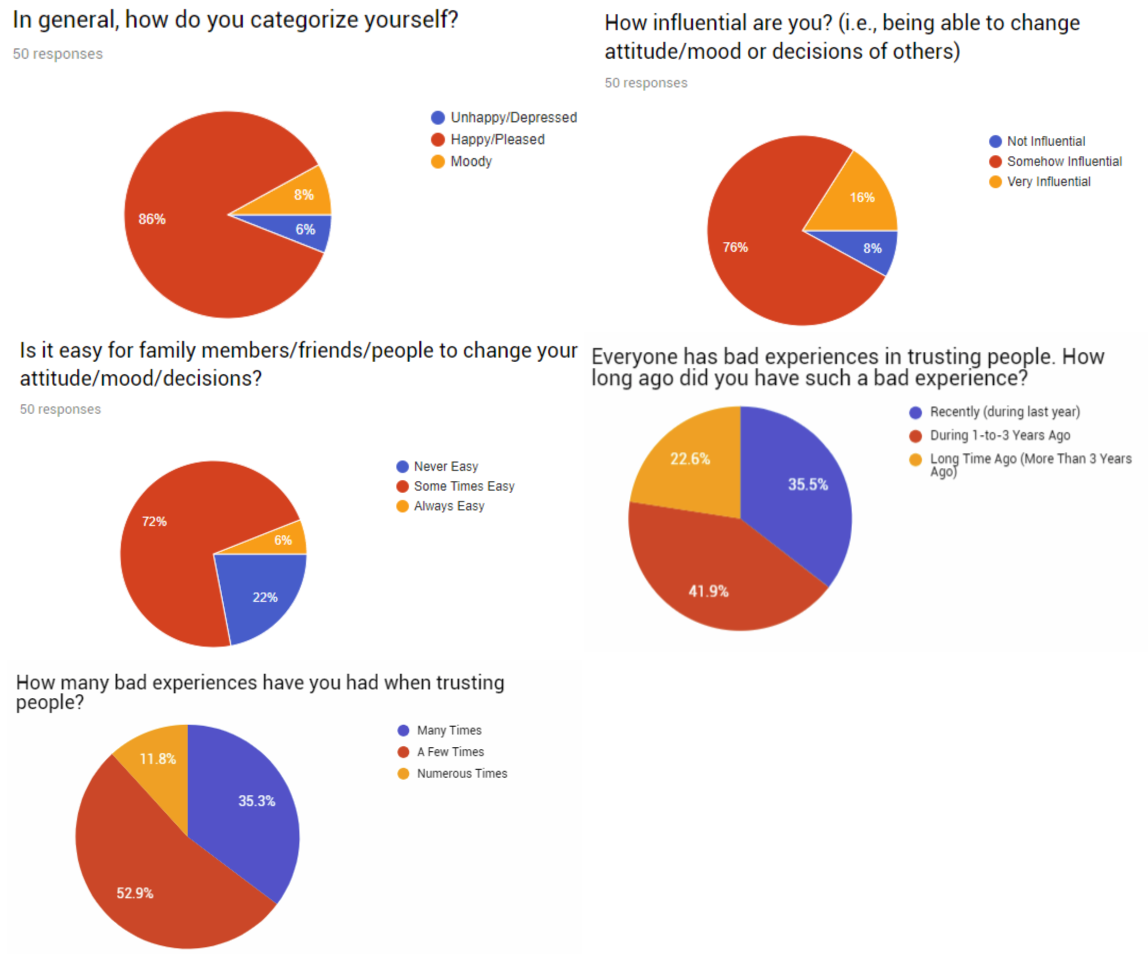

Figure 4.4: Survey Questions 8-11

Figure 4.5 shows the results of combined participant responses for questions 12-16.

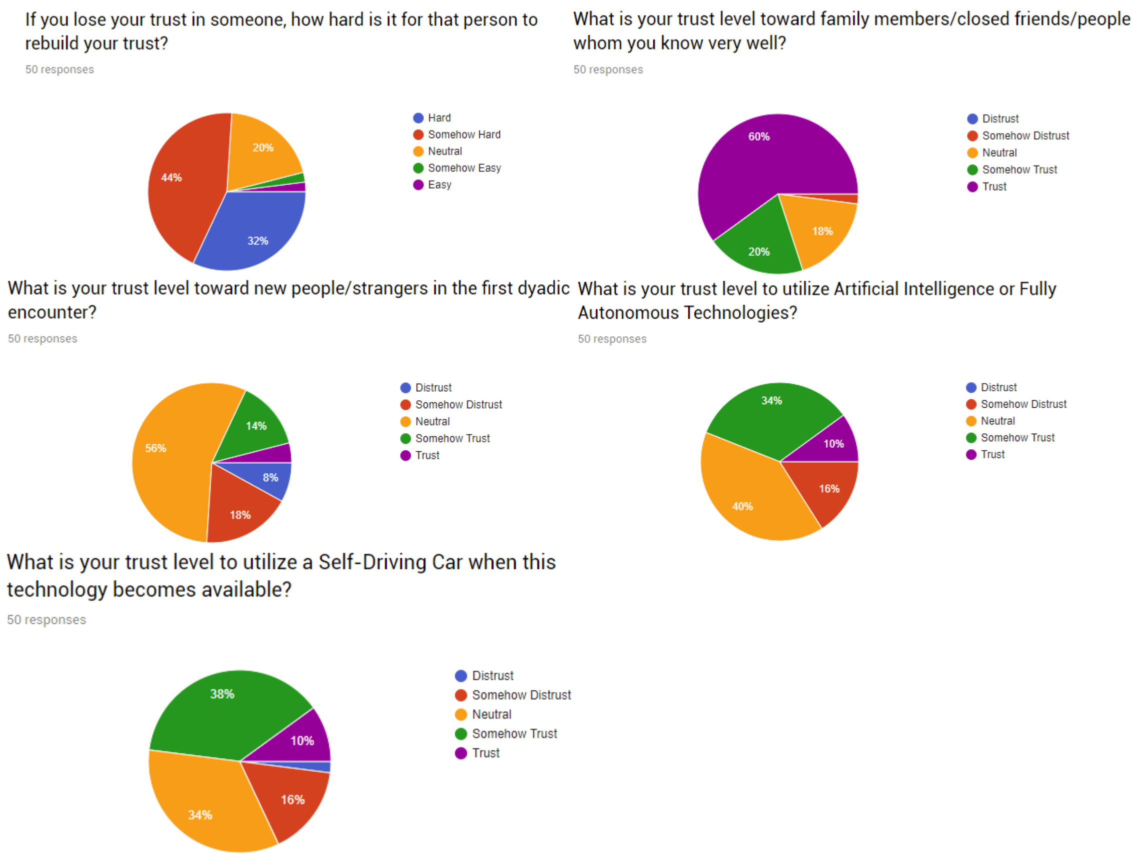

Figure 4.5: Survey Questions 12-16

Figure 4.6 shows the breakdown of the demographic information of Group 1 and Group 2 by gender, age, ethnicity, education, martial status, employment status, and income range.

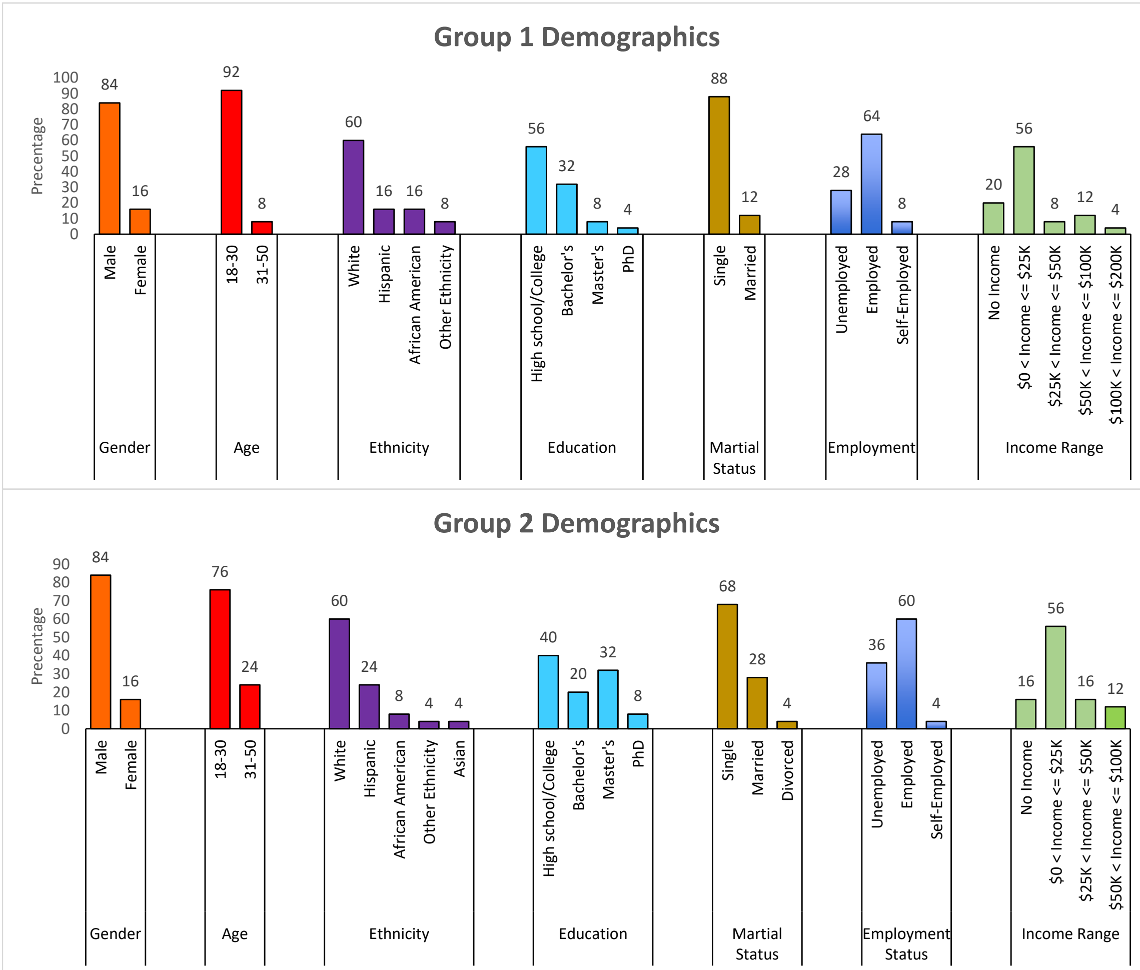

Figure 4.6: Group 1 vs. Group 2 Demographics

Figure 4.7 shows how the participants categorized themselves in terms of happiness, the ability to influence others, the ease of getting influenced by others, and recent bad/trust damaging experiences. [3]

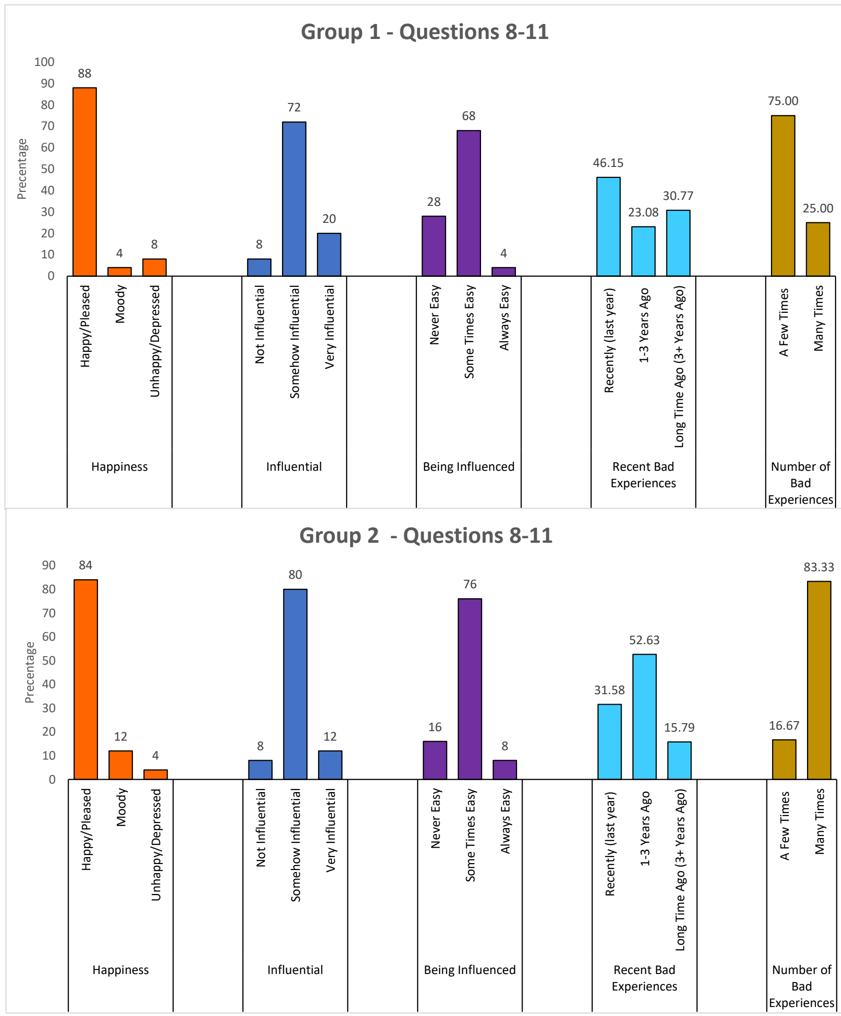

Figure 4.7: Group 1 vs. Group 2 - Psychological Questions 8-11

[3] Question 11 was split into two questions.

Figure 4.8 shows how the participants in Group 1 and 2 categorized themselves in terms of how easily they can rebuild their damaged trust in people, how much they trust their close friends/family, strangers, AI/autonomous systems, and the SDC technology.

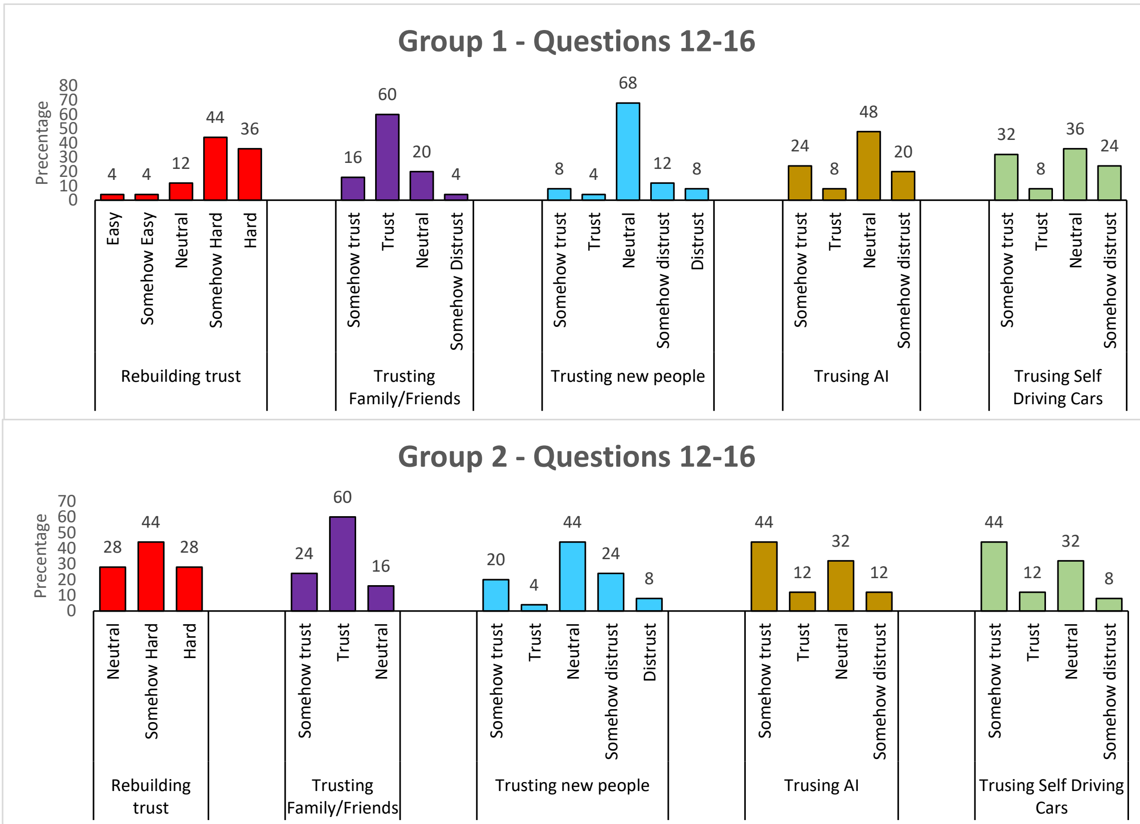

Figure 4.8: Group 1 vs. Group 2 - Psychological Questions 12-16

4.3.2 Scenario 1

Figure 4.9 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.

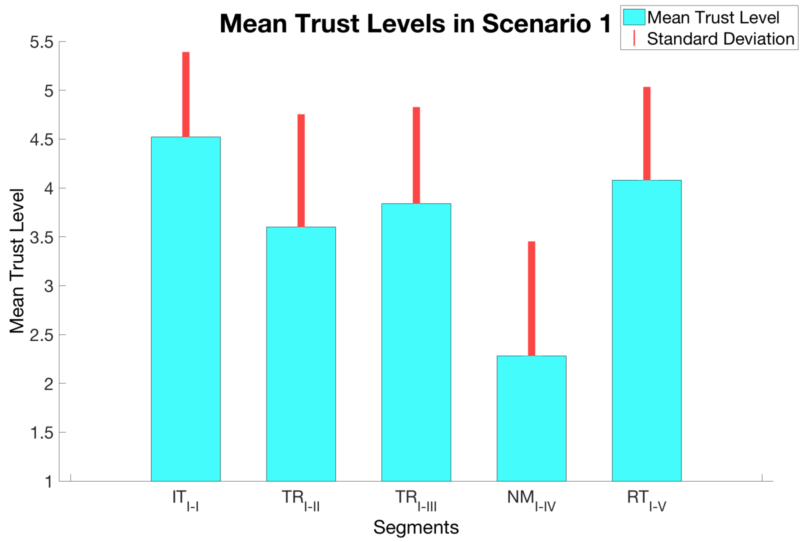

Figure 4.9: Mean Trust Levels Across Segments in Scenario 1

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 4.10. 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}, or TR_{I-III}. All other comparisons show statistically significant changes in trust.

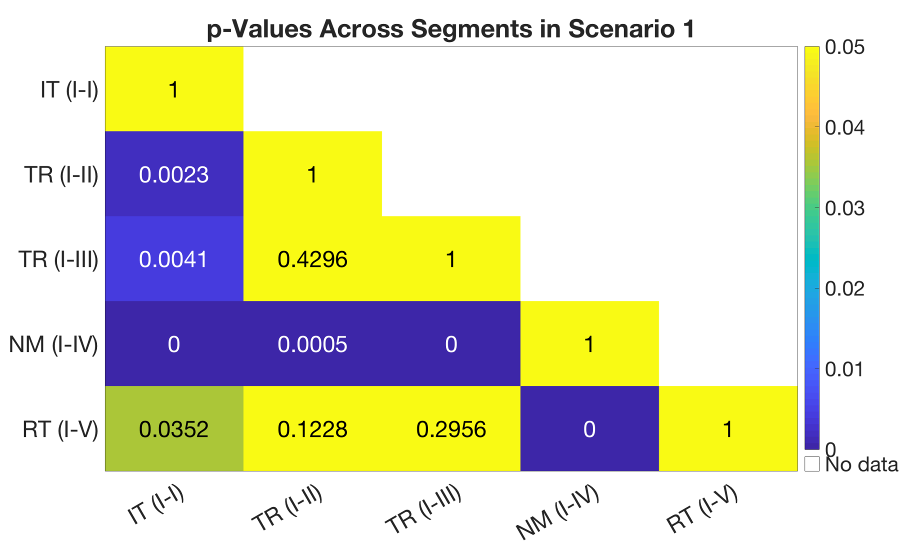

Figure 4.10: Wilcoxon Rank Sum Test P-Values Across Segments in Scenario 1

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 acclimation as the participant becomes more comfortable in the simulation. As expected, the negative trust mutation had the lowest trust levels and was significantly lower than all other segments. This indicates the simulations effectiveness in reducing participants trust.

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.

4.3.3 Scenario 2

Figure 4.11 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}). Here, we can observe that the trust increase from the negative trust mutation (NM_{II-IV}) to Rebuild Trust (RT_{II-V}) segments is not very sharp. This is due to the fact that after experiencing the trust reduction (TR_{II-III}) segment, the participants already had a low trust level.

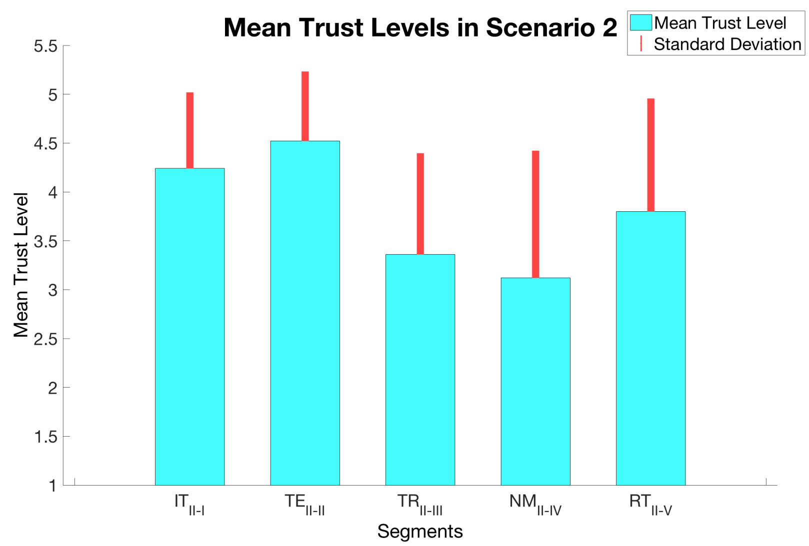

Figure 4.11: Mean Trust Levels Across Segments in Scenario 2

Figure 4.10 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 TR_{II-III} and TE_{II-II}. All other inter-comparisons show statistically significant changes in trust.

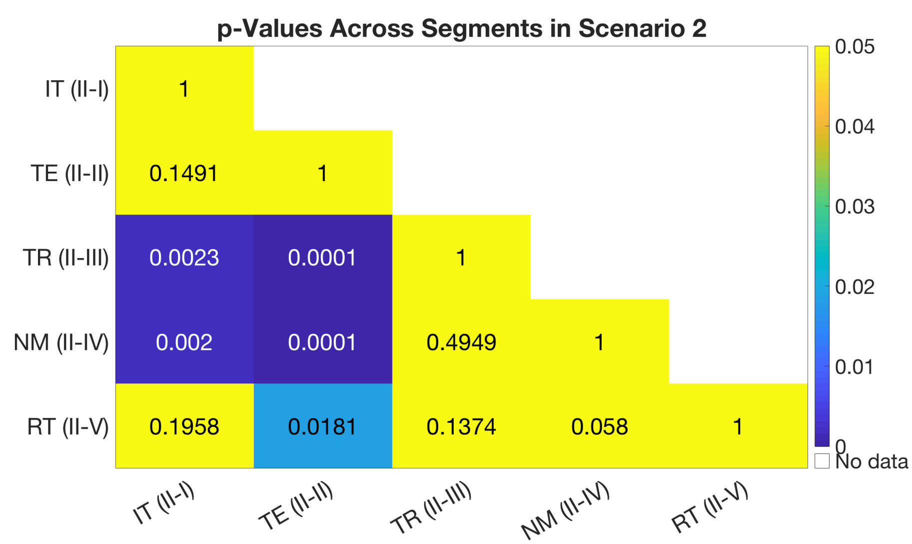

Figure 4.12: Wilcoxon Rank Sum Test: P-Values Across Segments in Scenario 2

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 running through multiple stop signs and interacting near a pedestrian. Participants reported low levels of trust after this segment. Many participants reported that they did not trust the SDC near pedestrians. These results seem to indicate that people do not fully trust SDCs around pedestrians.

Like Scenario 1, in Scenario 2 there is a difference between the initial trust segment and the final segment designed to rebuild trust. While participants score their level of trust after $\mathrm{RT}_{\mathrm{II-V}}$ at 3.80 $\pm$0.23, a high value, it is significantly lower than the initial trust segment (4.24 $\pm$0.15), representing a 13.50% decrease. This seems to indicate that participants trusted the SDC less after being exposed to trust reducing segments.

4.3.4 Scenario 1 vs 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 4.13 illustrates this comparison. Our analysis indicated that participants from both groups reported similar expected trust levels after positive and negative trust segments. The only significant difference that we observed was the reported trust levels for Negative Trust Mutation segments ($\mathrm{NM}_{\mathrm{I-IV}}$ and $\mathrm{NM}_{\mathrm{II-IV}}$). We believe that this is due to the fact that the Negative Trust Mutation segment in Scenario 2 ($\mathrm{NM}_{\mathrm{II-IV}}$) 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 $\mathrm{NM}_{\mathrm{I-IV}}$.

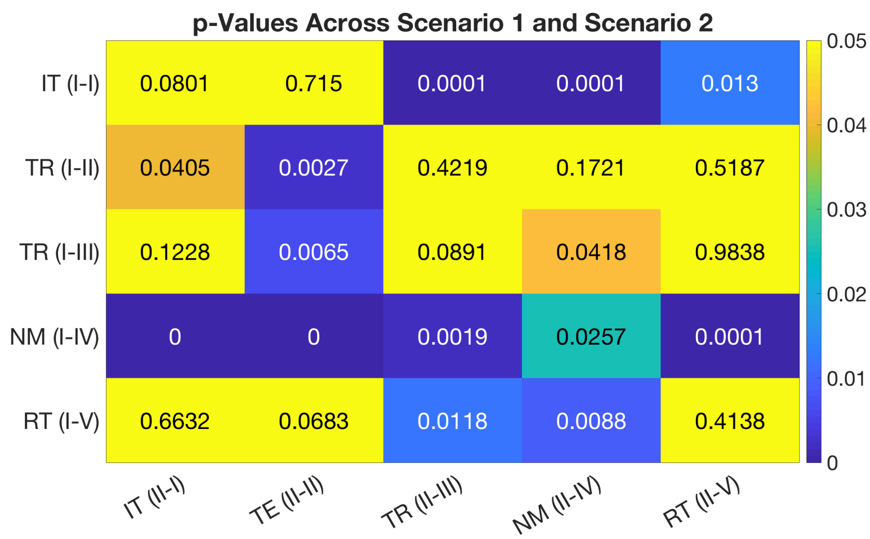

Figure 4.13: Wilcoxon Rank Sum Test: P-Values Across Segments in Scenario 1 & 2

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. Finally, participants in both groups were able to relatively rebuild their trust after the trust damaging Negative Trust Mutation segments.

4.3.5 Data Correlation Between Demographic Data and reported Trust Values

Our data analysis on the demographic and psychological data that we collected from our participants revealed some interesting facts. According to our data, females gained or lost their trust with a slower pace compared to males, who had with a faster pace. We calculated the average value of the reported trust levels of of males and females before and after they experienced damaging simulated segments (NM_{I-IV} in group 1 and NM_{II-IV} in group 2), and calculated their value of their slopes Δt. We discovered that after viewing the trust damaging experience, females had Δt = -0.63, which indicates a slow decline. Similarly, after viewing the trust re-building segments (RT_{I-V} in group 1 and RT_{II-V} in group 2), females had a Δt = 1.29, which is slightly slower than the trust recovery of males (Δt = 1.29). Figure 4.14 visualizes this trust difference.

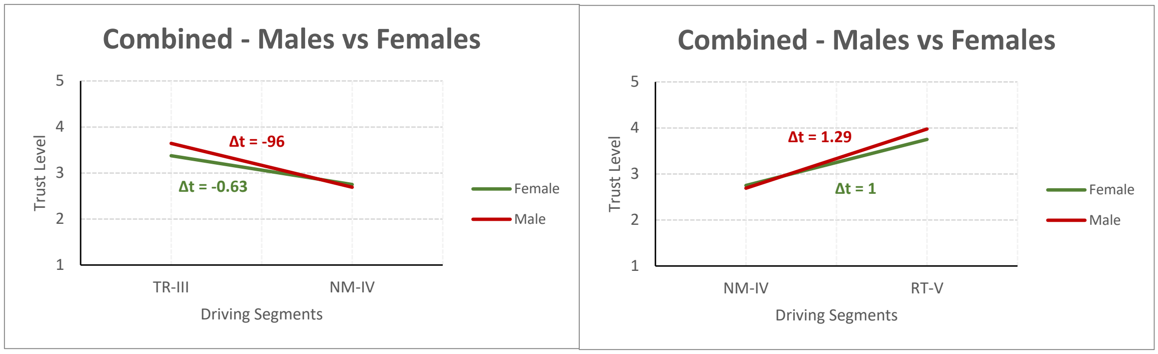

Figure 4.14: Average trust Level Fluctuation Comparison in Males vs Females

Another interesting discovery was that on average, the majority of the participants who were of Hispanic/Latino origin reported higher levels of trust in both groups. Additionally, their trust fluctuations indicated that they gained or lost their trust with a slower pace compared to others with a faster pace. This finding matches the findings of Akash in [21]. Figure 4.15 illustrates the fluctuations of trust levels in the Hispanic/Latino participants compared to participants of other ethnic backgrounds. We are comparing the sharpness of the slope Δt from segment Trust Reduction (TR_{III}) to Negative Trust Mutation (NM_{IV}) and from Negative Trust Mutation (NM_{IV}) to Rebuild Trust (RT_V).

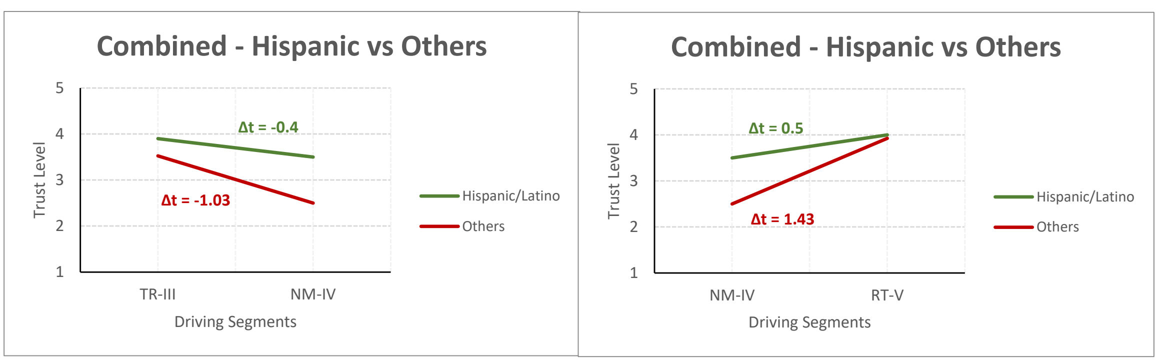

Figure 4.15: Average Trust Fluctuation of Hispanic/Latino Participants vs Others

We also observed that subjects who expressed that their trust can be easily rebuilt, regained their trust with a faster pace versus subjects who mentioned otherwise. See Figure 4.16 for the comparison of the slopes. As shown, participants who declared that they can easily rebuild their damaged trust had a sharper positive slope (Δt = 1.5) compared to the participants who reported that they usually have difficulty regaining their trust in people ($\Delta t = 1.26$).

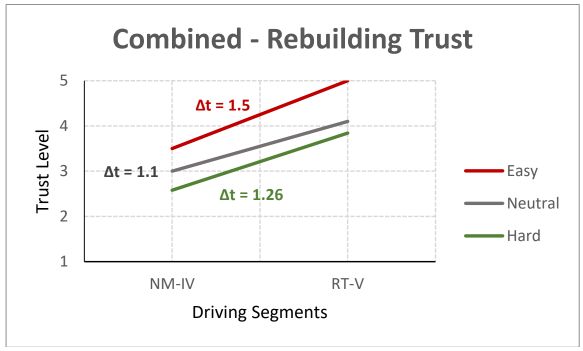

Figure 4.16: Average Trust Fluctuation Comparison for Rebuilding Trust

Subjects with recent bad trust-related experiences gained or lost their trust with a faster pace whereas, subjects who had bad trust-related experiences long time ago, gained or lost their trust with a slower pace. Illustrated in Figure 4.17, After experiencing the Negative Trust Mutation driving segments (NM_{IV}), subjects with recent bad experiences had $\Delta t = -1$, which indicates a steep decline for the trust decrease when compared to $\Delta t = -0.77$ and $\Delta t = -0.73$. Similarly, after the participants went through the Rebuild Trust segment (RT_V), subjects with recent bad experiences had $\Delta t = 1.5$ compared to $\Delta t = 1.077$ and $\Delta t = 0.86$.

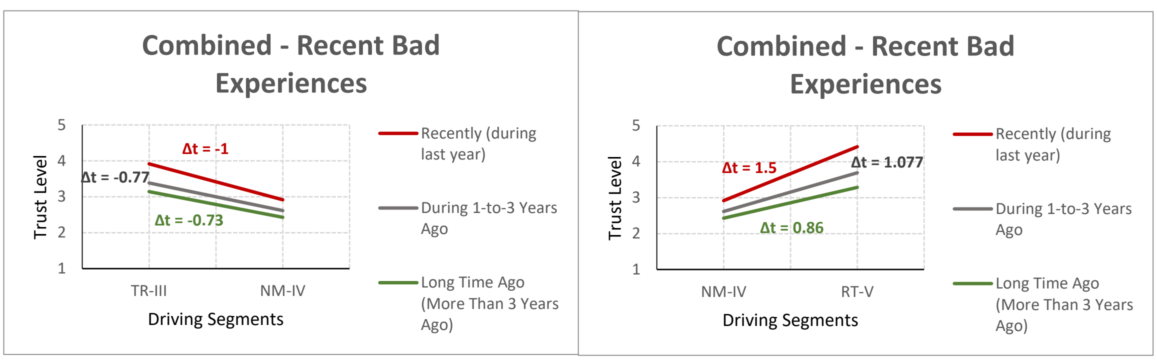

Figure 4.17: Recent Trust Damaging Experiences in the Participants

Subjects with a high-level of trust in SDC technology lost their trust with a slower pace and regained their trust with a faster pace whereas, subjects with a low-level of trust in this technology, lost their trust with a faster pace and regained their trust with a slower pace. Figure 4.18 illustrates this discovery.

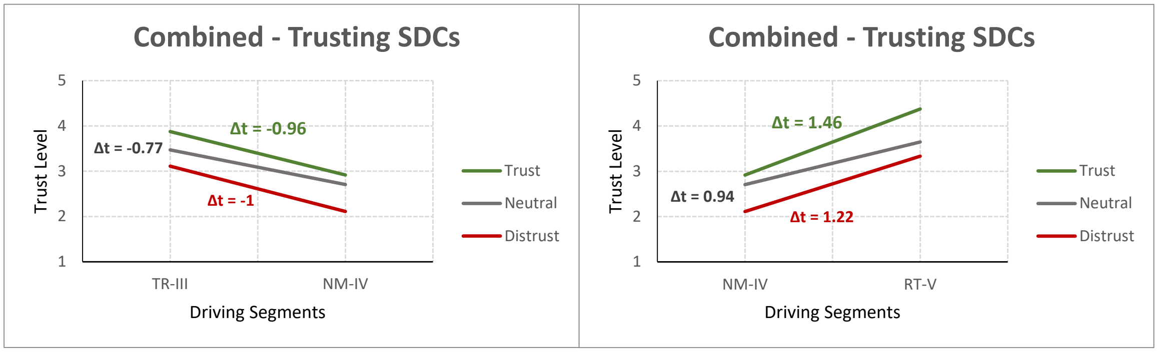

Figure 4.18: Participants Who Trust SDCs vs Others

It is worth mentioning our experiment did not reveal any meaningful results for the data correlation between the age of the participants, their income levels, their employment status and the reported trust values. This is due to the fact that the majority of our participants were unemployed 18-30 year old college students.

4.4 CONCLUSION AND FUTURE WORK

This chapter focuses on understanding fluctuations in the trust levels of passengers of self-driving cars. The data indicating the trust levels of human subjects were collected using an advanced self-reporting system after they experienced a VR simulation which consisted of a series of simulated autonomous driving videos. The results of the experiment indicated that the humans trust levels change depending on the autonomous driving style. Furthermore, we discovered that most of the participants were able to moderately rebuild their trust in the simulated SDC after faulty and erratic behaviors, and passengers are usually distrusting of the SDC near pedestrians.

After analyzing our collected demographic data, we discovered that certain attributes such as cultural background, gender, the number of past incidents, trust in social situations, and the current perception of the autonomous driving technology directly affect they way people trust the simulated SDC, and the recovery of this trust after trust damaging incidents.

In the next iterations of our research, we will improve our SDC VR simulation scenarios to make them more realistic. There were some minor deficiencies such as the motion of the motion simulator going slightly out of sync with the car movements in some parts. Additionally, some of the participants stated that they cannot see the side mirrors or the rear-view mirrors in the simulation, and this could have had an impact on their trust levels. Another area that we would like to improve is introducing segments with hazardous conditions such as heavy rain, storms, and snow to see if they would have a direct impact on the passengers trust levels. Furthermore, we would like to replace our trust self-reporting system with automatic detection of trust level via psychological sensors and non-intrusive devices such as EEG, heart rate sensors, and a facial recognition module.

CHAPTER 5

CONCLUDING REMARKS AND FUTURE DIRECTIONS

In this thesis, a new approach was introduced for measuring the trustworthiness of autonomous vehicles in real time. This approach employs a new structured data collection method, an advanced self-reporting trust system, and a realistic VR self-driving car simulator to study the fluctuations of trust levels of the passengers of self-driving cars. To test this novel approach, an empirical experiment was designed and conducted on 50 human subjects. The analysis of the experimental results indicated that the autonomous driving style directly influences the trust of the passengers in the system. For instance, aggressive driving diminishes trust, and defensive, predictable driving increases (builds) trust. Furthermore, most human subjects indicated signs of rebuilding their broken trust in the system after experiencing aggressive, and reckless driving scenarios.

We also discovered that certain traits such as cultural background, gender, past trauma, the psychological state, and the current perception of the autonomous driving technology by the test subjects have a high degree of correlation with their reported levels of trust in the simulated system. For instance, we discovered that females have a slower trust recovery and loss rate, the Hispanic/Latino passengers are more likely to have more trust in the system, and subjects recent negative trust-related experiences gain or lose their trust faster.

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 level of trust and psychological responses when exposed to different driving scenarios. We believe that the presented approach in this thesis is novel, and can potentially lay the foundation of a wide variety of future research in the context of trust, HRI, and autonomous vehicles. 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 in this domain is more important than ever before.

Within the context of trust and HRI, we are interested in further research and refinements to our proposed approach, and simulation technology. 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 the primary future research directions that we are planning to explore in the next iterations of our research [20] [78] [79].

APPENDICES

APPENDIX A: DEMOGRAPHIC QUESTIONS

Demographic Information

Please do not provide any personal information on this form so that we can keep your identity fully anonymous.

(01) What is your gender?
❑ Male ❑ Female

(02) What is your age?
❑ 18-30 ❑ 31-50 ❑ 51-over

(03) What is your ethnicity origin/race?
❑ White ❑ Hispanic/Latino ❑ Black/African American
❑ Asian ❑ Native (American) Indian ❑ Other

(04) What is your education?
❑ None ❑ High School/College Deg. ❑ Bachelor’s Degree
❑ Master’s Degree ❑ PhD/Postdoctoral ❑ MD/Other Professional Degrees

(05) What is your marital status?
❑ Single, Never Married ❑ Married ❑ Domestic Partnership
❑ Widowed ❑ Divorced ❑ Separated

(06) What is your employment status?
❑ Unemployed ❑ Employed ❑ Self-Employed
❑ Household ❑ Unable to Work ❑ Retired

(07) What is your income range?
❑ No Income ❑ \$0 < Income <= \$25K ❑ \$25K < Income <= \$50K
❑ \$50K < Income <= \$100K ❑ \$100K < Income <= \$200K ❑ \$200K < Income

(08) In general, how do you categorize yourself?
❑ Unhappy/Depressed ❑ Happy/Pleased ❑ Moody

(09) How influential are you? (i.e., being able to change attitude/mood or decisions of others)
❑ Not Influential ❑ Somehow Influential ❑ Very Influential

(10) Is it easy for family members/friends/people to change your attitude/mood/decisions?
❑ Never Easy ❑ Some Times Easy ❑ Always Easy

(11) Everyone has bad experiences in trusting people. How long ago did you have such a bad experience? And for how many times in your life?
❑ Recently (during last year) ❑ During 1-to-3 Years Ago ❑ Long Time Ago (More Than 3 Years Ago)
❑ A Few Times ❑ Many Times ❑ Numerous Times

(12) If you lose your trust in someone, how hard is it for that person to rebuild your trust?
❑ Hard ❑ Somehow Hard ❑ Neutral ❑ Somehow Easy ❑ Easy

(13) What is your trust level toward family members/closed friends/people whom you know very well?
❑ Distrust ❑ Somehow Distrust ❑ Neutral ❑ Somehow Trust ❑ Trust

(14) What is your trust level toward new people/strangers in the first dyadic encounter?
❑ Distrust ❑ Somehow Distrust ❑ Neutral ❑ Somehow Trust ❑ Trust

(15) What is your trust level to utilize Artificial Intelligence or Fully Autonomous Technologies?
❑ Distrust ❑ Somehow Distrust ❑ Neutral ❑ Somehow Trust ❑ Trust

(16) What is your trust level to utilize a Self-Driving Car when this technology becomes available?
❑ Distrust ❑ Somehow Distrust ❑ Neutral ❑ Somehow Trust ❑ Trust

APPENDIX B: PUBLICATIONS

The results of this research appear in the following publications:

[1] Originally Published in: Shervin Shahrdar, Luiza Menezes, and Mehrdad Nojoumian. "A Survey on Trust in Autonomous Systems". IEEE Computing Conference (CC). London, UK: IEEE, 2018. Reproduced with permission from IEEE

[2] To be Published in: Shervin Shahrdar, Corey Park, Mehrdad Nojoumian "Real-Time Trust Measurement in Human-Autonomous Vehicle Interactions Using a VR-Motion Simulator and a New Structured Data Collection Approach". Submitted to International Journal of Human-Computer Interaction, Special Issue: Human-System Cooperation in Autonomous Driving (IJGT) (Accepted With Revision)

[3] To be Published in: Corey Park, Shervin Shahrdar, Mehrdad Nojoumian "EEG-Based Classification of Emotional State Using an Autonomous Vehicle Simulator". Submitted to the 10th IEEE Sensor Array and Multichannel Signal Processing Workshop (Under Review)

BIBLIOGRAPHY

[1] "Consumer concerns about self-driving cars." https://www.statista.com/chart/5950/concerns-about-self-driving-cars/. Accessed March 31st, 2018.

[2] article36, "ban-autonomous-armed-robots/," *article36*, 2012.

[3] P. A. Hancock, D. R. Billings, K. E. Schaefer, J. Y. Chen, E. J. De Visser, and R. Parasuraman, "A meta-analysis of factors affecting trust in human-robot interaction," *The Journal of the Human Factors and Ergonomics Society*, vol. 53, no. 5, pp. 517–527, 2011.

[4] "Atomic motion systems sign manufacturing licence agreement with talon simulations." http://www.atomicmotionsystems.com/?p=11600. Accessed April 6, 2018.

[5] "Oculus rift official website." https://www.oculus.com/rift/#oui-csl-rift-games=star-trek. Accessed April 7, 2018.

[6] "Cleararmor official website." https://cleararmor.biz/. Accessed April 7, 2018.

[7] "360fly review." https://www.tomsguide.com/us/360fly-4k,review-3734.html. Accessed April 7, 2018.

[8] "Gopro." https://gopro.com/fusion. Accessed January 22, 2018.

[9] J. Beer, A. D. Fisk, and W. A. Rogers, "Toward a framework for levels of robot autonomy in human-robot interaction," *Journal of Human-Robot Interaction*, vol. 3, no. 2, p. 74, 2014.

[10] R. Parasuraman and V. Riley, "Humans and automation: Use, misuse, disuse, abuse," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 39, no. 2, pp. 230–253, 1997.

[11] J. D. Lee and K. A. See, "Trust in automation: Designing for appropriate reliance," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 46, no. 1, pp. 50–80, 2004.

[12] J. Koo, J. Kwac, W. Ju, M. Steinert, L. Leifer, and C. Nass, "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)*, vol. 9, no. 4, pp. 269–275, 2015.

[13] J. K. Choi and Y. G. Ji, "Investigating the importance of trust on adopting an autonomous vehicle," *International Journal of Human-Computer Interaction*, vol. 31, no. 10, pp. 692–702, 2015.

[14] P. E. Ross, "Ces 2017: Nvidia and audi say they'll field a level 4 autonomous car in three years," Jan 2017.

[15] A. Sage and P. Lienert, "Ford plans self-driving car for ride share fleets in 2021," Aug 2016.

[16] S. Gaudin, "Autonomous cars will arrive within 10 years, intel cto says," Oct 2012.

[17] M. S. Carlson, M. Desai, J. L. Drury, H. Kwak, and H. A. Yanco, "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*, pp. 20–27, 2014.

[18] "Tesla car that crashed and killed driver was running on autopilot, firm says." https://www.theguardian.com/technology/2018/mar/31/tesla-car-crash-autopilot-mountain-view. Accessed April 5, 2018.

[19] "Ubers self-driving cars were struggling before arizona crash." https://www.nytimes.com/2018/03/23/technology/uber-self-driving-cars-arizona.html. Accessed April 5, 2018.

[20] S. Shahrdar, L. Menezes, and M. Nojoumian, "A survey on trust in autonomous systems," in *To appear in IEEE Computing Conference (CC)*, 2018.

[21] K. Akash, W.-L. Hu, T. Reid, and N. Jain, "Dynamic modeling of trust in human-machine interactions," in *American Control Conference (ACC), 2017*, pp. 1542–1548, IEEE, 2017.

[22] W.-L. Hu, K. Akash, N. Jain, and T. Reid, "Real-time sensing of trust in human-machine interactions," *IFAC-PapersOnLine*, vol. 49, no. 32, pp. 48–53, 2016.

[23] K. E. Schaefer, *The perception and measurement of human-robot trust*. PhD thesis, University of Central Florida Orlando, Florida, 2013.

[24] autonomy., *Full Definition of autonomy*. merriam-webster, 2016.

[25] G. A. Bekey, *Autonomous Robots: From Biological Inspiration to Implementation and Control (Intelligent Robotics and Autonomous Agents)*. The MIT Press, 2005.

[26] M. Robinson, "Science and technology in the industrial revolution," *University of Toronto Press*, 1969.

[27] S. O.-R. A. V. S. Committee *et al.*, "Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems," *SAE International*, 2014.

[28] B. D. Adams, L. E. Bruyn, and S. Houde, *Trust in Automated Systems, Literature Review*. Humansystems Incorporated, 2003.

[29] M. Nojoumian and T. C. Lethbridge, “A new approach for the trust calculation in social networks,” in *E-business and Telecommunication Networks: 3rd International Conference on E-Business, Best Papers*, vol. 9 of *CCIS*, pp. 64–77, Springer, 2008.

[30] M. Nojoumian, “Trust, influence and reputation management based on human reasoning,” in *4th AAAI Workshop on Incentives and Trust in E-Communities, WIT-EC’15*, pp. 21–24, 2015.

[31] M. Nojoumian and D. R. Stinson, “Social secret sharing in coud computing using a new trust function,” in *10th IEEE Annual International Conference on Privacy, Security and Trust (PST)*, pp. 161–167, 2012.

[32] Trust., *Definition of trust in English:*. oxford-dictionaries., 2016.

[33] A. Uggirala, A. K. Gramopadhye, B. J. Melloy, and J. E. Toler, “Measurement of trust in complex and dynamic systems using a quantitative approach,” *International Journal of Industrial Ergonomics*, vol. 34, no. 3, pp. 175–186, 2004.

[34] T. Helldin, G. Falkman, M. Riveiro, and S. Davidsson, “Presenting system uncertainty in automotive uis for supporting trust calibration in autonomous driving,” in *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pp. 210–217, ACM, 2013.

[35] D. Howard and D. Dai, “Public perceptions of self-driving cars: The case of berkeley, california,” in *Transportation Research Board 93rd Annual Meeting*, 2014.

[36] M. Kyriakidis, R. Happee, and J. De Winter, “Public opinion on automated driving: results of an international questionnaire among 5000 respondents,” *Transportation research part F: traffic psychology and behaviour*, vol. 32, pp. 127–140, 2015.

[37] M. Wagner and P. Koopman, “A philosophy for developing trust in self-driving cars,” in *Road Vehicle Automation 2*, pp. 163–171, Springer, 2015.

[38] V. Butakov and P. Ioannou, “Driving autopilot with personalization feature for improved safety and comfort,” in *18th International Conference on Intelligent Transportation Systems*, pp. 387–393, IEEE, 2015.

[39] W. Payre, J. Cestac, and P. Delhomme, “Fully automated driving impact of trust and practice on manual control recovery,” *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 58, no. 2, pp. 229–241, 2016.

[40] R. A. Daziano, M. Sarrias, and B. Leard, “Are consumers willing to pay to let cars drive for them? analyzing response to autonomous vehicles,” *Transportation Research Part C: Emerging Technologies*, vol. 78, pp. 150 – 164, 2017.

[41] B. M. Muir, “Trust between humans and machines, and the design of decision aids,” *International Journal of Man-Machine Studies*, vol. 27, no. 5-6, pp. 527–539, 1987.

[42] D. Power, *Types of Decision Support Systems*. gdrc.org, 2016.

[43] P. Madhavan and D. A. Wiegmann, “Similarities and differences between human–human and human–automation trust: an integrative review,” *Theoretical Issues in Ergonomics Science*, vol. 8, no. 4, pp. 277–301, 2007.

[44] J. Lee and N. Moray, “Trust, control strategies and allocation of function in human-machine systems,” *Ergonomics*, vol. 35, no. 10, pp. 1243–1270, 1992.

[45] B. M. Muir, “Trust in automation: Part i. theoretical issues in the study of trust and human intervention in automated systems,” *Ergonomics*, vol. 37, no. 11, pp. 1905–1922, 1994.

[46] B. M. Muir and N. Moray, “Trust in automation. part ii. experimental studies of trust and human intervention in a process control simulation,” *Ergonomics*, vol. 39, no. 3, pp. 429–460, 1996.

[47] I. Dassonville, D. Jolly, and A. Desodt, “Trust between man and machine in a teleoperation system,” *Reliability Engineering & System Safety*, vol. 53, no. 3, pp. 319–325, 1996.

[48] N. Moray and T. Inagaki, “Laboratory studies of trust between humans and machines in automated systems,” *Transactions of the Institute of Measurement and Control*, vol. 21, no. 4-5, pp. 203–211, 1999.

[49] M. Itoh, G. Abe, and K. Tanaka, “Trust in and use of automation: their dependence on occurrence patterns of malfunctions,” in *Systems, Man, and Cybernetics, 1999. IEEE SMC’99 Conference Proceedings. 1999 IEEE International Conference on*, vol. 3, pp. 715–720, IEEE, 1999.

[50] M. T. Dzindolet, S. A. Peterson, R. A. Pomranky, L. G. Pierce, and H. P. Beck, “The role of trust in automation reliance,” *International Journal of Human-Computer Studies*, vol. 58, no. 6, pp. 697–718, 2003.

[51] S. M. Merritt and D. R. Ilgen, “Not all trust is created equal: Dispositional and history-based trust in human-automation interactions,” *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 50, no. 2, pp. 194–210, 2008.

[52] R. R. Murphy, D. Riddle, and E. Rasmussen, “Robot-assisted medical reachback: a survey of how medical personnel expect to interact with rescue robots,” in *Robot and Human Interactive Communication, 2004. ROMAN 2004. 13th IEEE International Workshop on*, pp. 301–306, IEEE, 2004.

[53] J. Carlson and R. R. Murphy, “An investigation of mml methods for fault diagnosis in mobile robots,” in *Intelligent Robots and Systems, 2004.(IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on*, vol. 1, pp. 180–186, IEEE, 2004.

[54] R. Parasuraman and C. A. Miller, "Trust and etiquette in high-criticality automated systems," *Communications of the ACM*, vol. 47, no. 4, pp. 51–55, 2004.

[55] D. P. Stormont, "Analyzing human trust of autonomous systems in hazardous environments," in *Proc. of the Human Implications of Human-Robot Interaction workshop at AAAI*, pp. 27–32, 2008.

[56] D. Li, P. P. Rau, and Y. Li, "A cross-cultural study: Effect of robot appearance and task," *International Journal of Social Robotics*, vol. 2, no. 2, pp. 175–186, 2010.

[57] R. E. Yagoda and D. J. Gillan, "You want me to trust a robot? the development of a human–robot interaction trust scale," *International Journal of Social Robotics*, vol. 4, no. 3, pp. 235–248, 2012.

[58] J. Penders, P. Jones, A. Ranasinghe, and T. Nanayakara, "Enhancing trust and confidence in human robot interaction," tech. rep., 2013.

[59] C. Castelfranchi and R. Falcone, "Trust and control: a dialectic link," *Applied Artificial Intelligence*, vol. 14, no. 8, pp. 799–823, 2000.

[60] Y. Wang, Z. Shi, C. Wang, and F. Zhang, "Human-robot mutual trust in (semi) autonomous underwater robots," in *Cooperative Robots and Sensor Networks 2014*, pp. 115–137, Springer, 2014.

[61] M. Abd, I. Gonzalez, M. Nojoumian, and E. Engeberg, "Impacts of robot assistant performance on human trust, satisfaction, and frustration," in *To appear in RSS: Morality and Social Trust in Autonomous Robots*, 2017.

[62] M. Abd, I. Gonzalez, M. Nojoumian, and E. Engeberg, "Trust, satisfaction and frustration measurements for real-time human-robot interaction," in *30th Florida Conference on Recent Advances in Robotics*, pp. 89–93, 2017.

[63] P. de Vries, C. Midden, and D. Bouwhuis, "The effects of errors on system trust, self-confidence, and the allocation of control in route planning," *International Journal of Human-Computer Studies*, vol. 58, no. 6, pp. 719–735, 2003.

[64] C. Kelly, "Guidelines for trust in future atm systems-principles," tech. rep., 2003.

[65] X. Jiang, M. T. Khasawneh, R. Master, S. R. Bowling, A. K. Gramopadhye, B. J. Melloy, and L. Grimes, "Measurement of human trust in a hybrid inspection system based on signal detection theory measures," *Int. J of Industrial Ergonomics*, vol. 34, no. 5, pp. 407–419, 2004.

[66] Y. Seong and A. M. Bisantz, "The impact of cognitive feedback on judgment performance and trust with decision aids," *Int. J of Industrial Ergonomics*, vol. 38, no. 7, pp. 608–625, 2008.

[67] S. R. Winter, S. Rice, R. Mehta, I. Cremer, K. M. Reid, T. G. Rosser, and J. C. Moore, "Indian and american consumer perceptions of cockpit configuration policy," *Journal of air transport management*, vol. 42, pp. 226–231, 2015.

[68] “Atomic motion systems.” http://www.atomicmotionsystems.com/. Accessed January 22, 2018.

[69] “Video presentation of hybrid reality work performed.” https://www.youtube.com/watch?v=BV3cc8ZXCTg&ab_channel=NASAHybridRealityLab. Accessed April 7, 2018.

[70] “Talon simulation clients.” https://www.talonsimulations.com/clients.html. Accessed April 7, 2018.

[71] “Facebook purchases oculus for 2 billion.” https://goo.gl/Kk5HSB. Accessed April 7, 2018.

[72] C. Moro, Z. Stromberga, and A. Stirling, “Virtualisation devices for student learning: Comparison between desktop-based (oculus rift) and mobile-based (gear vr) virtual reality in medical and health science education.,” *Australasian Journal of Educational Technology*, vol. 33, no. 6, 2017.

[73] “Simtools v2 - video ride player.” https://www.xsimulator.net/community/marketplace/simtools-v2-video-ride-player.175/. Accessed April 7, 2018.

[74] “Why do all the vr videos today look so pixelated?.” https://medium.com/visbit/why-do-all-the-360-vr-videos-today-look-so-pixelated-b1ab3cba6f95. Accessed April 7, 2018.

[75] “Motionlogger.” https://goo.gl/ZQGpFC. Accessed April 7, 2018.

[76] “Sensordata.” https://itunes.apple.com/us/app/sensor-data/id397619802?mt=8. Accessed April 7, 2018.

[77] J. J. LaViola Jr, “A discussion of cybersickness in virtual environments,” *ACM SIGCHI Bulletin*, vol. 32, no. 1, pp. 47–56, 2000.

[78] S. Shahrdar, P. Corey, and M. Nojoumian, “Real-time trust measurement in human-autonomous vehicle interactions using a vr-motion simulator and a new structured data collection approach,” in *International Journal of Human-Computer Interaction Special Issue: Human-System Cooperation in Autonomous Driving (IJGT) (Accepted With Revision)*, 2018.

[79] C. Park, S. Shahrdar, and M. Nojoumian, “Eeg-based classification of emotional state using an autonomous vehicle simulator,” in *IEEE Sensor Array and Multichannel Signal Processing Workshop (Under Review)*, 2018.