ASSESSING THE DEVELOPMENT OF OPERATOR TRUST IN AUTOMATION

A Thesis

by

MARGARET JANE FOWLER

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Farzan Sasangohar
Co-Chair of Committee,	Maryam Zahabi
Committee Members,	Robert Brydia
Head of Department,	Alfredo Garcia

May 2021

Major Subject: Industrial Engineering

ABSTRACT

Miscalibrated relationships between operator trust and automation can lead to accidents, some even fatal. If an operator either over or under trusts the system's capability, their overall assessment of the system's reliability can be inaccurate and potentially lead to poor decision making. As autonomous vehicles emerge, understanding the natural trust formation process as it occurs over time between drivers and these vehicles is crucial to increase safety and reliability, as well as identifying any factors that can affect this process. To fill this gap, an autonomous vehicle was observed as it operated on Texas A&M University's campus in mixed traffic for an 8-week demonstration.

Throughout the deployment, the vehicle was operated autonomously and used four safety operators from the student population to take over shuttle operations, as necessary. Research personnel collected daily and weekly surveys and hosted interviews to investigate how operators' trust developed and changed over time and to study the relationship between trust and operational factors. Preliminary findings established a potential relationship between trust and the number of vehicle errors. Interview data also suggested that trust was dependent on situational circumstances affected by the operator’s emotional comfort and familiarity with the vehicle.

DEDICATION

"At Texas A&M you learn first to follow, then to develop and practice your leadership skills, and finally you become someone that others want to follow."

- James R. Thompson, Class of 1968

This thesis is dedicated to the strong women in my life, my Mom and sister, Camella, and my friends Cassidy, Elma, Eliana, Kyndall, Kyra, Sarah, and Shelby.

May you all forever be honorary members of the BB club.

ACKNOWLEDGEMENTS

In 1965 my father, Neal Vester Fowler Jr., attended Allen Academy in Bryan, TX and later studied engineering at Texas A&M University. Whenever it was time for me to apply to college, like most teenagers, I protested the idea of following in his footsteps. However, I am here to state in writing that I was wrong. Not only has Texas A&M University and the City of Bryan/College Station provided so much opportunity, but it has allowed me to work and learn from leaders like no other. Thank you, Dad.

Of these leaders, first, I would like to thank Bob Brydia for not only mentoring me throughout my graduate program but providing me an opportunity to work alongside him at the Texas A&M Transportation Institute. Without you, this thesis would not have been possible, and I am forever grateful for the opportunities and advice you have provided me. I am proud to have you as a mentor and hope I can repay the same kindness and support forward one day. Second, I would like to express my gratitude to Dr. Sasangohar. Before taking my first ever Human Factors course with you, I never considered research or thought I would ever attend graduate school. Your class inspired me and forever changed my life. Thank you for all your support and for instilling in me the confidence to complete my thesis.

Lastly, I could not have completed my thesis without the Industrial and Systems Engineering Department faculty and staff's support. Thank you Dr. Zahabi, Dr. Currie-Gregg, and Dr. Ferris for making the courses I had with you engaging and exciting. It has been a pleasure working with you all.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a thesis committee consisting of Professor Farzan Sasangohar and Professor Maryam Zahabi of the Department of Industrial and Systems Engineering and Robert Brydia, a Senior Research Scientist at the Texas A&M Transportation Institute.

Funding Sources

Graduate study was supported by a fellowship from the Texas A&M Transportation Institute as well as the work presented in this thesis.

NOMENCLATURE

ADAS	Advanced Driving Assistance System
ADS	Automated Driving Systems
AI	Artificial Intelligence
AIC	Akaike Information Criterion
AV	Autonomous Vehicle
FAA	Federal Aviation Administration
GPWS	Ground Proximity Warning Systems
GPS	Global Positioning System
IMU	Inertial Measurement Unit
LIDAR	Light Detection and Ranging
NHTSA	National Highway Traffic Safety Administration
NTSB	National Transportation Safety Board
VESR	Vehicle Error Severity Rating
RADAR	Radio Detection and Ranging
SA	Situational Awareness
SAE	Society of Automotive Engineers
SART	Situation Awareness Rating Technique
TTI	Texas A&M Transportation Institute

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

Chapter I

INTRODUCTION

Over the last several decades, many industries have increasingly automated tasks once performed by a human operator to optimize cost, time, and efficiency. One study found that enterprises that implement automation can have cost-saving benefits of nearly $4M while potentially saving 360 work hours annually (Wald, 2017). As automation continues to advance, industry and workforce performance are rapidly changing, and utilization is expected to increase (McKinsey Global Institute, 2017). The McKinsey Institute predicts that nearly 30% of jobs will be replaced by automation by 2030 (French, Duenser, & Heathcote, 2018; McKinsey Global Institute, 2017). However, although many use automation to achieve their business endeavors, many such as aviation, medical, transportation, nuclear, and maritime industries implement automated systems to increase safety and reliability (Lee & See, 2004).

Despite the benefits that automation offers, it is often wrongly described as perfect and believed to be a fix-all solution. Instead, responsibilities are organized differently between operators and technology, creating even more complex, dynamic relationships. For instance, although operators have less manual work to complete, they are still required to monitor the system. Supervision tasks can potentially make it more difficult to spot errors, and over time, the operator can become deskilled since they do not practice their skillset as often (Bainbridge, 1983). Designers ironically still expect operators to resolve errors at a moment's notice despite their diminished ability. Furthermore, highly automated systems are still designed by humans making the system inherently flawed or biased but not error-proof (Bainbridge, 1983).

As systems continue to become more complex and or automated, the operator's trust in the system's reliability is becoming more essential for safety purposes especially in high-risk environments. Operator's trust must be calibrated to system performance otherwise operators may not rely on the system appropriately (Lee & See, 2004). If operator trust is miscalibrated, operators may distrust automation, also known as disuse, by rejecting the system when it is, in fact, reliable (Parasuraman & Riley, 1997). Additionally, operators could also over-trust automation, also known as misuse, leading to the operator being overconfident even when the technology underperforms (Parasuraman & Riley, 1997). These miscalibrated relationships between operators and automation can then have the potential to lead to accidents, some even fatal, and is a relevant problem across many domains.

For example, in 1995, the crew of the Royal Majesty cruise ship misused the ship's automated navigation system. The crew believed the ship was on course for several hours when it was not, and the ship ultimately ran aground (Lee & Sanquist, 2000; Lee & See, 2004; National Transportation Safety Board, 1997). More recently, an automated Uber test vehicle crashed with a pedestrian crossing the street while the driver was watching a show on their phone in Temple, Arizona (National Transportation Safety Board (NTSB), 2018). Parasuramen and Riley's paper (1997) also refer to examples of disuse of automation, including ignoring automated alerting systems like Ground Proximity Warning Systems (GPWS) in early aircraft designs due to their frequent false alarms. Unfortunately, this same issue exists in the medical industry. Also referred to as alarm fatigue, medical staff can become desensitized to alarms if the frequency of false alarms is high despite the alarm being accurate (Sendelbach & Funk, 2013).

Given the emergence of automated vehicles, which some say should be on the market by 2030, investigating miscalibrated trust in automated vehicles is timely and crucial to ensuring safety (Mckinsey

& Company, 2016). This requires understanding factors that can affect trust formation in automated vehicles over time as well as understanding the alignment between the levels of trust and system capability and reliability; a construct known as resolution (Cohen et al., 1999). However, while trust in automation has been investigated by many (e.g., Bisantz & Seong, 2001; Dzindolet et al., 2001; Dzindolet et al., 2003; Lee & Moray, 1992; Parasuraman & Wickens, 2008; Wiegmann, Rich, & Zhang, 2001), the natural trust formation process and resolution has not been studied longitudinally and or naturalistically for automated vehicles.

This thesis will document a naturalistic study to address this research gap by observing four safety operators' trust as they operate a Level 3 autonomous vehicle in mixed traffic over time. The primary purpose of this study is to observe how operators' trust will adapt over an 8-week period as drivers operate and supervise an autonomous shuttle. Specifically, this study aims to

1. Develop a new methodology to investigate trust-capability relationship while factoring in the effects of time and experience.
2. Identify any correlations between the operator's trust with the system's capability and performance.

For the remainder of this thesis, I will discuss further background information relating to these research aims, the methods used to explore them, and results. Specifically, in Chapter 2, I will review existing literature studying autonomous vehicles, operator trust in automation, trust formation, and other factors that affect trust. I will present the methodology implemented to meet specified research goals in Chapter 3. Finally, I will conclude by presenting the study results in Chapter 4 and 5, and a discussion further examining these findings in Chapter 6.

Chapter II

BACKGROUND

Automated Vehicles

National Highway Traffic Safety Administration (NHTSA) defines an automated vehicle (AV) as a vehicle that is capable of all driving functions in any condition, without the need for human intervention (NHTSA, n.d.). AVs are autonomous because of their dependence on AI (Artificial Intelligence) which makes decisions based on information gathered from the vehicle's multiple sensors, cameras, and location (Ondruš et al., 2020). Specifically, with LIDAR (Light Detection and Ranging) sensors, mounted on top of the vehicle, the vehicle can create a 3D, 360-degree map of its surroundings using either laser, ultraviolet, visible light, or infrared light to image objects. Obstacles are then monitored, and the distance between them and the AV is calculated through RADAR (Radio Detection and Ranging), ultrasonic sensors, and video cameras (Ondruš et al., 2020). GPS (Global Positioning System) combined with IMU (Inertial Measurement Unit) helps the vehicle determine its location and uses this information as a reference point to other surrounding objects.

However, not all autonomy is equal and exists on a continuum. The autonomy level depends on the vehicle's functionality, capability, and the human operator's responsibilities. The Society of Automotive Engineers (SAE) International's (2018) J3016 standard defines 5-levels of autonomy in autonomous vehicles, as seen in Figure 1. Level 0 in this taxonomy describes a vehicle completely dependent on the human operator, while a Level 5 vehicle is completely capable of performing all driving tasks without operator assistance.

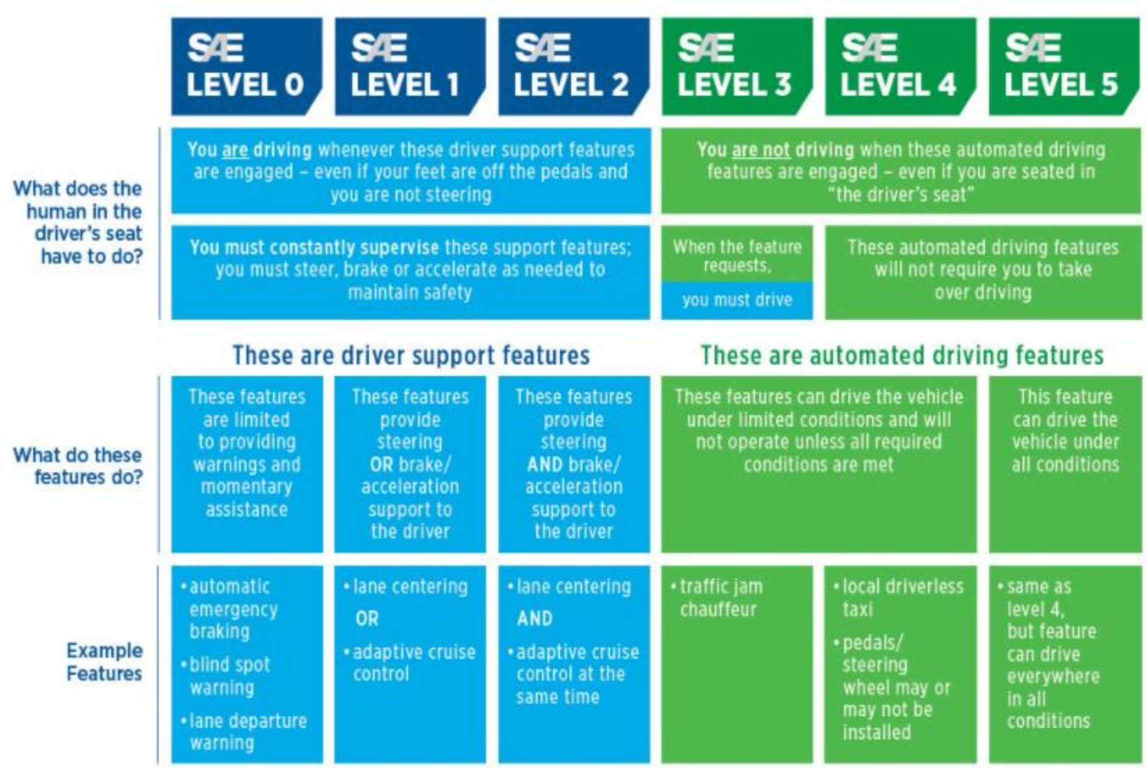

Figure 1. Levels of Autonomy (SAE, 2018)

Vehicles between Level 0 and 2 are ones that possess ADAS (Advanced Driving Assistance System) features but are not autonomous. ADAS features include but are not limited to cruise control, blind spot detection, self-park, highway autopilot, rearview systems, and so forth (NHTSA, n.d.). While ADAS features are utilized to make specific driving tasks easier, they do not remove the task of driving altogether. Instead, the human operator is still primarily responsible for driving throughout these levels. Alternatively, vehicles that are considered autonomous or have ADS (Automated Driving Systems) utilize the same technology between Level 3 and 5 as described earlier, but mostly differentiate in capability. For instance, Level 3 and 4 vehicles can only operate autonomously in certain road conditions, while Level 5 vehicles can operate autonomously in any condition at any time.

Currently, Level 0 to 2 vehicles described by SAE's taxonomy are commercially available, but Level 3 and 4 prototypes are still undergoing rigorous testing while Level 5 vehicles do not yet exist (Bertoncello & Wee, 2015; NHTSA, n.d.). Once autonomous vehicles, or Level 3 to 5 vehicles, are commercially available, they are expected to advance communities and help private vehicle owners. Benefits of automated vehicles would include reducing traffic upwards of 50 minutes per day per driver, which would decrease driver stress and greenhouse gas emissions (United States Department of Transportation (USDOT), n.d.). Additionally, because a Level 5 autonomous vehicle would not require human intervention, those unable to drive will become mobile, potentially improving the quality of life of millions who live with a disability (NHTSA, n.d.; USDOT, n.d.).

Among these benefits, though, the most crucial is the increase in safety expected from the nationwide adoption of autonomous vehicles. NHTSA reports that 94% of all motor vehicle-related deaths and injuries per year are caused by human error, supporting the common belief that automation could save many lives since the primary source of unreliability or the human operator would be replaced with automation instead (National Center for Statistics and Analysis, 2019). Achieving these benefits will depend not only on whether the technology is available but also on whether drivers choose to operate and trust autonomous vehicle technology. Trust is essential, primarily since it is often used to help operators make decisions in times of uncertainty or an incomplete understanding of the system's complexity (Kramer, 1999; Lee & See, 2004). Now that the possibility of autonomous vehicles entering the market is within reach, understanding driver trust is needed to design safe, reliable systems.

Trust in Automation

Designing a reliable system requires an appropriate level of trust (Lee & See, 2004). As autonomous vehicles emerge, trust must be examined to determine how these systems will affect driver behavior and decision making. Many researchers have claimed a relationship between trust and behavior, including how and if the automation is used by the operator (Masalonis & Parasuraman, 1999; Muir & Moray, 1996; Lee & Moray, 1992; Sheridan & Parasuraman, 2005). Nevertheless, research is limited in driver trust in automated vehicles, but many strides have been made identifying what factors can affect trust formation and the consequences if the relationship between operator trust and system performance is miscalibrated.

Currently, there are many definitions of trust in the context of automation. For example, Madsen and Gregor (2000) define trust in decision aids as "the extent to which a user is confident in, and willing to act on the basis of the recommendations, actions, and decisions of an artificially intelligent agent" (p. 1). Boon and Holmes (1991) define trust as "a state involving confident predictions about another's motives with respect to oneself in situations entailing risk" (p. 194). Meanwhile, Mayer, Davis, and Schoorman (1995) provide an organizational perspective, defining trust as "the willingness of a party to be vulnerable to the outcomes of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party" (p. 712). Most definitions discuss the trustor's vulnerability but will vary depending on the author's domain and whether they believe trust is a behavior (Meyer, 2001), attitude (Rempel et al., 1985), or intention (Mayer, Davis, & Schoorman, 1995).

In one of the most highly cited definitions of trust, which I will use for this study, by Lee and See (2004), trust is described as "the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability" (p. 51) (i.e., Hoff & Bashir, 2015; Parasuraman & Manzey, 2010; Wickens et al., 2015, etc.). However, according to Parasuraman and Riley (1997), operator judgement is not always accurate during times of uncertainty. In fact, poor judgement can often lead to misuse, disuse, and abuse, depending on the operator's trust for the system (Parasuraman & Riley, 1997). Misuse refers to the operator's overtrust in the system regardless of if the system is accurate. In contrast, disuse is whenever the operator distrusts the automated system and ignores commands even when correct. Lastly, abuse occurs when the system designer does not allocate functions to the automated system and human operator in an organized, responsible manner and can lead to misuse and disuse. In other words, designers do not consider the operator's performance and or behavior before function allocation. Instead, operator behavior results from the automated system created (Dzindolet et al., 2001).

Misuse is a very concerning behavior in driver-autonomous vehicle relationships. Some studies have demonstrated that if the driver is overtrusting the system, they can become complacent and unaware of any potential danger (Skitka, Mosier, & Burdick, 1999, 2000). There is also concern that drivers will not have enough time to react if mentally detached initially and can cause harm to those around them. Furthermore, operators will not take appropriate action due to skill degradation from constant supervising practice rather than manual control (Bainbridge, 1983; Hancock & Scallen, 1996; Parasuraman & Riley, 1997; Sheridan, 1992). For example, the first accident involving a Level 3 AV was an Uber test vehicle that crashed into a pedestrian crossing the street in Tempe, Arizona (NTSB, 2018). Rather than supervising the road, the operator trusted the vehicle enough to watch a show on their phone instead. Once the operator saw the pedestrian, there was unfortunately not enough time to respond. Other examples with similar circumstances have involved the misuse of Tesla vehicles, which are considered Level 2 AVs, in Williston, Florida (2016) and Delray, Florida (2019) (NTSB, 2016; NTSB, 2019).

On the other hand, disuse is equally as concerning as misuse. In some cases, disuse can also lead to fatal accidents, especially if the automated system is reliable and accurate. Although it is uncertain if there are any accidents involving the disuse of an autonomous vehicle, there are many other accidents involving the disuse of other automated systems. For example, nearly 5% to 10% of aviation accidents are caused by spatial disorientation, and 90% of these accidents are fatal (Federal Aviation Administration (FAA), 2011). According to the FAA, spatial disorientation is defined as "discrepancies between visual, vestibular, and proprioceptive sensory inputs [resulting] in a sensory mismatch that can produce illusions" (FAA, 2011, p. 1). Accidents often occur while the pilot is experiencing spatial disorientation, primarily since the illusions can be convincing to the point where pilots no longer trust their accurate instruments (Kallus & Tropper, 2004; Lyons et al., 2016).

While misuse and disuse are significant concerns for autonomous vehicle manufacturers and designers, it is difficult to predict either behavior. The likelihood of misuse or disuse will vary depending on the system's perceived reliability (Dzindolet et al., 2001). In a study conducted by Lee and Moray (1992), they found that system reliability affected operator trust and performance. Over three days for 2-hour sessions, they observed groups of participants as they monitored a medium-fidelity simulation of an orange juice pasteurization plant. On the first day of trials, operators experienced zero failures, while they experienced one failure on the 26th trial and a failure at each trial on the second and third day, respectively. Failure rates varied for each group (15%, 20%, 30%, 35%), but otherwise, the experimental setup was consistent for all participants.

Participants were free to utilize manual and automatic controls while supervising plant operations and could practice using both controls during the first 10 trials of each testing day. Operators were monetarily rewarded based on how much orange juice was successfully pasteurized and if the plant safely operated. After each trial, operators were then asked to complete a subjective scale to rate the system's predictability and dependability and their faith and trust in the overall system. Results showed that as operators became more experienced with the system and learned to accommodate system faults, there was an increase in trust. If they experienced a system fault, trust would initially decrease but would recover quickly. The magnitude of failure did impact trust even if the failure did not affect overall system performance.

Similarly, in Muir and Moray's (1996) work, experimental results showed a relationship between trust and reliability. In this study, operators supervised a computer-controlled milk pasteurization plant and were tasked with maximizing the output of milk within specified safety constraints. Operators could take manual or automatic control of the pump system as often and as long as needed to improve system performance. Experimental conditions varied based error magnitude (0, 5, 10, 20, 40 liters greater than inflow) and variability (zero or small variability) of the automatic pump. Each operator practiced manual and automatic controls until performance reached asymptote in terms of performance scores. Afterwards, each operator completed 8 trials for each experimental condition and then rated their trust using a subjective rating scale in the automatic pump and confidence in their trust ratings.

Results showed that manipulations of system competence in terms of error magnitude and variability impacted operator trust. Trust decreased even after the pump displayed a small amount of variability even if it did not necessarily affect overall system performance. Furthermore, as the magnitude of the error increased trust would diminish. Overall if operators trusted the system, they utilized its automated features, but they performed the task manually if they possessed distrust. These findings indicate that trust does not affect long term use but instead can vary the operator’s behavior moment to moment. These relationships are significant and illustrate that perception of system incompetence can change the behavior of those using the system.

Trust Formation Models

In Lee and See's framework, trust is considered an attitude, but within their trust formation model, they incorporate the influence of beliefs and intentions and their effects on behavior, unlike previous models (2004). This framework is grounded in the Theory of Planned Behavior (Ajzen, 1991; Fishbein & Ajzen, 1975) which posits that attitudes shape intentions and drive behavior. As shown in Figure 2, Lee and See describe trust formation as a dynamic, closed-loop process determined by context, the individual operator, and the automation, including the interface. Context includes individual, organizational, cultural, and environmental factors and their influence on an individual's beliefs, trust, intentions, and behavior.

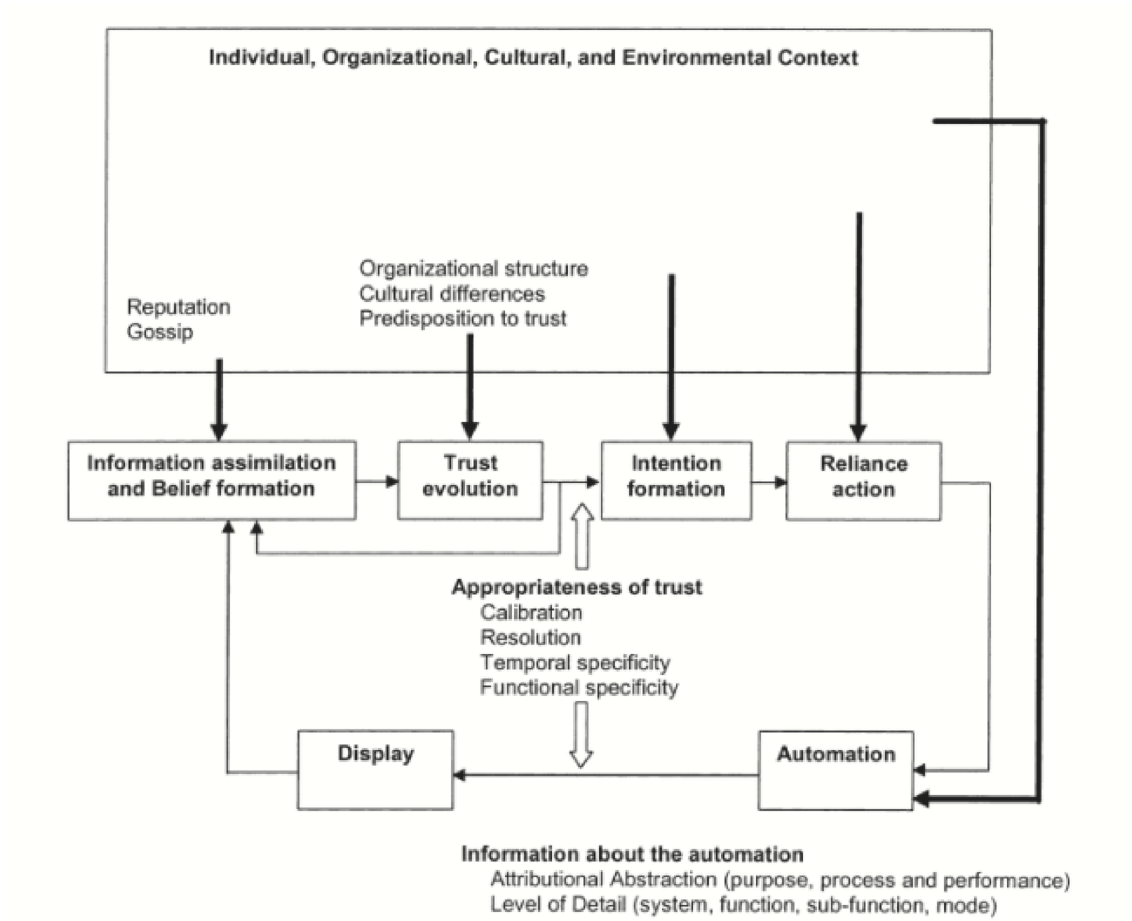

Figure 2. Lee & See's Trust Formation Model (2004)

The model also accounts for factors that could affect the automation's capability and affect the information the display provides. In turn, this would affect the operator's beliefs and, ultimately, their trust and, eventually, behavior. These factors not only determine trust but also its appropriateness for the specific system. In other words, how well a person's trust matches the system's true reliability (Lee & See, 2004). According to this framework, several metrics such as calibration, resolution, and temporal and or functional specificity, can be used to determine the operator's trust's appropriateness. Definitions of these metrics are defined in the table below.

Table 1. Metrics of Appropriateness (Lee & See, 2004)

Metrics of Appropriateness	Definition
Calibration	"The correspondence between a person's trust in the automation and the automation's capabilities" (p. 55).
Resolution	The precision of the judgment of trust differentiates the levels of automation capability.
Functional Specificity	"The differentiation of functions, subfunctions, and modes of automation" (p. 56).
Temporal Specificity	"A generic change over time as the person's trust adjusts to failures with the automation" (p. 56).

Although Lee and See's model is comprehensive, it has often been criticized for being too complicated to replicate in an experimental setting (French, Duenser, & Heathcote, 2018). In contrast, and more recently, authors Hoff and Bashir proposed their trust formation model (2015), which is more straightforward and highly valued since it is based on empirical evidence on factors that influence trust (French, Duenser, & Heathcote, 2018). In Hoff and Bashir's trust formation model (2015), trust is determined by dispositional, situational, and learned trust. These different aspects of trust account for the individual operator differences, environmental factors, and the specific automation in use.

For instance, dispositional trust entails the operator's tendency to trust the system based on predisposed individual differences such as culture, age, gender, and other unique personality traits (Hoff & Bashir, 2015). Situational trust will emerge from the situation at that specific moment in time but will vary depending on the individual operator and system. Lastly, learned trust "represents an operator's evaluations of a system drawn from past experience or the current interaction" and is determined by preexisting knowledge and system performance (Hoff & Bashir, 2015, p. 420). In a combination of dispositional, situational, and learned trust, trust is established and will determine the operator's initial reliance. However, dynamic learned trust will influence reliance as the operator continues to interact with the automation.

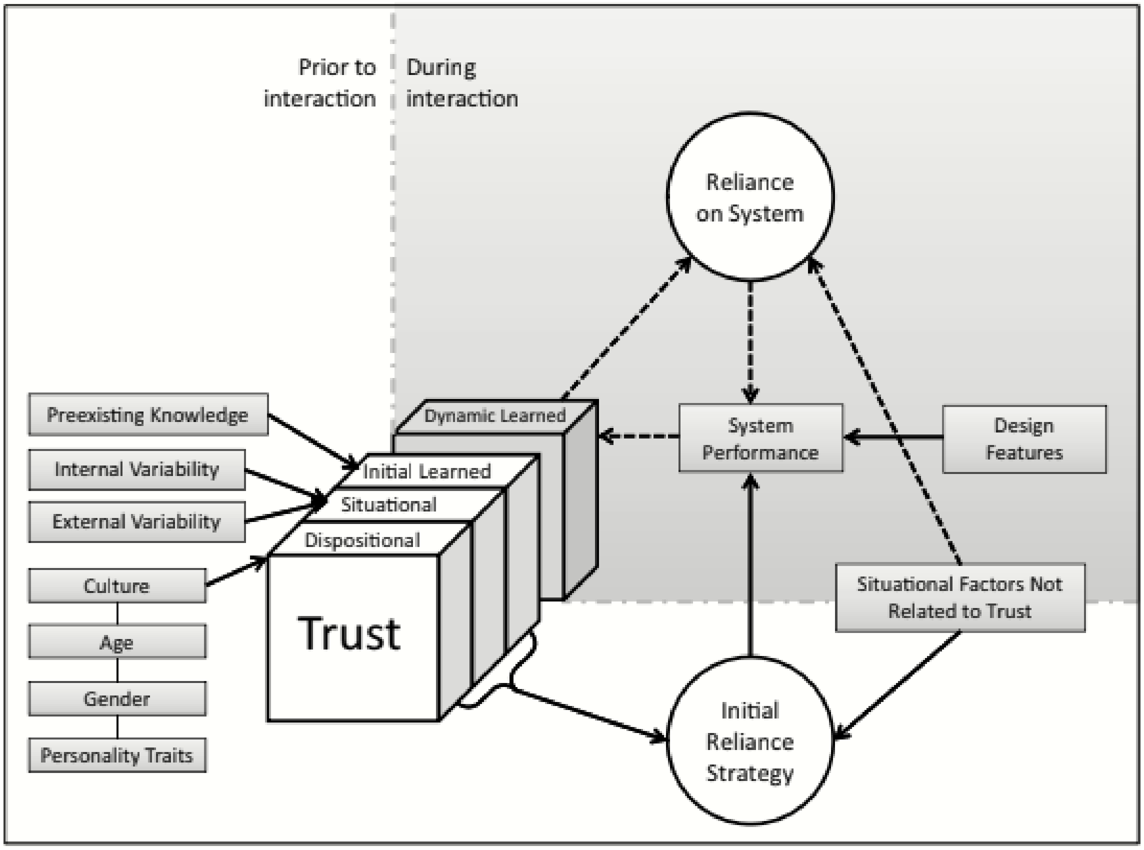

Figure 3. Hoff & Bashir's Trust Formation Model (2015)

Trust formation models serve as representations on how trust development is multifaceted. Much research has identified specific factors that influence trust, yet there are many variables that have not been observed such as time and experience. Because of the complexity of trust formation, it is important to observe the trust development process naturally using the specific system of interest.

Situational Awareness Influence on Trust in Automation

Relationships between trust in automation and situational awareness (SA) are beginning to emerge as new areas of research. The definition of situational awareness as proposed by Endsley (2000) is “the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future” (p.529). Current research suggests that increasing situational awareness can promote trust especially since autonomous vehicles could help the driver better understand their environment and predict future actions (Miller et al., 2014). Other studies have also demonstrated that even driving assistance systems that support SA will facilitate and promote trust, but more work needs to be done in this field, especially as it relates to more automated vehicles and or driving functions (Kridalukmana, Lu & Naderpour, 2020; Petersen et al., 2019).

Individual Factors Influencing Trust in Automation

Much of current research has involved identifying individual factors that influence operators' trust in automation. For example, studies show that subjects who are considered experts often do not trust decision aids as much as novices (Sanchez, Rogers, Fisk, & Rovira, 2014). This finding may be explainable since a subject matter expert is generally more confident compared to someone lacking experience. Many studies have also shown that when trust for the system exceeds confidence, automation will be utilized, but if not, the operator will manually complete the task (de Vries, Midden & Bouwhuis, 2003; Dishaw, Strong, & Bandy, 2002; Lee & Moray, 1994; Madhavan & Phillips, 2010). Other researchers have focused on other individual factors, such as specific personality traits. Findings suggested that those who possess a more positive attitude may be overly confident in the automated system's capabilities (Bailey, Scerbo, Freeman, Mikulka, & Scott, 2006; French, Duenser, & Heathcote, 2018; Merritt, Heimbaugh, LaChapell, & Lee, 2013).

From these findings, it is evident that trust is a multifaceted process that is influenced by many factors, including the human operator, the automated system, and operating conditions. Because of the complexity in which trust forms, it is necessary to observe it naturally with the specific system of interest. Because of the prevalence and lack of research surrounding human interaction with autonomous vehicles, this study's focus surrounds driver trust overtime. The following chapter will discuss the methodology implemented to capture the impact of system capability and performance onto operator trust naturally. Final chapters 4,5 and 6 will present quantitative and qualitative findings and a final discussion, respectively.

Chapter III

METHODOLOGY*

Automated Vehicle Under Investigation

The Texas A&M Transportation Institute and Texas A&M University partnered with NAVYA, an industry leader in autonomous vehicle development, to host an autonomous shuttle (see Figure 4) on campus for 12 weeks in the 2019 Fall semester in mixed traffic. During this demonstration, the vehicle was operated on campus by four student drivers and was available to both the public and student body to ride. The vehicle traveled on a 1.4-mile fixed, squared route which included two stops for passengers to board.

The vehicle's demonstration on campus was an opportunity for the student body to become exposed to innovative transportation methods and observe operators' trust with high ecological validity. To our knowledge, this is the first longitudinal study of an automated shuttle focusing on operator trust and behavior. The remaining sections will discuss the specifics of the design of this study and the methodology implemented to meet the proposed research objectives.

---

*Part of this chapter is reprinted with permission from Assessing the Development of Operator Trust in Automation: A Longitudinal Study of an Autonomous Campus Shuttle by Margaret Fowler, Farzan Sasangohar, Robert Brydia, 2020. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Volume 64, pp. 1421–1425, Copyright 2020 by Human Factors and Ergonomics Society.

Figure 4. Autonomous Shuttle (left) and Route (right) used for Study

Vehicle Design

The vehicle used in this study was NAVYA's Autonom Shuttle, which is a 4-wheel, electric test vehicle that can hold up to 15 passengers (including the safety operator) with 4 standing and 11 sitting (see Figure 5).

Figure 5. Operator Seat

The shuttle detects obstacles and tracks its location by integrating three technologies: odometry and inertial measurement unit (IMU), fixed Global Navigation Satellite System (GNSS) base with 3G/4G mobile network, and LIDAR recognition. The vehicle can be operated autonomously and manually under a certified safety operator's supervision using a hand-held controller. All information needed for travel, including the vehicle's operational status, is communicated using a dashboard as seen in Figure 6.

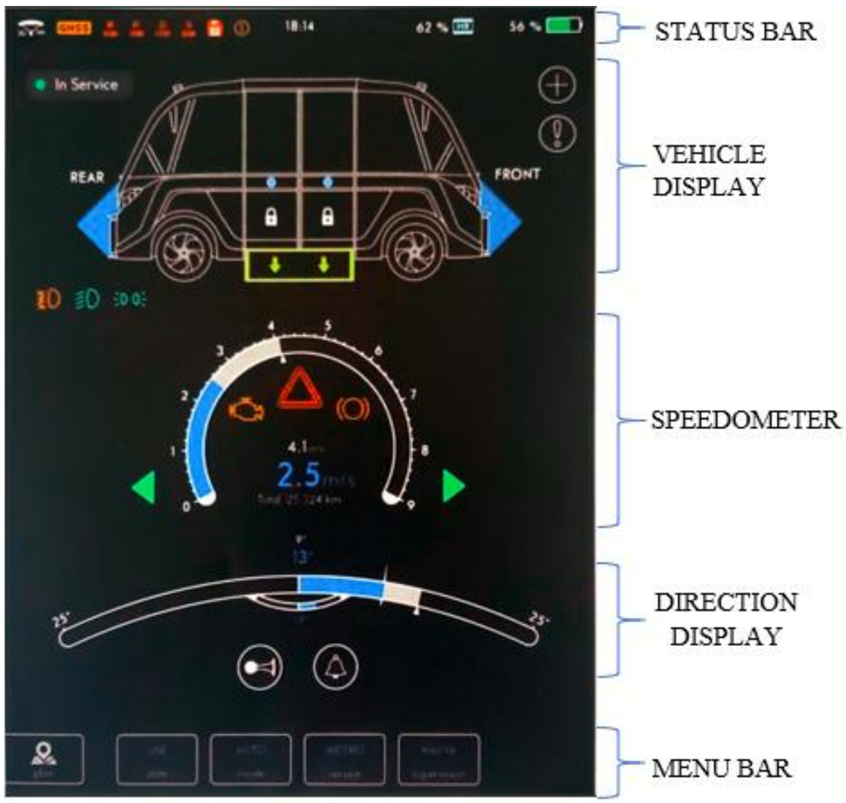

Figure 6. Vehicle Dashboard

Participants

Five safety operators were recruited from Texas A&M University's undergraduate student population (2 females, 3 males, M = 23 years old, SD = 1.09). However, Operator 5 was excluded from analysis due to missing data. Each operator was required to possess a valid driver's license, be at least 18 years of age, and be operator certified by NAVYA. The certification process entailed a week-long training course to become familiar with how to operate the shuttle manually and autonomously. The final examination included a written and practical segment to test each operator's driving skills and knowledge of the vehicle.

Study Design and Procedure

Interviews

Pre- and post-deployment interviews were conducted either on campus or at the Texas A&M Transportation Institute's Headquarters building. Interviews and data collection began after research personnel received the operator's informed consent. The initial semi-structured interview involved questions relating to the operator's previous experience with and knowledge of autonomous vehicles. Other questions focused on their trust and opinions on the vehicle's safety, reliability, and potential benefits. Similar questions were then repeated during post training and deployment interviews (see Appendix A and B).

Surveys

As the campus demonstration officially began, each operator was required to submit daily and weekly surveys tracking their situational awareness and levels of trust towards the vehicle, respectively, throughout week 4 to 12 of the deployment. Trust surveys were also administered before the deployment began and after it concluded. Situational awareness was measured throughout the deployment using the Situation Awareness Rating Technique (SART) (SART; Taylor, 1989) as seen in Appendix C. The SART survey involves rating 10 dimensions of situational awareness using a seven-point scale (1 = Low, 7 = High). The dimensions are then summed to solve for summing understanding, demand, and supply. The final situational awareness score is found by subtracting the difference of demand and supply from understanding.

Trust was recorded using one subjective scale (Madsen & Gregor, 2000). Madsen and Gregor's (2000) subjective trust scale include twenty-five positively framed questions that can be rated using a 7-point Likert scale (Strongly Agree = 3 to Strongly Disagree = -3) as seen in Appendix D. Trust scores can range from -75 (most distrustful) to 75 (most trustful). The variety of questions will score trust by evaluating the user's perceived reliability, technical competence, understandability, faith, and personal attachment regarding the system.

Finally, each operator was asked to wear a Tobii Pro Glasses 2 eye-tracker for the first hour of each scheduled shift to record and monitor their eye movements. A GoPro Hero 7 was also utilized and mounted to the inside of the shuttle to record interactions and behaviors, but the analysis of video and gaze behavior is outside the scope of this thesis and will be reported elsewhere.

Analysis

A quantitative analysis of trust surveys was conducted using R Studio version 4.0.3 (RStudio Team, 2020) and Microsoft Excel (2011) software (Microsoft Corporation, 2018). MAXQDA 12 (VERBI Software, 2018) was utilized to complete a thematic analysis of operator interviews from before the deployment started to after it commenced.

The quantitative analysis was completed for four operators from week 4 to 12 of the deployment since this was the timeframe trust surveys were collected. R Studio and Microsoft Excel were used to calculate important metrics, create visuals to display meaningful relationships and or findings and build models predicting trust. Table 2 presents the variables calculated and observed throughout the analysis.

Table 2. Variables Measured throughout Study

Variable	Definition
Vehicle Error Type	The type of error that caused the vehicle to malfunction.
Total Number of Errors	The total number of errors the vehicle experienced throughout the deployment.
Number of Errors per Operator	The total number of errors the vehicle experienced throughout the deployment per operator.

Table 2 Continued

Variable	Definition
Total Repair Time	The total amount of time required to resolve vehicle errors that occurred throughout the deployment.
Repair Time per Operator	The total amount of time required of each operator to resolve vehicle errors that occurred throughout the deployment.
Operational Time Loss	Time loss in operations due to repairs to vehicle errors.
Vehicle Error Severity Rating (VESR)	A rating describing the degree of error the vehicle experienced and repair time required during operations.
Situational Awareness	Final situational awareness score per operator collected from SART survey (Taylor, 1989).

Vehicle errors were categorized by type and ranged from software to mechanical failures (see Appendix E). Prior to finalizing the analysis, all error definitions were reviewed carefully by NAVYA and TTI (Texas A&M Transportation Institute) personnel. Furthermore, any errors that could not be classified by the research team were discussed with TTI staff to appropriately categorize.

To calculate VESR, first a codebook was developed to operationalize 5 levels of vehicle error severity categorized as low (Level 1), moderate (Levels 2&3), and high (Levels 4&5) in collaboration between research personnel and TTI staff, who were heavily involved in managing and supervising operations for the shuttle (see Table 3). Next, transcripts of operator communications with NAVYA's technical support team for each day of vehicle operation were reviewed carefully by three coders and rated by severity using the following subjective scale. Coders included project lead (Margaret Fowler) and two undergraduate research assistances part of the Dwight College of Engineering and were sponsored by the Aggie Research Scholar's Program. Both undergraduate students were also trained in qualitative data analysis prior analysis and were familiar with the project. After training, research personnel independently reviewed transcripts. Following the initial assignment of vehicle error severity ratings (1-5), there was 74% agreement between the three coders. After discussion, coders reached a consensus on the final severity ratings.

Table 3. Vehicle Error Severity Rating Criteria

<table>
<tr><th>Severity Category</th><th>Severity Rating</th><th>Definition</th></tr>
<tr><td>Low</td><td>1</td><td>Operations ran smoothly with possible minor errors that the operators were able to resolve on their own quickly.</td></tr>
<tr><td rowspan="2">Moderate</td><td>2</td><td>There were problems that required help from technical support, but the issues did not cause operations to stop and were resolved quickly.</td></tr>
<tr><td>3</td><td>Operations paused for more than a few minutes so the operator could spend more time troubleshooting the error. However, operations were ultimately not suspended.</td></tr>
<tr><td rowspan="2">High</td><td>4</td><td>There were errors on the shuttle that required significant time to resolve, and/or shut down operations for the day.</td></tr>
<tr><td>5</td><td>The shuttle was out of service for hardware and/or maintenance reasons.</td></tr>
</table>

Chapter IV

QUANITATIVE RESULTS

Overview of Operations

Vehicle Error

A total of 129 vehicle errors were recorded throughout the entire 12-week deployment, while 84 were experienced throughout the 8-week period operator trust was observed. Table 4 shows the number of errors experienced by each operator. Of the 84 vehicle errors, the four most common types of error the vehicle experienced included LIDAR Relocalization (43.86%), Unusual Vehicle Behavior (14.91%), Vehicle Stuck in Standby Mode (9.65%), and Loss of GNSS Signal (8.77%) (see Appendix E).

Table 4. Vehicle Error, Repair Time, and Operational Time Loss per Operator

OP.	Number of Vehicle Errors per Operator	Repair Time per Operator	Operational Time Loss
1	15	3.17 hours	33.17 hours
2	24	3.37 hours	21.37 hours
3	19	4.30 hours	34.30 hours
4	26	8.62 hours	26.62 hours

The time spent resolving errors during operations between all operators totaled 19.46 hours throughout the 8-week observation period. It is estimated that the vehicle lost 115.46 hours of operational time due to repair time as well. As shown in Table 4, Operator 4 experienced the most errors and the highest repair time during the study. Operator 1, meanwhile, had the least number of errors and spent the least amount of time repairing the vehicle. Of the total hours, Operator 3 had the highest loss of 33.17 hours, while Operator 2 had the least with a total of 21.37 hours.

Two one-way ANOVA tests were conducted to compare operator's repair time and number of vehicle errors experienced at an alpha level of .05. Results showed that there was not a significant difference between operator's repair time or number of vehicle errors [$F(3, 32) = 1.347$, $p_{\text{Repair Time}} = .277$; $F(3, 32) = .215$, $p_{\text{Number of Errors}} = .885$]. Figures 7 and 8 illustrate the number of vehicle errors and repair time each operator experienced per week throughout the deployment. As seen in Figure 7, most operators experienced a similar error frequency between weeks 4 and 7 and 10 and 12. However, between week 7 and 9, more variation between the operators appears. During week 8 and 9, Operators 4 and 2 experienced an unusually high number of errors, while Operator 3 experienced more errors during weeks 9-11.

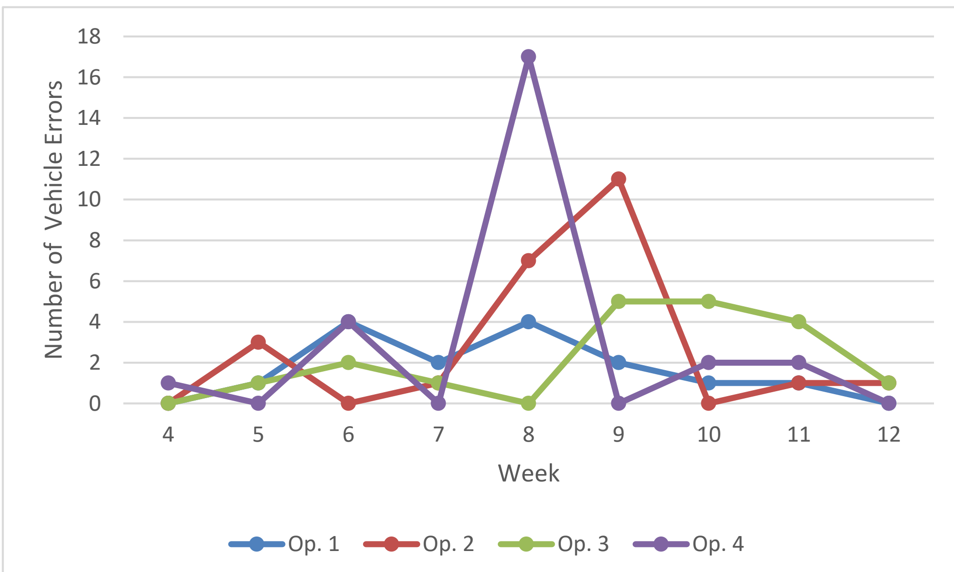

Figure 7. Weekly Vehicle Errors per Operator

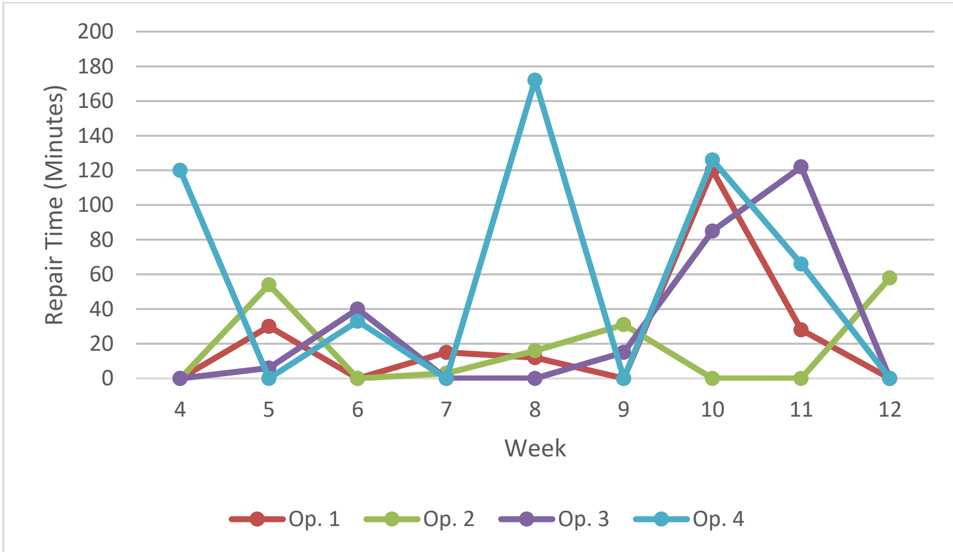

Figure 8. Weekly Repair Time per Operator

On the other hand, Figure 8 shows the time each operator spent repairing vehicle error per week. Operator 4 spent the most time resolving error throughout the deployment, while the other operators remained similar in time instead. By week 9 however, all operators except for Operator 2 were consistent with one another until the end of the deployment.

Vehicle Error Severity Rating

The average VESR for the 8-week observation period was 2.72 and the average VESR for the entire deployment was 2.58. The operators, on average, experienced errors daily they could not resolve independently and required assistance from NAVYA's technical support team. As shown in Table 5, most vehicle errors (51.81%) were rated as moderate in severity, while only 21.69% were considered low in severity. A one-way ANOVA test was used to compare operator's VESR at an alpha level of .05. Results revealed no significant differences between VESR among operators [$F(3, 31) = .730$, $p = .542$]. Two one-way ANOVAs were then conducted to determine differences between operators for moderate and high VESR. However, results did not yield any significant differences among operators [$F(3, 17) = .609$, $p_{MODERATE} = .618$; $F(3, 8) = .104$, $p_{HIGH} = .955$] .

Table 5. Percentage Breakdown of all shifts VESR by Operator

OP.	Low	Moderate	High
1	8.43%	18.07%	6.02%
2	2.41%	12.05%	6.02%
3	7.23%	12.05%	7.23%
4	3.61%	9.64%	7.23%
Total	**21.69%**	**51.81%**	**26.51%**

As seen in Table 6, Operator 4 and 2 experienced the highest percentage of shifts that possessed an VESR of low severity. Instead, Operator 3 and 1, respectively, experienced the least. Operator 1 experienced the least number of shifts (20%) with an VESR of 4 or 5, while Operator 4 experienced the highest (33%). Operator 3 and 1 experienced the highest percentage of shifts with an VESR score of 1, respectively, while Operator 2 and 4 experienced the least.

Table 6. Percentage Breakdown of Operator Shifts by VESR

OP.	Low Severity	Moderate Severity	High Severity
1	25.71%	54.29%	20.00%
2	14.29%	57.14%	28.57%
3	34.48%	37.93%	27.59%
4	14.29%	52.38%	33.33%

Figures 9 represents the average VESR each operator experienced per week throughout the deployment. VESRs remained consistent between all operators until week 11 of the deployment excluding week 8, but this was due to Operator 3's absence of work. However, during week 11, Operator 4 experienced the highest VESR compared to all operators, which continued to increase to week 12. In contrast, Operator 1 experienced the lowest VESR through week 11 and 12, while Operators 2 and 3 were comparable during this same time frame.

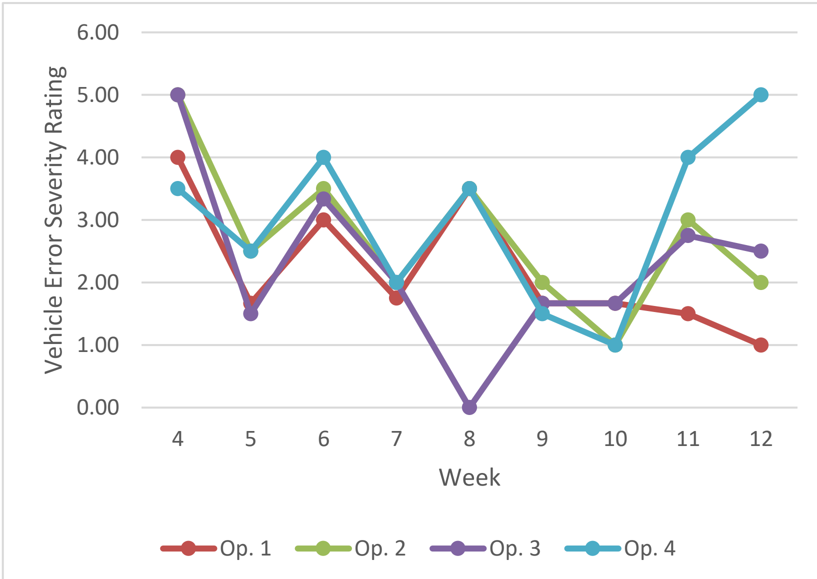

Figure 9. Average Weekly VESR per Operator

Survey Results

Operator Trust

As shown in Figure 10, each operator experienced a slight decrease in trust from week 4 to 12. Between all operators, trust averaged 4.93 throughout weeks 4-12 of the deployment (SD = 36.32, MIN = -59, MAX= 57). However, Mann-Kendall trend tests revealed there were not any significant trends among operators ($p_1 = 1$, $p_2 = .807$, $p_3 = .454$, $p_4 = .536$). Overall Operator 4 possessed the most distrust towards the vehicle during the demonstration with an average score of -42. Operator 2 and 3 trusted the vehicle the most with a score of 40.6 and 35.88, respectively. Operator 1 remained relatively neutral with an average trust score of -3.5 between week 4 and 12.

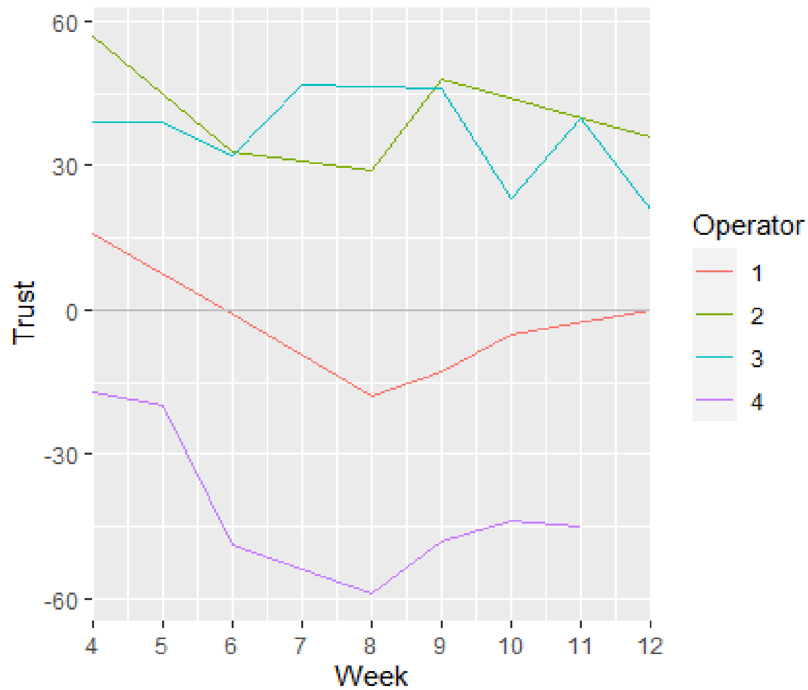

Figure 10. Operator Trust Throughout Deployment

A one-way ANOVA test was conducted to compare weekly trust scores between operators at an alpha level of .05. Test results yielded statistical significance [$F(3, 23) = 69.38$, $p = < .001$]. Post hoc comparisons using the Tukey HSD test showed significant differences between Operator 1 and 2, 3, and 4 ($M_{12} = -44.100$, $SD_{12} = 7.516$; $M_{13} = -39.375$, $SD_{13} = 6.703$; $M_{14} = 38.500$, $SD_{14} = 6.703$), between Operator 2 and 4 ($M_{24} = 82.600$, $SD_{24} = 7.076$), and Operator 3 and 4 ($M_{34} = 77.875$, $SD_{34} = 6.206$). However, Operators 2 and 3 did not significantly differ ($M_{23} = 4.725$, $SD_{23} = 7.076$).

Additionally, trust from before the deployment started to after it commenced also decreased for all operators, but the decrease was slight as seen in Figure 11 ($M_{10} = 8$, $M_{11} = -3.5$, $M_{12} = -2$; $M_{20} = 45$, $M_{21} = 40.6$, $M_{22} = 42$; $M_{30} = 37$, $M_{31} = 35.88$, $M_{32} = 30$; $M_{40} = -43$, $M_{41} = -42$, $M_{42} = -44$). The greatest change in trust was experienced by Operator 1. Their trust for the vehicle decreased by 10 points from before the deployment to after. The smallest change in trust was demonstrated by Operator 4 whose score dropped only by 1 point according to survey results. The average change for all operators from before the deployment to after was $\Delta M_{02} = -5.25$ points.

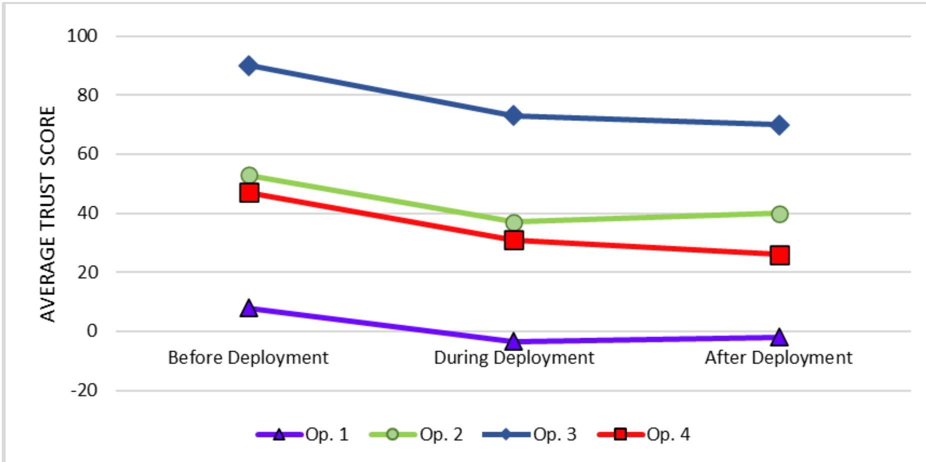

Figure 11. Average Operator Trust Before, During, and After the Deployment

Situational Awareness

Operator 1 and 4 ($M_1 = 13.83$, $M_2 = 14$) appear to possess low SA scores throughout the deployment, while Operator 2 and 3 ($M_2 = 20$, $M_3 = 33.619$) show much higher SA scores as seen in Figure 12.

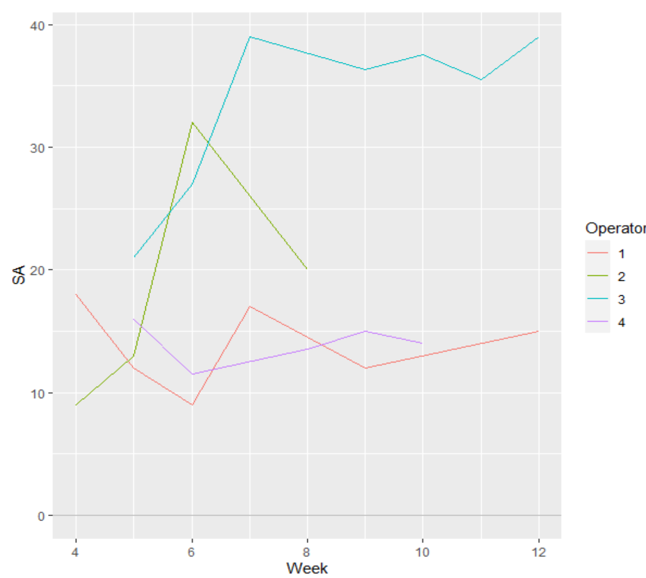

Figure 12. Operator SA Throughout Deployment

A one-way ANOVA test was conducted to compare SA scores between operators at an alpha level of .05. Test results yielded statistical significance [$F(3, 19) = 14.969$, $p = < .001$]. Post hoc comparisons using the Tukey HSD test only showed significant differences between Operator 1 and 3 (M_{13} = -19.786, SD_{13} = 3.394) and between Operator 3 and 4 (M_{34} = 19.619, SD_{34} = 3.572).

Figures 13-14 illustrate trends between each operator's trust and SA provided by SART. As shown in Figure 13, at weeks 5-8 as trust decreased, Operator 1's SA in increased. Similarly, there were some points that were inverse between Operator 2's trust and SA such as week 4-6. Figure 14 did not demonstrate any obvious relationships between trust scores and SA for Operator 3 nor Operator 4.

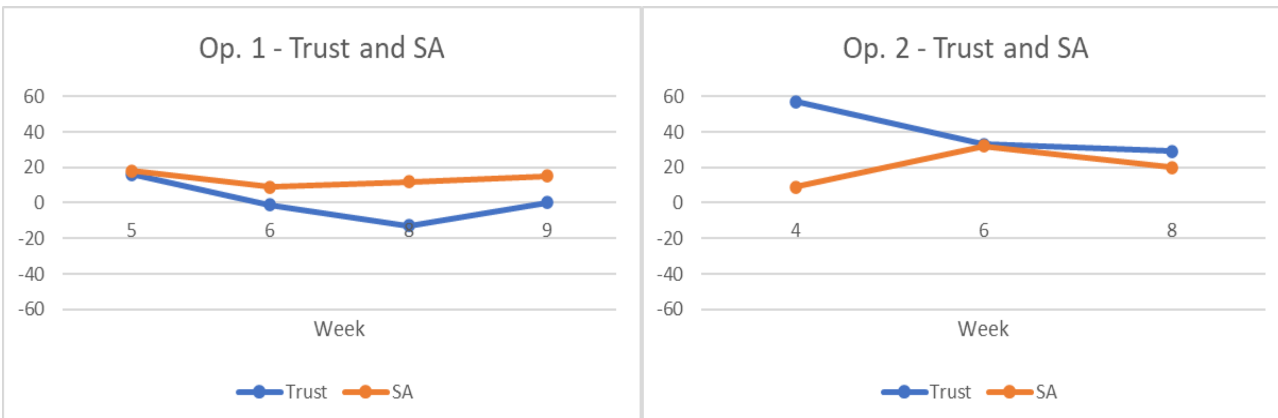

Figure 13. Trust and SA Overtime of Operator 1 and 2

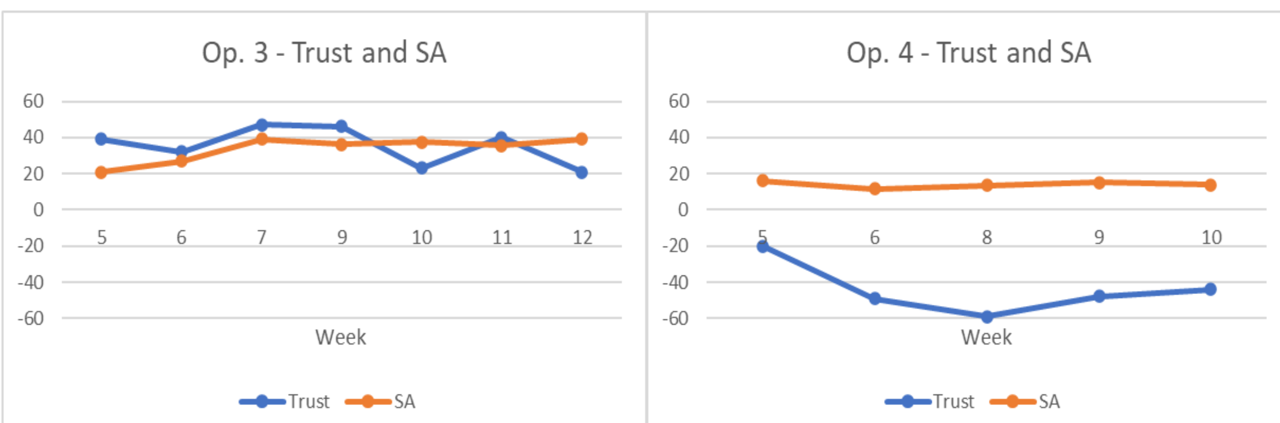

Figure 14. Trust and SA Overtime of Operator 1 and 2

Vehicle Error

Figures 15-16 show trends of operator's trust and the number of errors they individually experienced, and total recorded throughout the deployment. There are many instances where operator trust and vehicle error possess an inverse relationship from visual examination. At weeks 4-8 as trust decreased for Operator 1, as shown in Figure 15, vehicle error increased. However, as trust continued to increase from week 8-12 for Operator 1, vehicle error decreased, like results showed for Operator 3. Operator 4 also presented a similar inverse relationship as seen from week 4-6 and week 7-11 in Figure 16. Although the inverse relationship between trust and error is not as evident in Figure 16 for Operator 3, there are still weeks where this relationship exists. For example, between week 4-6, 6-7, and 10-12.

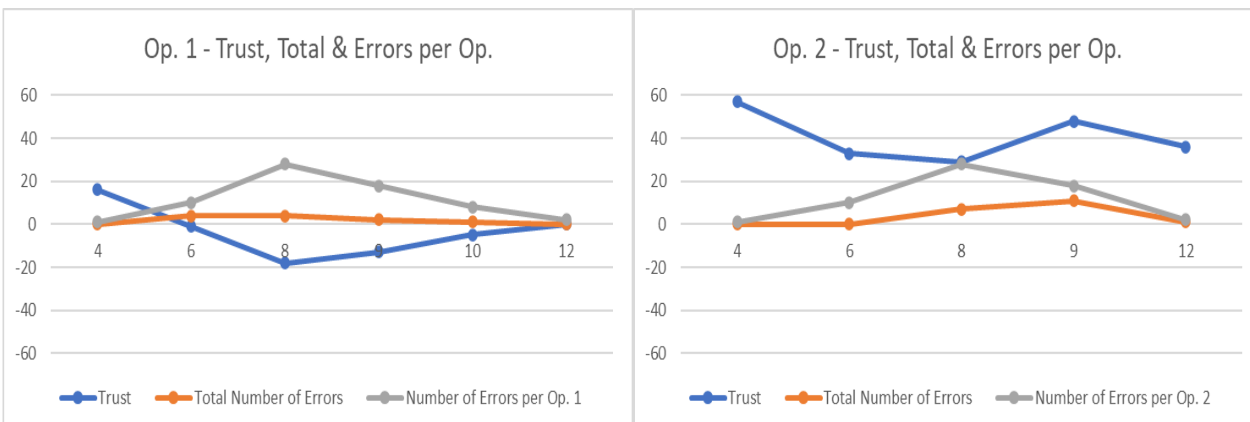

Figure 15. Trust, Total Number of Errors, and Errors per Op. of Operator 1 and 2

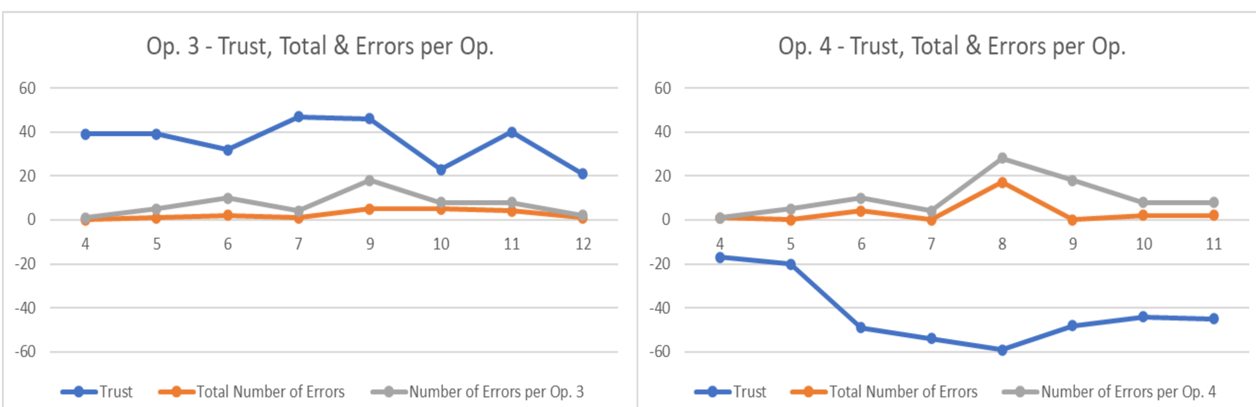

Figure 16. Trust, Total Number of Errors, and Errors per Op. of Operator 3 and 4

Operator Trust Model

The correlation matrix provided in Table 7 shows the correlations between the metrics observed throughout this study. Moderate to high correlation values are bolded and include the relationship between trust and SA (R^2 = .5839), as well as trust and repair time per operator (R^2 = -0.3084). Additionally, between repair time per operator and the number of vehicle errors per operator (R^2 = 0.5091) and between errors per operator and the total number of vehicle errors (R^2 = .5995). The coefficient of determination (R^2) was calculated to examine the proportion of the variance in the operator's individual trust scores that is predictable from the metrics of interest.

Table 7. Correlation Matrix

	Trust	*SA*	*Repair Time per Op.*	*Errors per Op.*	*VESR*	*Total Errors*	*Total Repair Time*
Trust	1.0000						
Situational Awareness (SA)	**0.5839**	1.0000					
Repair Time per Operator	**-0.3084**	0.0058	1.0000				
Number of Vehicle Errors per Operator	-0.1152	-0.0447	**0.5091**	1.0000			
VESR	0.1042	-0.2308	0.0427	0.0202	1.0000		
Total Number of Vehicle Errors	-0.2286	-0.0893	0.1157	**0.5995**	0.0338	1.0000	
Total Repair Time	-0.1859	0.0430	**0.5099**	0.1360	0.0837	0.2268	1.0000

Table 8 presents correlations per operator and shows that Operator 3 did not have any significant correlations except between their trust and total repair time throughout the deployment. All other operators had a moderate to high correlation between their trust and situational awareness and total number of errors. Operators 1 and 4 also had a moderate correlation between the number of errors they experienced individually and trust, while Operators 1 and 2 had at least moderate correlation between trust and VESR values.

Table 8. R-Squared Values per Operator

OP.	SA	VESR	Total Repair Time	Repair Time per Op.	Total Number of Errors	Number of Errors per Op.
1	***0.4819**	***.148**	0.03	0.0153	***0.7788**	***0.3974**
2	***0.6035**	***.569**	0.0818	0.039	***0.2442**	0.0004
3	0.0126	0.001	***0.1952**	0.0214	0.0836	0.0017
4	***0.423**	0.038	0.0042	0.007	***0.418**	***0.2423**
Note: Bolded values with asterisk (*) represent correlations that are moderate to high						

R Studio was then implemented to complete univariate and multivariate linear mixed effects models. Univariate models found a statistically significant relationship ($\alpha = .05$) between trust and the total number of errors experienced throughout the observation period ($p = .0098$). Other univariate tests did not find any statistically significant relationships between operator trust ($p_{SA} = .8076$, $p_{VESR} = .1868$, $p_{\text{Repair Time per Op.}} =.5594$, $p_{\text{Total Repair Time}} = .3215$, $p_{\text{Number of Errors per Op.}} = .115$).

Multivariate models considered the combination of all variables collected throughout the deployment. However, to determine the most accurate model predicting trust, all possible combinations of models of at least one variable were tested. The final three models examined possessed the lowest AIC (Akaike Information Criterion) out of all 43 possible combinations. The models with the lowest AICs are listed below in ascending order ($M_{AIC} = 198.55$, $AIC_1 = 155.82$, $AIC_2 = 158.16$, $AIC_3 = 160.52$):

(1) Trust ~ SA + Number of Errors per Op. + VESR

(2) Trust ~ SA + Total Number of Errors + VESR

(3) Trust ~ SA + VESR

Results of each model produced in R Studio are shown in Figure 17. Models 1 and 2 show a significant relationship between operator trust and number of errors per operator and total errors, respectively ($r(12) = -1.793583$, $p_{\text{Number of Errors per Op.}} = .0274$; $r(12) = -.800449$, $p_{\text{Total Number of Errors}} = .0403$) using an alpha level of .05.

MODEL 1

```
> trust_mixed <- lme(Trust ~ SA + Num_Errors + Severity ,random=~1|Operator,data=trust, na.action=na.omit)
Fixed effects:  Trust ~ SA + Num_Errors + Severity
                Value Std.Error DF    t-value p-value
(Intercept) -3.856509 23.477665 12 -0.1642629  0.8723
SA           0.013420  0.460852 12  0.0291201  0.9772
Num_Errors  -1.793583  0.714552 12 -2.5100796  0.0274
Severity     6.061875  3.678697 12  1.6478320  0.1253
```

MODEL 2

```
> trust_mixed <- lme(Trust ~ SA + Severity + Total_Errors ,random=~1|Operator,data=trust, na.action=na.omit)
Fixed effects:  Trust ~ SA + Severity + Total_Errors
                 Value Std.Error DF    t-value p-value
(Intercept)  11.092085 24.951342 12  0.4445486  0.6646
SA           -0.048630  0.472667 12 -0.1028853  0.9198
Severity      2.296941  3.651949 12  0.6289630  0.5412
Total_Errors -0.800449  0.353666 12 -2.2632895  0.0430
```

MODEL 3

```
> trust_mixed <- lme(Trust ~ SA + Severity ,random=~1|Operator,data=trust, na.action=na.omit)
Fixed effects:  Trust ~ SA + Severity
                Value Std.Error DF    t-value p-value
(Intercept) -1.946528 25.956661 13 -0.0749915  0.9414
SA           0.005898  0.538221 13  0.0109586  0.9914
Severity     3.564485  4.136498 13  0.8617156  0.4045
```

Figure 17. RStudio Results for Linear Mixed Effects Model(s) 1-3

Chapter V

QUALITATIVE RESULTS

Initial Interview

Significant themes that emerged from the first set of interviews was operator self-confidence and their faith and concerns in automation (See Table 9). Self-confidence refers to the operator's confidence in their own abilities and judgement versus the vehicles. Meanwhile, faith described the operator's belief in the shuttle or autonomous vehicles' technological ability to provide safe and reliable transportation despite their lack of experience riding in an AV and knowledge of the technology. In contrast, automation concerns included the operator's beliefs in the technology's shortcomings and doubts regarding its ability.

Table 9. Code System for Pre-Training Interviews

Theme	Subtheme	Frequency	Definition
Self-Confidence	Low Confidence	6	Anecdotes of operators concerned in their ability to operate the AV.
	High Confidence	14	Anecdotes of operators confident in their ability to operate the AV.
Faith in automation	Optimism	16	Operators expressed optimism for AV technology’s current and future abilities.
	Human Interference and Fault	5	Rather than criticize AVs, operators defend the technology and instead hold themselves, other operators, and or pedestrians responsible for shortcomings.
	Improved Performance	4	Operators make comments that AVs would improve their driving performance.

Table 9 Continued

Theme	Subtheme	Frequency	Definition
Automation Concerns	Doubt	6	Anecdotes of operators describing that AV technology may not be fully developed.
	Uncertainty	8	Operators describing their uncertainty of AV technology and their response to it.
	Preference for Manual Operation	3	Statements of the operator describing their preference for manual operation.

Self-Confidence

The table 10 presents the number of statements each operator made regarding their own self confidence in the vehicle during the initial interview and includes their trust scores prior to training. However, it should be noted that Operator 2 was not available for initial interview so was excluded from this portion of the analysis. Interestingly, the operators who were more confident in their operating skills before training and were more reluctant of the vehicle's technology were less trusting of the vehicle. Operators who were more trusting were very optimistic about the vehicle's safety and were more willing to relinquish vehicle control in times of uncertainty.

Table 10. Operator's Level of Self-Confidence in the System

Initial Trust Score (Before)	8	45	37	-43
	Op. 1	Op. 2*	Op. 3	Op. 4
High Self-Confidence	5	-	1	7
Low Self-Confidence	0	-	6	0

High Self-Confidence

Operators that possessed high self-confidence thought of themselves as highly capable in their ability to operate the vehicle and make decisions. For example, Operator 4, the most distrusting operator among the group, was the most confident in their own ability as the ultimate decision maker rather than the vehicle. Because of previous experience in the trucking driving industry, they made several mentions of being ready to take over manually ahead of time rather than waiting on the vehicle to decide. Operator 3 also made similar remarks.

"I don't need help. I'm pretty confident in my driving ability." – Op. 4

"Yeah, if someone's coming towards me, yeah, I'm taking over." – Op. 4

"I'll probably be quick to take control cause it is more, not necessarily safer, but I would be able to respond to a more diverse set of circumstances that maybe an autonomous vehicle won't." – Op. 3

Low Self-Confidence

In contrast to high-confidence, some statements of low-confidence were made during the first interview which placed more emphasis on the vehicle's ability in comparison to the operator. Unlike Operator 1 and 4, Operator 3, who was one of the most trusting operators, believed the vehicle was safe. They were more willing to allow the vehicle to make critical decisions in risky situations before using their own judgement as seen by their statements below.

"I think there are also negative talk on how unsafe it is especially because we drive our cars really fast and self-driving cars going to be the way of the future once they can work at 80 mph. And you know self-driving probably more safe than people outside the car controlling and human operated vehicles."

"I've never been in a self-driving car, operated a larger vehicle, or heavy machinery besides my own car, and…help conducting your car that can potentially hurt somebody. So, kind of like the liability, I'm not too but I'm looking forward to training like, I'm excited to see what happened, but I would be lying if I said I wasn't hesitant about what I'm doing."

Faith in automation related to the operator’s belief that the autonomous vehicle could provide safe and reliable transportation despite the operator’s lack of personal experience riding or operating an autonomous vehicle. The level of faith was compromised of multiple subthemes, including the operator’s optimism for and believe that an autonomous vehicle would improve the operator’s driving performance. Statements were also made that any accidents involving an autonomous vehicle were due to human interference and fault rather than the vehicle’s ability. According to Table 11, the most trusting operator, Operator 3, possessed the most faith in automation, while Operator 4 and 1 possessed the least, respectively.

Table 11. Operator's Level of Faith in Automation

Initial Trust Score (Before)	8	45	37	-43
	Op. 1	Op. 2*	Op. 3	Op. 4
Optimism	5	-	8	3
Human Interference and Fault	1	-	4	0
Improved Performance	1	-	3	0

Optimism

Although none of the operators had any prior personal experience with AVs before the deployment, many were optimistic about the technology. Operators 1 and 3 believed that the vehicle was not only safe and reliable but could be available on the market within the next ten years. Not only were most operators more optimistic about owning an autonomous vehicle and its safety at the beginning of the deployment but believed AVs would be ready for consumer use sooner.

In fact, all operators increased their timelines of when AVs would be ready for market from their initial answer at the first interview. Furthermore, operators reported that news coverage regarding

AVs has generally been positive and the vehicles could provide many benefits such as increased safety and mobility for different populations.

"It won't be impossible to design a vehicle that would be a hundred percent foolproof all the time." – Op. 1

"I think that if we want to make progress, we're going to have to use them and figure it out. So, I would say that they're reliable, is safe and a very interesting mode of transportation. – Op. 3

"I think a lot of people are excited about [autonomous vehicles]. Conducted research shows that they are for the most part reliable." – Op. 3

Human Interference and Fault

During the interviews, there were several instances where operators would defend autonomous vehicles whenever the topic of accidents and negative news coverage emerged during the discussion. Rather than fault the autonomous vehicle's technological capability, operators would rather blame outside interference or the vehicle operator instead. Operators also reported not completely trusting news covering AV accidents as they believed the media was generally negatively bias.

"…avoidable in the sense that maybe an operator should have been paying more attention." – Op. 1

"…probably due to outside interference that was out of not the shuttle car or autonomous car technology but of people being aware of their surroundings." – Op. 3

"…self-driving probably more safe than people outside the car controlling and it so human operated vehicles…. avoidable in the sense that maybe an operator should have been paying more attention." - Op. 3

Improved Performance

Operators also discussed that owning an autonomous vehicle would not only improve their visibility but increase their safety. Specifically, Operator 3 reported the vehicle would help them make decisions during times of uncertainty and help improve stress while driving.

"But I think we're approaching a point where I find most vehicles would be able to perform those steps better than if not magnitudes better." – Op. 1

"I think I just reflecting that the vehicle knows what it's doing. And it's like following procedures that I don't have to be like super stressed about my environment." – Op. 3

Automation Concerns

Expressed doubt, uncertainty, and manual control preference were considered general concerns for automation in the analysis. Overall, automation concerns were negative anecdotes of the technology's shortcomings and the operator's uncertainty of its ability. Table 12 provides a breakdown of statements relating to the operator's automation concerns and subthemes. From the table, Operators 1 and 4 possessed the most doubt in the system during their initial interview, while Operator 3 possessed the least. Operators 3 and 1 were also the most uncertain of the vehicle's ability compared to Operator 4. Interestingly, Operator 4 was the only operator that openly discussed their preference to drive their vehicle rather than an autonomous one.

Table 12. Operator's Concern for Automation

Initial Trust Score (Before)	8	45	37	-43
	Op. 1	Op. 2*	Op. 3	Op. 4
Doubt	3	-	1	2
Uncertainty	3	-	4	1
Preference for Manual Operation	0	-	0	3

Doubt

Operators expressed doubt concerning the shuttle's ability and its safety. It was discussed that because autonomous vehicle technology is still upcoming and not available on the market yet, operators believed this meant that autonomous vehicles were not yet deployable and understood the vehicle may have issues. However, they remained open minded to the technology especially since the vehicle was approved to operate on a college campus in mixed traffic.

"But yeah, I really only see having benefits to [autonomous vehicles] but I can see it having problems because it's just a new technology." – Op. 3

"Yeah, I would assume that right now [autonomous vehicles are] not [safe], since they're not out available to a lot of people." – Op. 4

Uncertainty

Anecdotes of the operator admitting their lack of knowledge or experience with autonomous vehicles resulted in their uncertainty in how they felt about the technology and or how they would operate the vehicle during the deployment. Operators wanted assurance they knew the vehicle was safe before considering purchasing one or letting a family or friend use the vehicle. Operator's 1 and 3 were admitted that if they were unsure whether the vehicle would stop in time for a pedestrian, they wanted to let the vehicle decide first but ultimately it depended on the situational circumstances.

"If I could be sure that [autonomous vehicles] would be safer than even me driving my own vehicle. I absolutely, I would [own an autonomous vehicle]. But I also understand that people don't like agency bans taken away from them people more, even if they know that it's safer than if they were to drive their own vehicle, even magnitude safer, they might still have some level of distress." – Op. 1

"I would probably drive it at a pretty low speed though." – Op. 3

"I would say [autonomous vehicles are safe] but I think that is also in theory assuming that I know how they run." – Op. 3

"Yeah, I think that goes back to like if, if it's, if it's safe, like if they know like, okay, we have this fully functioning car and you know, it drives itself, it doesn't really, it doesn't have any problems." – Op. 4

Preference for Manual Operation

Operator 4 was the only operator that confidently discussed their preference to manually control the vehicle during times of uncertainty. They asserted that if they saw a pedestrian coming towards the vehicle, they would not give the vehicle any time to respond first. Also whenever asked if they would own an autonomous vehicle, they reported they would still like to own and use their vehicle often.

“So that's kinda hard because I like driving my car. I have a fun part of drive and uh, I guess so we're going to have it alongside and it was like fully functioning and you know, didn't have any problems.”

“Yeah, if someone's coming towards me, yeah, I'm taking over.”

“I'm taking over. I am not going to let something bad happen.”

Focus Group

Throughout the focus group session, it was common for operators to discuss their situational trust in the vehicle. Statements relating to situational trust were mentioned 74 times throughout the focus group sessions (see Table 13). In these statements, operators described that their trust would vary depending on the specific circumstances and determine whether they would drive the vehicle manually or autonomously instead. Further analysis determined that specific circumstances often comprised of the operator's familiarity and emotional comfort level (see Figure 20).

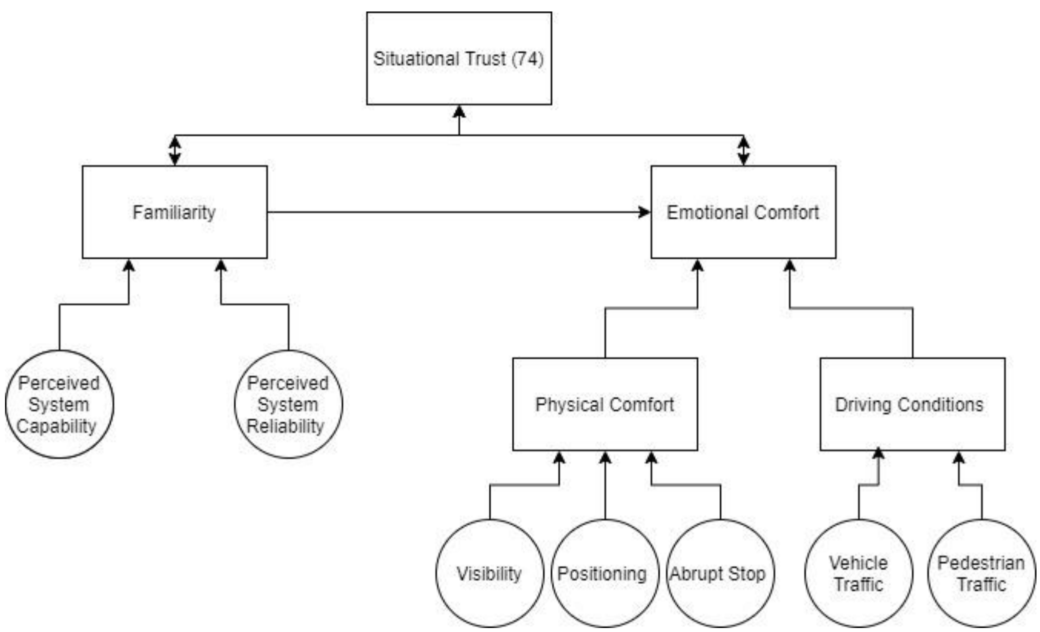

Figure 18. Breakdown of Comments Relating to Situational Trust

Table 13. Code System for Focus Group

Theme	Subtheme	Definition
Familiarity	Perceived System Capability	The operator's perception of the system's ability to make appropriate decisions.
	Perceived System Reliability	The operator's perception of the vehicle's reliability.
Emotional Comfort	Physical Comfort	The operator's physical comfort within the vehicle.
	Road Conditions	Discomfort imposed on the operator caused by specific road conditions while the vehicle was in operation.

Emotional Comfort

Operator's trust for the vehicle to decide and act independently primarily appeared to be related to whether the operator felt comfortable emotionally. Operators reported they were not comfortable emotionally if stress emerged from road conditions like high vehicle or pedestrian traffic in combination to the lack of visibility they possessed from the vehicle's physical structure.

"It was good at stopping, but in order to make it more comfortable, it was often easier to just drive manual." – Op. 1

"I think both, some of it was like me stopping it manually or the vehicle just experiencing too many errors, like getting off of it and I'm like solid ground. Yeah!" – Op. 2

Physical Comfort

Physical comfort would also influence the operator's emotional comfort level. For instance, if the operator felt they had poor visibility due to the vehicle's physical design or if the vehicle became too close to other cars and pedestrians while operating in autonomous mode, it would cause operators also to feel uncomfortable. Operators reported that due to the vehicle's design, they often thought they could not achieve a comfortable position to operate the shuttle and attend to passengers.

The driver seat often caused the operators to feel uncomfortable. Since the seat would position operators to sit facing away from the vehicle's dashboard, operators were not comfortable since they could not access critical information quickly, such as the vehicle's operating mode, location, and error messages. Operators also reported that the beveled edges and general shape of the vehicle caused them to have poor visibility.

"The driver to feel comfortable [inaudible] to the shuttle. Sometimes you would turn to face the road directly neglecting passengers, which would be one, one of our duties. So, we'd regularly encounter problems, so when it would break hard, I'm a little bit of a taller person. So, my knees immediately hit the side of the seats, that was always, and I felt like the engineers and the designers of the shuttle sacrifice a lot of things that in traditional vehicles, would make it more user friendly, for it being symmetrical or being innovative for it being, having a cooler design." – Op. 1

"I just feel, also, that the screen is a very bad blind spot for an operator." – Op. 2

"So, none of those spots were friendly to the operator because if you're sitting in the front and you face, your back would be to the road. If you are sitting on the side, then you're looking this way and like your neck hurts and then you can't sit over there and the back, because when you can't see anything in the controller, doesn't go that far. So, like, it was just tough, starting as an operator." – Op. 4

Road Conditions

Specific road conditions would also heavily influence physical comfort and ultimately emotional comfort. Operators reported that during times of heavy vehicle or pedestrian traffic, they would wait for extended periods of time before operating the vehicle or resume operations manually instead. These conditions would also cause stress especially if the vehicle were to error during times of high traffic and cause emotional discomfort.

"So that was predominantly at that pretty heavy crosswalk where like when students would get out, there'd be thoughts of students milling about, and so we would just sit there for two to three minutes, and just wait for it to subside before pushing the go button, regardless of if there were students there. So, I would, I would do the same if there was student walking in front, hit that stop button too." – Op. 3

"I had a lot less faith in the vehicle than [Operator 3] did. So, if those people on the road more times than not, then I would just stop." – Op. 4

Familiarity

However, sometimes, depending on the operator's familiarity or experience, operators would act ahead of time before the vehicle could potentially error. Based on the vehicle's previous demonstration of capability and reliability, operators would often foresee circumstances that would make them uncomfortable. Operators would use their previous experience and familiarity to forecast these inappropriate vehicle decisions to avoid discomfort altogether.

"Sudden stops, so we already knew sometimes when the stops would occur or stuff like that, so I would tell them that." – Op. 2

"I would just forecast in advance so that they're not as uncomfortable as someone who has no background knowledge about it. But I think it was just because we had driven enough circuits to know where some of those stops would be." – Op. 3

"If I didn't, I didn't trust [the vehicle] I knew I could take control of it." – Op. 4

Perceived System Capability and Reliability

Operators understood the vehicle had both limited capability and reliability but admitted that there were specific situations where the vehicle was capable and reliable enough to provide emotional comfort. Other times the vehicle was not capable or reliable enough. Still, it depended on the situational circumstances, and over time operators were able to forecast these scenarios, improving their familiarity with how the vehicle operated.

For example, although the operators were in consensus that the vehicle's stopping capability was consistent and reliable, they agreed the vehicle was not capable of always stopping appropriately and had limited ability. Often the vehicle would stop abruptly and not account for the speed of other cars. All operators expressed concern that they were less worried about the vehicle hitting an object, but instead more worried other objects would not have enough time to respond to the vehicle's stop. Operators also expressed that even if the vehicle would make the right decision to stop, sometimes the stop's abruptness was disproportionate. In other words, the vehicle would stop suddenly as if the object were in very close proximity to vehicle when, the object was not even close to the vehicle. Some objects were not even moving, like traffic cones and or tree branches.

"So, the safety was more about the people outside the vehicle and the vehicle kind of having a mind of its own." – Op. 1

"…if we were driving down around and cars had come into our lane and being a hard stop, the car drove in front of it, no matter how far ahead that vehicle was." – Op. 3

"…we're going slower than everybody else. And so, the cars that are going faster that want to get around us, they'll change the lane and merge into the lane that you're in and it would just sense that there was a car or there was something, an object in the road and it was just hard stop. And it wouldn't take into account how fast the other car was going, like we're going 15 miles an hour, they're going 18 miles an hour." - Op. 4

Additionally, the vehicle was unreliable at specific locations on the operational route. The vehicle would lose the GNSS signal at these locations, causing the vehicle to stop abruptly and experience other errors. Operators described foreseeing these scenarios and reported they were expecting these issues ahead, and overtime became confident in their ability to resolve these problems. However, there were other times where the vehicle would stop for no apparent reason according to operators causing them discomfort.

"Yeah, so there was definitely, there were definitely points where I did not feel safe in the vehicle. Cause if you got somebody that was a little more aggressive and then it would just suddenly stop, then they would run to the back of you." – Op. 3

"So, they had a station set up, they would provide signal to the shuttle so that the shuttle would know where exactly on the route it would be. Okay, so that's, if the signal went out, like the shuttle was lost and then it would just stop. So, there were instances where it would just stop randomly, like the signal would just go out randomly..." – Op. 4

Chapter VI

DISCUSSION AND CONCLUSIONS

Given the emergence of automated vehicles, investigating miscalibrated trust is crucial to designing safe and reliable systems to ensure a successful adoption and driver acceptance. If trust is miscalibrated, drivers may not appropriately rely on autonomous vehicles, leading to unsafe driving practices and potentially fatal accidents. However, to further understand the alignment between trust and system capability, trust must be observed naturally since trust formation between people and technology is complex and uncertain by many. This study aimed to close this research gap by observing four safety operators' trust over 8 weeks as they operated a Level 3 autonomous vehicle in mixed traffic. Preliminary results have shown a correlation between trust and error frequency as well as repair time. Furthermore, interviews have revealed that individual operator characteristics may drive trust, resulting in the emergence of three specific personas: Distrusting, Distrusting-Neutral, and the Trusting Operator.

Distrusting Operator

Distrusting operators possessed high self-confidence prior to beginning NAVYA's training course. Operator 4, the most distrusting operator, possessed the most self-confidence and concern for automation. Operator 4 reported they had previous truck driving experience and possessed a commercial driving license. This participant led certification courses on the weekends for others to obtain a CDL and reiterated their emphasis on highway safety and driver responsibility multiple times throughout interviews. This participant's distrust for the vehicle could be due to their years of experience in the trucking industry, allowing them to be more confident in their driving ability compared to the shuttle (de Vries, Midden & Bouwhuis, 2003; Dishaw, Strong, & Bandy, 2002; Lee & Moray, 1994; Madhavan & Phillips, 2010). Furthermore, Operator 4 also expressed that if they felt uncomfortable at any time while operating the shuttle, they would not wait on the shuttle to respond and planned to take over operations manually each time. Additionally, this operator was the only one to report their love for driving manually and their personal vehicle.

Distrusting-Neutral Operator

Operator 1 was similar to Operator 4 regarding self-confidence and their concerns for automation but was not as distrusting. Operator 1 acknowledged that autonomous vehicle technology is still immature, making it inherently flawed, and recognized that upcoming technology requires testing time and may only work in specific circumstances. However, Operator 1 was still more willing to own an autonomous vehicle and was less attached to their personal vehicle than Operator 4, especially if the vehicle could prove safer than their vehicle. They were also still confident in their ability to operate the vehicle. This participant had an engineering background which might have led to more familiarity with automated systems, including their shortcomings. The participant generally appreciated the benefits of automation more so than Operator 4 and possessed more faith in automation. Regardless, Operator 1 was also willing to take over operations in times of uncertainty but did not necessarily believe an AV was the safer option since they thought they could handle a broader range of scenarios and conditions.

Trusting Operator

Meanwhile, Operators 2 and 3 were the most trusting of operators and generally positive when discussing automated vehicles during interviews. Operator 3 did demonstrate low confidence and hesitation operating the vehicle for liability reasons before the deployment. Operator 3, however, seemed confident of the vehicle’s safety and reliability at the initial interview and did not pose as many questions about its potential shortcomings in comparison to the distrusting operators. Operator 3 also made the most statements relating to the subtheme of human interference and fault. In other words, they defended autonomous vehicle technology whenever discussing accidents the most. They would instead fault the human operator rather than the technology.

Interestingly, Operator 3 was the only operator who admitted they would wait for the vehicle's response during an uncertain or potentially unsafe moment during operations. Unfortunately, Operator 2 was not available for an initial interview. However, during the focus group session, they were very optimistic about autonomous vehicle technology. They still held the belief that autonomous vehicles could benefit the public and improve road safety.

Trust and System Performance

Operators trust generally decreased from before to after the deployment commenced. Of the four operators included in this study, both Operator 2 and 3 remained relatively trusting of the vehicle, while Operators 1 and 4 were neutral to distrusting and distrusting, respectively, according to survey results. ANOVA results comparing operators and their trust for the vehicle confirmed differences among operators excluding Operator 2 and 3.

However, although trust decreased for operators, the Mann-Kendall test did not yield any significant trends implying the decrease in trust was not substantial. This result is interesting since trust was expected to significantly change considering the number of errors and high VESRs each operator experienced throughout the deployment. The analysis revealed nearly 78% of all shifts were considered to possess moderate to high severity indicating operations ranged from operators needing help from technical support to ultimately ending operations due to high severity in vehicle errors. Additionally, operators lost about 29 hours on average of operational time throughout the eight weeks trust was observed due to vehicle error.

These results could be because trust is comprised of dispositional, situational, and learned trust, as proposed by Hoff and Bashir (2015). There are similarities between operator reports and the model proposed, such as the idea that trust is situational and learned over time based on system performance. For example, during focus group sessions, operators described that trust was, in fact, situational and depended primarily on their emotional comfort. They reported factors that would influence their emotional comfort, including their physical comfort within the vehicle and driving conditions.

High traffic conditions combined with poor visibility from the vehicle's design would yield low emotional comfort and cause the operators to drive the vehicle manually or pause operations. For example, many operators felt the vehicle's beveled edges were blind spots and would hinder their road visibility. Operators also reported feeling uncomfortable sitting in front of the vehicle's dashboard since the vehicle's information was not easily accessible and would require them to face away from both the road and passengers.

As their familiarity with the vehicle's capability and reliability increased, they could predict these situations and act ahead. Like what Hoff and Bashir describe as learned trust, operators could forecast what factors would lead to discomfort over time. Operators learned to predict error throughout specific locations in the operational route by bracing themselves or warning passengers. Preliminary findings also established a potential relationship between trust and the number of errors. Despite the small sample size, results from multivariate modeling and R-squared values reveal a correlation between error frequency and repair time, supporting the construct of situational and learned trust.

Findings from initial interviews also support Hoff and Bashir's idea of dispositional trust. Operators had prior dispositions before the deployment, such as faith and concerns in automation and their self-confidence to operate the vehicle. It was found that operators that possessed more self-confidence were less trusting in the vehicle, which is a concept that has been supported by many (de Vries, Midden & Bouwhuis, 2003; Dishaw, Strong, & Bandy, 2002; Lee & Moray, 1994; Madhavan & Phillips, 2010). Operators that were less confident or nervous about operating the vehicle were more trusting.

Trust and Situational Awareness

Unlike initial interview findings, the relationship between situational awareness and trust was unexpected. Before data collection, it was hypothesized that as trust decreased, there would be increased situational awareness. Instead, more trusting operators possessed higher situational awareness than operators with low trust, but their self-confidence may explain this result. For example, operators with high self-confidence, like Operator 1 and 4, both possessed low situational awareness throughout the deployment and vice versa for Operator 2 and 3. Operators may have had low situational awareness because they were confident in their ability to resume control of the vehicle quickly compared to operators with low self-confidence.

Overall Contribution

Results from this study provided preliminary evidence that trust can be affected by system performance. This study demonstrated an inverse relationship between trust and error frequency and repair time from statistical analysis. Additionally, qualitative results showed that trust formation might fit with Hoff and Bashir's trust formation model (2015) by revealing examples of dispositional, situational, and learned trust. Other preliminary findings have also shown a positive relationship between trust and situational awareness.

To our knowledge, this is the first longitudinal study of an automated shuttle focusing on operator trust and behavior. This study not only serves as a contribution to fill this gap but will continue to guide future researchers interested in studying trust in automated vehicles naturalistically and or longitudinally. The work performed in this study also serves as a practical contribution to vehicle manufacturers involved in the development of automated vehicles. Preliminary evidence that trust can be affected by system capability and performance should motivate industry to implement testing protocols relating to trust in automation similar to the methods used in this study to ensure drivers use these systems appropriately.

Finally, this work has contributed VESR as a new taxonomy to evaluate vehicle error severity. Because this study's primary focus was to observe the relationship between operator trust and system capability and performance, developing VESR was necessary to quantify this relationship since this type of taxonomy did not previously exist. VESR can be used in future work and serves as a new methodology to examine trust-system capability and performance relationships.

Future Work and Study Limitations

This study had several notable limitations that may affect the scope of statistical inference and generalizability of findings. The most important limitation was the sample size. Unfortunately, the sample size could not be controlled by research personnel and resulted from budget constraints from funding sources. Secondly, data was pulled from transcripts rather than collected in real-time such as error frequency and VESR. Because these metrics were collected retrospectively, it is possible some errors were not reported, or some severity ratings were inaccurate. Future work should consider allowing operators to complete VESR themselves and have operators note the number of errors they experienced in real-time. Finally, interviews and focus groups were coded by one person, and interpretations might be subject to the coder’s biases.

Based on interviews, dispositional trust played a significant factor in whether the operator would be trusting or not. Operator's support this claim based on individual differences such as self-confidence, faith in automation, and automation concerns before the deployment began. Although individual differences determined overall trust, week to week differences in trust are observed with evidence of an inverse relationship to error frequency.

The relationship between trust and error could perhaps be due to the operator's situational trust, which primarily depended on their emotional comfort and familiarity with the vehicle. There are currently no studies focusing on comfort, but preliminary findings suggest that it could be an operator's motivation to decide to trust or not. Given the small sample size, more work is needed to verify the patterns observed in this study. Future work should also consider allowing operators to assess VESR themselves in real-time rather than researchers assign scores retrospectively.

REFERENCES

Ajzen, I., & Fishbein, M. (1975). A Bayesian analysis of attribution processes. Psychological bulletin, 82(2), 261.

Ajzen, I. (1991). The theory of planned behavior. Organizational behavior and human decision processes, 50(2), 179-211.

Bailey, N. R., Scerbo, M. W., Freeman, F. G., Mikulka, P. J., & Scott, L. A. (2006). Comparison of a brain-based adaptive system and a manual adaptable system for invoking automation. Human factors, 48(4), 693-709.

Bainbridge, L. (1983). Ironies of automation. In Analysis, design and evaluation of man–machine systems (pp. 129-135). Pergamon.

Bertoncello, M., & Wee, D. (2015). Ten ways autonomous driving could redefine the automotive world. McKinsey & Company, 6.

Bisantz, A. M., & Seong, Y. (2001). Assessment of operator trust in and utilization of automated decision-aids under different framing conditions. International Journal of Industrial Ergonomics, 28(2), 85-97.

Boon, S. D., & Holmes, J. G. (1991). The dynamics of interpersonal trust: Resolving uncertainty in the face of risk. Cooperation and prosocial behavior, 190-211.

Cohen, M. S., Parasuraman, R., and Freeman, J. (1999) “Trust in Decision Aids: A Model and Its Training Implications,” Technical Report USAATCOM TR 97-D-4, Cognitive Technologies, Arlington, VA.

De Vries, P., Midden, C., & Bouwhuis, D. (2003). The effects of errors on system trust, self-confidence, and the allocation of control in route planning. International Journal of Human-Computer Studies, 58(6), 719-735.

Dishaw, M., Strong, D., & Bandy, D. B. (2002). Extending the task-technology fit model with self-efficacy constructs. AMCIS 2002 proceedings, 143.

Dzindolet, M. T., Pierce, L. G., Beck, H. P., Dawe, L. A., & Anderson, B. W. (2001). Predicting misuse and disuse of combat identification systems. Military Psychology, 13, 147–164.

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. International journal of human-computer studies, 58(6), 697-718.

Endsley, M. R., & Garland, D. J. (Eds.). (2000). Situation awareness analysis and measurement. CRC Press.

Federal Aviation Administration (FAA). (2011). Spatial Disorientation: Visual Illusions. Retrieved from https://www.faa.gov/pilots/safety/pilotsafetybrochures/media/SpatialD.pdf

Fowler, M., Sasangohar, F., Brydia, R.E. (2020). Assessing the Development of Operator Trust in Automation: A Longitudinal Study of an Autonomous Campus Shuttle. Accepted for publication in Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

French, B., Duenser, A., & Heathcote, A. (2018). Trust in Automation.

Hancock, P. A., & Scallen, S. F. (1996). The future of function allocation. Ergonomics in design, 4(4), 24-29.

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. Human Factors, 57(3), 407-434.

Jian, J., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. International Journal of Cognitive Ergonomics, 4(1), 53-71.

Kallus, K. W., & Tropper, K. (2004). Evaluation of a spatial disorientation simulator training for jet pilots. International journal of applied aviation studies, 4(1), 45-55.

Kramer, R. M. (1999). Trust and distrust in organizations: Emerging perspectives, enduring questions. Annual Review of Psychology, 50, 569–598.

Kridalukmana, R., Lu, H. Y., & Naderpour, M. (2020). A supportive situation awareness model for human-autonomy teaming in collaborative driving. Theoretical Issues in Ergonomics Science, 21(6), 658-683.

Lee, J., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. Ergonomics, 35(10), 1243-1270.

Lee, J. D., & Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. International Journal of Human-Computer Studies, 40(1), 153-184.

Lee, J. D., & Sanquist, T. F. (2000). Augmenting the operator function model with cognitive operations: Assessing the cognitive demands of technological innovation in ship navigation. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 30(3), 273-285.

Lee, J. D., & See, K. A. (2004). Trust in technology: Designing for appropriate reliance. Human Factors, 46(1), 50-80.

Lyons, J. B., Ho, N. T., Koltai, K. S., Masequesmay, G., Skoog, M., Cacanindin, A., & Johnson, W. W. (2016). Trust-based analysis of an Air Force collision avoidance system. Ergonomics in design, 24(1), 9-12.

Madhavan, P., & Phillips, R. R. (2010). Effects of computer self-efficacy and system reliability on user interaction with decision support systems. Computers in Human Behavior, 26(2), 199-204.

Madsen, M., & Gregor, S. (2000). Measuring human-computer trust. Australasian Conference on Information Systems (ACIS) Proceedings, 53, 6-8.

Masalonis, A. J., & Parasuraman, R. (1999, September). Trust as a construct for evaluation of automated aids: Past and future theory and research. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 43, No. 3, pp. 184-187). Sage CA: Los Angeles, CA: SAGE Publications.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. Academy of management review, 20(3), 709-734.

Mayer, A. (2008). The manipulation of user expectancies: Effects on reliance, compliance, and trust using an automated system (Unpublished master's thesis). Georgia Institute of Technology, Atlanta.

Mayer, A. K., Sanchez, J., Fisk, A. D., & Rogers, W. A. (2006). Don't let me down: The role of operator expectations in human-automation interaction. Proceedings of the Human Factors and Ergonomics Society, 50, 2345-2349.

McKinsey & Company (2016, January). Automotive revolution – perspective towards 2030 How the convergence of disruptive technology-driven trends could transform the auto industry. Retrieved from https://www.mckinsey.com/~/media/mckinsey/industries/automotive%20and%20 assembly/our%20insights/disruptive%20trends%20that%20will%20transform%20the%20auto %20industry/auto%202030%20report%20jan%202016.pdf

McKinsey Global Institute (2017, December). Jobs lost, jobs gained: Workforce transitions in a time of automation. Retrieved from https://www.mckinsey.com/~/media/McKinsey/Global%20Themes/Future%20of%20Organizations/What%20the%20future%20of%20work%20will%20mean%20for%20jobs%20skills%20and%20wages/MGI-Jobs-Lost-Jobs-Gained-Report-December-6-2017.ashx

Merritt, S. M., Heimbaugh, H., LaChapell, J., & Lee, D. (2013). I trust it, but I don’t know why: Effects of implicit attitudes toward automation on trust in an automated system. Human factors, 55(3), 520-534.

Meyer, J. (2001). Effects of warning validity and proximity on responses to warnings. Human Factors, 43, 563–572.

Microsoft Corporation. (2018). Microsoft Excel. Retrieved from https://office.microsoft.com/excel

Miller, Dave, Annabel Sun, and Wendy Ju. "Situation awareness with different levels of automation." 2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE, 2014.

Muir, B. M., & Moray, N. (1996). Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. Ergonomics, 39(3), 429-460.

National Center for Statistics and Analysis. (2019). Distracted driving in fatal crashes, 2017. Washington, DC: National Highway Traffic Safety Administration.

National Highway Traffic Safety Administration. (n.d.). Automated Vehicles for Safety. Retrieved from https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety

National Transportation Safety Board. (1997). Marine accident report – Grounding of the Panamanian passenger ship Royal Majesty on Rose and Crown Shoal near Nantucket, Massachusetts, June 10, 1995 (NTSB/MAR97/01). Washington, DC: Author.

National Transportation Safety Board. (2016, July 16). Preliminary Report Highway: HWY16FH018 (NTSB No. HWY16FH018-prelim). Retrieved from https://www.ntsb.gov/investigations/Accident Reports/Pages/HWY16FH018-preliminary.aspx

National Transportation Safety Board. (2018, March 24). PRELIMINARY REPORT HIGHWAY HWY18MH010. Retrieved from https://www.ntsb.gov/investigations/AccidentReports/Reports/HWY18MH010-prelim.pdf

National Transportation Safety Board. (2019, May 16). PRELIMINARY REPORT HIGHWAY HWY19FH008. Retrieved from https://www.ntsb.gov/investigations/AccidentReports/Pages/HWY19FH008-preliminary-report.aspx

Ondruš, J., Kolla, E., Vertaľ, P., & Šarić, Ž. (2020). How Do Autonomous Cars Work?. Transportation Research Procedia, 44, 226-233.

Petersen, L., Robert, L., Yang, J., & Tilbury, D. (2019). Situational awareness, driver's trust in automated driving systems and secondary task performance. SAE International Journal of Connected and Autonomous Vehicles, Forthcoming.

Rempel, J. K., Holmes, J. G., & Zanna, M. P. (1985). Trust in close relationships. Journal of Personality and Social Psychology, 49(1), 95–112.

RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.

Society of Automotive Engineers (SAE) International. (2014). Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles.

Sanchez, J., Rogers, W. A., Fisk, A. D., & Rovira, E. (2014). Understanding reliance on automation: effects of error type, error distribution, age and experience. Theoretical issues in ergonomics science, 15(2), 134-160.

Sendelbach, S., & Funk, M. (2013). Alarm fatigue: a patient safety concern. AACN advanced critical care, 24(4), 378-386.

Sheridan, T. B. (1992). Telerobotics, automation, and human supervisory control. MIT press.

Sheridan, T. B., & Parasuraman, R. (2005). Human-Automation Interaction. Reviews of Human Factors and Ergonomics, 1(1), 89–129. https://doi.org/10.1518/155723405783703082

Skitka, L. J., Mosier, K. L., & Burdick, M. (1999). Does automation bias decision-making?. International Journal of Human-Computer Studies, 51(5), 991-1006.

Skitka, L. J., Mosier, K., & Burdick, M. D. (2000). Accountability and automation bias. International Journal of Human-Computer Studies, 52(4), 701-717.

United States Department of Transportation. (n.d.). CV Basic Facts. Retrieved from https://www.its.dot.gov/cv_basics/cv_basics_facts.htm#fact3.

Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. Human factors, 52(3), 381-410.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. Human Factors, 39(2), 230-253.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. IEEE Transactions on Systems, Man, and CYBERNETICS-Part A: Systems and Humans, 30(3), 286-297.

Parasuraman, R., & Wickens, C. D. (2008). Humans: Still vital after all these years of automation. Human Factors, 50(3), 511-520.

Taylor, R.M. (1989, October). Situational awareness rating technique (SART): The development of a tool for aircrew systems design. In North Atlantic Treaty Organization, Proceedings of the NATO Advisory Group for Aerospace Research and Development (AGARD) Situational Awareness in Aerospace Operations Symposium (AGARD-CP- 478) (p. 17). Loughton, Essex, UK: AGARD. Retrieved from https://apps.dtic.mil/dtic/tr/fulltext /u2/a223939.pdf

Wald, J. (2017). How Automation Could Save Your Business $4 Million Annually. Forbes. Retrieved from https://www.forbes.com/sites/waldleventhal/2017/08/03/how-automation-could-save-your-business-4-million-annually/?sh=318f94b33807

Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2015). Engineering psychology and human performance. Psychology Press.

Wiegmann, D. A., Rich, A., & Zhang, H. (2001). Automated diagnostic aids: The effect of aid reliability on users' trust and reliance. Theoretical Issues in Ergonomics Science, 2(4), 352.

VERBI Software. (2018). MAXQDA 12 [computer software]. Berlin, Germany: VERBI Software.

APPENDIX A

- Have you ever ridden in an autonomous vehicle?
- In general, what are your thoughts on autonomous vehicles?
- What are things you have heard about autonomous vehicles from other people or news sources? Is the news mostly positive or negative, and what is your opinion?
- When do you think autonomous vehicles will be available on the market for consumer use?
- Would you recommend letting a loved one or friend use or ride an autonomous vehicle?
- What are your feelings or thoughts about operating an autonomous vehicle?
- If traffic were high while operating the vehicle and you saw a cyclist and or car approaching fast, would you allow the AV to respond to avoid a potential collision? Or do you think you would take over manually?
- Do you have any thoughts on NAVYA's training course? Expectations?

APPENDIX B

1. General Feelings about Autonomous Vehicles

- Did you feel like you were knowledgeable about the technology?
- Did you think it was safe?
- What made you think it was safe or unsafe?
- Did you think it was safer in certain conditions? Which ones?
- Has your opinion changed since the demonstration?

2. General feelings about the Smart Shuttle demonstration

- Did the vehicle operate the way you expected from the training?
- Was there anything unexpected that occurred?
- Did you feel like you could easily take control, when needed?
- Were you concerned about the technology? What were those concerns?
- Did you feel safe as an operator? Do you think the passengers felt safe?
- What were some of the most common comments you heard about the demonstration? Do you agree or disagree?
- Did you hear anything about the Smart Shuttle demonstration from people other than passengers? What was the overall sentiment?
- What were your general feelings about the training?
- Did you feel confident about your abilities to operate the Smart Shuttle after training?
- Did this change over the deployment period?
- What ways could the training be improved?
- What are your thoughts about shifts and scheduling? Were the shifts too long? Did you feel like you lost focus during a shift?

3. General feelings about the Smart Shuttle vehicle

- Did you consistently feel safe while operating the shuttle? If not, describe a specific time that you did not feel safe.
- What were some of the common problems that you experienced operating the shuttle?
- How reliable to do you think the Smart Shuttle is? Speak specifically about components of reliability (travel time, ability to operate autonomously)
- Tell me about your experiences with troubleshooting errors with the Smart Shuttle. Were you able to do that on your own? How did the NAVYA representatives help?
- What concerns do you have about the NAVYA vehicles? How can these concerns be addressed?

4. Thoughts about future deployments

- Was that too long? Too short?
- How could this demonstration have been improved?
- If there were to be another demonstration, what could NAVYA, or Texas A&M do to encourage more ridership on the Smart Shuttle?
- Do you think a different route might have more ridership? If so, which route?

5. Trust of technology

- Do you believe your trust in the vehicle's technology has changed since the beginning of the deployment? If so, how?
- What aspects of the vehicle's technology do you believe effected your trust/perception the most and why?
- Can you describe a time where the vehicle violated your trust? If so, how? Did you operate the vehicle differently after this incident?
- What year do you think autonomous vehicles would be ready for consumer use, did your estimate change from the start of the deployment to now?
- Would you still purchase an autonomous vehicle if it were on the market and you could afford it?

APPENDIX C

SITUATION AWARENESS RATING TECHNIQUE (SART; Taylor, 1990)

Instability of Situation
How changeable is the situation? Is the situation highly unstable and likely to change suddenly (High) or is it very stable and straightforward (Low)?

1 2 3 4 5 6 7

Complexity of Situation
How complicated is the situation? Is it complex with many interrelated components (High) or is it simple and straightforward (Low)?

1 2 3 4 5 6 7

Variability of Situation
How many variables are changing within the situation? Are there a large number of factors varying (High) or are there very few variables changing (Low)?

1 2 3 4 5 6 7

Arousal
How aroused are you in the situation? Are you alert and ready for activity (High) or do you have a low degree of alertness (Low)?

1 2 3 4 5 6 7

Concentration of Attention
How much are you concentrating on the situation? Are you concentrating on many aspects of the situation (High) or focussed on only one (Low)?

1 2 3 4 5 6 7

Division of Attention
How much is your attention divided in the situation? Are you concentrating on many aspects of the situation (High) or focussed on only one (Low)?

1 2 3 4 5 6 7

Spare Mental Capacity
How much mental capacity do you have to spare in the situation? Do you have sufficient to attend to many variables (High) or nothing to spare at all (Low)?

1 2 3 4 5 6 7

Information Quantity
How much information have you gained about the situation? Have you received and understood a great deal of knowledge (High) or very little (Low)?

1 2 3 4 5 6 7

Familiarity with Situation
How familiar are you with the situation? Do you have a great deal of relevant experience (High) or is it a new situation (Low)?

1 2 3 4 5 6 7

APPENDIX D

1. ***Perceived Reliability***
   R1 - The system always provides the advice I require to make my decision.
   R2 - The system performs reliably.
   R3 - The system responds the same way under the same conditions at different times.
   R4 - I can rely on the system to function properly.
   R5 - The system analyzes problems consistently.

2. ***Perceived Technical Competence***
   T1 - The system uses appropriate methods to reach decisions.
   T2 - The system has sound knowledge about this type of problem built into it.
   T3 - The advice the system produces is as good as that which a highly competent person could produce.
   T4 - The system correctly uses the information I enter.
   T5 - The system makes use of all the knowledge and information available to it to produce its solution to the problem.

3. ***Perceived Understandability***
   U1 - I know what will happen the next time I use the system because I understand how it behaves.
   U2 - I understand how the system will assist me with decisions I have to make.
   U3 - Although I may not know exactly how the system works, I know how to use it to make decisions about the problem.

   U4 - It is easy to follow what the system does.
   U5 - I recognize what I should do to get the advice I need from the system the next time I use it.

4. ***Faith***
   F1 - I believe advice from the system even when I don't know for certain that it is correct.
   F2 - When I am uncertain about a decision I believe the system rather than myself.
   F3 - If I am not sure about a decision, I have faith that the system will provide the best solution.
   F4 - When the system gives unusual advice I am confident that the advice is correct.
   F5 - Even if I have no reason to expect the system will be able to solve a difficult problem, I still feel certain that it will.

5. ***Personal Attachment***
   P1 - I would feel a sense of loss if the system was unavailable and I could no longer use it.
   P2 - I feel a sense of attachment to using the system.
   P3 - I find the system suitable to my style of decision making.
   P4 - I like using the system for decision making.
   P5 - I have a personal preference for making decisions with the system.

APPENDIX E

Error Type	% Total Vehicle Error	Definition
LIDAR Relocalization Error	43.86%	LIDAR sensors are unable to detect vehicle's position.
Unusual Vehicle Behavior	14.91%	The vehicle's behavior could not be categorized by NAVYA or the operator.
Vehicle Stuck in Standby Mode	9.65%	The vehicle would not allow the operator to switch from 'standby' to 'use' mode or the operator forgot to switch to 'use' mode.
Loss of GNSS Signal	8.77%	The vehicle loss GNSS signal while in operation.
LIDAR Malfunction	3.51%	LIDAR sensors malfunction due to weather conditions or other external factors.
Problems with Suspension	3.51%	Operator had problems with changing the mode of the suspension for towing or regular operations.
Construction	2.63%	Construction occurring on the route results in the suspension of operations. The vehicle is not programmed to navigate around the construction work.
Wrong Operational Mode	1.75%	The operator would forget to switch the vehicle from manual to autonomous mode.
Overheating	1.75%	The vehicle's battery would overheat causing it to turn off suddenly or malfunction.
Wrong Passenger Mode	1.75%	Operator had vehicle in metro mode instead of demand mode.
PC Offline	0.88%	One of the PCs in the shuttle was offline so it could not operate

Wheel Sensor Malfunction	0.88%	Problems with the wheel sensor(s) were causing the vehicle to have hard stops or not operate smoothly.
NAVYA Supervision	0.88%	NAVYA could not provide sufficient technical support.
NAVYA Supervision Connection	0.88%	NAVYA Supervision lost connection to the vehicle while operating.
Vehicle Stuck in Autonomous Mode	0.88%	Vehicle could not be switched to manual mode.
Problem with Motor	0.88%	Humming coming from motor compartment.
Problem with Doors Closing	0.88%	The doors to the vehicle would not close while attempting to tow the vehicle.
Wheel Damage	0.88%	The wheel of the vehicle suffered physical damage.
Loss of Ethernet Connection	0.88%	Ethernet losses connection to vehicle.