MEASURING THE INFLUENCE OF AUTOMATION ON SITUATION AWARENESS IN HIGHLY AUTOMATED VEHICLES

A Thesis
Presented to
The Academic Faculty

by

Zoe M. Becerra

In Partial Fulfillment
of the Requirements for the Degree
Masters of Science in Psychology in the
School of Psychology

Georgia Institute of Technology
May of 2020

MEASURING THE INFLUENCE OF AUTOMATION ON SITUATION AWARENESS IN HIGHLY AUTOMATED VEHICLES

Approved by:

Dr. Bruce N. Walker, Advisor
School of Psychology
Georgia Institute of Technology

Dr. Jamie Gorman
School of Psychology
Georgia Institute of Technology

Dr. Richard Catrambone
School of Psychology
Georgia Institute of Technology

Date Approved: April 17, 2020

ACKNOWLEDGEMENTS

To my lab mates, thank you for supporting me and answering my endless questions. Thank you so much to family and friends who continued to encourage me when challenges arose. Most importantly, to my partner, thank you for your continued support and confidence if my ability to accomplish my goals.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Higher levels of automation, such as adaptive cruise control (ACC) and automated lane keeping (ALK), are becoming more and more common in vehicles. With the inclusion of these automated features, the role of the driver is shifting from an active, operator role to a passive, supervisory role. As drivers enter this transition, it is critical they understand how the automation is performing and remain aware of the roadway environment. Situation awareness (SA) is the understanding of what is going on around you. Previous research has shown how a driver's SA is impacted by many factors including: age, driving experience, distraction, and secondary task engagement. Little work has explored the direct influence of level of automation on SA or how best to measure SA in an automated vehicle. To address these issues, this study examined how SA changes as a function of level of automation in the driving domain using three measures of SA. Participants completed two twenty-minute simulated drives with two levels of automation: low automation (ALK only); and high automation (ALK and ACC). The order of the drives were counterbalanced. Throughout the drives, the Situation Present Assessment Method (SPAM) and secondary task engagement were used to measure SA. SPAM is a query-based measure in which questions about the situation are periodically presented; the situation remains present and the participant continues to perform the task. Secondary task engagement was measured by the total time voluntarily spent playing a game of Tetris, a visuospatial task. After each drive, participants completed the SART questionnaire to subjectively measure their perceived SA. Additionally, the NASA-TLX and a Trust in Automation Scale were administered after each drive to measure subjective workload and trust. Results showed between the three administered measures of SA, query-based measures (SPAM) and subjective measures (SART) were more sensitive compared to performance measures (secondary task engagement). Further, there was evidence to suggest a combination of query-based and subjective measures is best to assess SA in the automated driving context. Concerning the impact of automation level on SA, high automation systems supported higher SA compared to low automation systems. The results also indicated the patterns of SA were different in the low and high automation drives. There were no significant changes in the pattern of SA during the low automation drive. However, the results suggested a quadratic trend best described the pattern of SA in the high automation drive. These insights will provide guidance to develop better standardized measures of SA for future research. In addition, these findings can inform the design of interventions to support driver SA, especially in low automated vehicles.

CHAPTER 1. INTRODUCTION

In 2017, ninety-four percent of all serious vehicle crashes involved either impaired driving, distracted driving, speeding, or other dangerous, illegal driving maneuvers (US Department of Transportation, 2018). As automation features continue to become more common in vehicles, it is important to consider the unintended effects these systems may have on safety. Operating a vehicle, manual or automated, is a dynamic task that requires the operator to remain engaged and aware of their environment. Further, different levels of automation changes what tasks the human is required to do and attend to. Current research is investigating how to improve take-over request performance, create understandable displays, and increase situation awareness in automated vehicles (Kunze, Summerskill, Marshall, & Filtness, 2019; May, Noah, & Walker, 2017; Noah, 2018; Noah & Walker, 2017; Wintersberger, Sawitzky, Frison, & Riener, 2017). However, little is known about the relationship between levels of automation and situation awareness. This study begins to explore this relationship to determine the impact trust, workload, and level of automation have on a driver's knowledge of the roadway environment.

1.1 Automation

Automation is a system that performs tasks a human previously completed; it is typically used to improve safety of the operator, increase performance, and decrease operator workload (Parasuraman, Sheridan, & Wickens, 2004). There are different levels of automation based on how tasks are allocated to the automated system and the operator. The more tasks allocated to an automated system instead of the human operator, the higher the automation level. In a low automated system, the human operator will have practically all of the tasks, whereas a highly automated system will have the majority of the tasks and the human will have very few responsibilities.

1.1.1 Operating Automated Vehicles

Society of Automotive Engineers (SAE) International (2018) has defined six levels of vehicle automation to distinguish manual vehicles (Level 0), semi-automated vehicles (Level 2), and fully automated vehicles (Level 5) (see APPENDIX A). Levels 0 to 2 describe low automation vehicles; the person in the driver's seat is still considered the driver and must constantly supervise the automated system *and* the roadway environment. While Level 0 captures manual vehicles, they may contain support features such as blind spot warning systems. Vehicles containing an adaptive cruise control (ACC) *or* automated lane keeping (ALK) system is considered a Level 1 automated vehicle, but if it contains both systems it is classified as Level 2. High automation vehicles (Levels 3 to 5) do not consider the person in the driver's seat to be the driver. However, Level 3 automated vehicles can still request the operator to take over the driving task. Fully automated vehicles (Level 5) will be able to operate under all conditions and will not require the driver to take over control. For the purposes of this study, simulations of Level 1 and Level 3 automated vehicles will be used.

As automation levels increase, the way drivers interact with the system and driving task changes (Endsley, 2017b). Drivers shift from an active role in manual driving (Level 0) to a passive role as automation levels increase (Endsley, 2012; Endsley, 1996). While the driver is being supported by these systems, it creates a new supervisory control task for drivers. Research has shown humans are not particularly successful performing these tasks, such as monitoring system performance (Sheridan, 2012). Supervisory control tasks essentially become vigilance tasks where arousal levels decrease. When operators are asked to complete a low arousal task that requires sustained attention it can be fatiguing and lead to performance decrements. There are typically four causes of performance decrement in vigilance tasks: time, event salience, signal rate, and arousal level (Wickens, 2004). We can consider how these factors affect performance in vehicles with lower levels of automation (Levels 0-2) where drivers are still required to remain attentive throughout the entire drive.

While operating an automated vehicle, longer drives require an extended period of sustained attention and a driver's arousal level may decrease. Additionally, if the system is reliable, the likelihood of an automation failure is rare (i.e., low signal rate) which could cause expectancy to decrease and bias to occur. Automation bias, believing a system is too reliable or unreliable, can lead to the misuse or disuse of a system (Parasuraman & Riley, 1997). This combination of time, low arousal level, and low signal rate increases the likelihood a driver will miss errors in the automation's performance as a result of fatigue. Drivers may not understand which tasks or actions the automation is completing, putting them at risk of becoming out-of-the-loop (Durso & Gronlund, 1999; Endsley, 1995b, 1995a). A driver lacking necessary knowledge about the automation may be unable to take over control if the system cannot function in current conditions. It is important to understand when people are out-of-the-loop at different levels of automation, in order to help prevent drivers feeling, or being, unprepared to take back manual control.

1.2 Situation Awareness

Situation awareness (SA) was originally studied in aviation to explore a pilot's understanding during tactical flight operations. As the field of SA has expanded to understanding an operator's comprehension of different dynamic systems, different definitions for SA among researchers and practitioners have emerged. Some define SA as a construct, or a process while others consider it to be a product (Durso & Gronlund, 1999). The most commonly cited definition of SA is "the perception of 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," (Endsley, 1995b, pg. 36). Endsley views SA as a hierarchical process of perceiving information in the environment and ultimately using this knowledge to predict future states. Others consider SA to reflect "updated, meaningful knowledge of an unpredictably-changing, multifaceted situation that operators use to guide choice and action when engaged in real-time multitasking" (Gugerty, 2011, pg. 265). Following that definition, SA is claimed to be the knowledge an operator uses to make decisions or take certain actions—it is a product. Generally, definitions describe SA as a person's knowledge or understanding of a dynamic environment and their ability to use this information to make future predictions or actions.

1.2.1 Situation Awareness Measurement

The absence of an agreed upon, formalized definition of SA has led to the development of several SA measures, both subjective and objective. This study will compare different subjective and objective measures of SA.

1.2.1.1 Subjective SA Measures

Subjective measures of SA have been criticized for their inability to accurately measure true SA; participants cannot rate or know their lack of knowledge. While people may perceive their own SA level to be higher than reality, subjective measures provide insight into their confidence completing a task (Jones, 2000). Subjective measures of SA should be used alongside other performance data or objective measures due to these limitations. Observer ratings and self-report scales are examples of different types of subjective measures of SA. This study will utilize the Situation Awareness Rating Technique (SART) as it has been thoroughly tested in previous studies (see APPENDIX B). SART aims to quantify an operator's understanding of situations and how this influences decision making (Taylor, 1990). The scale contains ten constructs to measure different aspects of SA: instability of situation, variability of situation, complexity of situation, arousal, spare mental capacity, concentration, division of attention, information quantity, information quality, and familiarity. These constructs cluster to form three general domains: demand of attentional resources (D), supply of attentional resources (S), and understandability of the situation (U) (see Table 1.1 for detailed definitions of the constructs). Participants are asked to rate each construct on a Likert scale ranging from "Low" (1) to "High" (7) based on their performance of a task. The ratings are then used to create an overall SA score.

Table 1.1
Definitions of SART constructs and distributions for each domain.

Domain	Construct	Definition
Demand of Attentional Resources	Instability of situation	Likeliness of situation to change suddenly
	Variability of situation	Number of variables which require one's attention
	Complexity of situation	Degree of complication (number of closely connected parts) of situation
Supply of Attentional Resources	Arousal	Degree to which one is ready for activity (sensory excitability)
	Spare Mental Capacity	Amount of mental ability available to apply to new variables
	Concentration	Degree to which one's thoughts are brought to bear on the situation
	Division of Attention	Amount of division of attention in the situation
Understanding of the Situation	Information Quantity	Amount of knowledge received and understood
	Information Quality	Degree if goodness or value of knowledge communicated
	Familiarity	Degree of acquaintance with situation experience

1.2.1.2 Objective SA Measures

There are many ways to objectively measure SA including performance measures and query-based measures. Performance measures involve a task the participant should complete or an event requiring the participant's response (De Winter, Happee, Martens, & Stanton, 2014; Kaber, Jin, Zahabi, & Pankok, 2016; Kass, Cole, & Stanny, 2007; Ruiqi Ma & Kaber, 2005). That particular method makes the assumption that errors or performance decrements indicate lower SA and successfully completing the task implies high SA. In the driving context, examples of performance measures are accuracy on a secondary task, or reaction time to a lead vehicle suddenly braking (De Winter et al., 2014).

Query-based measures present questions about an environment to the user periodically throughout a task. That approach brings light to a user's knowledge of a particular environment and how they can use information to make predictions about future events (Durso & Gronlund, 1999; Durso & Dattel, 2004; Endsley, 1995b). The Situation Awareness Global Assessment Technique (SAGAT), a popular query-based technique, was initially developed to assess pilot SA (Endsley, 1995b, 1995a; Endsley, 1988). SAGAT is administered throughout a task by freezing and blacking out the environment from the user. While the environment is cleared, the user is presented with questions about the situation or about potential future events (Endsley, 1995b). That technique has been validated to measure SA, but may not be the most appropriate method for all environments. In a dynamic task, such as driving, it could be too disruptive to freeze the task to question the participant. Additionally, that approach relies on memory to evaluate SA. In reality, a user would still have their environment available when making critical decisions. Durso and Dattell (2004) addressed these issues in their query-based approach, the Situation Present Assessment Method (SPAM). Rather than freezing the task and clearing the environment, SPAM is administered while the task continues, thereby leaving the information in the environment available to the participant. SPAM also introduces a unique probe item (Ready Prompt) for participants to accept or decline a question; this separates workload from SA. That procedure (i.e., task, ready prompt, query) allows for real-time evaluation of SA. There are three dependent variables measured using SPAM: response time to ready prompt (workload), response time to query (SA), and accuracy of response to query (SA).

1.3 Trust in Automation

How people interact with different automated systems can depend on how much they trust or distrust a system. Trust can be defined as the "willingness of a person to be vulnerable to the actions of another, with the expectation the other will perform a specific action" (Davis, Mayer, & Schoorman, 1995, pg. 712). While that definition was originally developed to describe interpersonal trust, it can be extended to describe trust between humans and automation. It is important to emphasize that trust in automation is meant to be dynamic and continuously reflect a system's performance. If an automation system is performing poorly, a person's trust in the system should decrease. Similarly, an operator should have higher trust levels in a reliable automation system. This is known as trust calibration.

Appropriate trust calibration is necessary to properly and safely operate an automated vehicle (Hoff & Bashir, 2015; Lee & See, 2004; Noah, 2018; Noah & Walker, 2017). If a driver believes the system is extremely unreliable, they could disuse the system and miss the safety benefits of the system. Alternatively, a driver could become reliant on the automated system and misuse or abuse it (Parasuraman & Riley, 1997). As previously mentioned, automation bias can occur if a person overtrusts or undertrusts an automated system and lead to misuse, abuse, or disuse of a system (Parasuraman & Riley, 1997). As automation levels increase, inappropriate trust calibration in an automated vehicle has more severe, even lethal, consequences.

1.3.1 Subjective Trust Measure

Subjective measures of trust in automation are commonly used. The Trust in Automation scale (APPENDIX C), developed and validated by Jian, Bisantz, and Drury (2000), will be utilized in this study. The scale was created to evaluate trust levels between people and automation generally, to be used across different contexts. The scale consists of both positively and negatively framed statements to measure trust and distrust of an automated system (e.g., “The system is deceptive,” “The system provides security”). Participants rate twelve statements using a 7-point Likert scale to describe their feelings and impressions of the automated system. The range for responses is “Not at all” (1) to “Extremely” (7). The results can be evaluated based on the trust and distrust subscales or as an overall trust score by averaging the ratings.

1.4 Cognitive Workload

Cognitive workload is the relationship between the mental effort necessary and the mental resources available to complete a task (Hart & Staveland, 1988). Research has shown workload directly influences task performance (Wickens, 2004). Many factors can contribute to the increase or decrease of cognitive workload, such as task demands or environmental factors (Wickens, 2004). In the driving context, this could include traffic levels, weather, secondary tasks, or road conditions. There are performance decrements if cognitive workload is too high (Wickens, 2004).

The inclusion of automation has been claimed to decrease operator workload (De Winter et al., 2014; Endsley, 1996; Ruiqi Ma & Kaber, 2005; Turner, Young, Walker, Stanton, & Randle, 2004). The goal of adding automation is to reduce the amount of mental resources necessary to complete a task. Automation that integrates or organizes information for the operator allows relevant information to be easily identified, decreasing workload (Parasuraman, Sheridan, & Wickens, 2000). Automation does not always result in lower workload. In some cases, if workload is already low then including automation can produce low arousal; or if the automation is difficult to initiate, that can increase workload (Parasuraman et al., 2000). Automation systems in vehicles should be carefully considered to prevent creating low arousal situations or making challenging situations, such as take over requests, more difficult.

1.4.1 Subjective Cognitive Workload Measure

There are many different metrics to measure various aspects of cognitive workload, such as subjective or physiological measures. A common tool to measure subjective cognitive workload is the NASA-TLX (Hart & Staveland, 1988). It is a multidimensional questionnaire consisting of six dimensions: mental demand, physical demand, temporal demand, performance, effort, and frustration. Typically, participants are asked to rate how much they experienced each of the dimensions of workload after completing a task. Following these ratings, participants complete pairwise comparisons of all six dimensions. They are asked to choose which dimension contributed to their workload most during the task. The comparisons are used to create a weighted rating for each of the six dimensions, which in turn can be used to calculate an overall workload score.

1.5 Driver Situation Awareness

To make safe driving maneuvers, it is necessary for drivers to maintain high SA. Previous research has explored the impact of different factors on a driver's SA in manual vehicles (Kass et al., 2007; Young, Salmon, & Cornelissen, 2013). Specifically, Kass and colleagues (2007) showed driver distraction and driving experience level have an effect on their overall SA. Participants were classified as either novice (not licensed) or experienced (licensed) drivers. Participants were randomly assigned to one of two distractor conditions (hands-free cell phone or no distractor), and were asked to complete a direction-following task. A freeze query-based method was used at three points during the drive to assess SA.

Cell phone use impaired SA for both novice and experienced drivers. Experienced drivers could not overcome the added demands of talking on a cell phone while driving, and performed similar to non-distracted novice drivers (Kass et al., 2007). There is further evidence that a secondary task alters driver SA (Young et al., 2013). A study used an on-road driving test and a distractor task (Visual Detection Task) to investigate the impact of a secondary task on SA (Young et al., 2013). Participants were asked to perform a verbal protocol to describe what they were seeing and doing throughout the on-road driving task. The verbal protocol data were used to measure driver SA. The authors found distracted drivers did not necessarily take in less information or have difficulty comprehending information from their environment, but their focus changed. Distracted drivers engaged with elements of their environment to complete specific control tasks, rather than with the entire roadway environment (Young et al., 2013). Based on that previous work, it is evident distraction changes how a driver achieves and maintains SA.

1.5.1 Need for Standardized Situation Awareness Measures in Driving Research

Methods should be comparable across studies to fully understand how various factors influence and change a driver's SA. Unfortunately, there is not a standardized measure used amongst the driving research community. Across the literature, many methods have been explored, but they lack consistency and are challenging to compare across studies (De Winteret al., 2014; Durso & Dattel, 2004; Endsley, 2017a; Gugerty, 2011; Gugerty, 1997; Loft et al., 2015; Loft, Morrell, & Huf, 2013; Lu, Coster, & de Winter, 2017; Ma & Kaber, 2007; Schömig & Metz, 2013; Walker, Stanton, & Young, 2008). Performance measures are commonly used in driving tasks under the assumption that if a driver has high SA, they will be able to successfully perform a task. There is a wide range of performance metrics; some studies use object detection and comprehension in the driving environment, uptake of a non-driving related task (NDRT), or reaction to critical events to evaluate SA (De Winter et al., 2014; Schömig & Metz, 2013). These measures alone do not give us a full or exclusive understanding of a driver's SA. It is difficult to separate the influence of workload or reaction time based on secondary task uptake and response to critical events. Performance metrics are narrow and only capture a small portion of a driver's SA. They do not evaluate a driver's knowledge of their environment or how they use information available to them to make decisions.

Query-based measures reveal how a user integrates information from their environment to achieve SA, but they are not commonly used in the driving domain. Specifically, SPAM has not been heavily applied in this context, but offers clear advantages to improve our understanding. SPAM does not rely on the memory of the participant, it clearly distinguishes workload, and does not pause or freeze the dynamic task to assess SA. That technique has been successfully applied in other contexts, such as submarine task management and air traffic control (Durso & Dattel, 2004; Loft et al., 2015, 2013; Pierce, 2012). Utilizing SPAM to asses SA in automated vehicles can provide insight about a driver's knowledge of the environment and automated system without disrupting or distracting the driver. To continue to make progress in our understanding of SA, future studies must use more rigorous and sensitive measures.

Research regarding automated vehicles has primarily focused on how the driver interacts with the automated systems and how we can improve this interaction (Becerra, Holthausen, & Walker, 2019b, 2019a; Beller, Heesen, & Vollrath, 2013; Gable, Tomlinson, Cantrell, & Walker, 2017; Helldin, Falkman, Riveiro, & Davidsson, 2013;

Noah, 2018; Noah & Walker, 2018). Some research has considered SA's role in this interaction, but the emphasis was learning about SA's role in take over requests or situations of automation failure (Merat & Jamson, 2009; Merat, Jamson, Lai, & Carsten, 2012; Noah, 2018). Studies have compared driver SA in a manual driving task and a highly automated driving task and determined the manual driving task led to higher SA (De Winter et al., 2014; Merat & Jamson, 2009). That research demonstrates the importance of a driver remaining engaged in the driving task to safely overcome critical events and maintain SA (De Winter et al., 2014; Merat & Jamson, 2009). The understanding of SA while operating an automated vehicle is severely limited by using such a narrow scope. Utilizing critical events to measure and learn about SA in automated vehicles only allows one to discover a driver's knowledge in that moment—we are ignoring their SA (or lack of) leading up to this point. It is still unclear what information drivers remain aware of at various levels of automation. Even further, there has been little investigation to understand how lower levels of automation (Level 1-2) affect SA. While there is a general understanding of how automation impacts SA, there is not enough evidence to discern how SA is influenced in the automated vehicle context.

1.6 Current Study

The purpose of this current research is to understand how SA changes at different levels of automation and to compare different methods to assess SA. This study investigated how SA transformed throughout a drive in a simulated low automation vehicle (Level 1) and a simulated high automation vehicle (Level 3). In addition to this characterization, this study determined differences in SA between low (Level 1) and high (Level 3) automation levels. Lastly, this study compared and evaluated three different measures of SA (query-based measure, subjective measure, and performance-based measure).

Specifically, this research aimed to address the following research questions:

RQ1: How do different subjective and objective measures of situation awareness compare in their measurement and sensitivity of situation awareness?

RQ2: How does situation awareness change throughout a drive utilizing a low automation system?

RQ3: How does situation awareness change throughout a drive utilizing a high automation system?

RQ4: What are the differences in situation awareness between a low automation and high automation drive?

RQ5: What other factors can help predict situation awareness?

CHAPTER 2. METHODS

This study consisted of one independent variable: automation level (within-subjects). The automation level varied across two conditions: (1) low automation (ALK only initiated); and (2) high automation (ALK and ACC both initiated). Participants were randomly assigned to a condition and the conditions were counterbalanced. There were five dependent variables: objective measure of SA (SPAM; Durso & Dattel, 2004); performance measure of SA (voluntary secondary task engagement); subjective measure of SA (SART; Taylor, 1990); subjective measure of workload (NASA-TLX; Hart & Staveland, 1988); and subjective measure of trust in automation (Trust in Automation Scale; Jian, Bisantz, & Drury, 2000).

2.1 Participants

Georgia Institute of Technology psychology students enrolled in the SONA system participated in this study. For their participation, they received partial course credit. The following inclusion criteria were used: normal or corrected-to-normal vision and hearing; sufficient mobility to operate a vehicle; a valid US driver's license; and at least two years of driving experience. The minimum driving experience requirement ensured all participants had adequate driving experience, and prevented novice drivers from biasing the results.

Sixty-three participants were recruited for this study. Only 50 participants' data were analyzed for the purposes of this research. Participants were excluded from analysis for a variety of reasons including, failing to meet the driving experience requirement (n= 2), turning off the automated system and/or crashing the vehicle (n= 8), or technical errors (n= 3).

Table 2.1.
Self-reported prior experience with automation features.

Automation System Familiarity: Automated Lane Keeping (ALK)		
	Frequency	**Percent**
Own a vehicle with ALK	6	12
Driven a vehicle with ALK	17	34
Passenger in a vehicle with ALK	14	28
Familiar with ALK	11	22
Never heard of with ALK	2	4
Automation System Familiarity: Adaptive Cruise Control (ACC)		
	Frequency	**Percent**
Own a vehicle with ACC	14	28
Driven a vehicle with ACC	12	24
Passenger in a vehicle with ACC	11	22
Familiar with ACC	9	18
Never heard of with ACC	4	8

The final group of participants (N= 50) ranged in age from 18 to 28 years-old (*M*= 20.12, SD= 1.75). Participants had an average of 3.59 years of driving experience (SD= 1.45). There were 28 male participants (56%) and 22 female participants (44%) in the study. As seen in Table 2.1, most participants had direct experience with automated lane keeping systems as either owning a vehicle (12%), having previously driven a vehicle (34%), or being a passenger in a vehicle (28%) with an ALK system. Similarly, many participants had direct experience with adaptive cruise control systems as either owning a vehicle (28%), having previously driven a vehicle (24%), or being a passenger in a vehicle (22%) with an ACC system. As college students, this sample of may have limited experience owning or driving automated vehicles than an adult. These results may not representative of the general population's familiarity with automated features.

Figure 2.1. Photo of the NADS MiniSim quarter-cab driving simulator used in this study.

2.2 Materials

2.2.1 Driving Task

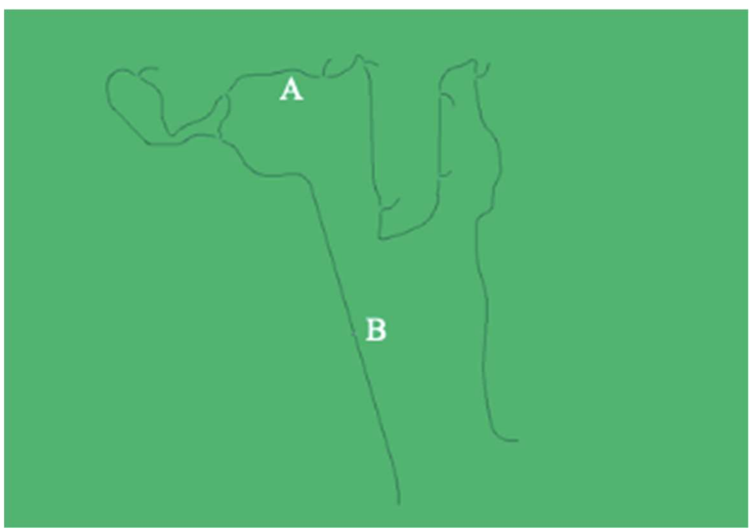
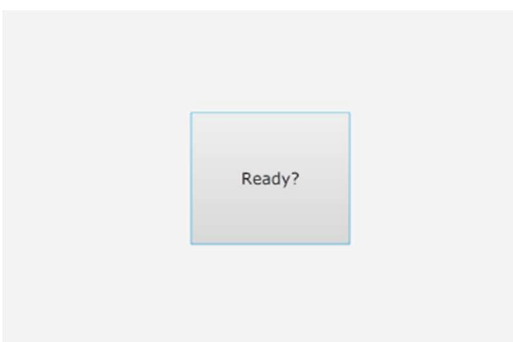


Figure 2.2. Map of the simulated driving scenarios. Point A is the starting point in the low automation drive (Level 1) and point B is the approximate end point. In the high automation drive (Level 3), point B is the starting point and point A is the approximate end point.

Participants completed two simulated drives in the NADS MiniSim quarter-cab driving simulator (Version 2.2.1), as seen in Figure 2.1. The simulated drives took place on a rural, curvy road with low to moderate traffic. Their task was to maintain a speed of 55 miles per hour to the best of their ability. A lead vehicle remained in front of the participant to ensure they maintained a speed of 55 miles per hour. During the low automation drive, participants drove with ALK, a system that maintains vehicle position within the lane. The ALK system turned on at the beginning of the drive and the participant controlled the speed of the car. Participants began at point A and ended approximately at point B (Figure 2.2). During the high automation drive, however, participants utilized both ALK and ACC, a system that maintains the speed of the vehicle and distance from a lead vehicle. Both ALK and ACC were activated at the beginning of the drive. Participants started at point B and ended at approximately point A (Figure 2.2). Both drives ended after twenty minutes had elapsed, which is why the endpoints are estimated. The drives were counterbalanced.

2.2.2 Measures of Situation Awareness

2.2.2.1 <u>Situation Present Assessment Method</u>



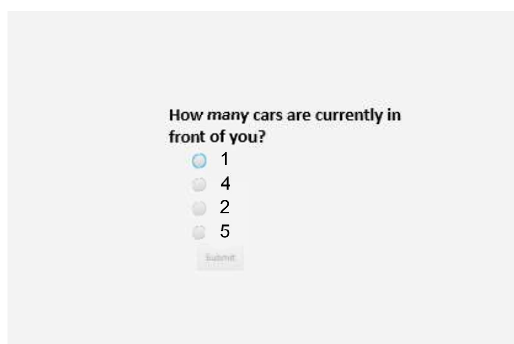


Figure 2.3. Presentation of the SPAM prompts throughout each drive. The picture on the right shows the ready prompt, while the picture on the right displays an example query.

The Situation Present Assessment Method (SPAM) was administered throughout the course of both drives to assess drivers' real-time SA (Durso & Dattel, 2004). The ready prompt and queries were presented visually on a small touch screen mounted to the right of the participant (where a radio would be located). The driving task continued while participants responded to SPAM prompts. Participants had five seconds to accept the ready prompt, otherwise it timed out and was recorded as a miss. To accept the ready prompt, participants tapped the “Ready” button on the screen. The time to respond to the ready prompt can be used as an objective measure of workload (Durso & Dattel, 2004). After accepting the ready prompt, a query appeared on the screen; all queries were multiple choice questions. There was a total of six queries throughout each drive, about one query every 2-5 minutes. The queries asked participants about past (2), present (2), or future (2) events or stimuli in the drive (APPENDIX E). An example of how the SPAM prompts were presented can be found in Figure 2.3. The time to respond to a query and the accuracy of the response was used to objectively measure the participant’s SA.

2.2.3 Secondary Task Engagement

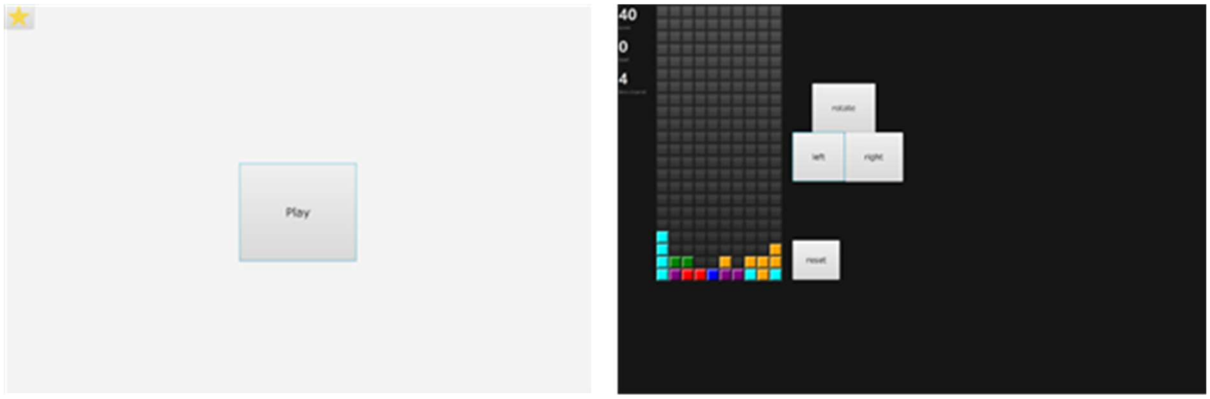


Figure 2.4. The secondary task provided to participants throughout both automated drives. The picture on the left is the play button participants will press to begin the secondary task. The picture on the right shows the Tetris game and controls.

During the drive, participants had the option to engage in a secondary non-driving related task (NDRT), namely the game Tetris (Figure 2.4). Tetris is a visuospatial task that requires players to strategically place different geometrical shapes to complete a row. Once a row is completed it disappears and 10 points are awarded. The shapes can be manipulated by being shifted horizontally or rotated clockwise in 90-degree increments. If there is not enough space for new shapes to be placed, the trial will reset by clearing. Participants were instructed to earn as many points as possible and were informed their score would be recorded. They were also instructed that the driving task was their primary task and they were only to engage with the secondary task when they felt it was safe (see APPENDIX F). Tetris was presented on a touchscreen to the right of the participant (same screen as the SPAM prompts). A start screen was displayed until participants choose to play the game. If the participant chose to play, they controlled where the shapes fell using controls on the touchscreen. The game blacked out when a SPAM ready prompt was displayed. After participants either missed a SPAM ready prompt or responded to a SPAM query, a new trial of Tetris began with the play screen for a total of seven Tetris trials per drive. The time to start a Tetris trial, the score, the number of resets, and time engaged in the secondary task were recorded.

2.2.3.1 SART

The SART scale (Taylor, 1990) was used to measure a participant's subjective SA. It was administered at the conclusion of each drive. The SART scale was created on Qualtrics and administered electronically on a provided iPad. Participants remained seated in the driving simulator while completing all post-drive questionnaires. When all questionnaires were finished, the researcher retrieved the iPad from the participant. Each of the scales listed below were administered in the same manner.

2.2.4 Measure of Cognitive Workload

2.2.4.1 Subjective Workload

The NASA-TLX scale (Hart & Staveland, 1988) was used to measure participant's subjective workload. Participants completed the measure after they finished each drive. Participants completed the scale electronically on a provided iPad while remaining seated in the driving simulator.

2.2.5 Measure of Trust in Automation

2.2.5.1 Trust in Automation Scale

The Trust in Automation scale (Jian, Bisantz, & Drury, 2000) was administered before the first drive to measure initial trust levels and after each drive to measure how trust levels changed after experience with the automated system. Participants responded to the scale electronically on a provided iPad while seated in the driving simulator.

2.3 Procedure

Following completion of the informed consent form (APPENDIX G) and demographics survey (APPENDIX H), participants completed the Georgia Tech Simulator Sickness Screening Protocol (SSSP; Gable & Walker, 2013). The SSSP helped ensure participants did not have any physical discomfort during the experiment. The Trust in Automation scale was administered when the participant passed the simulator sickness screening~~.~~ Participants completed a practice trial of the secondary task, Tetris, and one SPAM query to become familiar with the measures. The first drive began in one of two automation conditions (low automation; high automation). In the low automation condition, ALK turned on when the drive initiated. In the high automation condition, both ALK and ACC turned on when the drive initiated. Throughout the drive, participants chose to engage in the secondary task, Tetris, if they felt it was safe. Periodically, SPAM prompts (ready prompts; queries) were presented for participants to respond to. At the completion of the first drive, SART, the NASA-TLX scale, and the Trust in Automation scale were administered. The second drive included the remaining automation condition. Again, during the drive participants chose when to engage in the secondary task; and SPAM prompts were presented throughout the drive. Upon completing the second drive, SART, the NASA-TLX scale, and the Trust in Automation scale were administered once more. Finally, participants were debriefed (APPENDIX K), released, and awarded credit for their participation.

CHAPTER 3. ANALYSES AND RESULTS

The hypotheses and analyses have been grouped by research question. The Greenhouse-Geisser epsilon estimate was used to adjust for sphericity assumptions in repeated measures ANOVA analyses. When appropriate, Bonferroni corrections to alpha levels were used to adjust for family-wise Type I error.

3.1 How do different subjective and objective measures of situation awareness compare in their measurement and sensitivity of situation awareness?

3.1.1 Hypothesis 1a

The secondary task engagement measure will be related to workload.

Table 3.1.
Descriptive statistics of all dependent variables.

	Mean	SD	Minimum	Maximum
SPAM Query RT	6.64	2.41	0.70	16.68
SPAM Query Accuracy	0.55	0.16	0.08	0.83
SPAM Correct Query RT	7.04	3.46	2.40	23.72
Secondary Task TOT	997.45	242.62	38.30	1152.26
Secondary Task Score	30.39	10.62	0.71	48.57
SART Score	15.30	4.16	8.50	26.50
Combined Trust Score	5.49	1.02	2.17	7.00
Distrust Subscale Score	1.78	0.87	1.00	5.70
Trust Subscale Score	4.97	1.32	1.86	7.00
NASA-TLX Score	4.09	1.25	2.03	6.37

Table 3.2.
Pearson correlations between NASA-TLX and secondary task engagement data.

	NASA-TLX Score	Secondary Task TOT	Secondary Task Score
NASA-TLX Score	-		
Secondary Task TOT	-.02	-	
Secondary Task Score	.01	**.77****	-

Multiple correlations were calculated to determine the relationship between the secondary task engagement and NASA-TLX. This hypothesis was tested with the secondary task data, total time spent on task (TOT) and mean score, and calculated workload scores from the NASA-TLX data. No significant correlations were found between secondary task engagement TOT and NASA-TLX, $r(48)= .23$, $p= .11$, or secondary task engagement mean score and NASA-TLX, $r(48)= .15$, $p= .30$, as seen in Table 3.2. This does not support convergent validity and does not support the hypothesis (1a) that the secondary task engagement measure would be related to workload.

3.1.2 Hypothesis 1b

The SPAM measure and the SART questionnaire will not be related to the secondary task engagement measure.

To test this hypothesis, multiple correlations were conducted between the mean RT to a SPAM query, SPAM accuracy, SART scores, secondary task engagement TOT, and the secondary task mean scores to determine the relationship between these SA measures. An overall SART score was calculated by first summing the items measuring understanding of the situation (U), the items measuring the supply of attentional resources

(S), and the items measuring the demands on attentional resources (D). Using these sums, a SART score for each participant was determined by the following equation: U- (D-S).

Table 3.3.
Pearson correlations among SPAM, SART, and secondary task engagement.

	SPAM Query RT	**SPAM Query Accuracy**	**SART Score**	**Secondary Task TOT**	**Secondary Task Score**
SPAM Query RT	-				
SPAM Query Accuracy	.12	-			
SART Score	**-.36***	-.16	-		
Secondary Task TOT	.01	**.29***	-.02	-	
Secondary Task Score	-.07	.21	.01	**.77****	-

**. Correlation is significant at the 0.01 level.
*. Correlation is significant at the 0.05 level.

A statistically significant positive correlation was found between SPAM accuracy and secondary task engagement TOT, $r(48)= .29$, $p= .04$. No other significant correlations were found between SPAM and secondary task engagement, see Table 3.3 for correlations. This suggests SPAM and secondary task engagement may not be related or measuring the same construct. There were also no significant correlations between SART and secondary task engagement, see Table 3.3 for correlations. Additionally, there was a statistically significant negative correlation between SART scores and SPAM query RT, $r(48)= -.36$, $p= .01$. This suggests as SART scores increase, indicating high SA, SPAM query RT decrease, also demonstrating high SA. This result shows that SART and SPAM are related and both measuring the same construct. This confirms the hypothesis (1b) that the SART and secondary task engagement measures would not be related and this could be evaluating different constructs.

3.1.3 Exploratory Factor Analysis

Table 3.4.
Communalities matrix.

Communalities		
	Initial	Extraction
SART (1) Instability of situation	0.679	0.583
SART (2) Complexity of situation	0.748	0.938
SART (3) Variability of situation	0.426	0.408
SART (4) Arousal	0.546	0.678
SART (5) Concentration	0.683	0.935
SART (6) Division of Attention	0.494	0.487
SART (7) Spare Mental Capacity	0.148	0.073
SART (8) Information Quantity	0.327	0.345
SART (9) Familiarity	0.357	0.241
SPAM Average Query RT	0.423	0.362

An exploratory factor analysis (EFA) was performed using SPAM and SART, based on the correlation analysis, to determine how these measures contribute to the underlying structure of SA. Specifically, the average RT to a query and the nine SART items were used in the EFA. There are several ways to determine the number of factors to include in the model. First, a principal component analysis (PCA) was conducted as a way to determine the appropriate number of factors to retain in the EFA. The PCA identified three factors to retain by consideration of the scree-plot and Kaiser's eigenvalue criterion. An EFA was then conducted, using a correlation matrix, to identify the factor structure. Factors were extracted by the principal axis method and rotated by a Promax rotation. The communalities matrix revealed how correlated items were with each other, which identifies how likely items are to load onto factors (see Table 3.4). The number of factors in the EFA was determined by consideration of the scree-plot and Kaiser's eigenvalue criterion. A total of two factors were extracted and rotated.

Table 3.5.
Structure matrix for coefficients.

	Factor	
	1	2
SART (1) Instability of situation	**0.741**	0.336
SART (2) Complexity of situation	**0.948**	0.111
SART (3) Variability of situation	**0.622**	0.135
SART (4) Arousal	0.362	**0.810**
SART (5) Concentration	0.421	**0.738**
SART (6) Division of Attention	0.482	0.346
SART (7) Spare Mental Capacity	-0.092	-0.053
SART (8) Information Quantity	0.484	0.419
SART (9) Familiarity	0.223	0.457
SPAM Average Query RT	0.026	**-0.563**
% of variance explained	32.90%	12.43%

Rotation Method: Promax with Kaiser Normalization.

Factor structures of the EFA results can be found in Table 3.5. Factor 1 explained 32.90% of the variance and factor 2 accounted for 12.43% of the variance, and a cumulative variance of 45.32%. SART items (6), (7), (8), and (9) had a factor loading of less than 0.5 for both factors. Factor 1 was comprised of SART items (1), (2), and (3) and all had factor loadings greater than 0.6. These specific SART items make up the demand of attentional resources domain. Factor 2 included SART items (4) and (5) and the SPAM average RT to a query and all had factor loadings greater than 0.5. SART items (4) and (5) are related to the supply of attentional resources domain and the RT to a query indicates how much participants used their environment to answer a query compared to information in their working memory. These results suggest the structure of SA is comprised of the unpredictability of a situation and the attentiveness given to the situation.

3.2 How is situation awareness impacted by different levels of automation?

3.2.1 Hypothesis 2a

Situation awareness will decrease throughout the course of the drive utilizing a low automation system.

Table 3.6.
SPAM query time points for the low automation drive.

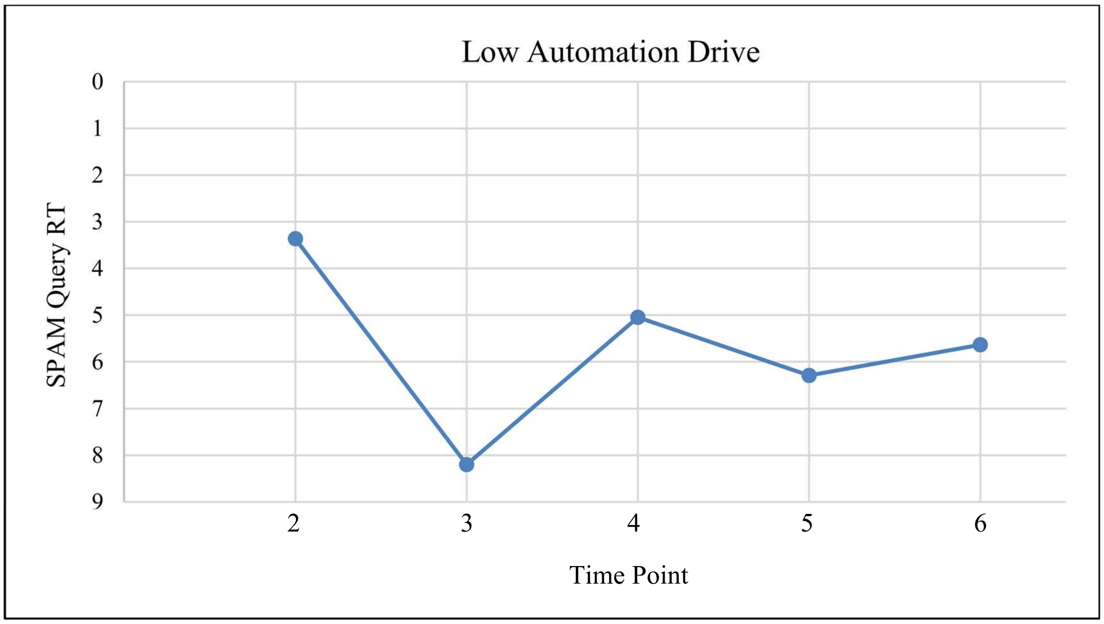
Low Automation Drive Time points					
Q1	Q2	Q3	Q4	Q5	Q6
2:00	5:22	7:55	10:46	13:25	16:56

*Approximate timing based on maintaining 55 mph

The drive was separated into segments to evaluate how SA changes during the low automation drive. Based on the results from Hypothesis 1b, only SPAM data were used to assess this hypothesis. A one-way repeated measures ANOVA was performed to determine if there were differences in SA across six time points in the low automation drive, based on the six SPAM queries presented (see Table 3.6). Specifically, the reaction time (RT) to a SPAM query was the dependent variable.

While analysing the data, it became evident the first time point was an outlier in the data set; the average RT at the first time point was 13.39 seconds (SD= 10.42). Further review of the query revealed there were issues with the apparatus, specifically, the event used for the query. The event was presented earlier than the query and this led to differences in behavior (longer RT). This issue did not occur for any other time points in the low automation drive. For these reasons, these data were excluded from analysis leaving a total of five time points to be compared.

Mauchly's test for sphericity indicated the sphericity assumption was violated, χ^2 (9)= 43.964, $p< .001$, therefore a Greenhouse-Geisser correction was used. There was a statistically significant main effect of time on SA in the low automation drive as measured by the RT to a SPAM query, $F(2.85, 139.65) = 16.35$, $p< .001$. See Figure 3.1 for results.

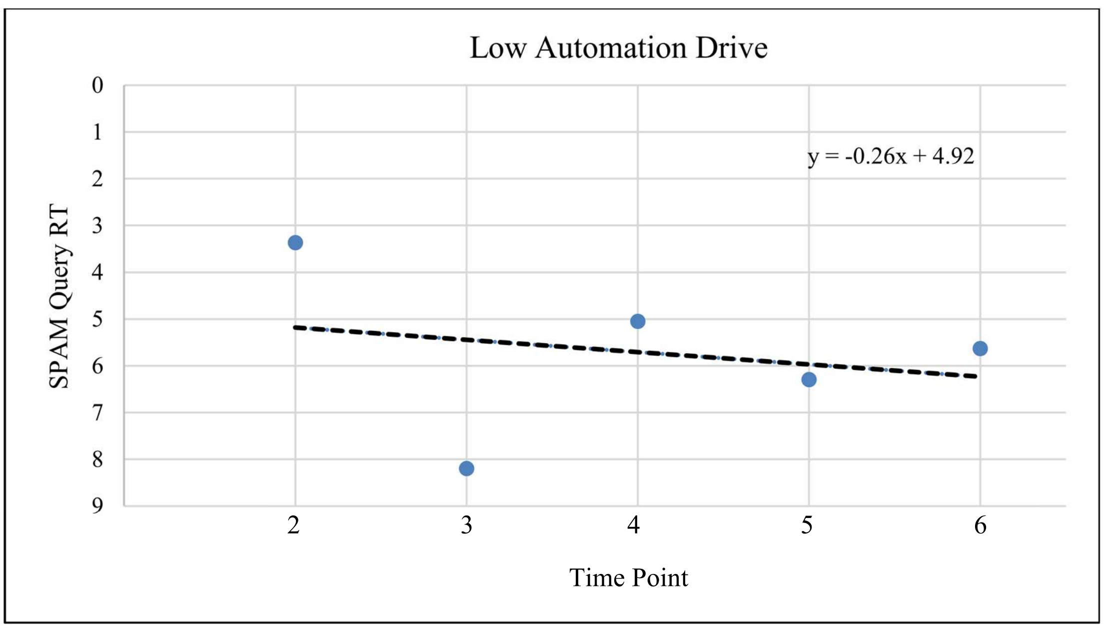


Note: The y-axis has been reversed to reflect the decrease in SA (i.e., higher query RT indicates lower SA).

Figure 3.1. One-way repeated measures ANOVA results for low automation drive.

This result indicates SA was significantly different throughout the low automation drive. Post-hoc paired t-tests for all pairwise comparisons were conducted to further determine differences in SA throughout the low automation drive. Bonferroni corrections to alpha levels were used to adjust for family-wise Type I error.

SA was higher at the Time Point 1 (M= 3.36, SD= 3.18) compared to Time Point 2 (M= 8.20, SD= 4.78), $t(49)$= -5.74, p< .001. Participants' SA was also significantly higher at Time Point 1 (M= 3.36, SD= 3.18) compared to Time Point 4 (M= 6.29, SD= 3.66), $t(49)$= -4.15, p< .001, and significantly higher than at Time Point 5 (M= 5.63, SD= 3.03), $t(49)$= -3.40, p= .001. At Time Point 2 (M= 8.20, SD= 4.78), SA was significantly lower compared to Time Point 3 (M= 5.05, SD=3.15), $t(49)$= 4.65, p< .001. SA was also significantly lower at Time Point 2 (M= 8.20, SD= 4.78) than at Time Point 5 (M= 5.63, SD= 3.03), $t(49)$= 4.71, p< .001. Lastly, SA was higher at Time Point 3 (M= 5.05, SD= 3.15) compared to Time Point 4 (M= *6.29*, SD= 3.66), $t(49)$= -3.14, p= .003. All other pairwise comparisons were not statistically significant.



Note: The y-axis has been reversed to reflect the decrease in SA (i.e., higher query RT indicates lower SA).

Figure 3.2. Trend analysis results for the low automation drive.

A trend analysis was also conducted to determine general patterns of SA over the course of the low automation drive (see Figure 3.2). The trend analysis shows SA decreased over time (because SPAM Query RT increased). However, the slope of the trend line (B= 0.26) was not significantly different from zero, $t(4)= .419$, $p= .70$. This does not provide statistical evidence for the hypothesis (2a) that SA decreased throughout the low automation drive. Although there was not statistical support, the results do indicate there is a trend where SA decreases over time. It is possible the drive was not long enough to reveal significant differences in this study.

3.2.2 Hypothesis 2b

Participants will have low situation awareness throughout the course of the drive utilizing a high automation system.

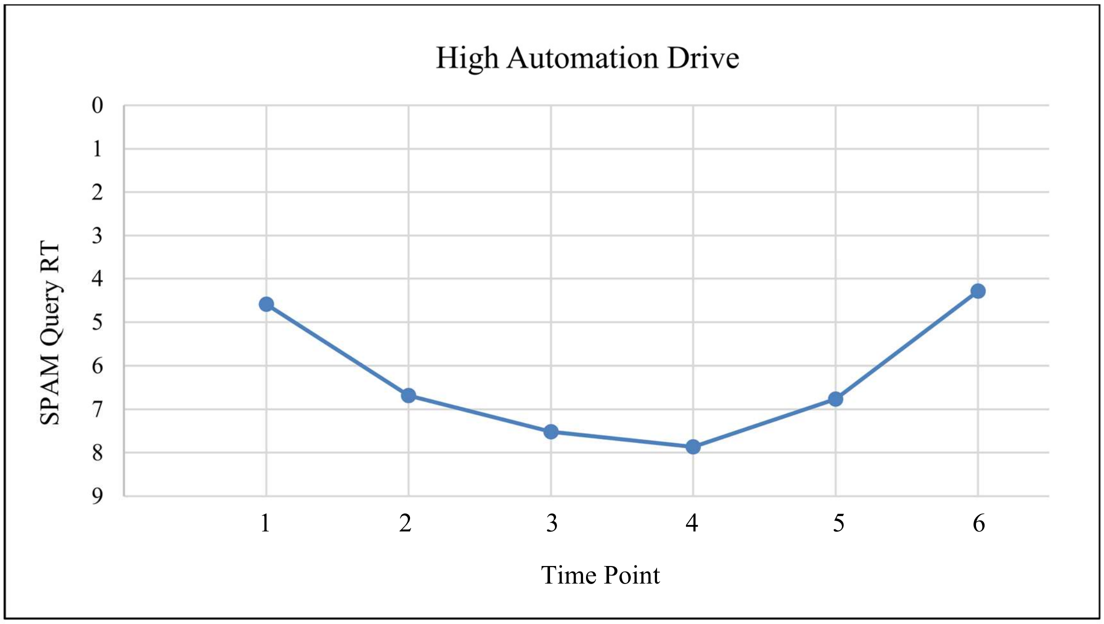
Table 3.7.
SPAM query time points for the high automation drive.

High Automation Drive Time points					
Q1	Q2	Q3	Q4	Q5	Q6
2:56	6:00	8:11	12:09	14:02	18:03

*Timing based on ACC maintaining 55 mph

Again, the drive was separated into segments to determine how SA changes during the high automation drive. A one-way repeated measures ANOVA was conducted to test if there were differences in SA across six time points in the high automation drive, based on the six SPAM queries presented (see Table 3.7). The SPAM query RT data were used to test this hypothesis. There were not outliers identified in this data set.

Mauchly's test for sphericity indicated the sphericity assumption was violated, χ^2 (14)= 24.515, p= .04, thus a Greenhouse-Geisser correction was used. A statistically significant main effect of time on SA in the high automation drive as measured by the RT to a SPAM query was found, $F(4.19, 205.22) = 8.68$, $p < .001$. See Figure 3.3 for results.



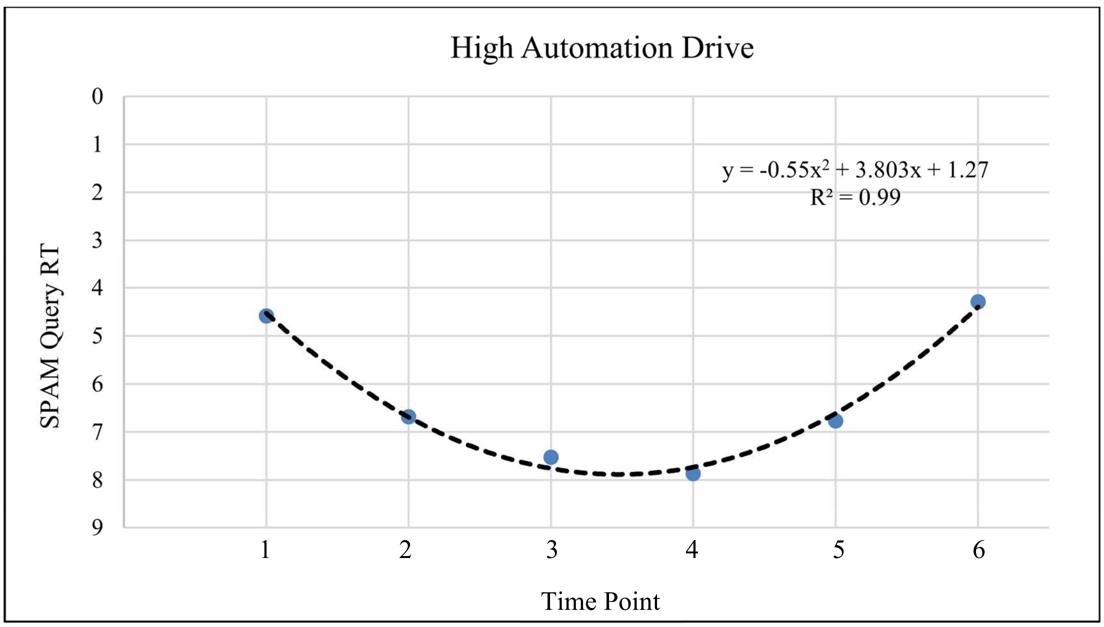
Note: The y-axis has been reversed to reflect the decrease in SA (i.e., higher query RT indicates lower SA).

Figure 3.3. One-way repeated measures ANOVA results for high automation drive.

This result suggests SA was significantly different throughout the high automation drive. Post-hoc paired t-tests for all pairwise comparisons were conducted to further determine differences in SA throughout the high automation drive. Bonferroni corrections to alpha levels were used to adjust for family-wise Type I error.

During the high automation drive, SA was significantly higher at Time Point 1 (M= 4.59, SD= 3.26) than Time Point 3 (M= 7.52, SD= 3.53), $t(49)$= -4.72, $p < .001$. Participants' SA at Time Point 1 (M= 4.59, SD= 3.26) compared to Time Point 4 (M=

7.87, SD= 4.61) was also found to be significantly higher, $t(49)= -4.039$, $p< .001$. SA was significantly better at Time Point 1 ($M= 4.59$, SD= 3.26) compared to Time Point 5 ($M= 6.77$, SD= 3.53), $t(49)= -3.58$, $p= .001$. Participants' SA at Time Point 2 ($M= 6.69$, SD= 4.57) was significantly lower than at Time Point 6 ($M= 4.29$, SD= 3.24), $t(49)= 3.172$, $p= .003$. At Time Point 6 ($M= 4.29$, SD= 3.24), SA was significantly higher than at Time Point 3 ($M= 7.52$, SD= 3.53), $t(49)= 4.80$, $p< .001$, and significantly higher compared to Time Point 4 ($M= 7.87$, SD= 4.61), $t(49)= 4.18$, $p< .001$. SA was significantly higher at Time Point 6 than ($M= 4.29$, SD= 3.24) Time Point 5 ($M= 6.77$, SD= 3.53), $t(49)= 3.50$, $p= .001$. All other pairwise comparisons were not statistically significant.



Note: The y-axis has been reversed to reflect the decrease in SA (i.e., higher query RT indicates lower SA).

Figure 3.4. Trend analysis results for the high automation drive.

Additionally, a trend analysis was performed to determine general patterns of SA over the course of the high automation drive (see Figure 3.4). The trend analysis suggests a quadratic trend best describes the changes in SA throughout the high automation drive,

$y = 1.27 + 3.80x - .55x^2$, $R^2 = 0.99$. These results do not confirm the hypothesis (1b) that low SA would be maintained during the high automation drive. Rather, it indicates there was high SA at both the beginning (Time Point 1) and end (Time Point 6) of the drive and SA decreased throughout the middle of the drive.

3.2.3 Hypothesis 2c

As the automation level increases, from low to high, situation awareness will decrease.

To test this hypothesis, paired t-tests were conducted between the low and high automation drives using SPAM data (RT to a query, accuracy) and SART scores. The secondary task engagement measure was not utilized based on the correlation analysis in Hypothesis 1b. Bonferroni corrections to alpha levels were used to adjust for family-wise Type I error.

The first paired t-tests were conducted between the average RT to a query on the low automation and high automation drives, as well as between the accuracy scores on the low automation and high automation drives. In addition, the average RT to a query answered correctly (RT to correct query) was compared between the low and high automation drives. Two participants had a drive without any correct responses to a query, either due to missing or rejecting a ready prompt or answering the query incorrectly, and were excluded from this analysis.

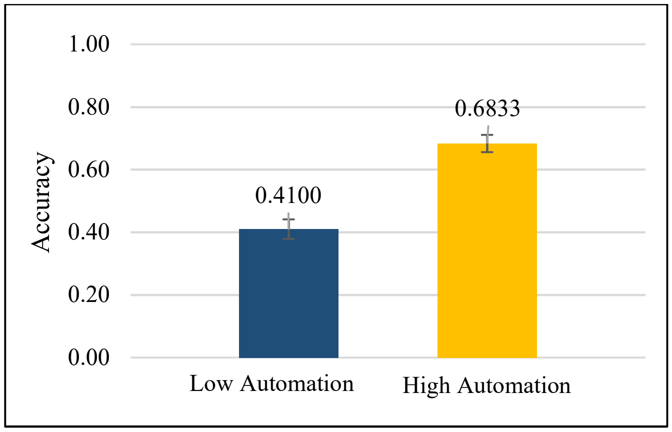


Figure 3.5. SPAM Accuracy paired t-test results.

See Figure 3.5 for the SPAM accuracy paired t-test results. SPAM accuracy was significantly worse throughout the low automation drive (*M*= .41, *SD*= .22) compared to the high automation drive (*M*= .68, *SD*= .19), *t(*49)= -7.43, *p*< .001.

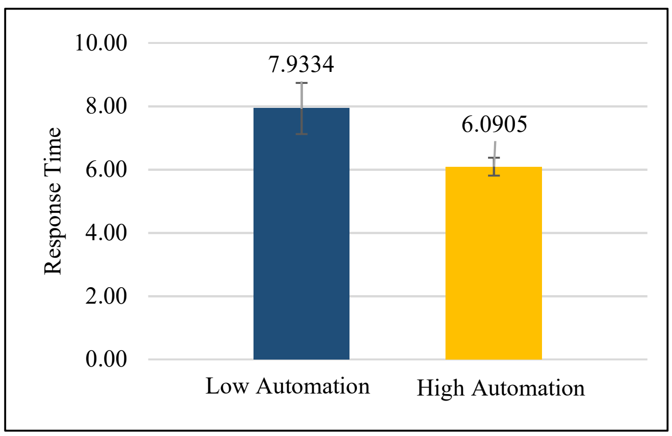


Figure 3.6. SPAM RT to a correct query paired t-test results.

See Figure 3.6 for the SPAM RT to correct query paired t-test results. The RT to a correct query was quicker in the high automation drive (*M*= 6.09, *SD*= 1.95) than in the low automation drive (*M*= 7.93, *SD*= 5.58) revealing significantly higher SA in the high automation drive, *t(*47)= 2.822, *p*= .007.

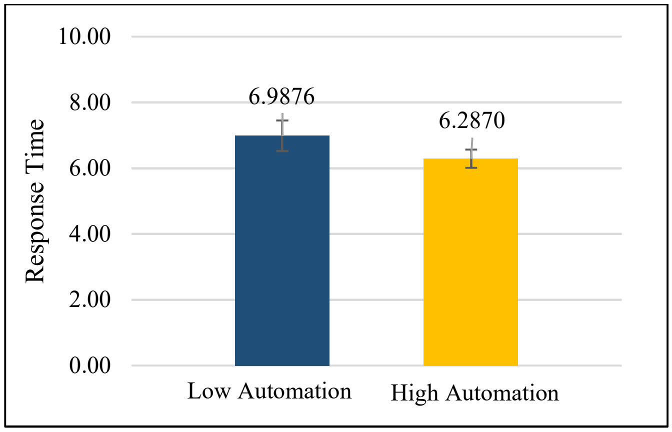


Figure 3.7. SPAM RT to a query paired t-test results.

See Figure 3.7 for the SPAM RT to a query paired t-test results. There was not a significant difference in RT to a query, $t(49)= 1.97$, $p= .055$, despite SA being higher in the high automation drive ($M= 6.29$, $SD= 1.96$) compared to the low automation drive ($M= 6.99$, $SD= 3.30$).

These results do not confirm the hypothesis (2c) that SA, as measured by SPAM, decreases as the automation level increases. Rather, the results suggest the opposite, that SA was significantly higher with high level automation.

The next test used the data collected from the SART questionnaire. The SART scores for the low automation drive and the high automation drive were compared using a paired t-test. See Figure 3.8 for the SART paired t-test results.

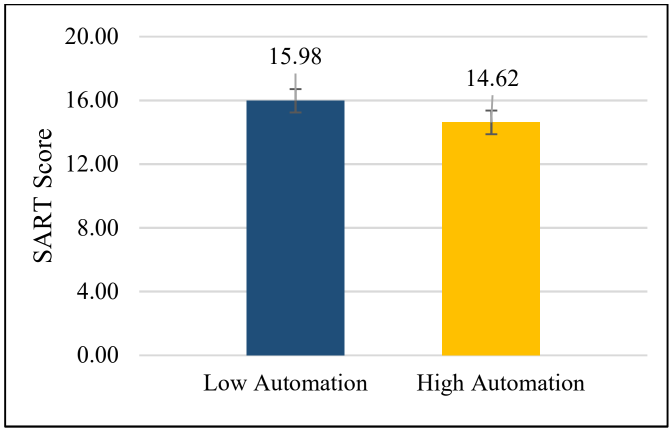


Figure 3.8. SART scores paired t-test results.

There was not a significant difference found between the SART scores, $t(49)= 1.54$, $p= .13$, in the low automation drive ($M= 15.98$, $SD= 5.18$) and the high automation drive ($M= 14.62$, $SD= 5.24$). This suggests SA was similar, as measured by SART, throughout the low and high automation drives. These results do not confirm the hypothesis (2c).

3.3 What factors can help predict situation awareness?

3.3.1 Hypothesis 3a

Trust in automation and workload can help predict situation awareness.

Hypothesis 3a was not evaluated as there were no significant correlations between trust, workload, and SA as measured by the Trust in Automation scale, NASA-TLX, SPAM, SART, or secondary task engagement. See Table 3.8 for the Pearson correlations.

Table 3.8.

Pearson correlations among Trust in Automation scale, NASA-TLX, and SA measures.

	Overall Trust Score	Distrust Subscale Score	Trust Subscale Score	NASA-TLX Score	SART Score	2nd Task TOT	2nd Task Score	SPAM Query RT	SPAM Query Accuracy
Overall Trust Score	-								
Distrust Subscale Score	**-.79****	-							
Trust Subscale Score	**.96****	**-.57****	-						
NASA-TLX Score	.06	.23	.19	-					
SART Score	.18	-.17	.16	**.40****	-				
2nd Task TOT	.07	-.01	.08	-.02	.23	-			
2nd Task Score	-.05	.09	-.02	.01	.15	**.77****	-		
SPAM Query RT	-.04	-.05	-0.07	**-.36***	**-.33***	.01	-.07	-	
SPAM Query Accuracy	-.07	-.06	-.12	-.16	-.18	**.29***	.21	.12	-

**. Correlation is significant at the 0.01 level.

*. Correlation is significant at the 0.05 level.

CHAPTER 4. DISCUSSION

Automated systems are becoming increasingly available to drivers with the promise of a new, relaxing driving experience. However, current vehicles with automation (Levels 0-2) require drivers to remain engaged with both the automated system and the roadway environment. Previous research has found when drivers have poor SA, they are not prepared to take over manual control (Merat et al., 2012). There is no exploration to help us understand how SA is impacted by the inclusion of automated systems or how different levels of automation change driver SA. Even further, researchers and practitioners use a variety of methods to measure SA, making it even more challenging to compare studies.

This research addressed these issues by evaluating three measures of SA (SPAM, Secondary Task Engagement, and SART) and measure driver SA at two levels of automation: low automated vehicle (Level 1) and high automated vehicle (Level 3). This research revealed secondary task engagement is not a sufficient measure of SA. This study provides evidence to use dynamic, query-based measures and subjective measures of SA in the automated vehicle context. Further, when considering the components of SART and SPAM, attentional demand of a situation and a person's concentration of attention were found to be key factors underlying SA.

This research also provided a deeper understanding of how SA changes across a low automation drive (where the driver completes part of the driving task), and a high automation drive (where the driver does not complete any of the driving task). Specifically, when utilizing the low automation system, SA did not significantly change throughout the drive. Throughout the high automation drive, SA was high at both the start and end of the drive, but worsened throughout the middle of the drive. Lastly, multiple measures identified significant differences in SA between the low and high automated drives. Participants had significantly higher SA when using the high automation system compared to the low automation system.

4.1 Theoretical Contributions: Situation Awareness Measurement

Three measures of SA were administered and compared: SPAM, SART, and secondary task engagement. Secondary task engagement has been used to assess driver SA (De Winter et al., 2014; Schömig & Metz, 2013), however secondary task engagement was not found to be largely related to either standardized measure of SA in this study. A relationship was identified between time spent engaged in the secondary task and SPAM accuracy scores. This is not unexpected as the secondary task required drivers to divide their attention, which can impact their accuracy on SPAM queries. However, in this study, more time engaged in the secondary task was related to higher query accuracy. The secondary task could have helped participants fight fatigue and increased arousal. Outside of this relationship with SPAM accuracy scores, secondary task engagement was not related to other aspects of SPAM or SART.

These findings suggest secondary task engagement is measuring a different construct. Secondary task engagement was also compared to workload scores, but a significant relationship was not found. This study demonstrates that secondary task engagement is not evaluating SA or workload, and, therefore, should not be used to measure these constructs.

Additionally, the two standardized measures (SPAM and SART) were compared and evaluated on their ability to capture SA. These measures were found to be related. Specifically, SPAM query RT and multiple SART items, subscales, and the overall SART score were correlated. This indicates both measures are evaluating SA and should be used in future research when assessing SA.

The exploratory factor analysis revealed how each of these measures individually contributes to the underlying structure of SA. The *demand* subscale of SART loaded onto a single factor, whereas the second factor contained items from the *supply* subscale of SART and SPAM query RT. This demonstrates the attentional demand of a situation and how attention is focused during a task is contributes to the structure of SA. Portions of each measure are crucial to understanding SA; it is advantageous to administer both measures in future research.

This measure comparison provided many insights into how different methods evaluate SA. This study identified standardized measures of SA, compared to performance measures, more accurately captures SA. Researchers should consider these findings and use more rigorous and consistent methodology across future studies. Administering standardized measures, such as SPAM and SART, ensures SA is being assessed. Multiple measures of SA should be administered to fully explain SA of a given situation or task until new measures are developed and validated.

4.2 Practical Contributions: Situation Awareness Measurement

This study revealed limitations of some of the SA measures, specifically SPAM. SPAM provides researchers the ability to evaluate SA dynamically, in real-time and separates workload from SA through the use of the ready prompt. However, this study determined there are limits to the interpretability of the ready prompt. Currently, there is no way to distinguish if a participant intentionally rejects a ready prompt or misses one. Both responses are recorded as a miss, which impacts the overall SPAM accuracy score. If a participant intentionally rejects the ready prompt, this does not necessarily indicate poor SA. Rather, intentionally rejecting a ready prompt could provide additional information about a participant's workload. Regardless, the ability to evaluate SA is lost when a ready prompt is not accepted. In high workload situations, using SPAM to evaluate SA could lead to unreliable results. To mitigate this issue, future research should add alternative responses the ready prompt (i.e., yes, no), as well as collect qualitative data if "no" is selected. This information could provide additional insight when evaluating SA dynamically.

Considering the identified limitations of SPAM and the previously acknowledged limitations of SART, it is evident a new standardized measure needs to be developed. This study exhibited parts of each measure are important in our understanding of SA, but individually do not fully evaluate SA. A new standardized measure should consider how to better assess the attentional demand of situation and the focus of attention, as identified in this study. This measure should also leverage the methods of both SPAM and SART to capture objective and subjective assessments of SA. With these factors in mind, the new measure will allow researchers to more comprehensively assess SA.

4.3 Theoretical Contributions: Automation Influence on Situation Awareness

This study identified the impact of low and high automation systems on SA, as well as differences in SA between automation levels. SA decreased throughout the low automation drive, however the trend was not statistically significant. At the start of the drive, participants had high SA, as indicated by SPAM query RT. The short RTs suggest participants had necessary information in their working memory and did not need to search their environment. As the drive continued, driver SA slightly decreased and did not increase back to the level recorded at the beginning of the drive. In this study, low levels of automation did not lead to large changes in SA throughout the drive. This pattern could be due to the participants continued engagement in the driving task. Participants were still required to control and maintain their speed in the low automation drive, which could have helped them from becoming out-of-the-loop (Durso & Gronlund, 1999). The speed maintenance task could have aided drivers in maintaining necessary knowledge and led to insignificant differences in SA throughout the drive.

A different pattern of SA was identified in the high automation drive. A quadratic trend was found where participants had high SA at the beginning and end of the drive, as measured by SPAM query RT. Again, short query RTs suggest participants did not search their environment to find information to answer the query. Participants' SA worsened throughout the middle of the drive, as indicated by longer query RT. The high automation system led to large fluctuations in SA, revealing the negative consequences for SA in highly automated vehicles. Although drivers were able to regain high levels of SA in this study, it is not clear if this is always the case. Further research is necessary to better understand and identify trends in SA while operating a highly automated vehicle for an extended period of time.

Comparing the two levels of automation revealed significant differences in SA. The high automation drive led to significantly higher driver SA compared to the low automation drive. This was evident by both SPAM accuracy scores and average correct query RT. Accuracy was higher during the high automation drive; the average RT was lower in the high automation drive. These results were unexpected, as it was predicted the high automation system would lead to worsened SA. However, previous literature has described how higher levels of automation could lead to increased SA (Parasuraman et al., 2004). As the automation level increases, it removes tasks from the operator, which can cause workload to decrease and SA to increase (Parasuraman et al., 2004). The reported subjective workload was low in this study, so it is plausible the low workload led to increased SA in the high automation drive. It is important to note, although this difference was significant, the overall SA was still poor in both drives. The average accuracy scores were below 68% at each automation level. This level of SA does not ensure the safety of drivers operating automated vehicles. Higher levels of automation may lead to better SA, but further research is needed to understand how to improve SA when operating these systems.

4.4 Practical Contributions: Automation Influence on Situation Awareness

The insights from this study clearly identified automation has negative consequences for driver SA. The degree in which SA worsens is impacted by the level of automation, where higher levels of automation lead to better SA. This finding provides hope for future automation systems; however, it is critical to closely consider the impact low automation systems have on SA. Current automated vehicles on the road are operating at lower levels of automation, using ALK or ACC systems. Additionally, drivers are not trained on how to safely use these systems and are not properly educated on their limitations and capabilities. This lack of information puts drivers are at risk of diminished SA during long drives if they do not constantly supervise the system, and could ultimately result in more fatal accidents. Research must continue investigating low automation systems to ensure drivers are safe before consideration of high or fully automated systems. Especially studies to develop and evaluate systems to help drivers consistently maintain SA, more informative automation displays (e.g., uncertainty displays, reliability displays), and a training for drivers to understand the limitations and capabilities of these systems.

4.5 Predictors of Situation Awareness

This study explored the ability of trust in automation and workload to predict SA. There was no significant relationship found between trust or workload and SA. This research was not able to identify predictors of SA, but it does suggest these constructs may not have a causal relationship with SA. Other constructs should be considered and investigated as predictors of SA, such as confidence in automation or conscientiousness.

4.6 Limitations

This study was designed to accomplish three goals: determine the impact of different automation levels on driver SA, compare various SA measures, and identify predictors of SA. The first two goals were addressed, but the study had a few limitations.

This research evaluated SA of the roadway environment exclusively and was unable to assess SA of the automated system or automation displays. The driving simulator used for this study does not currently have the capabilities to change or add automation displays. The default automation displays only provide state information (if a system is on or off). For these reasons, this study did not evaluate SA of other aspects of the environment. Future research should include queries about the automation system and displays to understand how drivers split their attention.

Lastly, this study did not find any predictors for SA. This could be a result of how trust in automation and workload were measured. Both trust and workload were assessed subjectively at the end of each drive with appropriate, standardized scales. However, measuring trust and workload at the conclusion of the drive meant participants could have referred to their average experience, a particular moment, or a combination when responding to each scale. This could have impacted the relationship between SA, trust, and workload. Future studies should explore the use of real-time measures of trust and workload to understand the relationship between these constructs.

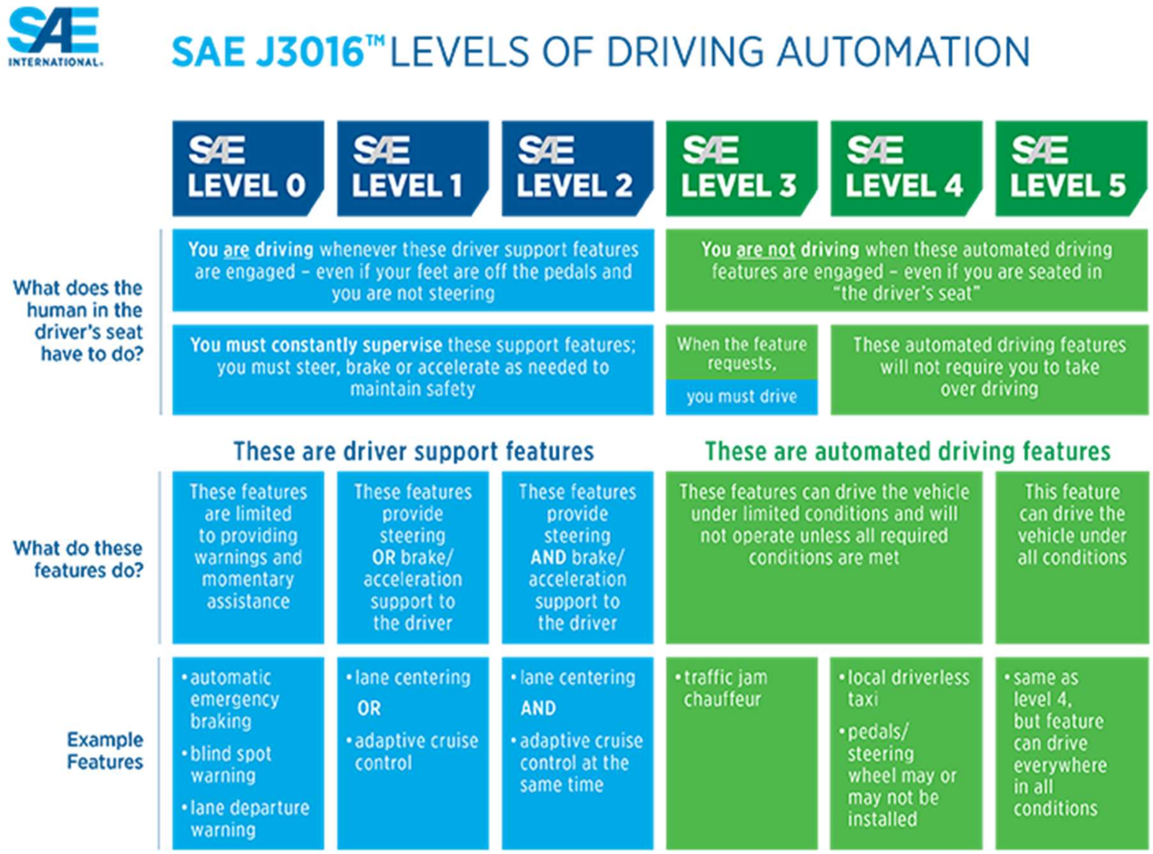
4.7 Future Research

The current study showed operating an automated vehicle is an extremely dynamic environment and there are many factors to consider when studying this domain. There are many paths for future research including the investigation of other contexts within the automated vehicle environment, the evaluation of interventions to assist driver SA, and the development of a framework to describe the impact of different secondary tasks.

Future research should consider assessing SA of the automation system and displays and any secondary task, in addition to the roadway environment. As discussed in the limitations section, this study was unable to explore these other contexts. Utilizing a measure, such as SPAM, to evaluate additional aspects of the environment could provide insights regarding where drivers are naturally attending. With this information, researchers can develop informed interventions to give drivers relevant information to maintain SA throughout a drive. For example, if drivers are mainly attending to the automation system, an intervention could provide information about the roadway to the driver. It is critical to determine how different interventions can improve drivers' ability to maintain SA while operating an automated vehicle. Designing interventions to support driver SA could improve the success of takeover requests and the overall safety of drivers. Evaluation of new interventions or displays should employ a combination of dynamic, query-based assessments and subjective measures of SA.

It is also important to consider the impact of the secondary task on driver SA, among other constructs such as trust, risk, and workload. The current study utilized Tetris since it is a visuospatial task, but it's possible the nature of the secondary tasks could influence how drivers interact with automated vehicles differently. The nature of the secondary task could be influenced by factors, such as consequences for poor performance, benefits or rewards for excelling, or varying levels of difficulty. These factors could affect driver SA, trust, workload, perceived risk, and driving performance in different ways. In order to better understand the complexities of drivers performing secondary tasks while operating an automated vehicle, a new framework should be developed to systematically categorize these secondary tasks. This knowledge not only helps to further understand the impact of different secondary tasks, but could provide insight into the risk for performing each type of task while operating an automated vehicle.

APPENDIX A. SAE AUTOMATION TAXONOMY



Sae.org

APPENDIX B. SITUATIONAL AWARENESS RATING TECHNIQUE (SART)

Based on the drive you just completed, please fill out the following questions:

Instability of Situation

	1 (1)	2 (2)	3 (3)	4 (4)	5 (5)	6 (6)	7 (7)
How changable is the situation? Is the situation highly unstable and likely to change suddenly (High) or is it very stable and straightforward (Low)? (1)	❍	❍	❍	❍	❍	❍	❍

Complexity of Situation

	1 (1)	2 (2)	3 (3)	4 (4)	5 (5)	6 (6)	7 (7)
How complicated is the situation? Is it complex with many interrelated components (High) or is it simple and straightforward (Low)? (1)	❍	❍	❍	❍	❍	❍	❍

Variability of Situation

	1 (1)	2 (2)	3 (3)	4 (4)	5 (5)	6 (6)	7 (7)
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)	❍	❍	❍	❍	❍	❍	❍

Arousal

	1 (1)	2 (2)	3 (3)	4 (4)	5 (5)	6 (6)	7 (7)
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)	❍	❍	❍	❍	❍	❍	❍

Concentration of Attention

	1 (1)	2 (2)	3 (3)	4 (4)	5 (5)	6 (6)	7 (7)
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)	❍	❍	❍	❍	❍	❍	❍

Division of Attention

	1 (1)	2 (2)	3 (3)	4 (4)	5 (5)	6 (6)	7 (7)
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)	❍	❍	❍	❍	❍	❍	❍

Spare Mental Capacity

	1 (1)	2 (2)	3 (3)	4 (4)	5 (5)	6 (6)	7 (7)
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)	❍	❍	❍	❍	❍	❍	❍

Information Quantity

	1 (1)	2 (2)	3 (3)	4 (4)	5 (5)	6 (6)	7 (7)
How much information have you gained about the situation? Have you received or understood a great deal of knowledge (High) or very little (Low)? (1)	❍	❍	❍	❍	❍	❍	❍

Familiarity with Situation

	1 (1)	2 (2)	3 (3)	4 (4)	5 (5)	6 (6)	7 (7)
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)	❍	❍	❍	❍	❍	❍	❍

APPENDIX C. TRUST IN AUTOMATION SCALE

Based on the drive you just completed, please fill out the following questions where 1 = not at all and 7 = extremely.

	1 (1)	2 (2)	3 (3)	4 (4)	5 (5)	6 (6)	7 (7)
The system is deceptive. (1)	❍	❍	❍	❍	❍	❍	❍
The system behaves in an underhanded manner. (2)	❍	❍	❍	❍	❍	❍	❍
I am suspicious of the system's intent, action, or outputs. (3)	❍	❍	❍	❍	❍	❍	❍
I am wary of the system. (4)	❍	❍	❍	❍	❍	❍	❍
The system's actions will have a harmful or injurious outcome. (5)	❍	❍	❍	❍	❍	❍	❍
I am confident in the system. (6)	❍	❍	❍	❍	❍	❍	❍
The system provides security. (7)	❍	❍	❍	❍	❍	❍	❍
The system has integrity. (8)	❍	❍	❍	❍	❍	❍	❍
The system is dependable. (9)	❍	❍	❍	❍	❍	❍	❍
The system is reliable. (10)	❍	❍	❍	❍	❍	❍	❍
I can trust the system. (11)	❍	❍	❍	❍	❍	❍	❍
I am familiar with the system. (12)	❍	❍	❍	❍	❍	❍	❍

APPENDIX D. NASA-TLX DEFINITIONS

Title	Endpoints	Description
Mental Demand	Low/High	How much mental and perceptual activity was required (e.g. thinking, deciding, calculating, remembering, looking, searcihing, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
Physical demand	Low/High	How much physical activity was required (e.g. pushing, pulling, turning, controlling, activating, ect.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
Temporal demand	Low/High	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
Performance	Good/Poor	How successful do you think you were accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
Effort	Low/High	How hard did you have to work (mentally and physically) to accomplish your level of performance?
Frustration level	Low/High	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

APPENDIX E. SPAM QUERIES

Low Automation Drive:

Past

1. What highway did you just pass?
2. What color was the car going in the opposite direction?

Present

1. How many cars are in your rearview mirror?
2. How many cars are on the shoulder?

Future

1. Where will the *orange* car merge? (In front of me; **behind me**)
2. How long will it take to get to XX? (i.e., the highway, a particular street, construction zone)

High Automation Drive:

Past

1. What information was on the billboard you just passed? (**antiques,** attorney, restaurant, tourist attraction)
2. What type of building/business did you pass? (**gas station**)

Present

1. How many cars are currently in front of you?
2. What zone are you in? (**construction**, school, passing, animal crossing)

Future

1. Where will the *orange* car merge? (In front of me; **behind me**)
2. How long will it take to get to XX? (i.e., the highway, a particular street, construction zone)

APPENDIX F. SECONDARY TASK ENGAGEMENT INSTRUCTIONS

During the drive you will again have the option to engage in a game on the side tablet screen. You are only to play the game if you feel it is safe to do so. Your primary task is to drive safely. A screen with a “Play” button will be presented and will remain on the screen unless you decide it is safe to play the game. If you press the “Play” button, the game will begin.

The game presented on the side tablet is Tetris. During the game, different shapes will slowly fall from the top of the screen. Your goal is to complete as many rows as possible with the shapes. Once you complete a row, it will disappear and you will earn points. To control the where the shapes fall use the “left” and “right” buttons on the right side of the screen; this will control the shape’s left and right movement. You may also rotate each shape clockwise, in 90-degree increments using the “rotate” button. If the screen fills with shapes, it will automatically reset. Do you have any questions?

APPENDIX G. CONSENT FORM

Consent to be a Research Participant, GT School of Psychology

Project:	Situation Awareness in Automated Vehicles
Principal Investigator:	Dr. Bruce N. Walker (404-894-8265)
Experimenters:	Georgia Tech Students
Location:	J.S Coon
Duration of each Session:	1.5 hours
Number of Sessions:	1.0
Total Compensation:	**1**.5 credit hours (if students)
Number of Participants:	100

Participation Limitations: Participants must have at least two years of driving experience with a full driver's license to participate in this study. Participants must have normal or corrected-to-normal vision and hearing.

Study Description: You are being asked to participate in a research study. Participation is voluntary and may be discontinued at any time without penalty. The goal of this study is to understand how a person's awareness of their driving environment changes while driving an automated vehicle. We will be taking a variety of measurements, such as eye movement, driving performance, your awareness of the road conditions, and task workload.

Procedures: You will be taken through a simulator sickness screening to make sure the driving simulator will not cause you physical discomforts. When you finish the sickness screening, you will be asked to complete two driving tasks. You will be asked to fill out questionnaires throughout the study. At the conclusion of the experiment, you will be debriefed on our study, and released.

Compensation: If you are taking part in this study as part of SONA systems you will receive compensation of 0.5 credits per half hour of participation. If you should choose to withdraw early, you will still receive credit. All other participants will not be compensated.

Costs to You: There are no costs to you, other than time, for being in this study. There are no anticipated risks associated with your participation in this study.

Benefits: You are not likely to benefit in any way from joining this study. We hope that what we learn will contribute to our understanding of situation awareness while operating an automated vehicle.

Confidentiality: The following procedures will be followed to keep your personal information confidential in this study: The data that is collected about you will be kept private to the extent required by law. Identifiable information will only be collected from Georgia Tech students who are receiving course credit through Psychology courses. After course credit has been awarded for these individuals, the identifiable information will be removed from the data set. To protect your privacy, your records will be kept under a code number rather than by name. Your records will be kept in password protected files and only study staff will be allowed to look at them. Your name and any other fact that might point to you will not appear when results of this study are presented or published. To make sure that this research is being carried out in the proper way, the Georgia Institute of Technology IRB may review study records. The Office of Human Research

Protections may also look over study records during required reviews. Again, your privacy will be protected to the extent required by law.

Injury/Adverse Reaction: Reports of injury or reaction should be made to Dr. Bruce Walker (404-894-8265). Neither the Georgia Institute of Technology nor the principal investigator has made provision for payment of costs associated with any injury resulting from participation in this study.

Contact Persons: If you have questions about this research, call or write Dr. Bruce Walker at 404-894-8265; School of Psychology, GA Tech, 654 Cherry Street, Atlanta, GA 30332-0170.

Participant Rights: Taking part in this study is completely voluntary. You have the right to change your mind and leave the study at any time without giving any reason and without penalty. Any new information that may make you change your mind about being in this study will be given to you. You may print out a copy of this consent form to keep. You do not waive any of your legal rights by participating in this study. If you have any questions about your rights as a research volunteer, call or write: The Institutional Review Board, Office of Research Integrity Assurance, 505 Tenth Street, Campus 0420. Phone: 404-385-2175; Fax: 404-385-2081.

Signatures: A copy of this form will be given to you. By signing your name below, you are consenting to participate in this research.

Participant's Signature:

____

Date: ______________________

Person Obtaining Consent:

____

Date: ______________________

APPENDIX H. DEMOGRAPHICS QUESTIONNAIRE

1. Do you have a driver's license?
    a. Yes (If yes, how many years have you held a license for?)
    b. No (If no, the following message appears: "Sorry, you are not eligible for the study. Please see the researcher.")
2. How many years have you held a license for?
    a. Less than two years (If no, the following message appears: "Sorry, you are not eligible for the study. Please see the researcher.")
    b. Two years or more
3. How many years have you held your license for?
4. On average, how many hours do you drive each week when you're on campus?
5. On average, how many hours do you drive each week when you're not on campus?
6. What is your age?
7. What is your gender?
    a. Male
    b. Female
    c. Other
    d. Choose not to identify
8. What is your primary language?
9. What other languages do you speak?

10. What is your level of familiarity with automated safety features such as automated lane keeping? Automated lane keeping systems automatically steer the vehicle to maintain position within a lane.
    a. I own a vehicle with one or more automated safety features
    b. I have driven a vehicle with one or more automated safety features
    c. I have been a passenger in a vehicle with one or more automated safety features
    d. I am familiar with automated safety features
    e. I have never heard of automated safety features prior to this study
11. What is your level of familiarity with automated safety features such as automated lane keeping? Automated lane keeping systems automatically steer the vehicle to maintain position within a lane.
    a. I own a vehicle with one or more automated safety features
    b. I have driven a vehicle with one or more automated safety features
    c. I have been a passenger in a vehicle with one or more automated safety features
    d. I am familiar with automated safety features
    e. I have never heard of automated safety features prior to this study

APPENDIX I. PARTICIPANT DRIVING INSTRUCTIONS

<u>Driving Task 1:</u>

<u>The Drive</u>
The driving course will last about 20 minutes. For this drive, we would like you to maintain a speed of 55 miles per hour for the duration of the drive to the best of your ability.

- For participants in the **low automation condition (ALK only):**
  - During this drive you will be using automated lane keeping. Automated lane keeping systems keep the vehicle in the center of the lane so that you do not have to steer. Putting your hands on the steering wheel and turning it even a small amount can turn off the automated lane keeping system. To avoid doing so inadvertently, please keep your hands off of the steering wheel unless you feel that you must take control of the vehicle to avoid an accident.
  - You will know that the automated lane keeping system is on if the green, nearly parallel lines are present in the dashboard. If this display is not present, the automated lane keeping system is not on.

- For participants in the **high automation condition (ALK and ACC):**
  - During this drive you will be using automated lane keeping and adaptive cruise control. Automated lane keeping systems keep the vehicle in the center of the lane so that you do not have to steer. Adaptive cruise control systems maintain a preset speed and adjust based on vehicles directly in front of you so that you do not have to accelerate or decelerate. Putting your hands on the steering wheel and turning it even a small amount can turn off the automated lane keeping system. Similarly, if you press either the gas or brake pedal, the adaptive cruise control system will turn off. To avoid doing so inadvertently, please keep your hands off of the steering wheel and your foot off the pedals unless you feel that you must take control of the vehicle to avoid an accident.
  - You will know that the automated lane keeping system is on if the green, nearly parallel lines are present in the dashboard. Similarly, you will know if the adaptive cruise control system is on if a car icon is present in the dashboard. If these displays are not present, the automated lane keeping system is not on.

During the drive you will have the option to engage in a game on the side tablet screen; the game is Tetris. You are only to play the game if you feel it is safe to do so. Your primary task is to drive safely. A screen with a “Play” button will be presented and will remain on the screen unless you decide it is safe to play the game. If you press the “Play” button, the game will begin. In Tetris, different shapes fall from the top of the screen and your goal is to completely fill a row. Once you fill a row, it disappears. If you run out of space, then the game is over and a new round will start. The control buttons will be on the right side of the screen. Periodically, you will be presented questions on the tablet screen. You will know a question is being presented when the “Ready” button appears on the tablet. If you are ready to answer the question, you will simply press the “Ready” button and answer the question on the tablet. After you answer the question, the “Play” screen for Tetris will reappear. Do you have any questions?

Please complete the drive as safely as possible.

Completion
Upon completion, we will stop the driving scenario and present you with questionnaires to complete prior to beginning the second drive.

Driving Task 2:

The Drive
The driving course is similar to the last drive you just completed. For this drive, we would again like you to maintain a speed of 55 miles per hour for the duration of the drive to the best of your ability. You will also be using the automated lane keeping system in this drive.

- For participants in the **low automation condition (ALK only):**
    - During this drive you will be using automated lane keeping. Automated lane keeping systems keep the vehicle in the center of the lane so that you do not have to steer. Remember, putting your hands on the steering wheel and turning it even a small amount can turn off the automated lane keeping system. To avoid doing so inadvertently, please keep your hands off of the steering wheel unless you feel that you must take control of the vehicle to avoid an accident.
    - You will know that the automated lane keeping system is on if the green, nearly parallel lines are present in the dashboard. If this display is not present, the automated lane keeping system is not on.

- For participants in the **high automation condition (ALK and ACC):**
    - During this drive you will be using automated lane keeping and adaptive cruise control. Automated lane keeping systems keep the vehicle in the center of the lane so that you do not have to steer. Adaptive cruise control systems maintain a preset speed and adjust based on vehicles directly in front of you so that you do not have to accelerate or decelerate. Remember, putting your hands on the steering wheel and turning it even a small amount can turn off the automated lane keeping system. Similarly, if you press either the gas or brake pedal, the adaptive cruise control system will turn off. To avoid doing so inadvertently, please keep your hands off of the steering wheel and your foot off the pedals unless you feel that you must take control of the vehicle to avoid an accident.
    - You will know that the automated lane keeping system is on if the green, nearly parallel lines are present in the dashboard. Similarly, you will know if the adaptive cruise control system is on if a car icon is present in the dashboard. If these displays are not present, the automated lane keeping system is not on.

During the drive you will again have the option to engage in a game on the side tablet screen. You are only to play the game if you feel it is safe to do so. Your primary task is to drive safely. A screen with a “Play” button will be presented and will remain on the screen unless you decide it is safe to play the game. If you press the “Play” button, the game will begin. Periodically, you will be presented questions on the tablet screen. You will know a question is being presented when the “Ready” button appears on the tablet. If you are ready to answer the question, you will simply press the “Ready” button and answer the question on the tablet. After you answer the question, the “Play” screen for Tetris will reappear. Do you have any questions?

Please complete the drive as safely as possible.

Completion

Upon completion, we will stop the driving scenario and present you with the final set of questionnaires to complete.

APPENDIX J. PARTICIPANT INSTRUCTIONS

Thanks and Introduction
First of all, thank you for your participation in this study. We are members of Sonification Lab in school of psychology.

Purpose of Experiment
This research is investigating how situation awareness changes as automation increases in vehicles.

Procedure

Consent
The consent form presented to you is to inform you of the content of this experiment. Please read through it, and ask any questions you have before you sign it. During the experiment, please let us know if you have questions, concerns, discomforts, or would like to withdraw from the experiment. You can do so without penalty.

General Instructions
Before this experiment, we will ask you to complete a simulator sickness screening first. This is to ensure that you do not encounter any motion sickness during the experiment. Then you will be asked to complete a set of questionnaires. Next, you will complete the first of two driving scenarios which will be followed by another set of questionnaires. The second drive will follow with a final set of questionnaires. The session should last no longer than an hour and a half, and the experimenter(s) will help you throughout the session.

Sim Sickness Screening
To make sure the driving simulator will not cause you any physical discomfort, we will conduct a screening procedure. This procedure includes a pre-drive survey, a short drive, and a post-drive survey. If for any reason, you feel sick during the procedure, this session will end and you will receive full credit for your time here.

Questionnaires and Driving Tasks
We will be collecting data during two separate driving tasks. Both driving courses will be about 20 minutes in length. Each drive will have its own set of instructions that the experimenters will go over with you before the drive. We will be collecting a number of measurements, such as eye movement, driving performance, situation awareness, and workload. The driving scenario will be similar for both of the driving tasks.

Debrief
Once the final set of questionnaires is completed after the second drive, you will be debriefed on the experiment and released. We will then assign you credit for your participation.

APPENDIX K. DEBRIEF FORM

Thanks and Introduction
First of all, thank you for your participation in this experiment. We are members of Sonification Lab in the School of Psychology.

Purpose of Experiment
The purpose of this experiment was to investigate how situation awareness is affected by different levels of automation. In each of the two drives, we measured your eye movements, pupil size, driving performance, workload, awareness of the driving environment, trust in the automated system, and feelings toward the automated system.

Meaning of Expected Results
We expect that as automation increases participants' situation awareness, their knowledge of the driving environment, will worsen. We expect that analysis of eye movements, driving performance, workload, awareness of the driving environment, and trust will help identify what information participants attend to in highly automated vehicles and how it influences awareness of the driving environment. These results will be used to establish guidelines for the design of displays for automated driving.

Confidentiality and Anonymity
The results of your experiment will be used for only psychological study and never used for any other purposes. The data that is collected from you will be kept private to the extent required by law. To protect your privacy, your records will be kept under a code number rather than by name. Your records will be kept in locked files and only research staffs will be allowed to look at them. Your name and any other fact that might point to you will not appear when results of this study are presented or published. To make sure that this research is being carried out in the proper way, the Georgia Institute of Technology IRB will review study records. Again, your privacy will be protected to the extent required by law.

Conclusion
All of the experiment procedures are finished. We very much appreciate your efforts again.

Contact Information
For further information of this research, contact:

Principal Investigator
Dr. Bruce Walker (bruce.walker@psych.gatech.edu)

Experimenters
Zoe Becerra (zbecerra3@gatech.edu)
Sanghavi Gaddam (sanghavig@gatech.edu)

Sahar Ali (saharnazimali@gmail.com)
Brittany Noah (brittany.noah@gatech.edu)

APPENDIX L. STUDY RECRUITMENT

Recruitment text:

The Sonification Lab at Georgia Tech is looking for research participants to complete a study about driving automated vehicles. Participants must have normal or corrected to normal vision, mobility, and hearing and have 2 years minimum of driving experience and a valid license. Participation in this survey is voluntary. Georgia Tech students completing this study through SONA will receive 1.5 credits.

REFERENCES

Becerra, Z. M., Holthausen, B. E., & Walker, B. N. (2019a). Bringing the Thrill to Automated Vehicles: An Evaluation of Thrill- Seeking Driving Displays. In *Proceedings of the 11th International ACM Conference on Automotive User Interfaces and Interactive Vehicular Application (AutomotiveUI 2019)* (pp. 381–386). Utrect, Netherlands. https://doi.org/https://doi.org/10.1145/3349263.3351328

Becerra, Z. M., Holthausen, B. E., & Walker, B. N. (2019b). Designing displays in automated vehicles for the thrill-seeking driver: Characterizing driving profiles to create personalized driving displays. *In Technology, Mind, and Society*.

Beller, J., Heesen, M., & Vollrath, M. (2013). Improving the driver-automation interaction: An approach using automation uncertainty. *Human Factors*, *55*(6), 1130–1141. https://doi.org/10.1177/0018720813482327

Davis, J. H., Mayer, R. C., & Schoorman, F. D. (1995). An Integrative Model of Organizational Trust, *20*(3), 709–734.

De Winter, J. C. F., Happee, R., Martens, M. H., & Stanton, N. A. (2014). Effects of adaptive cruise control and highly automated driving on workload and situation awareness: A review of the empirical evidence. *Transportation Research Part F: Traffic Psychology and Behaviour*, *27*(PB), 196–217. https://doi.org/10.1016/j.trf.2014.06.016

Durso, F. T., & Dattel, A. R. (2004). SPAM: The real-time assessment of SA. In S. Banbury & S. Tremblay (Eds.), *A Cognitive approach to situation awareness: theory and application* (pp. 137–154). Aldershot, UK: Ashgate.

Durso, F. T., & Gronlund, S. D. (1999). Situation Awareness. In F. T. Durso, R. S. Nickerson, R. W. Schvaneveldt, S. T. Dumais, D. S. Linday, & M. T. H. Chi (Eds.), *Handbook of Applied Cognition* (pp. 283–314). John Wiley & Sons Ltd. https://doi.org/10.1002/9781118131350

Endsley, M. R. (1988). SITUATION AWARENESS GLOBAL ASSESSMENT TECHNIQUE (SAGAT). In *Proceedings of the IEEE 1988 National Aerospace and Electronics Conference* (pp. 789–795).

Endsley, M. R. (1995a). Measurement of Situation Awareness in Dynamic Systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(1), 65–84. https://doi.org/10.1518/001872095779049499

Endsley, M. R. (1995b). Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors*, *37*(1), 32–64. https://doi.org/10.4324/9781315092898-13

Endsley, M. R. (1996). Automation and Situation Awareness. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp. 163–181). Mahwah, NJ: Lawrence Erlbaum.

Endsley, M. R. (2012). Situation Awareness. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics* (pp. 553–568). John Wiley & Sons, Incorporated.

Endsley, M. R. (2017a). Autonomous Driving Systems: A Preliminary Naturalistic Study of the Tesla Model S. *Journal of Cognitive Engineering and Decision Making*, *11*(3), 225 –238. https://doi.org/10.1177/1555343417695197

Endsley, M. R. (2017b). From Here to Autonomy: Lessons Learned from Human-Automation Research. *Human Factors*, *59*(1). https://doi.org/10.1177/0018720816681350

Gable, T. M., Tomlinson, B. J., Cantrell, S., & Walker, B. N. (2017). Spindex and Spearcons in Mandarin: Auditory Menu Enhancements Successful in a Tonal Language. *23rd International Conference on Auditory Display (ICAD 2017)*, (Icad), 27–34. https://doi.org/10.21785/icad2017.025

Gable, T. M., & Walker, B. N. (2013). Georgia Tech Simulator Sickness Screening Protocol. *Georgia Institute of Technology Technical Report*.

Gugerty, L. J. (2011). Situation Awareness in Driving. *Handbook of Driving Simulation for Engineering, Medicine, and Psychology*, 265–272. https://doi.org/10.1201/b10836-20

Gugerty, L. J. (1997). Situation Awareness during Driving: Explicit and Implicit Knowledge in Dynamic Spatial Memory. *Journal of Experimental Psychology: Applied*, *3*(1), 42–66. https://doi.org/10.1037/1076-898X.3.1.42

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index):

Results of empirical and theoretical research. *Advances in Psychology*, *52*, 139–183.

Helldin, T., Falkman, G., Riveiro, M., & Davidsson, S. (2013). Presenting system uncertainty in automotive UIs for supporting trust calibration in autonomous driving. *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '13*, 210–217. https://doi.org/10.1145/2516540.2516554

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*, *57*(3), 407–434. https://doi.org/10.1177/0018720814547570

Jian, J., Bisantz, A.M., Drury, C. G. (2000). Foundations for an Empirically Determined Scale of Trust in Automated Systems. *Intenational Journal of Cognitive Ergonomics*, *4*(1), 53–71.

Jones, D. G. (2000). Subjective Measures of Situation Awareness. In M. R. Endsley & D. J. Garland (Eds.), *Situation awareness analysis and measurement* (pp. 113–128). CRC Press.

Kaber, D., Jin, S., Zahabi, M., & Pankok, C. (2016). The effect of driver cognitive abilities and distractions on situation awareness and performance under hazard conditions. *Transportation Research Part F: Traffic Psychology and Behaviour*, *42*, 177–194. https://doi.org/10.1016/j.trf.2016.07.014

Kass, S. J., Cole, K. S., & Stanny, C. J. (2007). Effects of distraction and experience on situation awareness and simulated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*. https://doi.org/10.1016/j.trf.2006.12.002

Kunze, A., Summerskill, S. J., Marshall, R., & Filtness, A. J. (2019). Automation transparency: implications of uncertainty communication for human-automation interaction and interfaces. *Ergonomics*, *62*(3), 345–360. Retrieved from https://doi.org/10.1080/00140139.2018.1547842

Lee, J. D., & See, K. A. (2004). Trust in Automation: Designing for Appropriate Reliance. *The Journal of the Human Factors and Ergonomics Society*, *46*(1), 50–80. https://doi.org/https://doi.org/10.1518/hfes.46.1.50.30392

Loft, S., Bowden, V., Braithwaite, J., Morrell, D. B., Huf, S., & Durso, F. T. (2015). Situation awareness measures for simulated submarine track management. *Human*

Factors, *57*(2), 298–310. https://doi.org/10.1177/0018720814545515

Loft, S., Morrell, D. B., & Huf, S. (2013). Using the situation present assessment method to measure situation awareness in simulated submarine track management. *International Journal of Human Factors and Ergonomics*, *2*(1), 33–48. https://doi.org/10.1504/ijhfe.2013.055975

Lu, Z., Coster, X., & de Winter, J. (2017). How much time do drivers need to obtain situation awareness? A laboratory-based study of automated driving. *Applied Ergonomics*, *60*. https://doi.org/10.1016/j.apergo.2016.12.003

Ma, R., & Kaber, D. B. (2005). Situation awareness and workload in driving while using adaptive cruise control and a cell phone. *International Journal of Industrial Ergonomics*. https://doi.org/10.1016/j.ergon.2005.04.002

Ma, R., & Kaber, D. B. (2007). Situation awareness and driving performance in a simulated navigation task. *Ergonomics*, *50*(8), 1351–1364. https://doi.org/10.1080/00140130701318913

May, K. R., Noah, B. E., & Walker, B. N. (2017). Multimodal Heads Up Displays to Augment Autonomous Vehicle Supervision. *In Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications Adjunct - AutomotiveUI '17*, 246–246. https://doi.org/https://doi.org/10.1145/3131726.3131877

Merat, N., & Jamson, A. H. (2009). Is Drivers ' Situation Awareness Influenced by a Fully Automated Driving Scenario ? In *Human Factors, Security and Safety. Human Factors and Erogonomics Society Europe Chapter Conference*. Soesterberg, the Netherlands: Shaker Publishing.

Merat, N., Jamson, A. H., Lai, F. C. H., & Carsten, O. (2012). Highly automated driving, secondary task performance, and driver state. *Human Factors*, *54*(5), 762–771. https://doi.org/10.1177/0018720812442087

Noah, B. E. (2018). *Understanding Automation Handoff Impacts on Workload and Trust When Mitigated By Reliability Displays*. Georgia Institute of Technology. Retrieved from http://hdl.handle.net/1853/60205

Noah, B. E., & Walker, B. N. (2017). Trust Calibration through Reliability Displays in Automated Vehicles. *Proceedings of the Companion of the 2017 ACM/IEEE*

International Conference on Human-Robot Interaction - HRI '17, 361–362. https://doi.org/10.1145/3029798.3034802

Noah, B. E., & Walker, B. N. (2018). Personalized Driving Displays of Vehicle Automation to Improve Trust and Acceptance. *Proc. of CHI 2018 Workshop - Interacting with Autonomous Vehicles: Learning from Other Domains*. Retrieved from https://hci.sbg.ac.at/sites/autonomous-vehicles-chi2018/

Parasuraman, R., & Riley, V. (1997). Humans and Automation : Use , Misuse , Disuse , Abuse. *Human Factors*, *39*(2), 230–253. https://doi.org/https://doi.org/10.1518/001872097778543886

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A Model for Types and Levels of Human Interaction with Automation. *IEEE Transaction on Systems, Man, and Cybernetics- Part A: Systems and Humans*, *30*(3), 286–297.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2004). 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.

Pierce, R. S. (2012). The effect of SPAM administration during a dynamic simulation. *Human Factors*, *54*(5), 838–848. https://doi.org/10.1177/0018720812439206

Schömig, N., & Metz, B. (2013). Three levels of situation awareness in driving with secondary tasks. *Safety Science*, *56*, 44–51. https://doi.org/10.1016/j.ssci.2012.05.029

Sheridan, T. B. (2012). Human Supervisory Control. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics* (pp. 990–1015). John Wiley & Sons, Incorporated.

Taylor, R. M. (1990). Situational Awareness Rating Techinque (SART): The development of a tool for aircrew systems design. In *Situational Awareness* (pp. 111–128). Routledge: Routledge.

Turner, H., Young, M. S., Walker, G. H., Stanton, N. A., & Randle, S. (2004). Automating the Driver's Control Tasks. *International Journal of Cognitive Ergonomics*, *5*(3), 221–236. https://doi.org/10.1207/s15327566ijce0503_5

US Department of Transportation. (2018). *Preparing for the Future of Transportation*.

Walker, G. H., Stanton, N. A., & Young, M. S. (2008). Feedback and driver situation awareness (SA): A comparison of SA measures and contexts. *Transportation Research Part F: Traffic Psychology and Behaviour*, *11*(4), 282–299. https://doi.org/10.1016/j.trf.2008.01.003

Wickens, C. (2004). Stress and Workload. In C. D. Wickens, S. E. Gordon, & Y. Liu (Eds.), *An introduction to human factors engineering* (2nd ed., pp. 324–350). Pearson Education, Inc.,.

Wintersberger, P., Sawitzky, T. von, Frison, A.-K., & Riener, A. (2017). Traffic Augmentation as a Means to Increase Trust in Automated Driving Systems. https://doi.org/https://doi.org/10.1145/3125571.3125600

Young, K. L., Salmon, P. M., & Cornelissen, M. (2013). Missing links? The effects of distraction on driver situation awareness. *Safety Science*, *56*, 36–43. https://doi.org/10.1016/j.ssci.2012.11.004