Summer 2019

Enhancing situational awareness in highly automated vehicles through driver monitoring

Omeed Kashef
University of Iowa

Follow this and additional works at: https://ir.uiowa.edu/etd
Part of the Industrial Engineering Commons

ENHANCING SITUATIONAL AWARENESS IN HIGHLY AUTOMATED VEHICLES THROUGH DRIVER MONITORING

by

Omeed Kashef

A thesis submitted in partial fulfillment
of the requirements for the
Master of Science
degree in Industrial Engineering in the
Graduate College of
The University of Iowa

August 2019

Thesis Supervisor: Associate Professor Daniel McGehee

“Reason is, and ought to be the slave of the passions, and can never pretend to any other office than to serve and obey them.”

- David Hume
Treatise of Human Nature

ACKNOWLEDGEMENTS

I would like to thank my advisors Dr. Daniel McGehee and Dr. John Gaspar for their guidance throughout this process as well as the faculty and staff at the National Advanced Driving Simulator for providing me an invaluable experience at the University of Iowa.

I'd also like to thank Dr. Stephen Baek for his mentorship throughout the years and exceptional insight into the field of Human-Machine Interaction.

I'd like to give a special thanks to Dr. Chris Schwarz for his mentorship throughout the research project and for introducing me to the faculty and staff at the National Advanced Driving Simulator. Furthermore, I'd like to thank all the Industrial and Systems Engineering faculty for their contributions to my educational development over the past 2 years.

I would like to thank Toyota Collaborative Safety Research Center, SaferSim, and the National Advanced Driving Simulator for affording me the opportunity to pursue a graduate degree at the University of Iowa.

Lastly, I'd like to thank Mom, Dad, and Brother for their love, support, and patience throughout this journey.

ABSTRACT

With the development of level 3 AVs, drivers can now disengage from the driving task for extended periods of time. However, drivers are still responsible for the overall safety of their drive. Moreover, when drivers are not engaged in their monitoring task, they lose situational awareness. This leaves drivers vulnerable when they have to retake control from the AV. This research looks to advance the development of camera-based driver monitoring systems that measure situational awareness. In addition, this research examines the effect of adaptable warning systems on driver situational awareness and takeover performance.

In this study, we use situational awareness as ground truth to compare adaptable warning systems that reengage drivers in the monitoring task. Camera-based driver monitoring systems that measure gaze behavior can be used to adapt warning systems. Twenty-four participants split into three groups were asked to drive for approximately 40 miles in a level 3 AV simulator while completing a visual-manual secondary task. During the drive, participants experienced four events in which they had to disengage from the secondary task and take back control from the AV. Two interface designs based on gaze behavior were compared to a baseline warning system. The Attentional Maintenance group was given an alert throughout the drive after a fixed amount of time in which their gaze was directed away from the road. The State-Contingent Takeover group was given an alert only before takeover events after a fixed amount of time in which their gaze was directed away from the road. Results show that attentional maintenance alerts can increase situational awareness and takeover response time during automation failure. Future research to increase situational awareness is discussed in terms of advancements in cognitive control and bilateral communication between the driver and the AV.

PUBLIC ABSTRACT

Drivers in automated vehicles (AVs) have a tendency to look away from the road and driving environment in favor of other tasks such as texting. With the development of highly automated vehicles, drivers can now take their hands off the wheel and eyes off the road for extended periods of time. This leaves drivers vulnerable if the AV fails. However, drivers are still responsible for the overall safety of their drive. Recent developments in camera-based driver monitoring systems can be used to help drivers safely take back control during automation failure.

In this study, we split drivers into three different groups. Each group was assigned a warning system. Two groups received alerts from the camera-based driver monitoring system based on their gaze behavior. The Baseline group did not receive alerts based on their gaze behavior. The Attentional Maintenance group was given an alert throughout the drive after a fixed amount of time in which their gaze was directed away from the road. The State-Contingent Takeover group was given an alert only before takeover events after a fixed amount of time in which their gaze was directed away from the road. Results show that attentional maintenance alerts can improve awareness of the driving situation and response time during automation failure.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 – INTRODUCTION

On a clear, dry day in May, 2016, in Williston, Florida, a Tesla collided with a tractor-trailer, killing the Tesla driver. This was the first death in the U.S. by an automated vehicle (AV). The National Transportation Safety Board (NTSB) examined the data to determine the potential causes. They found that the Tesla was operating in Autopilot mode, which includes adaptive cruise control and lane-keeping systems. Lane-keeping systems keep the car within its lane of travel. Adaptive cruise control maintains a specific distance away from a lead vehicle by applying brakes and accelerating/decelerating to a set cruise speed when there is no obstruction in front of the vehicle. In addition, the instrument panel shows icons that represent activity levels of the adaptive cruise control and lane-keeping systems. However, on this occasion, the Tesla's sensors were not able to detect a white tractor trailer against a bright, sunny background (NTSB, 2017). More importantly, the driver was rarely engaged in the driving task according to the vehicle performance data. The data showed that for the 37 minutes Autopilot was active, the driver applied torque on the steering wheel for only a total of 25 seconds. There was no driver interaction with Autopilot, no change in steering angle, and no brakes applied for 1 minute 51 seconds before collision, which suggests the driver may have been distracted.

More recently, another Tesla in Autopilot mode crashed on March 23, 2018 in Mountain View, California, which also resulted in the death of the driver. The National Transportation Safety Board released a preliminary report that cited the driver was inattentive according to the crash data. Autopilot mode was engaged for approximately 19 minutes before crashing into an impact attenuator (NTSB, 2018a). Although the driver had his hands on the wheel for 34 seconds of the 60 seconds prior to the crash, the driver did not have his hands on the wheel during the 6 seconds before impact. The Tesla increased speed from 62 mph to 70.8mph within 3 seconds up to the time of impact and no pre-crash braking or evasive steering movement was detected.

In both these crashes, it is unclear whether the driver was engaged in a non-driving related task. It is possible that automation of the driving task led to disengagement from safe driving behavior and monitoring complacency, making it much more difficult for the drivers to take back control during these hazardous situations.

Poor monitoring behavior due to complacency was more clearly captured the night of March 18, 2018 in Tempe, Arizona when a distracted driver in a highly automated Uber struck a woman walking her bicycle across the street. This was the first ever pedestrian death involving an automated vehicle (NTSB, 2018b). It was a pitch-black night and the in-vehicle video showed the driver looking away from the road toward the self-driving system interface. According to Uber, the self-driving system relies on an attentive operator to monitor the system's interface as well as to intervene if the system fails to perform appropriately. The driver engaged in the steering wheel less than a second before impact and began braking less than a second after already hitting the woman walking her bicycle. It appears that because automation was responsible for the majority of the driving task, the monitoring job of the driver became more difficult.

Tomorrow's vehicles will increase the monitoring responsibilities of drivers as driving tasks become more automated. The Society of Automotive Engineers (SAE) outlines the five levels of vehicle automation shown in Table 1, which highlight the changing roles as automation increases. Level 1 AVs control either lateral or longitudinal tasks such as cruise control, but never both simultaneously. Level 2 AVs, such as the Tesla in Autopilot mode, execute both lateral and longitudinal control and allow for temporary hands-free driving, which can be immediately disengaged upon the driver's request. Unlike level 3 and higher, in level 2 AVs the driver is solely responsible for detecting and responding to objects and events. Human factors engineers are now working to understand system requirements to implement level 3 automated vehicles.

Table 1. Levels of Automated Vehicles adapted from SAE (2018).

Level	Name	Narrative Definition	Lateral and Longitudinal Vehicle Control	Monitoring of Driving Environment	Fallback Controller of Driving Task	ODD/System Limits
0	No Automation	The performance by the driver of the entire driving task even when enhanced by active safety systems	Driver	Driver	Driver	N/A
1	Driver Assistance	The *sustained* and *ODD*-specific execution by a *driving automation system* of either the *lateral* or the *longitudinal vehicle motion control* subtask of the DDT (but not both simultaneously) with the expectation that the *driver* performs the remainder of the *DDT*.	Driver and System	Driver	Driver	Limited
2	Partial Driving Automation	The *sustained* and *ODD*-specific execution by a *driving automation system* of both the *lateral* and *longitudinal vehicle motion control* subtasks of the *DDT* with the expectation that the *driver* completes the *OEDR* subtask and *supervises* the *driving automation system*.	**System**	Driver	Driver	Limited
3	Conditional Driving Automation	The *sustained* and *ODD*-specific performance by an *ADS* of the entire DDT with the expectation that the *DDT fallback-ready user* is *receptive* to *ADS*-issued *requests to intervene*, as well as to *DDT performance- relevant system failures* in other *vehicle* systems, and will respond appropriately.	System	**System**	Driver	Limited
4	High Driving Automation	The *sustained* and *ODD*-specific performance by an *ADS* of the entire *DDT* and *DDT fallback* without any expectation that a *user* will respond to a *request to intervene*.	System	System	**System**	Limited
5	Full Driving Automation	The *sustained* and unconditional (i.e., not *ODD*- specific) performance by an *ADS* of the entire *DDT* and *DDT fallback* without any expectation that a *user* will respond to a *request to intervene*.	System	System	System	**Unlimited**

Because level 3 AVs free drivers from the responsibility of detecting objects and events, the role of the driver changes from level 2 to level 3. In level 3 and higher AVs, such as the Cadillac when Super Cruise is engaged, automation is sustained long enough to complete the driving task within its operational design domain (ODD). The driver is not responsible for monitoring the environment within the ODD. The ODD is defined as operating conditions under which a given driving automation system, or feature thereof, is specifically designed to function. This includes, but is not limited to, environmental, geographical, and time-of-day restrictions, and/or the requisite presence or absence of certain traffic or roadway characteristics. For example, the Cadillac Super Cruise, a level 3 AV, should not be used when lane markings are poor, in tunnels or construction zones, or during adverse weather conditions such as rain or snow.

Under certain conditions in level 3 AVs, the driver can remove his hands from the wheel, his feet off the pedals, and his eyes from the road for extended periods of time. When the AV falls outside its ODD, automation should prompt the driver who is then expected to serve as the fallback controller. To prompt the driver in a level 3 AV, NTSB suggests using eye tracking technology (NTSB, 2017). They argue that because drivers in level 3 AVs have their hands off the wheel for extended periods of time, a system that uses the steering wheel to monitor the state of the driver is highly ineffective. In order to prompt the driver in a level 3 AV, driver monitoring systems that use head and eye tracking to measure drivers' attention allocation are better equipped to transition inattentive drivers to take back control from automation. For example, when the Cadillac Super Cruise is in its ODD, the user interface lets drivers know when they can engage automation. Similarly, when Super Cruise exits its ODD, the vehicle prompts the driver to reengage manually. In addition, Super Cruise uses a small camera equipped with infrared technology to track the driver's head position. If drivers take their eyes off the road for too long, the system gives a series of warnings that increase in intensity until the driver takes back manual control from automation. The Super Cruise driver monitoring system is designed to keep drivers engaged in the monitoring task.

It is important to note the difference between a level 3 and level 4 automated vehicle. A level 4 AV, such as the self-driving Uber, does not require the driver to take back control under any circumstances. Although a level 4 may request the driver to intervene if it falls outside its ODD, there is no expectation that drivers will respond to this request unless they choose to operate the vehicle manually. Level 4 AVs serve as their own fallback controller when they fall outside their ODD and therefore are expected to take safe action if the driver chooses not to intervene. Because drivers are not expected to intervene unless they choose to, drivers' monitoring expectations are toward the AV itself rather than the road. For example, the Uber driver was advised to monitor the diagnostic messages that appeared on an interface and not necessarily the environment.

Takeover situations in level 3 AVs are unique because drivers are held fully liable even though they disengage from physical control. During transfer of control from automation to driver in a level 3 AV, the system goes through a mandatory transition process that requires the driver to be capable of switching from their previous task to manual control when prompted. This is known as an automation-initiated driver in control transition and will be referred to as "takeover" for the rest of the paper (Lu and colleagues, 2016). However, according to the SAE (2018), the driver does not need to supervise a level 3 advanced driving system within its ODD, even though it may fail. The discrepancy between AV limitations and the driver's understanding of those limitations creates confusion as to the driver's role. Moreover, the varying conditions that fall outside of the ODD for level 3 AVs can complicate a driver's understanding of the ODD. The mismatch between what AV designers expect of the driver and what the drivers believe their AV is capable of can lead to situations in which awareness needed to safely takeover is inadequate. Nevertheless, properly designed and transparent AVs that adapt to a driver's behavior may allow drivers to safely engage in secondary tasks while maintaining necessary situational awareness.

Maintaining good situational awareness while driving requires the driver to monitor objects and events such as traffic patterns, construction, and weather. The driver forms a dynamic mental representation of the driving environment (i.e., a mental model), that allows them to perceive critical changes, make decisions, and execute responses. Endsley (1988) defined situational awareness as "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future." As seen in Figure 1, level 1 situational awareness reflects a driver's perception of the elements such as other vehicles and pedestrians. Level 2 reflects a driver's comprehension of those objects' movements and intentions. Finally, level 3 reflects drivers' ability to predict the future state of their vehicle and other vehicles/pedestrians based on previous and current information. Each level builds on top of the other. Level 1 failures cause drivers to develop uninformed meanings of situations due to missing information, which may lead to an inaccurate prediction of a dangerous outcome. Even if drivers maintain level 1 situational awareness, level 2 failures can occur if a driver is not able to comprehend the significance of the changes in space and time. Level 2 failures also lead to a failure to correctly predict unfolding situations. Even when a driver may have a firm understanding of their environment, level 3 failures may unfold when drivers inaccurately predict the outcome of changes in the environment. Failures in any of the levels of situational awareness can lead to unsafe decision making. Therefore, higher levels of situational awareness are necessary for optimal decision making while driving. Drivers who do not safely monitor the environment lack perception of objects and events and thus overall situational awareness, which can lead to poor AV takeover performance. These drivers are referred to as "out of the loop" drivers.

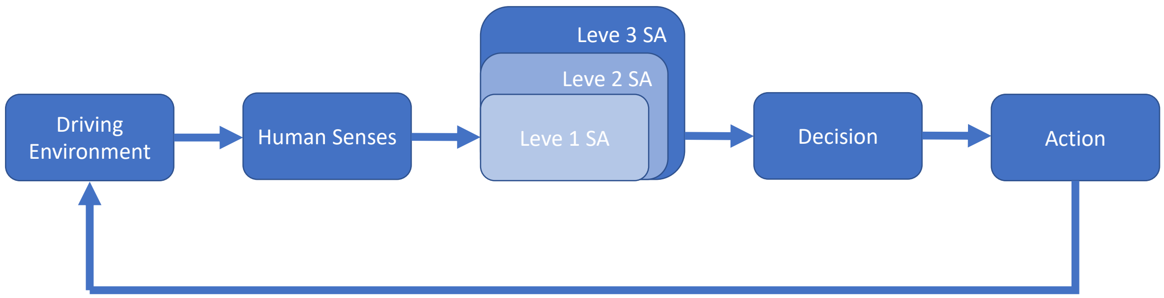

Figure 1. Human Control System in Automated Driving

When drivers are out of the loop, driver situational awareness and human-system interaction is limited (Endsley and Kiris, 1995). The concept of the loop originated from Control Theory and more generally refers to components of control and their connections. As shown in figure 1, these elements include environmental, cognitive, and behavioral mechanisms and connections that work to achieve a goal state (Merat and colleagues, 2018). Drivers that are in the loop are in physical control of the vehicle and are monitoring the driving situation (Merat and colleagues, 2018). Drivers that are on the loop are not in physical control of the vehicle but are monitoring the driving situation (Merat and colleagues, 2018). Finally, drivers that are out of the loop are neither in physical control or monitoring the driving situation, or they are in physical control but not monitoring the driving situation (Merat and colleagues, 2018).

In order to design level 3 AVs that promote situational awareness and keep drivers on the loop, human factors engineers need to understand the type of driver behavior that is required to enhance situational awareness. However, there is a dearth of data on how drivers physically behave behind the wheel of an AV. To better understand whether Autopilot mode appropriately supports drivers' situational awareness, Banks and colleagues (2018) collected video observations as part of an on-road study using a Tesla Model S. The study showed that drivers displayed behaviors indicative of automation complacency and over-trust, such as turning around in their seat or drinking coffee for extended periods of time without their hands on the wheel. Drivers were happy to engage in non-driving related tasks because they felt comfortable enough with their hands off the wheel for extended periods of time. Another study by Buckley and colleagues (2018) had participants drive a level 3 AV in a simulator. The experiment involved five transfer of control scenarios: missing line markings, presence of a police vehicle (parked on the roadside), traffic light failure, police-attended crash, and no obvious reason. After the experiment, participants were given a survey to understand how they perceived their role as a driver. Many of the participants saw the AV as a way to engage in secondary tasks, drive impaired, or as a means to relax (Buckley and colleagues, 2018; see also Payre and colleagues, 2014).

Even when drivers of highly automated vehicles do not engage in a non-driving related task their takeover performance can suffer. Merat and Jamson (2009) used a simulator to compare manual drivers' performance to automated drivers' takeover performance during critical events. Automated mode controlled both lateral and longitudinal movements. This was followed by a questionnaire for automated drivers to see how they felt about the automated system. Overall, drivers had positive opinions of the automated system and in-vehicle interface and driver response to all critical takeover events was found to be 1.5 seconds slower on average in the automated driving condition compared to manual driving condition. In another study, Strand

and colleagues (2014) used a simulator to compare a vehicle with only longitudinal automation (semi-automated) to a vehicle with both longitudinal and lateral automation (highly automated). The group engaged in highly automated driving had slower braking reactions than those driving a semi-automated vehicle. The highly automated driving group also had shorter minimum times to collision, which is defined as “the minimum time until a collision between the simulator car and a vehicle in front given that the vehicles maintained their current speed” (Strand and colleagues, 2014). These controlled studies, as well as the real-world crash examples, suggest that as the level of automation increases, takeover performance deteriorates.

As the driving task becomes increasingly automated and drivers assumes less control, drivers’ roles change and their responsibility to re-engage when prompted becomes more challenging. In order to design AVs that promote higher levels of situational awareness and keep drivers on the loop, human factors engineers need a better understanding of the limits of humans’ monitoring the driving task. Properly designed AVs that can detect the state of the driver and provide constructive feedback can help increase driver situational awareness in automated vehicles. In this study, we examine how camera-based driver monitoring systems that detect the direction of a driver’s gaze can help reinforce monitoring behavior as well as enhance situational awareness and takeover performance in level 3 AVs.

CHAPTER 2 – BACKGROUND AND LITERATURE REVIEW

Humans Monitoring Automated Tasks

Throughout time, humans have taken note of ways to complete tasks more efficiently. For example, in the 1900s mechanical engineer Frederick Winslow Taylor found that the amount of coal shoveled by workers increased as the weight of the shovel and required movements of the workers decreased. However, Taylorism was highly criticized for methods that portrayed humans as machines, ignoring fatigue, safety, and the well-being of workers (Sheridan, 2002). Before World War I, Human Factors was concerned with using the field of psychology to choose and train the right people for the job (Fitts, 1947). However, the demands of war shifted the perspective towards the design of equipment. Engineers studied the design of knobs, levers, and gauges in fighter plane cockpits to minimize workload and errors and to enhance pilot performance. Today, the rise of computers and automation shifts the use of psychology and the focus of Human Factors research toward an understanding of the way in which our minds perceive, comprehend and act on information and ways to enhance human monitoring behavior. One very important problem, the antithesis to optimal human performance, is the problem of vigilance decrement, which leads to a degradation of monitoring behavior and information perception. Vigilance decrement can be measured as a decrease or failure in physiological or psychological readiness to react to a given task over a period of time (Mackworth, 1948). Although vigilance decrement can be minimized, as long as we remain human the problem will persist.

The limits of human monitoring have been apparent in a number of other domains besides driving. On June 10, 1995 the Royal Majesty cruise ship ran aground 17 miles from where the watch officers thought the vessel was located (NTSB, 1995). Evidence showed that the satellite-based position data was interrupted about an hour after the ship left its port. Although there were three crew members to monitor the navigation through independent sources, they all seemed to over rely on the automated navigation system because the navigation system had proven to be highly reliable and accurate over the previous 3.5 years. Based on the crew's testimony, the navigation system was believed to be superior to other onboard systems. Because the master did not ask for deliberate cross checks between the GPS and the Loran-C, or make any comparisons himself, the National Transportation Safety Board concluded that the master's methods for monitoring the progress of the voyage did not account for the technical capabilities and limitations of the automated equipment. The grounding resulted in $7 million in damages. Nevertheless, properly designed and transparent automation should help operators understand their roles and the limitations of the overall system.

In order for drivers to understand situational limitations of the overall human-AV system, automation must help engage drivers in the monitoring task. Studies from rail transportation also look to understand vigilance decrement and ways to engage operators of automated systems in the monitoring task. Using a simulated, semi-automated rail control, Rees and colleagues (2017) conducted a 45 minutes study where participants monitored representations of railway lines on a screen and identified whether the planned routes were correct. The control group had no break for the entire 45 minutes while the other groups had 5-minute breaks after 20 minutes in which they undertook a music listening task, a music watching task, no task at all or a task of their choosing. Results showed that compared to the control group, the response latency to misrouted trains was lower for all other break conditions and that any activity that drew operators' attention from the primary rail control task enabled improvements in performance compared to the control group. In the context of driving, this study suggests that drivers may need to disengage from the monitoring task for at least short periods of time to maintain a high level of monitoring performance. However, because most automation takes drivers out of the loop, maintaining driver engagement, especially during long, straight, monotonous routes, becomes difficult.

Vigilance decrement can affect drivers within short periods of time and has been observed in nuclear power plant operators as well. Nuclear power plant operators have to constantly monitor the state of the plant through an array of control panels and computer displays. Reinerman Jones and colleagues (2016) had nuclear power plant operators perform checking, detection, and response implementation tasks. The checking task required a one-time inspection of an instrument to verify that it was in a correctly specified state. The detection task required participants to correctly locate a control and then continuously monitor that control parameter for identification of a specified change. There were twelve random changes per minute totaling 60 changes per detection task. The response implementation task required participants to correctly identify a control, and then open or shut a switch on that control. The study showed that vigilance decrement and higher subjective workload was induced during the 5-minute detection tasks.

Because vigilance decrement has been observed in a variety of domains from shipping to nuclear power plants, there have been countless studies to understand why humans are poor monitors of automation. These studies started during WWII when radar detection required pilots to stay vigilant in case of enemy attacks, and pilots were required to monitor screens for prolonged periods of time. It was suspected that working efficiency was deteriorating due to overlong spells at the radar screens. However, the elusive nature and complexity of the human mind makes human monitoring and vigilance decrement a challenge to study. Although solutions to the problem exist, they are never perfect, nor are the variety of theories used to address the problem. Nevertheless, countless academics within interdisciplinary fields have conducted studies to inch our way toward a deeper understanding of vigilance in the context of human-automation interaction. In terms of automated driving, because vigilance decrement leads to low situational awareness and poor takeover performance, understanding these mechanisms are essential for developing safe level 3 AVs.

The factors that cause vigilance decrement in level 3 automated vehicles can be understood using an older study conducted by Yerkes and Dodson (1908). Their law described the relationship between stimulus strength and habit-formation for tasks varying in difficulty (see Figure 2). Yerkes and Dodson's experiment consisted of a black box and a white box in which mice were to discriminate between. To see how perception of a task changes the act of vigilance, electric shocks of varying intensity were used to study the rate of learning the mice went through. If the mice went into the white box, they received a shock whereas if they entered the black box they received no shock. They observed that the learning rate was the lowest when very low intensity shocks and very high intensity shocks were applied, with moderate intensity leading to the highest rate of learning. Moreover, they found that the optimal shock intensity depends on the level of difficulty of the task. In terms of level 3 automated vehicles, providing drivers with optimal stimuli may help reinforce monitoring behavior. Furthermore, stimulus strength that is too low or too high can lead to inefficient or poor reinforcement of monitoring behavior. Nevertheless, humans are much more complicated and researchers were still not sure if these changes resulted from psychological concepts such as punishment, motivation, arousal, anxiety, or stress. Their law seemed to reflect a basic relationship between a variety of psychological variables. In addition, their model varies drastically depending on the type of task.

A more detailed explanation of how different stimuli and tasks relate to vigilance decrement was needed.

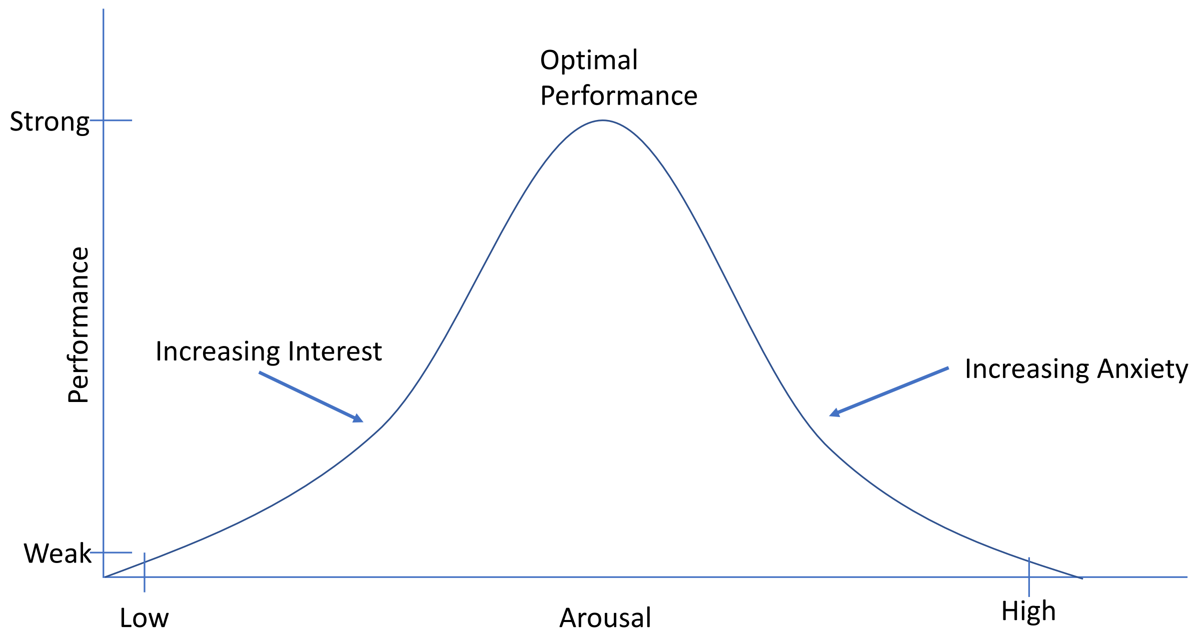

Figure 2. Yerkes-Dodson inverted-U law.

As research on pilot vigilance decrement advanced so did the definitions used to describe vigilance. Fraser (1957) conducted a series of experiments to test the effects of multi-day prolonged flights, differences between nighttime and daytime flying, differences between continuous and intermittent flying, and increasing heat stress on pilots. Based on his studies, Fraser differentiated vigilance into two types. The first type occurs when subjects must observe infrequent and significant signals appearing amongst frequent and insignificant signals over a long period of time. Fraser defines the second type as vigilance required for subjects to observe very frequent stimuli in a short period of time. Although vigilance decrement may be observed in the second type, there is no change in reaction time. It may be described that the first type is related to fatigue, a physical form of vigilance decrement. The second type is related to stress, a cognitive form of vigilance decrement. Nevertheless, cognitive stress can eventually lead to physically visible fatigue (Hancock and Warm, 1989). In fact, vigilance 1 may just be an exaggerated version of vigilance 2 (Fraser, 1957). Furthermore, the two types of vigilance also reflect Yerkes-Dodson's law in which type 1 requires vigilance in the context of a weak stimuli load and type 2 requires vigilance during a heavy stimuli load. Both lead to vigilance decrement and suboptimal learning rates. Nevertheless, because driving is already a satisficing task, the problem of vigilance decrement in automated vehicles thus far seems to be a type 1 vigilance problem, in which hazardous situations are infrequent and the signals from the environment are constantly unique and changing over a long period of time.

Type 1 vigilance and the standard for future vigilance tasks originated from a study by Mackworth (1948). Mackworth conducted one of the first vigilance experiments in which subjects watched the arm of a clock move a certain distance. At irregular intervals, the clock would go through a double jump, and the participants were to identify when these irregular movements took place. Only during the practice period did participants receive feedback of their results. After a 2-hour experimental session, Mackworth found that the number of signals participants began to miss dramatically increased after the first half hour and that lapses in visual perception become more frequent after a certain period of time. Using Conditioned Response theory similar to Yerkes and Dodson, Mackworth went one step further and attributed the observed vigilance decrement in their study to a lack of reinforcement due to an absence of immediate feedback of results. His study highlights the fact that human attention deteriorates over time and that monitoring performance becomes worse without reinforcement feedback based on performance. Therefore, in order to maintain a driver's situational awareness, level 3 AVs need to provide feedback to the driver based on their performance. Nevertheless, Mackworth's results still do not explain why different cognitive tasks are more vulnerable to vigilance decrement than others. Understanding what factors hold the attention of drivers in level 3 AV needs further examination.

When people take up under arousing tasks, it is not a matter of sufficient resources but how subjects allocate these resources to other thoughts and tasks that are more stimulating. Comparatively, drivers in automated vehicles have the resources needed to engage in monitoring responsibilities just as they would while driving manually, yet the nature of the monitoring task and the reduction of physical control lead to under arousing stimuli and a lack of feedback. To prevent drivers from engaging in non-driving related tasks, not only do level 3 AVs need to provide performance feedback to the driver, they need to activate mechanisms to sustain effortful attention.

To better design level 3 AVs that help maintain driver situational awareness and enhance driver takeover performance, methods that sustain and extend effortful attention may prove useful. As several studies have already mentioned, performance feedback deters vigilance decrement. In addition, Massar and colleagues (2016) used theories from Effort Allocation models, more commonly used in economics, to better understand other aspects of attentional control. Participants performed sustained attention tasks under reward conditions of either 0, 1, or 10 cents for fast responses as well as a discounting task to estimate the subjective value assigned to monetary reward. The sustained attention task required participants to press a button as quickly as possible upon appearance of a target. Participants were rewarded if their response was faster than the median. For the discounting task, participants were offered a choice of a low reward for short durations of the vigilance task and a high reward for long durations. After each choice, monetary reward for low reward was increased if the long duration option was chosen and decreased if the short duration option was chosen. Pupil diameter was also monitored throughout the experiment. Results showed that reward value influenced both task performance and the willingness to engage in task performance, which is in line with Effort Allocation theories. Pupil size was larger only during rewarded task runs based on good performance but not when rewards were provided at random. Furthermore, these results suggest that the motivational value of a task significantly increased attentional effort in terms of allocation of attention and duration of attention. Similarly, drivers may weigh rewards and risks of multitasking while driving at any level of automation. As automation increases, the risks involved in disengaging from the monitoring task may appear less likely.

When automation reduces the physical workload associated with driving, maintaining the same level of cognitive workload needed to maintain the physical workload may seem frivolous because the requirement of physical control motivates drivers to monitor their environment. When automation takes this control away, drivers may lose the motivation to maintain situational awareness. In order for drivers to maintain situational awareness in level 3 AVs, designers need to implement tasks that stimulate higher levels of cognitive control. In other words, the decisions that drivers make involving costs and rewards significantly affect how they distribute their attention while the vehicle is in automation.

An older study by Adams (1961) also looked to understand how decision making affects sustained attention. Adams looked towards the activationist hypothesis to provide an operational definition of stimulation in terms of environmental and response-produced stimuli. Under the activationist hypothesis, human alertness is a function of stimulation level. A semi-automatic air defense surveillance task was used for the vigilance experiment with characteristics of visual monitoring behavior in complex tasks like those found in other modern semi-automatic systems. Stimulation level was defined by the number of visual stimulus sources to be monitored. The study consisted of visual monitoring required in a simulated air defense surveillance task for a 3-hour period. There were four independent groups that received one of four load/response combinations. The two visual loads were expressed as six and thirty-six symbol stimulus sources and the two complexity values of response were Detection (D) and Evaluation (E). Detection involved finding changes in the environment and pressing the button labeled F. In the Evaluation condition subjects were to detect the signal change and also make a four-choice evaluation of the symbol on the display and indicate their decision by first pressing either button A, B, C or D, and then following it with the buttons for F and the three numbers of the symbol. The subject had to decide whether the symbol whose G had changed to F ended in an odd or an even number, and whether the symbol was above or below the horizontal line across the center of the display. No vigilance decrement was found in terms of percent of signals correctly detected. Nevertheless, response latency declined significantly for groups that had simple response conditions but not for groups with complex response requirements. The effects of waning vigilance, as revealed in an increase of response latency, was associated only with the Detection condition, not the Evaluation condition. Results suggest that although the information rate may not necessarily affect vigilance decrement, response-produced stimulation has a significant effect on the state of alertness. In terms of effort allocation, decision making seems to increase stimulation levels by providing motivational value to engage in the task. Therefore, when drivers are not involved in the decision-making process, their monitoring performance can suffer due to a lack of motivation and arousal. Simply giving drivers information without a purpose to use that information does not provide drivers the proper stimulation or meaning needed to stay on the loop.

Many years later, Adams results were corroborated by Endsley who directly measured the effect of passive and active information processing on situational awareness. As previously mentioned, Endsley defined situational awareness as “the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future.” Endsley’s theory allowed for the development of a protocol to measure and compare different tasks and system designs in order to objectively choose systems that result in higher vigilance. To objectively compare human situational awareness in different systems, Endsley derived and used a method referred to as the Situational Awareness Global Assessment Technique (SAGAT). The basic idea of SAGAT is that any simulation can be frozen for no more than 5 minutes. During the freeze-probes, participants are asked questions that directly relate to different levels of their situational awareness. Because their answers can be compared to the simulation, this is an objective measurement of situational awareness and can also be compared to other participants and between groups. Endsley and Kiris (1995) used SAGAT to further explore the loss of situational awareness during an experiment in which participants monitored an automobile navigation task in a simulator. A paragraph of text describing a decision task was then presented. The subject's task was to observe the presented scenarios and decide on one of three possible actions. In the manual condition, probabilities of the best action were not given. The decision support system group was given probabilities for each action in terms of the best performance. The consensual AI group was given the best action and participants could either accept or reject that action. In the monitored AI group, the best action would be carried out automatically but could be overruled by the subject if desired. The fully autonomous group was not able to intervene or make decisions in any way. Results showed that increasing the level of automation and shifting the subject from active to passive information processing resulted in loss of level 2 situational awareness. Endsley refers to this performance decrement from increasing automation as the out-of-the-loop problem and directly linked it to a loss of manual skill and awareness of the state and system processes.

The importance of decision making from the previous studies can also be conceptualized using Merat and colleagues' (2018) multi-level driving control model. Merat and colleagues argue that there are two potential ways drivers of level 3 AVs transition from in the loop to out of the loop. First, transferring manual control from the driver to automation naturally reduces situational awareness by eliminating physical control and the multi-sensory cues used for physical control (Merat and colleagues, 2018). Because the driver does not need to control the vehicle using the steering wheel, modes of feedback to the driver, such as proprioceptive feedback, are also eliminated in level 3 AVs. Nevertheless, the driver can still be considered on the loop. However, being fully in the loop and out of the loop is viewed as a spectrum (Merat and colleagues, 2018). As momentum shifts from the driver being in the loop and engaged in the driving task to on the loop, the momentum appears to continue. The momentum then pushes drivers out of the loop and towards other loops. In other words, when drivers are not in the loop they have a tendency to engage in other tasks. This brings up the second way and most important reason taking away manual control reduces situational awareness. As shown in Figure 2, to require manual control means that drivers must operate and monitor on a scale from milliseconds to seconds. All levels of situational awareness are involved in each of its three control loops (Merat and colleagues, 2018). Moreover, if the situational awareness box in Figure 1 is replaced by the control system in Figure 3, we see that when the need for the sampling rate associated with lateral and longitudinal movement is eliminated so is the need to make decisions at this rate. When decisions are not needed at a higher sampling rate, drivers tend to allocate their attentional resources towards other tasks that may be more motivating or stimulating.

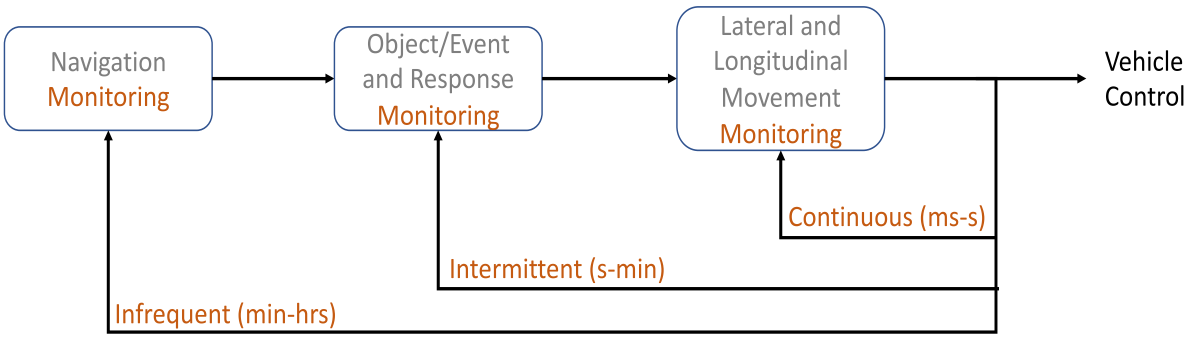

Figure 3. Monitoring inherent to multi-level control in driving (Merat and colleagues 2018).

Although there are a variety of theories that attempt to describe the underlying concepts of vigilance, there is no single theory that can be applied to every system and situation. Nevertheless, providing performance feedback and increasing decision making capabilities of driver's in level 3 AVs may help reinforce monitoring behavior and enhance situational awareness. The frequency at which these decisions must be made also affects situational awareness. In addition, Endsley's methods allow researchers to compare system designs and choose systems that deter vigilance decrement by enhancing situational awareness. Understanding the strengths and weaknesses of the aforementioned studies and their underlying theories will better equip human factors engineers to design systems that prevent vigilance decrement and enhance situational awareness in highly automated vehicles.

Human-Machine Interaction and Transfer of Control

As we begin to understand how humans behave while monitoring the driving task in a highly automated vehicle, naturally we ask if there are ways to optimize driver behavior as well as create adaptable human-vehicles systems that are robust in times of failure. Based on previous vigilance decrement research, properly motivating or reinforcing the driver to allocate attention toward the driving task can help maintain situational awareness. The type of feedback continues to appear as a prevalent factor responsible for reinforcing sustained attention and preventing out of the loop performance (Mackworth, 1948; Massar and colleagues, 2016; Endsley and Kiris, 1995). Because humans are poor monitors of automation, AVs that properly provide feedback to promote monitoring behavior and increase situational awareness can help enhance takeover performance.

Wickens (1979) observed the importance of feedback when he measured responses of participants engaged in single task computer-controlled trials. Participants were required to engage in a tracking task with their right hand and return the system to normal by clicking a button if a change was detected. The tracking target followed a random path and disturbances happened at random intervals as well. The step disturbances would also change the dynamic of the control stick in the manual condition. Thus, in manual mode, disturbances were not only detectable with visual information but also proprioceptive information coming from both joint position and muscle-tension related receptors. Each trial lasted about 150s. Wickens found that detection performance in the manual group was faster but slightly less accurate than the monitoring group. Wickens attributed the fast reactions in the manual group to proprioceptive feedback. A lower detection accuracy in the manual group was attributed to the fact that slow gains or small disturbances are less sensitive to proprioceptive input because humans are visually dominant. This study provides evidence that having more than one mode of feedback can help humans respond more quickly to automated system failures.

Nevertheless, human factors engineers not only need to think about the most effective ways to convey information but also how people feel about feedback from automation. Inadequate feedback may lead to over reliant behavior. On the other hand, too much feedback may cause an increase in workload as well as an under reliance on automation. Parasuraman and Riley (1997) refer to these cases as automation misuse and disuse, respectively. Because most automated vehicles appear highly reliable, automation complacency and automation bias will most likely lead to failures of monitoring and reduced situational awareness. Automation complacency can be operationally defined as poorer detection of system malfunctions under automation than manual control, and automation bias can be defined as the outcome of people's reliance on an automated decision aid as a heuristic replacement for vigilant information seeking. These effects are found in both naïve and expert participants, occur under conditions of multiple-task load, cannot be overcome with simple practice, and have been found to relate to allocation of attentional resources (Parasuraman and Manzey, 2010).

Because reliable automation leads to poor situational awareness and takeover performance, varying the reliability of automated may help enhance human-vehicle performance. Parasuraman and colleagues (1993) had participants engage in a 2 day flight simulation task for a total of 1 hour each day divided into 2 sessions. The overall task included monitoring, tracking, and fuel-management subtasks. Only the monitoring subtask was automated. Under normal operating conditions, the automated monitor detected and corrected the system malfunctions. However, from time to time the automation failed to detect a malfunction and the operator was responsible for detecting this failure. All groups experienced 16 failures every 10 minutes. The two groups tested were variable and constant reliability groups and each of these groups were split into two sub-groups. Automated monitoring for the two variable sub-groups alternated every 10 minutes between low reliability, in which the automated system would correct 9 of the 16 malfunctions, or high reliability, which corrected 14 of the 16 system malfunctions. The constant reliability sub groups consisted of one group which performed that task under low reliability the entire time and one which performed the task under high reliability. Parasuraman found that constant automation reliability at both low and high failure rates did not properly allocate the operators attention toward the monitoring task whereas variable automation reliability improved performance. The signal rate did not seem to be a factor that influenced allocation of attention whereas signal consistency had significant affects. Results also suggest that a variable reinforcement schedule of automation failures led to the enhancement in performance. Unfortunately, the enhancement of performance may come from a lack of trust in automation, which could lead to disuse during implementation.

In light of cases of misuse and disuse of automation, research on calibration of trust has gained much attention within the domain of automation and beyond. Trust can be defined as an attitude that shapes one's mental model of how well an agent can carry out a task in certain situations to achieve a goal (Lee and See, 2004). In order to trust in general, one must decide to forgo control and be willing to put themselves in a vulnerable or risky position. In a review of research on trust in automation, Lee and See (2004) urge for design of automation that can appropriately calibrate a person's trust in automation with the automation's capabilities, or trustworthiness. They argue that good calibration, high resolution, and high specificity of trust that is based on motives, rather than reliability, can lead to proper use of automation. Furthermore, Lee and See (2004) note that trust is an effective response and an emotion that can guide people's attention and decision making. Adaptive function allocation that transfers control based on human performance may provide a more practical solution to properly calibrate trust and update an operator's memory of the automated task than variable automation.

To begin understanding adaptive function allocation, one must look into theories that consider different levels of human control. Broadbent (1977) emphasizes that human information processing takes place on multiple levels simultaneously with different processors. To test this hypothesis, Broadbent gave each of his subjects the task of controlling transportation in a city by either altering time intervals between buses entering the city or altering the amount charged for parking in the city. The subjects were given an initial stable state in which circumstances demanded a change in load and free parking spaces in order to return to the stable state. After each situation, participants were asked how much to increase the time between buses and raise the parking fee to drive the bus load and number of empty parking spaces to a desired target. After each decision made by the participants, they were given the outcome of their decisions and asked to make another decision until the two target values for bus load and parking spaces were reached. Performance in actually controlling the system was judged by the number of trials to reach target values. Based on the responses of the participants tested under step and ramp inputs, Broadbent found that the simplest control system that models these results is an adaptive controller.

An adaptive controller is one that initially receives an output with large error from the lower levels, which causes the upper levels to reduce that error. The dominant activity seems to be a transfer of control from lower level processors to higher level processors in which the higher level processors eliminate output errors by altering the nature of the lower levels, rather than merely supplying them with input. In Broadbent's study, we observe that the higher level corrects the lower level by modifying the transfer function after every decision based on the amount proportional to the error. The lower levels obey different parts of the memory until the higher level's operations are done. Once the error is completely eliminated, the upper level disengages from the lower level. Furthermore, the lower level system continues to produce the desired output on its own using the last updated transfer function. This allows the higher level to engage in other purposes. Once this kind of control mechanism has been exposed to a situation repeatedly, it will react more rapidly and efficiently. In other words, adjustment of the lower level's transfer function can be thought of as a learning mechanism and once learning has taken place, feedback from the upper level is abolished. The upper level is needed only in the stage of adjustment. In addition, we can observe that the lower levels react to momentary input whereas the higher level operates on a longer time scale as it sets the value of the gain based on the error throughout the experiment.

It is important to note that regardless of the controller in Broadbent's study, without decision making there can be no error feedback and therefore no modification of the transfer function. Broadbent argues that humans preoccupied with unrelated tasks when primary tasks are automated show open chain behavior running off in the absence of higher level control, which becomes clear in response to novel events that need correction of automated errors. Therefore, design of automation should carefully introduce decision making as to continuously reengage driver's higher levels of control. Systems that help reengage a driver's higher level of control, even in the absence of error feedback, can help maintain situational awareness and safe driving behavior in highly automated vehicles.

Mouloua and colleagues (1993) conducted one of the first studies that looked at how adaptive function allocation may change monitoring behavior and enhance automation failure detection. Subjects simultaneously performed tracking and fuel management tasks while the system monitoring task was automated. The automation detected and corrected system malfunctions but would occasionally fail. Subjects were responsible for detecting any automation failure of the system monitoring task. The first study examined two different types of adaptive automation. The first type of adaptive automation was model-based and allocated the system monitoring task to the operator during a fixed period of time. The second type was performance-based and control of the system monitoring task was transferred from automation to operator when the operator's detection rate of automation failures fell below a certain criterion. The system monitoring task remained automated the entire time for the control group. Although there was not a significant difference between the two types of adaptive automation, Mouloua and colleagues found that in general, multiple adaptive changes were able to decrease monitoring inefficiencies and sustain overall performance for longer periods of time compared to the control group. Further examination of the possible explanations for the improvement in performance led Parasuraman and colleagues (1996) to believe that adaptive automation led to an improvement in performance because it provided an update of the operator's memory for the task. Parasuraman and colleagues explain that forgetting occurs when a task is automated and that adaptive automation, as well as variable automation, help maintain the operator's mental model of the task (Parasuraman and colleagues, 1996). Regardless of the cognitive mechanisms involved in optimal human-machine interaction, it is clear that in order to avoid negative consequences of automating tasks such as vigilance decrement, adaptive automation and decision making are key elements that help properly calibrate trust and reengage drivers in optimal monitoring behavior.

Driver Monitoring Systems

In the context of highly automated vehicles, it is important to know whether the driver is on the loop or out of the loop in order to design adaptable systems that optimize situational awareness. In other words, driver monitoring systems that measure the actual state of the driver may also be used to transition the driver into an expected state of behavior. Using situational awareness and takeover performance measurements as ground truth, we can test the effects of new technologies on attention allocation, safe driving behavior, and overall system performance.

Measurements of attention allocation that correlate with situational awareness and takeover performance are especially important for highly automated vehicles. Driver inattention is described as insufficient, or no attention, to activities critical for safe driving. Therefore, it is important to understand when attention is and is not allocated on activities critical for safe driving. Multiple resource theory claims that there are multiple resources running simultaneously but that these resources are all limited and performance greatly suffers when two stimuli occupy the same resource. Because vision is the most dominant and rich source of information for humans, situational awareness is most affected by where humans are looking. By measuring where humans are focusing their vision, human factors engineers can design level 3 AVs that adapt to changes in visual attention to allow for safe transfer of control when AVs fall outside their ODD. Nevertheless, because attention is a multidimensional theory like vigilance, there are a variety of ways driver monitoring systems can measure how and where attention is occupied.

Researchers notably differentiate driver monitoring systems that detect inattention into two categories: distraction and fatigue. May and Baldwin (2009) conducted a literature review and outlined 3 types of driver fatigue. Sleep-related fatigue results from accumulated sleep debt and prolonged wakefulness, which is resistant to most intervention strategies. Active task-related fatigue and passive task-related fatigue on the other hand are caused by mental overload and underload, respectively. These types of fatigue depend on the demands set forth by the driving task and/or environment and can be mitigated with properly designed AVs. Many of the measurements for active and passive task-related fatigue are similar to measurements of distraction and will be discussed in more detail in the following sections.

Driver distraction can be differentiated between visual distraction and cognitive distraction. Liang and Lee (2010) explain that visual distraction reflects demand for visual attention and cognitive distraction reflects shared demand of driving and secondary tasks. Cognitive distraction affects driving by disrupting the allocation of visual attention to the driving scene and the processing of attended information (Liang and Lee, 2010). There are a variety of other sets of driver inattention definitions found in the literature. For example, Koesdwiady and colleagues (2017) differentiate cognitive distraction into external, such as auditory distraction, and internal, such as mind wandering. Moreover, inattention is not always categorized as distraction or fatigue, such as drivers who are under the influence of drugs. Inattention due to inexperience has also been classified as its own subcategory (Koesdwiady and colleagues, 2017). Because there are a variety of ways to define inattention, there are also a variety of technologies used to capture inattention. Driver monitoring systems that can capture a broader definition of attention and accurately classify drivers as on the loop or out of the loop will prove to be more useful.

Distraction and fatigue are both abstract concepts still being studied. The main methods used by driver monitoring systems to study these concepts include vehicle dynamics, driver physiology, driver behavior, and subjective measures (see Koesdwiady and colleagues, 2017; Aghaei et al., 2016). However, for the purpose of this study we exclude discussing driver monitoring systems that use physiological measures such as heart rate monitoring or electroencephalogram because they are too invasive and only used for research purposes. Likewise, vehicle dynamics can give valuable information about the state of a driver but are only useful for AVs during takeover situations and thus can rarely be obtained from real-world scenarios. Subjective measures may also be invasive during real-time driver state detection and are only used for highly controlled experiments. The most sophisticated and information rich techniques available to measure the state of the driver involve image processing. In this paper we will focus on camera-based measures of driver behavior because they are noninvasive and can capture a variety of out-of-the-loop behaviors in lab settings as well as real-world settings.

Although camera-based driver monitoring systems that accurately and consistently measure the position of the head and eye in real-time are still in development, it is important to understand what measurements can be used and how to use them to detect different states of the driver. Doshi and Trivedi (2012) conducted a post-hoc analysis using a stereo-camera-based non-intrusive commercial eye gaze tracker to differentiate between top-down and bottom-up attention shifts in a level 3 AV simulation. Top down shifts in attention are internal shifts uninfluenced by any external stimuli vs bottom-up shifts which result from changes in the environment that catch a driver's attention. They found that bottom-up visual cues evoke different eye-head movement pattern latencies than top-down, attention shifts. Doshi and Trivedi (2012) found that head position seems to be a better indicator of attention than eye position. More specifically, head motions prior to gaze saccade indicate top-down processing, which is the same process observed prior to lane changes indicating a "task-oriented" attention shift. These findings are highly relevant for identifying the state of the driver's attention.

Another study by Yang and colleagues (2018) conducted a post-hoc analysis using a camera-based driver monitor to examine gaze behavior during different types of cognitive distractions. In this study participants drove approximately 25 km and were also to engage in several cognitively demanding tasks that demanded a range of resources such as long-term memory, working memory, and graph-based cognitive strategies. Percent road center (PRC) was a measure used to understand gaze behavior and is defined as the percentage of gaze fixations or gaze directions falling inside the road center area. Subjective reports of the difficulty of the secondary tasks were also conducted. Interestingly, PRC and subjective reports were inconsistent and may reflect a nonlinear relationship between cognitive loads and cognitive distractions. PRC proved to be an effective metric for detecting cognitive distraction, although it was not able to distinguish between the different types of cognitive tasks. The study also found that 8-12 degrees for the radius of the road center was optimal for detecting cognitive distractions using PRC.

Yamani and colleagues (2015) studied glance sequences of distracted drivers to compare gaze behavior of experienced and novice drivers. The participants drove in a simulator and performed a variety of in-vehicle and out of vehicle tasks such as adjusting the temperature of the air-conditioning or searching for a road on a map. The results indicated that the proportion of glances longer than 2s away from the road among untrained drivers was almost double the number of such glances for the trained drivers. Moreover, the researchers found that the duration of off- road glances varies as a function of their order in a sequence of glances. These results offer important insights into the visual demands imposed on a driver that can be used by driver monitoring systems to classify out-of-the-loop behavior.

Louw and Merat (2017) also conducted a post-hoc analysis using gaze distribution to identify when a driver is and is not capable of retaking control from automation. Participants drove with both manual and level 3 automated vehicles in a simulator and encountered six events. Each drive contained two critical events and four non-critical events. The lead vehicle would accelerate or change lanes during non-critical events and decelerate during critical events resulting in a 3s time-to-collision. To induce varying levels of the out-of-the-loop state during the automated drives, participants drove under conditions of no fog, light fog, heavy fog, and heavy fog plus a questionnaire task. Results showed that, during automation, drivers' horizontal gaze was more dispersed than that observed during manual driving. Drivers looked around more when their view of the driving scene was completely blocked by an opaque screen in the heavy fog condition. By contrast, horizontal gaze dispersion was more concentrated when drivers performed a visual secondary task. However, once the manipulations ceased and an uncertainty alert captured drivers' attention towards an impending incident, a similar gaze pattern was found for all drivers with no carry-over effects observed after screen manipulations. Results showed that drivers' understanding of the automated system increased as time progressed, and that scenarios that encourage drivers to focus their gaze towards the road center are more likely to increase situational awareness during high levels of automation.

Although the previous studies do not test driver monitoring systems' potential to enhance situational awareness in real-time, they portray important measurements that may one day be implemented in more advanced camera-based driver monitoring systems. Table 2 shows a comprehensive list of common video-based measures used to detect driver distraction and fatigue (Gonçalves and colleagues 2015). All of the measurements are based off head and eye detection and have been used in a variety of studies to understand the out-of-the-loop performance problem at varying levels of automated driving. For example, Boverie and Giralt (2008) created a driver vigilance monitoring system in level 1 AVs that captured the opening and closing of eyelids for fatigue detection. Their system detects features of the driver's face and eyes such as eyebrows, eye corners, and nostrils. They created four categories of blink classification based on the duration of eyelid closing: Short, Long, Very Long and Sleepy blinks. In their experiment, 12 drivers each drove about 360 km for a duration of about four hours. The degradation of the driver state all along the experiment was well detected by the system and verified by EEG and vehicle metrics.

Table 2. Adapted from Gonçalves and colleagues (2015).

Distraction Type	Distraction Metric	Fatigue Type	Fatigue Metric
Visual	Glance Pattern	**Eye Based**	Blink duration
	Mean Glance Duration		Blink Frequency
	Eyes-Off-Road Duration		Microsleep rate
Auditory	Pupil Diameter	**Behavior Based**	Yawning
	Blink Frequency		Nodding
Mechanical	Head direction		Slouching
Cognitive	Pupil Diameter		Eyebrow rising

Assuming driver state monitoring can detect the state of the driver perfectly, how can we use that information to enhance driver situational awareness and what are the consequences of certain interventions? For example, Ahlström and colleagues (2013) found that using the AttenD algorithm to issue warnings during manual driving decreased the visual time sharing between the driving task and a secondary task. The AttenD algorithm is a camera-based driver monitoring algorithm that attempts to classify drivers as out-of-the-loop in real-time if their visual attention is off the road. The system extracts facial features and determines areas of interest the driver may be focused on using gaze. Gaze is estimated as a vector perpendicular to the front of the face projecting outward. If the drivers gaze is directed away from the field relevant for driving, their visual attention is considered to be away from the road. AttenD identifies visual distraction in real time based on single long glances as well as repetitive glances (Kircher and Ahlström, 2013). Glances can be described as gaze duration and gaze frequency on a particular area of interest. When the driver's glances are directed away from the field relevant for driving for a certain period of time, such as 2 seconds, the algorithm may classify the driver as distracted (Kircher and Ahlstrom 2009). Algorithms such as AttenD may be used in level 3 AVs to reinforce monitoring behavior and enhance situational awareness. The results indicated that the total duration of potentially dangerous off-road glances was shortened by AttenD. The following study looks to use similar head tracking and algorithms to understand how driving behavior in automated vehicles changes and how these changes in behavior can help enhance situational awareness.

CHAPTER 3 – RESEARCH QUESTIONS AND HYPOTHESES

A dense history of research shows that requiring humans to monitor automated tasks leads to vigilance decrement. When automated vehicles fail, the resulting consequences could be fatal. Sustained attention toward the driving task is needed in order to optimize situational awareness and transfer of control when automation fails. In order to automate driving tasks without reducing driver situational awareness, driver's need to be kept on the loop through other driving related tasks that provide feedback. Giving drivers decision-making abilities may help reinforce behavior associated with higher situational awareness. In this study, we use a camera-based driver monitoring system to monitor the driver's head behavior and provide feedback. The driver monitoring system requires drivers to make decisions about their own behavior. We believe that driver monitoring systems can help direct drivers' gazes toward the road and their attention towards the field relevant for driving (FRD). The FRD is considered the road, objects, and events outside of the vehicle and within the driver's field of vision. Moreover, these camera-based driver monitoring systems can be used to enhance situational awareness as well as takeover performance in highly automated vehicles. This study looks to answer and confirm the three following questions and hypotheses:

1. Can the DMS interface for the AM group be used to direct gaze back to the FRD and enhance situational awareness of drivers while drivers are on the loop and engaged in a secondary task?

   Hypothesis: The AM DMS interface can be used to increase SAGAT scores, the frequency of glances toward the FRD, the average duration of glances toward the FRD, and the total duration of time spent looking toward the FRD compared to the Baseline and SCT groups.

2. Can the AM and SCT DMS interfaces help drivers react sooner and with smaller lateral deviations after being prompted to takeover while engaged in a secondary task during level 3 automated driving?

   Hypothesis: Both AM and SCT groups' DMS interfaces can be used to reduce hand to steering wheel and steering response times and maximum lateral deviations during level 3 AV takeovers compared to the Baseline Group. Drivers in the SCT group will return their hands to the steering wheel and steer sooner with less braking and smaller maximum lane deviations than the other two groups.

3. How do DMS interfaces affect driver complacency in the level 3 AV?

   Hypothesis: Drivers in the Baseline group and SCT group will report higher levels of overall comfort, comfort during AutoDrive, and comfort while their eyes are off the road. in the system compared to the AM group. Drivers in the SCT group will report higher levels of takeover comfort. There should be no significant difference between group comfort levels for transitioning from manual to AutoDrive.

In order to test these hypotheses, we have conducted a level 3 AV study and have collected data using the NADS-1 high fidelity simulator, a camera-based driver monitoring system, as well as situational awareness data during freeze-probe events. The goal of the study was to understand if camera-based driver monitoring systems can be used to enhance driver situational awareness and takeover performance in level 3 automated vehicles and to discover other consequences of driver monitoring systems in level 3 automated vehicles as well as implications for future research and implementation.

CHAPTER 4 – MATERIALS AND METHODS

This experiment was designed to understand how driver monitoring systems can be implemented to enhance situational awareness and takeover performance of distracted drivers in level 3 AVs.

Participants

Participants were recruited from the NADS IRB-approved registry and contacted by email or phone. They were provided a general overview of the study and screened to verify eligibility. The sample population consisted of 24 participants (12 male and 12 female) within the age range of 21-45 that had at least three years of driving experience and were in good general health. A between subject analysis was set up with three groups of eight subjects, and each subject was randomly placed in one group. The three groups include the Baseline (control) group, the State Contingent Takeover group, and the Attentional Maintenance group. There were 4 males and 4 females in each group.

Apparatus

Driving simulators are necessary to test AV technologies unfit for implementation on roads and testing during real-world hazardous situations. Performance in simulators have been verified to reflect real-world driving performance in the past. A series of experiments at the Iowa Driving Simulator examined the differences between simulators compared to real-world driving and found that driver response times were statistically equivalent after perception of a threat (McGehee et al., 2000). Furthermore, strong positive correlations were found between simulators and real-world driving conditions for drivers transferring control to and from automation with no significant differences with regard to workload, perceived usefulness and satisfaction (Eriksson et al., 2017).

This study used the high-fidelity full-motion NADS-1 simulator at the National Advanced Driving Simulator at the University of Iowa as shown in Figure 4. The NADS-1 simulator consists of a 24-foot-diameter dome, which encloses a full-size sedan. The 13 degree-of- freedom motion system provides drivers accurate acceleration, braking, and steering cues as if they were actually driving. Motion is critical to testing countermeasure effectiveness because drivers rely on these physical cues to maintain vehicle position in the lane and to detect deviations or lane departures. The NADS-1 uses sixteen high definition (1920x1200) LED (light emitting diode) projectors to display seamless imagery on the interior walls of the dome with a 360-degree horizontal 40-degree vertical field of view. The simulator cab is a 2014 Toyota Camry equipped with active feedback on steering, braking, and accelerating as well as a fully operational dashboard. Data are sampled at 240 Hz. Due to the importance of motion cueing while the vehicle is under automated control, this study will make use of the NADS-1 simulator with a passenger vehicle sedan equipped with automated driving functionality and a driver monitoring system. Additionally, wireless and cellular capabilities are available in the cab to allow the driver to engage with carry-in electronic devices.

Figure 4. Exterior and interior views of the NADS-1 simulator.

Scenario Development

Using the Interactive Scenario Authoring Tool (ISAT) at University of Iowa's National Advanced Driving Simulator, a simulation was designed that consisted of approximately 40 miles of interstate driving as shown in Figure 5. The interstate consisted of two lanes running in each direction with varying traffic density ranging from three to six surrounding cars.

The drive included four freeze probes and four takeover events. Two of the takeovers were due to construction and the other two were due to failure in automation (sudden dropout). Each type of takeover event happened once during low traffic demand (3-4 cars) and once during high traffic demand (5-6 cars) (see Table 3).

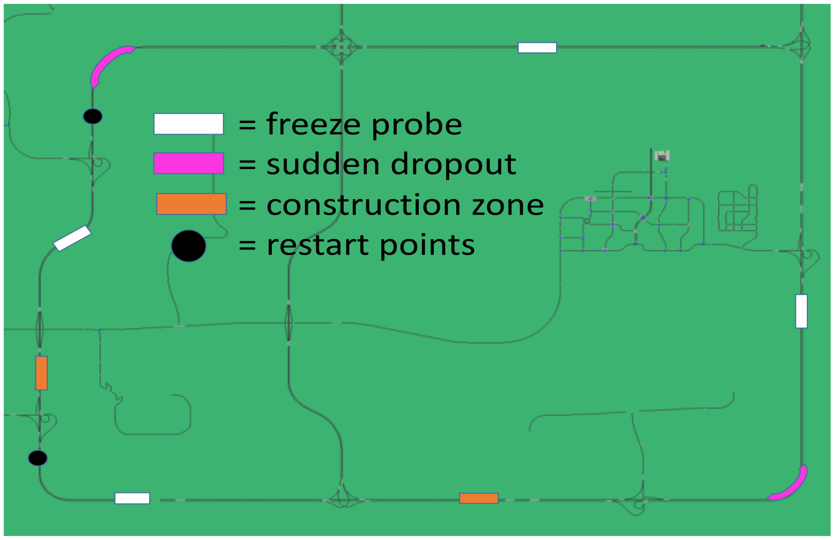

Figure 5. Bird's eye view of simulation track.

Table 3. Freeze probes and hazardous events.

Event	Event Name
Event 1	Freeze 1
Event 2	Dropout high demand
Event 3	Freeze 2
Event 4	Work zone high demand
Event 5	Freeze 3
Event 6	Work zone low demand
Event 7	Dropout low demand
Event 8	Freeze 4

Procedure

Participants were asked to avoid consuming alcohol or other drugs not prescribed by a physician within the 24 hours before the experiment. They were informed that they will be driving an automated vehicle and that they may engage in other activities while it is in operation. Upon arrival, the study was explained, and participants were asked to provide informed consent. This was followed by a general survey on trust of technology, a general demographics survey, and a presentation on the automated vehicle and general simulator procedures. Participants were then escorted to the simulator where they were provided a brief overview of the cab layout and were allowed to adjust their seat, steering wheel, and mirrors. Although participants were encouraged to engage in a secondary task, and other activities they felt safe engaging in while the vehicle was under automated control, they were reminded that they were responsible for the overall safety of the drive. Headphones or earbuds while driving were not allowed.

The familiarization of the simulator took place at the beginning of the first drive and lasted approximately 10 minutes, during which drivers had a chance to accelerate, engage in automation, brake, and steer. During this period, they were also introduced to freeze probe events as well as attention and takeover alerts. AutoDrive was engaged by pressing a button on the steering wheel with their right thumb (see Figure 6). When automation was engaged, drivers were alerted by an auditory signal and a visual icon on the instrument panel behind the steering wheel. The level 3 automated system included longitudinal and lateral control with adaptive cruise control, lane keeping assistance, and automated lane changing. At the end of the practice drive, participants were also asked to complete a simulator sickness survey before beginning the main drive.

Figure 6. AutoDrive button to engage automation followed by icon and audio.

At the beginning of the main drive, participants were asked to drive to 65mph and to engage automation. Automated auditory messages reminding the driver to engage automation were also provided at the beginning of the driver and after takeover events. After completing the main drive, participants were escorted from the simulator to complete a post-drive survey that evaluated potential simulator symptoms, trust and acceptance in the automated vehicle, mental models, and realism of the main drive.

Secondary Task

Drivers of level 3 AVs tend to disengage from driving tasks to engage in non-driving related tasks. In order to examine the potential of driver monitoring systems to enhance situational awareness and takeover performance of a distracted driver in a level 3 AV, participants were asked to hold an iPad and engage in a trivia game subject of the participant's choosing (see Figure 7). Studies have found that visual and motoric tasks are most demanding and lead to lower quality takeover performance (Gold et al., 2015). The trivia game was provided by TriviaPlaza.com. In order to have drivers naturally engage in the secondary task as they would in the real world, a system of monetary incentives was used to replicate the motivational tradeoffs of a distracted driver. Although all participants received additional compensation, participants were originally deceived into believing additional compensation would be received only if they answered 100 trivia questions correctly within the study drive. Participants were asked to engage in the secondary task as much as they felt comfortable while the car was in automation.

Figure 7. Driver engaged in secondary task during level 3 automated driving.

Takeover Scenarios

Participants experienced two dropout takeover events and two construction takeover events while engaged in the secondary task. The construction takeover events involved a work zone that blocked the lane the vehicle was in as shown in Figure 8. Drivers were issued a takeover request and were required to take back control from the vehicle in order to safely maneuver into the adjacent lane and avoid collision with the construction. The vehicle remained in automation until the driver engaged manually. Automation was disengaged anytime drivers intervened using the steering wheel, the breaks, or by pressing the automation button on the steering wheel.

The dropout events involved gradual steering disturbances on a curved road that eventually led to a sudden and total failure in the automated system as shown in Figure 9. A takeover request was issued at the point of the complete automation failure. At this point, participants were required to takeover control of the vehicle manually to prevent the car from leaving the lane. All groups received a baseline takeover warning 6 seconds prior to collision with a construction zone and a takeover warning precisely at the time of complete automation failure during the dropout events. After each takeover event, an automated audio message reminded the driver to put the vehicle back into automation. Data collected from the takeover events include human input to the vehicle as well as the DMS.

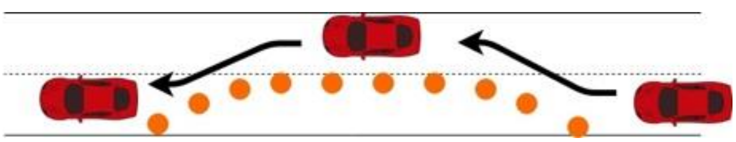

Figure 8. Maneuver required by driver to avoid construction zone.

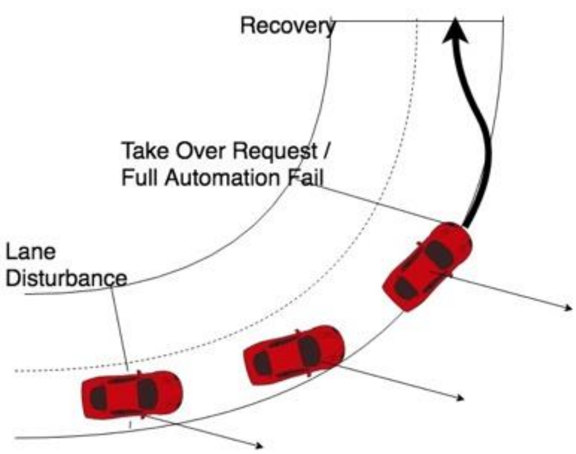

Figure 9. Maneuver required by driver to maintain control after automation dropout.

SAGAT

Four freeze probes were implemented to measure situational awareness throughout the drive. These freeze probes consisted of a stretch of highway where the car remained in automation, but there was no visibility of the scenario as the screens turned completely white. The freeze probes were based on Endsley's SAGAT method (Endsley, 1988a). During the freeze probes, a driver's situational awareness was measured through questions about the environment. During each freeze probe, participants were immediately handed a sheet by the researcher and asked to draw the traffic configuration before the driving scene continued. Participants were asked to write in the color of the car and the location of the vehicle in boxes which represent spatial proximity (see Figure 10). The freeze probes lasted approximately 30 seconds before the driving scene resumed in which the participant was then required to immediately hand back the sheet to the researcher. No other data was collected during the freeze probes. Traffic configurations did not change during or immediately after the end of the freeze probes. SAGAT was not found to affect performance and has been shown to be an objective measurement of situational awareness that can be used to compare human in the loop systems (Endsley, 1988a). The information from the freeze probes was used as ground truth to compare the overall situational awareness of each group and for comparison to DMS data.

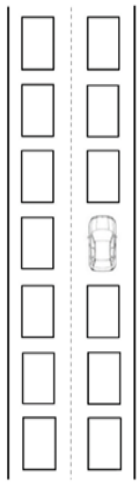

Figure 10. Freeze
Probe Test

DMS and Algorithms

A driver monitoring system provided by Aisin was integrated into the NADS-1 cab and used to classify the state of the driver. The driver monitoring system detected drivers' eyelids and face orientation and determined the direction of their gaze throughout the drive except during freeze-probes. Calibration of the drivers' faces and the center of their gaze took place while they were operating the vehicle manually before transferring control to the automated system. A simple yet effective algorithm was implemented that alerted drivers based on their group condition after a number of seconds of gaze fixation off the road. An algorithm similar to the AttenD algorithm was used by adjusting the field relevant for driving and the time buffer that decrements when the driver is looking away from the field relevant for driving (Kircher and Ahlström, 2013).

If the participant's gaze lied inside the FRD, defined as a circle with radius ten degrees from the center of calibration, the time buffer fell quickly to zero. When the participant's gaze lied outside the FRD, or if the camera was not tracking the driver's gaze, the time buffer incremented up to 60. An attention alert was given to the driver when the buffer exceeded 30, which indicated the driver's gaze lied outside the FRD for at least 30 seconds. The signal could not repeat at a rate faster than once every 30 seconds. Equation 1 was used to determine when a signal would be given, where θ and ϕ are pitch and yaw of the driver's face. When the participants gaze lies inside the FRD, the value of ρ is less than one and greater than one if their gaze lies outside the FRD.

$$\rho = \sqrt{\frac{(\theta-\theta_0)^2}{100} + \frac{(\phi-\phi_0)^2}{100}} \qquad \text{Eq.1}$$

In addition to gaps in research studying driver monitoring systems in highly automated vehicles, there are currently no studies that establish situational awareness ground truth for comparison to these algorithms in the presence of a secondary task.

Experimental Design

The experiment comprised of three groups with 4 males and 4 females randomly selected for each DMS group. All drivers were warned with a standard warning system similar to that of current commercial warning systems. Previous studies have shown that drivers need about 5-8 seconds to take back control from automation (Mok et al., 2015). In this study, a 6 second baseline takeover warning was used before each construction zone. Takeover warnings were issued through a red and black icon in the heads-up display behind the steering wheel as shown in Figure 11. The visual warning was also accompanied with an auditory warning.

In addition to takeover warnings, the State Contingent Takeover and Attentional Maintenance groups were given attention alerts. The attention alert was also given visually in the heads-up display as shown in Figure 11 and was accompanied with an auditory warning different from that of the takeover request. The point at which the driver was warned depended on the

DMS algorithm. For the State Contingent Takeover group, the algorithm alerted the driver at most 4 seconds before the baseline 6 second warning. The driver could have been warned at any time within the extra 4 second interval as soon as the DMS detected their gaze was away from the field relevant for driving for more than 10 seconds. The Attentional Maintenance group was given an attention alert anytime the driver's gaze was outside the field relevant for driving for more than 30 seconds and was not dependent on the timing of takeover events.

All groups received a takeover warning at the time in which automation completely failed during dropout events as shown in Figure 12. Prior to complete failure, slight disturbances could be felt from the automated vehicle as soon as 10 seconds prior to dropout. The state contingent group received an attention alert at most 3 seconds before dropout using the same DMS algorithm as for the construction zone events. The control group received no attentional alerts from the DMS at any point in the drive.

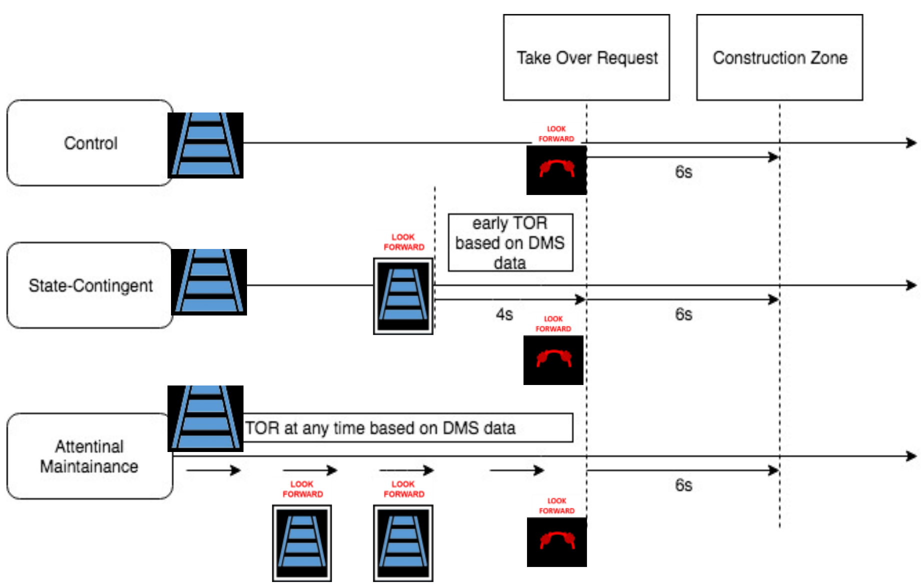

Figure 11. Takeover process for each DMS group before a construction zone.

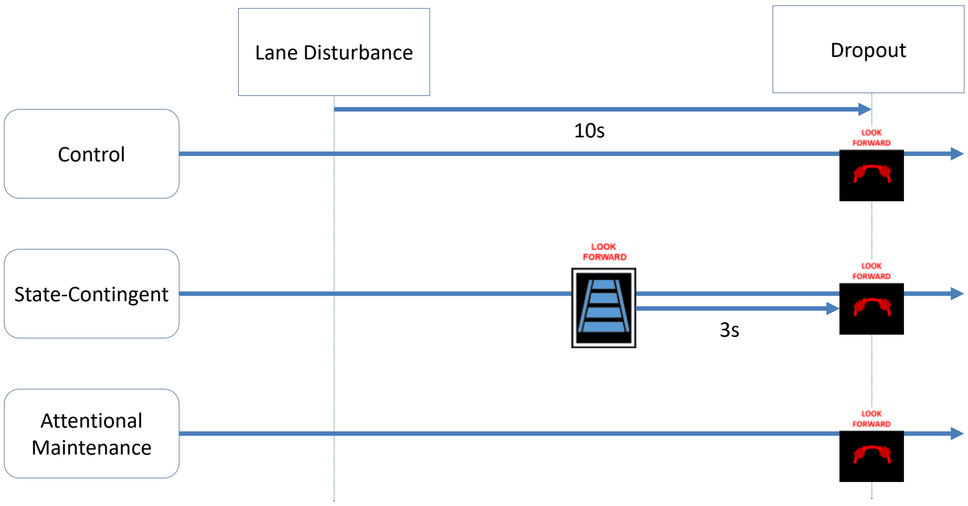

Figure 12. Takeover process for each DMS group before automation dropout.

Data Analysis

The purpose of this study was to determine the effects of the AM and SCT interfaces on drivers in a level 3 AV. The interface for the AM group was designed to use driver state feedback to keep drivers on the loop. The purpose of the AM group was to test how attention alerts based on gaze behavior affect situational awareness and takeover performance, as well as gaze behavior itself, compared to drivers in a level 3 automated vehicle who did not receive attention alerts. The SCT interface used driver state information to modify takeover requests when drivers were distracted. The SCT group was designed for two purposes. Its first purpose was to understand how an interface that alerted drivers before a TOR effected takeover performance. The second reason for the SCT group was for comparison to drivers in the AM group who were not given attention alerts based on takeover events but were alerted throughout the drive based on if their gaze was directed toward or away from the field relevant for driving (FRD). We were interested in how the AM group's interface prepared drivers for takeovers compared to the SCT interface which alerted drivers before a takeover request.

The independent variables used for analysis include the DMS interface between groups, the traffic configuration for the freeze-probes within groups, and the location and type of takeover event within groups. Analysis of dependent variables will be divided into four main data sets shown in Table 4. The first set will include results from the freeze-probe questionnaire derived from SAGAT. The results of the questionnaire will be used to help establish situational awareness ground truth for comparison to the other data sets. The second set will include data from takeover events, which includes human input to the vehicle and DMS. The takeover events data set includes all data starting several seconds before the takeover request up to the first point in time in which the driver presses the AutoDrive button. From this data set, we will look to analyze response times and takeover quality. The third set includes DMS data from all other sections in which the vehicle is in automation, except during the freeze-probes. In these automated sections, data from the DMS will be used to analyze gaze behavior as the participant transitions back and forth between the monitoring task and the trivia task. The fourth data set includes subjective measurements of trust from post-drive surveys. RStudio version 1.0.143 was used for all data analysis.

Table 4. Data sets used for analysis.

Main Data Sets	**Purpose and Content**
1. Freeze Probe Questionnaires	Establishes situational awareness ground truth. Based on fixed traffic configurations.
2. Takeover Events	For takeover performance evaluation. Includes DMS data and vehicle data
3. Automated Driving Sections	For analyzing gaze behavior data from the Driver Monitoring System
4. Post-Drive Surveys	For evaluation of trust based on subjective measurements

Question 1: Can the DMS interface for the AM group be used to direct gaze back to the FRD and enhance situational awareness while drivers are on the loop and engaged in a secondary task?

In level 3 AVs, drivers are taken out of the control loop for long periods of time, which makes monitoring responsibilities increasingly important for tomorrow's drivers. Because humans are historically bad at monitoring automation, AV interfaces that use DMSs aim to increase situational awareness by directing the driver's attention back to tasks relevant to driving. Although the freeze probe questionnaires are the only objective measurement of situational awareness used in this study, driver head movement and gaze can be used to detect whether the driver is paying attention to task-relevant objects (Doshi and Trivedi, 2012). Dependent measurements are shown below in Table 5.

Table 5. Dependent variables used to answer research question 1.

Dependent Measures 1	Calculation Per Subject
Situational awareness scores (%)	Accuracy averaged for all four freeze probes
Frequency of glances (glances/min)	Total number of glances toward the FRD divided by the total time of the drive
Duration of glances toward the FRD (s)	Geometric mean of all glance durations toward the FRD
Percent of time drivers look toward the FRD (%)	Total time DMS classified driver as looking toward the FRD divided by the total time

It was predicted that drivers in the AM group would have higher accuracy on freeze-probe tests, transition gaze between the driving task and the secondary task more frequently, glance longer toward the FRD, and spend a greater percent of time looking toward the FRD. Freeze-probe questionnaires used to measure the driver's situational awareness were scored using an answer key. The answer key was marked with the correct position of the surrounding cars relative to the driver for each freeze probe. After scoring the percent of surrounding vehicles drivers correctly marked on their freeze-probe sheets, the scores were then aggregated across the three DMS conditions and a one-way ANOVA was used to compare the accuracy between groups.

In addition to objective measurements of situational awareness, data from the DMS was used to measure a variety of gaze metrics throughout the drive. Because the driver was either engaged in the secondary task or the monitoring task during automation, gaze was only analyzed on a binary scale in which gaze is directed toward driving related tasks if the drivers gaze is determined by the DMS to be inside the field relevant for driving, or on non-driving related tasks if the drivers gaze was determined to be outside the field relevant for driving. These measurements are based on similar measurements used by Ahlstrom and colleagues (2013) to study the AttenD algorithm.

The frequency of glances toward the FRD measures how often drivers shift their attention between the secondary task and the driving task. This was calculated using the total of number of glances toward the FRD divided by the time of each half of the drive. The average duration of glances toward the FRD was calculated for each driver. This measurement gives information on the length of time the driver stayed engaged in the driving task. Frequency of glances and duration of glances have both been used to determine driver distraction (Kircher and Ahlström, 2009). We also looked at the total percent of time drivers look toward the FRD as an overall measure of how much time was spent monitoring automation and driving related tasks. Gaze measurements were compared between groups using a one-way ANOVA and Tukey's test.

Question 2: Can the AM and SCT DMS interfaces help drivers react sooner and with smaller lateral deviations after being prompted to takeover while engaged in a secondary task during level 3 automated driving?

In order to understand how DMS interfaces affect takeover performance, we used the takeover events data set to compare the DMS conditions using the dependent variables shown in Table 6. A previous study by Gold and colleagues (2013) showed that as the timing of takeover requests increase, steering maneuvers increase and hands on wheel response times decrease. Here we examined how different DMS conditions affect hands on wheel time and steering response time when takeover alert timing remains constant. Hands on wheel time was identified through video coding and measured as the first instance drivers placed their hands on the steering wheel. Steering response time was measured by the first instance of a steering wheel movement greater than three degrees in magnitude.

Table 6. Dependent variables used to answer research question 2.

Dependent Measures 1	Calculation Per Subject
Hand to steering wheel response time (s)	Arithmetic mean for each type of takeover
Steering response time (s)	Arithmetic mean for each type of takeover
Maximum lane deviation (ft)	Arithmetic mean for each type of takeover

It was predicted that drivers of both AM and SCT groups would have their hands on the wheel and initiate steering sooner than in the Baseline group, but that the SCT group would have the earliest response times. Hands on wheel time and steering response time will be averaged over each type of takeover event for each subject for comparison between groups using a one-way Analysis of Variance (ANOVA). One-way ANOVA provides the best comparison of means between multiple groups as it controls for the overall rejection of the null hypothesis. In other words, ANOVA corrects for Type I error, which is the probability of finding an effect that is not there. ANOVAs have been used in a number of papers to analyze response times and lateral lane deviations during takeovers (Gold and colleagues, 2015; Louw and colleagues, 2015; Zeeb and colleagues, 2017; Mok and colleagues, 2015). Levene's test for homogeneity of variances is included in the output. If groups were statistically different according to the ANOVA, Tukey's honestly significant difference was used to test significance between group means. Games Howell post hoc test replaced Tukey's if the homogeneity of variances assumption failed.

Measures of maximum lateral deviation were also recorded for each takeover event and used as indicators of control and stability. Maximum lateral deviation was measured as the maximum distance from the center of the lane during each takeover event. Greater maximum lateral deviation values represent less stable control (Louw and colleagues, 2015). We predicted that the Baseline group would deviate from the center of the lane during takeover events more than the AM and SCT groups and that the SCT group would have the smallest maximum lane deviation.

Question 3: How do DMS interfaces affect driver complacency in the level 3 AV?

Operators tend to over rely on poorly designed automation to carry out tasks, which can lead to automation complacency. When drivers are complacent they may use the AV in a way the designer did not intend for the system to work, especially during automation failures.

Automation can induce complacency as secondary tasks compete with the monitoring task for the driver's attention (Parasuraman and Manzey, 2010). In order to avoid cases of misuse, AVs must be designed to avoid automation complacency. In addition, a system's trustworthiness must be transparent to drivers to help properly calibrate their trust in the system. Although the word 'trust' was not used in the post-drive survey, complacency and over-trust have been used interchangeably in the literature (Parasuraman and colleagues, 1993). Nevertheless, we will refer to the concepts of automation complacency and trust as different yet correlated constructs in which initially high trust toward automation can lead to automation complacency (Parasuraman and Manzey, 2010).

In addition, Hergeth and colleagues (2016) found a consistent relationship between gaze behavior and trust in automated vehicles after automation failures. They found that participants who reported higher trust in automated vehicles monitored the FRD less frequently. In combination with previous gaze measures, we used subjective measurements of comfort to infer how DMS interfaces affect the driver's trust and complacency levels in the AV. Based on our earlier predictions of gaze behavior, we predicted that drivers in the Baseline and SCT group would report higher levels of overall comfort, comfort during AutoDrive, and comfort while eyes were off the road than the AM group. Because drivers in the SCT group will have a state-contingent attention alert before the takeover request, we predicted that they would be more comfortable taking back control during takeover events than the Baseline and AM groups. Because the HMI was the same for all groups while transitioning from manual to AutoDrive, there should have been no significant differences. Dependent measurements of comfort are shown in Table 7. Comfort was measured subjectively on a scale from 1-7 using a post-drive questionnaire, found in Appendix A. Comfort scores were analyzed using a one-way ANOVA.

Shapiro-Wilk's method was used to test for normality. If the assumption of normality failed, we used a two-sampled Wilcoxon rank sum test with Bonferroni correction, which is a non-parametric t-test.

Table 7. Dependent variables used to answer research question 3.

Dependent Measurements 3	**Questions (1=low and 7=high comfort)**
Overall Comfort	In general, how comfortable did you feel during the drive?
Transferring Control to Automation	How comfortable did you feel transferring the vehicle into AutoDrive?
Taking Control from Automation	How comfortable did you feel resuming manual control from AutoDrive during the drive?
Comfort during Automation	Compared to driving manually, how comfortable did you feel driving in AutoDrive?
Comfort taking eyes off the road	Compared to commuting in your regular vehicle, how likely is it that you would take your eyes off the road for several seconds while driving with AutoDrive engaged?

CHAPTER 5 – RESULTS

Situational Awareness and Gaze Measurements

Situational Awareness Accuracy

Freeze-probe accuracy was averaged over all four freeze-probe events for each participant. Freeze-probe accuracy reflects level 1 situational awareness. Figure 13 shows freeze-probe accuracy for each group. An ANOVA comparing the freeze-probe averages showed there was a significant difference between groups ($F(2,20)= 4.671$, $p = 0.0217$). Normality and homogeneity of variance assumptions were met using Levene's and Shapiro-Wilk's tests at a 95 percent confidence interval. Post-hoc analysis using Tukey's method showed a significant difference ($p=0.0208668$) between the AM group ($M=80.52754$, $SD=8.404729$) and the SCT group ($M=69.35757$, $SD= 8.600540$). Although the difference between the AM and Baseline groups was not significant ($p=0.1078324$), possibly due to weak statistical power from low number of subjects, the mean freeze-probe accuracy for the AM group was also much higher than the Baseline group's ($M= 72.14524$, $SD = 4.793792$). The difference between freeze-probe accuracy of the SCT and Baseline groups was also not significant ($p=0.7603024$).

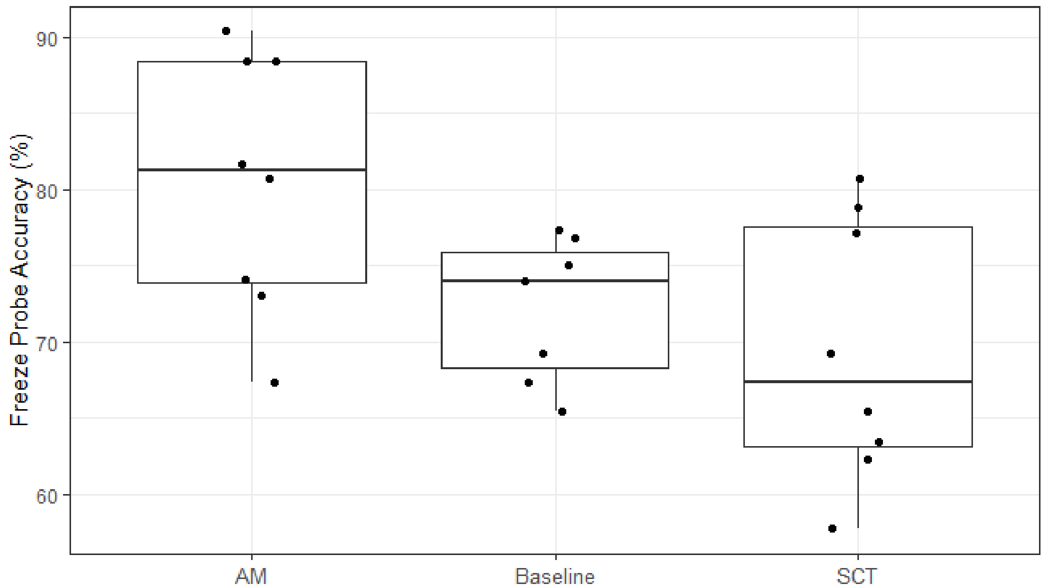

Figure 13. Boxplot showing median and 25-75 percent quantiles for freeze-probe accuracy with each point representing subject accuracy.

Gaze Measurements

Binary gaze behavior (toward or away from the FRD) was visualized for each subject. Before analyzing gaze measurements, subject 4 in the Baseline group was removed because the DMS was not able to capture head behavior during the second half of the drive based on visuals of the binary gaze data.

Figure 14 shows the percent of time drivers were looking toward the FRD. An analysis of variance showed a significant difference between group means ($F(2,19) = 3.115$, $p=0.0676$). Normality and homogeneity assumptions were met at a 95 percent confidence interval using Levene's and Shapiro-Wilk's tests. Post-hoc results using Tukey's methods showed drivers in the AM group ($M=0.3464620$, $SD=0.21069765$) looked toward the FRD more than the Baseline group ($M=0.1384349$, $SD=0.05247121$; $p=0.0586145$). There were no significant differences between the AM and SCT ($M=0.2283317$, $SD=0.14194410$) groups ($p=0.3089627$) or the SCT and Baseline groups ($p=0.5474926$).

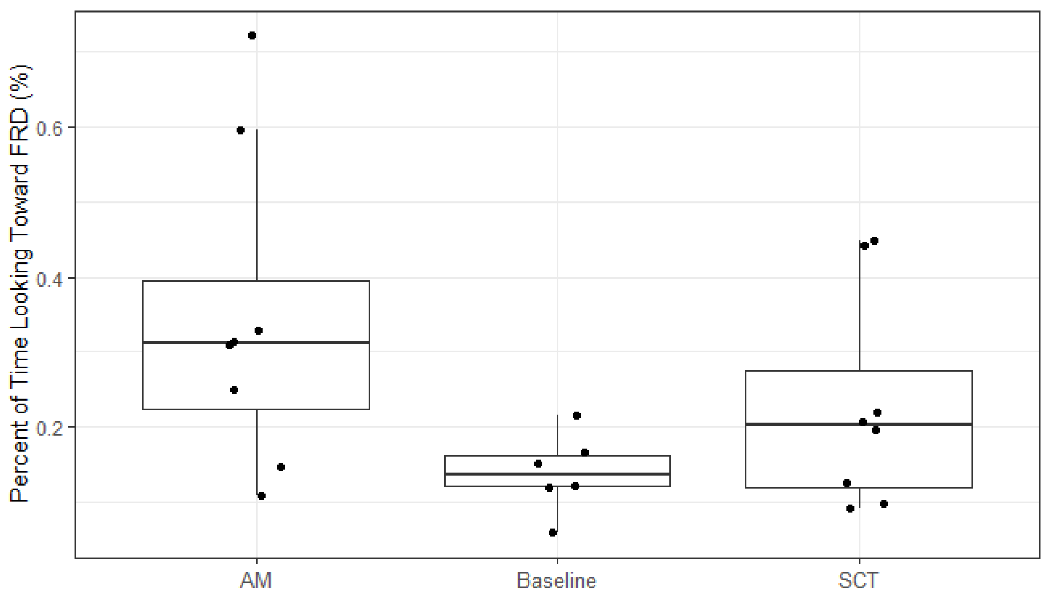

Figure 14. Boxplot showing median and 25-75 percent quantiles for the amount of time subjects spent looking toward the FRD with each point representing subject percentage.

There was no significant difference between groups for the frequency of glances up toward the FRD ($F(2,19)=2.181$, $p=0.14$). There was also no significant difference in glance duration toward the FRD between groups ($F(2,18)=.734$, $p=.493$; $F(2,19)=0.252$, $p=0.78$). Boxplots for frequency of glances (Figure A1) and glance duration (Figure A2) are shown in Appendix A. All non-significant visualization of measures are shown in Appendix A.

Takeover Response time and Quality

All takeover time and quality measures were analyzed using the arithmetic mean for each type of takeover. Before analyzing the response times and lane deviations for the dropout events, subject 8 was removed. The reasoning was that the hand response time was more than 10 seconds before the dropout takeover request, which is more than two standard deviations away from the group mean (M=-0.6577, SD=4.514). Moreover, because there was no visual indication that a dropout was approaching, the dropout could not have been predicted in advance.

Figure 15 shows hand response times for each group during the dropout events. Larger values represent slower hand to steering wheel response times and smaller values represent faster response times after a takeover request. An analysis of variance showed there were significant differences between group means ($F(2,18) = 5.021$, $p=0.0185$). Homogeneity of variance and normality assumptions were verified for all ANOVAs using Levene's and Shapiro-Wilk's tests and were met using a 95 percent confidence interval.

A post-hoc Tukey test showed that drivers in the AM group (M=1.164323, SD = 0.2313972) had significantly faster hand response times (p=0.0570690) than the Baseline Group (M=1.591369, SD=0.5064851). The SCT group (M=1.048264, SD=0.1186762) also had significantly faster hand response times (p=0.0226716) than the Baseline Group. Hand response time was not significantly different between the AM and SCT groups (p=0.7962625).

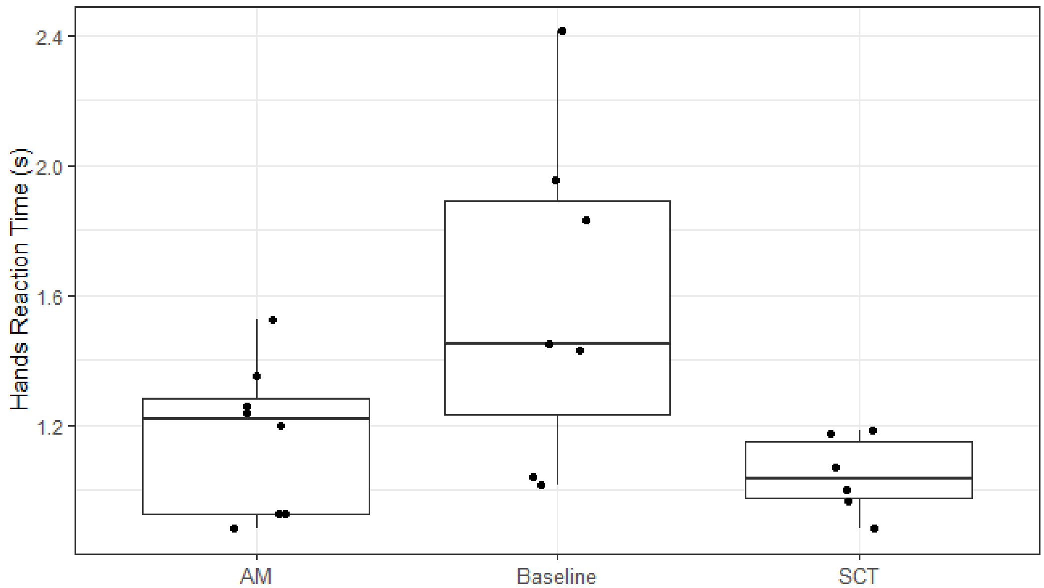

Figure 15. Boxplot showing median and 25-75 percent quantiles for hand response time during dropout events with each point representing subject mean.

Figure 16 shows steering response time during the dropout events. Larger values represent slower steering response times and smaller values represent faster steering response times after a takeover request. ANOVA results showed a significant difference between groups (F(2,18)= 4.078 , p=0.0346). Tukey post-hoc results showed drivers in the AM group (M=1.377083, SD=0.2366996) had significantly faster response times (p=0.0680899) than the Baseline group (M=1.716369, SD= 0.3734020). Steering response time was also significantly faster (p=0.0512596) in the SCT group (M=1.329514, SD=0.1553358) than the Baseline group. No significant difference in steering response time was found between the AM and SCT groups (p=0.9447233).

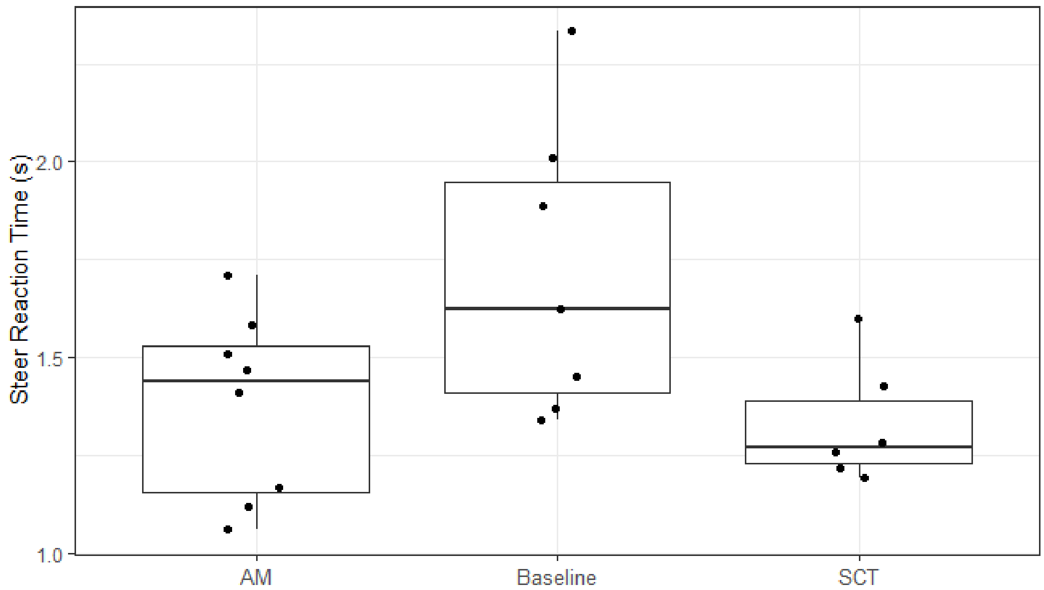

Figure 16. Boxplot showing median and 25-75 percent quantiles for steering response time during dropout events with each point representing subject mean.

Maximum lane deviation for dropout events (Figure A4 in Appendix A) was not significantly different between groups ($F(2,18)= 1.526$, $p=0.244$). The main effects between groups for hand response time during work zone events (Figure A3 in Appendix A) were not significant ($F(2,19)= 1.548$, $p=0.238$). The main effects between groups for steering response time during work zone events (Figure A5 in Appendix A) were not significant; $F(2,18)= 1.554$, $p=0.239$). Maximum lane deviation for work zone events (Figure A6 in Appendix A) was not significantly different between groups ($F(2,19)= 0.549$, $p=0.586$).

Subjective Measurements of Complacency

As shown in Figures 17-19, drivers in the Baseline group were generally more comfortable than drivers in the AM and SCT groups. Comfort was scored on a scale from 1-7, seven being the most comfortable and one being the least comfortable. Violin plots were used to visualize the distribution of comfort scores and their probability density. The red dot in the middle of the violin plot for each group represents the group mean. Levene's and Shapiro-Wilk's tests showed that assumptions of analysis were not violated for any ANOVAs on comfort.

ANOVA results show there was a significant difference in overall comfort level between groups ($F(2,18)=3.703$, $p=0.045$). Drivers in the Baseline group ($M= 3.571429$, $SD =0.9759001$) were significantly more comfortable with the overall drive (see Figure 17) than drivers in the AM group ($M=2.142857$, $SD =0.6900656$; $p=0.0409828$). No significant differences were found between the SCT ($M=2.571429$, $SD=1.2724180$) and AM groups ($p=0.7104766$) or the SCT and Baseline groups ($p=0.1802529$).

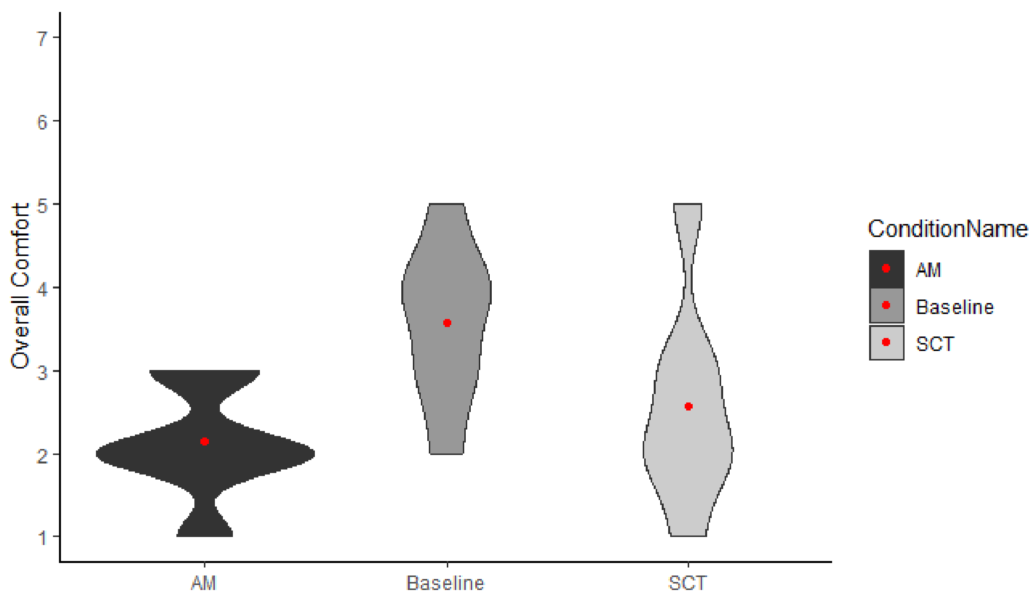

Figure 17. In general, how comfortable did you feel during the drive?

ANOVA results also showed a significant difference ($F(2,18)= 7.304$, $p=0.00444$) in takeover comfort levels between groups. Drivers in the Baseline group ($M=5.714286$, $SD =0.7559289$) were more comfortable taking over control from automation (see Figure 18) than drivers in the SCT ($M=2.714286$ $SD =1.6035675$; $p=0.0034042$) and AM groups ($M=3.875000$, SD =1.8077215; p=0.0663241). No significant difference was found between takeover comfort levels in the AM and SCT groups (p=0.3071887).

Figure 18. How comfortable did you feel when AutoDrive failed and you had to retake control?

ANOVA results also showed a significant difference (F(2,18)= 3.118, p=0.0702) in AutoDrive comfort levels between groups. Drivers in the Baseline group (M=4.333333, SD =1.5055453) felt more comfortable with AutoDrive (see Figure 19) than drivers in the SCT group (M=2.714286, SD =0.9511897; p=0.0856860). No significant differences were found between the AM group (M= 2.857143, SD = 1.3451854) and SCT group (p= 0.9760881) nor the between AM and Baseline groups (p=0.1235269).

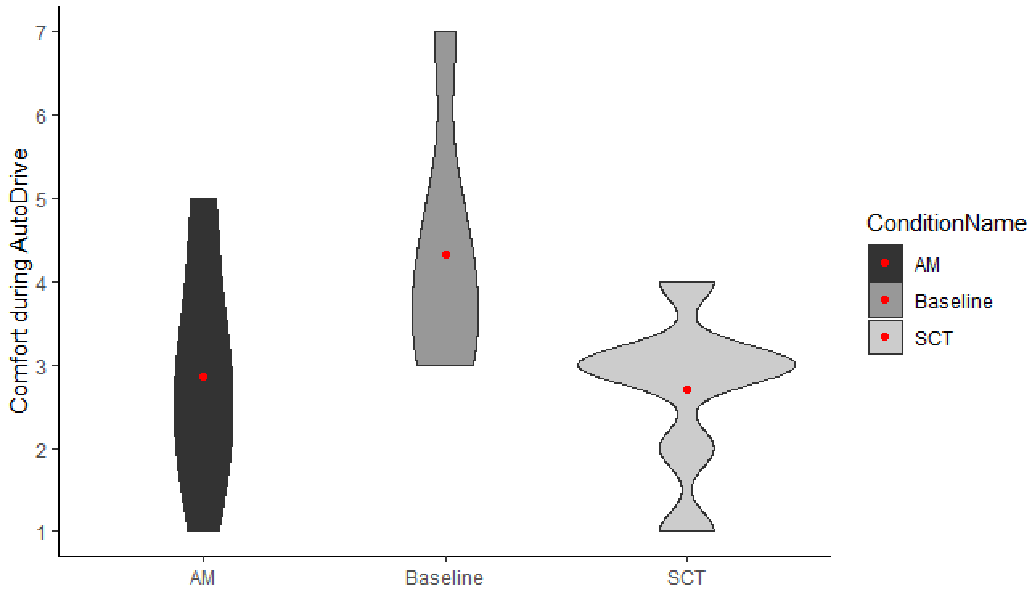

Figure 19. Compared to driving manually, how comfortable did you feel driving in AutoDrive?

Comfort with transferring control from manual to automation (Figure A7 in Appendix A) and comfort with taking eyes off the road (Figure A8 in Appendix A) were not found to be significantly different between groups (F(2,17)= 1.255, p=0.31; F(2,15)=0.19, p=0.829).

CHAPTER 6 – DISCUSSION

Situational Awareness and Gaze

We were correct in predicting that drivers in the AM group would have higher situational awareness than the SCT group. Results of the freeze-probe tests showed a significant difference in the percent of correctly identified vehicles between the AM and SCT groups, but not between the AM and Baseline groups. The freeze-probe test was designed to test Level 1 situational awareness, which is the foundation for overall situational awareness. The difference between the AM and SCT group's freeze-probe accuracies suggests that a DMS interface that uses attentional maintenance alerts can increase situational awareness.

We predicted that the AM group would look toward the FRD a higher percent of the time than the SCT and Baseline groups. Results showed that drivers in the AM group looked toward the FRD significantly more than the Baseline group, but not more than the SCT group, which is inconsistent with our results on situational awareness between groups. However, standard deviations were much larger between the AM and Baseline groups for the percent of time drivers directed their gaze toward the FRD than freeze-probe accuracies. For freeze-probe measures, the standard deviation of the AM group was almost twice as large, yet the standard deviation for the percent of time drivers looked toward the FRD was more than three times as large in the AM than the Baseline group. This suggests that although there is a large variance from attentional maintenance alerts to help direct a driver's visual attention toward the FRD, attentional maintenance alerts are less of a factor for the variance of cognitive resources toward the FRD. Studies on mind wandering show cognitive attention does not always correlate with visual attention. When visual resources are directed toward driving related tasks, cognitive resources are not necessarily directed toward the driving task (Geden and colleagues, 2017; Kaber and colleagues, 2012; Liang and Lee, 2010). Furthermore, Yang and colleagues (2018) found that the percent of time driver's direct their gaze toward the FRD can significantly increase with cognitive distraction. Results show that the AM group had higher situational awareness than the SCT group even though there was no difference between the percent of time drivers spent looking toward the FRD. In other words, when drivers directed their visual attention towards the FRD, the AM group may have been more likely to direct their cognitive attention towards the FRD than the SCT group. Because there is no statistical significance between the two groups in the percent of time drivers spent looking toward the FRD, the difference in freeze-probe scores suggests that the SCT group may have kept their cognitive attention on the trivia task when looking up toward the FRD. Therefore, larger variations in gaze measures compared to freeze-probe measures suggest that attentional maintenance alerts may not only cue drivers to look up but may help allocate cognitive resources toward the FRD. The results in terms of Merat and colleagues (2018) monitoring model suggest that attentional maintenance alerts helped drivers make decisions about the appropriateness of their monitoring behavior and adapt their interaction with the secondary task to direct cognitive resources toward the FRD (Schömig and Metz, 2018). It is important to note that, as predicted, there were no differences between the Baseline and SCT groups for either gaze percentage toward the FRD or freeze-probe accuracy.

Had a larger number of participants been used, the difference between the AM and Baseline freeze-probe scores may also have been significant as the p value was on the boarder of being significant. On the other hand, had we described significance using only a 95 percent confidence interval, there would be no significant differences in gaze behavior between groups using Tukey's post-hoc analysis. Nevertheless, using a more liberal criterion we see that attentional maintenance alerts help increase situational awareness and eyes on road time, which is consistent with our predictions. However, it is important to note that the difference between the AM and SCT group for measures of total eyes on road time was nowhere near significant, which is inconsistent with our predictions. A lack of a difference between the AM and the SCT group for measures of gaze will be explained further in terms of comfort.

Because monitoring an AV is considered an underload task, the attentional maintenance alerts may have increased performance by increasing arousal (Yerkes and Dodson, 1908). Moreover, Mackworth (1948) suggested that vigilance decrement can be mitigated using feedback performance to reinforce monitoring. The AM interface warned drivers based on their own monitoring performance, which may have increased arousal. In addition, the attentional maintenance warnings may have updated the driver's memory (Parasuraman and colleagues, 1996). Because drivers were encouraged to engage in the trivia task while the vehicle was in AutoDrive, they may have forgotten to monitor the environment and road or were reminded to shift their attention back toward the FRD using gaze performance feedback. When drivers hear the attentional maintenance alert, not only do they return their gaze back to the road, they are reminded their attention is not directed toward the FRD and to allocate cognitive resources toward the road. Attentional maintenance alerts remind drivers of the monitoring task which may help engage higher-level cognitive processes that transition visual and cognitive attention from the trivia task back toward the monitoring task and reinforce attention toward the FRD (Broadbent, 1977). Further analysis, discussed in sections related to comfort and trust, is needed to understand whether the stimuli changed behavior as a result of punishment, motivation, arousal, or stress. Nonetheless, the transition in visual and cognitive resources back to the monitoring task helped drivers in the AM group maintain situational awareness and react faster than the Baseline group during takeover events.

Takeovers

We predicted that the SCT group would have the fastest response times and smallest deviations from the center of the lane and that the Baseline group would have the slowest response times and largest deviations from the center of the lane during takeover events. However, there was no significant difference between the AM and SCT groups during the dropout events. In addition, response times for both the AM group and the SCT group were significantly faster than the Baseline group. These results suggest that attentional maintenance alerts helped prepare drivers to react just as quickly to takeover events as state-contingent alerts. Because drivers in the AM group had more alerts to respond to, they may have been better able to transition their visual and cognitive attention back toward the FRD to regain situational awareness during takeovers.

The takeover quality during dropout events between groups was not significant. Nevertheless, results suggest that attentional maintenance alerts can help increase situational awareness and help drivers return to the control loop more quickly during automation failure. Although situational awareness was not measured during takeover events, drivers in the SCT group who were out of the loop were given more time to react to takeover requests from state-contingent alerts, which suggests that the takeover response time variance in the SCT group may be due to a lack of situational awareness and individual differences in the allocation of cognitive resources directed toward the FRD (Vlakveld and colleagues, 2018).

Comfort and Trust

We predicted that drivers in the Baseline and SCT groups would report higher levels of overall comfort than the AM group. Overall comfort scores between the AM and Baseline group were correctly predicted. These results also align with results from the percent of time drivers looked toward the FRD (Körber and colleagues, 2018). A variety of concepts such as workload, arousal, and stress can be associated with an increase in performance and discomfort (Teigen, 1994; Yerkes and Dodson, 1908). An increase in workload provides one explanation for the lower levels of overall and takeover comfort compared to the Baseline group. Adaptive automation that briefly allocates tasks to the operator has been shown to increase workload along with monitoring behavior (Parasuraman, 1996). Attentional maintenance alerts significantly increased the percent of time drivers monitored the FRD, which may have increased workload and reduced comfort.

The percent of time drivers looked toward the FRD suggests an increase in monitoring workload from attentional maintenance alerts because the AM group spent more time monitoring the FRD. However, because the AutoDrive comfort scores were not significantly different between the AM and Baseline group, the difference in overall comfort scores may not reflect lower comfort from the physical sound but from expectations and interpretations of the attentional maintenance alerts. When participants were asked to describe additional factors that made them uncomfortable, not a single person mentioned discomfort from attention alerts. One person in the AM group mentioned they would have preferred an attention alert before a takeover event. Other comments were related to automation failure, takeovers, and a lack of control. Therefore, the difference in overall comfort may reflect a difference in the way the AM interface affected takeover expectations. Furthermore, a lack of significant differences between groups for eyes off road comfort also suggests the AM interface did not affect comfort while subjects were engaged in the secondary task during AutoDrive. Although the workload for AM drivers may have been higher, comfort scores suggests drivers may have been accepting toward attentional maintenance alerts to help increase their situational awareness and feel in control.

State-contingent warnings theoretically give a time advantage to regain situational awareness before a takeover request (TOR). Therefore, we predicted the SCT group would report higher levels of takeover comfort because they would have higher situational awareness during takeovers (Petersen and colleagues, 2019). However, the Baseline group was significantly more comfortable taking over control from automation compared to both the SCT and AM groups. Because drivers could have experienced takeover events with or without a state-contingent attention alert before a takeover request, drivers may not have known what to expect. If the driver's gaze was toward the FRD within 10 seconds before a takeover request, no state-contingent attention alert would have been given. During these 10 seconds, drivers could have directed their attention back to the secondary task. Even if a SCT driver's gaze was directed toward the FRD, their cognitive resources may have been occupied by the trivia task. A TOR without an attention alert could have come as a surprise. Drivers may have expected an attention alert prior to all takeover alerts. Failure of these expectations may have led to low comfort levels in the SCT group compared to the Baseline. It is also important to note that drivers in the SCT group could only interact with the DMS interface a maximum of four times, once for each takeover event.

A transparent warning system is one in which drivers understand how, when, and why warnings are administered. Transparent warning systems are more trustworthy (Lee and See, 2004). Therefore, a lack of transparency of the SCT interface to warn drivers of takeover events could have caused low levels of AutoDrive comfort as well as the large variation in overall comfort. Because the Baseline group did not receive attention alerts, the overall system may have been more transparent and comfortable during takeover events. Similarly, low transparency of the AM interface could have contributed to the lower takeover comfort experienced by the AM group. Driver's may have expected alerts to help with takeovers. If expectations failed, attentional maintenance alerts may have decreased comfort compared to the Baseline group.

Complete transparency of automation does not always promote proper calibration of trust (Reinmueller and colleagues, 2018). Over time, if drivers begin to understand the DMS, they may over rely on warnings to remind them to look up toward the FRD. According to Hergeth (2016), this type of monitoring behavior demonstrates higher trust in the AM interface. A study by Reinmueller and colleagues (2018) showed that drivers may over rely on adaptive warning systems when they are aware of the occurrence, functionality and reliability of the strategy in order to optimize resources for engagement in a secondary task. If the AM interface was fully transparent and comfortable enough to maintain situational awareness, drivers may have reduced their monitoring workload in order to allocate resources toward the secondary task. This may help explain the variance in the time AM drivers spent looking toward the FRD. A lack of comfort in the interface warning system may also explain the lack of significance in gaze measures between the AM and SCT groups. Lower comfort in the SCT group may have directed visual attention toward the road. However, cognitive attention may have remained on the secondary task as demonstrated by measures of situational awareness. Nevertheless, limitations in the sample size suggest further analysis is required to understand how drivers distribute cognitive attention as there were no significant differences between the SCT and Baseline group for either situational awareness or gaze measurements.

Limitations

There were several limitations of the experiment that should be considered for future studies. Although ANOVA corrects against false significance, the only way to avoid non-significant results that may be significant is to increase the power of the test with more subjects.

More participants would also help keep the variance between groups more consistent, which would help account for individual differences.

Due to time constraints, cross balancing was not an option. Future studies will also need to cross balance events to minimize the effect of the order of events. Non-significant work zone events may be due to the fact that the work zones could be seen before takeover requests were issued. Individual differences in gaze behavior right before the work zone events may have contributed to the lack of significant differences between groups.

It is important to note that although there was no significant difference between the AM and Baseline groups' freeze-probe accuracies, the mean accuracy in the Baseline group was only slightly larger than the mean accuracy for the SCT group. Large variations of situational awareness in the AM and SCT groups may have skewed results. There were also large variations in the AM and SCT groups for the percent of time drivers directed their gaze toward the FRD. Standard deviations of percent of time drivers looked toward the FRD for the SCT group were more than twice as large as the Baseline group. Measures of situational awareness may have been limited by the subjects' working memory. Similarly, measures of trust may be limited by the subjects' long-term memory since they were conducted after the study drive. In addition, participants were not required to answer every question, which may have reduced statistical power. Moreover, large variations suggest that designers need to consider how individual differences affect situational awareness, trust, and interaction with the DMS interface.

A larger study with more participants may find that AM drivers have higher situational awareness and look toward the FRD a higher percent of the time compared to both the SCT an Baseline groups. However, a longer study may also show behavioral adaptations to the attention alerts that decrease gaze percentage and situational awareness over time while an increase in the number of takeover events may increase gaze percentage and situational awareness in the SCT and Baseline groups. Over time, these percentages may converge toward a common number between all groups. Because automation failure is a lot less frequent in the real world, attentional maintenance alerts may still help drivers look toward the FRD more than the Baseline group. Nevertheless, AM drivers may over rely on the DMS warnings to look up toward the FRD. Therefore, although attentional maintenance alerts help engage drivers in their monitoring task, improvements in the level of engagement are warranted to avoid drivers over relying on the DMS warning system.

CHAPTER 7 – CONCLUSIONS

General Conclusion

No other studies to date have used driver monitoring systems to measure gaze behavior and adapt the interface in a level 3 AV to increase situational awareness. This study used Endsley's methods to objectively measure situational awareness and understand the effect of adaptive, camera-based driver monitoring interfaces on drivers in level 3 AVs. The goal was to increase driver situational awareness while drivers were out of the loop and increase takeover response time as well as takeover quality during transfer of control from the AV to the driver.

Future Research

More sophisticated analyses of gaze can help measure situational awareness before takeovers. Louw and Merat (2017) have previously shown that both vertical and horizontal gaze dispersion increase while gaining situational awareness. Gaze has also been shown to highly correlate with measures of trust (Hergeth and colleagues, 2016). Gaze dispersion can also be used as a sensitive measure for cognitive workload (Gold and colleagues, 2016). Future studies may also want to consider using time series analysis to measure gaze and test the effect of DMS interfaces over a longer period of time. A longer study could test how drivers adapt and rely on the interface and could include more takeovers, which would increase statistical power. Comparing gaze behavior and situational awareness over a longer period of time may also provide insight into whether attentional maintenance alerts help drivers direct both cognitive and visual attention toward the FRD or just visual attention. Similar adjustments could be made to test the effects of camera-based driver monitoring warning systems on fatigued drivers.

Measures of situational awareness may be limited by the subjects' working memory. Although there are no better ways to objectively measure situational awareness, subjective measures of situational awareness, such as the Situational Awareness Rating Technique should also be used in combination with SAGAT. Subjective measures of situational awareness provide information on the confidence of one's own situational awareness and have been found to correlate with measures of objective situational awareness and trust (Rousseau and colleagues, 2010; Petersen and colleagues, 2019). Just as studies on trust have shown that proper calibration of trust with trustworthiness can enhance performance (Helldin and colleagues, 2013), proper calibration of actual situational awareness (objective) with perceived situational awareness (subjective) can help enhance performance (Lee, 1999; Rousseau and colleagues, 2010). As Endsley suggests, although subjective situational awareness may be completely independent of objective situational awareness, assessments of subjective situational awareness may provide a critical link between situational awareness and performance (Endsley and colleagues, 1998). Future studies involving feedback from DMSs should consider calibrating subjective situational awareness with objective situational awareness using methods derived from the Qualitative Analysis of Situational Awareness technique as ground truth (Edgar and colleagues, 2018). Driver monitoring interface systems that do not account for individual differences through measures of subjective situational awareness may not be able to properly calibrate situational awareness.

Another limitation of this study lies in measurements of trust, which may be limited by the subjects' long-term memory since they were conducted after the study drive. Just as camera-based driver monitoring systems that measure gaze behavior can be used to calibrate confidence in situational awareness with objective situational awareness, they may also be used to calibrate trust with trustworthiness. Future studies of driver monitoring warning systems may be able to indirectly measure trust by adapting and analyzing a change in the time decrement used to issue attentional alerts. In this study, the time decrement remained constant, but advances in adaptable warning algorithms may enhance measurements of trust and benefit drivers who adversely adapt to constant attention alerts. In other words, future DMSs not only need to help adapt driver trust to the system's trustworthiness, but also adapt the system's transparency, or trustworthiness, to avoid overreliance. However, constant attention alerts from the AM interface could still lead to disuse of the DMS, especially under highly reliable AV capabilities. Parasuraman and Riley (1997) define disuse of automation as the neglect or underutilization of automation, caused by alarms that activate falsely. Overtime, drivers may view DMS alerts as false alarms and find ways to turn them off or ignore the alerts.

In order to avoid disuse and poor calibration of trust and perceived situational awareness, designers must properly develop cooperation between the AV and the driver. Communication and shared control promote cooperation (Flemisch and colleagues, 2016). In this study, communication was limited to binary gaze behavior and attention alerts. Nevertheless, advancements in communication may give drivers a sense of shared control. Because drivers may over rely on a transparent system, communication with the AV may help maintain engagement in the monitoring task. Furthermore, Parasuraman and colleagues (2000) identified four functions automation can control that may or may not be shared with the driver. These are information acquisition, information analysis, action selection, and action implementation. A number of prominent authors have recommended higher levels of automation for information acquisition and analysis, and lower levels of automation for action selection and implementation (Parasuraman and colleagues, 2000). Although AVs disengage drivers from physical control, and therefore, disengage drivers from action implementation, cognitive control (action selection) in the form of decision making can still be shared. Indeed, several authors have shown that decision making is essential in maintaining optimal performance (Endsley and Kiris, 2004; Adams, 1961; Onnasch and colleagues, 2014). In terms of Merat and colleagues (2018) findings, drivers may have difficulty remaining on the loop when action selection and decision making are completely left to the AV. Adaptive automation can be considered a form of nonverbal communication or shared control that helps facilitate decision making and cooperation. However, studies have shown that situational awareness decreases during times when automation is in control (Chen and colleagues, 2017). Therefore, future interface systems that use camera-based driver monitoring systems should help facilitate cooperation through decision making even while the vehicle is in automation in order to enhance situational awareness in level 3 AVs. Future driver monitoring systems may be able to use emotional cues to facilitate communication as well (Izquierdo-Reyes and colleagues, 2018).

Final Conclusion

Humans have historically performed poorly on monitoring tasks, and the problem persists as demonstrated by the Tesla crashes in Willison, Florida and Mountain View, California. Because of the complexity of level 3 AVs, drivers may over rely on the system, which leads to drivers looking away from the FRD. A literature review on vigilance decrement and human monitoring limits looked to apply old and new human factors engineering principles to enhance driver monitoring performance in a level 3 AV. The literature review highlights the importance of feedback based on performance, adaptable automation, decision making, and calibration of trust to avoid vigilance decrement and out of the loop syndrome. However, drivers may also begin to over rely on driver monitoring systems. This creates a cyclical dilemma in which new technology is introduced in order to avoid the problems of old technology, yet the new technology requires additional technology for the same reasons. Findings warrant further research into how to break this cycle of introducing new technology to fix the old.

Nevertheless, camera-based driver monitoring systems have potential to break this trend. Till now, communication between a level 3 AV and driver was unilateral or involved very low levels of bilateral communication. Advancements in decision-based interface design and artificial intelligence may help make the monitoring task more satisfying and engaging. The idea is that when physical tasks are taken away, a mode of feedback is not the only aspect of the driving tasks that is taken away. When physical control is taken away, so is cognitive control. The monitoring task is a low cognitive workload task because there is no opportunity for cognitive control. Therefore, the cognitive task of monitoring associated with physical tasks of driving loses meaning. In other words, because reinforcement requires decision making and a lack of reinforcement leads to vigilance decrement, when there is no opportunity for decision making (cognitive control), motivation for situational awareness may decrease. Drivers then turn toward other tasks in which they do have cognitive control over. However, these tasks may be non-driving related and hazardous during takeovers. Therefore, driving task decisions are necessary to avoid vigilance decrement and maintain situational awareness because they provide cognitive control. Cognitive tasks that require decision making directed toward the overall driving task as a form of cognitive control may enhance awareness of the driving situation. Future research should look to understand how an increase in cognitive control can be designed using driver monitoring systems in order to comfortably reengage drivers in their monitoring responsibilities.

Equally important is the automated vehicle community's vision for designing safe and satisfying AVs. It appears that before moving in any direction, the automated vehicle community needs to decide whether AVs should support safety through satisfying engagement in the driving task or satisfying and safe multi-tasking behavior. These are two views that may contradict each other. Louw and colleagues (2015) suggest that until there is an effective strategy to help drivers regain situational awareness during transitions of control from automation, drivers should be motivated to monitor the driving situation at all times. Moreover, as research has shown, drivers tend to engage in non-driving related tasks in highly automated vehicles (Banks and colleagues, 2018; Buckley and colleagues 2018). Therefore, it appears that for level 3 AVs, multitasking behavior should be discouraged through design of a more satisfying monitoring experience.

The research goal was to establish situational awareness ground truth for the application of adaptable level 3 AV interfaces using camera-based driver monitoring systems. An increase in situational awareness, the percent of time drivers spent looking toward the FRD, and takeover response times in the AM group validified these objectives. Moreover, this research shows the potential of camera-based driver monitoring systems to disengage drivers from secondary tasks and transition attention towards monitoring the FRD in level 3 and higher AVs. Driver monitoring systems can help facilitate bilateral communication between the driver and the AV system. However, comfort scores suggest further research needs to consider how to increase situational awareness without sacrificing user satisfaction. Future research on the frequency of driving related decisions, shared control, and bidirectional communication may direct the design of AVs toward a more situationally aware and satisfying monitoring experience. These preliminary results provide a foundation for the advancement of interface design methodologies necessary to increase situational awareness and takeover performance in level 3 AV.

REFERENCES

Adams, J. A., Stenson, H. H., & Humes, J. M. (1961). Monitoring of Complex Visual Displays— II. Effects of Visual Load and Response Complexity on Human Vigilance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *3*(4), 213–221. https://doi.org/10.1177/001872086100300401

Aghaei, S.A., Donmez, B., Liu, C. C., He, D., Liu, G., Plataniotis, K. N., … Sojoudi, Z. (2016). Smart Driver Monitoring: When Signal Processing Meets Human Factors: In the driver's seat. *IEEE Signal Processing Magazine*, *33*(6), 35–48. https://doi.org/10.1109/MSP.2016.2602379

Ahlstrom, C., Kircher, K., & Kircher, A. (2013). A Gaze-Based Driver Distraction Warning System and Its Effect on Visual Behavior. *IEEE Transactions on Intelligent Transportation Systems*, *14*(2), 965–973. https://doi.org/10.1109/TITS.2013.2247759

Banks, V. A., Eriksson, A., O'Donoghue, J., & Stanton, N. A. (2018). Is partially automated driving a bad idea? Observations from an on-road study. *Applied Ergonomics*, *68*, 138–145. https://doi.org/10.1016/j.apergo.2017.11.010

Boverie, S., Giralt, A., & Le Quellec, J. M. (2008). Diagnostic fusion for in vehicle Driver vigilance assessment. *IFAC Proceedings Volumes*, *41*(2), 12825–12830. https://doi.org/10.3182/20080706-5-KR-1001.02169

Broadbent, D. E. (1977). Levels, Hierarchies, and the Locus of Control. *Quarterly Journal of Experimental Psychology*, *29*(2), 181–201. https://doi.org/10.1080/14640747708400596

Buckley, L., Kaye, S.-A., & Pradhan, A. K. (2018). A qualitative examination of drivers' responses to partially automated vehicles. *Transportation Research Part F: Traffic Psychology and Behaviour*, *56*, 167–175. https://doi.org/10.1016/j.trf.2018.04.012

Chen, S. I., Visser, T. A. W., Huf, S., & Loft, S. (2017). Optimizing the balance between task automation and human manual control in simulated submarine track management. *Journal of Experimental Psychology: Applied*, *23*(3), 240–262. https://doi.org/10.1037/xap0000126

Doshi, A., & Trivedi, M. M. (2012). Head and eye gaze dynamics during visual attention shifts in complex environments. *Journal of Vision*, *12*(2), 9–9. https://doi.org/10.1167/12.2.9

Edgar, G. K., Catherwood, D., Baker, S., Sallis, G., Bertels, M., Edgar, H. E., … Whelan, A. (2018). Quantitative Analysis of Situation Awareness (QASA): modelling and measuring situation awareness using signal detection theory. *Ergonomics*, *61*(6), 762–777. https://doi.org/10.1080/00140139.2017.1420238

Endsley, M. R. (1988a). Design and Evaluation for Situation Awareness Enhancement. *Proceedings of the Human Factors Society Annual Meeting*, *32*(2), 97–101. https://doi.org/10.1177/154193128803200221

Endsley, M. R. (1988b). *Situation awareness global assessment technique (SAGAT)*. 789–795. https://doi.org/10.1109/NAECON.1988.195097

Endsley, M. R., & Kiris, E. O. (1995). The Out-of-the-Loop Performance Problem and Level of Control in Automation. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(2), 381–394. https://doi.org/10.1518/001872095779064555

Eriksson, A., Banks, V. A., & Stanton, N. A. (2017). Transition to manual: Comparing simulator with on-road control transitions. *Accident Analysis & Prevention*, *102*, 227–234. https://doi.org/10.1016/j.aap.2017.03.011

Fitts, P. M. (1947). Psychological Research On Equipment Design. (AD 651795). Washington, DC: Clearinghouse for Federal Scientific and Technical Information.

Flemisch, F., Abbink, D., Itoh, M., Pacaux-Lemoine, M.-P., & Weßel, G. (2016). Shared control is the sharp end of cooperation: Towards a common framework of joint action, shared control and human machine cooperation. *IFAC-PapersOnLine*, *49*(19), 72–77. https://doi.org/10.1016/j.ifacol.2016.10.464

Fraser, D. C. (1957). A Study of Vigilance and Fatigue. Unpublished doctoral dissertation, University of Edinburgh

Geden, M., Staicu, A.-M., & Feng, J. (2017). The impacts of perceptual load and driving duration on mind wandering in driving. *Transportation Research Part F: Traffic Psychology and Behaviour*. https://doi.org/10.1016/j.trf.2017.07.004

Gold, C., Damböck, D., Lorenz, L., & Bengler, K. (2013). Take over!" How long does it take to get the driver back into the loop? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *57*(1), 1938–1942. https://doi.org/10.1177/1541931213571433

Gold, C., Berisha, I., & Bengler, K. (2015). Utilization of Drivetime – Performing Non-Driving Related Tasks While Driving Highly Automated. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *59*(1), 1666–1670. https://doi.org/10.1177/1541931215591360

Gold, C., Körber, M., Lechner, D., & Bengler, K. (2016). Taking Over Control From Highly Automated Vehicles in Complex Traffic Situations: The Role of Traffic Density. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *58*(4), 642–652. https://doi.org/10.1177/0018720816634226

Gonçalves, J., & Bengler, K. (2015). Driver State Monitoring Systems– Transferable Knowledge Manual Driving to HAD. *Procedia Manufacturing*, *3*, 3011–3016. https://doi.org/10.1016/j.promfg.2015.07.845

Hancock, P. A., & Warm, J. S. (1989). A Dynamic Model of Stress and Sustained Attention. *Human Factors*, 20.

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

Hergeth, S., Lorenz, L., Vilimek, R., & Krems, J. F. (2016). Keep Your Scanners Peeled: Gaze Behavior as a Measure of Automation Trust During Highly Automated Driving. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *58*(3), 509–519. https://doi.org/10.1177/0018720815625744

Izquierdo-Reyes, J., Ramirez-Mendoza, R. A., Bustamante-Bello, M. R., Navarro-Tuch, S., & Avila-Vazquez, R. (2018). Advanced driver monitoring for assistance system (ADMAS): Based on emotions. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, *12*(1), 187–197. https://doi.org/10.1007/s12008-016-0349-9

Kaber, D. B., Liang, Y., Zhang, Y., Rogers, M. L., & Gangakhedkar, S. (2012). Driver performance effects of simultaneous visual and cognitive distraction and adaptation behavior. *Transportation Research Part F: Traffic Psychology and Behaviour*, *15*(5), 491–501. https://doi.org/10.1016/j.trf.2012.05.004

Kircher, K., & Ahlström, C. (2009). Issues related to the driver distraction detection algorithm AttenD. First International Conference on Driver Distraction and Inattention. Gothenburg, Sweden, 2009.

Kircher, K., & Ahlström, C. (2013). The driver distraction detection algorithm AttenD. In Driver Distraction and Inattention: Advances in Research and Countermeasures (Vol. 1, pp. 327-348). Ashgate Publishing.

Koesdwiady, A., Soua, R., Karray, F., & Kamel, M. S. (2017). Recent Trends in Driver Safety Monitoring Systems: State of the Art and Challenges. *IEEE Transactions on Vehicular Technology*, *66*(6), 4550–4563. https://doi.org/10.1109/TVT.2016.2631604

Körber, M., Baseler, E., & Bengler, K. (2018). Introduction matters: Manipulating trust in automation and reliance in automated driving. *Applied Ergonomics*, *66*, 18–31. https://doi.org/10.1016/j.apergo.2017.07.006

Lee, J. D. (1999). Measuring Driver Adaptation to In-Vehicle Information Systems: Disassociation of Subjective and Objective Situation Awareness Measures. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *43*(18), 992–996. https://doi.org/10.1177/154193129904301810

Lee, J. D., & See, K. A. (2004). Trust in Automation: Designing for Appropriate Reliance. *Human Factors*, 31.

Liang, Y., & Lee, J. D. (2010). Combining cognitive and visual distraction: Less than the sum of its parts. *Accident Analysis & Prevention*, *42*(3), 881–890. https://doi.org/10.1016/j.aap.2009.05.001

Louw, T., Merat, N., & Jamson, H. (2015). *Engaging with Highly Automated Driving: To be or Not to be in the Loop?* 190–196. https://doi.org/10.17077/drivingassessment.1570

Louw, T., & Merat, N. (2017). Are you in the loop? Using gaze dispersion to understand driver visual attention during vehicle automation. *Transportation Research Part C: Emerging Technologies*, *76*, 35–50. https://doi.org/10.1016/j.trc.2017.01.001

Lu, Z., Happee, R., Cabrall, C. D. D., Kyriakidis, M., & de Winter, J. C. F. (2016). Human factors of transitions in automated driving: A general framework and literature survey. *Transportation Research Part F: Traffic Psychology and Behaviour*, *43*, 183–198. https://doi.org/10.1016/j.trf.2016.10.007

Mackworth, N. H. (1948). The Breakdown of Vigilance during Prolonged Visual Search <sup/>. *Quarterly Journal of Experimental Psychology*, *1*(1), 6–21. https://doi.org/10.1080/17470214808416738

Massar, S. A. A., Lim, J., Sasmita, K., & Chee, M. W. L. (2016). Rewards boost sustained attention through higher effort: A value-based decision making approach. *Biological Psychology*, *120*, 21–27. https://doi.org/10.1016/j.biopsycho.2016.07.019

May, J. F., & Baldwin, C. L. (2009). Driver fatigue: The importance of identifying causal factors of fatigue when considering detection and countermeasure technologies. *Transportation Research Part F: Traffic Psychology and Behaviour*, *12*(3), 218–224. https://doi.org/10.1016/j.trf.2008.11.005

McGehee, D. V., Mazzae, E. N., & Baldwin, G. H. S. (2000). Driver Reaction Time in Crash Avoidance Research: Validation of a Driving Simulator Study on a Test Track. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *44*(20), 3-320-3–323. https://doi.org/10.1177/154193120004402026

Merat, N., & Jamson, A. H. (2009). *How Do Drivers Behave in a Highly Automated Car?* 514–521. https://doi.org/10.17077/drivingassessment.1365

Merat, N., Seppelt, B., Louw, T., Engström, J., Lee, J. D., Johansson, E., … Keinath, A. (2018). The “Out-of-the-Loop” concept in automated driving: proposed definition, measures and implications. *Cognition, Technology & Work*. https://doi.org/10.1007/s10111-018-0525-8

Mok, B. K.-J., Johns, M., Lee, K. J., Ive, H. P., Miller, D., & Ju, W. (2015, June). *Timing of unstructured transitions of control in automated driving*. 1167–1172. https://doi.org/10.1109/IVS.2015.7225841

Mouloua, M., Parasuraman, R., & Molloy, R. (1993). Monitoring Automation Failures: Effects of Single and Multi-Adaptive Function Allocation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *37*(1), 1–5. https://doi.org/10.1177/154193129303700102

NTSB (National Transportation Safety Board). (1995). Marine Accident Report, Grounding of the Panamanian Passenger Ship Royal Majesty on Rose and Crown Shoal near Nantucket, Massachusetts, June 10, 1995. Marine Accident Report NTSB/MAR-97/01. Retrieved from http://www.ntsb.gov

NTSB (National Transportation Safety Board). (2017). Collision Between a Car Operating With Automated Vehicle Control Systems and a Tractor-Semitrailer Truck Near Williston, Florida May 7, 2016 . Highway Accident Report NTSB/HAR-17/02 . Washington, DC: NTSB. Retrieved from http://www.ntsb.gov

NTSB (National Transportation Safety Board). (2018a). Preliminary Report Issued for Investigation of Fatal, Mountain View, California, Tesla Crash. Preliminary Report Highway HWY18FH011. Retrieved from http://www.ntsb.gov

NTSB (National Transportation Safety Board). (2018b). Preliminary Report Released for Crash Involving Pedestrian, Uber Technologies, Inc., Test Vehicle. Preliminary Report Highway HWY18MH010. Retrieved from http://www.ntsb.gov

Onnasch, L., Wickens, C. D., Li, H., & Manzey, D. (2014). Human Performance Consequences of Stages and Levels of Automation: An Integrated Meta-Analysis. *Human Factors: The*

Journal of the Human Factors and Ergonomics Society, *56*(3), 476–488. https://doi.org/10.1177/0018720813501549

Parasuraman, R., Molloy, R., & Singh, I. L. (1993). Performance Consequences of Automation-Induced “Complacency.” *The International Journal of Aviation Psychology*, *3*(1), 1–23. https://doi.org/10.1207/s15327108ijap0301_1

Parasuraman, R., Mouloua, M., & Molloy, R. (1996). Effects of Adaptive Task Allocation on Monitoring of Automated Systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *38*(4), 665–679. https://doi.org/10.1518/001872096778827279

Parasuraman, R., & Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *39*(2), 230–253. 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 Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, *30*(3), 286–297. https://doi.org/10.1109/3468.844354

Parasuraman, R., & Manzey, D. H. (2010). Complacency and Bias in Human Use of Automation: An Attentional Integration. Human Factors: The Journal of the Human Factors and Ergonomics Society, 52(3), 381–410.

Payre, W., Cestac, J., & Delhomme, P. (2014). Intention to use a fully automated car: Attitudes and a priori acceptability. *Transportation Research Part F: Traffic Psychology and Behaviour*, *27*, 252–263. https://doi.org/10.1016/j.trf.2014.04.009

Petersen, L., Robert, L., Yang, J., & Tilbury, D. (2019). Situational Awareness, Driver’s Trust in Automated Driving Systems and Secondary Task Performance. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3345543

Rees, A., Wiggins, M. W., Helton, W. S., Loveday, T., & O'Hare, D. (2017). The Impact of Breaks on Sustained Attention in a Simulated, Semi-Automated Train Control Task: Breaks and sustained attention. *Applied Cognitive Psychology*, *31*(3), 351–359. https://doi.org/10.1002/acp.3334

Reinerman-Jones, L., Matthews, G., & Mercado, J. E. (2016). Detection tasks in nuclear power plant operation: Vigilance decrement and physiological workload monitoring. *Safety Science*, *88*, 97–107. https://doi.org/10.1016/j.ssci.2016.05.002

Reinmueller, K., Koehler, L., & Steinhauser, M. (2018). Adaptive warning signals adjusted to driver passenger conversation: Impact of system awareness on behavioral adaptations. *Transportation Research Part F: Traffic Psychology and Behaviour*, *58*, 242–252. https://doi.org/10.1016/j.trf.2018.06.013

Rousseau, R., Tremblay, S., Banbury, S., Breton, R., & Guitouni, A. (2010). The role of metacognition in the relationship between objective and subjective measures of situation awareness. *Theoretical Issues in Ergonomics Science*, *11*(1–2), 119–130. https://doi.org/10.1080/14639220903010076

SAE. (2018). Surface Vehicle Recommended Practice: Taxonomy and Definitions for Terms Related to Driving Automation Systems for On- Road Motor Vehicles, SAE Standard J3016, USA, 2018. Retrived from: https://www.sae.org

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

Siebert, F. W., Oehl, M., Höger, R., & Pfister, H.-R. (2013). Discomfort in Automated Driving – The Disco-Scale. In C. Stephanidis (Ed.), *HCI International 2013 - Posters' Extended Abstracts* (pp. 337–341). Springer Berlin Heidelberg.

Sheridan, T. B. (2002). Humans and Automation: System Design and Research Issues. Hoboken, NJ, US: John Wiley & Sons Inc; Santa Monica, CA, US: *Human Factors and Ergonomics Society*.

Strand, N., Nilsson, J., Karlsson, I. C. M., & Nilsson, L. (2014). Semi-automated versus highly automated driving in critical situations caused by automation failures. *Transportation Research Part F: Traffic Psychology and Behaviour*, *27*, 218–228. https://doi.org/10.1016/j.trf.2014.04.005

Teigen, K. H. (1994). Yerkes-Dodson: A Law for all Seasons. *Theory & Psychology*, *4*(4), 525–547. https://doi.org/10.1177/0959354394044004

Vlakveld, W., van Nes, N., de Bruin, J., Vissers, L., & van der Kroft, M. (2018). Situation awareness increases when drivers have more time to take over the wheel in a Level 3 automated car: A simulator study. *Transportation Research Part F: Traffic Psychology and Behaviour*, *58*, 917–929. https://doi.org/10.1016/j.trf.2018.07.025

Wickens, C. D., & Kessel, C. (1979). The Effects of Participatory Mode and Task Workload on the Detection of Dynamic System Failures. *IEEE Transactions on Systems, Man, and Cybernetics*, *9*(1), 24–34. https://doi.org/10.1109/TSMC.1979.4310070

Yamani, Y., Horrey, W. J., Liang, Y., & Fisher, D. L. (2015). Sequential In-Vehicle Glance Distributions: An Alternative Approach for Analyzing Glance Data. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *57*(4), 567–572. https://doi.org/10.1177/0018720814560225

Yang, S., Kuo, J., & Lenné, M. G. (2018). Analysis of Gaze Behavior to Measure Cognitive Distraction in Real-World Driving. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *62*(1), 1944–1948. https://doi.org/10.1177/1541931218621441

Yerkes, R. M., & Dodson, J. D. (1908). The Relation of Strength of Stimulus to Rapidity of Habit-Formation. *Journal of Comparative Neurology and Psychology*, *18*(5), 459–482. https://doi.org/10.1002/cne.920180503

Zeeb, K., Härtel, M., Buchner, A., & Schrauf, M. (2017). Why is Steering Not the Same as Braking? The Impact of Non-Driving Related Tasks On Lateral and Longitudinal Driver Interventions During Conditionally Automated Driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *50*, 65–79. https://doi.org/10.1016/j.trf.2017.07.008

APPENDIX A: NON-SIGNIFICANT MEASURES AND RESULTS

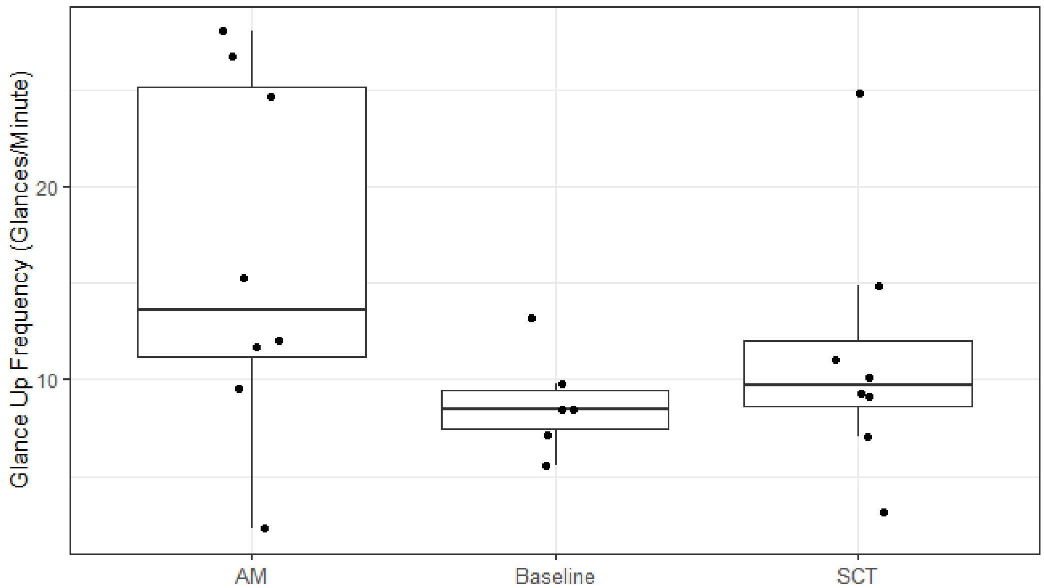

Figure A 1. Boxplot showing median and 25-75 percent quantiles for frequency of glances toward the FRD per minute.

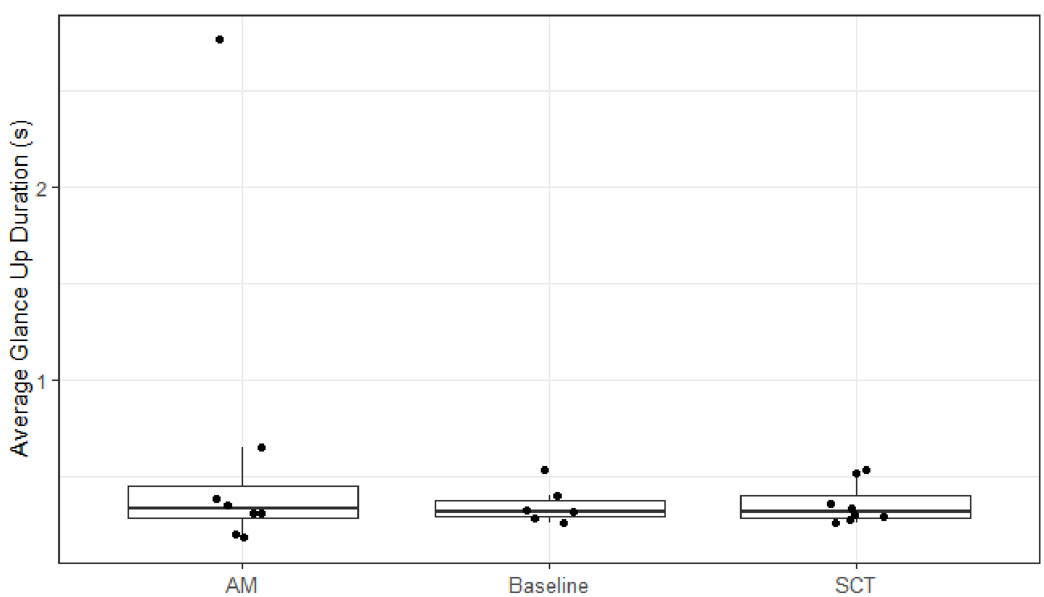

Figure A 2. Boxplot showing median and 25-75 percent quantiles for the duration of glances toward the FRD with each point representing subject average.

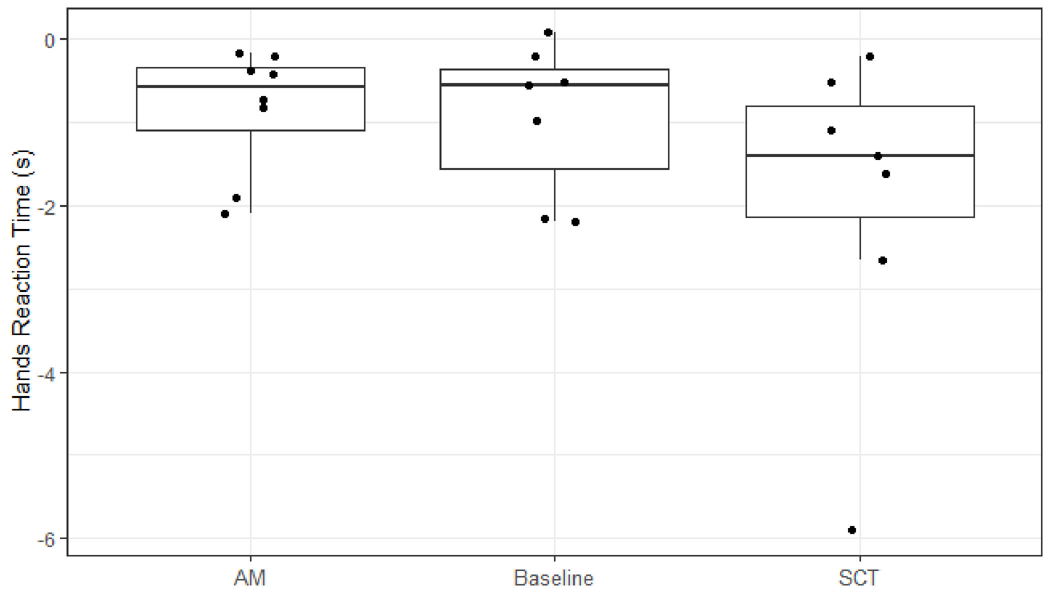

Figure A 3. Boxplot showing median and 25-75 percent quantiles for hand response time during work zone events with each point representing subject mean.

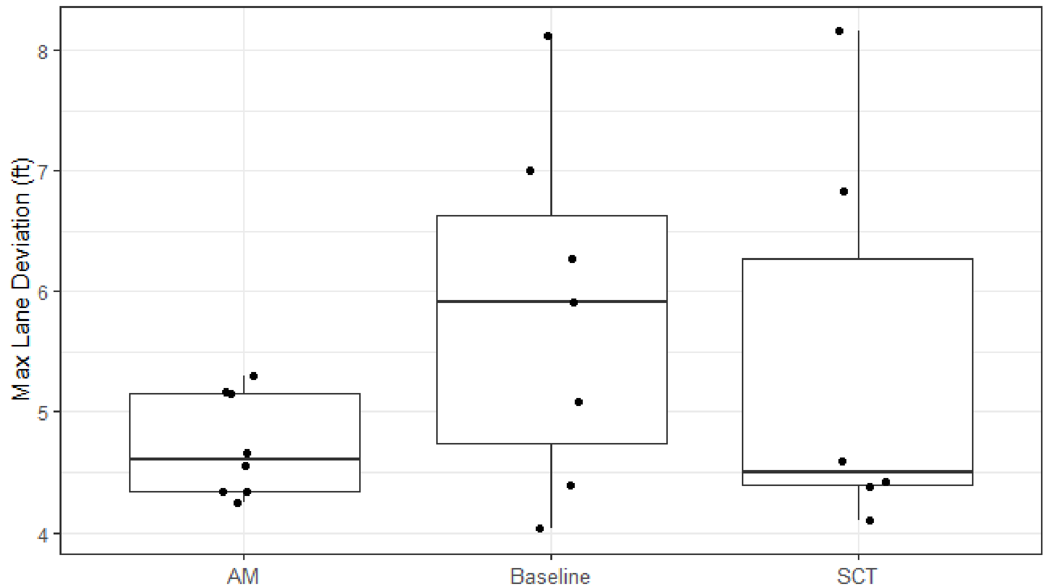

Figure A 4. Boxplot showing median and 25-75 percent quantiles for maximum lane deviation during dropout events with each point representing subject mean.

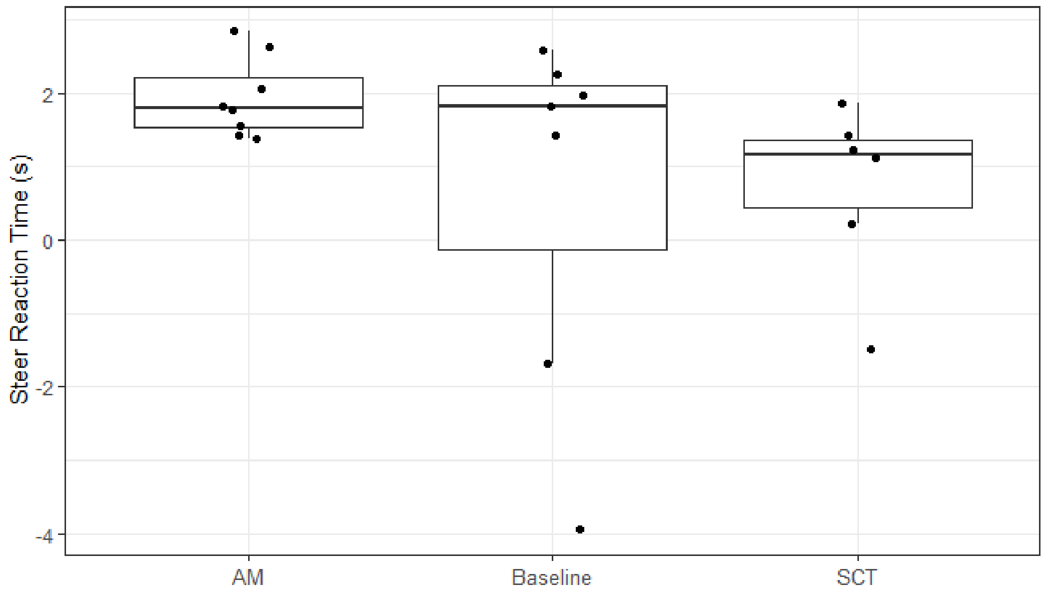

Figure A 5. Boxplot showing median and 25-75 percent quantiles for steering response time during work zone events with each point representing subject mean.

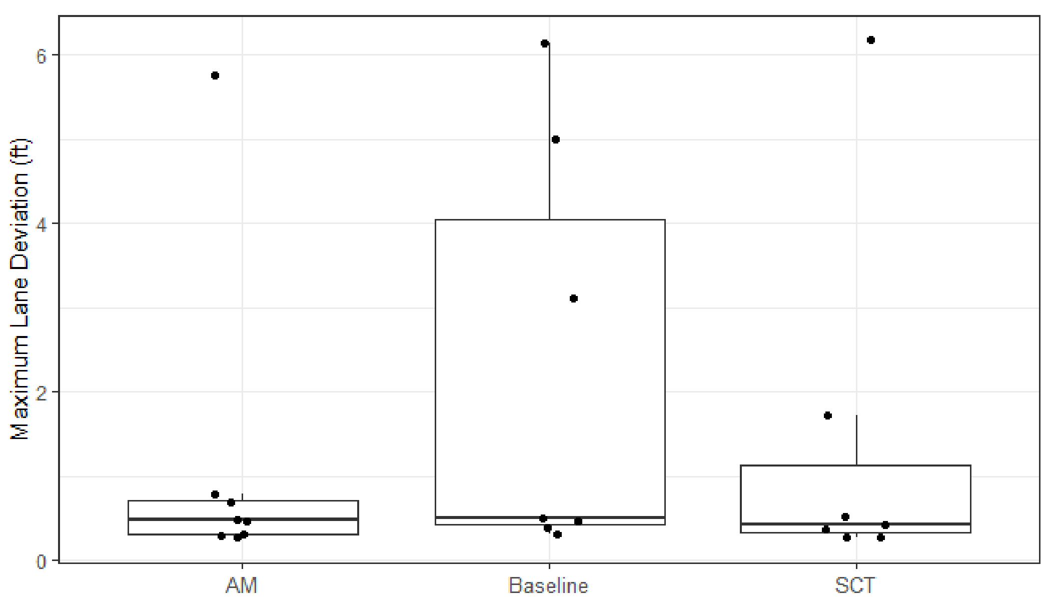

Figure A 6. Boxplot showing median and 25-75 percent quantiles for maximum lane deviation during work zone events with each point representing subject mean.

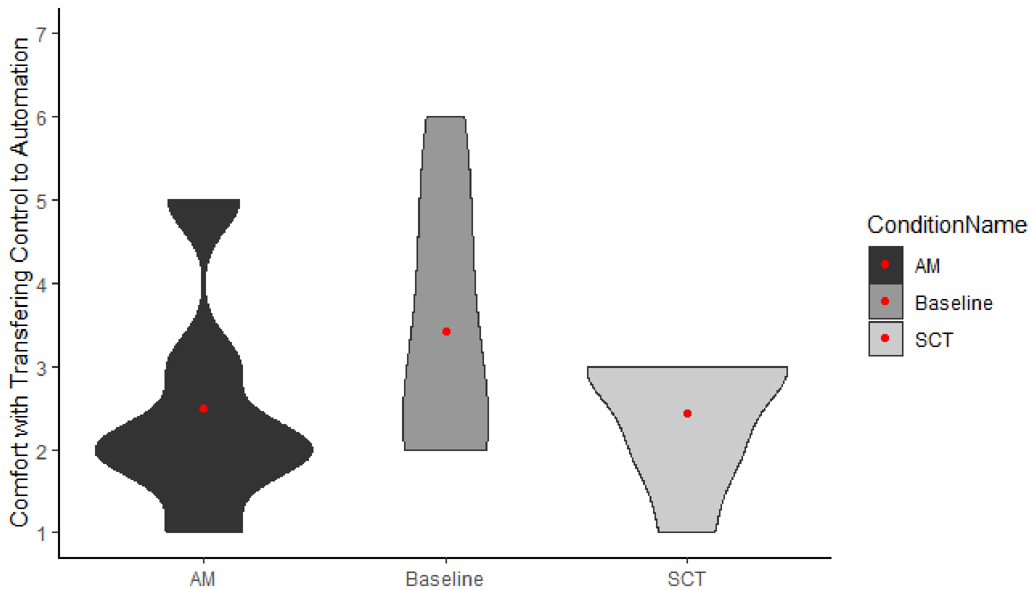

Figure A 7. How comfortable did you feel transferring the vehicle into AutoDrive?

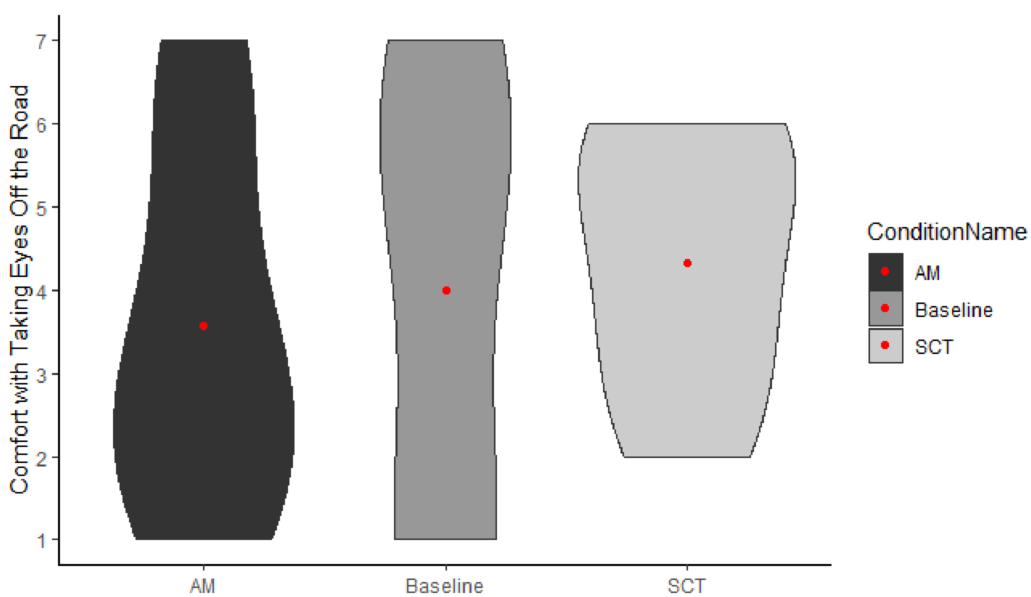

Figure A 8. Compared to commuting in your regular vehicle, how likely is it that you would take your eyes off the road for several seconds while driving with AutoDrive engaged?

APPENDIX B: SUBJECTIVE EVALUATION OF HUMAN TRUST IN AV SYSTEMS

Please rate using this 1-7 scale:

In general, how comfortable did you feel during the drive?

How comfortable did you feel transferring the vehicle into AutoDrive?

How comfortable did you feel resuming manual control from AutoDrive during the drive?

How comfortable did you feel when AutoDrive failed and you had to retake control?

Compared to driving manually, how comfortable did you feel driving in AutoDrive?

Did the vehicle drive differently in AutoDrive than when you were manually controlling it?

Compared to commuting in your regular vehicle, how likely is it that you would take your eyes off the road for several seconds while driving with AutoDrive engaged?