Behavioral Adaptations of Drivers to Autonomous Systems: Evaluating Intermediate and Carryover Effects

Erika E. Miller

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2018

Reading Committee:

Linda Ng Boyle, Chair

John D. Lee

Don MacKenzie

Program Authorized to Offer Degree:

Civil & Environmental Engineering

University of Washington

Abstract

Behavioral Adaptations of Drivers to Autonomous Systems:
Evaluating Intermediate and Carryover Effects

Erika E. Miller

Chair of the Supervisory Committee:
Chair and Professor Linda Ng Boyle
Industrial & Systems Engineering and Civil & Environmental Engineering

Autonomous vehicle systems have elicited the attention of car manufacturers, consumers, policy makers, and the media as they offer societal, environmental, and economic benefits. However, prolonged exposure to these systems may lead drivers to adapt to these systems in ways not anticipated by the designer; resulting in unintended safety consequences. To explore this issue, a longitudinal driving simulator study was conducted to evaluate behavioral adaptations due to exposure to an active lane keeping system. In this study, performance before, during, and after exposure to the semi-autonomous system was compared. Forty-eight participants (30 treatment, 18 control) completed a series of eight drives across three separate days. Treatment participants were exposed to approximately 40 minutes of baseline [manual] driving, 80 minutes of semi-automated driving, and 40 minutes of post-automated [manual] driving. A control group was exposed to approximately 160 minutes of manual driving, but otherwise identical study procedures, in order to provide a reference for time on task effects. Changes in secondary task engagement (number completed and accuracy), driving performance (SDLP and TTC), cognitive workload (TDRT response time and miss rate) and eye glance behavior (mean glance duration, 90$^{\text{th}}$ percentile glance duration, total eyes-off-road time, and percent long glances) were modeled using generalized linear mixed models. Cluster analysis techniques were used to examine the effects of trust in automation on behavioral adaptations. The findings of this dissertation suggest that drivers began to rely on automation for support and experienced adverse effects when the system

was removed. Moreover, drivers with higher self-reported trust in the autonomous system experienced the largest degradations in performance and were associated with inherently more risky driving habits. By identifying the associations between trust and behavioral adaptations over time, vehicle systems, infrastructure, and educational programs can be designed to support appropriate use and attention allocation, in order to minimize adverse effects during handover and takeover of vehicle control.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Number Page

LIST OF TABLES

GLOSSARY

ACC: Adaptive Cruise Control

ADAS: Advanced Driver Assistance System

ANOVA: Analysis of Variance

DRT: Detection Response Task

EOR: Eyes-Off-Road

GD: Glance Duration

GEE: Generalized Estimating Equation

GLMM: Generalized Linear Mixed Model

IRB: Institutional Review Board

IRR: Incidence Rate Ratio

ISO: International Organization for Standardization

IVIS: In-Vehicle Information System

LK(S): Active Lane Keeping System

LMM: Linear Mixed Model

MPH: Miles Per Hour

MS: Milliseconds

NADS: National Advanced Driving Simulator

NASA TLX: NASA Task Load Index

NHTSA: National Highway Traffic Safety Administration

SD: Standard Deviation

SDLP: Standard Deviation of Lateral Position

TDRT: Tactile Detection Response Task

TTC: Time to Collision

TUKEY HSD: Tukey's Honest Significant Difference

ACKNOWLEDGMENTS

To my advisor, Dr. Linda Ng Boyle – I would like to express my most sincere gratitude for all of the guidance, encouragement, patience, and opportunities you have provided me. I will always be grateful for the fantastic example you have been as a mentor.

To each member of my committee – I appreciate all of your support throughout my graduate studies. Dr. John D. Lee, thank you for every opportunity you shared with me to learn about human factors, teaching, and research. Dr. Don MacKenzie, thank you for your great mentorship on statistical applications and teaching techniques. Dr. Franziska Roesner, thank you for extending your expertise and time to provide insightful feedback.

To all my peers in the Human Factors and Statistical Modeling Lab – Thank you for all your countless hours of support, inspiration and friendship. Yiyun Peng, Yuqing Wu, Huimin Xiong, Tolu Abe, Xingwei Wu, Edith Guo, Chun-Cheng Chang, Steven Hwang, Ning Li, and Kelsey Gragg, I appreciate all of the feedback and discussions that helped me define my research. Warren Sprecher, Yoel Tekle, Madison Keck, and David Prendez, I appreciate all of your help with data collection and data reduction.

To my friends and family – Thank you for all the love, faith, and laughs along this journey.

This dissertation would not be possible without each and every one of you. Thank you.

Chapter 1
INTRODUCTION

As market penetration of autonomous vehicle systems continues to increase, the roles of all transportation users will continue to evolve. Humans are likely to adapt their driving behaviors in both positive and adverse ways as their interactions with these systems changes. Previous research suggests that these adaptations can manifest as misplaced trust (J. D. Lee & See, 2004), misuse and disuse of systems (Parasuraman & Riley, 1997), skill atrophy (Casner et al., 2014), and increased risky behavior (Wochinger et al., 2008). Moreover, vehicle fleet turnover is relatively slow and different manufacturers design systems with differing functionality and limitations. This resulting mixed equipage vehicle fleet has the potential to undermine the benefits and efficiency that automation affords, as vehicles of varying levels of automation interact and driver expectations change based on the vehicle they are in.

There is limited research in quantifying adaptations in driver behavior due to prolonged exposure and transfer effects of removing or changing these systems, particularly with active lane keeping assistance systems. It is important to understand how the human operator will adapt their behaviors, in order to effectively design systems, infrastructure, and policy for successful exchange of control between the vehicle and the driver.

A longitudinal driving simulator study was conducted with 48 drivers from the Seattle, Washington area. There were 30 participants exposed to an active lane keeping system and 18 participants used as a control group with no exposure to automation. The control group provided a baseline for expected changes in performance due to increased familiarity with study procedures (i.e., time on task effects). Drivers in both groups completed eight drives, across three different days within a seven-day period. During drives two through seven, participants completed visual-manual distracting tasks on an In-Vehicle Information System (IVIS). Participants in the treatment group were exposed to the lane keeping system in drives three through six. Changes in drivers' driving performance, cognitive workload, risk perception, and trust in automation were examined across drives. Driving performance was measured using a driving simulator, cognitive workload was assessed using a Tactile Detection Response Task (TDRT), risk perception was evaluated based on engagement with IVIS tasks and eye glance patterns, and trust in automation was quantified using a questionnaire.

1.1 Research Aims

There were three research aims in this study, targeted at understanding behavioral adaptations and the role that trust plays in driving these adaptations. In this dissertation, behavioral adaptations are characterized as changes overtime in an individuals' driving performance (i.e., lateral and longitudinal control of the vehicle), cognitive workload (as measured by their performance in responding to a detection task), engagement in a secondary distracting task (i.e., frequency and accuracy of engagement), and eye glance behavior (i.e., durations of glances off road). These behavioral adaptations were measured as *(1)* immediate effects, such as occurring immediately during exposure to the lane keeping system, and *(2)* carryover effects, such as occurring after exposure and withdrawal of the lane keeping system.

Aim 1: Are there changes in performance and risk perception during exposure to a lane keeping system? Baseline performance measures for each driver were collected in drives 1 and 2, which were used as baseline for comparison of performance in subsequent drives. Changes in performance were quantified using driving performance measures and cognitive workload measures. Changes in drivers' willingness to engage in risky behaviors was quantified using IVIS task completion measures. This research aim evaluates the immediate effects of exposure.

Aim 2: Do behavioral adaptations persist after the lane keeping system is withdrawn? The last two drives (i.e., drives 7 and 8) required drivers to return back to manual driving. Changes in performance and risk perception in these two drives relative to baseline were considered carryover effects due to exposure and withdrawal of the active lane keeping system. Similar to Research Aim 1, driving performance, cognitive workload, and IVIS task completion measures were evaluated. Additionally, changes in eyes-off-road patterns were examined to quantify adaptations in risk perception. This research aim evaluates the carryover effects of exposure.

Aim 3: How do drivers self-reported levels of trust impact behavioral adaptations? Drivers in the treatment group (i.e., exposed to the automation) were asked to fill out a questionnaire regarding their trust in the lane keeping system at the end of each of the three days. Their responses to these questions were used to quantify their trust in the system and group participants based on similar levels of trust. Differences in the effect sizes of behavioral adaptations and inherent measures of riskiness were then compared across trust groups.

Chapter 2
BACKGROUND

This chapter summarizes the current state of knowledge regarding vehicle automation and the relation to driver behavior. Previous research has evaluated the potential benefits and implications of assimilating large share market penetration of autonomous vehicle systems. However, the majority of this research to date forecasts these effects without considering the behavioral adaptations of drivers due to prolonged exposure to automation. This chapter highlights the importance of these issues pertaining to human-automation interactions and current gaps in research.

2.1 Vehicle Automation

2.1.1 Levels of Automation

The term autonomous vehicle systems refers to the technology within the vehicle that aids the vehicle in sensing, and potentially responding, to its environment. There are various degrees that a vehicle can be equipped with these systems, and as such SAE defined six levels of automation, ranging from 0 to 5, in an effort to provide conformity to the industry. The definitions for the SAE levels of vehicle automation are summarized below, as adapted from SAE J3016_201609 (2016):

- *Level 0 - No Automation*: Human performs all driving tasks at all times.
- *Level 1 - Driver Assistance*: Vehicle can aid with either lateral or longitudinal control, but not both, and human performs remainder of driving task.
- *Level 2 - Partial Automation*: Vehicle can perform lateral and longitudinal control, but the human must monitor the environment and supervise the system at all times.
- *Level 3 - Conditional Automation*: Vehicle can perform driving task, but human must be ready to takeover control.

- *Level 4 - High Automation*: Vehicle can perform entire driving task under limited conditions, with no expectation of the human to respond to a request to intervene.
- *Level 5 - Full Automation*: Vehicle provides full control and there is no expectation of the human to intervene.

By removing the human from the control loop and replacing this control with automation, it reduces the likelihood of human error occurring or changes the types of human errors. If properly designed and implemented, automation has the potential to increase safety and network efficiency. However, even through level 4 automation, there still exists the potential for the human to intervene with the driving task. Thus, it is still necessary to consider the human in the control loop, as well as the reality that machines are prone to errors and limitations as well (e.g., speed limitation, poor visibility, faded pavement markings).

2.1.2 Advanced Driver Assistance Systems, ADAS

Advanced Driver Assistance Systems (ADAS) are often referenced when discussing vehicle automation. ADASs are systems that aid the driver with the task of driving. These systems can range from providing real-time advisory information (e.g., alerting of hazards), to intervening [partial] vehicle control (e.g., crash avoidance, parking assist), to providing full control in the longitudinal and/or lateral direction for extended periods of time. Examples of real-time advisory information include blind spot monitoring, navigation cues, and rear-view cameras. Support and control technologies can provide longitudinal, lateral, and/or angular rate control (Gustafsson, 2009). Longitudinal control includes, but is not limited to, cruise control, adaptive cruise control, brake assistance, forward collision warning, and forward collision mitigation. Examples of lateral control include lane departure warning and lane-keeping assistance. Systems for angular rate include roll stability control and rollover detection. Autopilot, summon, and platoon driving are examples of conditional to high automation. Figure 2.1 further illustrates how these countermeasure technologies relate to a hypothetical crash sequence.

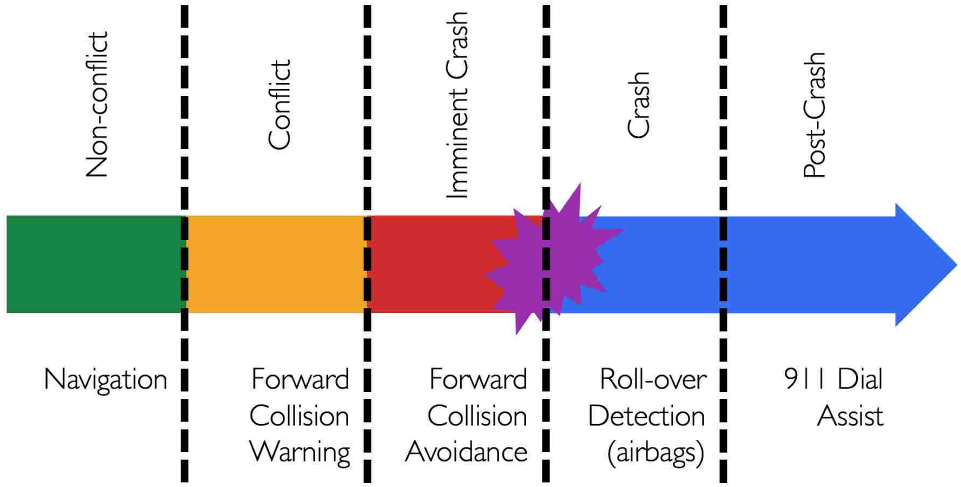

Figure 2.1: Example of ADAS for Crash Sequence, Adapted from Gordon et al. (2010)

2.1.3 Takeover and Handover

Takeover and handover are important constructs of human-automation interaction. Handover, or handoff, refers to the vehicle system releasing control, with the expectation that the driver will assume control. Takeover refers to the driver initiating the transfer of control away from the vehicle in order to intervene with the automation. In instances of handover, it is imperative to capture the drivers' attention, as it is likely they are not prepared to resume control. That is, as drivers remain out of the control loop for an extended period of time, their vigilance to the driving task will likely decrease. Therefore, their reaction time and execution of the appropriate maneuver will likely be slow and less effective. Takeover often has less of a safety critical implication, as it often refers to the driver choosing to intervene. This generally occurs based on the driver's understanding of the system's capabilities. If their mental model of the system's limitations is accurate, then takeover is warranted; but if they have an inaccurate overestimation of the system's limitations, then their takeover may negate the potential benefits of the autonomous system.

The overall task of takeover can be summarized by the human information processing model (see Figure 2.2), which is adapted from human information processing models presented by J. D. Lee and See (2004) and J. D. Lee et al. (2017). This model describes the process of a driver to decide to takeover control of the vehicle. As one would expect, this is highly dependent on the context; for example a driver might takeover control when in a school zone, but not for a similar situation on a rural highway. Additionally, this decision to takeover is highly dependent on the perception capabilities of the driver, such as what information they are able to detect through their senses. All of this information filters into the cognitive process to help formulate a decision; their decision making is also affected by past experiences, time and workload constraints, and their confidence in their own driving and the systems capabilities. Ultimately, they formulate a decision, and in the context of takeover, that decision is to distribute vehicle control between some percentage of automated control and some percentage of manual control (e.g., yield longitudinal control to the vehicle and lateral control to manual). This is an iterative process, where the driver receives feedback from the system (e.g., lead vehicle brakes, approach intersection, adverse event changes their trust), and has to reiterate this decision making process.

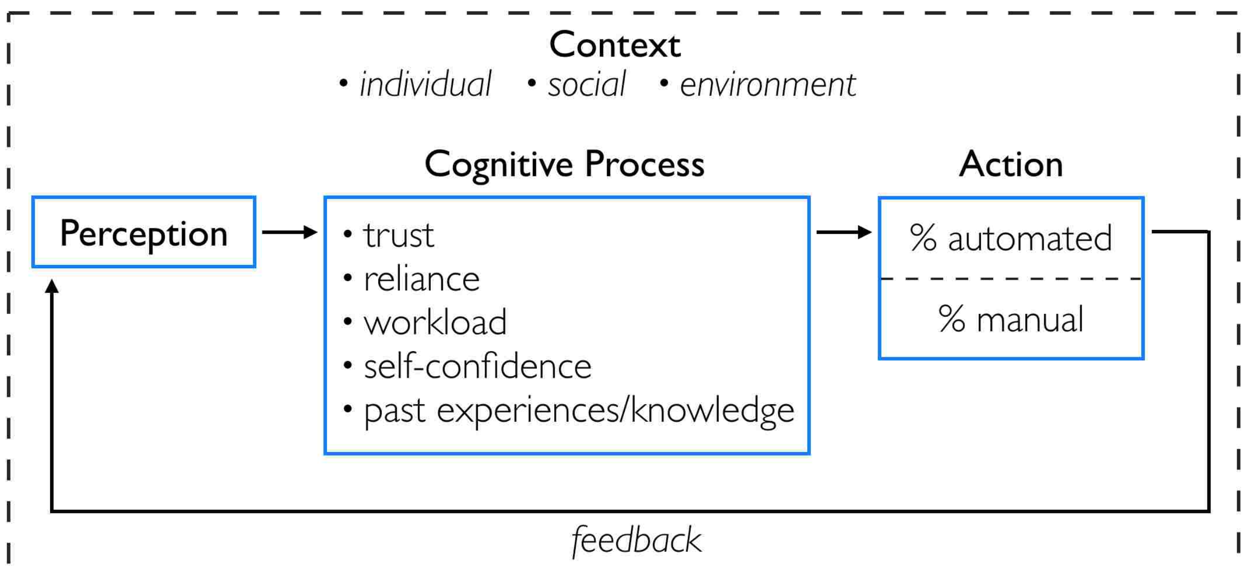

Figure 2.2: Human Information Processing Model for Takeover, Adapted from J. D. Lee and See (2004) and J. D. Lee et al. (2017)

2.1.4 Effects of Automation

The design of autonomous systems have been focused on the assistance to the driver and the ability to alleviate high driving demands. Under mass market penetration and assuming minimal system malfunctions, it is projected that automation can reduce crash frequency and the associated economic cost of crashes, reduce fuel consumption, increase route efficiency, expand parking possibilities, and increase mobility to under-served populations (e.g., young, elderly, disabled) (Fagnant & Kockelman, 2015; NHTSA, 2011; Skottke et al., 2014; Levitan & Bloomfield, 1998; Gouy et al., 2014; VanArem et al., 2006). For example, automation can assist drivers with safety concerns related to sight obstructions, and help direct attention to critical elements, detecting activity deficits, and supporting longitudinal and lateral control (Staubach, 2009; J. D. Lee et al., 2008). Automation can also provide additional support by enhancing the skills of vulnerable driving groups, such as those with disability, inexperience, or old age (Brookhuis et al., 2001). This is of particular importance as human error tends to account for the primary factor in over 90% of vehicle crashes, where approximately 41% are attributed to recognition errors, 33% are attributed to decision errors, and 11% are attributed to performance errors (NHTSA, 2015).

Alternatively, some literature suggests that the integration of vehicle automation may alter travel demands and patterns enough to offset many of the potential gains. For example, increased mobility for previously non-driving demographic groups may lead to new vehicle trips (Wadud et al., 2016; Sivak & Schoettle, 2015; Litman, 2017). New trips with empty vehicle occupancy may also be generated as autonomous vehicles are summoned, parking remotely, or picking-up/dropping-off car sharing individuals (Litman, 2017). Additionally, alleviating the driving task from the human occupant can allow travel time to be better used to increase productively; likely encouraging travelers to choose longer trips and lead to increased sprawl (Stephens et al., 2015; Litman, 2017). Ultimately, this increase in vehicle trips could actually lead to an increase in energy consumption and emissions (Wadud et al., 2016). Understanding how behaviors will shift is a critical component to successful integration of autonomy into the vehicle fleet.

Behavioral Adaptations

Prolonged exposure to these systems may lead drivers to adapt to the system in ways not anticipated by the designer. These unintended safety consequences can relate to changes in driving performance, risk compensation, shifts in attention allocation, mode confusion, or misplaced trust. This phenomenon of the user changing their behavior in a way unintended by designers is often referred to as behavioral adaptation (OECD, 1990).

These behavioral adaptations can manifest as net positive, neutral, or negative effects on driving performance, see Figure 2.3. Ideally, these systems should be designed such that they work synergistically with the human operator, where the system aids the driver and the human is available to intervene when needed. However, these behavioral adaptations have the potential to shift towards a negative impact on driving performance if the driver begins to misuse and disuse these systems. This dissertation is aimed at understanding how these negative adaptations may materialize. Behavioral adaptations occur over time; changes over the initial exposure period as operators learn the system is often referred to as the learning curve. After long durations of exposure, automaticity of the behavior begins to occur, which is often referred to as habit. Previous studies have suggested that habit formation occurs, on average, after approximately 66 days (Lally et al., 2009). Behavioral adaptations that occur during this middle period may be attributed to malleable attentional resource theory, which posits that attentional capacity fluctuates with demand (Young & Stanton, 2002). For example, a driver may experience reduced attentional resources as a result of vehicle automation reducing the driving task demand.

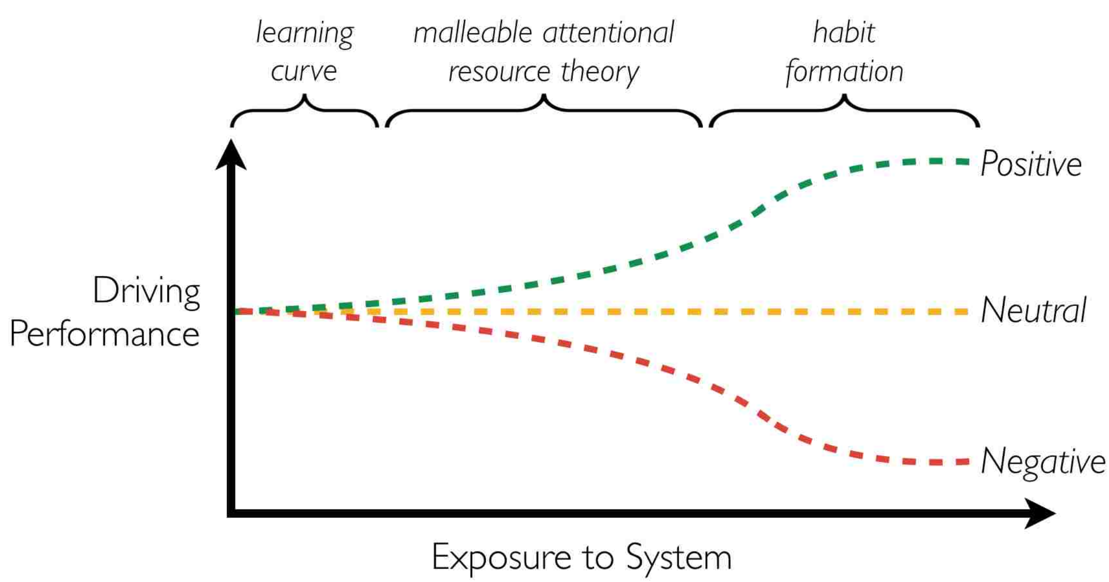

Figure 2.3: Theoretical Effect of Exposure to Automation on Driving Performance

There has been previous research in comparing driving behaviors under manual, semi-automated, and highly automated driving conditions. Past studies that examine semi-automated systems have focused on Adaptive Cruise Control (ACC), but the findings were not always consistent. Some studies show that situation awareness (Ma & Kaber, 2005) and detection accuracy of a stimulus (Funke et al., 2007) increase (i.e., better performance) for drivers when ACC is engaged (DeWinter et al., 2014). In contrast to those findings, Rudin-Brown and Parker (2004) found that ACC increased response time to a hazard detection task (i.e., worse performance). However, they did note that their finding could have been confounded with secondary task engagement, as drivers also had increased secondary task accuracy under ACC (Rudin-Brown & Parker, 2004). These studies suggest that the presence of ACC may alleviate some of the demands of the driving task while it is engaged, but that drivers may also offset the positive performance increase with more secondary task engagement.

Other studies have evaluated ACC in conjunction with Lane Keeping Systems (LK or LKS) as a proxy to measure highly automated conditions in a driving simulator. In a between-subjects design, Strand et al. (2014) found that drivers under highly automated conditions had lower minimum time to collision (TTC), lower minimum time headway (THW), and longer response time when responding to an automation failure as compared to drivers using only ACC. Gouy et al. (2014) showed that drivers in manual mode adapted their time headway in correlation to an automated vehicle platoon in the next lane. Interestingly, once the platoon was not present, drivers would revert back to their baseline driving. In the meta-analysis by DeWinter et al. (2014), all studies (n=9) that examined secondary tasks showed an increase in secondary task involvement under highly-automated driving when compared to manual. This increase in engagement with secondary tasks can lead to safety implications when handover or takeover is needed. As Merat et al. (2012) reported, drivers performed poorly when having to takeover control from highly automated driving while engaging in a secondary task. However, when not engaged in a secondary task, drivers driving with automation were equally able to respond to critical events (e.g., reduce speed, change lanes) as drivers driving manually. Several studies have also found that highly automated driving conditions lead to decreased driver arousal, increased drowsiness, slower reaction times, and reduced eyes-on-road gazes (Jamson et al., 2013; Barnard & Lai, 2010; Carsten et al., 2012; DeWinter et al., 2014). These findings suggest that the increase in automation may adversely effect safety when the driver needs to regain control after being out of the control loop for an extended period of time.

The majority of these studies were based on cross sectional designs and focused on behavioral differences for driving with the autonomous system engaged versus manual driving. However, less information is known about the long-term effects of exposure to automation and more importantly, the impact on performance when the automation is removed. This latter situation is crucial to understand as the vehicle may not be able to handle all critical situations and there may be times when the driver is in a vehicle with a different level of automation (e.g., car sharing service, rental car, family member's car).

Moreover, the current implementation of autonomous systems across vehicle makes and models varies greatly in terms of distribution and functionality (e.g., capabilities and limitations), even for seemingly similar systems. These discrepancies across systems and interfaces can cause the driver to become confused or unknowingly react inappropriately to a cue (Robinson et al., 2011). As the research presented above suggests, humans are highly adaptive and thus likely to adapt their behaviors as the role of the driver changes with the dissemination of automation. These differences in system functionality are likely to elicit confusion and these adaptations in behavior due to exposure may lead to failures of trust. As a result, drivers may unintentionally misuse different vehicles systems, and thereby undermine the safety benefits that automation affords.

2.1.5 Trust and Acceptance of Automation

Designing the technology and infrastructure is not the only construct in ensuring successful implementation of autonomous vehicle systems. It is also imperative to understand how humans will interact with the system as they adapt over prolonged use. Even aggressive forecasts do not have full saturation of fully autonomous vehicles occurring within the next 20 years (Lavasani et al., 2015; Litman, 2017). Therefore, this transition period towards higher levels of automation will extend a substantial amount of time and designing for handoff and takeover must be considered.

The decision of the driver to takeover control and their performance during handover is influenced by their trust. Trust can be indicative of how drivers interact with a system (Kircher et al., 2014); drivers are likely to intervene when their trust falls below some threshold (Muir, 1994; Rudin-Brown & Parker, 2004), while they are likely to become complacent and fail to intervene when they trust the system too much (Parasuraman, 2000; Rudin-Brown & Parker, 2004). J. D. Lee and See (2004, p. 51) defines trust as "the attitude that an agent will help achieve an individuals goal in a situation characterized by uncertainty and vulnerability."

Ghazizadeh et al. (2012) used the Technology Acceptance Model to discuss how perceived usefulness and ease of use are predictors in behavioral intent to use, acceptance, and attitudes towards automation. Ghazizadeh et al. (2012), as well as J. D. Lee and See (2004), contend that automation should be trustable (e.g., understandable and transparent) in order to reduce automation misuse, disuse, and complacency. In this context, misuse refers to inappropriate [over]-reliance on automation while disuse refers to the rejection of the automations capabilities (J. D. Lee & See, 2004).

By developing appropriate trust between the system and the driver, the benefits of automation can be maximized. A study by Seppelt and Lee (2007) revealed that drivers were able to respond more quickly and more consistently when provided continuous information about the state of the Adaptive Cruise Control (ACC) system. This concept of providing feedback and transparency of the automation can facilitate appropriate trust. In a similar context, Verberne et al. (2012) found that drivers were more trusting and accepting of ACC systems that provided information during takeover, as opposed to systems with takeover events that provided no information.

However, as drivers begin to trust, or even over trust their system, they may become unavailable to regain control when needed. Drivers are likely to engage in non-driving related activities with increased automation (Strand et al., 2014; Carsten et al., 2012). It is important to understand how drivers allocate attention to the driving task as they are removed from the control loop. Takeover and handover often require the driver to act within milliseconds but yet, this reentry into the control loop can take seconds depending on the system design and driver's attentiveness (Merat & Lee, 2012). This shift of attention away from the driving task yields more safety concerns as many drivers are also often unaware of system limitations. In a survey study with ACC owners, 72% of respondents reported that they were not aware of any limitations about their ACC system and a large percentage had incorrectly thought their ACC system would help them avoid a collision when the system in fact would not be effective (Jenness et al., 2008).

It is also important to consider the human-automation interaction as dynamic; as summarized in the previous section, drivers are likely adapting their behaviors with increased exposure to the automation. This was demonstrated in the above mentioned studies that evaluated changes in driving behaviors as drivers move between manual, semi-automated (e.g., ACC), and highly automated driving conditions. Additional studies that support this notion have found that drivers reduced their time headways for periods of manual driving after decoupling from fully automated convoys (Eick & Debus, 2005). Young and Stanton (2007) noted that automation led to a significant increase in secondary task performance. Similarly, Flemisch et al. (2008) reported that drivers increased their engagement with secondary tasks for increasing levels of automation, and that drivers reported lower workload with higher automation, as measured through NASA TLX.

2.2 In-Vehicle Information Systems

In-Vehicle Information Systems (IVIS) integrate information, entertainment, and communication into a central location within the vehicle that has traditionally not been available to the driver. These systems can provide information about the driving environment (e.g., navigation, traffic), media control (e.g., radio, connect external media players), climate control (e.g., cabin temperature), communications (e.g., phone calls, text messaging), and even Internet access. A well-designed IVIS interface can improve safety, efficiency, situational awareness, and reduce anxiety (Vashitz et al., 2008). However, this increase in access and quantity of information can also increase mental workload and distraction (Kaber et al., 2012). Similar to vehicle automation, IVISs are becoming increasingly more prevalent in vehicles, which is why it is important to evaluate driver behavior to automation with the potential influence of IVISs.

IVISs are predominantly a visual-manual interface, as they display visual information and provide controls for the driver to manually manipulate. Therefore these systems are of particular concern as visual-manual tasks have the highest crash odds ratios as compared to other secondary tasks (i.e., visual, manual, cognitive, and pairwise combinations) (Heikkinen et al., 2013; NHTSA, 2013). In the 100-Car Naturalistic Driving Study, it was reported that engagement in visual and/or manual complex tasks were associated with a three times higher crash/near-crash risk, as compared to attentive driving (i.e., no engagement in distraction) (Klauer et al., 2006). As drivers engage in tasks associated with their IVIS, their driving performance may be degraded. Thus, when drivers are exposed to automation, IVIS tasks can be a confounding covariate to performance and behavioral adaptations.

This can be further explained by Multiple Resource Theory, which describes the limited capacity of humans to perceive, process, and respond to tasks across modalities and channels (Wickens, 2002; Kahneman, 1973). According to this theory, task interference occurs when concurrent tasks overlap across the dimensions of mental resources - stages, sensory modalities, codes, and visual channels (Wickens, 2002). This resource competition can occur as drivers engage in secondary tasks while driving, such as with IVISs.

2.3 Measuring Driver State and Performance

Driver state and performance variables can be used to measure behavioral adaptations. In the driving literature, a common method to assess driving performance is to use vehicle kinematic data, such as in the studies previously discussed evaluating measures such as time headway, speed, lateral deviations, etc. Metrics that also directly measure the drivers' state, such as cognitive workload and eye glance patterns, can provide further insight in understanding confounding factors that may influence their interaction with autonomous systems.

2.3.1 Cognitive Workload

The driving task is a complicated task, requiring drivers to perceive, process, and respond to complex information in a dynamic context. As such, the driving task requires a great deal of attention from the driver. However, increased cognitive workload can impair attention allocation and information processing capabilities (Brookhuis & deWaard, 2010;

Faure et al., 2016). Cognitive workload refers to the demand imposed by tasks on the humans information processing capabilities (i.e., limited mental resources) (Wickens, 2008). Cognitive workload has implications on transportation safety, as increased driver workload has been associated with driving impairments (Strayer et al., 2015).

Detection Response Tasks (DRT), which are a variant of Peripheral Detection Tasks, are commonly adopted in the driving domain to evaluate cognitive workload. In a Detection Response Task, participants press a button every time they identify a somewhat randomly appearing target/stimulus. Cognitive workload is quantified based on response time and detection accuracy. The most common Detection Response Tasks include the Tactile Detection Response Task (TDRT), Head-mounted Detection Respond Task (HDRT), and Remote Detection Response Task (RDRT). Where the TDRT utilizes a vibrating stimulus taped to the participant's body (generally their neck), and an illuminating light stimulus either mounted on the participant's head (HDRT) or remotely in their forward vision (RDRT). Previous studies have used TDRT, HDRT, and RDRT to evaluate visually demanding secondary tasks (Cooper et al., 2016) and auditory-vocal tasks (Ranney et al., 2014) in the driving context. Ranney et al. (2014) reported that TDRT, as compared to HDRT and RDRT, was more sensitive to changes in cognitive workload in a driving simulator setting. TDRT has also been used in studies measuring load induced by voice control systems and n-back tasks (Large et al., 2016; Chang et al., 2017; Miller et al., 2018).

2.3.2 Eye Glance Behavior

Previous studies have shown that eye glance behavior is correlated with crash risk, which is not surprising since driving is a highly demanding visual task. Peng et al. (2013) reported that eyes-off-road time influenced lane-keeping ability, or more specifically was associated with increased Standard Deviation of Lane Position (SDLP). Results from the 100-Car Naturalistic Driving Study showed that 2 second glances away from the roadway increased near-crash/crash risk by at least two times, as compared to baseline [non-distracted] driving (Klauer et al., 2006). Similarly, a naturalistic study on commercial motor vehicle drivers reported a positive correlation between tasks associated with increased risk and long eyes-off-road time (Olson et al., 2009). NHTSA's Visual-Manual Driver Distraction Guidelines suggest using eyes-off-road glances in evaluating in-vehicle systems, as this metric often correlates with crash risk (NHTSA, 2013; Boyle et al., 2013).

2.4 Gaps in Literature

Much of the current research, as summarized previously, has focused on ACC or approximated to highly automated systems, rather than evaluating lane keeping systems alone. This is relevant, as several vehicle manufacturers offer lane keeping systems that do not need to be used in conjunction with ACC. While there has been substantial research on driver behavior using ACC, lane keeping (LK) systems can also foster negative behavioral adaptations (Breyer et al., 2010). Moreover, advancements in technology have made lane keeping systems a somewhat basic system to implement in the vehicle, as evident by the numerous manufacturers that currently offer these systems at affordable prices or even standard; including, but not limited to, Acura, Audi, Buick, Cadillac, Chevrolet, Ford, Honda, Jeep, Lexus, Subaru, Toyota, Volkswagen, Volvo. Lane keeping systems are an important consideration as these systems have a great potential to improve safety. Lane departure events often provide minimal forgiveness, as lateral clearance on roadways is often restricted. In fact, lane departure crashes constitute a high proportion of the total number of crashes, and are often more severe. The Federal Highway Administration (FHWA), based on the NHTSA Fatality Analysis Reporting System (FARS), estimated that 54% of all traffic fatalities in the US in 2014 were a result of roadway departure crashes. This corresponds to 17,791 fatalities, where a roadway departure was defined as a vehicle crossing an edge line or centerline (FHWA, 2016).

Also, as noted previously, there is a general inconsistency across findings relating to the effects of automation on driver behavior, and this may be largely due to experimental design. Specifically, several studies have utilized between subject designs to compare manual versus automated conditions. However, without meticulous matching of participants, it is difficult to infer that differences are specific to the effects of automation. Even in studies that use within subject designs, much of the studies collect cross sectional data (e.g., one continuous drive or data measured within one day). As such, these studies may not be capturing behavioral adaptations, but rather effects of learning a new system. Many of these cross sectional designs compare pre-exposure (manual driving) with during exposure (automated), but do not consider carryover effects of removing or changing these systems (e.g., decoupling from automation). Several studies have also considered the effects of automation on secondary task engagement, however many of the secondary tasks used are not representative of common tasks drivers would likely divert their attention to as automation aids their driving. For example, tasks have included visual tracking of a target, n-back tasks, or tests used in cognitive assessments (e.g., trail making). Rather, it is more likely drivers will begin to engage with infotainment tasks (e.g., music, texting, reading).

Previous studies have had narrow scopes in automation type, adaptation time (i.e., exposure and withdrawal), performance measures, and inclusion of confounding factors. This dissertation is unique in that it uses a holistic approach to measure and analyze behavioral adaptations of drivers. This is accomplished by administering extended periods of exposure and withdrawal time, using between subject measures to account for time on task effects, and using within subject measures to account for behavioral adaptations and risk perception across a broad range of driver state and performance measures.

2.5 Research Objectives

The main objectives of this dissertation is to evaluate the relation between behavioral adaptations, changes in risk perception, and trust in automation as drivers are exposed to an active lane keeping system. The scope of this dissertation includes evaluating intermediate and carryover effects of exposure, in order to understand how drivers may perform during handover and takeover in future contexts. These research objectives are further defined by three research aims.

Aim 1: Are there changes in performance and risk perception during exposure to a lane keeping system? This dissertation aims at establishing a methodology for an experimental design and analytical method that can capture a broad range of behavioral adaptations, such that a similar methodology could be deployed to evaluate various other autonomous vehicle systems. It is hypothesized that behavioral adaptations will manifest in terms of changes in driving performance, cognitive workload, and willingness to engage in a distracting task while drivers are exposed to a semi-autonomous system. It is also expected that drivers will exhibit changes in behavior as they gain comfort and familiarity with study procedures, therefore a control group is used to provide a reference for these expected time on task effects.

Aim 2: Do behavioral adaptations persist after the lane keeping system is withdrawn? This dissertation seeks to expand the focus of behavioral adaptations to include carryover effects as drivers return to manual driving. It is hypothesized that drivers will learn to rely on these systems, potentially in ways they are not even aware of. Therefore, it is important to consider how drivers will perform when they are required to return to manual driving.

Aim 3: How do drivers self-reported levels of trust impact behavioral adaptations? Trust is an important construct in the decision to takeover and performance during handover. This dissertation aims at providing a method for capturing trust and correlating trust with behavioral adaptations. It is hypothesized that different degrees of trust will impact interactions with these systems, and as such it is important to consider the effect of trust on intermediate and carryover effects of exposure on behavioral adaptations.

Chapter 3
EXPERIMENTAL DESIGN

This chapter outlines the experimental design employed for collecting the data. This study utilized a longitudinal driving simulator study, which was conducted at the University of Washington in Seattle, WA in February-April 2017. An active lane keeping system was used to evaluate behavioral adaptations and trust to a semi-autonomous vehicle system. A control group was used to measure baseline changes in performance, commonly referred to as time on task effects (i.e., gaining familiarity with the study). An In-Vehicle Information System (IVIS) was used to induce distraction, a Tactile Detection Response task (TDRT) was used to measure cognitive workload, a driving simulator was used to collect driving performance measures, a video camera was used to capture eye glance behaviors, and a questionnaire was used to quantify trust. This study was approved by the Institutional Review Board (IRB) at University of Washington. Informed written consent was obtained from each participant at the beginning of the first day of their involvement.

The study was a 2 (Automation Group: Control, Treatment) $\times$ 3 (Drive: Pre, During, Post Automation) $\times$ 3 (Age Group: Younger, Middle, Older) $\times$ 2 (Gender: Male, Female) $\times$ 4 (Task Difficulty: Easy, Medium, Hard, Hardest) $\times$ 3 (Road Type: Straight, Curve, Hill Curve) repeated measures mixed factorial design.

3.1 Participants

The study included 48 participants balanced across three age groups (younger: 25-34, middle: 35-44, older: 45-54 years old). There were 18 (6 younger, 6 middle, 6 older aged) participants in the control group (i.e., only manual driving throughout). There were 30 (10 younger, 10 middle, and 10 older aged) participants in the treatment group (i.e., intervention of a lane keeping system). Gender was balanced across age group and automation treatment groups, see Table 3.1. All participants had a valid US driver's license, drove at least 3,000 miles annually, and were fluent in English. Participants were compensated \$20 for the first day, \$30 for the second day, and \$50 for the third day, for a total of \$100 for completing entire study.

Table 3.1: Number of Participants in Each Group

Automation Group	**Younger** *(25-34)* **Male**	**Female**	**Middle** *(35-44)* **Male**	**Female**	**Older** *(45-54)* **Male**	**Female**	**Total**
Control	3	3	3	3	3	3	18
Treatment	5	5	5	5	5	5	30

3.2 Apparatus

3.2.1 Driving Simulator

A National Advanced Driving Simulator (NADS) miniSim fixed-based quarter cab driving simulator was used in this study (see Figure 3.1). The simulator display was composed of three plasma 42-inch widescreen monitors with a horizontal field of view of 140 degrees and vertical field of view of 30 degrees. Data was collected from the simulator at 60 Hz.

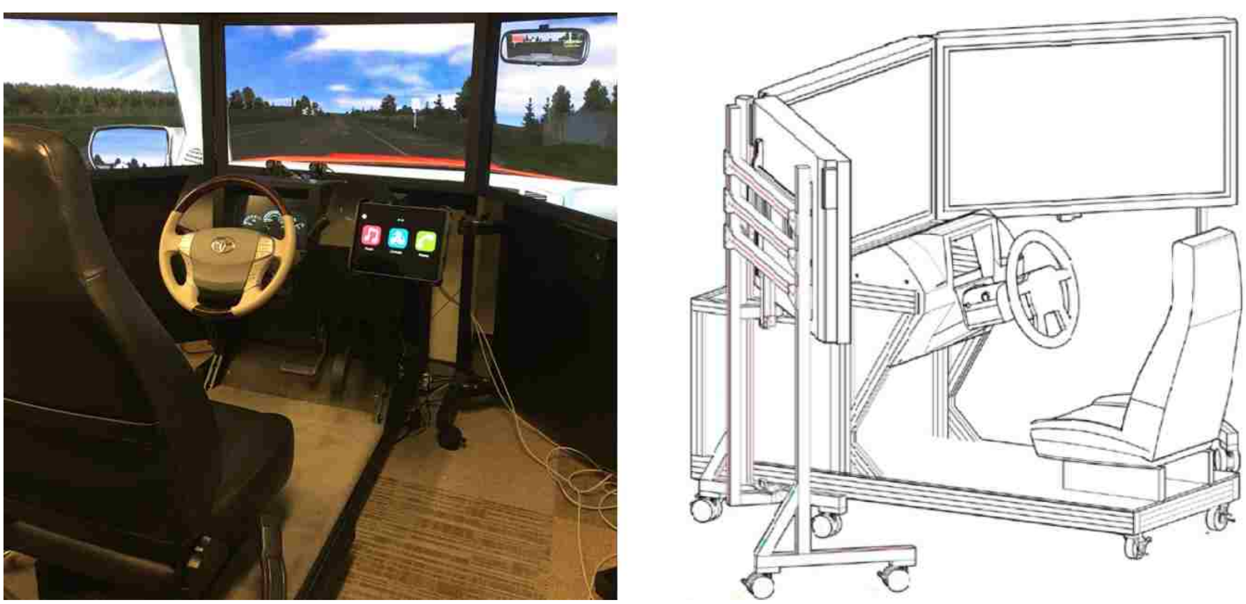

Figure 3.1: Driving Simulator Used for the Study *(left)* and Schematic *(right, from NADS)*

The driving scenario used in this study was developed specifically for this dissertation. The miniSim software Tile Mosaic Tool, TMT, (version 1.7.5.4) was used for creating the visual environment and Interactive Scenario Authoring Tool, ISAT, (version 1.8.0) was used for developing the scenario logic. There were two versions of the driving scenario, one with lane keeping enabled throughout and the second without the lane keeping algorithm, but otherwise these two scenarios were identical.

The driving scenario was a two–lane (one lane in each direction) undivided rural road with one lead vehicle and no intersections. The geometric configuration included straight segments, wide curved segments, and two moderate hills (each with an uphill and downhill grade) with slight horizontal curvature (see Figure 3.2). Each drive took approximately 20 minutes to complete; the end of the drive was based on distance traveled rather than time, thus some participants finished slightly faster or slower than 20 minutes. The posted speed limit was 50 miles per hour (mph) and the lead vehicle traveled at a normally distributed speed of 50 mph with a standard deviation of 2.5 mph. However, in order to allow the participant to reach target speed in the beginning of the drive, the lead vehicle was programmed to maintain a 3 second gap for the first three-fourths of a mile. Participants were instructed to keep a safe following distance behind the lead vehicle and never overtake it. Each vehicle was approximately 5.8 feet wide and the lanes were 12 feet wide.

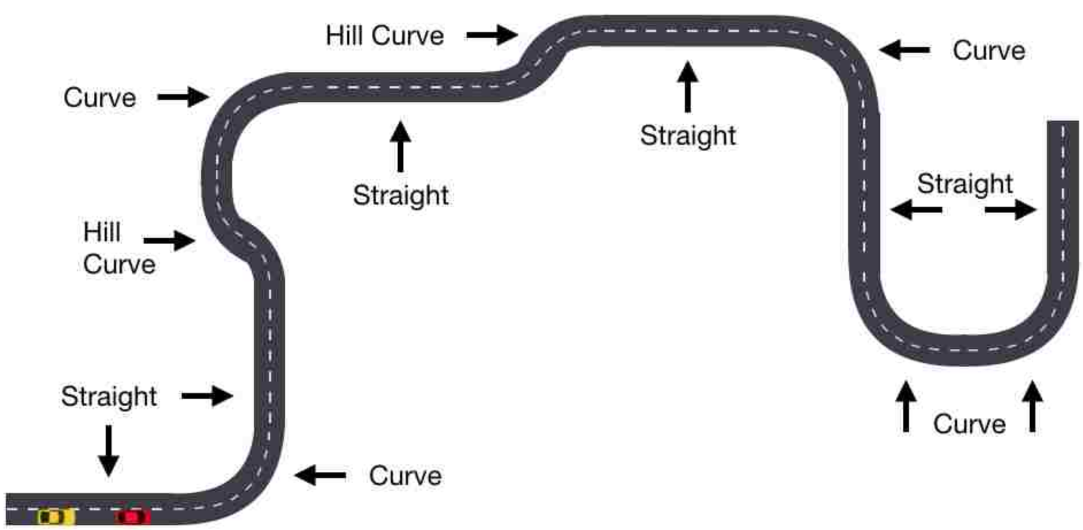

Figure 3.2: Scenario Roadway Configuration (not to scale)

The algorithm for the active lane keeping system was programmed such that the vehicle would follow the center of the lane, which meant participants did not need to steer the vehicle. Participants were able to override the system by moving the steering wheel, however if the vehicle deviated beyond the lane boundaries, the system provided force feedback on the steering wheel to keep it in the lane. No audio alert was provided for lane departure events. For drives with the lane keeping system enabled, the participants only had to control the longitudinal control (i.e., brake and accelerator) of the vehicle. Participants were told this information about how the lane keeping system operated and given the freedom to choose how they engaged with the steering wheel during these drives (i.e., steer or not steer).

Each drive was partitioned into 13 different segments for most of the analyzes (i.e., mean values for performance measures of each participant were aggregated for each of the 13 segments). The 13 segments were defined based on the geometric roadway type, but each section were approximately 1.25 miles long. There were four straight, five curved (2 right, 3 left), and two hill (each with an uphill and downhill component) segments included in analysis. There were two segments excluded from analysis, which were the straight segments at the beginning and end of the drive. The beginning segment was excluded to allow drivers to achieve target speed and get comfortable with the scenario tasks. The end segment was excluded as to avoid any differences in driving behavior when bringing the vehicle to a stop.

There were three practice scenarios developed for this study, one performed at the beginning of each day. All three of the practice scenarios were designed to take approximately ten minutes to complete. The roadway was similar to the main study drive, with a lead vehicle and was predominately composed of straight road segments. One practice drive was only manual driving (used for the control group on all three days); one practice drive had the lane keeping system turn on approximately halfway through the drive (used for the treatment group on day one); the third practice drive had the lane keeping system engaged for the entire drive (used for the treatment group on day two and three). The practice drive provided the driver with exposure to manual driving, then added the TDRT, then added IVIS tasks. The lane keeping system was then also engaged for the first practice drive for the treatment group, otherwise the lane keeping was on the entire time (practice drive 2 and 3 for treatment) or never engaged (all practice drives for control group).

3.2.2 In-Vehicle Information System, IVIS

A 7–inch (800 x 480 pixels) capacitive touchscreen was mounted to the right of the steering wheel at about a 20 degree visual angle for the driver. This touchscreen was used in drives 2 through 7 to provide visual-manual distracting tasks. An application was developed specifically for this study using C# on Microsoft Visual Studio. The application was designed to emulate an In-Vehicle Information System (IVIS) that was representative of systems currently embedded in vehicles. All tasks were completed using visual-manual input (i.e., no voice control).

There were five different types of tasks ranging across four difficulty levels (see Table 3.2): Contact (easy), Playlist (easy), Radio (medium), Climate (hard), and Dial (hardest). Task difficulty was determined during pilot testing on four undergraduate students and assessed using NASA TLX. There were a total of 150 unique tasks (30 different tasks for each task type) and the order was randomly generated by the application for each day, such that no task was repeated within the same day. The application logged task activity to a CSV file. Each time a task was completed, data was recorded for the task start time, end time, task identification number, and whether the task was completed correctly.

Table 3.2: IVIS Task Descriptions

Task	**Difficulty**	**Example**
Contact	Easy	"Call contact Julie Reed"
Playlist	Easy	"Select song Rocket Man, by Elton John, from music playlist"
Radio	Medium	"Tune to FM station 98.1, and set volume to 21"
Climate	Hard	"Set temperature to 68°, fan level to high, and rear defrost on"
Dial	Hardest	"Dial phone number 1-800-blue-car"

Participants were instructed to complete IVIS tasks at a pace they were comfortable with. Participants were cued to begin the tasks approximately two minutes into the drive. A computer automated voice provided task instructions and participants were given five seconds after the instruction to repeat the audio once more. The task would begin by displaying the home screen (see Figure 3.3) after the audio instruction was finished. A task was completed once the participant pressed the submit button at the top right of the touchscreen. The next task automatically began ten seconds after the submit button was pressed.

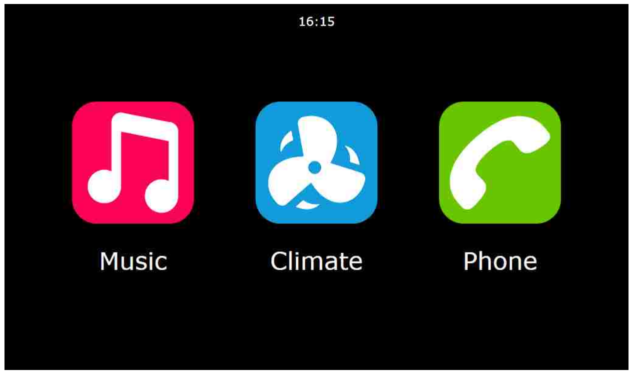

Figure 3.3: IVIS App Home Screen

The music screen was used for the radio and playlist tasks, see Figure 3.4. For each new task, the application would randomly generate a set of 15 songs from a database and assigned it to the playlist menu (in a random order). The music database contained 260 different songs, which were generated using the billboard top charts across various years and genres. The songs were selected to be representative of different genres and eras as to not give any participant group an advantage in song recognition over another. For the radio tasks, participants were instructed to tune to either an FM or AM station, and adjust the volume to a specific value. For the playlist tasks, participants were provided the song and artist name for a specific song.

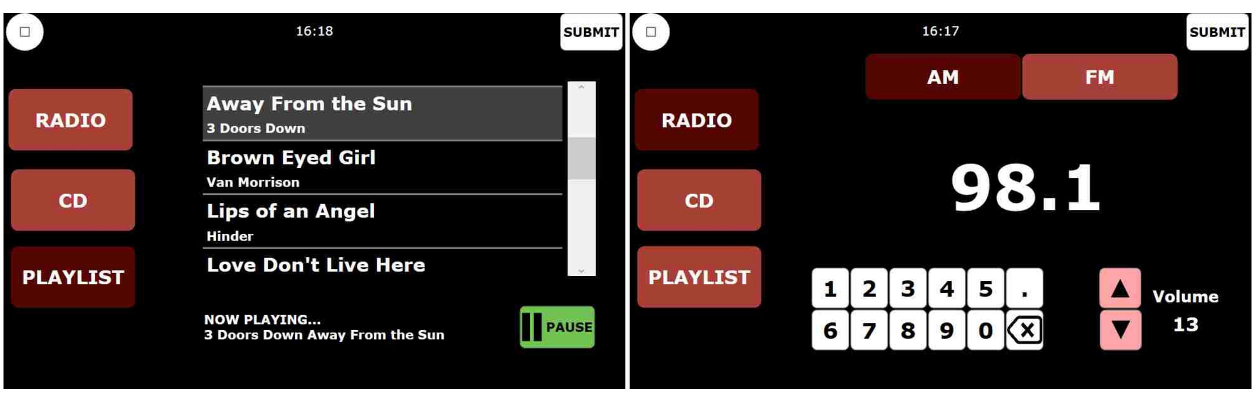

Figure 3.4: IVIS App Playlist *(left)* and Radio *(right)* Screens

Similarly to the playlist menu, the default state for the settings on the climate screen were randomly generated for each new task (see Figure 3.5). Logic was programmed into the application such that the default setting never included the state instructed by the task instructions (e.g., the same temperature, fan direction, etc.). For each climate task, participants were given three climate functions to adjust, including any three of the following settings: cabin air temperature, fan direction, fan level, air conditioning, front defrost, and rear defrost.

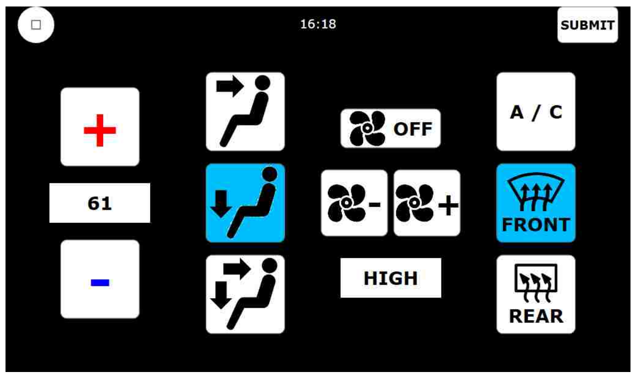

Figure 3.5: IVIS App Climate Screen

The phone screen was used for the contacts and dial tasks (see Figure 3.6). For each new task, the list of contacts in the phonebook was randomly generated by selecting a set of 15 names from the phone database. The database contained a list of 180 randomly generated names, with equal male and female names. The names in the contacts list were in a random order, rather than alphabetically, in order to make the task similar in structure to the playlist task. The dial task instructed the participant to dial a phone number, where the instruction was an alpha phrase (e.g., 1-400-Red-Sock) and the participant had to dial the number using the numeric pad (e.g., 1-400-733-7625). The use of an alpha phrase as the phone number was selected because it yielded a reasonable amount of information for the participant to remember, rather than a random string of 11 digits. Each alpha phrase was specifically selected to reflect words that could be easy to understand from the audio instruction and spell at a 7th grade reading level.

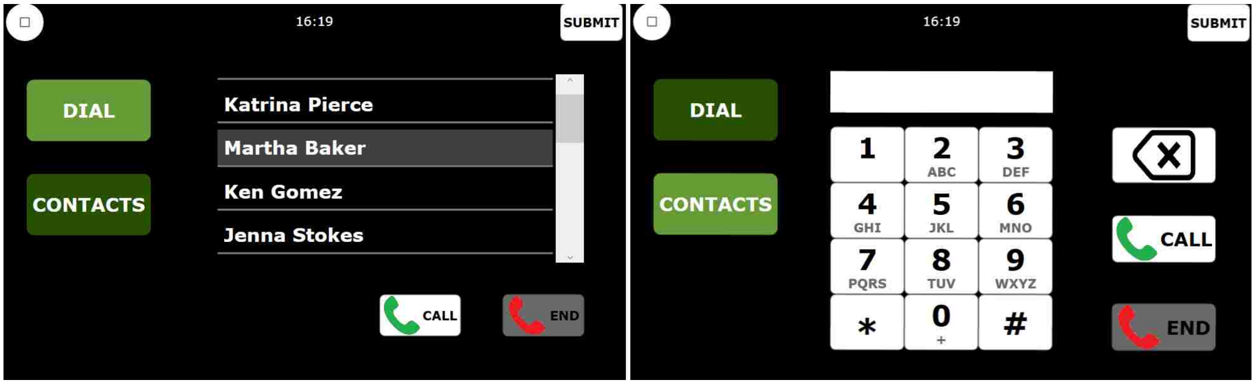

Figure 3.6: IVIS App Contact *(left)* and Dial *(right)* Screens

3.2.3 Tactile Detection Response Task, TDRT

A Tactile Detection Response Task (TDRT) was used to measure cognitive workload during the drives. The TDRT was implemented and analyzed per the protocol defined in the ISO 17488:2016 standard (ISO, 2016). A small motor, which was taped to the participant's neck, randomly vibrated once every 3-5 seconds. Participants responded to the vibration by pressing a small button as soon as they felt the vibration. This button was strapped to their left index finger, see Figure 3.7. More specifically, cognitive workload was evaluated using response time (i.e., time between vibration to button press) and miss rate (i.e., number of stimulus misses divided by total number of stimulus events).

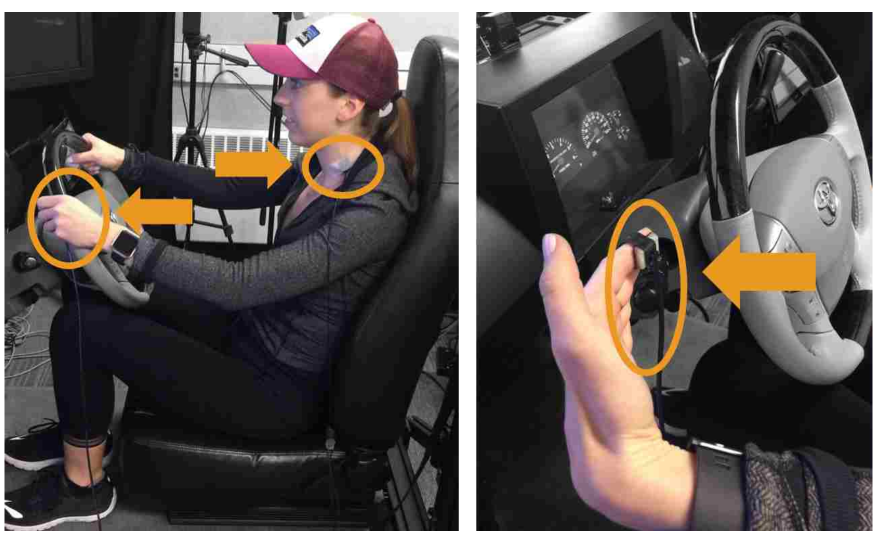

Figure 3.7: TDRT Motor and Button Configuration on Participant

3.2.4 Video Camera

A GoPro Hero4 was mounted on the driving simulator above the IVIS touchscreen. The camera was focused on the participant's face for all drives, in order to capture eye glance patterns to and away from the driving scene. The camera was set to collect video data at 30 frames per second at 720p. The Morae Manager software (version 3.3.4) was used to identify eyes on and off road glances from the video data.

3.2.5 Trust Questionnaire

A questionnaire was developed to evaluate trust in the lane keeping system. This questionnaire was administered at the end of each of the three days for the treatment participants. The trust questionnaire was a series of eight 10–point Likert scale questions regarding various dimensions of trust, where 1 indicated very low trust and 10 was very high trust. The survey was adapted from a survey previously used to evaluate trust in ACC (Moeckli et al., 2015). The questions were developed and refined based on previous research evaluating key aspects of human-computer trust, such as capturing reliability, confidence, faith, predictability, and dependability (Jian et al., 2000). The eight trust questions were as follows:

1. To what extent does the lane keeping system perform the task it was designed to do?
2. To what extent can the lane keeping system's behavior be predicted from moment to moment?
3. To what extent does the lane keeping system respond similarly to similar circumstances at different points in time?
4. What is your degree of faith that the lane keeping system will be able to cope with future driving situations?
5. What is your degree of trust in the lane keeping system to respond accurately?
6. What is your degree of self-confidence to manually intervene with the lane keeping system?
7. What is your overall degree of trust in the lane keeping system?

8. How confident do you feel about your previous trust ratings?

3.2.6 NASA TLX Questionnaire

The NASA Task Load Index (TLX) was used to measure perceived workload for each of the five IVIS task types (i.e., contact, playlist, radio, climate, dial). This questionnaire was originally developed to assess subjective workload by the Human Performance Group at NASA (Hart, 1986). Participants were asked to fill out the NASA TLX survey immediately after completing the last drive on each of the three days. This assessment measures workload based on six scales: mental demand, physical demand, temporal demand, performance, effort, and frustration. The questions were answered based on a 20–point Likert scale. A weighted average for the workload was computed by asking each participant to rank the relative importance for the 15 pairwise contrasts of the scales. An overall workload score was computed for each participant on each day for each of the five task types; the score could range from 0 to 100, with higher values indicating higher workload.

3.3 Procedure

Participant recruitment was conducted using Craigslist and through student outreach at the University of Washington. All study procedures took place at the Human Factors and Statistical Modeling Lab in the Engineering Library at the University of Washington - Seattle campus. Upon arrival to the lab, the study procedures were explained to the participant and written informed consent was obtained.

Participants completed a total of eight drives: three on the first day, two on the second day, and three on the third day. All three study days were completed within a seven-day period. There was also a practice drive at the beginning of each day. The eight study drives were approximately 20 minutes each and the practice drives were 10 minutes each. There was a ten minute break between each drive. After participants completed the study drives each day, they were asked to fill out a series of surveys relating to demographics, driving history, task difficulty, trust, and simulator realism.

Drive 1 provided baseline measures for driving performance for each participant. During the second through seventh drives, all of the participants engaged in visual-manual distracting tasks on the IVIS application. Drive 2 provided baseline measures with a secondary task. In drives three through six, participants from the treatment group were exposed to the lane keeping system. Drive 7 was used to quantify changes in behavior after exposure and withdrawal of the lane keeping system. Drive 8 was used to further measure any carryover effects on baseline driving. Fully manual driving (i.e., no automation) was required in drives 1, 2, 7, and 8 for both groups. See Table 3.3 for this sequence by group.

Table 3.3: Drive Order

Automation	**Vehicle**	**Drive Number**							
Group	**System**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Control Treatment	IVIS	No	Yes	Yes	Yes	Yes	Yes	Yes	No
Control Treatment	LK	No	No	No Yes	No Yes	No Yes	No Yes	No	No

Chapter 4
ANALYTICAL METHODS

As discussed in the previous chapter, this study used a longitudinal experimental design to evaluate behavioral adaptations to the lane keeping system. Longitudinal data often has inherent correlation within its structure, which occurs due to multiple observations (i.e., repeated measures) across time for each participant. These repeated measures lead to correlated observations and as such violates the assumption of independence required for traditional regression methods (e.g., Ordinary Least Squares, OLS). Using OLS with longitudinal data often results in inaccurate standard error estimates (Hubbard et al., 2010). To account for this covariance structure, this dissertation utilized generalized linear mixed models.

4.1 Independent Variables

This study utilized both between-subject and within-subject measures to evaluate behavioral adaptations across various driving demands and driver populations.

4.1.1 Age Group (between-subject): 3 levels

Previous literature suggests that there are differences across age groups in regards to driver behavior (Korber et al., 2016; Xiong & Boyle, 2012) and trust (Sanchez et al., 2004) with autonomous vehicle systems. Differences in crash risk associated with secondary task engagement have also been correlated to age (Guo et al., 2017). Therefore three age groups were included in the study: *(1)* younger (25-34 years old), *(2)* middle (35-44 years old), and *(3)* older (45-54 years old). There were 16 participants in each age group.

4.1.2 Gender (between-subject): 2 levels

Similarly, previous studies have shown that there are effects of gender on trust in automation (Hoff & Bashir, 2015) and attitudes towards automation (Payre et al., 2014). Therefore gender, *(1)* male and *(2)* female, were balanced within each age group.

4.1.3 Automation Group (between-subject): 2 levels

Participants were randomly assigned to the *(1)* treatment group or *(2)* control group. Both groups were exposed to the same study procedures, except the treatment group had the addition of the lane keeping system engaged in drives 3-6. The control group was used to account for time on task effects.

4.1.4 Drive Number (within-subject): 8 levels

Each participant completed 8 drives: *(1-2)* the first two were before automation (i.e., manual driving) and were used to collect baseline measures; *(3-6)* the middle four were exposure to automation (for the treatment group) and used to collect intermediate effects of exposure; and *(7-8)* the last two were withdrawal of automation (i.e., manual driving) and captured carryover effects of exposure.

4.1.5 Road Type (within-subject): 3 levels

There were three different geometric alignments implemented in the driving scenario. These were used to assess the driving skill aided by the semi-autonomous system (i.e., lateral control). The alignments were *(1)* straight road, *(2)* curved road (i.e., horizontal alignment), and *(3)* curved hill road (i.e., horizontal and vertical alignment).

4.1.6 IVIS Task Difficulty (within-subject): 4 levels

There were four levels of task difficulty for the secondary tasks on the IVIS: *(1)* easy (contact and playlist), *(2)* medium (radio), *(3)* hard (climate), and *(4)* hardest (dial). The tasks were *a priori* designed to yield different levels of workload and were further calibrated based on pilot testing.

4.2 Dependent Variables

The dependent variables were collected using the IVIS application (secondary task engagement), TDRT (cognitive workload), driving simulator (driving performance), video camera (eyes-off-road behavior), and surveys (trust in automation). The driving performance measures were aggregated at eleven different intervals across each drive, corresponding to the geometric alignment. Each of the 11 segments were approximately 1.25 miles long and included four straight, five curved (2 right, 3 left), and two hill (each with an uphill and downhill) segments. The first and last mile of the drives were excluded from analysis in order to remove accelerating from and decelerating to a parked position. The cognitive workload and eye glance measures were aggregated for each IVIS task within the drive in order to account for each IVIS task difficulty.

4.2.1 IVIS Task Completion Count

The total number of tasks completed per participant for each drive were collected in order to capture risk perception. Participants were instructed to complete the tasks at their own pace, thus the number of tasks completed was used as a measure of riskiness.

4.2.2 IVIS Task Accuracy

A binary indicator of whether the task was completed correctly or not was used to gauge performance of secondary tasks over time and influence of automation.

4.2.3 TDRT Response Time

Response time was measured, in milliseconds (ms), from the onset of the tactile stimulus until the participant pressed the button with their finger. In all analysis involving response time, only correct responses (i.e., non-misses) were included. A larger response time is associated with an increase in cognitive workload (ISO, 2016; Victor et al., 2008).

4.2.4 TDRT Miss Rate

A miss was classified as a response slower than 2500 ms (i.e., unrequested response), quicker than 100 ms (i.e., premature response), or more than one button press for a single stimulus (i.e., repeated response). The miss rate was computed as the total number of misses divided by the total number of tactile stimulus events over the respective time segment. A larger miss rate is associated with an increase in cognitive workload (ISO, 2016; Victor et al., 2008). Miss rate was used rather than hit rate in order to reduce dissonance between the two TDRT metrics, so that an increase in both TDRT measures correlated to an increase in cognitive workload.

4.2.5 Standard Deviation of Lateral Position, SDLP

Standard Deviation of Lateral Position (SDLP) was used as a driving performance measure because it directly related to the driving skill aided by the semi-autonomous intervention (i.e., lane keeping). Lateral position of the vehicle was measured as the difference between the center of the vehicle and the center of the lane (in inches). SDLP was then computed using the equation defined in SAE J2944_201506 (2015), see Equation 4.1.

$$SDLP = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2} \tag{4.1}$$

where,

$x_i = i^{th}$ observation of lane position,

$\bar{x}$ = mean lane position of the segment,

N = number of data points in the segment.

4.2.6 Time to Collision, TTC

Time to Collision (TTC), in seconds, measured the time required for the participant's vehicle to strike the lead vehicle, factoring in speed, acceleration, and distance. A cutoff value of 20 seconds was used for TTC, as supported by the SAE operational definitions (SAE, 2015). A cutoff value is often adopted in the literature as TTC can approach infinity when the two objects are traveling at the same speed. TTC was selected as a variable of interest because it was not a driving maneuver aided by the semi-autonomous intervention.

4.2.7 Eyes-Off-Road Behavior

Eyes-off-road was characterized as any eye glance to the IVIS touchscreen, in order to quantify willingness to engage in a distracting task. Participants were instructed to complete IVIS tasks at their own pace and the order was randomly generated for each drive. As a result, each participant completed a different number and combination of tasks. In order to make comparisons across participants, a set of 15 tasks per drive were extracted for analysis. Fifteen tasks were utilized as it provided three trials (repetitions) for each task type and thus an average of task performance could be computed. The average of task performance, rather than one single trial of a task, was preferred as it would reduce the impact of randomly selecting a task with outlier performance. For each drive, there were 3 contact, 3 playlist, 3 radio, 3 climate, and 3 dial tasks randomly selected.

There were four metrics of eye glance behavior analyzed in relation to eyes-off-road (EOR) glances to the IVIS: *(1)* mean glance duration per task, *(2)* 90$^{\text{th}}$ percentile glance duration per task, *(3)* proportion of glance durations that exceeded 2.0 seconds (i.e., long glances) per task, and *(4)* total eyes-off-road time per task.

The eyes-off-road measures were aggregated at the task level for each participant for each drive. Since each participant completed 3 trials of each task type, the four eye glance metrics were first computed at the trial level and then averaged across the three trials to get a task level average per participant. The equations are provided at the trial level for mean glance duration (Equation 4.2), percent long glances (Equation 4.3), and total eyes-off-road (Equation 4.4). The 90$^{\text{th}}$ percentile glance duration was computed as the 90$^{\text{th}}$ quantile glance duration (i.e., 90$^{\text{th}}$ longest glance), aggregated at the task level for each participant on each day.

$$Mean\ GD_i = \frac{\sum_{j=1}^{N_i} EOR_{ij}}{N_i} \tag{4.2}$$

$$Percent\ Long\ GD_i = \frac{N_i > 2.0\ seconds}{N_i} \times 100 \tag{4.3}$$

$$Total\ EOR_i = \sum_{j=1}^{N_i} EOR_{ij} \tag{4.4}$$

where,

i = trial, where $i = 1, 2, 3,$

$EOR_{ij} = j^{th}$ EOR duration in trial i,

N_i = total number of EOR glances in i^{th} trial.

Previous literature has used various combinations of mean glance duration, maximum glance duration, proportion of long glances, and total eyes-off-road time in evaluating safety (Peng & Boyle, 2015; Boyle et al., 2013; Korber et al., 2018). This current study used 90$^{\text{th}}$ percentile glance duration as opposed to maximum glance duration for robustness. The 2.0 second threshold for long glances was selected based on the NHTSA Distraction Guidelines (NHTSA, 2013) and previous research that indicates an increased crash risk for glances off road longer than 2 seconds (Klauer et al., 2006).

4.2.8 Trust in Automation

Trust in automation was a subjective measurement collected using the 10–point Likert scale trust questionnaire, which was a series of eight questions administered to the treatment participants at the end of each study day.

4.3 Statistical Models

All data reduction and analysis was conducted using the R statistical software program (version 3.4.0). Statistical significance was assessed at $\alpha = 0.05$.

4.3.1 Generalized Linear Mixed Model, GLMM

Theory

Generalized Linear Mixed Models (GLMMs) are an extension of both linear models and mixed models. These models can account for a broader range of distributions, unlike the linear model which is specific to the Gaussian distribution. The mixed model element of the model class incorporates both fixed effects and random effects; fixed effects are constant across individuals, while random effects can vary across individuals. Mixed models can therefore account for hierarchical data structures, such as repeated observations across individuals over time.

Generalized Linear Mixed Models utilize a link function, $g(\mu_i)$, which describes the relation between the expected value, $E(y_i) = \mu_i$, and the linear predictor, η_i, such that $\eta_i = g(\mu_i) = {X_i}^t\beta$. The link function can be applied to the random effects model introduced by Laird and Ware (1982), which is expressed in Equation 4.5.

$$y_i = X_i\beta + Z_i\gamma_i + \epsilon_i \tag{4.5}$$

where,

y_i = vector of size $N_i \times 1$ of the outcome variable for the i$^{\text{th}}$ participant,
X_i = design matrix for the fixed effects of size $N_i \times p$,
β = vector of size $p \times 1$ of the fixed effects coefficient estimates,
Z_i = design matrix for the random effects of size $N_i \times q$,
γ_i = vector of size $q \times 1$ of the random effects estimates,
ϵ_i = vector of size $N_i \times 1$ of the residuals, where $\epsilon_i \sim \mathcal{N}(0, R_i)$ and $R_i = \sigma^2 I_{n_i}$.

In this dissertation, only random intercepts are fit as the random effect and random slopes are not implemented. Thus, the GLMMs used in this dissertation are expressed in Equation 4.6 for the Gaussian model (i.e., continuous performance measures) and Equation 4.7 for the negative binomial model (i.e., count data).

$$\mu_i = \beta_0 + \gamma_{0i} + \beta X_i \tag{4.6}$$

$$log\Big(\frac{\mu_i}{\mu_i + \theta}\Big) = \beta_0 + \gamma_{0i} + \beta X_i \tag{4.7}$$

where,

β_0 = fixed intercept,

γ_{0i} = random intercept for i^{th} participant,

β = vector of fixed effects,

X_i = design matrix of fixed effects.

Example

GLMMs are used in this study to account for the repeated observations of individual participants over time (i.e., within and across drives). For example, consider the measured response times to the TDRT across participants; it is likely that participants have inherently different reaction times and a random intercept model can accommodate this. This is demonstrated in Figure 4.1 using a subset of data from this study. TDRT response times for one random participant across a 1.5 mile segment were plotted (top left). Then data for three additional random participants (i.e., four total) were plotted (top right). A linear model, using OLS, was fit to the data of all four participants, ignoring the violated independence assumption (bottom left). It is apparent from the plot that this linear model poorly fits the data, in fact the model has an adjusted $R^2 = 0.105$. Instead, a linear mixed model was fit to the data (bottom right) allowing individual intercepts for each participant. As it can be seen in this plot, the linear mixed model fits the data much better, allowing for meaningful analysis of the data. While this example uses actual data from this study, the data is plotted in its raw time series form. The data in this study was aggregated into bins in order to evaluate trends specifically pertaining to the independent variables discussed above, for example into 1.25 mile increments for driving measures and by each secondary task for cognitive workload and eye glance measures. As such, the serial correlation of the data in its time series form was removed.

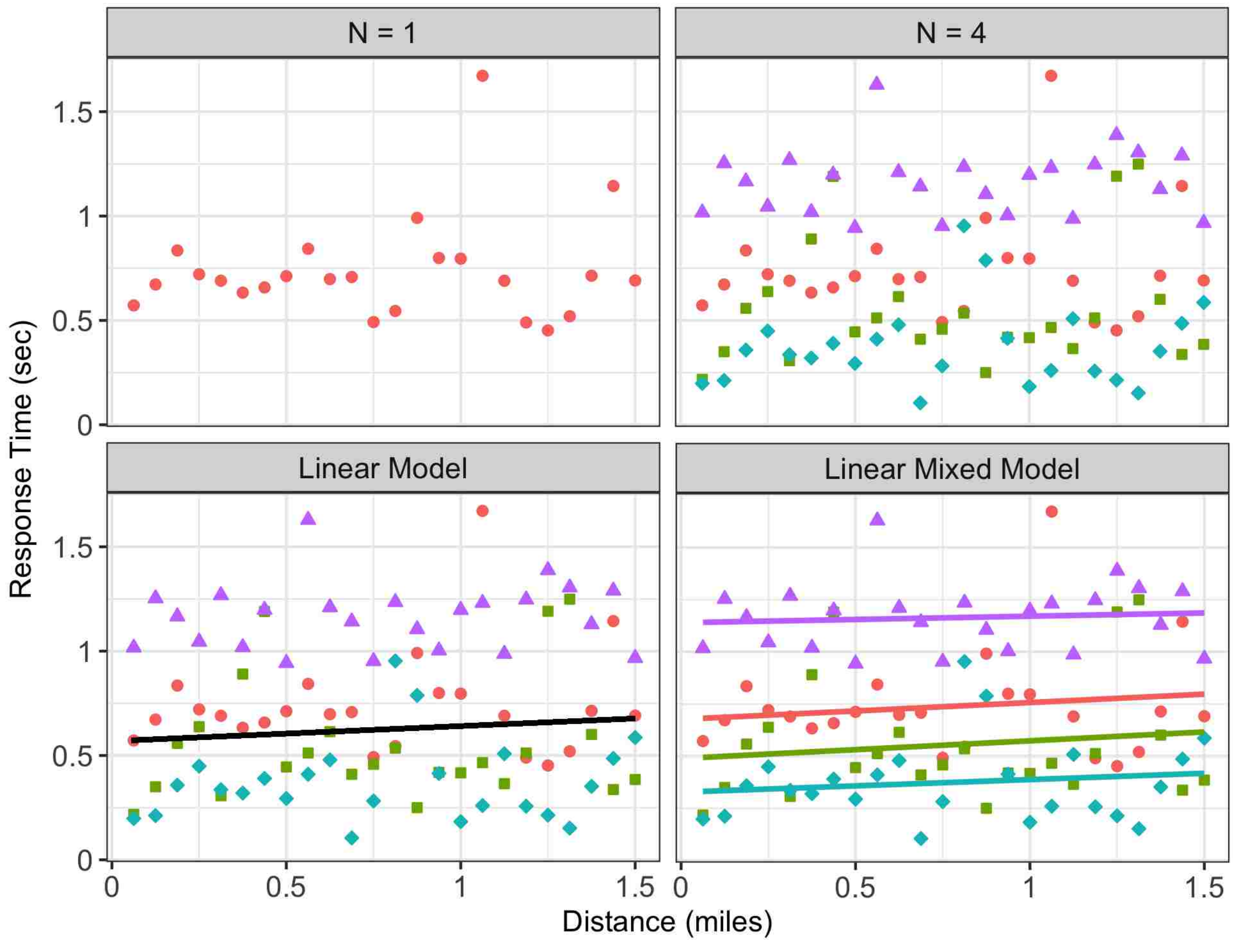

Figure 4.1: Subset of TDRT Data for Example with One Participant *(top left)*, Four Participants *(top right)*, Linear Regression Model Fit *(bottom left)*, and Linear Mixed Model Fit *(bottom right)*

Application

GLMMs were used to quantify behavioral adaptations in secondary task engagement, cognitive workload, and driving performance. Mixed effects linear regression (i.e., $g(\cdot) = \mu$) was used to model the outcome variables for IVIS task count, TDRT response time, SDLP, and TTC. Mixed effects negative binomial regression (i.e., $g(\cdot) = log\left(\frac{\mu}{\mu+\theta}\right)$) was used to model TDRT miss rate (i.e., miss count with exposure). These general equations can be expressed as:

$$\begin{aligned} TaskCount_{ij} = \beta_0 + \gamma_{0i} + \beta_{1i}Treatment + \beta_{2j}DriveNumber + \\ \beta_{3ij}Treatment \times DriveNumber + \\ \beta_{4i}AgeGroup + \beta_{5i}Gender + \epsilon_{ij} \end{aligned} \tag{4.8}$$

$$\begin{aligned} CognitiveWorkload_{ij} = \beta_0 + \gamma_{0i} + \beta_{1i}Treatment + \beta_{2j}DriveNumber + \\ \beta_{3ij}Treatment \times DriveNumber + \beta_{4j}TaskType + \\ \beta_{5i}BaselineDRT + \beta_{6i}AgeGroup + \beta_{7i}Gender + \epsilon_{ij} \end{aligned} \tag{4.9}$$

$$\begin{aligned} DrivingPerformance_{ij} = \beta_0 + \gamma_{0i} + \beta_{1i}Treatment + \beta_{2j}DriveNumber + \\ \beta_{3ij}Treatment \times DriveNumber + \beta_{4j}RoadType + \\ \beta_{5i}AgeGroup + \beta_{6i}Gender + \epsilon_{ij} \end{aligned} \tag{4.10}$$

where,

i = participant i, where i is 1, ..., 48,

j = drive number j, where j is 1, ..., 8 for Equation 4.10 and j is 2, ..., 7 for Equations 4.8 and 4.9,

$RoadType$ = curved or curved hill (reference straight),

$TaskType$ = playlist, radio, climate, or dial (reference contact),

$AgeGroup$ = middle or older (reference younger),

$BaselineDRT$ = TDRT response time or miss rate measured in drive 1,

γ_{0i} = random intercept for the i^{th} participant,

ϵ_{ij} = residual, Gaussian distributed for identity link and Gamma distributed for negative binomial link.

GLMM versus GEE

There are two common approaches for modeling longitudinal data: subject-specific (GLMM) and population-averaged (Generalized Estimating Equation, GEE) models. The main difference between these two methods are in the interpretation of the regression coefficients; whether the inference is associated with individual or population averaged outcomes. The population-averaged approach assumes heterogeneity of the population, whereas subject-specific allows for random effects across individuals (Zeger et al., 1988). GEE models are commonly applied in the health and medical sciences (Fitzmaurice et al., 2011), where cohorts are grouped based on shared characteristics (e.g., similar risk factors or patients of specific clinics or doctors) (Hardin & Hilbe, 2012).

These two statistical methods are also referred to as conditional models (GLMM) and marginal models (GEE). The interpretation of a GLMM and GEE coefficient both describe the mean change in response for one unit change in the covariate, holding all else equal across the other covariates. However, the GLMM is conditional on the random effects, thus the inference is specific to the individual. Whereas GEE describes the marginal response across all groups (i.e., mean change regardless of holding the individual random effects constant) (Y. Lee & Nelder, 2004). Previous literature recommends that conditional models (GLMM) be applied for inference focusing on within-subject differences (e.g., before, during, after an intervention) and GEE be applied for between-subject differences (e.g., effects of a condition or state) (Murray et al., 2004).

GLMMs were applied to the data in this dissertation instead of GEEs for two reasons. The first is that the objective of this dissertation was to evaluate the effects of exposure to a lane keeping system, thus relating to the within-subject changes due to an intervention. The second reason that GLMM was assumed to be more appropriate rather than GEE was to account for individual differences of drivers; it is likely that individual drivers exhibit different driving behaviors, have unique eye glance patterns, and experience different cognitive workload associated with the study tasks. Thus GLMM was applied to account for these inherent subject specific differences and within-subject specific behavioral adaptations.

4.3.2 Analysis of Variance, ANOVA

Repeated measures Analysis of Variance (ANOVA) was used to understand differences in means across various categorical independent variables for several of the continuous response variables. Repeated measures ANOVA is similar to the mixed models regression discussed previously, in that repeated measures ANOVA utilizes within-subject blocking to account for the correlated nature of longitudinal data. This method is different from GLMM, as the dependent variable in ANOVA must be continuous and follow a Gaussian distribution. Additionally, ANOVA utilizes the F statistic, which examines the variations within groups. While ANOVA tests for the difference in means between groups, it is not restricted to comparing two conditions like with the t-test. Therefore, the ANOVA was used in this dissertation to evaluate the effects of various factors on performance and behavior, while GLMM was used to quantify these effect sizes. For example, the output from the ANOVA provides insight across all groups (e.g, is there a difference between younger, middle, and older age groups), while the effects of a factor in the GLMM are in reference to another group (e.g., older versus younger or middle versus younger). Tukey's Honest Significant Difference (HSD) test was used as a post hoc test for the ANOVA, in order to evaluate the differences of specific contrasts (e.g., younger-middle, younger-older, middle-older age groups).

More specifically, the repeated measures ANOVAs used in this dissertation applied the blocking on the participant level (equivalent to the random intercept used in the GLMMs). The ANOVAs were used to evaluate behavioral adaptations in eye glance behavior between the two automation groups, where the response variables were mean GD, 90$^{\text{th}}$ percentile GD, percent long glances, and total EOR time. ANOVAs were also used to evaluate differences in behavioral adaptations based on trust, where the response variables were driving performance (SDLP), cognitive workload (DRT response time), and eye glance behavior (mean GD, 90$^{\text{th}}$ percentile GD, percent long glances, and total EOR time). The model equations are provided for driving performance in Equation 4.11, cognitive workload in Equation 4.12, and eye glance behavior in Equation 4.13. Similar model structures were used in the ANOVAs focusing on automation group (Chapter 6) and trust group (Chapter 7), where automation group (2 levels) was used in place of trust group (3 levels). The following equations demonstrate the ANOVA structures used for the analysis on trust (i.e., Chapter 7).

$$\begin{aligned} Driving_{ijklmn} = \mu + (Trust \times Drive)_{ij(k)} \; + Trust_i + Drive_{j(k)} + \\ RoadType_{l(k)} + Gender_m + AgeGroup_n + \epsilon_{ijklmn} \end{aligned} \tag{4.11}$$

$$\begin{aligned} Cognitive_{iokpmn} = \mu + (Trust \times Drive)_{io(k)} + Trust_i + Drive_{o(k)} + \\ TaskType_{p(k)} + Gender_m + AgeGroup_n + \epsilon_{iokpmn} \end{aligned} \tag{4.12}$$

$$\begin{aligned} EyeGlance_{iqkpmn} = \mu + (Trust \times Drive)_{iq(k)} + Trust_i + Drive_{q(k)} + \\ TaskType_{p(k)} + Gender_m + AgeGroup_n + \epsilon_{iqkpmn} \end{aligned} \tag{4.13}$$

where,

$Driving_{ijklmn}$ = observed driving performance on the k^{th} participant for the i^{th} trust group for the j^{th} drive for the l^{th} road type for the m^{th} gender for the n^{th} age group,
$Cognitive_{iokpmn}$ = observed cognitive measure on the k^{th} participant for the i^{th} trust group for the o^{th} drive for the p^{th} task type for the m^{th} gender for the n^{th} age group,
$EyeGlance_{iqkpmn}$ = observed eye behavior on the k^{th} participant for the i^{th} trust group for the q^{th} drive for the p^{th} task type for the m^{th} gender for the n^{th} age group,
μ = overall population mean,
$Trust_i$ = effect of the i^{th} trust group,
$Drive_j(k)$ = effect of the j^{th} drive number within the k^{th} participant (or o^{th} or q^{th} drive number),
$(Trust \times Drive)_{ij(k)}$ = effect of the i^{th} trust group for the j^{th} drive within the k^{th} participant,
$RoadType_{l(k)}$ = effect of the l^{th} road type within the k^{th} participant,
$Gender_m$ = effect of the m^{th} gender,
$AgeGroup_n$ = effect of the n^{th} age group,
$\epsilon_{ijklmn} = \epsilon_{iokpmn} = \epsilon_{iqkpmn}$ = individual deviation,
$i = 1, 2, 3$; $j = 1, ..., 8$; $k = 1, ..., 48$; $l = 1, 2, 3$; $m = 1, 2$; $n = 1, 2$; $o = 2, ..., 7$; $p = 1, ..., 5$; $q = 1, 2$; $N_{driving} = 1152$; $N_{cognitive} = 1436$; $N_{eye} = 480$.

4.3.3 Cluster Analysis

A cluster analysis was used to group participants based on their trust in automation across the three study days. The analysis was conducted on the responses to the questionnaire relating to trust in the lane keeping system. The questionnaire had eight questions and was administered at the end of each of the three days on the treatment group, resulting in 24 observations per each of the 30 treatment participants. The resulting matrix used in the cluster analysis was 30 rows (one per participant) by 24 columns (each trust question, Q_1 to Q_8, across the three days), see Equation 4.14. The cluster analysis was performed on the standardized values (i.e., z-scores, with mean 0 and SD 1) for the 10–point Likert scale trust questions.

$$\begin{array}{cc} Ss & \begin{array}{ccccccc} Q_{1Day1} & Q_{2Day1} & Q_{3Day1} & \cdots & Q_{1Day2} & \cdots & Q_{8Day3} \end{array} \\ \begin{array}{c} 1 \\ \vdots \\ 30 \end{array} & \begin{pmatrix} x_{1,1,1} & \cdots & \cdots & \cdots & \cdots & \cdots & x_{1,8,3} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{30,1,1} & \cdots & \cdots & \cdots & \cdots & \cdots & x_{30,8,3} \end{pmatrix} \end{array} \tag{4.14}$$

Hierarchical clustering with Ward's minimum variance method was used (Ward, 1963). This agglomerative method (i.e., starting with each observation as its own cluster) builds clusters based on minimizing the total within-group error sum of squares across observations for each iteration (Contreras & Murtagh, 2016). The distance matrix was computed using the Euclidean distance, see Equation 4.15.

$$d(i,j) = \sqrt{(X_i - X_j)^2 + (Y_i - Y_j)^2} \tag{4.15}$$

The within-cluster sum of squares calculates the variability of observations within a cluster. Such that a smaller within-cluster sum of squares value suggests a tighter cluster of observations. The definition for within-cluster sum of squares is provided in Equation 4.16.

$$SS_{within\ cluster} = \sum_{c=1}^{C} \sum_{i=1}^{n_c} (x_i - \bar{v}_c)^2 \tag{4.16}$$

where,

C = number of clusters,

n_c = number of observations in the c^{th} cluster,

$\bar{v}_c$ = mean of the c^{th} cluster.

The number of clusters was determined using the pseudo F statistic, which measures the ratio of the between-cluster variation to the within-cluster variation. The between-cluster variation provides a measure of how close the clusters are from each other, see Equation 4.17, which was adapted from (Yang et al., 2015).

$$SS_{between\ cluster} = \sum_{c=1}^{C} \sum_{i=1}^{n_c} n_c (v^c - \bar{v})^2 \tag{4.17}$$

where,

$v^c = \frac{1}{n_c}(x_1^c + x_2^c + ... + x_n^c)$,
$\bar{v} = \frac{1}{n}(x_1 + x_2 + ... + x_n)$.

The pseudo F statistic is shown in Equation 4.18, where a larger pseudo F value is generally optimal for clustering.

$$pseudo\ F = \frac{(n-c) \times \sum_{c=1}^{C} \sum_{i=1}^{n_c} n_c (v^c - \bar{v})^2}{(c-1) \times \sum_{c=1}^{C} \sum_{i=1}^{n_c} (x_i - \bar{v}_c)^2} = \frac{n-c}{c-1} \times \frac{SS_{between\ cluster}}{SS_{within\ cluster}} \tag{4.18}$$

where,

n = number of observations,

c = number of clusters.

Previous literature in the driving domain has used Ward's minimum variance technique for clustering driving behaviors and driver perceptions (Donmez et al., 2010; Xiong et al., 2012; Peng & Boyle, 2015).

Chapter 5

PARTICIPANT AND TASK SUMMARY STATISTICS

This chapter presents summary statistics on the participants included in the study. Descriptive statistics on study participants were evaluated in order to understand the generalizability of the cohort to the general driving population. This was also important in understanding any differences between groups (e.g., automation groups, age groups). The descriptive statistics on the IVIS tasks are also provided in this chapter, as it was considered important to understand differences in performance and workload across the different task types. Specifically, this provided validation in the *a priori* determined task difficulty levels and rationale for including task type as an independent variable in the analytical models.

5.1 Participant Demographics

There were a total of 48 participants included in the analysis, however there were a total of 52 participants involved in this study (i.e., four participants did not finish entire study procedures). One participant was scheduled, but never showed up for any of their sessions. Two participants experienced simulator sickness during their practice drive on the first day and did not complete any study procedures as a result. One participant completed the first two study days, but did not show up for their third session; this participant was not able to reschedule their third session within the seven day window needed to complete all three of their sessions.

There were 30 participants in the treatment group, ranging in age from 26 to 54 years old with a mean age of 38.7 years old (SD 9.6 years). The control group was comprised of 18 participants, who also ranged in age from 26 to 54 years old with a mean age of 38.3 years old (SD 9.1 years). The breakdown of demographics by age group (i.e., younger, middle, older) for each automation group is provided in Table 5.1.

Table 5.1: Participant Demographics

Group	Age *(years)* Mean ^{SD}	Range	Age at License	Education *n(%)* H.S.	Assoc.	Bach.	Grad.
Control							
Younger	27.7 $^{1.37}$	26-29	16.2 $^{1.17}$	1 (16.7)	0 (0.0)	2 (33.3)	3 (50.0)
Middle	38.8 $^{3.19}$	35-42	19.0 $^{3.95}$	0 (0.0)	2 (33.3)	3 (50.0)	1 (16.7)
Older	49.5 $^{4.04}$	45-54	16.5 $^{0.84}$	2 (33.3)	1 (16.7)	2 (33.3)	1 (16.7)
All	38.7 $^{9.62}$	26-54	17.2 $^{2.62}$	3 (16.7)	3 (16.7)	7 (38.9)	5 (27.7)
Treatment							
Younger	28.3 $^{2.71}$	26-33	17.5 $^{1.51}$	2 (20.0)	1 (10.0)	3 (30.0)	4 (40.0)
Middle	36.7 $^{2.24}$	35-42	17.2 $^{2.05}$	1 (10.0)	2 (20.0)	4 (40.0)	3 (30.0)
Older	48.7 $^{3.10}$	45-54	18.4 $^{4.93}$	1 (10.0)	2 (20.0)	5 (50.0)	2 (20.0)
All	38.3 $^{9.14}$	26-54	17.8 $^{3.25}$	4 (13.3)	5 (16.7)	12 (40.0)	9 (30.0)

Participants in both automation groups had similarities in average weekly driving frequencies, where the majority of the participants drove on average seven days a week (66.7% in the control group and 70.0% in the treatment group). However, two of the 30 participants in the treatment group (6.7%) drove on average less than one day a week, while none of the participants in the control group reported having this low of an average weekly driving frequency. The treatment group also had a slightly higher proportion of drivers with at least one moving violation within the past 5 years (43.3% for treatment versus 22.2% for control), but both groups had very similar 5 year crash histories, see Table 5.2.

Table 5.2: Participant Driving History

Group	Avg Weekly Driving Frequency, days *n(%)*			Moving Violations 5 years *n(%)*		Crashes in 5 years *n(%)*	
	7	6-1	<1	0	1+	0	1+
Control	12 (66.7)	6 (33.3)	0 (0)	14 (77.8)	4 (22.2)	13 (72.2)	5 (27.8)
Treatment	21 (70.0)	7 (23.3)	2 (6.7)	17 (56.7)	13 (43.3)	23 (76.7)	7 (23.3)

Previous driving experience with Advanced Driver Assistance Systems (ADAS) was similar across groups, where 22.2% of control group participants and 23.3% of treatment group participants had previously driven a vehicle with lateral assist; 22.2% of control and

26.7% of treatment group participants had previous experience with a forward assistance system; 38.9% of control group participants and 23.3% of treatment group participants had prior experience with a blind spot alert system (see Table 5.3). Participants were told that lateral assistance included either a lane keeping system or lane departure warning system (i.e., mitigation or alert system). Similarly, participants were informed that a forward assistance system could include a forward collision mitigation or warning system.

Table 5.3: Participant Experience with Advanced Driver Assistance Systems

Group	**Ever Driven** ***n(%)***			**Ever Owned** ***n(%)***		
	Yes	**No**	**Not Sure**	**Yes**	**No**	**Not Sure**
Lateral Assistance						
Control	4 (22.2)	14 (77.8)	0 (0.0)	2 (11.1)	16 (88.9)	0 (0.0)
Treatment	7 (23.3)	21 (70)	2 (6.7)	2 (6.7)	28 (93.3)	0 (0.0)
Forward Assistance						
Control	4 (22.2)	14(77.8)	0 (0.0)	3 (16.7)	15 (83.3)	0 (0.0)
Treatment	8 (26.7)	19 (63.3)	3 (10)	5 (16.7)	23 (76.7)	2 (6.6)
Blind Spot Alert						
Control	7 (38.9)	11(61.1)	0 (0.0)	2(11.1)	16(88.9)	0 (0.0)
Treatment	7 (23.3)	20 (66.7)	3 (10.0)	3 (10.0)	24 (80.0)	3 (10.0)

5.2 IVIS Task Performance

The IVIS tasks were performed in drives 2 through 7. There were a total of 9,435 tasks completed across all participants and all days, where 5,996 of these tasks were completed by treatment participants and 3,439 were completed by control participants. The total number of tasks completed in drives 2, 3, 4, 5, 6, and 7, respectively, was 1,265, 1,516, 1,630, 1,695, 1,750, and 1,579. The IVIS application randomly generated the task sequence for each driver, and as expected from this randomization, none of the tasks were over represented. Across all participants and all drives, there was a total of 1,240 contact tasks, 1,177 playlist tasks, 1,145 radio tasks, 1,200 climate tasks, and 1,234 dial tasks.

As mentioned previously, task difficulty was *a priori* defined based on calibration during pilot testing. Participants in this study were asked to complete a NASA TLX questionnaire relating to each of these five IVIS task types at the end of each study day. The rank ordering of the NASA TLX scores matched the intended task difficulty ranking (i.e., higher

NASA TLX score indicated increased difficulty). Summary statistics were also computed for task accuracy and time duration, which was collected using the IVIS application (see Table 5.4). These two task performance measures did not identically match the difficulty rank order. However, the two easiest tasks did have the highest accuracy rate (99.3%). One possible explanation for this could be that NASA TLX is a multidimensional assessment of workload, encompassing mental demand, physical demand, temporal demand, performance, effort, and frustration, and also allows each person to weight these individual scales. Thus, task accuracy and duration are just components of the overall NASA TLX assessment.

Table 5.4: Task Performance and NASA TLX by Task Type

Task Type	**Difficulty**	**NASA TLX Score** ***Mean*** ^{SE}	**Accuracy (%)** ***Mean*** ^{SE}	**Duration (sec)** ***Mean*** ^{SE}
Contact	Easy	$35.1^{2.80}$	$99.3^{0.41}$	$14.3^{0.64}$
Playlist	Easy	$35.8^{2.77}$	$99.3^{0.26}$	$17.2^{0.84}$
Radio	Medium	$57.3^{2.71}$	$69.5^{3.31}$	$20.3^{1.31}$
Climate	Hard	$59.4^{2.89}$	$71.5^{1.82}$	$13.9^{0.54}$
Dial	Hardest	$62.0^{2.60}$	$81.2^{2.40}$	$18.5^{0.97}$

TDRT performance by task type was also used to evaluate workload across the different tasks. On average, participants had the shortest response time and lowest miss rate (i.e., lower cognitive workload) to the TDRT stimuli while not engaging with the IVIS. The individual differences between task types are less obvious when aggregated at the entire cohort level (see Table 5.5). Given that, the subsequent chapters analyze TDRT measures at the individual participant's drive level to closer examine these differences over time.

Table 5.5: TDRT Performance by Task Type

Task Type	**Difficulty**	**Response Time (msec)** ***Mean*** ^{SE}	**Miss Rate (%)** ***Mean*** ^{SE}
None	—	$457.6^{30.83}$	$5.6^{0.88}$
Contact	Easy	$718.8^{41.11}$	$30.9^{1.85}$
Playlist	Easy	$714.7^{41.98}$	$31.0^{1.91}$
Radio	Medium	$738.1^{40.50}$	$33.5^{2.25}$
Climate	Hard	$747.1^{44.49}$	$36.1^{2.32}$
Dial	Hardest	$731.9^{41.42}$	$30.7^{1.97}$

Chapter 6
BEHAVIORAL ADAPTATIONS

The proportion of vehicles with autonomous systems is on the rise, with new technologies rapidly entering the market place. As a result, the demands of the driving task are evolving and as such drivers are likely adjusting their behaviors as their role as a driver [operator] changes. Although not all behavioral adaptations are disadvantageous, it is important to understand how these behavioral adaptations manifest in order to identify training and feedback systems needed to minimize adverse outcomes in safety critical situations (e.g, handover and takeover). This chapter evaluates behavioral adaptations of secondary task engagement, driving performance, cognitive workload, and eye glance patterns between baseline (i.e., manual driving), semi-automated, and post semi-automated (i.e., manual) driving conditions. The objective of this chapter is to address the first two research aims of this dissertation:

- *Aim 1: Are there changes in performance and risk perception during exposure to a lane keeping system?*
- *Aim 2: Do behavioral adaptations persist after the lane keeping system is withdrawn?*

6.1 Analytical Methods

Generalized Linear Mixed Models and repeated measures ANOVA with Tukey's HSD test were used to evaluate the immediate and carryover effects of automation exposure on drivers. Drives 1 and 2 provided baseline measures for before exposure to automation. Drives 3 through 6 were used to evaluate immediate effects, as these were the drives with the lane keeping system engaged. Drives 7 and 8 were used to evaluate the carryover effects, as these drives required returning back to manual driving.

6.2 Results

6.2.1 Secondary Task Engagement

Secondary tasks on the IVIS were completed in drives 2 through 7 for all participants. The mean number of IVIS tasks completed and mean completion accuracies were aggregated at the drive level by automation group (see Figure 6.1). These aggregated trends suggests that treatment participants completed more tasks while the lane keeping system was engaged, as compared to the control participants. However, trends in accuracy were similar between the two groups across the six drives, suggesting that the presence of automation did not enhance task accuracy.

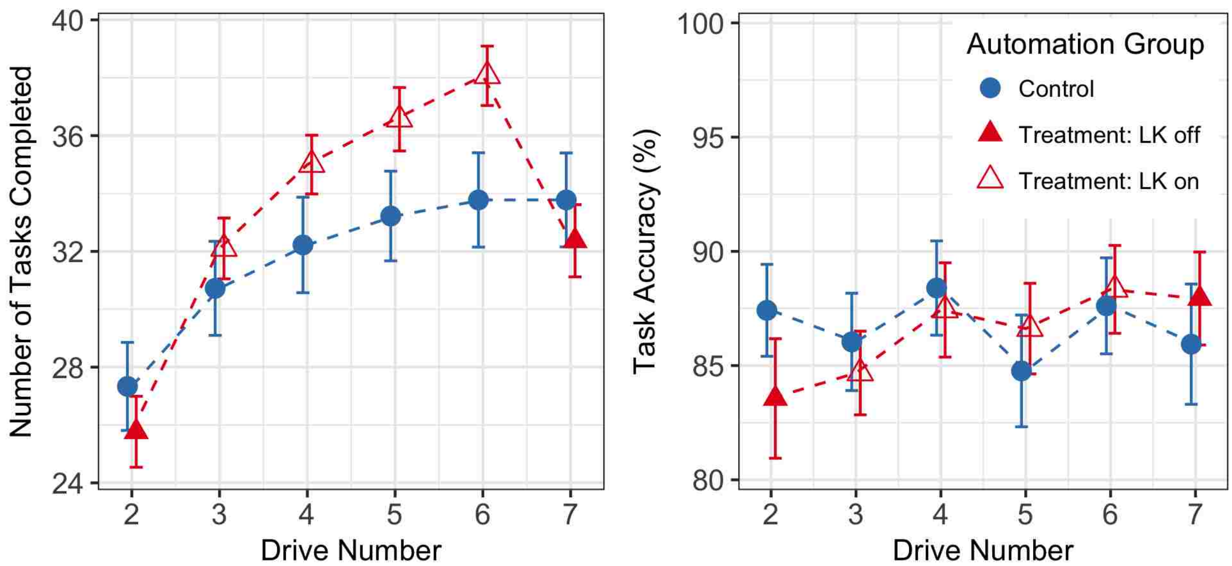

Figure 6.1: IVIS Task Performance for Completion *(left)* and Accuracy *(right)*

Wilcoxon rank sum tests were used to evaluate differences in mean task accuracies between the treatment and control groups across various drives (see Table 6.1). Wilcoxon rank sum was used due to the non-normality of the task accuracy distributions for the two groups. These pairwise tests support the hypothesis derived from the previous figure (Figure 6.1), that task accuracy was similar regardless of the presence of automation. Specifically, there was no difference between or within groups across all drives aggregated together (*All Drives*), before exposure to automation (*Drive 2*), during exposure (*Drive 6*), after withdrawal (*Drive 7*), or between the pairwise contrasts for these conditions.

Table 6.1: Wilcoxon Rank Sum Test on Task Accuracy by Drive and Group

Contrast	**W-statistic**	**p-value**
Control (All Drives) vs. Treatment (All Drives)	9354.0	0.593 (ns)
Control (Drive 2) vs. Treatment (Drive 2)	306.0	0.449 (ns)
Control (Drive 6) vs. Treatment (Drive 6)	242.5	0.565 (ns)
Control (Drive 7) vs. Treatment (Drive 7)	226.5	0.360 (ns)
Control (Drive 2) vs. Control (Drive 6)	164.0	0.962 (ns)
Control (Drive 2) vs. Control (Drive 7)	174.5	0.704 (ns)
Control (Drive 6) vs. Control (Drive 7)	147.0	0.656 (ns)
Treatment (Drive 2) vs. Treatment (Drive 6)	381.0	0.310 (ns)
Treatment (Drive 2) vs. Treatment (Drive 7)	378.5	0.294 (ns)
Treatment (Drive 6) vs. Treatment (Drive 7)	437.0	0.853 (ns)

The trends in secondary task engagement were further explored using a linear mixed model, where the outcome variable was total number of tasks completed. The fixed effects included the full interaction of automation group with drive number, where drive 2 (before automation) was the reference drive. Participant demographics were also accounted for in the model. A random intercept was fit on the participant level. The results of the model are summarized in Table 6.2. This model suggests that there was no initial or final significant difference in the mean number of tasks completed between the two groups, as expressed by the p-values for the *Treatment* and *Treatment* × *Drive 7* coefficients. However, the interaction term for *Treatment* × *Drive 3-6* suggests that the presence of the lane keeping system correlated with a significant increase in the number of tasks completed, relative to their performance before the automation exposure. There was also an increasing trend for the number of tasks completed regardless of automation group as the study progressed, as evident by the increasing magnitude in the *Drive 3-7* coefficient estimates. There was no significant difference between genders (see *Male* variable) or between middle and younger aged groups (see *Middle Aged Group* variable). However, the older aged participants completed significantly less tasks throughout all of the drives, as compared to younger individuals (see *Older Aged Group* variable).

Table 6.2: Linear Mixed Model on Task Count

Variable	**Estimate**	**SE**	**t-statistic**	**p-value**
(Intercept)	29.20	1.87	15.62	< 0.001
Treatment	-1.50	1.76	-0.85	0.398 (ns)
reference: Treatment × Drive 2	–	–	–	–
Treatment × Drive 3	2.94	1.41	2.09	0.038
Treatment × Drive 4	4.34	1.41	3.08	0.002
Treatment × Drive 5	4.91	1.41	3.48	< 0.001
Treatment × Drive 6	5.86	1.41	4.15	< 0.001
Treatment × Drive 7	0.16	1.41	0.11	0.912 (ns)
reference: Drive 2	–	–	–	–
Drive 3	3.39	1.12	3.04	0.003
Drive 4	4.89	1.12	4.38	< 0.001
Drive 5	5.89	1.12	5.28	< 0.001
Drive 6	6.44	1.12	5.78	< 0.001
Drive 7	6.44	1.12	5.78	< 0.001
reference: Younger	–	–	–	–
Middle Aged Group	-2.30	1.82	-1.26	0.214 (ns)
Older Aged Group	-5.99	1.76	-3.40	0.002
Male	1.80	1.46	1.23	0.225 (ns)
Model Fit	*AIC*	*LogLik*	*L Ratio*	*p-value*
At Convergence	1657.2	-811.6	216.4	< 0.001
Null	1845.6	-919.8		
Number of Observations = 288				
Number of Groups = 48				
Intraclass Correlation = 0.679				

Further visualization of this model is provided in Figure 6.2, where the effect size with 95% confidence intervals is provided for the estimated change in number of tasks completed for each factor. Although this plot expresses the same information as Table 6.2, it allows for clearer comparison of relative magnitudes and significance across variables. Specifically, this plot demonstrates how the effect of the lane keeping system was associated with a substantial increase in task engagement for the drives in which it was turned on. Additionally, being in the older age group was associated with one of the largest effects on number of tasks completed.

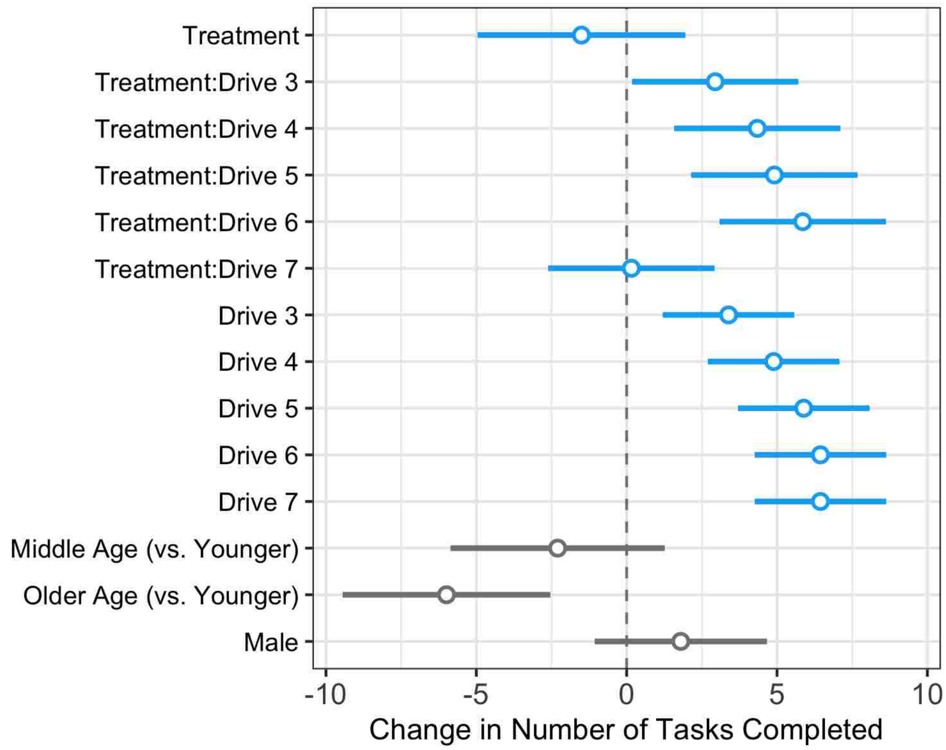

Figure 6.2: Coefficient Estimates with 95% CI for LMM on Task Count

6.2.2 Cognitive Workload

Cognitive workload was measured using the TDRT. The DRT data was first aggregated at the drive level for every participant to compare overall trends across time (see Figure 6.3). Only drives 2-7 are plotted as these were the drives with a secondary task, thus were relatively comparable drives for comparing workload measured by the TDRT. In this plot, each participant's baseline DRT performance, as measured in drive 1, was subtracted from their subsequent drives (e.g., $DRT_{drive2} - DRT_{drive1}$, $DRT_{drive3} - DRT_{drive1}$, etc.). This adjusted DRT measure, relative to their baseline, allowed for comparison between participants. There was an overall decreasing trend in cognitive workload (i.e., lower response time and lower miss rate) as the study progressed for both automation groups, as evident in the plot by the negative slopes between drives. However, participants in the treatment group had an increase in response time and miss rate (i.e., increase in cognitive workload) in drive 7, which was when the automation was withdrawn.

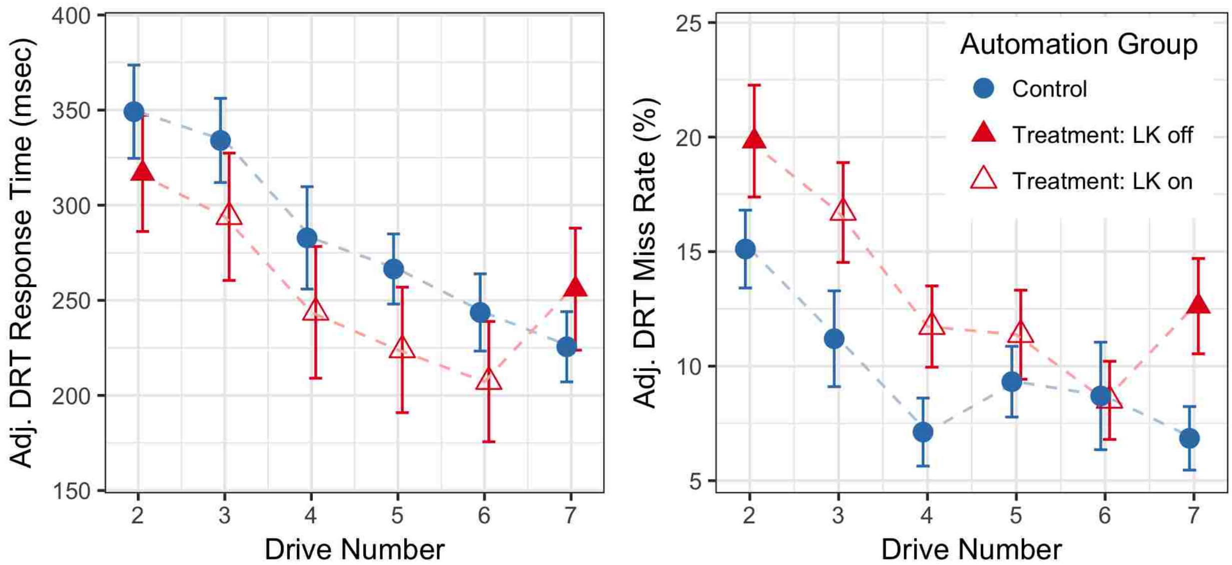

Figure 6.3: DRT Performance for Response Time *(left)* and Miss Rate *(right)*

A linear mixed model was fit on the TDRT response time data, aggregated at the task level for each participant's drive (see Table 6.3). There was no significant difference in response time between automation groups before or during exposure to the automation, as suggested by the p-values for *Treatment* and *Treatment* × *Drive 3-6* coefficients. However, the treatment group had a significant increase in response time in drive 7 (*Treatment* × *Drive 7*), and this factor was relative to their before exposure performance (*Treatment* × *Drive 2*). The increasingly more negative coefficients for *Drive 3* to *Drive 7* likely represents the time on task effect, where participants became more comfortable with the study procedures as they gained experience (i.e., less workload experienced for study tasks). There was no significant difference in response time between the two easy IVIS tasks (i.e, *Playlist Task* with *Contact* as reference task). The *Radio* and *Climate Tasks*, which were designed to be more difficult then the easy (*Contact*) task, were associated with longer response times. There was no significant difference between genders (*Male* coefficient) or middle and younger aged individuals (*Middle Age Group* coefficient). However, older drivers had significantly higher response time measures as compared to younger participants (*Older Age Group* coefficient). Each participants' inherently unique response time was accounted for in the model by including the continuous covariate *Baseline Response Time* (i.e., response time in drive 1 for each participant).

Table 6.3: Linear Mixed Model on DRT Response Time, msec

Variable	**Estimate**	**SE**	**t-statistic**	**p-value**
(Intercept)	299.35	80.877	3.70	< 0.001
Treatment	-9.22	58.474	-0.16	0.875 (ns)
reference: Treatment × Drive 2	–	–	–	–
Treatment × Drive 3	-17.32	22.591	-0.77	0.443 (ns)
Treatment × Drive 4	-16.55	22.577	-0.73	0.464 (ns)
Treatment × Drive 5	-22.22	22.577	-0.98	0.325 (ns)
Treatment × Drive 6	-22.35	22.577	-0.99	0.323 (ns)
Treatment × Drive 7	47.09	22.577	2.09	0.037
reference: Drive 2	–	–	–	–
Drive 3	-23.77	17.814	-1.33	0.182 (ns)
Drive 4	-76.91	17.814	-4.32	< 0.001
Drive 5	-95.51	17.814	-5.36	< 0.001
Drive 6	-110.17	17.814	-6.18	< 0.001
Drive 7	-125.06	17.814	-7.02	< 0.001
reference: Contact	–	–	–	–
Playlist Task	-3.35	9.967	-0.34	0.737 (ns)
Radio Task	27.55	9.976	2.76	0.006
Climate Task	20.57	9.976	2.06	0.039
Dial Task	11.12	9.976	1.11	0.265 (ns)
Baseline Response Time	0.97	0.118	8.21	< 0.001
reference: Younger	–	–	–	–
Middle Age Group	-29.83	66.464	-0.45	0.657 (ns)
Older Age Group	154.60	63.674	2.43	0.019
Male	18.46	52.402	0.35	0.726 (ns)
Model Fit	*AIC*	*LogLik*	*L Ratio*	*p-value*
At Convergence	18039	-8997.6	639.0	< 0.001
Null	18257	-9125.5		
Number of Observations = 1436				
Number of Groups = 48				
Intraclass Correlation = 0.693				

The coefficient estimates and corresponding 95% confidence intervals for their effect on response time, as derived from the above model, are plotted in Figure 6.4. This figure highlights the decreasing trend in cognitive workload as the study progressed (*Drive 3* to *Drive 7*) and significant increase in workload for the treatment participants once the automation was withdrawn (*Treatment* × *Drive 7*). The relatively wide confidence interval for the effect of older age on response time also shows how being in the older age group could be associated with as much as 280 msec slower response time.

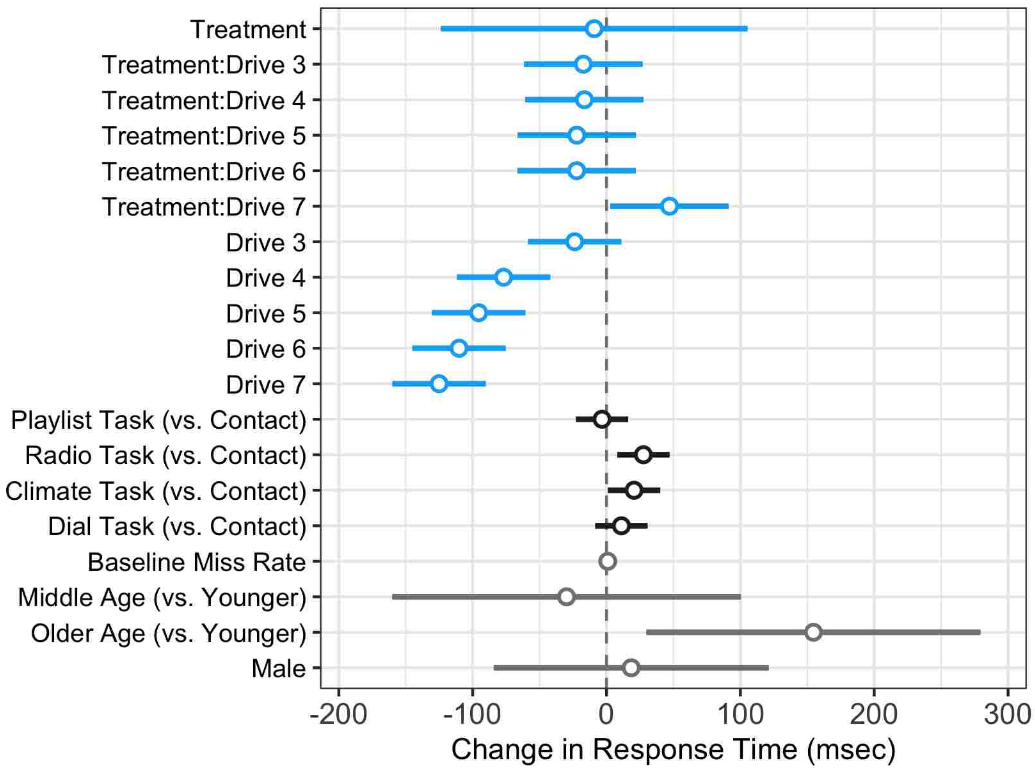

Figure 6.4: Coefficient Estimates with 95% CI for LMM on DRT Response Time

A mixed model was also fit on DRT miss count. This model structure utilized the negative binomial link function, as the response variable was a count variable and followed a negative binomial distribution (i.e., overdispersed). The total number of DRT stimuli within the given time interval (i.e., total number of vibrations during each task for the participant) was used as an exposure variable. Since the negative binomial link function utilizes the logarithmic scale, the miss count can also be expressed as miss rate (see Equation 6.1).

$$\begin{aligned} log(E(MissCount_i)) =& \beta X_i + log(Exposure_i) \\ log(E(MissCount_i)) - log(Exposure_i) =& \beta X_i \\ log\left(E\left(\frac{MissCount_i}{Exposure_i}\right)\right) = log(E(MissRate_i)) =& \beta X_i \end{aligned} \tag{6.1}$$

The model for DRT miss count is summarized in Table 6.4. The results are consistent with the estimated cognitive workload based on response time from Table 6.3. Specifically, there is no statistically significant difference between automation groups through drive 5 (i.e., *Treatment* and *Treatment* × *Drive 3-5*). While there is an associated increase in workload after the automation is removed (*Treatment* × *Drive 7*), and this is relative to before exposure (*Treatment* × *Drive 2*). This model also suggests that older drivers experienced higher cognitive workload compared to young drivers, as expressed by the *Older Age Group* variable. However, miss count appears more sensitive as compared to response time in identifying a reduction in workload in drive 6 (*Treatment* × *Drive 6*) and identifying differences in workload between the five IVIS tasks, where all four tasks are associated with higher workload as compared to the Contact task.

Table 6.4: Negative Binomial Mixed Model on DRT Miss Count

Variable	**Estimate**	**SE**	**t-statistic**	**p-value**
(Intercept)	-2.409	0.298	-8.09	< 0.001
Treatment	0.241	0.252	0.96	0.339 (ns)
reference: Treatment × Drive 2	–	–	–	–
Treatment × Drive 3	0.130	0.086	1.51	0.132 (ns)
Treatment × Drive 4	0.126	0.091	1.39	0.165 (ns)
Treatment × Drive 5	-0.071	0.089	-0.80	0.423 (ns)
Treatment × Drive 6	-0.338	0.088	-3.83	< 0.001
Treatment × Drive 7	0.236	0.091	2.59	0.010
reference: Drive 2	–	–	–	–
Drive 3	-0.272	0.071	-3.85	< 0.001
Drive 4	-0.568	0.075	-7.60	< 0.001
Drive 5	-0.425	0.072	-5.89	< 0.001
Drive 6	-0.335	0.070	-4.77	< 0.001
Drive 7	-0.635	0.075	-8.44	< 0.001
reference: Contact	–	–	–	–
Playlist Task	0.101	0.041	2.47	0.014
Radio Task	0.294	0.039	7.46	< 0.001
Climate Task	0.283	0.040	6.99	< 0.001
Dial Task	0.085	0.040	2.11	0.035
Baseline Miss Rate	0.043	0.012	3.49	< 0.001
reference: Younger	–	–	–	–
Middle Age Group	0.179	0.290	0.62	0.537 (ns)
Older Age Group	0.658	0.278	2.37	0.018
Male	-0.361	0.232	-1.56	0.120 (ns)
Model Fit	*AIC*	*LogLik*	*L Ratio*	*p-value*
At Convergence	7064.0	-3510.0	382.2	< 0.001
Null	7408.3	-3701.1		
Number of Observations = 1438; Number of Groups = 48				
Intraclass Correlation = 0.091				

The incidence rate ratios (IRRs) from the regression on miss count are plotted in Figure 6.5. This plot emphasizes the relative effects of each variable on cognitive workload based on the model output summarized in Table 6.4. Specifically, this plot highlights the decreasing trend in miss rates for the treatment group as the study progressed, and how the reduction in workload becomes statistically significant in drive 6. This suggests that by drive 6, the treatment participants may have had enough exposure to the system for it to effectively reduce cognitive workload. Additionally, the strongest effect on cognitive workload is associated with being in the older age group.

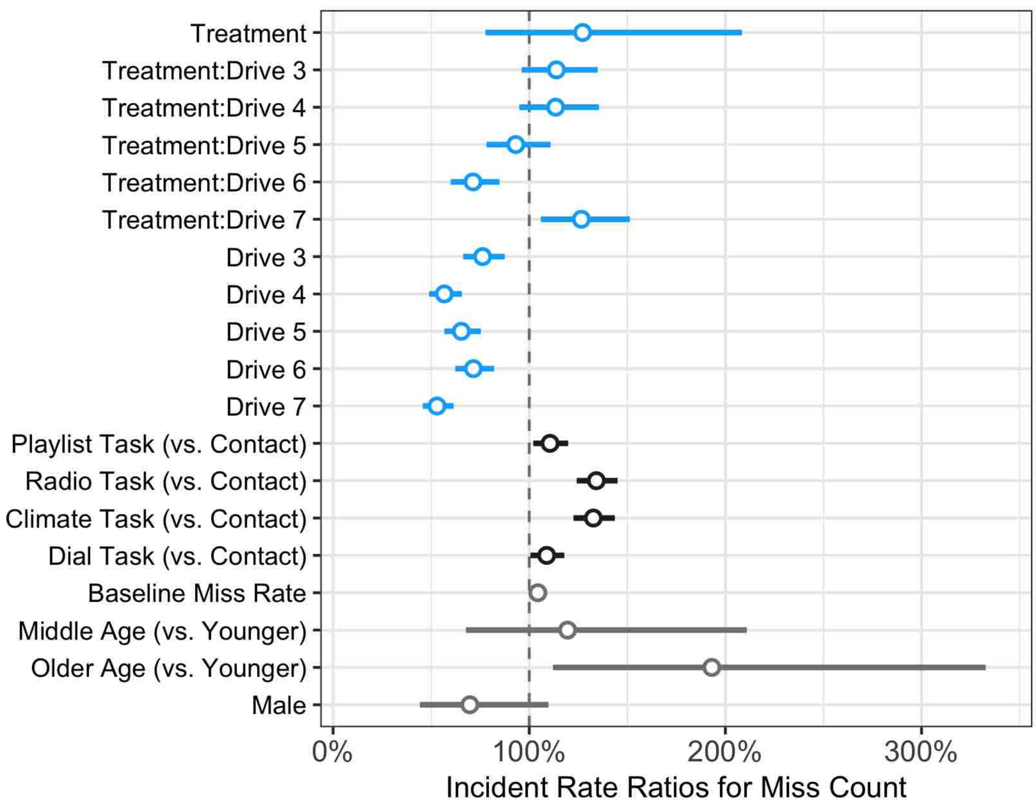

Figure 6.5: Incidence Rate Ratio Estimates with 95% CI on DRT Miss Count

6.2.3 Driving Performance

Driving performance was analyzed in terms of Standard Deviation of Lateral Position, SDLP, (i.e., driving maneuver directly aided by autonomous intervention) and Time to Collision, TTC, (i.e., driving maneuver not directly effected by autonomous intervention).

Standard Deviation of Lateral Position

SDLP was aggregated at the drive level across participants to compare general trends between the two automation groups. It is expected that different roadway curvature would have different values of SDLP, thus this was accounted for in Figure 6.6, while all road alignments were aggregated for simplicity in Figure 6.7. Participants in the control group had an overall decreasing trend in lateral deviation as the study progressed for comparable drives (i.e., drives with and without distracting tasks). However, participants in the treatment group had an increase in lateral deviation after the system was removed (drive 7). Participants in both groups had similar initial values of SDLP (i.e., means between groups within drive 1 and drive 2). The lane keeping system provided lateral control of the vehicle, hence the low values of SDLP in drives 3 through 6. The noticeable difference in mean SDLP between the two groups in drive 7 emphasizes this change in behavior after exposure. These trends were similar across all roadway alignments, with larger variations on the road segments with curvature.

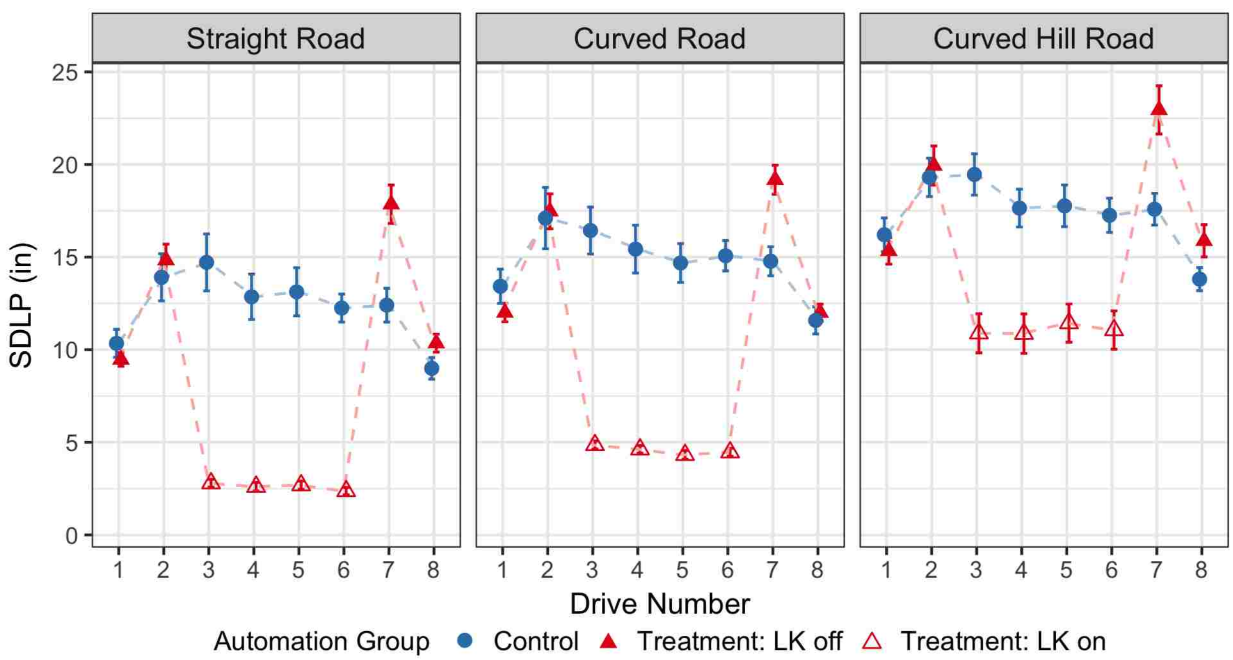

Figure 6.6: Observed SDLP ± 1 SE Across Drives by Automation Group and Road Type

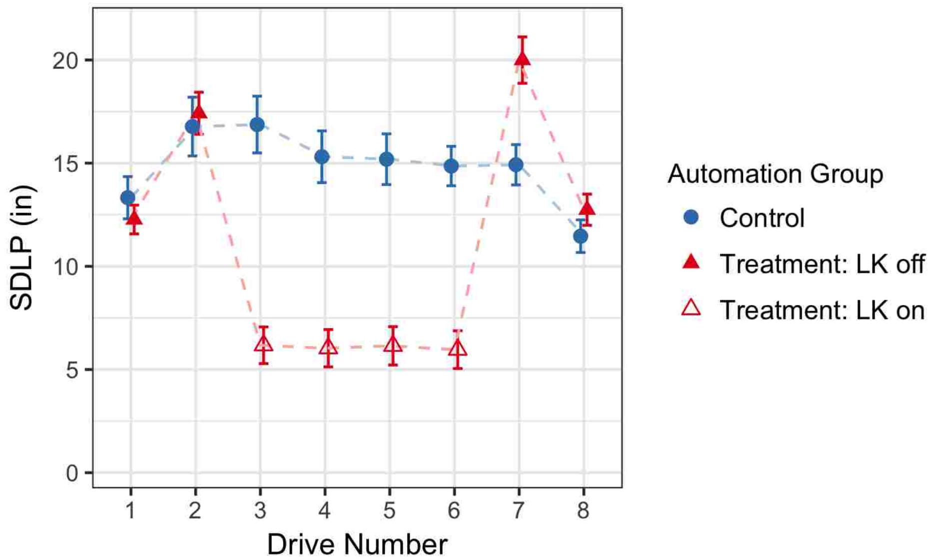

Figure 6.7: Observed SDLP ± 1 SE Averaged Across All Roadway Alignments

A linear mixed model was fit on the log transformed values of SDLP, which were transformed for normality, see Table 6.5. In this model, the *Treatment* × *Drive 2-8* variables estimate differences specifically observed in the treatment group that differ from their drive 1 measures. The results support the notion that there were no initial significant differences in SDLP between the two groups before exposure, as expressed by the p-values for *Treatment* and *Treatment* × *Drive 2*. Additionally, treatment participants had significantly larger lateral deviations after withdrawing the system (*Treatment* × *Drive 7-8*), relative to their SDLP values before exposure with no distracting tasks (drive 1). Participants had an increase in SDLP relative to drive 1 for the first two drives after adding the IVIS tasks *Drive 2* and *Drive 3*. However, *Drives 4-7*, which still included the IVIS tasks, was not associated with any additional increase in SDLP relative to the first drive. This could be capturing the time on task effect, where participants became better at managing the demands of the study tasks. The negative estimate for *Drive 8* further supports this time on task effect, where participants may have become more comfortable with the study environment and thus able to perform better. The roadway sections with curvature, *Curved Road* and *Curved Hill Road*, also had larger values of SDLP as compared to *Straight Road* segments. There were no significant differences in SDLP between genders or age groups.

Table 6.5: Linear Mixed Model on log(SDLP, inches)

Variable	**Estimate**	**SE**	**t-statistic**	**p-value**
(Intercept)	2.152	0.083	26.05	< 0.001
Treatment	-0.079	0.083	-0.95	0.346 (ns)
reference: Treatment × Drive 1	–	–	–	–
Treatment × Drive 2	0.124	0.090	1.39	0.166 (ns)
Treatment × Drive 3	-1.123	0.090	-12.51	< 0.001
Treatment × Drive 4	-1.031	0.090	-11.49	< 0.001
Treatment × Drive 5	-1.028	0.090	-11.46	< 0.001
Treatment × Drive 6	-1.086	0.090	-12.11	< 0.001
Treatment × Drive 7	0.369	0.090	4.12	< 0.001
Treatment × Drive 8	0.182	0.090	2.03	0.043
reference: Drive 1	–	–	–	–
Drive 2	0.221	0.071	3.12	0.002
Drive 3	0.229	0.071	3.23	0.001
Drive 4	0.096	0.071	1.35	0.176 (ns)
Drive 5	0.098	0.071	1.38	0.167 (ns)
Drive 6	0.122	0.071	1.73	0.085 (ns)
Drive 7	0.122	0.071	1.72	0.086 (ns)
Drive 8	-0.145	0.071	-2.04	0.042
reference: Straight	–	–	–	–
Curved Road	0.323	0.027	12.15	< 0.001
Curved Hill Road	0.700	0.027	26.32	< 0.001
reference: Younger	–	–	–	–
Middle Age Group	-0.005	0.070	-0.07	0.948 (ns)
Older Age Group	0.036	0.068	0.53	0.599 (ns)
Male	0.068	0.056	1.22	0.229 (ns)
Model Fit	*AIC*	*LogLik*	*L Ratio*	*p-value*
At Convergence	1084.9	-519.4	1399.8	< 0.001
Null	2444.6	-1219.3		
Number of Observations = 1152; Number of Groups = 48				
Intraclass Correlation = 0.191				

These coefficient estimates with 95% confidence intervals are also visualized in Figure 6.8, which are expressed as a percent change in SDLP due to the log transformation of SDLP in the model. This plot emphasizes the relative magnitude of effect for each factor, such as the largest values of SDLP were associated with the drive after removing the system (*Treatment × Drive 7*) and the roads with curvature (*Curved Road* and *Hill Road*).

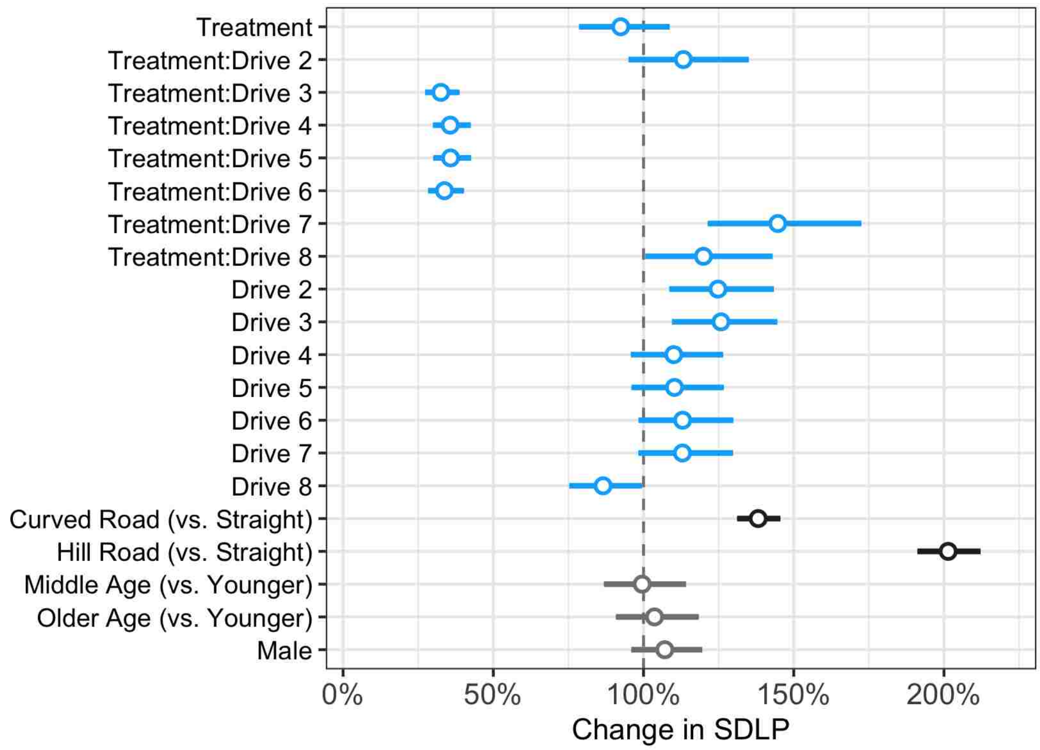

Figure 6.8: Coefficient Estimates with 95% CI for LMM on SDLP

Time to Collision

TTC was also aggregated at the drive level across participants to compare general trends between the two automation groups, see Figure 6.9 for differentiation by roadway alignment and Figure 6.10 for trends averaged across all roadway alignments. Although participants in the treatment group appeared to have lower initial mean values of TTC, there still appears to be an effect associated with the lane keeping system. Participants in both groups had an increase in TTC from drive 1 to drive 2, which corresponded to driving under manual conditions only (drive 1) to manual driving with the addition of the secondary task (drive 2). This is to be expected, as it is likely that drivers would increase their TTC when engaging in a secondary task in an effort to provide a larger safety margin between them and the car in front of them. However, mean TTC for the treatment group reduced after the lane keeping was engaged (drives 3 through 6) and remained low even after the system was withdrawn (drives 7 and 8). Whereas, the control group had relatively consistent mean TTC values across all drives with the IVIS distracting tasks (drive 2 through 7).

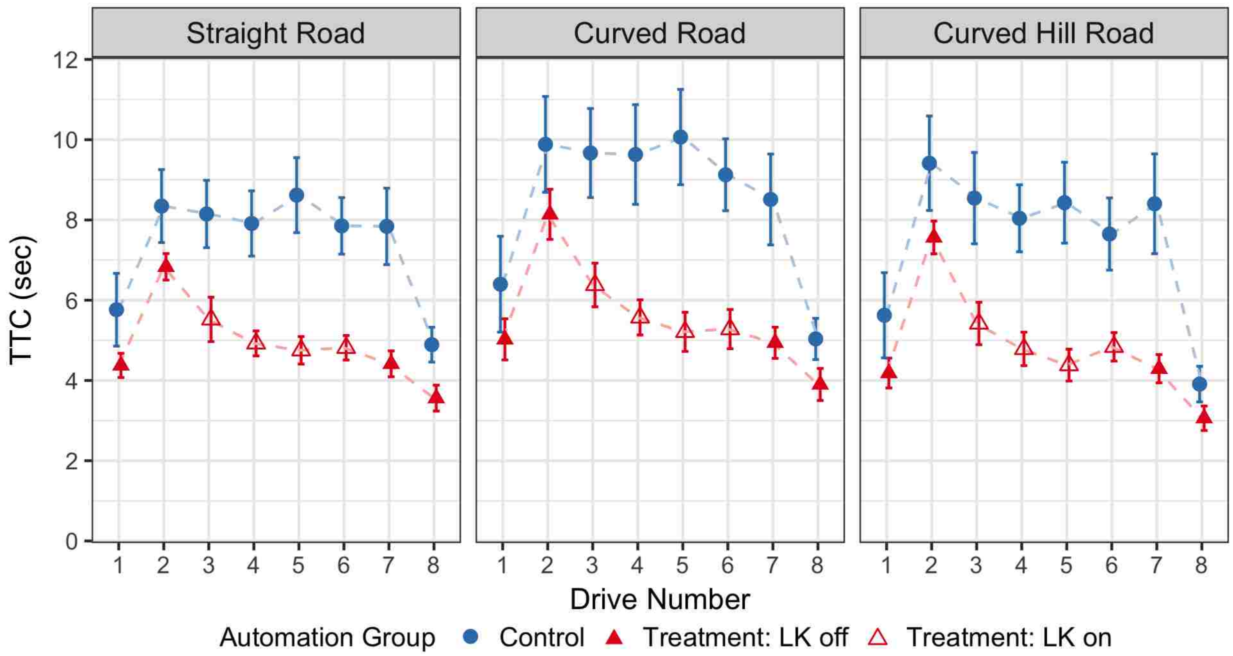

Figure 6.9: Observed TTC ± 1 SE Across Drives by Automation Group and Road Type

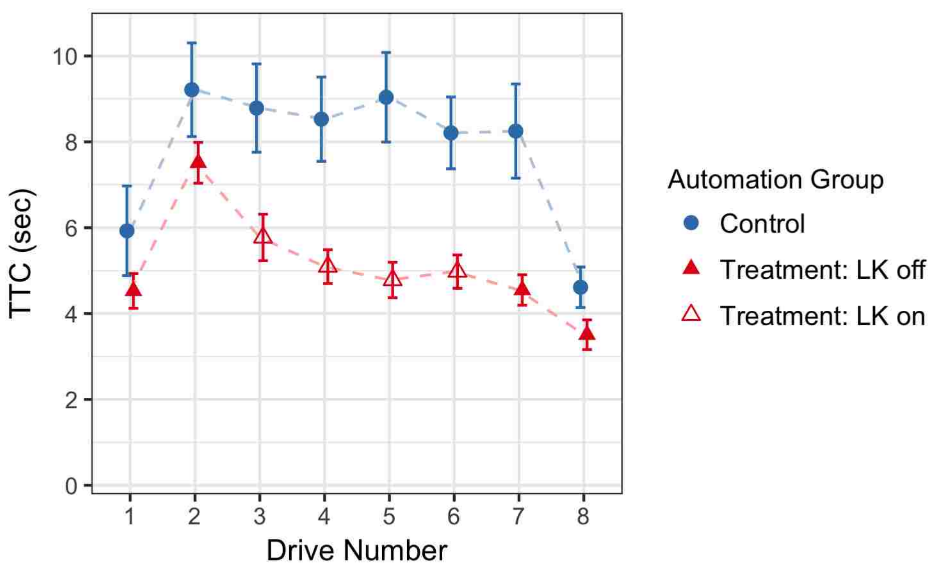

Figure 6.10: Observed TTC ± 1 SE Averaged Across All Roadway Alignments

A linear mixed model was fit on the log transformed TTC observations, see Table 6.6. There was no significant difference in TTC between the two groups before the automation intervention, as suggested by the p-values for the *Treatment* and *Treatment* × *Drive 2* coefficients. Exposure to the lane keeping system (*Treatment* × *Drive 3-6*) was associated with lower TTC as compared to their values of TTC in drive 1, and a carryover effect of this lower TTC was found for the drive immediately following withdrawal (*Treatment* $\times$ *Drive 7*). Meanwhile, the coefficient estimates for *Drive 2-7* (i.e., drives with IVIS tasks) suggests that the control participant increased their TTC values relative to their drive 1 values. There were also differences in TTC associated with demographic effects, where older participants had larger values of TTC as compared to the younger drivers (see *Older Age Group*). Additionally, *Males* had smaller values of TTC as compared to females.

Table 6.6: Linear Mixed Model on log(TTC, sec)

Variable	**Estimate**	**SE**	**t-statistic**	**p-value**
(Intercept)	1.498	0.115	12.98	< 0.001
Treatment	-0.151	0.109	-1.39	0.172 (ns)
reference: Treatment × Drive 1	–	–	–	–
Treatment × Drive 2	0.019	0.089	0.22	0.828 (ns)
Treatment × Drive 3	-0.307	0.089	-3.46	< 0.001
Treatment × Drive 4	-0.365	0.089	-4.11	< 0.001
Treatment × Drive 5	-0.509	0.089	-5.74	< 0.001
Treatment × Drive 6	-0.364	0.089	-4.10	< 0.001
Treatment × Drive 7	-0.423	0.089	-4.77	< 0.001
Treatment × Drive 8	-0.170	0.089	-1.92	0.055 (ns)
reference: Drive 1	–	–	–	–
Drive 2	0.552	0.070	7.86	< 0.001
Drive 3	0.517	0.070	7.37	< 0.001
Drive 4	0.490	0.070	6.98	< 0.001
Drive 5	0.548	0.070	7.80	< 0.001
Drive 6	0.471	0.070	6.71	< 0.001
Drive 7	0.434	0.070	6.19	< 0.001
Drive 8	-0.121	0.070	-1.73	0.084 (ns)
reference: Straight	–	–	–	–
Curved Road	0.087	0.03	3.302	< 0.001
Curved Hill Road	-0.050	0.026	-1.918	0.055 (ns)
reference: Younger	–	–	–	–
Middle Age Group	0.184	0.110	1.674	0.101 (ns)
Older Age Group	0.251	0.11	2.351	0.023
Male	-0.201	0.09	-2.274	0.028
Model Fit	*AIC*	*LogLik*	*L Ratio*	*p-value*
At Convergence	1105.1	-529.5	447.8	< 0.001
Null	1512.9	-753.4		
Number of Observations = 1152; Number of Groups = 48				
Intraclass Correlation = 0.399				

These coefficient estimates with 95% confidence are also plotted in Figure 6.11 to further visualize their relative effect on percent change in TTC. Percent change was used because TTC was log transformed for normality in the regression model, thus the coefficient estimates have a multiplicative effect. This plot particularly shows the pronounced effect of secondary task engagement on TTC (*Drive 2-7*).

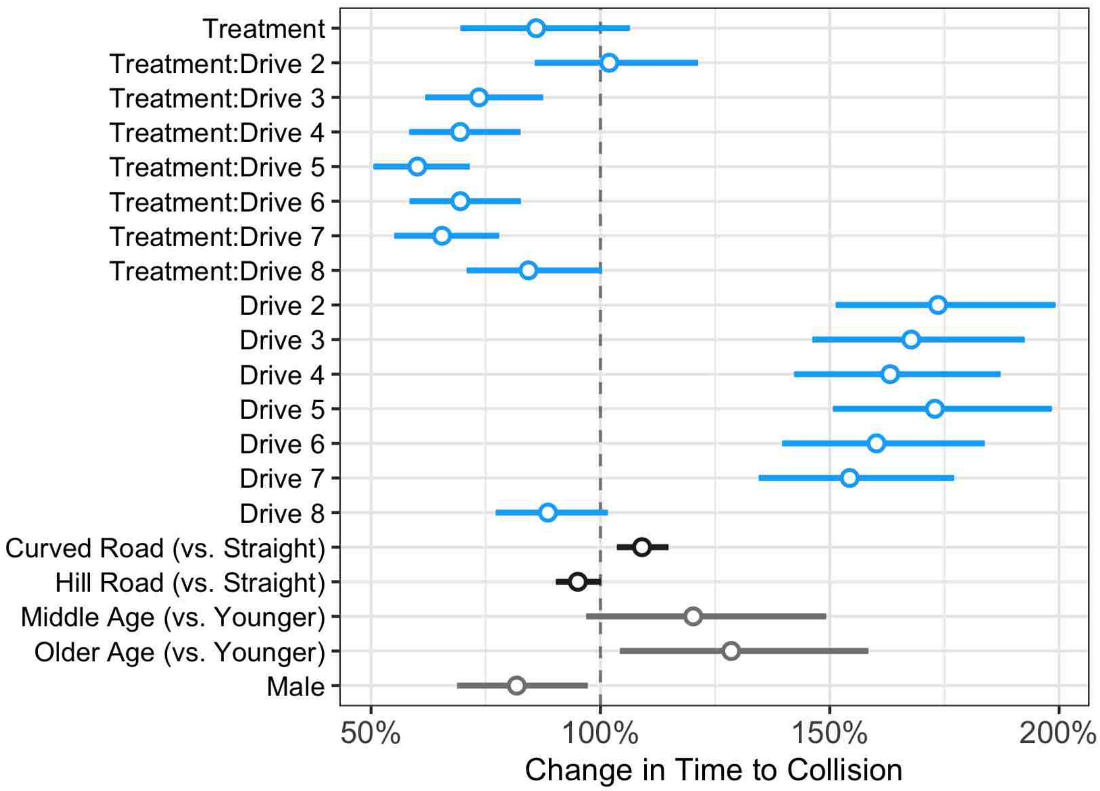

Figure 6.11: Coefficient Estimates with 95% CI for LMM on TTC

6.2.4 Eye Glance Behavior

Eyes-off-road glances to the IVIS touchscreen were compared between drive 2 (before exposure) and drive 7 (after withdrawal). Repeated measures ANOVAs were used to assess differences between the various factors and Tukey's HSD test was used to specifically understand the differences in means for the contrasts of *Automation Group (2 levels: treatment, control)* × *Drive (2 levels: drive 2, drive 7)*.

Mean Glance Duration

Mean glance durations for each IVIS task type were compared between the two automation groups, see Figure 6.12. Most of the mean glances were below 2.0 seconds. However, the average mean GD for the treatment group in drive 7 for the Contact, Playlist, Radio, and Dial tasks increased to approximately 2 seconds. Meanwhile, the control group mean GDs appeared to remain stable or decrease in duration from drive 2 to drive 7.

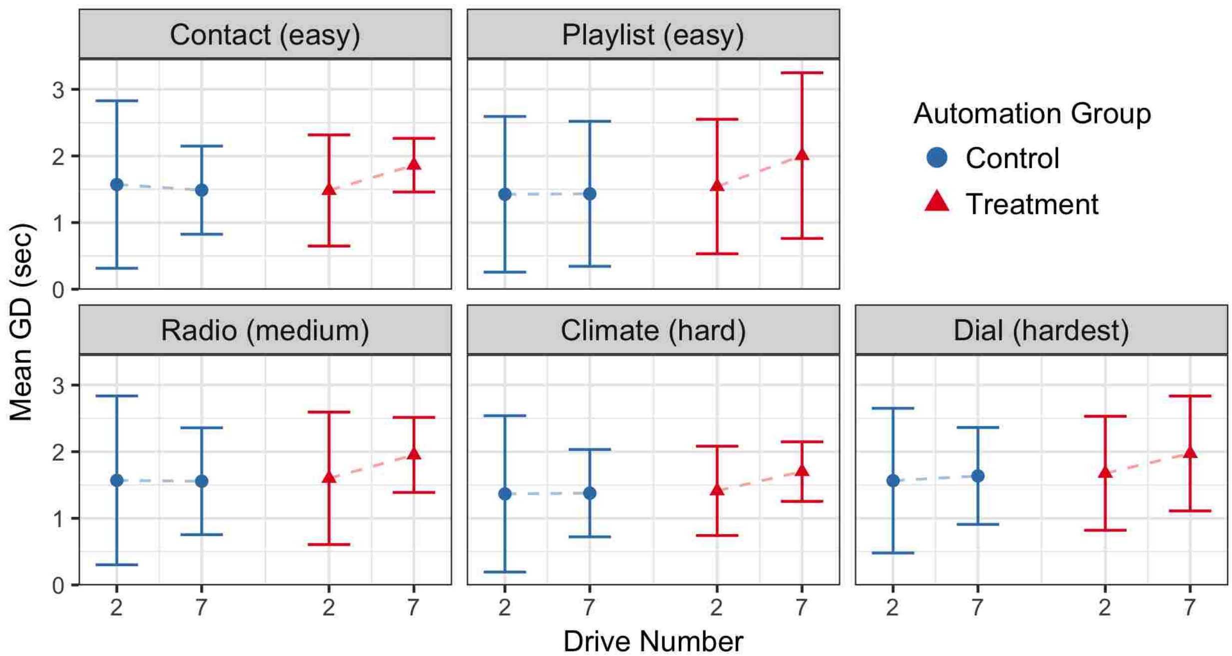

Figure 6.12: Observed Mean Glance Durations ± 1 SE by Task Types for Before (Drive 2) and After (Drive 7) Automation Exposure

As suggested by the trends in Figure 6.12, there was a significant effect for the interaction of drive and automation group ($F(1, 416) = 29.29$, $p < 0.001$). There was also a significant difference in mean GDs between drives ($F(1, 416) = 74.39$, $p < 0.001$) and task type ($F(4, 416) = 12.07$, $p < 0.001$). There was no effect of automation group, gender, or age group. The results of the post hoc test on the comparisons between drive and automation group are summarized in Table 6.7. There were no differences between mean GDs for the control participants in drive 2 or 7, or between the two automation groups in drive 2. However, treatment participants in drive 7 had significantly higher mean values as compared to their glances in drive 2 and as compared to control participants in drives 2 and 7.

Table 6.7: ANOVA Contrasts of Mean Glance Durations, seconds

		Tukey's HSD p-value		
Interaction	**Mean** $^{\mathrm{SD}}$	*Control × Drive 2*	*Control × Drive 7*	*Treatment × Drive 2*
Control × Drive 2	1.499 $^{0.61}$	–		
Control × Drive 7	1.498 $^{0.44}$	ns	–	
Treatment × Drive 2	1.543 $^{0.53}$	ns	ns	–
Treatment × Drive 7	1.898 $^{0.69}$	0.014	0.034	< 0.001

90th Percentile Glance Duration

The 90$^{\text{th}}$ percentile glance durations for drives 2 and 7 by automation group were aggregated by IVIS task type, see Figure 6.13. The control group had fairly comparable 90$^{\text{th}}$ percentile glance durations in drives 2 and 7 across all five task types. However, the treatment group had noticeably higher values of 90$^{\text{th}}$ percentile glance durations in drive 7 as compared to drive 2 across tasks.

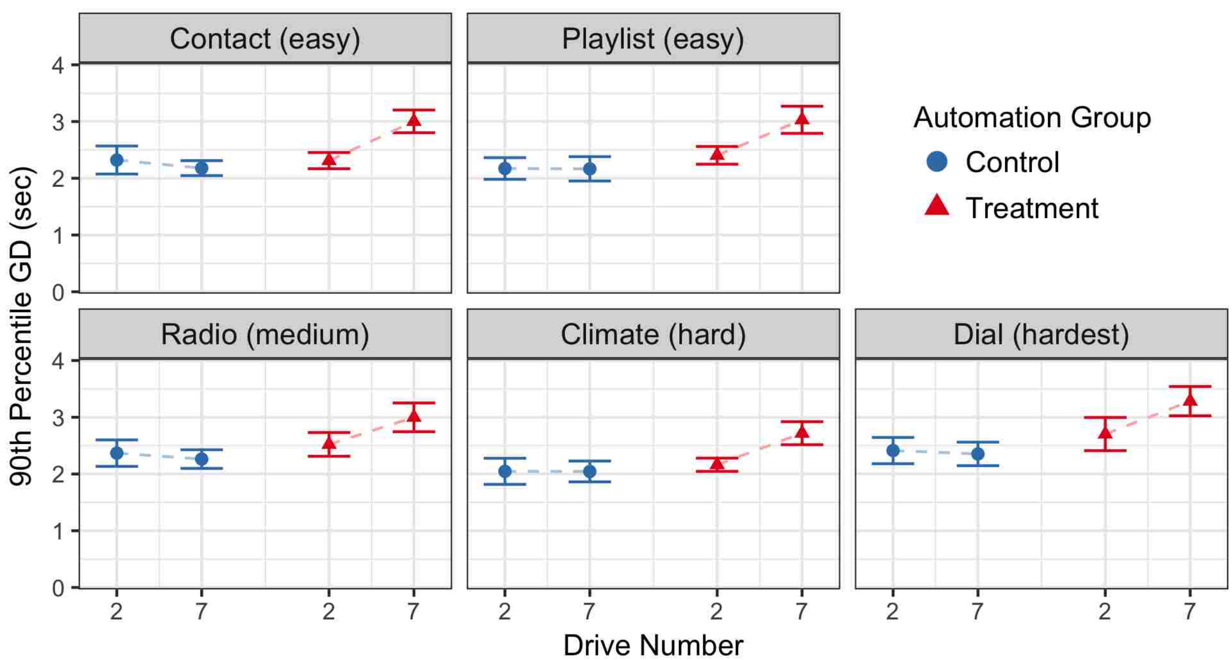

Figure 6.13: Observed 90$^{\text{th}}$ Percentile Glance Durations ± 1 SE by Task Types for Before (Drive 2) and After (Drive 7) Automation Exposure

The ANOVA on 90th percentile glance durations indicated a significant effect for drive ($F(1, 416) = 55.94$, $p < 0.001$), automation group ($F(1, 42) = 5.30$, $p = 0.026$), the interaction of drive and automation group ($F(1, 416) = 36.62$, $p < 0.001$), and task type ($F(4, 416) = 11.32$, $p < 0.001$). There was no significant effect of gender or age group. Tukey's HSD test was used to further evaluate the difference between means for the interaction of drive and automation group, see Table 6.8. The results are similar to the findings for mean glance duration, where only the contrasts involving treatment participants in drive 7 had significantly different mean values of 90th percentile glance durations. While there was no difference in initial values between automation groups (i.e., in drive 2) or between the two drives for the control group participants.

Table 6.8: ANOVA Contrasts of 90th Percentile Glance Durations, seconds

		Tukey's HSD p-value		
Interaction	**Mean ^{SD}**	*Control × Drive 2*	*Control × Drive 7*	*Treatment × Drive 2*
Control × Drive 2	2.264 $^{0.95}$	–		
Control × Drive 7	2.201 $^{0.76}$	ns	–	
Treatment × Drive 2	2.423 $^{1.06}$	ns	ns	–
Treatment × Drive 7	3.007 $^{1.27}$	0.007	0.007	< 0.001

Total Eyes-Off-Road and Long Glances

Total eyes-off-road and percent long glances were aggregated at the drive level by automation group in order to more easily compare these two measures with each other, see Figure 6.14. Participants in both the treatment group and control group had similar total eyes-off-road time behavior across the two drives. Specifically, both automation groups showed a decreasing trend in total eyes-off-road time in drive 7 as compared to drive 2. This suggests that participants were able to, on average, complete each IVIS task quicker after gaining experience with the tasks. However, the two automation groups differed in their proportions of long glances off road, where the control group remained consistent and the treatment group had an increase in percent long glances from drive 2 to 7.

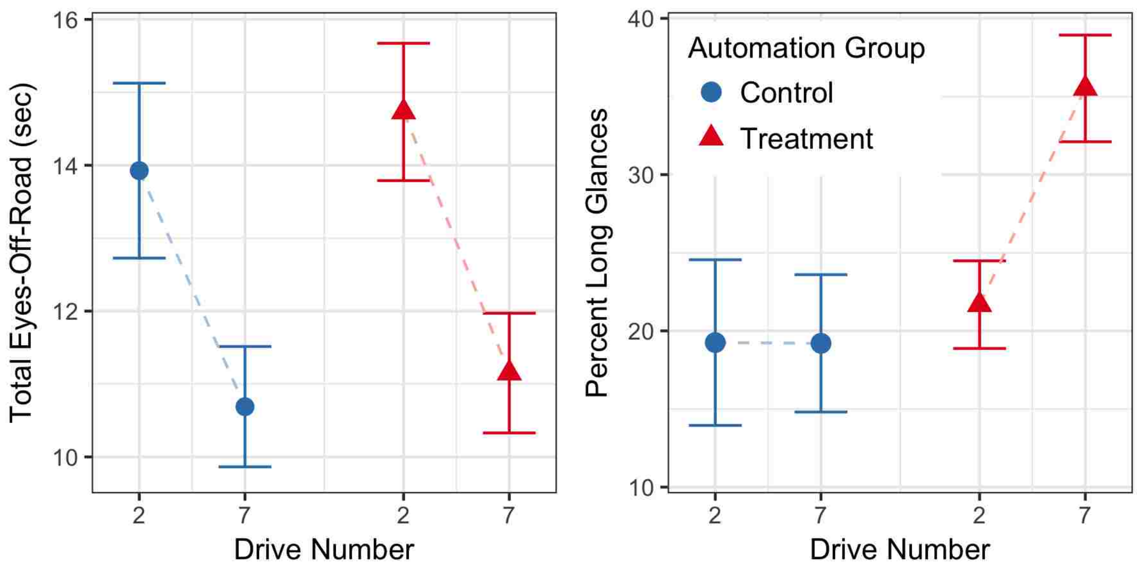

Figure 6.14: Observed Mean ± 1 SE of Total Eyes-Off-Road Time *(left)* and Percent Long Glances *(right)* for Before (Drive 2) and After (Drive 7) Automation Exposure

The results from the ANOVA on total eyes-off-road indicated a significant effect for drive ($F(1, 416) = 195.84$, $p < 0.001$), task type ($F(4, 416) = 36.45$, $p < 0.001$), and age group ($F(2, 42) = 3.48$, $p = 0.04$). There was no significant difference between automation groups, the interaction of automation group and drive, or gender. A Tukey test on age group showed that older drivers had longer mean total eyes-off-road time (mean 13.94, SD 5.63) as compared to younger drivers (mean 11.07, SD 3.94), $p = 0.038$. Tukey's post hoc test was not conducted on the contrasts for *Automation Group* × *Drive* since this interaction variable was not significant in the ANOVA.

The ANOVA on percent long glances showed a significant difference for the factors drive ($F(1, 416) = 56.77$, $p < 0.001$), automation group ($F(1, 42) = 9.90$, $p = 0.003$), interaction of automation group and drive ($F(1, 416) = 13.61$, $p < 0.001$), and task type ($F(4, 416) = 11.48$, $p < 0.001$). There was no significant effect of age group or gender. Tukey's post hoc comparisons of percent long glances for automation group by drive is provided in Table 6.9. Similar to the results for mean and 90th percentile glance durations, there were no differences in means for percent long glances between the two automation groups before exposure (i.e., drive 2), or between the two drives for the control group. However, the treatment group in drive 7 had a larger mean value of percent long glances as compared to in drive 2 and as compared to both drives for the control group.

Table 6.9: ANOVA Contrasts of Percent Long Glances

Interaction	**Mean** $^{\text{SD}}$	**Tukey's HSD p-value** *Control × Drive 2*	*Control × Drive 7*	*Treatment × Drive 2*
Control × Drive 2	19.25 $^{22.48}$	–		
Control × Drive 7	19.20 $^{18.64}$	ns	–	
Treatment × Drive 2	21.68 $^{15.35}$	ns	ns	–
Treatment × Drive 7	35.52 $^{18.70}$	< 0.001	0.002	< 0.001

6.3 Discussion

This chapter evaluated the first two research aims, which focused on intermediate (*Aim 1*) and carryover effects (*Aim 2*) of exposure to an active lane keeping system on driver behavioral adaptations. Changes in secondary task engagement and eyes-off-road patterns were used to quantify adaptations in risk perception; changes in driving performance and cognitive workload were used to quantify adaptations in performance.

6.3.1 Changes in Performance

As expected, drivers with the lane keeping system had less lateral deviations when the system was engaged as compared to manual drivers. However, after the lane keeping system was removed, drivers in the treatment group had a significant increase in lateral deviation relative to their before exposure performance. These results are comparable to Rudin-Brown and Noy (2002), who evaluated the effects of accurate and inaccurate lane departure warning systems, and reported that drivers had improved lane-keeping performance regardless of system accuracy as compared to manual driving. That study also found that drivers self-reported a high degree of trust in the inaccurate system. Those findings, in conjunction with the carryover effect of decreased lateral performance found in this current study, suggest that drivers may begin to over trust these lateral assistance systems.

It is also important to consider the time interval that SDLP was aggregated at for the regression model. For this study, the driving performance measures were aggregated to correspond with the roadway geometry to account for differences between straight and curved roads. Thus, SDLP was computed over segments that were approximately 1.25 miles in length, and took the driver, on average, 1.5 minutes to travel. The mean degradation in SDLP from drive 2 to drive 7 for the treatment participants across a straight segment was approximately 3 inches. In 1.5 minutes, this may only equate to 0.25 feet of variation, but in an average drive of about 30 minutes, that could be as much as an added 5 feet of deviation.

Drivers, regardless of automation group, tended to have an increase in TTC corresponding to the introduction of IVIS tasks on manual driving (i.e., drive 1 to drive 2). Once the lane keeping system was engaged, drivers [in the treatment group] had significantly lower values of TTC as compared to their own values of TTC under manual conditions. This decreased TTC for the treatment group persisted in the drives even after the lane keeping system was withdrawn (i.e., drive 7 and 8). These findings suggest that a lateral assistance system also has an impact on longitudinal vehicle control; that is, lateral and longitudinal driving performance are not independent tasks. This is in accordance with Strand et al. (2014), who also argue that longitudinal and lateral control are not independent; in a study evaluating automation failure in a single direction, drivers experienced adverse effects on both lateral and longitudinal vehicle control. In a similar context, Rudin-Brown and Parker (2004) found a similar relationship in the inverse, where drivers had increased lateral variability when using ACC as compared to manual driving. However, that study did not examine carryover effects.

There was a decreasing trend in cognitive workload for time on task, as measured by the TDRT miss rate and response time. This was observed in the parameter estimates for each consecutive drive in the regression model, where response time became quicker (i.e., coefficient estimates became more negative) for subsequent drives. In the first three drives with the lane keeping engaged (i.e., drives 3-5), there was no significant change in workload, which is somewhat surprising given that the system should alleviate some of the driving demands. This could be due to risk compensation. That is, the driver was increasing their level of engagement with the secondary task (i.e., completing more tasks) as they became more accustomed to the lane keeping system helping with lateral control. As a result, a decrease in cognitive workload would not be observed, as they would be offsetting the decreased driving demands with an increased load induced by completing more secondary tasks. Previous research has found that drivers will increase engagement with IVIS tasks under automated assisted driving (Jamson et al., 2013; DeWinter et al., 2014). Thus, it is possible that drivers in this study were increasing their attention to the secondary task as the driving component became less demanding and hence a change in cognitive workload could not be detected.

In the last drive with the lane keeping engaged (i.e., drive 6), there was a significant and negative interaction effect for treatment and drive on miss rate, which indicated a decrease in cognitive workload relative to before exposure to the system. This reduction in cognitive workload under the lane keeping assistance system is consistent with previous literature that has evaluated workload under ACC and highly-automated driving (DeWinter et al., 2014). This finding suggests that these autonomous systems can have an overall positive effect on safety, as they appear to reduce the overall driving load. This can be leveraged if systems are designed to capture the operators attention at crucial times or target their available attention to more important information.

There was a significant increase in both measures of cognitive workload (i.e., DRT response time and miss rate) for the treatment group when the lane keeping system was turned off, relative to their cognitive workload before exposure. This suggests that there was a carryover effect on cognitive workload from exposure to the lane keeping system, where the driving task may have become more demanding than drivers had previously experienced before exposure. These cognitive models only considered the carryover effect on drive 7, and did not include drive 8. This was because TDRT was used to measure cognitive workload and the ISO standard indicates that the DRT protocol is best intended for use with secondary tasks; there were no secondary tasks in drive 8 (ISO, 2016).

Cognitive workload was assessed using both TDRT miss rate and response time. These measures were mostly consistent with each other in measuring differences in workload across the various study factors. However, it is noteworthy that miss rate appeared to be more sensitive in identifying these changes in workload. One possible explanation for this could be in the relative magnitude that each metric captures workload. That is, response time only analyzes successful responses to the DRT stimulus; in other words, only when participants are able to notice and respond to the event in a timely manner. While miss rate analyzes those instances when drivers do not respond to the stimulus in a reasonable amount of time. Thus, miss rate may be capturing when drivers are too cognitively overloaded to respond. Hence, the possible differences in workload measured across the study for the various tasks.

6.3.2 Changes in Risk Perception

The IVIS tasks were used as a way to evaluate willingness to engage in a distracting task (i.e., riskiness). Participants were instructed to complete the tasks at a pace they were comfortable with and the IVIS application was designed to allow the participant to cue the frequency of the tasks. Thus, engagement with these tasks was considered a measure of risk perception.

The exposure of automation did not appear to improve drivers' task accuracy, as drivers in the control group and treatment group had similar accuracy across all drives. Regardless of automation group, both groups had a similar baseline number of tasks completed (i.e., treatment variable not significant in regression model). However, automation exposure was associated with an increase in task engagement, where drivers in the treatment group completed more tasks while the lane keeping system was engaged (drives 3-6). After the lane keeping system was removed (drive 7), treatment and control participants completed a similar number of tasks. Although this can be considered a good sense of risk compensation by the treatment participants, it is important to note that they were told before the drive that the lane keeping system would not be turned on for the entirety of drive. Hence, a different outcome may occur if the lane keeping system suddenly and unexpectedly handed off control to the driver.

The evaluation of eye glance behavior provided a further measure of risk perception, as eyes-off-road time can be an important indicator of safety. Previous research has shown that eyes-off-road glances (particularly glances greater than 2.0 seconds) can increase standard deviation of lane position (Peng et al., 2013; Liang & Lee, 2010) and increase crash risk (Klauer et al., 2006).

There were no changes in mean glance durations and 90^{th} percentile glance durations for the control group from drive 2 to drive 7. Additionally, there were no initial differences between the two automation groups (i.e., in drive 2). However, mean glance durations and 90th percentile glance durations were significantly higher for the treatment participants in drive 7 (after withdrawal) as compared to their matched glances in drive 2 (before exposure). This suggests that there was a carryover effect on glance durations after removing the system. Specifically, treatment group participants had a mean value of 3.0 seconds for 90th percentile glance durations in drive 7 as compared to the 2.2 second mean value in drive 7 for control participants. This can have implications on safety, particularly as these glances approach or exceed the 2.0 second threshold.

Participants in both automation groups were able to complete the tasks with less total eyes-off-road time in drive 7 as compared to drive 2. Specifically, there was an average decrease in total eyes-off-road time from drive 2 to drive 7 of almost 3.0 seconds for the control group and 4.0 seconds for the treatment group. While there was no significant change in the proportion of glances longer than 2.0 seconds for the control group, there was a significant increase observed for the treatment participants. In fact, treatment participants had an average increase from 21.7% in drive 2 to 35.5% in drive 7 of percent long glances. This is a meaningful change, as these glances are considered particularly unsafe as glances greater than or equal to 2.0 seconds are associated with increased crash risk.

6.3.3 Behavioral Differences Across Demographics

Participant demographics were also accounted for in the regression models and ANOVA analyzes in order to evaluate differences between driving populations. This is an important consideration, as different driving groups may benefit more or less from specific autonomous systems. Thus, systems may be more effective if they can be personalized to individuals or groups.

There were no differences observed in driving performance, cognitive workload, eye glance behaviors, or secondary task engagement between the younger (25-34 years old) and middle (35-44 years old) age groups. When comparing the older (45-54 years old) aged participants to the younger aged group, there were some noticeable differences across measures. Specifically, these older drivers had higher values of TTC and higher values of cognitive workload (i.e., DRT response time and miss rate) as compared to the younger drivers. This is similar to a finding by Korber et al. (2016), who found that younger drivers ($\leq$ 28 years old) were more likely to adapt TTC values below 1 second as compared to older drivers ($\geq$ 60 years old).

Older drivers also exhibited different behaviors associated with secondary task engagement. While there were no differences associated with age on mean glance duration, 90$^{\text{th}}$ percentile glance duration, or percent long glances, older drivers had significantly longer mean total eyes-off-road time per task as compared to younger drivers. This older population also completed less tasks during the drives. These behaviors (i.e., less engagement with distracting tasks) are consistent with safer driving habits.

There were no effects associated with gender on lateral deviation, cognitive workload, eye glance behavior, or secondary task completion. However, males had significantly lower values of TTC as compared to females, which is generally indicative of more risky driving.

Chapter 7
DRIVER TRUST IN AUTOMATION

Drivers' trust in an autonomous vehicle system can predicate how they will interact with the system. Lee and See (J. D. Lee & See, 2004) posits that trust guides reliance in a system, where drivers will rely on a system they trust and reject a system they do not trust. This chapter characterizes drivers based on their trust in the lane keeping system and then evaluates the influences of trust on adaptations in their driving performance, cognitive workload, and eye glance patterns. The objective of this chapter is to address the third research aim of this dissertation:

- *Aim 3: How do drivers self-reported levels of trust impact behavioral adaptations?*

7.1 Analytical Methods

Hierarchical clustering using Ward's minimum variance method was applied to the responses from the trust in the lane keeping system questionnaire. This analysis was conducted on the standardized values of the 24 observations per participant, on all 30 treatment participants.

Repeated measures Analysis of Variances (ANOVAs) were then used to evaluate changes in driving performance, cognitive workload, and eye glance behavior across trust cluster groups and drives. Changes in lateral vehicle control (SDLP) was compared between drive 2 (before exposure) and drive 7 (after system was removed). This provided insight on carryover effects of exposure based on trust levels. Cognitive workload (i.e., TDRT response time and miss rate) was compared between drive 6 (final drive with automation) and drive 7 (after system was removed). This was used to understand changes in workload experienced (e.g., reliance on automation) between semi-automated and post semi-automated driving, based on trust levels. Changes in eye glance behavior between drives 2 and 7 were used to evaluate carryover effects on risk perception based on the trust.

7.2 Results

7.2.1 Cluster Analysis

The pseudo F statistic indicated that three cluster groups were appropriate for clustering the data. The final clustering appeared to group participants into a low trust, moderate trust, and high trust in the lane keeping system. The three trust cluster groups are hereafter refereed to as "low", "middle", and "high". Table 7.1 provides a list of the eight questions used in the trust questionnaire and the median score for each trust group by day.

Table 7.1: Trust Survey Questions with Responses by Trust Group

Trust Question	**Day**	**Group *(median)***		
1 (not at all) - 10 (extremely high)		**Low**	**Middle**	**High**
1. To what extent does the LK perform the task it was designed to do?	1	6	7	10
	2	7	9	9
	3	7.5	8	10
2. To what extent can the LK's behavior be predicted from moment to moment?	1	6	6	8
	2	6	8	9
	3	6.5	8	9
3. To what extent does the LK respond similarly to similar circumstances at different points in times?	1	6	6	9
	2	6	8	9
	3	6.5	9	9
4. What is your degree of faith that the LK will be able to cope with future driving situations?	1	4	7	9
	2	5.5	7	9
	3	6	8	9
5. What is your degree of trust in the LK to respond accurately?	1	4.5	6	9
	2	6	7	9
	3	6	8	9
6. What is your degree of self-confidence to manually intervene with the LK?	1	5.5	8	9
	2	6.5	8	9
	3	6	8	9
7. What is your overall degree of trust in the LK?	1	4	7	9
	2	5.5	7	9
	3	6	8	9
8. How confident do you feel about your previous trust ratings?	1	6	7	9
	2	6.5	8	9.5
	3	7.5	9	9
Average Trust Score *(not included in cluster analysis)*	1	5.5	6.8	8.8
	2	6.3	7.7	9.1
	3	6.5	8.3	9.0

An average trust score per day was computed for each participant based on their responses, which was calculated as the mean value of the eight questions per day per participant (i.e., $\frac{1}{8}\sum_{i=1}^{8} Q_i$). This average trust score was not used in the cluster analysis, however was computed to help provide a simple visualization of the overall trust in the lane keeping system for each participant (i.e., single dimensionality response for the day). These average trust scores by day for each participant, grouped by trust cluster, are plotted in Figure 7.1. The within group trust scores were relatively consistent across days for the low and high trust groups. The middle trust group tended to be more spread across mid-range scores on the first day, but seemed to learn to trust the lane keeping system by the third day (i.e., scores were higher in day 3 relative to day 1).

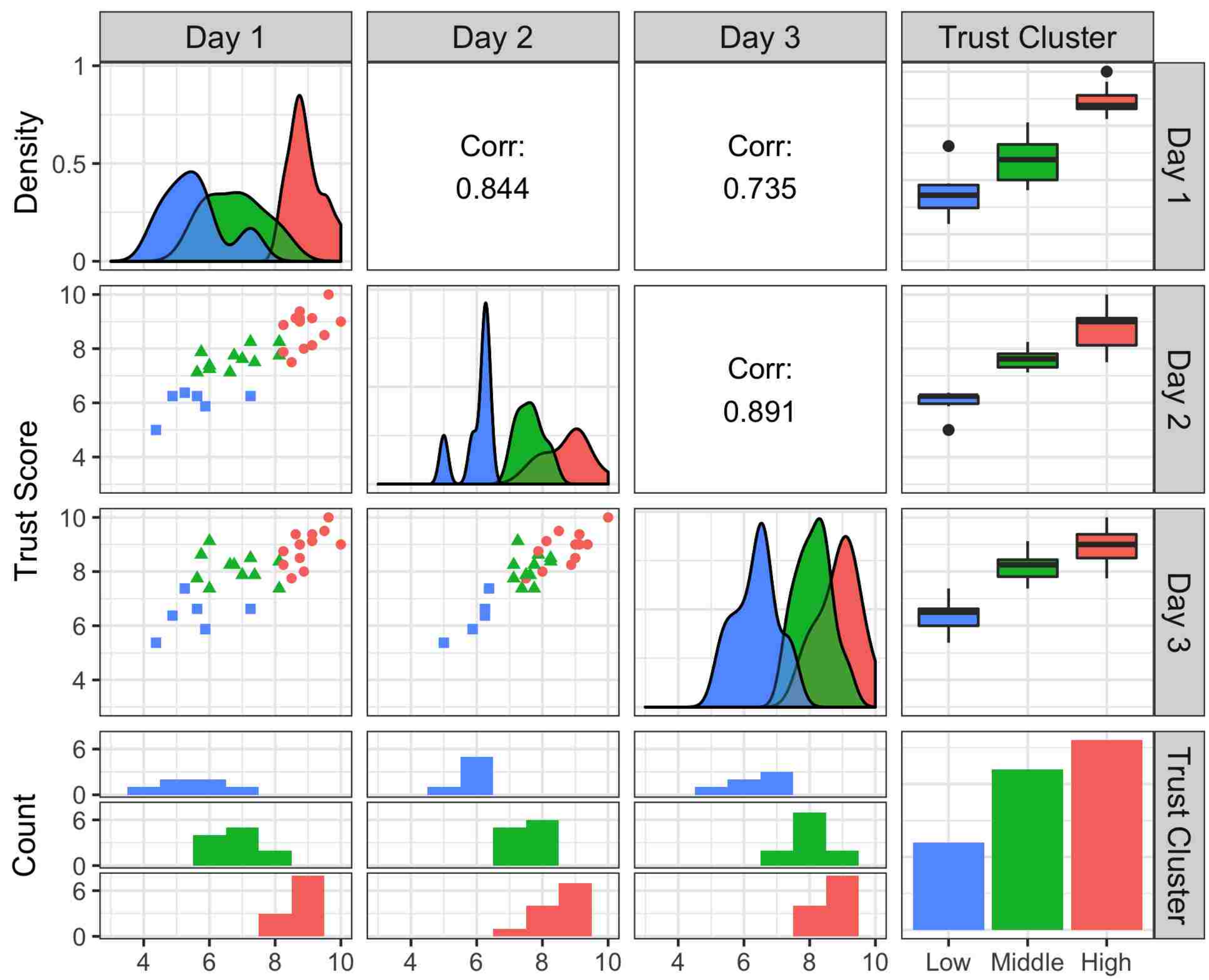

Figure 7.1: Summary of Average Trust Scores (1, low, to 10, high)

7.2.2 Trust Group Summary Statistics

There were 6 (4 male, 2 female) participants in the low trust group, 11 (4 male, 7 female) in the middle trust group, and 13 (7 male, 6 female) in the high trust group. A Fisher's exact test was used to evaluate differences in the proportions of gender by cluster, which indicated there was no significant difference among clusters ($p = 0.527$). The high trust group were the oldest on average (mean 40.2 years old, SD 8.90), followed by the lowest trust group (mean 38.2 years old, SD 10.19), and the middle trust group had the youngest average age (mean 35.7 years old, SD 8.90).

The middle and high trust groups had similar driving history profiles, see Table 7.2. The low trust group had the least prior experience with using a vehicle with lateral assistance compared to the other two groups. The low trust group also had proportionally less crashes in the past five years. Additionally, none of the drivers in the low trust group reported receiving a moving violation in the past five years, whereas the same could not be said about the middle and high trust groups.

Table 7.2: Trust Cluster Group Driving History

Trust Group	**Previous Experience with Lateral Assistance *n(%)***	**Moving Violations 5 years *n(%)***		**Crashes in 5 years *n(%)***	
		0	**1+**	**0**	**1+**
Low	1 (16.7)	6 (100)	0 (0.0)	5 (83.3)	1 (16.7)
Middle	3 (27.3)	5 (45.5)	6 (54.5)	8 (72.7)	3 (27.3)
High	3 (23.1)	6 (46.2)	7 (53.8)	10 (76.9)	3 (23.1)

7.2.3 Driving Performance

The Standard Deviation of Lateral Position (SDLP) before exposure (drive 2) and after withdrawal (drive 7) of the lane keeping system was examined, see Figure 7.2. The change in mean SDLP values were noticeably the lowest across all road segments for the low trust group as compared to the two higher trust groups.

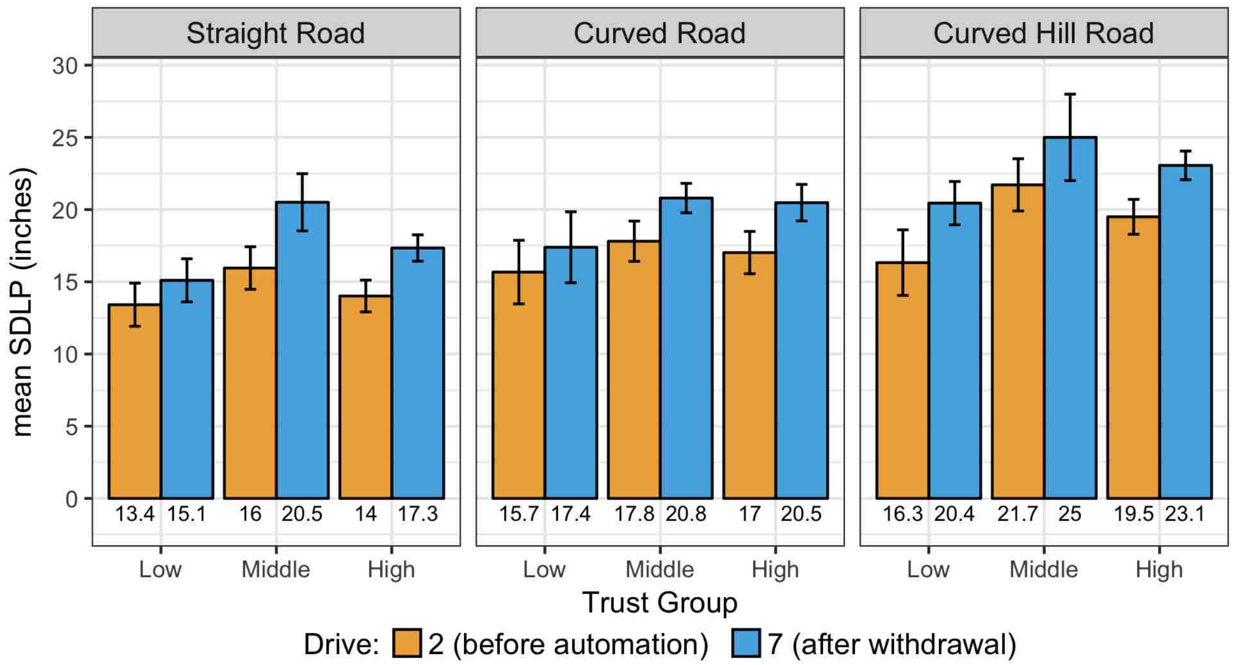

Figure 7.2: Observed Mean SDLP ± 1 SE by Trust Group for Before (Drive 2) and After (Drive 7) Automation Exposure

Based on the ANOVA, there was a significant difference in mean SDLP between drives ($F(1, 145) = 71.42, p < 0.001$) and road type ($F(2, 145) = 53.04, p < 0.001$). However there was no significant effect associated with cluster group ($F(2, 27) = 1.71, p = 0.200$) or the interaction of cluster group and drive ($F(2, 145) = 0.28, p = 0.754$). A Tukey HSD test indicated a significant difference between drive 2 (mean 17.11, SD 5.28) and drive 7 (mean 20.43, SD 5.86). There was also a difference between all three contrasts of road type, straight (mean 16.33, SD 4.97), curved (mean 18.51, SD 4.94), and curved hill (mean 21.46, SD 6.29).

7.2.4 Cognitive Workload

Cognitive workload was compared across cluster groups for the last drive with automation (drive 6) to the drive after withdrawal (drive 7), see Figure 7.3. The TDRT measures were adjusted to account for individual differences in reactions by subtracting each participants' respective baseline DRT measure (i.e., response time or miss rate), as measured in drive 1. There was a very slight difference in mean cognitive workload for the low trust group between when the lane keeping was engaged as compared to after it was removed;

while there was a noticeable increase in workload for the two higher trust groups after the system was removed.

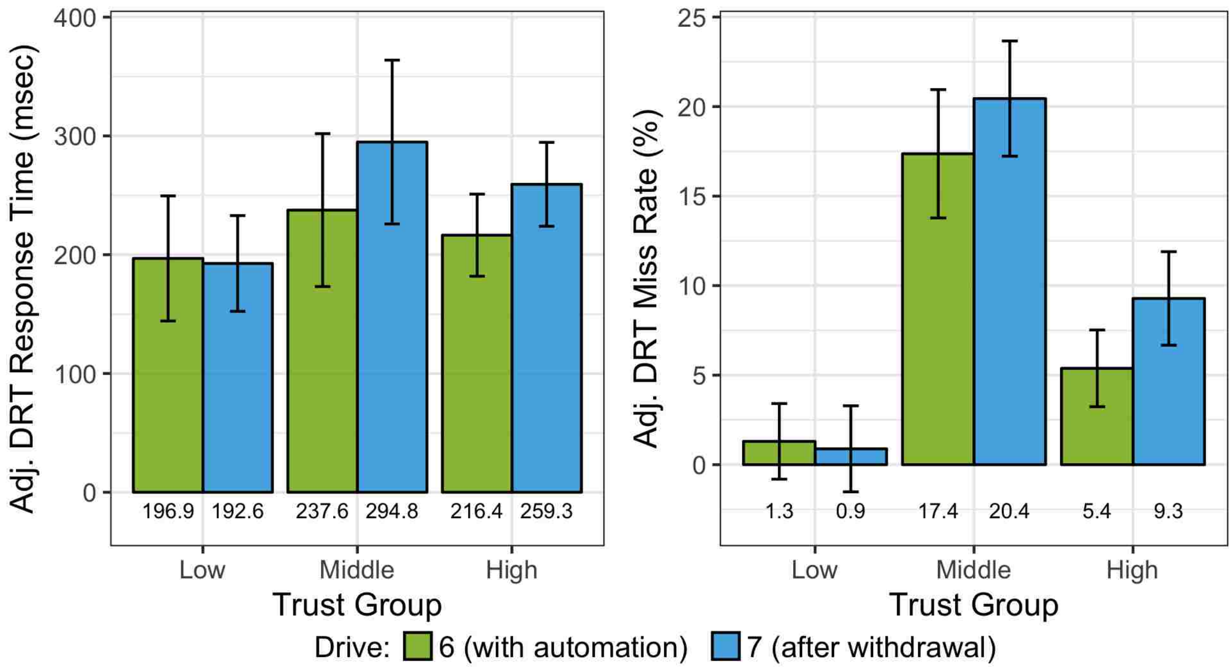

Figure 7.3: Mean DRT Response Times *(left)* and Miss Rates *(right)* $\pm$ 1 SE by Trust Group for During (Drive 6) and After (Drive 7) Automation Exposure

The ANOVA on adjusted DRT response times showed a significant difference between drives ($F(1, 27) = 8.55, p = 0.007$), but not between cluster groups ($F(2, 27) = 0.20, p = 0.823$) or the interaction of cluster and drive ($F(2, 27) = 1.43, p = 0.256$). The ANOVA on adjusted DRT miss rate indicated there was a difference between cluster groups ($F(2, 27) = 5.15, p = 0.012$), but not between drive ($F(1, 27) = 3.51, p = 0.071$) or the interaction of drive and cluster ($F(2, 27) = 0.61, p = 0.548$).

7.2.5 Eye Glance Behavior

There were four measures examined for eye glance behavior, coinciding with the four measures that were also evaluated in the previous chapter on behavioral adaptations:

1. *Mean glance duration* for each task type was computed based on the average of the three trials (i.e., repetitions) within the drive for each participant.

2. *90th percentile glance duration* was computed based on the 90$^{\text{th}}$ quantile of all off road glances per task type for each participant on the respective day.

3. *Proportion of long glances* was calculated for each task for each participant based on the total number of glances that exceeded 2.0 seconds divided by the total number of glances, and then averaged across the three repetitions for the participant.

4. *Total eyes-off-road time* for each participant by task was computed as the sum of the glances off road, and averaged across the three trials per task per participant.

These eye glance behaviors were aggregated by cluster group to compare glance durations and eyes-off-road time, see Figures 7.4 and 7.5.

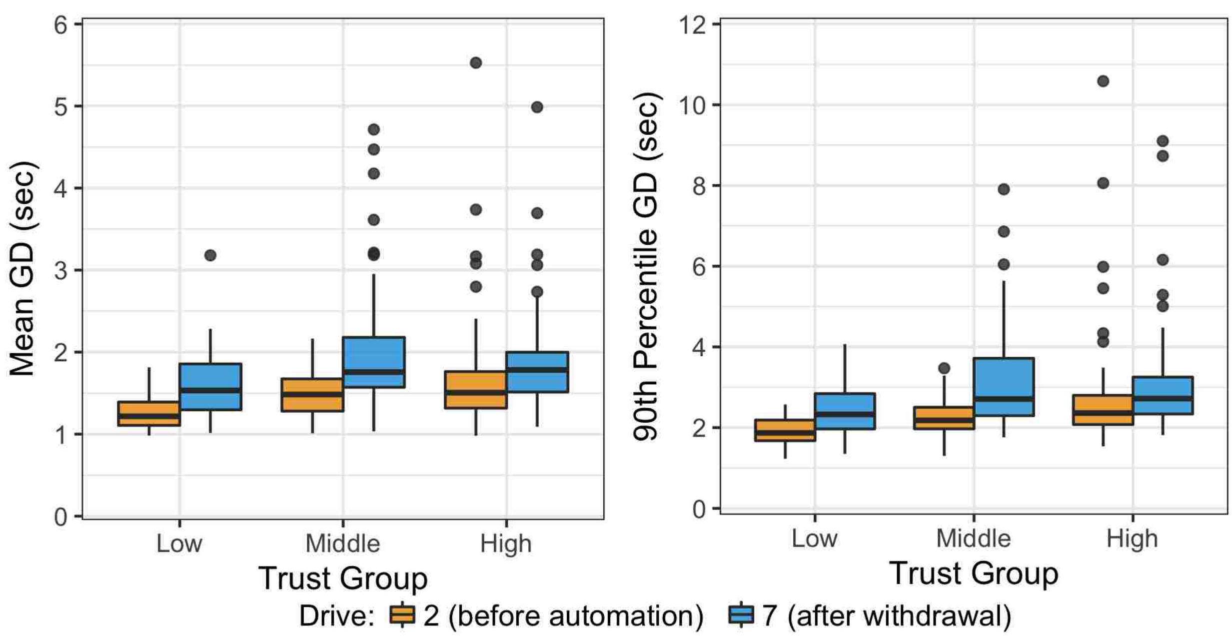

Figure 7.4: Eyes-Off-Road Glance Durations for Mean *(left)* and 90$^{\text{th}}$ Percentile *(right)* by Trust Group for Before (Drive 2) and After (Drive 7) Automation Exposure

The ANOVA on mean eye glance duration indicated a significant difference in means by drive ($F(1, 253) = 95.63, p < 0.001$), task type ($F(4, 253) = 7.24, p < 0.001$), and interaction of drive and cluster group ($F(2, 253) = 3.97, p = 0.020$). There was no significant difference between cluster groups ($F(2, 23) = 1.26, p = 0.301$), gender ($F(1, 23) = 0.59, p = 0.450$), or age groups ($F(2, 23) = 0.701, p = 0.506$). The interaction of day within each cluster group was particularly of interest and further evaluated using Tukey's HSD test. There was a significant difference ($p < 0.001$) between mean eye glance durations for the low trust group in drive 2 (mean 1.28, SD 0.23) and drive 7 (mean 1.63, SD 0.49), the middle group in drive 2 (mean 1.51, SD 0.29) and drive 7 (mean 2.03, SD 0.80), and the high trust group in drive 2 (mean 1.67, SD 0.70) and drive 7 (mean 1.89, SD 0.64).

The ANOVA on the 90th percentile eye glance duration similarly indicated a significant difference in means by drive ($F(1, 253) =, 88.58, p < 0.001$), task type ($F(4, 253) = 7.28, p < 0.001$), and the interaction of drive and cluster group $F(2, 253) = 6.75, p = 0.001$). There was no significant effect of cluster group ($F(2, 23) = 1.56, p = 0.232$), gender ($F(1, 23) = 0.60, p = 0.445$), or age group ($F(2, 23) = 0.81, p = 0.456$). The Tukey post-hoc test was also performed on the contrast of each drive and cluster; there was a significant difference ($p < 0.01$) between means for the low trust group in drive 2 (mean 1.94, SD 0.37) and drive 7 (mean 2.47, SD 0.72), middle trust group drive 2 (mean 2.26, SD 0.48) and drive 7 (mean 3.20, SD 1.33), and the high trust group drive 2 (mean 2.74, SD 1.44) and drive 7 (mean 3.05, SD 1.33).

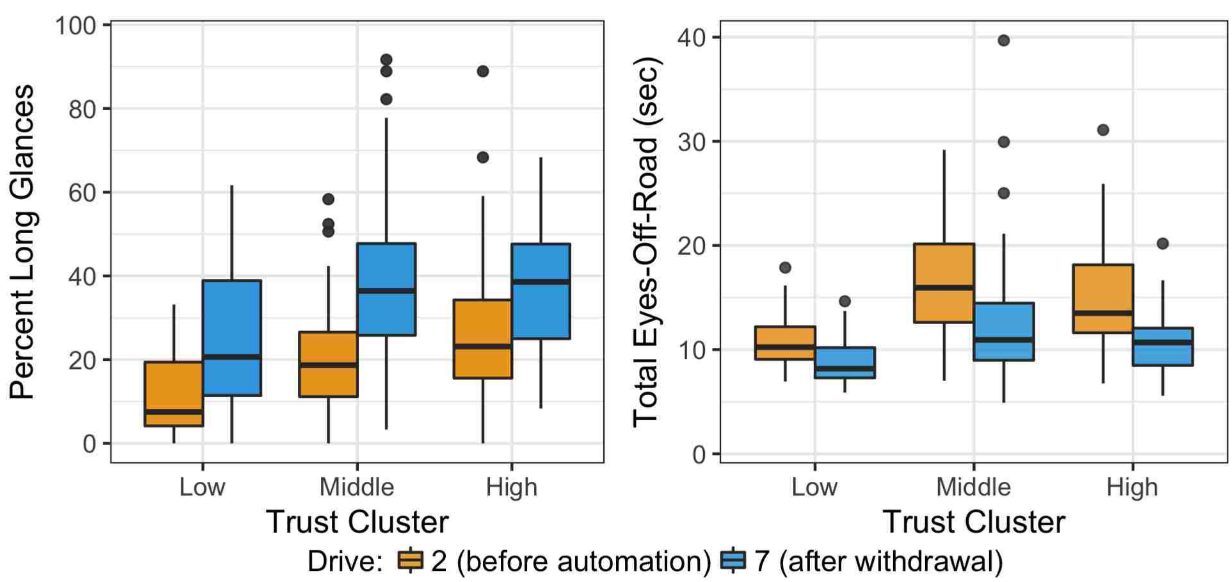

Figure 7.5: Percent Long Glances *(left)* and Total Eyes-Off-Road Time *(right)* by Trust Group for Before (Drive 2) and After (Drive 7) Automation Exposure

The ANOVA for percent long glances showed a significant difference for drive ($F(1, 253)$

$= 81.32, p < 0.001$) and task type ($F(4, 253) = 7.22, p < 0.001$). Unlike the previous two eye glance behavior metrics on glance duration, there was also a significant difference between cluster groups ($F(2, 253) = 3.75, p = 0.039$). There was no significant difference between gender ($F(1, 23) = 0.05, p = 0.818$), age group ($F(2, 23) = 0.55, p = 0.582$), or interaction of drive and cluster group ($F(2, 253) = 1.60, p = 0.205$). Based on the Tukey HSD results, there was a significant difference ($p = 0.029$) in percent long glances between the low trust cluster (mean 18.80, SD 15.26) and high trust cluster (mean 31.77, SD 17.17).

The ANOVA on the total eyes-off-road time also suggested there was a significance difference in means by drive ($F(1, 253) = 138.20, p < 0.001$), task type ($F(4, 253) = 30.98$, $p < 0.001$), and cluster group ($F(2, 23) = 4.61, p = 0.021$). There was no significant difference between the interaction of drive and cluster group ($F(2, 253) = 0.93, p = 0.396$), gender ($F(1, 23) = 0.034, p = 0.856$) or age group ($F(2, 23) = 3.07, p = 0.066$). The Tukey HSD test indicated there was only a significant difference between the low trust group (mean 9.96, SD 2.76) and high trust group (mean 14.60, SD 5.98) ($p = 0.033$), but not any of the other contrasts.

7.3 Discussion

This analysis revealed three cluster groups based on a multidimensional trust assessment: low trust, middle trust, and high trust. This approach was similar to a study by Xiong et al. (2012), who conducted a cluster analysis based on ACC use patterns and found that drivers could be grouped into high risk, moderate risk, and conservative. However, that study clustered on driving performance measures (i.e., ACC gap setting, speed setting, number of warnings, number of disengagements of the system), rather than subjective measures of trust.

The low trust group consistently remained low in their trust in the system and similarly, the high trust group remained steady in their high trust across all three days. The middle trust group had the largest shifts towards learning to trust the system. The change in median average trust scores were 1.0, 1.5, and 0.2 for the low, middle, high groups, respectively. Although the middle group had the largest median observed increase in trust, all three groups had various degrees of increases in trust for the lane keeping system. This is consistent with previous literature that has noted trust increases after using an autonomous system. Rudin-Brown and Parker (2004) reported that drivers trust in an ACC system increased after exposure to the system, despite even experiencing system failures. Merat and Lee (2012) also reported that trust in vehicle automation increases with increased familiarity, while acceptance does not increase with experience. Beggiato and Krems (2013) reported that drivers' trust and acceptance increased with experience for ACC systems that operated correctly, while drivers' trust and acceptance decreased for experiences with an ACC that provided incorrect descriptions and failures. This current study did not have any failures in the lane keeping system, and, as such, is consistent with the findings from these previous studies.

The three trust groups were also characterized by their demographics and driving history. The low trust group had the safest driving history (i.e., proportionally less crashes and no moving violations). The low trust group also had the least prior experience with lateral assistance systems (16.7% versus 27.3% for the middle and 23.1% for the high group). There were no significant differences in the gender distribution across the three groups, as determined using Fisher's exact test. Similarly, in a study by Xiong et al. (2012), there were no differences in genders between groups clustered on use patterns with ACC. The high trust group were the oldest on average, although the differences in age across the groups were not statistically significant. Previous literature has noted age related differences in use and acceptance of automation (Xiong & Boyle, 2012). Gold et al. (2015) found that older drivers self reported higher levels of trust in a highly automated vehicle as compared to younger drivers. Similar findings were observed in a survey study by Jenness et al. (2008), who showed that older drivers were more likely to increase following distance under manual driving as compared to ACC. They were also less likely to report problems or safety concerns as a result of using ACC. Ho et al. (2005) suggests that older individuals tend to have greater trust and reliance in automation due to deficits in cognitive abilities, increased workload, and decreased self-confidence in completing tasks manually.

All three trust groups had decreased driving performance (i.e., higher SDLP) in drive 7 as compared to drive 2, which represented the drives immediately before exposure to automation and immediately after withdrawing the automation. This concept of carryover effects of behavioral adaptations were also observed in a study by Skottke et al. (2014), where drivers decreased time headways for 10 km after being decoupled from highly automated driving. However, the exposure time to automation and duration measured for the carryover effect differed between these two studies. On average, the largest SDLP occurred on the curved hill road, followed by the curved flat road, while the straight segments had the least lateral variations. Although it was not statistically significant, drivers in the low trust group tended to have smaller degradations in lateral control (i.e., smaller increases in SDLP after automation withdrawal). This was particularly pronounced on the straight and curved roadway segments, while not as obvious on the curved hill roadway. Many participants appeared to cut the corner on the uphill portion of the curved hill road by driving over the lane center marker. This could have been a result of the absence of traffic in the on-coming lane and relatively sharp horizontal curvature, rather then poor performance. This geometric configuration and observed behavior could explain why there were noticeably larger variations on the curved hill segment across all trust groups.

Changes in cognitive workload (i.e., TDRT response time and miss rate) were also observed across groups and drives with and without automation. Drive 6 was the last drive with exposure to the lane keeping system, thus cognitive workload was evaluated during this drive to capture the effects of prolonged exposure and reliance on the system. Participants in the middle and high trust groups tended to have higher response times and miss rates (i.e., increased workload) in the drive immediately after withdrawing the system, as compared to during exposure. This suggests that these two higher trust groups experienced less workload while the system was engaged. Previous research supports this notion that automation leads to a decrease in workload as compared to manual driving. This has been shown through decreased heart rate (Carsten et al., 2012; DeWinter et al., 2014), lower NASA TLX scores (Dambock et al., 2013; Stanton & Young, 2005), lower stress using the Dundee Stress State Questionnaire (DSSQ) scale (Stanton & Young, 2005), and increased blink rate (Dambock et al., 2013). However, the low trust group had similar measures of cognitive workload in drives with automation and following withdrawal. This negligible change in workload suggests that drivers in the low trust group may not have been as reliant on the system during drives with automation. Overall, drivers in the low trust group had significantly lower miss rates as compared to the other two groups in both drives that were evaluated. One possible explanation for this difference in observed workload could be attributed to the confounding effects of engaging in a secondary task, where drivers in this study were encouraged to engage at their own pace during the drives.

Within all three trust groups, drive 7 (after withdrawal) had significantly longer mean and 90$^{\text{th}}$ percentile glance durations as compared to drive 2 (before exposure). The largest change in eye glance durations was observed in the middle trust group, which was the group characterized as learning to trust the automation. Specifically, the middle trust group had a mean increase of 0.52 seconds in mean glance duration and 0.94 second increase in 90$^{\text{th}}$ percentile glance duration from drive 2 to drive 7. There was also a noticeable, but not statistically significant, trend of increasing mean and 90$^{\text{th}}$ percentile glance durations for increasing trust groups (i.e., low trust had the shortest and high had the longest glances). A similar pattern was observed for percent long glances and total eyes-off-road time, however this trend was statistically significant. Specifically, the low trust group had the lowest percent long glances and total eyes-off-road time, while the high trust group had the highest. This, in conjunction with driving history trends discussed above, suggests that drivers who were more trusting of the automation, were also more risky drivers. In a similar methodology, Peng and Boyle (2015) used cluster analysis on longitudinal measures to group drivers into either high-risk or low-risk when driving while engaging with an in-vehicle device. The results from this current analysis are consistent with findings from Peng and Boyle (2015), who found that high risk drivers had longer maximum eyes-off-road glances and that this difference increased with time. Other studies have also evaluated the relationship between trust and eye glance behavior and reported comparable results. Korber et al. (2018) found that participants who self-reported higher trust in a conditionally automated drive spent more time with eyes-off-road to a non-driving related task as compared to those with lower trust. Similarly, Hergeth et al. (2016) evaluated monitoring frequency (i.e., eye glances) to the driving scene during highly automated driving, and found that drivers with higher self-reported trust in the automation were associated with decreased monitoring frequency.

In this analysis, drivers in the middle and high trust groups had similar driving behaviors and habits (e.g., crash frequency, eye glance behavior, cognitive workload). However, these groups differed predominately in their initial trust ratings (i.e., day 1 trust scores). By the third day, participants in the middle trust group appeared to have learned to trust the system and had trust ratings similar to the high trust participants. If this study had a longer duration, it is likely that these two trust groups would have converged into one group. Future applications should consider not only building appropriate trust, but also maintaining appropriate trust, as many drivers may also learn to over trust these systems over time.

Chapter 8
GENERAL CONCLUSIONS

This chapter provides an overall summary of the findings from this dissertation, the relevance and dissemination of these results, and future research and applications of the study methodology.

8.1 Overall Findings

The objective of this dissertation was to examine behavioral adaptations and changes in risk perception during exposure to and after removing an active lane keeping system, as well as examine the effect of trust on these adaptations. An active lane keeping system was used to capture adaptation to automation, while still requiring drivers to attend to the driving task. A longitudinal driving simulator study was conducted, providing drivers with approximately 40 minutes of baseline driving, 80 minutes of semi-automated driving, and 40 minutes of post-automated (manual) driving. A control group was exposed to approximately 160 minutes of manual driving, but otherwise identical study procedures, in order to provide a reference for time on task effects. Changes in driving performance and risk perception were measured using driving performance measures, cognitive workload, eye glance behavior, and quantifying secondary task engagement. Trust was quantified using a questionnaire targeted at capturing a multidimensional assessment of human-automation trust. The key findings from this dissertation are summarized below:

- *Lateral Vehicle Control*. Drivers in the treatment group tended to have an increase in lateral deviations (i.e., worse lateral control) after the lane keeping system was removed, relative to their performance before exposure. Meanwhile, drivers in the control group tended to have improved lateral control (i.e., lower SDLP) as the study progressed.

- *Longitudinal Vehicle Control.* On average, drivers increased their TTC under manual driving with the addition of IVIS tasks. However, a decrease in TTC was observed in the treatment group once the lane keeping system was engaged. This decreased TTC was comparable to their preferred TTC under manual control with no distracting tasks. Despite the fact that they were still engaging with the IVIS. A carryover effect on longitudinal control was observed after the lane keeping system was removed, such that drivers in the treatment group tended to keep their decreased TTC when they returned to manual driving while still engaging with the IVIS. There was no time on task effect observed for the control group, instead they kept similar values of TTC across all drives with secondary tasks.

- *Cognitive Workload.* There was a general decreasing trend in cognitive workload (DRT response time and miss rate) observed across all drivers as the study progressed. There was an additional decrease in workload observed specifically in the treatment participants for the fourth drive of exposure to the lane keeping system. However, once the lane keeping system was removed, drivers in the treatment group tended to have an increase in cognitive workload relative to their baseline measure.

- *Eye Glance Behavior.* Drivers in the control group did not have noticeable differences in their measures of eye glance behavior (mean glance durations, 90^{th} percentile glance durations, and percent long glances) as the study progressed. However, drivers in the treatment group appeared to have longer mean glance durations, longer 90^{th} percentile glance durations, and more long glances (i.e., over 2.0 seconds) off the road in the drive after removing the lane keeping system as compared to their eye glance behavior before exposure. Participants in both groups tended to have less total eyes-off-road time per task in drive 7 as compared to drive 2. This suggests that drivers were able to complete the tasks quicker after gaining experience with the tasks, and while the control participants did not change their glance durations, the treatment participants learned to complete the tasks by taking longer off road glances.

- *IVIS Task Engagement.* There was a time on task effect observed for secondary task completion, where drivers completed more tasks per drive as the study progressed.

However, there was an additional increase observed in the treatment group, where these participants appeared to complete even more tasks relative to manual drivers. The addition of the lane keeping system did not have any effect on task accuracy.

- *Trust: Behavioral Adaptations.* The smallest degradations in lateral vehicle control were observed in the low trust group after the lane keeping system was removed. Changes in cognitive workload suggest that drivers with higher trust in the lane keeping system had higher reliance on the system, as they tended to have increased cognitive workload after the system was removed, as compared to while the system was engaged. Meanwhile, participants in the low trust group tended to have similar measures of cognitive workload regardless of whether the lane keeping system was engaged or removed.

- *Trust: Risk Perception.* Drivers with higher self-reported trust in the lane keeping system were associated with riskier driving behaviors. Specifically, these high trust drivers self-reported more moving violations and more crashes within the past five years as compared to low trust drivers. Drivers in the high trust group also appeared to have a larger proportion of long glances off road (i.e., greater than 2.0 seconds) as compared to the lower trust group.

- *Gender Effects.* There were no statistical differences between males and females with regards to secondary task engagement, cognitive workload, eye glance behavior, or lateral vehicle control. However, males were associated with lower values of TTC, which is generally correlated with riskier driving.

- *Age Group Effects.* There were no statistical differences between the younger and middle aged groups across any of the variables measured. However, older participants, as compared to the younger age group, tended to complete less secondary tasks, have higher cognitive workload (DRT response time and miss rate), have longer total eyes-off-road time per task, and have higher values of TTC.

8.2 Theoretical Constructs

Behavioral adaptations pertain to changes in behavior as a result of some exposure or modification in a system, for example the lane keeping system used in this study. These adaptations can have a positive impact, such as leading to an increase in awareness. In this dissertation, positive effects of exposure to the lane keeping system were observed through decreased cognitive workload and decreased lateral vehicle variations while the system was engaged. However, behavioral adaptations can also have a negative impact, such as diverting attention away from a safety critical task. In this dissertation, this included reduced car following distances during exposure and increased eyes-off-road glances after removing the lane keeping system.

Behavioral adaptations that are not intended by the designer are an essential consideration in evaluating a systems effectiveness, particularly because they can undermine the systems objectives. In the transportation domain, this often means observing users (e.g., drivers, pedestrians, etc.) interacting with the system. This dissertation developed a driving simulator protocol that could capture a variety of behavioral adaptations, not only to a lane keeping system, but one that could be further applied to other internal or external vehicle systems. This variety of behavioral adaptations evaluated in this study included behaviors relating to driving performance (i.e., SDLP and TTC), risk perception (i.e., secondary task engagement and eyes-off-road durations), and cognitive workload. This dissertation balanced participants based on age and gender, but this approach could also be used to balance on various driver trait characteristics, to understand how these influence behavioral adaptations. Many systems, even beyond transportation, have shifted towards individualizing designs and experiences, thus it is important to continue to consider a collective evaluation of behavioral adaptations on an individual level.

Time is also an important construct, as behavioral adaptations may manifest differently during initial exposure (i.e., learning curve), habit formation after prolonged exposure, and after removing the system (i.e., carryover effects). Previous literature suggests that habit formation typically develops after about 66 days of routine (Lally et al., 2009). While this driving study only included approximately 160 hours of exposure across three separate days, it is still relevant that behavioral adaptations across a variety of measures were observed. Habit formation often presents as an asymptotic curve (Lally et al., 2009), and the trends observed in this study began to taper as one would expect. While habit formation was likely not captured in this study, the behavioral changes may be explained through malleable attentional resource theory. Under malleable attentional resource theory, attentional resources shrink as mental workload decreases (i.e., mental underload), which can explain the performance decrements observed under increased automation (Young & Stanton, 2002). The results from this dissertation support this notion, as drivers likely reduced their attentional capacity as a result of exposure to the lane keeping system, and were therefore mentally unprepared to regain full driving responsibilities after the lane keeping system was removed. As drivers were not required to engage their full attentional resources during exposure, this mental resource did not work as well when it was required after the system was removed. These behavioral changes that continue to occur throughout exposure emphasize the importance of training and design considerations to account for this iterative interaction between the driver and system.

8.3 Contributions

As autonomous vehicle systems become increasingly more prevalent, gaps in consistency across systems will continue to expand, as functionality and limitations often differ across manufacturers. This is further confounded by the relatively slow turn over of the vehicle fleet (i.e., 10-15 years per vehicle), thus leading to a mixed equipage fleet. Concerns relating to trust, misuse, disuse, and skill atrophy will also increase as the human operator continues to adapt to their changing role as a driver. This is relevant as it introduces the potential for negative transfer effects of learning, as drivers move from one [semi-]autonomous system to another, from a vehicle with automation to one with little to none, or instances when automation is not equipped to respond to a given situation. This unfamiliarity with system characteristics and misplaced trust may lead to misinterpretations and misapplications that may undermine safety.

Although the implementation of autonomous systems into the vehicle fleet has been long anticipated, there has been a recent influx in the availability of semi-autonomous systems. This sudden increase in market share has led to new concerns that have not yet been explored in the literature. One major limitation in the literature is that there is a general disregard for exploring active lane keeping systems alone. The majority of the research evaluates ACC and some studies have begun to explore ACC in conjunction with lane keeping assistance. However, there are several vehicle manufacturers that offer lane keeping systems that can be activated independently of other ADASs. Hence, there is a gap in knowledge pertaining to how drivers will interact with these systems; this was the premise for selecting a lane keeping system as the autonomous system intervention for this dissertation. There further exists a lack of knowledge relating to the unintended safety consequences that may occur due to prolonged exposure or by removing or modifying these autonomous systems. Therefore, this dissertation developed a methodology for measuring and quantifying multiple dimensions of behavioral adaptations in a controlled driving simulator setting. The findings from this study have been disseminated into three academic journal articles and presented at lectern sessions at three conferences.

8.3.1 Publications

The results relating to changes in driving performance (SDLP and TTC) and cognitive workload (DRT miss rate and response time) between between baseline, semi-automated, and post semi-automated driving conditions were aggregated to evaluate behavioral changes that directly effect the driving task. That is, behavioral adaptations that may indicate improper reliance on systems or that lead to skill atrophy. These findings have been submitted to *Human Factors* (Miller & Boyle, submitted a).

The results focusing on eye glance behavior and IVIS task engagement were used collectively to evaluate changes in driver attention allocation due to exposure and withdrawal of the lane keeping system. As drivers adapt to using automation, it is likely that they will learn to rely more on the automation in order to compensate for secondary task involvement. This becomes a concern as misplaced trust compels drivers to push these systems beyond their design limits. Consideration of this shift in attention allocation away from the driving task should be mediated by system feedback to minimize safety critical events. This manuscript is currently being prepared for submission to *Transportation Research Part F:*

Traffic Psychology and Behaviour (Miller & Boyle, in preparation).

Driver's trust in an autonomous system can provide insights on their interaction with the system, such as when or if they will choose to intervene. All measures of behavioral adaptations (i.e., driving performance, cognitive workload, and eye glance behavior) were aggregated to examine the relationship between self-reported trust, behavioral adaptations, and inherent risk propensity. These findings have been submitted to *IEEE Transactions on Intelligent Transportation Systems* (Miller & Boyle, submitted b).

A general overview of the experimental design and findings relating to behavioral adaptations and influences of trust was presented at the doctoral student research in transportation safety session at the *Transportation Research Board 97th Annual Meeting* (Miller & Boyle, 2018a). This session was a hybrid session, meaning that it included a 3 minute presentation followed by a poster session. At that session, this dissertation won the Best Doctoral Student Research Award from the Safety Data, Analysis and Evaluation Committee (ANB20) and from the Statistical Methods Committee (ABJ80).

A subset of 18 drivers (9 treatment, 9 control) were used in a deeper evaluation of IVIS task performance and cognitive workload (DRT miss rate and response time). In this analysis, changes in cognitive workload measures within easy and hard tasks were evaluated in an effort to understand how exposure and withdrawal of the lane keeping system may effect performance in different types of secondary tasks. These findings were presented and published in the *Proceedings of the Human Factors and Ergonomics Society 61st Annual Meeting* (Miller & Boyle, 2017). This paper received the HFES Surface Transportation Technical Group Best Student Paper Award.

Standard Deviation of Lateral Position is an important measure when considering lateral control. However, another application is to consider severity of the deviation, such as lane departure events. A subset of the data pertaining to one straight segment in drives 2 (before exposure) and drives 7 (after withdrawal) per participant was extracted. A model was fit with a first order autoregressive covariance structure on the time series data with a binary response variable indicating if it was a lane departure. A comparison between interpreting effects of exposure and withdrawal of a lane keeping system on SDLP versus lane departure events was presented at the *8th International Conference on Information*

Society and Technology (Miller & Boyle, 2018b).

8.4 Limitations

A study limitation was the time frame that data was collected (three days of driving in a one–week period). Given time and resources, a study that was conducted over one month or even one year may show larger effects in behavioral adaptations and correlates with trust. Behavioral adaptations were observed in this study, and these adaptations of drivers to autonomous vehicle systems may undermine the safety benefits that automation affords. However, it is also possible that the performance degradations observed in this study may also be capturing the learning curve to time sharing the various study tasks. Thus, a longer-term study could provide more insight on habit formation and trust due to exposure rather than time sharing effects.

Drivers in this study were told immediately prior to the drive that the lane keeping system would be either on or off. Hence, they may appear to maintain reasonable performance, but in the real world, if such feedback is not provided, their performance may actually worsen to an unacceptable threshold. In general, these findings are limited in terms of unexpected handover and there may be even more such concerns as automation within the vehicle increases. It is important to understand how these behavioral adaptations may manifest in order to design systems and training programs to minimize adverse adaptations.

There are trade-offs associated with conducting a study in a driving simulator versus on road. One of the benefits that the laboratory setting affords, is that independent variables can be controlled and manipulated. Thus, endogeneity becomes less of a problem and statistical models become more insightful on predicting associations. Additionally, drivers can be asked to do tasks while driving the simulator that may be considered unsafe to request study participants to complete on the road (e.g., secondary distracting tasks). Such tasks and experimental designs could make a study difficult to receive IRB approval. That being said, utility for safe driving is likely different for many drivers in a simulator as compared to on road. Although driving simulators are a commonly accepted method for transportation safety research, this dissertation tried to account for this potential limitation to real world application in the methods. Specifically, the majority of the analyzes were based on within-subject differences. Thus, a change in performance from one point in time to the next point in time provides insight on behavioral adaptations, focusing on the magnitude of change rather than quantifying performance at one point in time.

The performance degradations observed after withdrawing the automation suggests that drivers began to rely on the lane keeping system for support and experienced adverse effects when the system was removed. This study provided drivers with approximately 80 minutes of exposure to a lane keeping system (4 drives $\times$ 20 minutes each). Even in this relatively short time period, behavioral adaptations were observed. This dissertation also showed the importance of understanding trust based on individual user characteristics. By identifying the associations between trust and behavioral adaptations over time, vehicle systems and educational programs can be tailored for individual operator differences. This will help drivers be able to appropriately trust and engage with the automation.

8.5 Future Research

The results of this study suggest that behavioral adaptations occur as drivers gain experience to lane keeping assistance systems, and likely the broader range of [semi-]autonomous systems. This dissertation developed a methodology for measuring and quantifying behavioral adaptations. However, further manipulation of the driving demands, training, in-vehicle system designs, infrastructure design, or interaction with other autonomous systems could provide further insight on these behavioral adaptations.

Recall the figure presented in Chapter 2, on the human information processing model for takeover (see Figure 8.1). Future research should consider the implications of various contexts, perception limitations, cognitive precesses that effect decision making, and how this affects the drivers' action to takeover vehicle control.

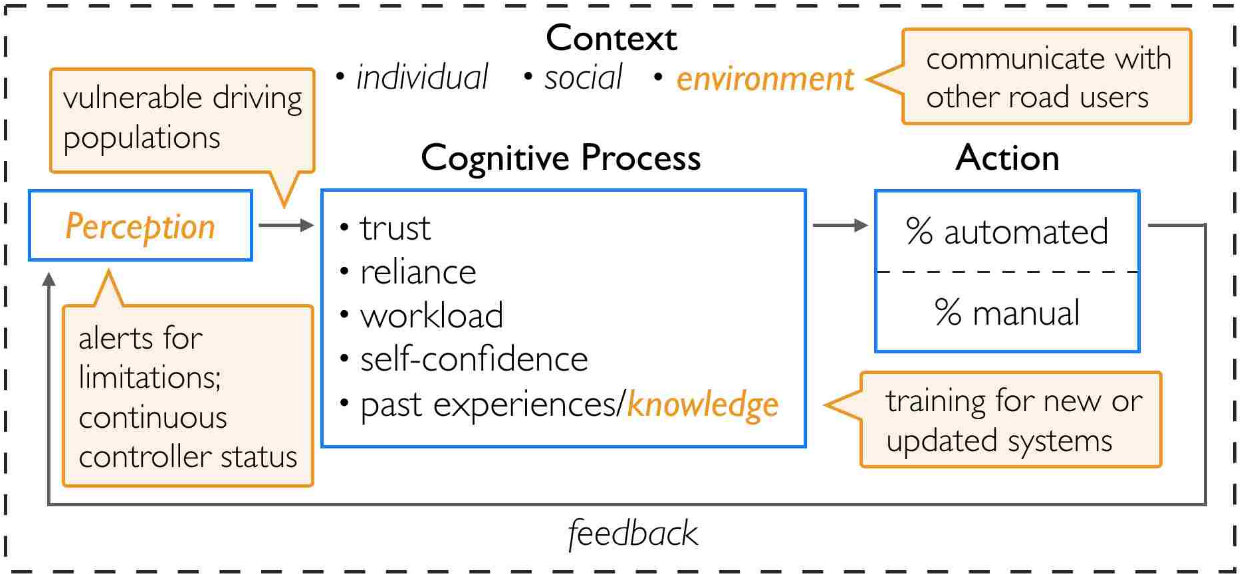

Figure 8.1: Future Research Relating to the Human Information Processing Model for Takeover, Adapted from J. D. Lee and See (2004) and J. D. Lee et al. (2017)

More specifically, future research should consider different environmental *contexts*, such as increased vehicle traffic, increased multi-modal traffic or the presence of various controlled and uncontrolled intersections. For example, vehicles in the adjacent oncoming lane would have provided a consequence for lane deviation and thus could effect these results. The role of human *perception* is also an important consideration, as autonomous systems that aid in perception limitations could diminish these adverse behavioral adaptations. For example, systems that provide alerts for system limitations across modalities based on the intensity of the situation, or systems that provide continuous controller status of the automations intent (e.g., planning to change lanes) could effect ones decision to takeover control or performance during handover. One of the benefits of automation is the aiding it can afford to vulnerable driving groups (e.g., inexperienced, elderly, or impaired drivers), and specific systems targeting these various groups should be considered in future research. Additionally, many vehicle manufacturers are shifting towards patching software updates remotely, via cloud connections (Azizian et al., 2017). As a result, systems can change with none to minimal training. Thus, future research should consider training methods for new or updated systems, as this likely has an impact on the *cognitive process* for deciding how or when to takeover control. It is also important to consider a broader range of autonomous vehicle systems, as it is likely different systems or different exposure time will effect the *action* of distributing control between manual or automated assistance. Additionally, future research should also consider sudden events (e.g., abrupt lead vehicle braking, work zones, etc.) and measure the reaction time of the driver responding to the sudden event.

REFERENCES

Azizian, M., Cherkaoui, S., & Hafid, A. S. (2017). Vehicle software updates distribution with SDN and cloud computing. *IEEE Communications Magazine*, *55*(8), 74-79.

Barnard, Y., & Lai, F. (2010). Spotting sheep in yorkshire: Using eye-tracking for studying situation awareness in a driving simulator. In D. deWaard, A. Axelsson, M. Berglund, B. Peters, & C. Weikert (Eds.), *Human factors: A system view of human, technology, and organisation* (p. 249-261). Maastricht, Netherlands: Shaker Publishing.

Beggiato, M., & Krems, J. F. (2013). The evolution of mental model, trust and acceptance of adaptive cruise control in relation to initial information. *Transportation Research Part F: Traffic Psychology and Behaviour*, *18*, 47-57.

Boyle, L. N., Lee, J. D., Peng, Y., Ghazizadeh, M., Wu, Y., Miller, E., & Jenness, J. (2013). *Text reading and text input assessment in support of the NHTSA visual-manual driver distraction guidelines* (Tech. Rep. No. DOT HS 811 820). Washington, DC: NHTSA.

Breyer, F., Blaschke, C., Farber, B., Freyer, J., & Limbacher, R. (2010). Negative behavioral adaptation to lane-keeping assistance systems. *IEEE Intelligent Transportation Systems Magazine*, *2*(2), 21–32.

Brookhuis, K. A., & deWaard, D. (2010). Monitoring drivers workload in driving simulators using physiological measures. *Accident Analysis and Prevention*, *42*, 898-903.

Brookhuis, K. A., deWaard, D., & Janssen, W. H. (2001). Behavioural impacts of advanced driver assistance systems-an overview. *European Journal of Transport and Infrastructure Research*, *1*(3), 245-253.

Carsten, O., Lai, F. C., Barnard, Y., Jamson, A. H., & Merat, N. (2012). Control task substitution in semiautomated driving. Does it matter what aspects are automated. *Human Factors*, *54*, 747-761.

Casner, S. M., Geven, R. W., Recker, M. P., & Schooler, J. W. (2014). The retention of manual flying skills in the automated cockpit. *Human Factors*, *56*(8), 1506-1516.

Chang, C. C., Boyle, L. N., Lee, J. D., & Jenness, J. (2017). Using tactile detection response tasks to assess in-vehicle voice control interactions. *Transportation Research Part F: Traffic Psychology and Behaviour*, *51*, 38-46.

Contreras, P., & Murtagh, F. (2016). Hierarchical clustering. In C. Hennig, M. Meila, & F. Murtagh (Eds.), *Handbook of cluster analysis* (p. 103-124). Boca Raton, FL: CRC Press.

Cooper, J. M., Castro, S. C., & Strayer, D. L. (2016). Extending the detection response task to simultaneously measure cognitive and visual task demands. In *Proceedings of the Human Factors and Ergonomics Society 60th Annual Meeting* (p. 1962-1966). Washington, DC, Sept 2016.

Dambock, D., Weißgerber, T., Kienle, M., & Bengler, K. (2013). Requirements for cooperative vehicle guidance. In *Proceedings of the 16th International IEEE Conference on Intelligent Transportation Systems* (p. 1656-1661). Hague, Netherlands, Oct 2013.

DeWinter, J. C. F., Happee, R., Martens, M. H., & Stanton, N. A. (2014). Effects of adaptive cruise control and highly automated driving on workload and situation awareness: A review of the empirical evidence. *Transportation Research Part F: Traffic Psychology and Behaviour*, *27*, 196-217.

Donmez, B., Boyle, L. N., & Lee, J. D. (2010). Differences in off-road glances: Effects on young drivers performance. *Journal of Transportation Engineering*, *136*(5), 403-409.

Eick, E. M., & Debus, G. (2005). Adaptation effects in an automated car-following scenario. In G. Underwood (Ed.), *Traffic and transport psychology. Theory and application* (p. 243-255). Amsterdam, Netherlands: Elsevier.

Fagnant, D. J., & Kockelman, K. (2015). Preparing a nation for autonomous vehicles: Opportunities, barriers and policy recommendations. *Transportation Research Part A: Policy and Practice*, *77*, 167-181.

Faure, V., Lobjois, R., & Benguigui, N. (2016). The effects of driving environment complexity and dual tasking on drivers mental workload and eye blink behavior. *Transportation Research Part F: Traffic Psychology and Behaviour*, *40*, 78-90.

FHWA. (2016). *Roadway departure safety.* Retrieved May 18, 2016, from http://safety.fhwa.dot.gov/roadway_dept

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2011). *Applied longitudinal analysis* (2nd ed.). Hoboken, NJ: Wiley.

Flemisch, F. O., Kelsch, J., Loper, C., Schieben, A., Schindler, J., & Heesen, M. (2008). Cooperative control and active interfaces for vehicle assistance and automation (Paper No. F2008-02-045). *Paper Presented at FISITA World Automotive Congress.* Munich, Germany, Sept 2008.

Funke, G., Matthews, G., Warm, J. S., & Emo, A. K. (2007). Vehicle automation: A remedy for driver stress. *Ergonomics*, *50*, 1302-1323.

Ghazizadeh, M., Lee, J. D., & Boyle, L. N. (2012). Extending the technology acceptance model to assess automation. *Cognition Technology and Work*, *14*, 39-49.

Gold, C., Korber, M., Hohenberger, C., Lechner, D., & Bengler, K. (2015). Trust in automation - before and after the experience of take-over scenarios in a highly automated vehicle. *Procedia Manufacturing*, *3*, 3025-3032.

Gordon, T., Sardar, H., Blower, D., Aust, M. L., Bareket, Z., Barnes, M., ... Theander, H. (2010). *Advanced crash avoidance technologies (ACAT) program - final report of the Volvo-Ford-UMTRI project: Safety impact methodology for lane departure warning - method development and estimation of benefits* (Tech. Rep. No. DOT HS 811 405). Washington, DC: NHTSA.

Gouy, M., Wiedemann, K., Stevens, A., Brunett, G., & Reed, N. (2014). Driving next to automated vehicle platoons: How do short time headways influence non-platoon drivers longitudinal control. *Transportation Research Part F: Traffic Psychology and Behaviour*, *27*, 264-273.

Guo, F., Klauer, S. G., Fang, Y., Hankey, J. M., Antin, J. F., Perez, M. A., ... Dingus, T. A. (2017). The effects of age on crash risk associated with driver distraction. *International Journal of Epidemiology*, *46*(1), 258-265.

Gustafsson, F. (2009). Automotive safety systems. *IEEE Signal Processing Magazine*, 32-47.

Hardin, J. W., & Hilbe, J. M. (2012). *Generalized estimating equations.* Boca Raton, FL: CRC Press.

Hart, S. G. (1986). *NASA task load index (TLX): v. 1.0* (Tech. Rep.). Moffett Field, CA: NASA Ames Research Center.

Heikkinen, J., Makinen, E., Lylykangas, J., Pakkanen, T., Vaananen-Vainio-Mattila, K., & Raisamo, R. (2013). Mobile devices as infotainment user interfaces in the car: Contextual study and design implications. In *Proceedings of the 15th International Conference on Human-Computer Interaction with Mobile Devices and Services* (p. 137-146). Munich, Germany, Aug 2013.

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*, *58*(3), 509-519.

Ho, G., Wheatley, D., & Scialfa, C. T. (2005). Age differences in trust and reliance of a medication management system. *Interacting with Computers*, *17*, 690-710.

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

Hubbard, A. E., Ahern, J., Fleischer, N. L., VanDerLaan, M., Lippman, S. A., Jewell, N., … Satariano, W. A. (2010). To GEE or not to GEE: Comparing population average and mixed models for estimating the associations between neighborhood risk factors and health. *Epidemiology*, *21*(4), 467-474.

ISO. (2016). 17488:2016. *Road vehicles - Transport information and control systems - Detection-Response Task (DRT) for assessing attentional effects of cognitive load in driving*, Geneva, CH: International Organization for Standardization.

Jamson, A. H., Merat, N., Carsten, O. M. J., & Lai, C. H. (2013). Behavioural changes in drivers experiencing highly-automated vehicle control in varying traffic conditions. *Transportation Research Part C: Emerging Technologies*, *30*, 116-125.

Jenness, J. W., Lerner, N. D., Mazor, S., Osberg, J. S., & Tefft, B. C. (2008). *Use of advanced in-vehicle technology by young and older early adopters. Survey results on adaptive cruise control systems* (Tech. Rep. No. DOT HS 810 917). Washington, DC: NHTSA.

Jian, J. Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive*

Ergonomics, *4*(1), 53-71.

Kaber, D. B., Liang, Y., Zhang, Y., Rogers, M. L., & Gangakhedkar, S. (2012). Driver performance effects of simultaneous visual and cognitive distraction and adaption behavior. *Transportation Research Part F: Traffic Psychology and Behaviour*, *15*, 491-501.

Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs, NJ: Prentice-Hall.

Kircher, K., Larsson, A., & Hultgren, J. A. (2014). Tactical driving behavior with different levels of automation. *IEEE Transactions on Intelligent Transportation Systems*, *15*(1), 158-167.

Klauer, S. G., Dingus, T. A., l. Neale, V., Sudweeks, J. D., & Ramsey, D. J. (2006). *The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data* (Tech. Rep. No. DOT HS 810 594). Washington, DC: NHTSA.

Korber, M., Baseler, E., & Bengler, K. (2018). Introduction matters: Manipulating trust in automation and reliance in automated driving. *Applied Ergonomics*, *66*, 18-31.

Korber, M., Gold, C., Lechner, D., & Bengler, K. (2016). The influence of age on the take-over of vehicle control in highly automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *39*, 19-32.

Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, *38*(4), 963-974.

Lally, P., vanJaarsveld, C. H. M., Potts, H. W. W., & Wardle, J. (2009). How are habits formed: Modelling habit formation in the real world. *European Journal of Social Psychology*, *40*(6), 998-1009.

Large, D. R., Burnett, G., Anyasodo, B., & Skrypchuk, L. (2016). Assessing cognitive demand during natural language interactions with a digital driving assistant. In *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (p. 67-74). Ann Arbor, MI, Oct 2016.

Lavasani, M., Jin, X., & Du, Y. (2015). Market penetration model for autonomous vehicles based on previous technology adoption experiences. *Journal of the Transportation Research Board*, *No. 2597*, 67-74.

Lee, J. D., Regan, M. A., & Young, K. L. (2008). Defining driver distraction. In M. A. Regan, J. D. Lee, & K. Young (Eds.), *Driver distraction: Theory, effects, and mitigation* (p. 31-40). Boca Raton, FL: CRC Press.

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

Lee, J. D., Wickens, C. D., Liu, Y., & Boyle, L. N. (2017). *Designing for people: An introduction to human factors engineering* (3rd ed.). Charleston, SC: CreateSpace.

Lee, Y., & Nelder, J. A. (2004). Conditional and marginal models: Another view. *Statistical Science*, *19*(2), 219-238.

Levitan, L., & Bloomfield, J. R. (1998). Human factors design of automated highway systems. In W. Barfield & T. A. Dingus (Eds.), *Human factors in intelligent transportation systems* (p. 131-163). Mahwah, NJ: Lawrence Erlbaum.

Liang, Y., & Lee, J. D. (2010). Combining cognitive and visual distraction: Less than the sum of its parts. *Accident Analysis and Prevention*, *42*, 881-890.

Litman, T. (2017). Autonomous vehicle implementation predictions: Implications for transport planning. *Victoria Transport Policy Institute*.

Ma, R., & Kaber, D. B. (2005). Situation awareness and workload in driving while using adaptive cruise control and a cell phone. *International Journal of Industrial Ergonomics*, *35*, 939-953.

Merat, N., Jamson, A. H., Lai, F. C. H., & Carsten, O. (2012). Highly automated driving, secondary task performance, and driver state. *Human Factors*, *54*(5), 762-771.

Merat, N., & Lee, J. D. (2012). Preface to the special section on human factors and automation in vehicles: Designing highly automated vehicles with the driver in mind. *Human Factors*, *54*(5), 681-686.

Miller, E. E., & Boyle, L. N. (2017). Driver adaptation to lane keeping assistance systems: Do drivers become less vigilant. In *Proceedings of the Human Factors and Ergonomics Society 61st Annual Meeting* (p. 1934-1938). Austin, TX, Oct 2017.

Miller, E. E., & Boyle, L. N. (2018a). Behavioral adaptations to autonomous systems: Evaluating intermediate and carryover effects. *Presented at Transportation Research Board 97th Annual Meeting*. Washington, DC, Jan 2018.

Miller, E. E., & Boyle, L. N. (2018b). Effects of exposure and withdrawal of a lane keeping system. *Paper Presented at the 8th International Conference on Information Society and Technology*. Kopaonik, Serbia, March 2018.

Miller, E. E., & Boyle, L. N. (in preparation). Exposure to automation: Adaptations in attention allocation. *Transportation Research Part F: Traffic Psychology and Behavior*.

Miller, E. E., & Boyle, L. N. (submitted a). Behavioral adaptations to lane keeping systems: Effects of exposure and withdrawal. *Human Factors*.

Miller, E. E., & Boyle, L. N. (submitted b). Trust in automation: Influences on driver behavioral adaptation. *IEEE Transactions on Intelligent Transportation Systems*.

Miller, E. E., Boyle, L. N., Jenness, J. W., & Lee, J. D. (2018). Voice control tasks on cognitive workload and driving performance: Implications of modality, difficulty, and duration. *Journal of the Transportation Research Board*, in press.

Moeckli, J., Brown, T., Dow, B., Boyle, L. N., Schwarz, C., & Xiong, H. (2015). *Evaluation of adaptive cruise control interface requirements on the national advanced driving simulator* (Tech. Rep. No. DOT HS 812 172). Washington, DC: NHTSA.

Muir, B. M. (1994). Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics*, *37*, 1905-1922.

Murray, D. M., Varnell, S. P., & Blitstein, J. L. (2004). Design and analysis of group-randomized trials: A review of recent methodological developments. *American Journal of Public Health*, *94*(3), 423-432.

NHTSA. (2011). *Crash warning interface metrics, Phase 2* (Tech. Rep. No. DOT HS 811 471). Washington, DC: National Highway Traffic Safety Administration.

NHTSA. (2013). Visual-manual NHTSA driver distraction guidelines for in-vehicle electronic devices. *Federal Register*, *78*(81), 24817-24890.

NHTSA. (2015). *Critical reasons for crashes investigated in the national motor vehicle crash causation survey* (Tech. Rep. No. DOT HS 812 115). Washington, DC: Author.

OECD. (1990). Behavioural adaptations to changes in the road transport system. *Road Transport Research Programme*, Paris, France: Organisation for Economic Co-operation and Development.

Olson, R. L., Hanowski, R. J., Hickman, J. S., & Bocanegra, J. (2009). *Driver distraction in commercial vehicle operations* (Tech. Rep. No. FMCSA RRR 09 042). Washington, DC: Federal Motor Carrier Safety Administration.

Parasuraman, R. (2000). Designing automation for human use: Empirical studies and quantitative models. *Ergonomics*, *43*, 931-951.

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

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.

Peng, Y., & Boyle, L. N. (2015). Drivers adaptive glance behavior to in-vehicle information systems. *Accident Analysis and Prevention*, *85*, 93-101.

Peng, Y., Boyle, L. N., & Hallmark, S. L. (2013). Drivers lane keeping ability with eyes off road: Insights from a naturalistic study. *Accident Analysis and Prevention*, *50*, 628-634.

Ranney, T. A., Baldwin, G. H. S., Smith, L. A., Mazzae, E. N., & Pierce, R. S. (2014). *Detection response task evaluation for driver distraction measurement application* (Tech. Rep. No. DOT HS 812 077). Washington, DC: NHTSA.

Robinson, R., Lerner, N., Jenness, J., Singer, J., Huey, R., Baldwin, C., . . . Monk, C. (2011). *Crash warning interface metrics: Task 2 report: Empirical studies of efforts of dvi variability* (Tech. Rep. No. DOT HS 811 470b). Washington, DC: NHTSA.

Rudin-Brown, C. M., & Noy, Y. I. (2002). Investigation of behavioral adaption to lane departure warnings. *Transportation Research Record: Journal of the Transportation Research Board*, *No. 1803*, 30-37.

Rudin-Brown, C. M., & Parker, H. A. (2004). Behavioural adaptation to Adaptive Cruise Control (ACC): Implications for preventive strategies. *Transportation Research Part F: Traffic Psychology and Behaviour*, *7*, 59-76.

SAE. (2015). SAE J2944_201506. *Operational Definitions of Driving Performance Measures and Statistics*, Detroit, MI: Society of Automotive Engineers.

SAE. (2016). SAE J3016_201609. *Taxonomy and Definitions for Terms Related to Driving*

Automation Systems for On-Road Motor Vehicles, Detroit, MI: Society of Automotive Engineers.

Sanchez, J., Fisk, A. D., & Rogers, W. A. (2004). Reliability and age-related effects on trust and reliance of a decision support aid. In *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting* (p. 586-589). New Orleans, LA, Sept 2004.

Seppelt, B. D., & Lee, J. D. (2007). Making Adaptive Cruise Control (ACC) limits visible. *International Journal of Human-Computer Studies*, *65*, 192-205.

Sivak, M., & Schoettle, B. (2015). *Influence of current nondrivers on the amount of travel and trip patterns with self-driving vehicles* (Tech. Rep. No. UMTRI-2015-39). Ann Arbor, Michigan: UMTRI.

Skottke, E., Debus, G., Wang, L., & Huestegge, L. (2014). Carryover effects of highly automated convoy driving on subsequent manual driving performance. *Human Factors*, *56*(7), 1272-1283.

Stanton, N. A., & Young, M. S. (2005). Driver behavior with adaptive cruise control. *Ergonomics*, *48*(10), 1294-1313.

Staubach, M. (2009). Factors correlated with traffic accidents as a basis for evaluating advanced driver assistance systems. *Accident Analysis and Prevention*, *41*, 1025-1033.

Stephens, T. S., Gonder, J., Chen, Y., Lin, Z., Liu, C., & Gohlke, D. (2015). *Estimated bounds and important factors for fuel use and consumer costs of connected and automated vehicles* (Tech. Rep. No. NREL/TP-5400-67216). Golden, CO: National Renewable Energy Laboratory.

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*, 218228.

Strayer, D. L., Turrill, J., Cooper, J. M., Coleman, J. R., Medeiros-Ward, N., & Biondi, F. (2015). Assessing cognitive distraction in the automobile. *Human Factors*, *57*(8), 1300-1324.

VanArem, B., VanDriel, C. J. G., & Visser, R. (2006). The impact of cooperative adaptive cruise control on traffic-flow characteristics. *IEEE Transactions on Intelligent*

Transportation Systems, *7*, 429-436.

Vashitz, G., Shinar, D., & Blum, Y. (2008). In-vehicle information systems to improve traffic safety in road tunnels. *Transportation Research Part F: Traffic Psychology and Behaviour*, *11*, 61-74.

Verberne, F. M. F., Ham, J., & Midden, C. J. H. (2012). Trust in smart systems: Sharing driving goals and giving information to increase trustworthiness and acceptability of smart systems in cars. *Human Factors*, *54*, 799-810.

Victor, T. W., Engstrom, J., & Harbluk, J. L. (2008). Distraction assessment methods based on visual behavior and event detection. In M. A. Regan, J. D. Lee, & K. Young (Eds.), *Driver distraction: Theory, effects, and mitigation* (p. 135-165). Boca Raton, FL: CRC Press.

Wadud, Z., MacKenzie, D., & Leiby, P. (2016). Help or hindrance? the travel, energy and carbon impacts of highly automated vehicles. *Transportation Research Part A: Policy and Practice*, *86*, 1-18.

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, *58*, 236244.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, *3*(2), 159-177.

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors*, *50*(3), 449-455.

Wochinger, K., Binder, S., Traube, E., Band, D., & Foley, J. (2008). *Human factors forum on advanced vehicle safety technologies-summary and proceedings* (Tech. Rep. No. DOT HS 810 918). Washington, DC: NHTSA.

Xiong, H., & Boyle, L. N. (2012). Drivers' adaptation to adaptive cruise control: Examination of automatic and manual braking. *IEEE Transactions on Intelligent Transportation Systems*, *13*(3), 1468-1473.

Xiong, H., Boyle, L. N., Moeckli, J., Dow, B. R., & Brown, T. L. (2012). Use patterns among early adopters of adaptive cruise control. *Human Factors*, *54*(5), 722-733.

Yang, S., Guo, J., & Sato, Y. (2015). Validity index mixed pseudo F for a new fuzzy cluster analysis algorithm. In *20th International Symposium on Artificial Life and*

Robotics (p. 506-509). Beppu, Japan, Jan 2015.

Young, M. S., & Stanton, N. A. (2002). Malleable attentional resources theory: A new explanation for the effects of mental underload on performance. *Human Factors*, *44*(3), 365-375.

Young, M. S., & Stanton, N. A. (2007). What's skill got to do with it? Vehicle automation and driver mental workload. *Ergonomics*, *50*(8), 1324-1339.

Zeger, S. L., Liang, K. Y., & Albert, P. S. (1988). Models for longitudinal data: A generalized estimating equation approach. *Biometrics*, *44*(4), 1049-1060.