Old Dominion University
ODU Digital Commons

Psychology Theses & Dissertations | Psychology

Fall 2019

The Effects of Vehicle Automation Level and Warning Type on Responses to Vehicle Hacking

Wyatt D. McManus
Old Dominion University, wyattmcmanus@gmail.com

Follow this and additional works at: https://digitalcommons.odu.edu/psychology_etds

Part of the Applied Behavior Analysis Commons, Human Factors Psychology Commons, and the Transportation Commons

Recommended Citation

McManus, Wyatt D.. "The Effects of Vehicle Automation Level and Warning Type on Responses to Vehicle Hacking" (2019). Master of Science (MS), Thesis, Psychology, Old Dominion University, DOI: 10.25777/pae6-8m73
https://digitalcommons.odu.edu/psychology_etds/335

This Thesis is brought to you for free and open access by the Psychology at ODU Digital Commons. It has been accepted for inclusion in Psychology Theses & Dissertations by an authorized administrator of ODU Digital Commons. For more information, please contact digitalcommons@odu.edu.

THE EFFECTS OF VEHICLE AUTOMATION LEVEL AND WARNING TYPE ON RESPONSES TO VEHICLE HACKING

by

Wyatt D. McManus
B.S May 2016, Purdue University

A Thesis Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

PSYCHOLOGY

OLD DOMINION UNIVERSITY
December 2019

Approved by:

Jing Chen (Director)

Yusuke Yamani (Member)

Violet Xu (Member)

ABSTRACT

THE EFFECTS OF VEHICLE AUTOMATION LEVEL AND WARNING TYPE ON RESPONSES TO VEHICLE HACKING

Wyatt D. McManus
Old Dominion University, 2019
Director: Dr. Jing Chen

Modern surface transportation vehicles often include different levels of automation. Higher automation levels have the potential to impact surface transportation in unforeseen ways. For example, connected vehicles with higher levels of automation are at a higher risk for hacking attempts, because automated driving assistance systems often rely on onboard sensors and internet connectivity. As the automation level of vehicle control rises, it is necessary to examine the effect of different levels of automation have on the driver-vehicle interactions. In addition, auditory warnings have been shown to effectively attract a driver's attention while performing a driving task, which is often visually demanding. The purpose of the current study was to examine the effect of level of automation and warning type on the driver's responses to vehicle hacking attempts. This goal was accomplished by manipulating level of automation (manual vs. automated) and warning type (non-semantic vs. semantic) and measuring drivers' responses to a vehicle hacking attempt using time to collision (TTC) values, maximum steering wheel angle, number of successful responses, and other measures of response. Our results revealed no significant effect of level of automation or warning type on TTC or successful response rate. However, there was a significant effect of level of automation on maximum steering wheel angle such that manual drivers had safer responses to the hacking attempt with smaller maximum steering wheel angles. In addition, an effect of warning type that approached significance was also found for maximum steering wheel angle such that participants who received a semantic warning had more severe and dangerous responses to the hacking attempt. The current results

suggest that level of automation and warning type may not significantly affect how quickly people respond to hacking scenarios when the warning is given in advance, but they may affect the quality of the driver’s response. These findings are important to future vehicle system design and subsequently the safety of future roadways.

Copyright, 2019, by Wyatt D. McManus, All Rights Reserved.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Jing Chen, for her constant guidance and assistance throughout the thesis process and my research assistants, Faye Wakefield, Petti Perkins, and Hannah Smith for their help with data collection. This research was performed under NSF grant #1760347. In addition, I would like to thank the lab manager and fellow graduate student, Scott Mishler, for his many discussions and advice as well as his assistance with learning the STISIM coding language. I would also like to thank my fellow graduate, Cat Turner, for her format proof reading and statistical assistance. Last, but certainly not least, I would like to thank my parents, Pamela, Jason, and Richard for their unfaltering support and guidance throughout my education, but particularly through the thesis process.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The implementation of automated systems in modern surface transportation vehicles is an ongoing process. Autonomous and semi-autonomous vehicles are potentially extremely helpful and safer than human operated vehicles (Nunes, Reimer, & Coughlin, 2018), but before autonomous vehicles can be used there are risks and concerns that need researched and understood. Research has been conducted examining the effect of different levels of vehicle automation on the driver's behavioral patterns (Gold, Damböck, Lorenz, & Bengler, 2013; Merat, Jamson, Lai, & Carsten, 2012; Shen & Neyens, 2017; Strand, Nilsson, Karlsson, & Nilsson, 2014). A vehicle hacking attempt, and the driver's response, is an example of a modern critical event that could change with the use of manual and automated driving technologies (Amoozadeh et al., 2015; Strand et al., 2014). Malicious hacking attempts could cause catastrophic accidents with the potential for injuries to those in and around the vehicle when the control of the vehicle is taken from the driver or automated driving system. To prevent these attempts or to remedy the potential consequences of the attacks, it is important to understand how drivers will respond to such attempts when using manual or automated driving.

Automated driving is different from manual driving in that the level of automation could affect the driver's perception of relevant information from their dynamic environment, referred to as situation awareness (SA), and subsequently, the driver's ability to respond to critical events such as cyberattacks. There has been little empirical research examining automated driving systems and manual driving in terms of which level of automation promotes safer responses to malicious hacking attempts. Amoozadeh et al. (2015) examined ways the vehicle could defend itself, such as lowering levels of automated control or potentially to manual control when an attack occurred. Although it is critical to design technology and automated systems to resist these attacks from a cyber-security perspective, it is also important to investigate how human drivers respond to hacking attempts at different levels of automation. In addition to level of automation, providing an auditory warning is one potential way to help drivers respond to critical events such as cyberattacks. Both sematic and non-semantic warnings have been used in different situations, with semantic warnings often containing additional information compared to non-semantic warnings (Chang, Lin, Fung, Hwang, & Doong, 2008; Hellier, Edworthy, Weedon, Walters, & Adams, 2002; McKeown & Isherwood, 2007). In the context of warnings about vehicle hacking attempts, the additional information provided by semantic warnings may improve responses to the hacking attempts that are relatively rare. The current study aimed to empirically explore the effect of level of automation and warning type on the operator's responses to vehicle hacking attempts in a driving simulator. The outcome of this experiment was expected to inform the design of future vehicle systems that ensure efficient prevention of automated vehicle (AV) hacking attempts.

Levels of Automation

The degree to which a human-machine system is referred to as automated depends on how much control of the system is allocated to the automation or to the human. In the current context, level of automation refers to the use of different automated and manual systems for operating the vehicle and the amount of vehicle control allocated to the human. There is a theoretical spectrum of system control that, when simplified, vary from complete human control to complete automation control (Flemisch, Kelsch, Loper, Shieben, & Schindler, 2008). The spectrum of system control varies widely between manual and fully automated control with many of the current driver assistance technologies falling somewhere near the middle of the spectrum. According to SAE (2018), autonomous vehicles can be categorized into six levels based on degree of automation. Level 0 has no driving automation and complete human control. As the level rises, more automation is implemented until Level 5, full system automation with no driver control required, is reached. For example, the introduction of automation to create more advanced driver assistance systems (ADAS) has created products that allow for automated control of different aspects of the vehicle system relevant to vehicle driving and control. Vehicles with one or more of these systems could serve as an example of a semi-autonomous system, for which control over the human-machine system is shared between the human and the automation.

Level 3 conditional automated vehicles are capable of controlling the lateral (lane keeping) and longitudinal (speed and braking) aspects of driving and only require the driver to be prepared as fallback after the vehicle issues a warning or request (Petermeijer, Doubek, & de Winter, 2017; SAE, 2018). Vehicles that are capable of level 3 conditional automated driving are designed to operate within the functional limits of the vehicle system (Petermeijer et al., 2017). When the automation is no longer capable of controlling one or more aspects of vehicle control or it leaves its operational domain, the driver is warned, and a takeover request is issued (Gold et al., 2013; Petermeijer et al., 2017). The human’s response to takeover requests could change based upon the level of automation the vehicle is operating at (Carsten, Lai, Barnard, Jamson, and Merat, 2012; de Winter, Happee, Martens, & Stanton, 2014). The current experiment compared human responses to a critical event when using between Level 2 and 3 automation and Level 0 manual driving.

For vehicles with conditional automated assistance technologies, monitoring the system’s performance is an additional task allocated to the driver (Petermeijer et al., 2017). The use of automated systems as support for the driving task may decrease the attentional resources required for the driving task and as such the human may decide to allocate their attentional resources to secondary tasks. The human operator's attentional resource allocation affects the operator's ability to monitor and cooperate with the automated system (Yamani & Horrey, 2018). For instance, Carsten, Lai, Barnard, Jamson, and Merat (2012) found that as higher levels of automation were used, participants became more likely to engage in a secondary task such as watching a DVD player. Engagement in a secondary task could lower the driver's ability to acquire appropriate SA for the driving task and inhibit their ability to correctly respond to critical events (Merat et al., 2012). Alternatively, if the system is not performing to the human's satisfaction, they may decide to allocate more attentional resources to monitoring the system or taking responsibilities back from the automation.

In summary, as shown by the currently expanding market of ADAS, surface transportation is moving toward fully automated driving. Despite this trend, it is critical to understand whether higher automation levels are more beneficial in terms of safety and SA compared to manual driving. This study compared drivers' responses in a manual driving mode and an automated driving mode when faced with a novel complex critical event (i.e., a potential cyber-attack).

Level of Automation and Situation Awareness

The level of automation used has the potential to affect the driver's ability to monitor for and appropriately respond to critical events as a result of differing levels of situation awareness (SA). Situation awareness refers to "a person's perception of relevant information from their dynamic environment within a given time and space, the comprehension of its meaning, and projection of their status in the near future" (Endsley & Garland, 2000, p. 3). Despite the intention of automation, many automated systems still require the human as a critical part of the system (SAE, 2018). In many automated systems, the human is required to monitor the system for failures and monitor the environment for scenarios that are outside of the automated systems capabilities (Endsley, 1996). However, the introduction of automation can negatively impact the operator’s cognitive processes leading to inaccurate or slow responses to novel critical events (Endsley & Garland, 2000). Higher levels of automation have been found to lower the operator’s SA, and consequently, result in detriments to the operator’s potential response to novel critical events (Endsley & Garland, 2000). Despite this, in the context of surface transportation, higher levels of vehicle automation are often considered desirable. Highly automated driving systems may eventually lead to increased driving safety and reliability by reducing the demands on the driver. Relieving the user of the driving task, in turn, could lead to higher levels of comfort for the driver (Endsley & Garland, 2000; Gold et al., 2013).

A central issue stemming from the use of automation in human-machine systems is the human-out-of-the-loop performance problem (Endsley, 1996; Kessel & Wickens, 1982). When humans are monitoring automation, they are often found to be slower at detecting issues that require them to intervene in the system potentially due to lower SA levels (Kaber & Endsley, 2004). Additionally, once the issue is detected, extra time is needed for the humans to become situationally aware enough to determine the state of the human-machine system and decide the appropriate response (Endsley, 1996). The level of control that the human has over the human-machine system could dictate the way the human responds to incidents that occur during the use of the system (de Winter, Happee, Martens, & Stanton, 2014). For instance, if the human is not at all engaged in controlling the system, then they may be less engaged in and attentive to what the system is doing as a result of their attention being allocated elsewhere.

The increased use of assistance technologies with automation in driving, and the subsequent rise in level of automation in vehicles, could lead to the operator allocating their attention to other non-driving related tasks (Carsten et al., 2012). According to Endsley (1995), if fewer attentional resources are allocated to the driving task then the driver may not achieve higher levels of SA. The driver's SA levels can be negatively impacted by automation which may lead to poor responses to novel critical events (Endsley & Garland, 2000). Until automation is completely self-sufficient, there will be instances where human input is required because the automation assistant is incapable of handling some aspect of the situation it is faced with. When a system inadequacy happens, input from the driver is often required and a takeover request is given to the driver. The amount of time that it takes for the driver to successfully resume control of the vehicle is thought to depend upon their SA and identifying information about their environment (Endsley, 1996). After the driver has gathered sufficient information about the current situation then they can decide how to resume control of the vehicle. This process is referred to as a takeover, and despite the positive aspects of automation in surface transportation, takeover and take-over requests take time for the driver to understand and respond to (Gold, Dambock, Lorenz, & Bengler, 2013).

The system's level of automation could affect the operator's SA level and subsequently affect the operator's performance. For example, Kaber and Endsley (2004) examined how levels of automation and adaptive automation characteristics affected the participant's SA, performance, and workload in a non-driving related context. The researchers had the participants do a simulation and gauge monitoring task. The experiment found that the primary task performance (i.e. the gauge monitoring task performance) was more closely associated with SA than with mental workload. The authors found that level of automation was the strongest predictor of primary task performance and the operator's SA level (Kaber & Endsley, 2004). Furthermore, this change in SA due to change in level of automation is thought to result from three major mechanisms (Endsley & Kiris, 1995). The first mechanism is the change in vigilance and complacency that occurs when humans monitor automation. The second mechanism is the human's assumed role change from active to passive in regard to vehicle control. The last mechanism is the change in quality or form of the feedback provided to the operator. To examine the effect of automation implementation on human performance, Wickens and Kessel (1979, 1981) performed experiments that demonstrated longer recovery times and poorer response accuracy for operators who had automation controlling the primary task before the critical event occurred. These studies showed the deficit in SA and subsequent responses caused by the implementation of automation in the primary task. For tasks such as driving, the negative effects automation has on human responses to critical events could lead to increased accidents and less safe roads.

The relationship between SA and level of automation is important to understanding responses to critical events occurring at different levels of automations in surface transportation. Parasuraman and Manzey's (2010) model of SA posits that SA is often lost as a result of complacency potential and attention bias during the processing of information about their situation. Both complacency potential and attentional bias are influenced by the level of automation that is in use (Strand et al., 2014). Strand and colleagues (2014) found that participants in the semi-automated vehicle were less likely to fail to take over control of the vehicle in time to prevent an accident when compared to the highly automated condition. This finding suggests that higher levels of automation can lead to degraded performance in the driving task potentially due to lower SA at higher levels of automation. Humans are poor monitors of automation and this could lead to a variety of safety and efficiency problem in the surface transportation domain as the field moves toward higher levels of automation (Strand et al., 2014). The driver's ability to respond to a critical event such as an automation failure is dependent on their level of SA at the time of the event and is important to ensure the safety of those in or around the vehicle. A driver's failure to notice and respond to the automation failure as a result of being "out-of-the-loop" at higher levels of automation, could lead to accidents and lower vehicle safety.

To gather a deeper understanding of how level of automation empirically affects human responses as a result of lowered SA, comparisons at different levels of automation need to be performed. Whereas Strand and colleagues (2014) compared semi-automated and highly automated driving, other research compared manual driving to driving with automation controlling the vehicle. As an example, Shen and Neyens (2017) compared conditional automated driving, with lane keeping and adaptive cruise control (conditional automated driving), to manual driving. The authors found that participants in the manual driving group responded to the system failure (critical event) significantly faster than participants in the automated conditions when looking at reaction time to the critical event as the dependent variable. Shen and Neyens suggest that this difference is a result of lower SA as the driver is further out-of-the-loop in the automated driving group. These findings further suggest there is potential for higher levels of automation to prevent the human from quickly and correctly responding to system failures. The relationship between levels of automation and successful response to critical events is still unclear, but a stronger understanding could lead to safer transition of vehicle control in the future.

The negative effect of higher automation on SA is an area of continued interest because of the continued push towards full automation in the driving domain. In an attempt to understand the relationship, Gold and colleagues (2013) compared automated driving to manual driving and reported higher levels of SA for the manual driving group. The authors posit this increase in SA, due to manual driving, led to approximately half as long reaction times to the critical event than those in the automated condition. The researchers also found that when the driver had less time to respond to the event, as a result of lower SA, they were more likely to use the brake than to steer with the wheel in response to the event. The authors posit that the use of the brake in low SA scenarios was to give the participant more time to analyze their dynamic environment and come to an appropriate decision. Gold et al.'s study suggests that differences in reaction time and response method are the result of differing levels of SA due to the level of automation present (Gold et al., 2013). Further research into the effect of levels of automation on SA and subsequently on driving performance is needed. The current study attempted to further understand how higher levels of automation affect SA and responses to novel critical events.

While some studies have found that higher levels of automation lead to lower levels of SA, other studies have found unclear results. For instance, Merat, Jamson, Lai, and Carsten (2012) examined whether the level of vehicle automation and the presence of a secondary task affected the participants driving performance level. The authors compared two levels of automation (manual and automated) and examined how the driver responded to either a critical incident (e.g., system failure) or no critical incident. The results indicated that the average driving speed was lower for automated driving portions of the experiment when compared to the manual driving portion. When a critical incident was present, average speeds were slower, and were not affected by the presence of a secondary task. Merat et al. found that drivers responded to the “critical incident” the same regardless of driving method (manual vs. automated) when no secondary task was present. However, the worst driver performance was obtained when the participant was distracted by the secondary task in the automated driving mode. It follows that this is the result of their attention being allocated elsewhere when the automation system needed them to resume control. The results of Merat et al.’s research help to provide information about how the vehicle level of automation affects driver SA, performance, and response to critical events. Specifically, the results suggest that higher levels of vehicle automation may have negative implications for driving safety.

Research in the driving domain that examined level of automation, SA, and driver’s responses to critical events suggests that higher automation levels may lead to lower SA and slower responses to critical events. Many of the experiments were conducted with repeated critical events such as a car stopping in front of the participants. Endsley and Garland (2000) suggested that automation could lead to slower responses to novel critical events. With possibility of vehicle hacking due to increased level of automation, hacking attempts can be the novel critical events that may solicit responses of human drivers that are distinct from the repeated critical events studied before. The current experiment explored the effect the relationship between automation level and SA has on human responses to novel hacking attempts.

Level of Automation and Cyberattacks

Autonomous and semi-autonomous vehicles, though helpful and potentially safer than human operated vehicles, have risks and concerns that need to be addressed. Research into the public concerns about driving with autonomous systems (Bazilinskyy, Kyriakidis, & de Winter, 2015; Kyriakidis, Happee, & de Winter, 2015; Schoettle & Sivak, 2014) has provided researchers with issues that need to be understood and accounted for as surface transportation continues to head towards higher levels of automation. One potential concern for the future of autonomous driving is the potential for vehicles with autonomous systems to be the subject of cyberattacks (Amoozadeh et al., 2015). A cyberattack implies deliberate actions “to alter, disrupt, deceive, degrade, or destroy adversary computer systems or networks or the information and (or) programs resident in or transiting these systems or networks” (National Research Council, 2009, p. 19).

As vehicles and related technologies continue to advance toward higher levels of automation, they will become increasingly at risk for vehicle hacking (Amoozadeh et al., 2015). Many of the systems that help a vehicle at higher levels of automation rely heavily on onboard sensors and internet connectivity in order to make real time vehicle maneuvering decisions. The connectivity and reliance on technology makes automated vehicles more vulnerable to cyberattacks. Amoozadeh et al.’s simulation showed that insider cyberattacks caused significant instability in the vehicle’s automated systems. Specifically, Amoozadeh et al. examined ways the vehicle could defend itself. One tactic highlighted was downgrading to lower levels of automated control or to manual control when a cyberattack occurred. This tactic of downgrading to lower automation levels is limited by the level of automation the vehicle is at when the attack occurs. It is critical to design technology and automated systems to resist cyberattacks, but it is also important to investigate how human drivers respond to hacking attempts at different levels of automation. Further understanding of how levels of automation affect human’s responses will allow researchers to look at whether humans are capable of making the necessary decisions to prevent a cyberattack from negatively affecting the vehicle.

There have been multiple indicators of the increasing potential for vehicle hacking attempts and their potential negative effects on driver safety. A public service announcement by the Federal Bureau of Investigation (FBI) in 2016 presented a step by step "How To" on what to do if you suspect your vehicle is the victim of a cyberattack, further suggesting the greater prevalence this issue may take in the near future (Federal Bureau of Investigation, 2016). Additionally, multiple surveys have found that respondents are concerned about the hacking of autonomous vehicles (Bansal, Kockelman, & Singh, 2016; Casley, Jardim, & Quartulli, 2013; Menon, 2015; and Schoettle et al., 2014).

In our prior study, 537 people recruited from Amazon Mechanical Turk (MTurk) were surveyed on their concerns with automated vehicle use and we found that less than 8% of the responses contained a hacking concern (McManus, Chen, & Mishler, 2018). The difference in responses shown in McManus et al. (2018) could be the result of the way the question was asked. The prior studies (Bazilinskyy et al., 2015; Kyriakidis et al., 2015) prompted the respondents about the possibility of autonomous vehicle hacking, whereas our pilot study used a free-response question that did not explicitly mention hacking. This difference as a result of how the question is asked, suggests that when the user is prompted about vehicle hacking, they view it as a serious problem, but much of the general public is unaware of the potential for vehicle cyberattacks without being informed of it. Research that examines the effect of automation level on novel vehicle hacking responses may provide detailed information about the effect automation level has on driving related cognitive processes.

While not directly related to vehicle hacking, cyber security research suggests that detecting a cyberattack is influenced by the amount of knowledge an individual has about cyber security threat detection. Ben-Asher and Gonzalez (2015) found that individuals with more knowledge about cyber security threat detection were better at detecting malicious attacks. Unfortunately, few people are aware of the potential for automated vehicle hacking without first being informed of the risk (McManus et al., 2018). If the findings on cyber security are shown to be similar in relation to vehicle hacking, then it could make a compelling case for educating the public about the potential for vehicle hacking and its effect on the driver's SA and responses. The information gained from research on the effect of levels of automation on SA during novel event responses could lead to the safer future implementation of automation in the surface transportation domain. The current experiment further explored the human aspect of vehicle hacking and will help to provide more insight into the understudied area of human responses to vehicle hacking at different levels of automation.

Although little empirical research pertaining to driver/passenger hacking responses has been performed in the surface transportation domain, some research has been conducted in the field of aviation. The field of aviation was influential in the creation of human factors, and human factors principles and research are often used to make aviation safer. Many of the findings in the aviation domain are applicable to surface transportation. Defining the exact parameters of a given hacking attempt is impossible due to the infinite ways and situations the hack or cyberattack could occur. For this reason, and the recent introduction of vehicle hacking as an issue, few researchers have attempted to examine the effect of novel vehicle hacks on human response. Those that have looked at the effect of hacking attempts on human responses have implemented a specifically designed control hack to examine its effect on operator response. (For the sake of simplicity across related fields, vehicle hacking and cyberattacks will be used interchangeably.) For example, Gontar and colleagues (2018) investigated the effect of cyberattacks on pilot performance. Half of the pilots in the experiment were exposed to a cyberattack while the other half were not. Their results suggested that the cyberattack on the pilot's navigational aid negatively affected the pilots' mental workload, eye movements, and behavior. This was shown in self-report data from questionnaire responses and in the deterioration of the pilot's instrument monitoring and overall performance.

In an attempt to further examine the effect of cyberattacks on pilots, Gontar et al. (2018) manipulated whether the pilot was warned about the incoming cyberattack in the experiment with half of the participants receiving an ambiguous semantic warning from the air traffic control tower and the remaining half not receiving the warning. The ambiguity of the warning was used to determine if there was a difference in responses to the warning when it was followed by a cyberattack or not. Warning the pilots about the possible cyberattack reduced the behavior and performance deficits caused by the cyberattack and subsequently raised workload when the attack occurred. Providing a warning when no attack was implemented resulted in higher workload and behavior deficits similar to when the attack was actually implemented. This suggests that warnings that precede an attack are helpful for mitigating the negative effects of the attack, but false alarm warnings have negative effects on pilot performance.

The current study, similar to the Gontar et al. study, focused on how humans responded to cyberattacks as well as whether warnings mitigated some of the negative effects of cyberattacks. The current study expected to find similar benefits of warning implementation due to an increase in SA on novel critical event response due to hacking attempts in the surface transportation domain.

In summary, the overall amount of responses that express concern about vehicle hacking found in McManus et al., (2018) suggests that the public believes vehicle hacking is a real issue and is concerned about its potential. Hacking related research in aviation that shows the negative effect of hacking attempts on human operators, is at least of equal concern in the increasingly automated surface transportation domain. With this in mind, researchers need to empirically examine the way people respond in the event of a vehicle hacking attempt and how responses are affected by the vehicle automation level.

Semantic Versus Non-Semantic Warnings in Surface Transportation

The predicted increase in vehicle hacking attempts is one of many issues that warnings could be used to remedy. It is essential that warnings used in vehicles reflect the urgency of the situation at hand and are representative of what they warn the user of (Politis, Brewster, & Pollick, 2015). Auditory warnings are beneficial in the surface transportation context because they are capable of attracting the driver's attention during visually demanding tasks like driving (Baldwin, 2011; Petermeijer et al., 2017). Moreover, speech-based warnings are capable of expressing additional information due to the inclusion of their meaning in the warning itself, and some messages convey more urgency than others based on the content of the warning (Hellier et al., 2002).

Warnings with speech include semantic information that has the added benefit of informing the driver about the hazard in addition to alerting them of the presence of the hazard. The additional information provided by warnings including speech is of extra benefit to the driver when the driver must perform a crash avoidance maneuver (Chang et al., 2008). The extra information included in semantic warnings could potentially boost the user's SA at the time of the warning when compared to warnings that do not contain clear semantic information. The use of auditory warnings to warn the driver of incoming vehicle hacking attempts could help to improve human responses to vehicle hacking attempts. In order to use them in future vehicles, a further understanding of how the semantic information in warnings affects the driver's SA is needed.

Acoustic and semantic aspects of auditory warnings have been shown to affect each other when used for collision avoidance (Baldwin & Moore, 2002; Baldwin & May, 2005). When comparing speech words such as "danger, caution, and warning", Bansal and May (2005) found that the word "danger" received a higher hazard rating than "caution" and "warning". The higher hazard rating for danger has been replicated by other research (Politis et al., 2015). According to Horswill and McKenna (2004), hazard perception can be considered SA for dangerous situations in the driving environment. A novel vehicle hacking attempt while driving or monitoring automated driving could serve as a dangerous situation in the driving environment. Politis et al. (2015) found that non-semantic (tone-based) auditory cues had faster recognition times in low criticality situations but that tone-based and speech-based cues have similar response times in response to critical events. In the same study, the authors suggested the use of speech-based warning due to a marginal increase in driving performance when speech-based warnings were implemented. These findings suggest that the semantic content of a warning may affect the driver's ability to respond and the effect of the semantic information may depend on the criticality of the situation.

The semantic information contained in a warning has been shown to dictate aspects of human response and lead to different levels of reported urgency and pleasantness. For example, McKeown and Isherwood (2007) found that non-semantic, tone-based sounds had the longest response times and lowest accuracy whereas speech and auditory icons had the lowest response times and highest accuracy. Additionally, in the same study, the participants reported speech warnings to be more pleasant and less urgent than non-semantic sounds. When comparing speech based and non-speech based auditory warnings, Edworthy and colleagues (2000) found no difference in perceived urgency of the warnings. Politis et al. (2015) claim that other than the Edworthy et al.'s study no other study had compared these two warning types with regards to urgency. Politis and colleagues compared language-based warnings with between three and six words to tone-based warnings at three different levels of urgency. They found no significant differences between the two warning types in number of mistakes made in response to the warnings, recognition time, or response time. Politis et al. posit that the lack of a difference in responses suggests language-based warnings perform just as well as tone-based warnings, which leads to a potential advantage for language-based warnings due to their higher pleasantry ratings and additional semantic content.

Past research has examined critical events that are often repeated and common such as braking in response to a vehicle. The similarity between semantic and non-semantic warnings may not be found when a novel critical event occurs because of the increased need for additional SA for unfamiliar situations. The addition of extra information in a semantic warning may increase SA for the driver enough to affect the quality of their response to a hacking attempt. The current study aimed to expand on the Politis et al.'s (2015) findings by comparing warnings with semantic information (e.g. Semantic, Verbal warnings) to warnings without semantic information (e.g. Non-Semantic, Tone-Based warnings) to examine if semantic information affected driver SA and responses to novel vehicle hacking attempts. The current study hoped to make this comparison and provide further information about the effect of semantic information in warnings on driver SA.

Current Study

Little empirical research on automated vehicle hacking has been done and research on how individuals respond to automation hacking in general is very limited (McManus et al., 2018). To bridge this gap in the literature, research on driver reactions to vehicle hacking at different levels of automation is essential. The current study compared human reactions to a vehicle hacking incident when the vehicle was operating at between Level 2 and Level 3 (SAE, 2018) automation (Automated Condition) to traditional manual driving (Manual Condition) using a driving simulator. To develop strategies to prevent the negative effects of automated vehicle hacking, there is a need to investigate how humans interact with the vehicle in the event of a hacking attempt. The present research examined how humans responded to a simulated aggressive vehicle hacking attempt.

A between-subject design was used to examine the effect of level of automation (Automated vs. Manual) and auditory warning type (Semantic vs. Non-Semantic) on how the drivers responded to the vehicle hacking attempt. The hacking attempt, in this experiment, was a loss of vehicle control as it drives off the road towards vehicles and buildings. This loss of control due to the simulated hacking attack was preceded by an auditory warning dependent upon which warning condition the participant was assigned to.

Hypothesis 1: Participants in the manual driving group were expected to have higher TTC values than the participants in the automated driving group. The rationale for this hypothesis was that participants in the manual condition were expected to have higher SA for the primary task (in this case, vehicle control) than those in the automated condition, leading to faster reaction times and subsequently higher TTC values, as found in past studies (Gold et al., 2013; Kaber & Endsley, 2007; Kessel & Wickens, 1979, 1981: Parasuraman & Manzey, 2010; Shen et al., 2017; Strand et al., 2014).

Hypothesis 2: Participants in the manual driving group were expected to have more successful responses than the participants in the automated driving group. This expectation was consistent with Hypothesis 1, wherein participants in the manual group were expected to have faster reaction times than those in the automated groups as a result of higher SA levels. Faster reaction times were expected to translate to more successful responses in the manual condition in our experiment (Gold et al., 2013; Shen et al., 2017; Strand et al., 2014).

Hypothesis 3: Participants in the manual driving group were expected to have different reaction methods (wheel turns, pedal presses, etc.) than the participants in the automated driving group. The elevated SA of manual drivers was expected to result in different responses to the critical event as found in past studies (Gold et al., 2013; Kaber & Endsley, 2007; Merat et al., 2012). It was expected that more pedal responses may be found in the automated condition as they brake to give themselves more time to gather information about the situation and decide upon the correct method of response.

Hypothesis 4: Participants in the semantic warning condition were expected to have larger TTC values than participants in the non-semantic warning condition. The semantic warning was expected to provide more information than the non-semantic tone-based warning and increase the driver's SA more than the non-semantic tone-based warning (Chang et al., 2008; Hellier et al., 2002; McKeown & Isherwood, 2007), resulting in higher TTC values for the semantic warning condition.

Hypothesis 5: There was expected to be an interaction between level of automation and warning type such that, in the manual driving group, the addition of the semantic warning would have less of a positive effect on TTC values than in the automated driving group. The addition of semantic information from the semantic warning is unlikely to be as helpful to the operator when they are already at a higher level of SA in the manual condition. Adding semantic information to the warning is likely to be more helpful to participants in the automated driving group because it will boost their SA more than in the manual driving group (Chang et al., 2008; Hellier et al., 2002; McKeown & Isherwood, 2007), leading to an interaction between level of automation and warning type.

CHAPTER II

METHOD

Participants and Recruitment

All participants for this study were recruited using the Old Dominion University SONA participant recruitment system (https://odupsychology.sona-systems.com). All participants were granted SONA credits towards their course requirements in various psychology courses. A pilot study that contained 28 participants was performed to ensure that the experiment and data collection process were in order. Next, a total of 72 undergraduates between the ages of 18 (required age for participation in the SONA participant recruitment system) and 25 participated in the experiment (for power analysis information please see Appendix A). The cutoff age of 25 was chosen based on the findings of Wetton et al.'s (2010) study. Wetton and colleagues found significant differences in hazard perception ability between 18-25 and 26-45-year-old drivers. In order to avoid contaminating our sample with groups of different hazard perception abilities it was decided to focus on a convenience sample of college age students (18-25). The participants were tested in individual sessions on a driving simulator. Each session lasted approximately 50 minutes. The experiment was approved by the Institutional Review Board at Old Dominion University.

The final sample contained 72 participants that were randomly assigned to one of the four experimental conditions (18 per groups). All participants reported normal or corrected to normal vision and hearing according to the demographics survey given to all participants. The final sample population was majority female (N = 56; 77.8%). All participants were between the ages of 18-23 years old with an average age of 19.44 (SD = 1.55) years. The sample ethnicity was

47.2% African/African American, 36.1% Caucasian, 8.3% More than one race, 5.6% Asian, and 2.8% Other/Unknown.

Of the 72 participants collected 65 (90.3%) reported a private vehicle as their primary mode of transportation, with the remaining seven (9.7%) reporting walking and cycling as their primary mode of transportation. The participants reported the average age when they received their first license as 16.63 years old (SD = 0.86). When asked whether they drove to school or not, 63 participants (87.5%) reported that they did drive to school whereas nine participants (12.5%) reported that they did not drive to school. When asked about their driving frequency over the last twelve months, 45 (62.5%) of the seventy-two participants reported driving every day, 19 (26.4%) reported driving often, six (8.3%) reported driving sometimes, and two (2.8%) reported driving rarely. The average amount of accidents in the last three years for the entire sample was 0.63 accidents (SD = .85). Overall, the sample contained participants who were familiar with driving and avoided accidents. The collected sample should serve as good basis for understanding how young drivers (18-25 years old) respond to a vehicle hacking attempt under different automation and warning conditions (Wetton et al., 2010).

Stimuli and Apparatus

The stimuli were presented to the participants using a dual monitor personal computer running the STISIM Drive 3 software (http://stisimdrive.com). The stimuli were presented to the participants using a 27-inch Dell P2717H LCD Monitor with 1920x1080 pixel resolution and an NVIDIA GeForce GTX 1080 graphic card. There were two monitors present. The additional monitor was used by the experimenter to control the programs and was not shown to the participants. Participant inputs were performed using a Logitech G29 driving wheel and pedals (See Appendix B).

Procedure

Participants were seated in front of a computer monitor with a steering wheel and foot pedals for driving based responses to the program. They were first asked to sign the informed consent and fill out a basic demographics survey (see Appendix C). The participants were then seated so that they had access to the driving wheel and pedals. Participants performed a five-minute driving simulator practice session that consisted of freely navigating through a city with no ambient traffic. This was done to allow the participants to familiarize themselves with the driving simulator and to make sure that simulator sickness was not an issue.

The participants were then assigned to one of four groups. The level of automation (automated vs. manual) and warning type (semantic vs non-semantic) were randomly assigned to each participant, and they served as the experiment's main independent variables. The groups included the following combinations of the IVs: Manual Driving and Semantic Warning, Manual Driving and Non-Semantic Warning, Automated Driving and Semantic Warning, and Automated Driving and Non-Semantic Warning. Participants were assigned to one of the four groups.

The automated groups contained participants that were in the automated driving group (between SAE Levels 2 and 3). Participants in the automated groups were asked to actively watch as the automated vehicle system drove them to class and instructed to keep their hands on the simulator steering wheel and feet ready at the pedals. The drive to class lasted approximately 20 minutes. This time was chosen based on the average commute time of 20 minutes for residence of Hampton Roads wherein our subjects reside (Old Dominion University Social Sciences Research Center, 2015). The manual groups contained participants in the manual driving group. Participants in the manual groups were asked to manually drive themselves in the same simulator set up and along the same route as the automated groups, to class/work. Both groups were asked to monitor their vehicle to ensure that their vehicle followed the rules of the road including all speed limits and were monitored by the researchers to ensure compliance.

The semantic warning group received a semantic auditory warning that stated "Danger, hacking attempt incoming". This warning lasted for two seconds and immediately preceded the hacking attempt. The non-semantic warning group received a 500 Hz tone that lasted for two seconds and immediately preceded the hacking attempt. Participants were instructed to respond to their respective warnings by safely taking control of the vehicle.

In addition, participants in the manual condition were instructed to follow all posted speed limits and the rules of the road. Each participant was asked to participate in two drives. During the first drive no hacking incident occurred, and this served as a route and system familiarization drive. The second drive contained the vehicle hacking event approximately three quarters of the way through the route.

For the second drive, the route began the same as in the first drive, but approximately three quarters of the way through the route a vehicle hacking attempt took place. The hacking attempt was preceded by the vehicle security system's warning that was dependent upon which warning group the participant was assigned to. Upon the completion of the warning, the vehicle abruptly angled off the road toward a parked car and buildings on either the left or right side of the road. The hacking attempt lasted for five seconds, then the vehicle made contact with a parked car if no response was made. The five-second hacking duration was decided upon based on a meta-analysis done by Eriksson and Stanton (2017) wherein the results suggested that when no secondary task was present, it took participants approximately five seconds to take control of the vehicle from the automation. Upon contact with the parked car, the simulator simulated a crash and the simulation ended. When this occurred, it was recorded as a failed response. If the driver reacted in time and responded, the type of response, and the TTC values were recorded. When participants responded in time to avoid a collision and resumed normal driving, it was recorded as a successful response. After completion of the driving portion of the experiment, the participants were asked additional questions about their driving experience (see Appendix D) and given a Participant Experiment Opinion Survey that contained questions about their understanding and interpretation of the experiment (see Appendix E). Next, credit was granted by the researcher and the participant was asked not to tell others about the experiment and allowed to leave.

Dependent Measures

The participant's reactions to the hacking attempt provided information on how humans respond in the event of an aggressive vehicle hacking attempt. The current study specifically examined the methods with which the participants responded. The dependent variables that were collected are: Method of response (wheel turn vs pedal press), successful response, and TTC (Time to Collision).The participant's responses to the hacking attempt and accompanied warning were scored as 1= a wheel turn only, 2= pedal press only, or 3= both wheel turn and pedal press. The type of response was scored by the researcher based on the first frame of response data for each participant where they responded with a wheel turn greater than two degrees or a change in pedal depression that causes a change in acceleration/deceleration of more than 1 m/s^2 (Hergath et al.,2017; Merat et al., 2012; Petermeijer et al., 2017; Radlmayr et al.,2014) . . Method of response served as the first dependent variable that was examined. A succeed/fail criterion which was used to determine whether the takeover response was successful or not and was based upon whether the participant responded in time to avoid a collision and resume normal driving. Participants who responded to the hack and accompanied warning in time to avoid an accident and resumed normal driving were scored as successful responses. Successful responses were denoted by a 0, and those who did not avoid an accident were scored as unsuccessful responses, denoted by a 1. This successful response variable was compared between the two different driving groups in order to see which level of automation resulted in the highest successful response rate and as a result avoided the accident most frequently. The number of successful responses served as the second dependent variable that was examined.

The TTC information was measured as the time remaining until impact with the vehicle or building and the TTC measurement was started after the vehicle hacking attempt warning tone was completed. As such, it provides information about how quickly the driver was able to respond to the cyberattack. Higher TTC's are preferable in our study as high TTC's indicate more time remaining before the accident occurs. The entire time from the warning to collision was approximately seven seconds. Time to collision data was compared between the four groups. TTC served as the third dependent variable that was examined, but TTC was only collected for the participants who responded before a collision occurred. This was done to ensure that zero TTC scores of those who failed to respond did not skew the TTC values that were examined and compared between groups. The information gleaned from these dependent measures may be helpful in the future development of safety systems in the surface transportation domain. Empirically comparing manual and automated driving and semantic and non-semantic warnings will allow future researchers and designers to optimize the driver vehicle relationship.

CHAPTER III

RESULTS

Missing data/outliers. Prior to analyses, the collected data were cleaned (i.e., missing data, outliers) and statistical assumptions for ANOVA were addressed. For the continuous variable TTC, there were two TTC data points collected that were more than three standard deviations outside the mean TTC value. Winsorizing data takes significant outliers that otherwise would not be able to be used as data points and converts them to the score of next lowest TTC value that is not a significant outlier. This method allowed for the inclusion of data that otherwise would have to be omitted and allowed the researcher to maintain an equal number of participants in each experimental group. The two significant TTC outliers (2.619 seconds, 2.629 seconds) were Winsorized to the next lowest TTC score (4.579 seconds) from the dataset (Ghosh & Vogt, 2012). The descriptive statistics, chi-squared distributions, assumption analyses, and ANOVA's were conducted using IBM SPSS Statistics version 25.

The dependent variable (DV) TTC is continuous, and reaction type and successful response rate are categorical. To test hypotheses pertaining to continuous dependent variables (Hypotheses 1, 4, & 5) a 2x2 factorial ANOVA was conducted to determine mean differences on each outcome measure (i.e., TTC values) between the different groups. To test hypotheses 2 and 3, the categorical hypotheses, chi-squared tests were performed to examine the differences in successful responses and reaction methods between manual and automated driving groups.

ANOVA assumptions. The assumptions of ANOVA are the following: normality, independence of observations, and homogeneity of variance. To test the assumption of normality, histograms and Q-Q plots were used to examine skewness and kurtosis. When testing the normality assumption for the TTC scores (Hypotheses 1, 4, and 5), the Winsorized TTC data was used. When checking for normality of the Winsorized TTC data, the data were normally distributed as shown by a skewness and kurtosis values of -.051 and -1.240 respectively which fall between the cutoff values of -1.96 and 1.96 which represent the 95% confidence interval of a standard normal z distribution (Gravetter & Wallnau, 2005). The assumption for independence of observations was addressed using participant randomization to study conditions (manual or automated driving, non-semantic or semantic warning). Each group consisted of different individuals who were not measured twice or specifically matched in any way. The assumption of homogeneity of variance was checked to ensure the variance of data in each condition was the same using a Levene's test. The results of the Levene's test on the Winsorized TTC was not significant with $F\ (3,68) = 0.11$, $p = .953$, and as such the assumption of homogeneity of variance was met.

ANOVA results. To test Hypotheses 1, 4, and 5, Winsorized TTC was examined across level of automation condition (Manual or Automated) and across warning condition (Non-Semantic or Semantic) using a 2x2 factorial ANOVA. It was found that there was no significant effect of level of automation on TTC with $F(3, 68) = 0.18$, $p = .673$, $\eta_p^2 = .00$. The mean TTC value for the manual driving group ($M = 4.63$, $SD = .03$) was not significantly different from the average TTC value for the automated driving group ($M = 4.63$, SD = .03). Based on the results of the statistical analysis the researcher was unable to provide support for Hypothesis 1. It was also found that there was no significant effect of warning condition on TTC with $F(3, 68) = 1.20$, $p = .278$, $\eta_p^2 = .02$. The average TTC value for the Non-Semantic warning condition ($M = 4.63$, $SD = .03$) was not significantly different from the average TTC value for the Semantic warning ($M = 4.62$, SD = .03). Based on the results of the statistical analysis our data were unable to provide support for Hypothesis 4. Lastly the interaction between and warning condition was not significant with $F(3, 68) = 0.57$, $p =.455$, $\eta_p^2 = .01$. Based on the result of the statistical analysis the researcher was unable to provide support for the interaction predicted in Hypothesis 5.

Hypothesized chi-squared results. In order to test Hypothesis 2, the hypothesis that pertained to successful responses a chi-squared analysis was performed. Three of the seventy-two participants successfully avoided crashing into a car or building and were able to regain control of the vehicle to safely resume normal driving. Two of the successful responses occurred for participants in the automated semantic condition and the other successful response occurred in the automated non-semantic condition. These results suggest that the vehicle hacking attempt used in the current study may have been too difficult for the participants to safely respond. The results indicate that level of automation does not have a significant effect on successful responses, $\chi^2(1) = 1.06$, $p = .303$, not supporting Hypothesis 2.

In order to test Hypothesis 3, a chi-squared analysis was performed. Sixty-four of the seventy-two participants responded with only a wheel turn as there first recorded response to the hacking attempt and accompanied warning (scored as a 1). The remaining eight responses were wheel turns and pedal presses (scored as a 3). The results of the chi-squared analysis showed that there was no significant effect of level of automation on response type, $\chi^2(1) = .00$, $p = 1.00$. The result of the chi-squared analysis fails to provide support for Hypothesis 3.

Qualitative survey response results. A post-experiment survey was given to all participants in an attempt to gain insight into their understanding and opinions about automated vehicles and the current experiment. For the first question of the post experiment survey, the participants were given one of following statements, “I was engaged in *driving the vehicle* for the entire ride” (Manual) or “I was engaged in *monitoring the automated driving system* for the entire ride” (Automated), and asked to rate their level of agreement with the statement on a five-point Likert scale from strongly disagree (1) to strongly agree (5). A 2x2 ANOVA was used to determine if there was a difference in reported engagement due to level of automation or warning type. There was no significant difference in reported engagements due to level of automation, $F(3, 68) = 1.84$, $p = .180$, $\eta_p^2 = .03$. There was no significant difference in reported engagements due to warning condition, $F(3, 68) = 0.14$, $p = .713$, $\eta_p^2 = .00$. There was no significant interaction between level of automation and warning condition for reported engagement as shown by $F(3, 68) = 0.74$, $p = .391$, $\eta_p^2 = .01$.

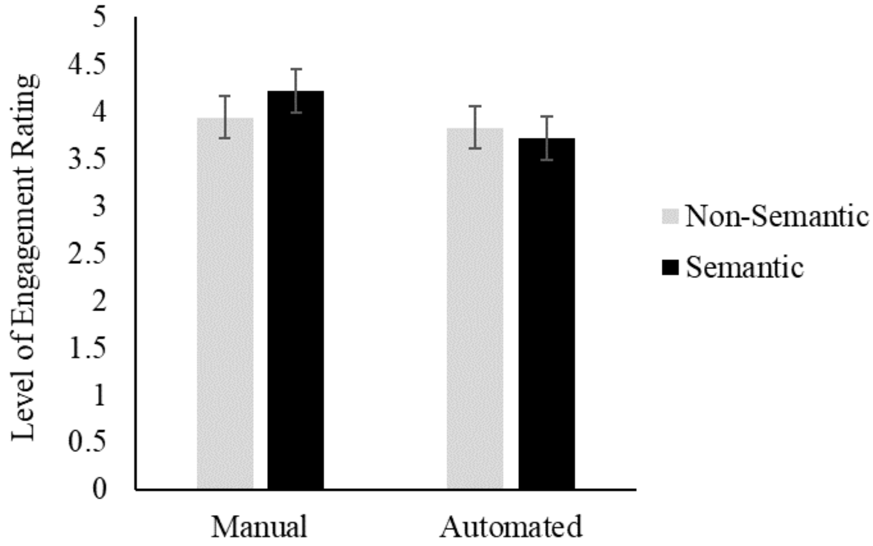


Figure 1. Reported engagement during drive (error bars represent 1 standard error of the mean).

Participants were also asked whether they noticed a warning during the drive. Chi-squared analyses were used to determine if there was a statistical difference in the number of participants that reported noticing a warning due to level of automation and warning condition. For level of automation there was not a significantly different number of participants that reported noticing the warning, $\chi^2(1) = 2.49$, $p = .114$. However, for warning condition, there was a statistically different number of participants that reported noticing the warning, $\chi^2(1) = 9.97$, $p = .002$.

The counts in Table 1 show that participants that received a semantic warning reported that they noticed a warning (Manual Semantic had only 3 no responses, Automated Semantic had only 1 no response) more than participants who received a non-semantic warning (Manual Non-Semantic had 10 no responses, Automated Non-Semantic had 6 no responses). This suggests that non-semantic warnings were more frequently not noticed as a warning compared to semantic warnings. Additionally, non-semantic warnings resulted in less reports of noticing the warnings (Manual Non-Semantic had 8 yes responses, Automated Non-Semantic had 12 yes responses) than participants who received semantic warnings (Manual Semantic had 15 yes responses, Automated Semantic had 17 yes responses). This suggests that participants who received a semantic warning more frequently reported noticing the warning as a warning than non-semantic warning recipients based on the counts.

Participants in the manual driving group reported not noticing the warning (Manual Non-Semantic had 10 participants, Manual Semantic had 3 participants) more frequently than the participants in the automated driving group (Automated Non-Semantic had 6 participants, Automated Semantic had 1 participant). Participants in the manual driving group reported noticing the warning (Manual Non-Semantic had 8 participants, Manual Semantic had 15 participants) less frequently than the participants in the automated driving group (Automated Non-Semantic had 12 participants, Automated Semantic had 17 participants).

Table 1

Count of Whether Participants Noticed Warning or Not

	Noticed Warning?	
Group	No	Yes
Manual Non-Semantic	10	8
Manual Semantic	3	15
Automated Non-Semantic	6	12
Automated Semantic	1	17
Total	20	52

The participants in both level of automation conditions and warning conditions that reported noticing the warning were then asked, “What did the warning you received represent?” with four response options (A system malfunction, an automation failure, a vehicle hacking attempt, or other). The participants’ responses to the question were broken down by level of automation (Manual or Automated) and the warning type they received (Semantic or Non-Semantic). A chi-squared analysis was used to determine if the perceived representation of the warning was significantly different across the four experimental groups (Manual Non-semantic, Manual Semantic, Automated Non-semantic, and Automated Semantic). The four groups did not have significantly different perceived representations of the warning, $\chi^2(3) = 6.09$, $p = .107$ (See Table 2).

Table 2

Perceived Warning Meaning Counts

Group	Perceived Warning Meaning			
	Vehicle Hacking Attempt	System Malfunction	Other	Automation Failure
Manual Non-Semantic	1	4	2	1
Manual Semantic	10	4	0	1
Automated Non-Semantic	0	2	1	9
Automated Semantic	8	6	1	2
Total	19	16	4	13

In addition, chi-squared analyses were used to determine if the perceived representation of the warning was significantly different based upon level of automation and warning type when the responses were scored as a vehicle hacking attempt or as not a vehicle hacking attempt. For level of automation there were not significantly different perceived representations of the warning, $\chi^2(1) = 2.99$, $p = .224$. However, for warning condition, there were significantly different perceived representation of the warning, $\chi^2(1) = 19.17$, $p < .001$. For specific counts of the perceived representations see Table 3 below. These findings suggest that level of automation did not significantly affect the perceived representation of the warning as a hack or not, but the warning type did.

Table 3

Perceived Warning Meaning Counts: Hacking Attempt vs. Not a Hacking Attempt

	Perceived Warning Meaning	
Group	Vehicle Hacking Attempt	Not a Vehicle Hacking Attempt
Manual Non-Semantic	1	7
Manual Semantic	10	5
Automated Non-Semantic	0	12
Automated Semantic	8	9
Total	19	33

Participants that reported noticing the warning were also given the statement, "I perceived the warning I received to be urgent" and asked to rate their level of agreement with the statement on a five-point Likert scale from strongly disagree (1) to strongly agree (5). A 2x2 ANOVA was used to determine if there was a difference in reported urgency due to level of automation or warning condition. There was no significant difference in reported urgency due to level of automation shown by $F(3, 68) = 2.25$, $p = .140$, $\eta_p^2 = .05$. There was a significant effect difference in reported urgency due to warning condition shown by $F(3, 68) = 4.43$, $p = .041$, $\eta_p^2 = .08$. There was also a significant interaction between level of automation and warning type for reported urgency shown by $F(3, 68) = 5.21$, $p = .027$, $\eta_p^2 = .10$. These findings suggest that non-semantic warnings are perceived as less urgent than semantic warnings but only in the manual driving condition.

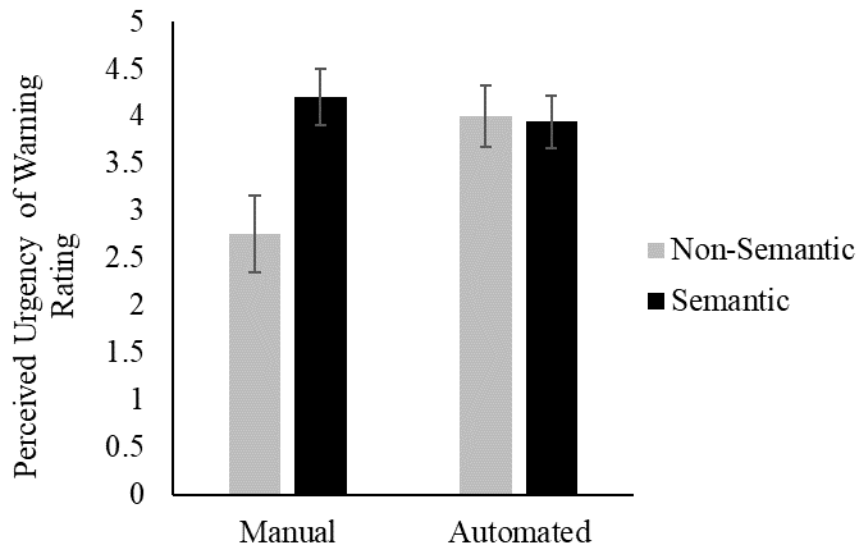


Figure 2. Reported perceived urgency of warning (error bars represent 1 standard error of the mean).

All participants in the study were asked “Were you aware of the potential for vehicle hacking before this experiment?”. Chi-squared analyses were used to determine if the number of participants with previous knowledge of vehicle hacking was significantly different due to level of automation or warning condition. For level of automation there was a significantly different number of participants with previous knowledge of vehicle hacking, $\chi^2(1) = 12.77$, $p < .001$. However, for warning condition there was not a significantly different number of participants with previous knowledge of vehicle hacking, $\chi^2(1) = 0.36$, $p = .551$. For specific counts of participants with previous hacking knowledge see Table 4 below. These findings suggest that there was a significant difference in the number of participants with previous knowledge of vehicle hacking attempts when examined based upon level of automation but not when examined based upon warning condition.

Table 4

Count of Participants Who Reported Previous Knowledge of Vehicle Hacking

	Previous Knowledge of Vehicle Hacking Potential?	
Group	No	Yes
Manual Non-Semantic	17	1
Manual Semantic	18	0
Automated Non-Semantic	13	5
Automated Semantic	10	8
Total	58	14

CHAPTER IV

EXPLORATORY RESULTS

The researcher performed several exploratory analyses that were not included in the proposed analyses to gain further insight into the effects of level of automation and warning condition on participants response quality. These measures of response quality included maximum steering wheel angle, maximum steering wheel angle over time, and maximum resulting acceleration. The measures of response quality used in the current study were included in order to gain more specific additional information about the participants responses to the vehicle hacking attempt. More detailed analysis of the quality of the responses may provide additional information about the hacking attempt responses above that provided by the hypothesized analysis.

Maximum steering wheel angle. The first exploratory analysis preformed pertained to maximum steering wheel angle which is measured in degrees. Maximum steering wheel angle was calculated by taking the difference of the angles of participant's initial wheel turn response greater than two degrees and the participant's largest (most severe) wheel turn across the duration of the hacking attempt (Shen & Neyens, 2017). Two degrees was used as the wheel response cutoff due to its use in the literature as the smallest number of degrees the steering wheel is required to move to count as a wheel response (Merat et al., 2012; Petermeijer et al., 2017; Radlmayr et al., 2014). Maximum steering wheel angle is an indication of response quality such that larger maximum steering wheel angle values are considered less safe.

Maximum steering wheel angle was analyzed using a 2x2 factorial ANOVA to determine if level of automation (Manual or Automated) or warning type (Semantic or Non-Semantic) had an effect on the participants maximum steering wheel angle. Before the ANOVA was conducted, the assumptions of normality, independence of observations, and homogeneity of variance were addressed for maximum steering wheel angle and the data was checked for significant outliers. There were no significant outliers and the skewness and kurtosis values of .94 and -.51 respectively suggest the data is sufficiently normal because they fall between the cutoff values of -1.96 and 1.96 which represent the 95% confidence interval of a standard normal z distribution (Gravetter & Wallnau, 2005). The assumption for independence of observations was addressed using participant randomization to study conditions (manual or automated driving, non-semantic or semantic warning). The assumption of homogeneity of variance was checked to ensure the variance of data in each condition was the same using a Levene's test. The results of the Levene's test on maximum steering wheel angle was significant with $F(3, 68) = 3.77$, $p = .015$, and as such the assumption of homogeneity of variance was not met. ANOVA is sensitive to violations of homogeneity of variance, but this sensitivity is mitigated due to the equal sample sizes for each level of each independent variable (Sawyer, 2009).

There was a significant effect of level of automation on maximum steering wheel angle, $F(3, 68) = 4.53$, $p = .037$, $\eta_p^2 = .06$ (See Figure 3). The mean maximum steering wheel angle for participants in the Manual driving group was $M = 505.12$ ($SD = 75.64$) and the mean maximum steering wheel angle for participants in the Automated driving group was $M = 548.45$ ($SD = 97.25$). There was an effect of warning condition was not significant, $F(3, 68) = 3.19$, $p = .079$, $\eta_p^2 = .05$ (See Figure 3). The mean maximum steering wheel angle for participants in the Semantic warning condition was $M = 544.95$ ($SD = 91.41$) and the mean maximum steering wheel angle for participants in the Non-Semantic condition was $M = 508.61$ ($SD = 84.34$). There was not a significant interaction between level of automation condition and warning condition for maximum steering wheel angle, $F(3, 68) = .026$, $p = .873$, $\eta_p^2 = .00$.

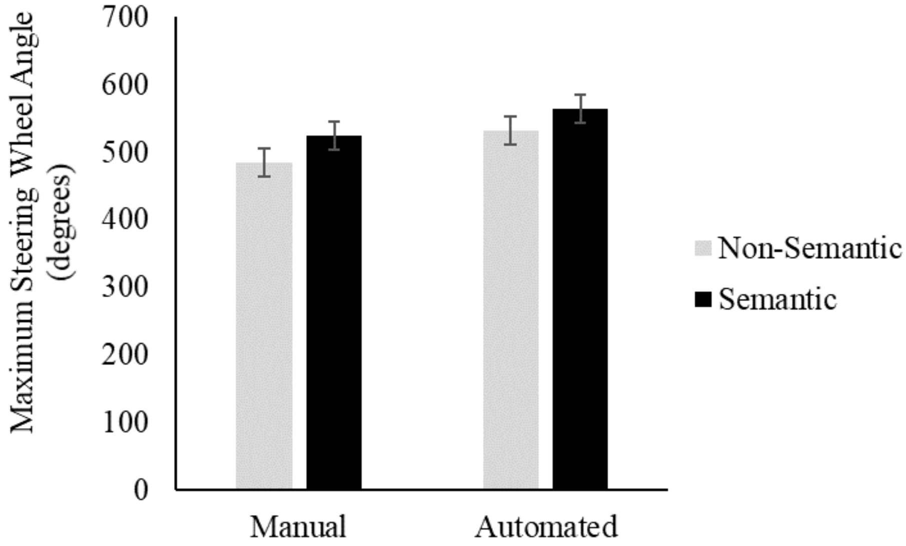


Figure 3. Maximum steering wheel angle (error bars represent 1 standard error of the mean).

Maximum steering wheel angle over time. In addition to maximum steering wheel angle, maximum steering wheel angle over time was calculated and is a unique use of the previously used maximum steering wheel angle dependent variable that was created to further analyze the results of the experiment. Maximum steering wheel angle over time was calculated by taking the calculated maximum steering wheel angle and dividing it by the time between the initial and largest reaction. Maximum steering wheel angle over time allowed researchers to quantify how quickly the participant turned the wheel from their initial response to their most severe response. Maximum steering wheel angle over time is a measure of response quality such that larger values are more dangerous responses based on how quickly the user turned the wheel from their initial turn to their most severe turn.

Maximum steering wheel angle over time was analyzed using a 2x2 factorial ANOVA to determine if level of automation (Manual or Automated) or warning type (Semantic or Non-Semantic) had an effect on the participants maximum steering wheel angle over time. Before the ANOVA was conducted the assumptions of normality, independence of observations, and homogeneity of variance were addressed for maximum steering wheel angle over time and the data was checked for significant outliers. There were no significant outliers and the skewness and kurtosis values of 1.29 and -.08 respectively suggest the data is sufficiently normal because they fall between the cutoff values of -1.96 and 1.96 which represent the 95% confidence interval of a standard normal z distribution (Gravetter & Wallnau, 2005). The assumption for independence of observations was addressed using participant randomization to study conditions (manual or automated driving, non-semantic or semantic warning). The assumption of homogeneity of variance was checked to ensure the variance of data in each condition was the same using a Levene’s test. The results of the Levene’s test on maximum steering wheel angle over time was significant with F (3,68) = 2.97, p = .04, and as such the assumption of homogeneity of variance was not met. The violation of ANOVA's assumption of homogeneity of variance is mitigated due to the equal sample sizes for each level of each independent variable (Sawyer, 2009).

There was not a significant effect of level of automation on maximum steering wheel angle over time with F (3,68) =.65, p = .424, η_p^2 = .01. The mean maximum steering wheel angle over time for participants in the Manual driving group was M = 202.19 (SD = 103.43) and the mean maximum steering wheel angle over time for participants in the Automated driving group was M = 221.66 (SD = 102.43). There was no significant effect of warning condition on maximum steering wheel angle over time with F (3,68) = 1.41, p = .240, η_p^2 = .02. The mean maximum steering wheel angle over time for participants in the Semantic warning condition was M = 197.59 (SD = 86.32) and the mean maximum steering wheel angle over time for participants in the Non-Semantic condition was M = 226.27 (SD = 116.24). There was not a significant interaction between level of automation condition and warning condition for maximum steering wheel angle over time with F (3,68) = 1.04, p = .312, η_p^2 = .02.

Maximum resulting acceleration. An additional measure of response quality was also collected. Maximum resulting acceleration is calculated by taking the square root of the sum of the squared values of the vehicles longitudinal and lateral accelerations and is presented in m/s^2 (Gold et al., 2013; Hergeth, Lorenz, & Krems, 2017). If maximum resulting accelerations are high and approach the physical limits of the vehicle and the vehicle's tires, the vehicle may become unstable. Thus, lower maximum resulting accelerations are safer and considered better quality responses.

Before the maximum resulting acceleration data was analyzed across driving and warning conditions the assumptions of ANOVA were checked. There were ten significant outliers in the maximum resulting acceleration data. The skewness and kurtosis values of 1.46 and .39 respectively suggest the data is sufficiently normal because they fall between the cutoff values of -1.96 and 1.96 which represent the 95% confidence interval of a standard normal z distribution (Gravetter & Wallnau, 2005). The assumption for independence of observations was addressed using participant randomization to study conditions (manual or automated driving, non-semantic or semantic warning). The assumption of homogeneity of variance was checked to ensure the variance of data in each condition was the same using a Levene's test. The results of the Levene's test on maximum resulting acceleration was not significant with F (3,68) = 2.30, p = .085 and as such the assumption of homogeneity of variance was met.

A 2*2 factorial ANOVA was performed to determine if level of automation (Manual or Automated) or warning type (Semantic or Non-Semantic) had an effect on the participants maximum resulting acceleration values. There was not a significant effect of level of automation on maximum resulting acceleration with F (3,68) = .53, p =.469, η_p^2 = .01. The mean maximum resulting acceleration value for participants in the Manual driving group was M = 24.71 (SD = 2.53) and the mean maximum steering wheel angle over time for participants in the Automated driving group was M = 25.19 (SD = 3.03). There was no significant effect of warning condition on maximum resulting acceleration with F (3,68) = 1.48, p = .229, η_p^2 = .02. The mean maximum resulting acceleration value for participants in the Semantic warning condition was M = 25.35 (SD = 2.95) and the mean maximum resulting acceleration value for participants in the Non-Semantic condition was M = 24.55 (SD = 2.58). There was not a significant interaction between level of automation condition and warning condition for maximum resulting acceleration with F $(3,68) = .008$, $p = .929$, $\eta_p^2 = .00$.

CHAPTER IV

DISCUSSION

The primary objective of this study was to determine if the level of automation used to control the vehicle and the type of auditory warning, affected the driver's response to novel critical events. This was accomplished by comparing manual driving and automated driving and providing either a semantic or non-semantic warning to the driver before the onset of the vehicle hacking attempt that served as the novel critical event. The driver's TTC, response method, the amount of successful responses, and three measures of response quality were analyzed based upon the level of automation and warning type that the driver was randomly assigned. The results of the experiment suggest that level of automation may not affect how quickly people respond to vehicle hacking attempts when the hacking attempt is preceded by an auditory warning. In addition, the way participants responded to the hacking attempt and the amount of successful responses to the hack, were very similar regardless of level of automation or warning type. However, exploratory analyses provided some evidence that manual driving may lead to safer responses to critical events as shown by the lower maximum steering wheel angles in the manual driving condition. Additionally, exploratory analyses suggested that semantic warnings are more often interpreted correctly overall and semantic warnings lead to higher reported urgency, specifically when manually driving the vehicle.

Level of Automation: TTC

The first major manipulation used in the current study was the level of automation used to control the vehicle. The current study hypothesized that level of automation would affect participants TTC values such that participants in the manual driving condition would have higher TTC values than participants in the automated driving condition (Hypothesis 1). The current experiment did not find a significant effect of level of automation on TTC, and as such, support for Hypothesis 1 was not found. This finding is not consistent with previous research in the field.

The null findings for TTC, while not hypothesized, do partially coincide with the findings of Merat et al. (2012) who also did not find a difference in how quickly participants responded based upon automation level when no secondary task was present. The presence of a secondary task has been shown to increase the SA deficit caused by the level of automation in the driving domain by further distracting the driver from the primary task. The current study also did not have a secondary task which may have been a reason why no TTC differences were found. When no secondary task is present, it is possible that the level of SA is not sufficiently lowered in the automated driving group to lead to significant differences in their responses to the hacking attempt.

Past research by Gold et al. (2013) and Shen and Neyens (2017) both examined the effect of manual and automated driving on participant responses to critical events similar to the current experiment. They both found significant advantages, in terms of reaction time, for the manual driving condition over the automated driving condition. In addition, when comparing semi-automated and highly automated driving, Strand et al. (2014) found faster reactions times for participants in the semi automation condition than in the highly automated condition. The researchers explained the results of these past experiments by positing that the faster reaction times for drivers in the manual and lower automation conditions could be due to higher levels of SA compared to the automated driving conditions.

For the current experiment, the TTC values were very similar for both levels of automation which does not support our hypothesis or match past findings in the literature. The lack of differences between the two levels of automation for TTC could suggest that there was not a difference in level of SA between manual and automated driving like there was in the past studies by Gold et al. (2013), Strand et al. (2014), and Shen and Neyens (2017). This lack of a difference in TTC and potentially SA, could be due to the auditory warning that was presented to all participants before the hack began. By providing a warning beforehand, it is possible that participants in each group had the duration of the warning (2 seconds) to gain SA, and as such, members in both groups were prepared to respond to the critical event by the time it occurred. In addition, Gold et al. (2013), Strand et al. (2014), and Shen and Neyens (2017) used more traditional driving event scenarios such as braking for a vehicle in front or changing lanes to avoid an upcoming obstacle as their critical event. These methodologies are different than the complex and novel hacking scenario used in the current study.

In summary, the novelty of the vehicle hacking attempt used in the current study may have led to the findings that are inconsistent with the results of Gold et al. (2013), Strand et al. (2014), and Shen and Neyens (2017). The current finding suggests that the level of automation present does not affect how quickly participants are able to respond to the critical event. This inconsistency could be due to the lack of a secondary task, resulting in less differences in SA in the current study, similar to Merat et al.'s. (2012) study. It could also be due to the addition of the warning before the critical event, which possibly eliminated the SA differences between the levels of automation. This may have led to a ceiling effect for TTC, such that all participants responded to the critical event very quickly, eliminating the potential differences between the driving groups. We did not find a significant difference between manual and automated driving for TTC which is contrary to the previous literature, potentially due to the warning implementation time and its effect on participants SA levels. Despite this, the benefit of providing warnings before the actual critical event was highlighted as shown by the extremely high TTC values found across all experimental groups.

Level of Automation: Successful Responses

The next goal of the study was to determine if the level of automation condition affected the participants ability to successfully respond to the hacking attempt and safely resume control of the vehicle. In order to examine this, the current study compared the amount of successful responses to the hacking attempt between the manual and automated driving groups. In the current experiment, participants in the manual driving group were expected to have more successful responses than participants in the automated driving condition. Only three of the seventy-two participants were able to successfully respond to the hacking attempt and all three were in the automated driving condition. The statistical analysis did not find a significant difference in successful responses between the two levels of automation. The current study did not find significantly different numbers of successful responses to the vehicle hacking attempt in the manual and automated driving conditions. Based on this finding, the current study did not find support for Hypothesis 2.

Strand et al. (2014), in addition to other previous studies (Gold et al., 2013, & Shen and Neyens, 2017) suggested that participants in the manual and semi-automated driving groups would have more successful responses to the critical event than participants in the automated driving group. The past authors attributed the higher amount of safe responses to the critical event to higher levels of SA for those in semi-automated or manual driving. They suggested that higher SA levels for participants was crucial for adequate responses and that higher automation levels lead to participants that are more “out-of-the-loop”. These findings were not replicated in the current experiment possibly due to the potential lack of differences in SA shown by the very similar TTC values for both level of automation conditions.

The overall low successful response rate across all participants (3 successful attempts in 72 critical events) suggests that the hacking scenario may have been too difficult for the participants to safely respond, due to its novel and complex nature. The past studies by Gold et al. (2013), Strand et al. (2014), and Shen and Neyens (2017) used simpler critical event scenarios such as a lane change due to a stalled vehicle which may contribute to their significant results. In contrast, they suggested the use of more complex and novel critical events for future research. The hacking scenario used in the current study included vehicles and buildings in a cityscape and the closeness of objects that needed to be avoided may have made the scenario too difficult for participants to avoid while also attempting to combat the effects of the hacking attempt.

In summary, there was no significant difference for successful responses between the two levels of automation which is not consistent with previous findings in the literature. This finding suggests that the level of automation present does not affect the participants likelihood of successfully responding to the critical event. Our null result is potentially due to the elevated difficulty and complexity of the scenario and the hacking attempt as shown by the very low number of successful responses as well as the elevated SA levels due to the early onset of the warning.

Level of Automation: Response Method

Next, we wanted to determine if the level of automation led to different response methods (wheel turns, pedal presses, etc.) to the vehicle hacking attempt critical event. We hypothesized that participants in the manual driving condition would have different response methods than the participants in the automated driving condition potentially due to a difference in SA between the two groups. It was expected that more pedal press responses would be found in the automated condition as they brake to give themselves more time to gather information about the situation and decide upon the correct method of response. The current study did not find significant differences in respond method in the manual and automated driving conditions. Approximately 89% of all responses to the vehicle hacking critical event were wheel turns and there were equal amounts of wheel turn and wheel turn/pedal press responses in the manual and automated driving groups. Based on these findings, we did not find support for Hypothesis 3.

Past research by Gold and colleagues (2013) found that participants in the automated driving group had slower reaction times due to lower levels of SA. Further, when the participants had less time to react to the critical event due to their lower SA levels the participants were more likely to use the brake than to use a wheel turn to steer the vehicle. Gold et al. posited that this tendency to brake rather than to steer as a first reaction was likely an attempt by the participants to give themselves more time to gather SA about the critical event and their surroundings before making the decision of how to appropriately respond. This explanation and accompanied results were not found in the current experiment, potentially due to a lack of differences in SA between the manual and automated driving conditions in the current study.

There are a couple of potential reasons for the inconsistency with Gold and colleague's study and the current experiment's null result. Implementing the warning before the hacking attempt potentially led to higher levels of SA across both levels of automation and provided a window of time for the participant to attend to the vehicle and analyze the situation before the hacking attempt actually began. This could explain why the participants used a wheel turn as a first response rather than a brake press since they may not have needed to create additional extra time to analyze their situation. The potentially higher SA levels across both levels of automation could also have led to the participants typically using a wheel turn as their response method to the hacking attempt.

Another potential reason for the null result was the way the hacking attempt was implemented. The hacking attempt itself caused the vehicle to change direction and veer off the road. Had the hacking attempt required the vehicle to avoid an object directly in front of them it is possible they would have been more likely to use the brake as found in Gold et al.'s (2013) study. In essence, the way that the hack occurred, and its change of direction may have been the reason why the participant predominately used a wheel turn to combat the direction change rather than a brake press.

In sum, there was no significant difference in response method between the manual and automated driving groups. This result suggests that the level of automation present does not affect the way drivers respond to critical events. This result is not consistent with past findings or the theory of SA (Endsley, 1995). The lack of a difference between the two groups is potentially due to elevated SA levels for both groups due to the preemptive warning and the nature of the hacking attempt. Further research in needed into this domain as the complexity of the current experiment may have led to results that are not indicative of all driving situations.

Warning Type

In addition to level of automation, the current study also manipulated the content of the auditory warning that each participant received. The warnings contained either a tone or a semantic statement and we aimed to determine if the semantic content within a warning affected the participants responses to the vehicle hacking attempt. It was hypothesized that participants in the semantic warning condition would have higher TTC values than participants in the non-semantic warning condition. This expectation was based upon the idea that warning type would affect participant’s responses such that semantic warnings would elevate the participants SA by informing them of the cause of the warning resulting in better warning responses. The current study did not find significantly different TTC values between the semantic and non-semantic warning suggesting that the semantic warning did not provide additional benefit over the non-semantic warnings. Based on this result, the current study was not able to support Hypothesis 4.

Past research that examined whether the content of a warning affected responses produced results not replicated in the current study. For instance, Politis et al. (2015) found that when auditory warnings were used, language-based warning recipients had faster reaction times than abstract tone-based warning recipients. Also, Baldwin and May (2005) suggested that verbal auditory warnings have the potential to reduce crash risk during hazardous situations. Chang et al. (2008) also posited that warnings that included speech are beneficial to the driver when they need to avoid a crash. Chang and colleagues suggested that the extra information provided by the semantic warning over the non-semantic warning would lead to better SA for the driver. These findings and others, (Hellier et al., 2002; McKeown & Isherwood, 2007) were the basis for the current studies expectations of different TTC values between the groups due differences in SA provided by the warnings. However, the current study did not replicate the findings of these studies as shown by nearly identical TTC values for participants in both the semantic and non-semantic warning conditions.

The results found in the current study, that do not match these previous studies, are potentially the result of a few experimental design differences. Past research on auditory warnings has focused on after-the-fact reported urgency and pleasantness ratings, whereas the current study attempted to see if the warning type produced tangible differences in how drivers responded to a novel and complex driving event. This complicated and potentially less warning centric design may have resulted in similar levels of SA for the drivers. It is possible that the null result of the current study is due to similarities between the warnings which were matched based upon volume and length, but control and experimental conditions are often matched as closely as possible so that any differences can be attributed to the one aspect that is different.

An additional possible cause of the null result is that the warnings were provided before the hacking attempt occurred and both warnings lead to sufficient boosts in SA for the participant to quickly respond to the hacking attempt. If drivers in both warning groups were warned ahead of time for an extended period, such as in this experiment, it may have led to sufficient time for the driver to gather SA in time to respond to the hacking attempt. This suggestion is supported by the very high mean TTC scores for both warning groups that were found in this study. In hindsight, presenting the warnings simultaneously with the hacking attempt may have provided a scenario wherein the semantic value of the warning had an effect because the participants would not already be at higher SA levels before the hacking attempt even began.

In conclusion, there was no significant difference for TTC between the semantic and non-semantic warnings. This finding suggests that the content of the warning does not affect the driver's ability to respond quickly to the warning, but these results are not in accordance with the theoretical and experimental findings in the warning and SA literature. This null result is possibly due to the similarity between the warnings and the fact that both warnings were provided before the hacking attempt occurred. The early implementation of the warnings may have elevated the participants SA levels sufficiently enough to lead to the overall large TTC values for both groups and masking the potential benefit of the semantic content that was found in past research.

Level of Automation and Warning Content

Our final hypothesis pertained to the potential interaction between warning type and level of automation such that the addition of the semantic warning in the manual driving condition would have less of a benefit for TTC values than the addition of the semantic warning in the automated driving condition. It was expected that the addition of semantic information would not benefit the manual group as much due to the manual drivers elevated SA levels, but it would benefit automated drivers who were more “out-of-the-loop” and had lower SA levels. The current experiment did not find a significant interaction between level of automation and warning conditions for TTC. As a result, we were unable to provide support for Hypothesis 5.

Previous research by Chang et al. (2008) proposed that the benefit of the semantic warning was due to the additional increase in SA that the extra content provided. The expectation that the SA change was the driving factor for the benefit of the semantic content is an important factor that can be found in the level of automation literature as well. Gold et al. (2013) proposed that the faster TTC values they found for the manual group were due to increased levels of SA for participants that were manually controlling the vehicle. If manual driving is predicted to lead to higher levels of SA than automated driving, it stands to reason that the SA benefit of the semantic warning may not be as beneficial for manual drivers due to their already elevated SA when compared to the automated driving group. However, the hypothesized interaction was not found in the results of the current study.

The results of the current experiment did not show a significant interaction between the level of automation condition and warning condition as was hypothesized and predicted based on previous research. This null result is potentially due to the fact that the manual and automated driving groups were not different based on TTC or response method suggesting similar levels of

SA for both levels of automation. In addition, there was no significant difference between the semantic and non-semantic warning conditions for TTC which suggests that both warnings provided sufficient boosts to SA prior to the hacking attempt which led to the high collected TTC values for response from both warning groups. If there was no benefit of manual driving that lead to higher SA levels and no benefit of semantic warning that lead to higher SA levels, then the expectation of a significant interaction between the two IVs is far less likely. It is also possible that by providing the warnings before the hacking attempt the participants were able to reach such elevated levels of SA that neither IV was able to significantly benefit one group over the other.

In summary, Hypothesis 5 regarding an interaction between level of automation and warning condition was not supported. This suggests that level of automation and warning type are not related but this finding is not in line with the theoretical and experimental findings in the related literature. This null finding is possibly due to a lack of main effects for each IV and the ceiling effect caused by the preemptive implementation of both warning types. Further research on how level of automation and warning type are related to one another is needed to bridge the gap between the warning and level of automation literature. This future research is of extra importance to transportation researchers, due to the high prevalence of auditory warning usage in the driving domain.

Reported Engagement

The first item on the post experiment survey given to the participants asked participants to rate their agreement with a statement suggesting that they were engaged in driving or monitoring the vehicle for the entire ride. The results showed no difference across the groups. This finding suggests that participants in each group reported that they were engaged in the drive/ride similarly regardless of level of automation and warning content. This finding also provides support that the random assignment used during the experiments worked sufficiently. In conclusion, participants reported being engaged regardless of experimental group.

Reported Warning Realization

The second item on the post experiment survey asked all participants if they noticed a warning during the drive and participants responded with a yes or no response. We did not find a significant difference in the number of participants who reported noticing the warning based upon the level of automation they were assigned. This null result suggests that noticing warnings is not affected by whether manual or automated driving is in use when the warning is implemented but this null finding may also be due to Type 1 error. However, we did find a significant effect of warning type on the number of participants who reported noticing the warning. Specifically, participants who received the non-semantic warning reported not noticing a warning more frequently than participants who received the semantic warning. This finding suggests that semantic warnings are superior if recognizing an auditory input as a warning is the goal. Based on this finding, we suggest that semantic warnings should be used when attempting to warn a participant about an upcoming critical event regardless of the vehicle’s level of automation. Using semantic warnings to warn participants of incoming critical events may increase the likelihood that individuals recognize the warning as a warning and respond to it rather than attributing the warning to ambient noise and ignoring the warning.

Reported Warning Interpretation

The third item on the post experiment survey asked the participants that did report noticing the warning what they interpreted the warning to represent. It was a multiple-choice question and the correct answer was a vehicle hacking attempt and all other responses were scored as not a vehicle hacking attempt. Understanding what the warning represents is essential to making the appropriate response in many cases and those who received semantic warnings appeared better at determining what the warning meant based on their self-report responses. There was no significant effect of level of automation on the number of correct interpretations of the warning, suggesting that warnings are interpreted regardless of the level of automation used in the vehicle. However, there was a significant effect of warning condition on the interpretation of the warning. Those who received the semantic warning were more likely to correctly interpret the warning as caused by a vehicle hacking attempt. This finding may potentially be due to the inclusion of what the warning represents in the semantic warning, but that is exactly what the addition of semantic content is supposed to provide.

The large number of incorrect interpretations of the warnings as system malfunctions or other issues may have influenced how quickly and successfully participants responded to the hacking attempt critical event. Had participants effectively interpreted the warnings as hacking attempt warnings, the urgency, method, and success of their responses may have changed or improved. In conclusion, the participants did not accurately interpret the cause of the warning, most of the time, and this inconsistency may have affected the way participants responded to the warnings across all aspects of the study. In summary, the level of automation present did not affect the interpretation of the warning, but the warning type received did. This finding suggests that semantic warnings are more often correctly identified than non-semantic warnings. Based on this, we suggest that semantic warnings should be used when the warning represents a novel or complex issue such as a vehicle hacking attempt. The additional information in semantic warnings may be of added benefit when novel situations occur. Any extra information about never experienced scenarios while driving may lead to better responses and more avoided accidents compared to generic auditory warnings that do not provide any context for the incoming critical event. More research on the effect of warning condition is needed to further explore these findings and provide additional advice for vehicle system development.

Reported Warning Urgency

Participants that responded that they noticed the warning on the previous question in the post experiment survey were also asked to give their level of agreement that the warning they received was urgent. The current study did not find a significant effect of level of automation on the reported urgency level of the received warning consistent with behavioral data. In contrast, the current study did find a significant effect of warning type on the reported urgency level of the received warning and a significant interaction between level of automation and warning type.

Hellier et al. (2002) proposed that warnings that contained speech/semantic information were more capable of expressing the urgency of the warning because the warning potentially includes a description of what the warning represents. Hellier et al.'s theory was partially supported in the current study by the significantly lower urgency rating for the non-semantic warning in the manual driving group. However, Hellier and colleagues' theory was not supported in the automated driving condition where the urgency ratings of both the semantic and non-semantic warnings were not significantly different.

Based on the self-report data from participants in the manual driving group, the semantic warnings were rated as urgent more frequently than the non-semantic warnings. The current results support this finding, suggesting that the semantic content within the auditory warning may lead to higher levels of perceived urgency compared to non-semantic warnings depending on the level of automation in use. In the current experiment, the implementation of a semantic warning lead to more urgent ratings when manual driving was in use, but not when automated driving was in use. This may suggest that semantic warnings are only beneficial in manual driving or that warnings are interpreted differently when automated driving is used. Further research is needed to understand how level of automation and warning type affect warning urgency in the driving domain.

Reported Knowledge of Vehicle Hacking

Another major question given to all participants in the experiment pertained to whether or not the participant was previously aware of the potential for vehicle hacking and the participants responded with a yes or no answer. Of the seventy-two participants that received the question, less than 20% had previously heard of vehicle hacking. This suggests that vehicle hacking is not an issue that most people may be aware of. This finding accords with McManus et al.'s (2018) finding wherein most of the participants were unaware of the potential for vehicle hacking unless told otherwise.

Next, the participants were asked whether they were more concerned with automated vehicle hacking, technical based automation failure, both, or neither. For participants in the manual and automated driving groups, most participants reported that they were concerned with both, but technical automation failure was the next most prevalent answer. The self-report data pertaining to concern suggest that while most participants had not heard of vehicle hacking previously, that it is of similar concern to technical automation failure. This result may be due to their recent experience with a simulated vehicle hacking attempt. While not conclusive, these results suggest that when participants are aware of the potential for vehicle hacking attempts it is an important concern they have with automated driving in the future.

Level of Automation and Maximum Steering Wheel Angle

In addition to the planned dependent variables, multiple exploratory measures of hacking attempt response quality were also collected. The first measure of response quality that was analyzed was maximum steering wheel angle (Shen & Neyens, 2017). The current study found a significant effect of level of automation on maximum steering wheel angle. Specifically, participants in the manual driving group had significantly smaller maximum steering wheel angle values than participants in the automated driving group. This finding suggests that the responses to the critical event were more severe when in the automated driving group than in the manual driving group.

The benefit of manual driving for maximum steering wheel angle may be due to a couple of reasons. For instance, the participants in the manual driving group had larger amount of previous experience with the simulator steering wheel even though participants in both groups were required to spend five minutes manually driving the vehicle in the practice portion of the experiment. In juxtaposition, it may be due to higher levels of SA for manual drivers. Regardless, this finding is consistent with the previous finding by Shen and Neyens (2017) and provides additional evidence that responses to critical events may be less severe for manual drivers than automated drivers. In conclusion, participants in the manual driving group had less severe responses to the vehicle hacking attempt based on maximum steering wheel angle. This finding is potentially due to past experience with the steering wheel or boosted levels of SA for manual drivers. Based on this finding we suggest that manual driving be used when vehicle hacking attempts are imminent in order to minimize the maximum steering wheel angle achieved when responding to a vehicle hacking attempt.

Warning Condition and Maximum Steering Wheel Angle

In addition to level of automation, maximum steering wheel angle was also exploratorily examined across warning condition. The results of the experiment showed an effect of warning condition that approached significance.

Though the effect only approached significance, semantic warning recipients had larger maximum steering wheel angles, potentially due to their understanding of the urgency of the situation. This suggestion is supported by the difference in how many people correctly identified the warning they received as a warning. Far less participants in the non-semantic warning condition reported that they received a warning at all. The higher frequency of understanding the warning and the recognition of its presence and urgency may have led to the more severe maximum steering wheel turn values for the semantic condition. This effect, that approached significance, may have been amplified by the participants lack of familiarity with the potential for vehicle hacking attempts.

Further research is needed to determine the effect of warning type on the safety of critical event response in order for recommendations to be made. Additional measures of response quality would provide further information about the quality of participant responses and provide further insight into how warnings affect driver’s response quality.

Maximum Steering Wheel Angle Over Time

In addition to maximum steering wheel angle, we examined maximum steering wheel angle over time. This measure may provide additional insight by incorporating how quickly the maximum wheel turn was performed with lower times suggesting more severe turns and potentially more dangerous driving. The current experiment did not produce a significant difference in maximum steering wheel angle over time between manual and automated driving.

The current experiment also did not find a significant difference for maximum steering wheel angle over time between the warning conditions. These null results are potentially due to the consistent seven second time from the onset of the warning until the end of the hack that resulted in a crash for the participants the vast majority of the time. In conclusion, there were no significant effects of warning type or level of automation conditions for maximum steering wheel turn over time which is potentially due to the relatively short time window for response of seven seconds. This finding may suggest that maximum steering wheel angle over time is not as useful as other measures of response quality, but further research using this dependent variable is needed to make that determination.

Maximum Resulting Acceleration

Maximum resulting acceleration is a measure of response quality calculated by using the lateral and longitudinal accelerations of the vehicle. The current experiment did not find a significant difference for maximum resulting acceleration due to level of automation. Maximum resulting acceleration was examined by Gold and colleagues (2013) and Shen and Neyens (2017) with varying results. Gold et al. (2013) found a significant effect of advanced warning time (5 seconds vs 7 seconds) for maximum resulting acceleration but did not find a significant difference between the manual and automated driving groups.

The current experiment did not produce a significant difference in maximum resulting acceleration between manual and automated driving. This is consistent with the result of Gold et al. (2013) who also did not find an effect of level of automation for maximum resulting acceleration. The current study did not measure maximum resulting acceleration at multiple points and used the largest maximum resulting acceleration for each participant which prevented direct comparison to the Shen and Neyens (2017) study.

The current experiment did not produce a significant difference in maximum resulting acceleration between the semantic and non-semantic warning conditions. It was not expected that there would be a significant difference for semantic and non-semantic warning based on the previous literature, but it was exploratorily analyzed since there was a significant effect for a warning related variable in the Gold et al. (2013) study. In conclusion, maximum resulting acceleration was not significantly affected by driving or warning condition in the current experiment which is neither consistent nor inconsistent with the past literature due to the varying results found in past experiments for maximum resulting acceleration.

Theoretical Implications

The level of automation and warning literatures contain numerous studies with differing interpretations of how the level of automation and warnings present in the vehicle affects participants responses to critical events. The current study sought to further explore this concept by manipulating the level of automation and warning type and implementing a novel and more difficult critical event scenario. The implementation of a novel and complex critical event was determined based on suggestions from previous literature and the growing potential for vehicle hacking. The results of the study suggest that level of automation and warning type do not significantly affect the method of response, number of successful responses, or TTC of the participants. The current study’s findings are in accordance with some past findings but not with others. This suggests that the effect level of automation and warning type have in critical events scenarios may vary depending on specific aspects of the critical event. The lack of consistent findings in this domain, including the current study, suggests that a clear theoretical understanding of how level of automation and warning type affect responses to critical events may not be present.

Past research has suggested that the theory of SA can be used to explain the differences due to level of automation and warning type and SA was the basis for the current study's expected results. SA was expected to be different based on the level of automation and warning used at the time of the critical event, but due to the current study not measuring SA, direct interpretations cannot be made. Future studies should include direct measures of SA in order to further understand the theoretical implication that level of automation and warning type have on SA in surface transportation. While SA was not directly measured in the current study, the results of the experiment may suggest that SA changes due to level of automation are not present during emergency response scenarios such as the vehicle hacking attempt used in the current study. If emergency scenarios result in SA not accurately explaining how participants responded, then further theory may be needed to explain participant responses in emergency scenarios.

In addition, the literature is unclear how warning type and level of automation affect SA when used simultaneously in response to a critical event. The current study found that early auditory warning implementation may have mitigated the potential difference in SA due to the level of automation present by capturing the participant's attention before the critical event. This led to very quick responses to the impending vehicle hacking attempt regardless of the level of automation in use. If preemptive auditory warnings lead to fast responses to the critical event, then the SA difference due to level of automation may be mitigated because the participant is already attending to their environment and are prepared to respond thanks to the preemptive warning they received. Further research is needed to understand how SA is affected by level of automation and warning type. When the warning is implemented may need to be accounted for in future research. SA is a complex concept and direct measurement of SA is needed in future studies to make clear interpretations of how different warning characteristics and levels of automation affect individuals SA during critical events.

Design Implications

The results of the current experiment were sporadic, and at times, not consistent with the level of automation, situation awareness, and warning literature in the surface transportation domain. Many of the expected differences were not found partly as a result of very large TTC values and a limited number of successful responses. Despite this, useful information can be gleaned in order to make suggestions for the design of future vehicle automation and warning systems. Firstly, the results of the experiment suggest that warnings given prior to the onset of the critical event lead to very high TTC values regardless of warning type or level of automation. Based on this finding, it may be beneficial to provide warnings in advance of the critical event when possible. Secondly, semantic warnings were found to be identified as warnings significantly more frequently than non-semantic warnings. Based on this finding, we suggest that semantic warnings should be used when auditory warnings are used to inform the driver that a critical event is about to occur. In addition, there was a benefit of semantic warnings when participants were asked to identify the cause of the warning. We suggest that when novel critical events, such as vehicle hacking attempts, are possible, semantic warnings should be used so the driver is more likely to appropriately diagnose the cause of the warning. When critical event response quality was measured using maximum steering wheel angle, we found a benefit of manual driving. Based on this finding, we suggest that manual driving should be implemented when complex vehicle responses are needed from the driver. In conclusion, multiple of the results from the current study suggest that manual driving and semantic warnings may be preferable when drivers need to respond to complex critical events, like vehicle hacking attempts. These findings should be taken into account when designing future warning and automation systems.

Study Limitations

The current experiment did not provide many of the expected significant differences that were predicted by the researcher's hypotheses and the previous literature. One major potential reason for this was the implementation of the warnings before the hack occurred which led to a ceiling effect and very high TTC values for all participants. Three of the experiments hypotheses pertained to TTC so the elevated TTC values across the board potentially led to the null results. The use of a low fidelity driving simulator could also have affected the results due to less tactile feedback pertaining to their control inputs and their effect on the vehicle compared to actual vehicles. Additionally, simulator use may have resulted in less actual perceived urgency and concern than in an actual vehicle due to a lack of genuine danger resulting from the vehicle hacking attempt.

There were also different amounts of participant in our experimental groups who reported previous knowledge of vehicle hacking. Those who were aware of vehicle hacking previously may have different responses than those who hadn't and the unequal distribution across experimental groups is a limitation of the study. The sample was also heavily female and African American and, as such, may not have been representative of the overall driving population. The study was also limited by the age of the participants who were all between the ages of 18-23 years old and as such does not accurately depict the overall driving population.

The novelty of vehicle hacking may also have played a part since most of the participants had never heard of vehicle hacking before the study. The fact that the participants likely never experienced any driving situation like this and its lack of similarities to past studies is a potential explanation for why the results of the current study did not match the previous literature or our hypotheses. Lastly, and potentially most importantly, the hacking attempt scenario presented to the drivers was difficult due to the complexity of the surrounding environment and the short time period given to the participants to maneuver through them. This may have led to the majority of participants failing to successfully respond to the hacking attempt despite the high TTC values found across all experimental groups.

Future Directions

The use of automation and warning in vehicles and the prevalence of vehicle hacking attempts are likely to continue to increase. Future research on level of automation and warning implementation can learn from this study's findings. Future studies may consider using a multiple vehicle hacking attempts and more diverse automation levels. The type of hacking attempt and its complexity is also a factor that could potentially be manipulated. Different warning types such as tactile warnings could be compared to auditory warnings and measures of urgency and pleasantness could be implemented. Importantly, the warning implementation time could also be manipulated such that warnings are implemented as the hacking attempt begins and compared to preemptive warnings like those used in this study. In addition, different driving populations and higher fidelity simulators may provide further insight to how level of automation and warning type affect driver responses to novel vehicle hacking attempts.

CHAPTER VI

CONCLUSIONS

The current experiment attempted to examine whether the level of automation and the content of the warning effected participants responses to a complex and novel take over response scenario in the form of a vehicle hacking attempt. The current results indicate that there was no significant effect of level of automation condition or warning condition on TTC values, successful response rate, or reaction method. These findings are contrary to the previous literature and are potentially due to various factors of the experimental scenario such as warning implementation time and hacking scenario complexity. Despite null results for the hypothesized analyses, multiple analyses found significant benefits for manual driving and semantic warnings as well as a significant interaction between level of automation and warning type for reported urgency. Additionally, preemptive warning implementation resulted in very high TTC values across the board suggesting warnings implemented prior to the critical event may result in quick responses to the warning. Further research is needed to clarify how level of automation and warning type interact and affect how participants respond to vehicle hacking attempts. However, the current study provides evidence that suggests manual driving and semantic warnings can be beneficial for warning interpretation and response quality when responding to a complex and novel critical event like a vehicle hacking attempt.

REFERENCES

Amoozadeh, M., Raghuramu, A., Chuah, C. N., Ghosal, D., Zhang, H. M., Rowe, J., & Levitt, K. (2015). Security vulnerabilities of connected vehicle streams and their impact on cooperative driving. *IEEE Communications Magazine*, *5*3, 126-132.

Baldwin, C. L. (2011). Verbal collision avoidance messages during simulated driving: perceived urgency, alerting effectiveness and annoyance. *Ergonomics*, *54*, 328-337.

Baldwin, C. L., & May, J. F. (2005). Verbal collision avoidance messages of varying perceived urgency reduce crashes in high risk scenarios. *Proceedings of the International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*, *3*, 128-133.

Baldwin, C. L., & Moore, C. (2002). Perceived urgency, alerting effectiveness and annoyance of verbal collision avoidance system messages. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *46*, 1848-1852.

Bansal, P., Kockelman, K. M., & Singh, A. (2016). Assessing public opinions of and interest in new vehicle technologies: an Austin perspective. *Transportation Research Part C: Emerging Technologies*, *67*, 1-14.

Bazilinskyy, P., Kyriakidis, M., & de Winter, J. (2015). An International Crowdsourcing Study into People's Statements on Fully Automated Driving. *Procedia Manufacturing*, *3*, 2534-2542.

Ben-Asher, N., & Gonzalez, C. (2015). Effects of cyber security knowledge on attack detection. *Computers in Human Behavior*, *48*, 51-61.

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.

Casley, S. V., Jardim, A. S., & Quartulli, A. M. (2013). *A study of public acceptance of autonomous cars*. (Unpublished bachelor's dissertation) Worcester Polytechnic Institute, Worchester, Massachusetts.

Chang, S. H., Lin, C. Y., Fung, C. P., Hwang, J. R., & Doong, J. L. (2008). Driving performance assessment: effects of traffic accident location and alarm content. *Accident Analysis & Prevention*, *40*, 1637-1643.

de Winter, J. C., 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.

Edworthy, J., Walters, K., Hellier, E., & Weedon, B. (2000). Comparing speech and nonspeech warnings. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *44*, 746-749.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, *37*, 32-64.

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

Endsley, M. R., & Garland, D. J. (2000). Theoretical underpinnings of situation awareness: A critical review. *Situation Awareness Analysis and Measurement* (pp. 1-24). Mahwah, NJ: Lawrence Erlbaum.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191.

Federal Bureau of Investigation. (2016, March 17) *Motor Vehicles Increasingly Vulnerable to Remote Exploits* [Public Service Announcement]. Retrieved from: https://www.ic3.gov/media/2016/160317.aspx.

Flemisch, F., Kelsch, J., Löper, C., Schieben, A., & Schindler, J. (2008). Automation spectrum, inner/outer compatibility and other potentially useful human factors concepts for assistance and automation. In D. de Waard, F.O. Flemisch, B. Lorenz, H. Oberheid, & K.A. Brookhuis (Eds.), *Human Factors for Assistance and Automation* (pp. 1-16). Maastricht, the Netherlands: Shaker Publishing.

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

Gontar, P., Homans, H., Rostalski, M., Behrend, J., Dehais, F., & Bengler, K. (2018). Are pilots prepared for a cyber-attack? A human factors approach to the experimental evaluation of pilots' behavior. *Journal of Air Transport Management*, *69*, 26-37.

Ghosh, D., & Vogt, A. (2012). Outliers: An evaluation of methodologies. In *Joint statistical meetings*, 3455-3460. San Diego, CA: American Statistical Association.

Gravetter, F. J., & Wallnau, L. B. (2005). Essentials of statistics for the behavioral sciences. Belmont. *CA: Wadsworth*.

Hellier, E., Edworthy, J., Weedon, B., Walters, K., & Adams, A. (2002). The perceived urgency of speech warnings: Semantics versus acoustics. *Human Factors*, *44*, 1-17.

Hergeth, S., Lorenz, L., & Krems, J. F. (2017). Prior familiarization with takeover requests affects drivers' takeover performance and automation trust. *Human factors*, *59*, 457-470.

Horswill, M. S., & McKenna, F. P. (2004). Drivers' hazard perception ability: Situation awareness on the road. In S. Barnbury & S. Tremblay (Eds.), *A Cognitive Approach to Situation Awareness: Theory and Application* (pp. 155-175). Aldershot, UK: Ashgate.

Kaber, D. B., & Endsley, M. R. (2004). The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. *Theoretical Issues in Ergonomics Science*, *5*, 113-153.

Kessel, C. J., & Wickens, C. D. (1982). The transfer of failure-detection skills between monitoring and controlling dynamic systems. *Human Factors*, *24*, 49-60.

Kyriakidis, M., Happee, R., & de Winter, J. C. (2015). Public opinion on automated driving: Results of an international questionnaire among 5000 respondents. *Transportation Research Part F: Traffic Psychology and Behaviour*, *32*, 127-140.

McKeown, D., & Isherwood, S. (2007). Mapping candidate within-vehicle auditory displays to their referents. *Human Factors*, *49*, 417-428.

McManus, W., Chen, J., Mishler, S., (2018). *Automated Vehicle Hacking: An Examination of Public Understanding and Priming.* Poster presented at 2018 American Psychological Association Technology Mind and Society Conference, Washington D.C.

Menon, N. (2015). *Consumer Perception and Anticipated Adoption of Autonomous Vehicle Technology: Results from Multi-Population Surveys* (Doctoral Dissertation).

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

National Research Council. (2009). *Technology, Policy, Law, and Ethics Regarding US Acquisition and Use of Cyberattack Capabilities*. National Academies Press.

Nunes, A., Reimer, B., & Coughlin, J. F. (2018). People must retain control of autonomous vehicles. *Nature*, *556*, 169-171.

Old Dominion University, Social Science Research Center. (2015). *Life in Hampton Roads Survey Press Release #4* [Press release]. Retrieved November 20, 2018, from https://www.odu.edu/content/dam/odu/col-dept/ssrc/docs/press-release-4-transportation-2015.pdf.

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

Petermeijer, S., Doubek, F., & de Winter, J. (2017). Driver response times to auditory, visual, and tactile take-over requests: A simulator study with 101 participants. In A. Basu, W. Perdrycz, & X. Zabuli (Eds.), Proceedings of the IEEE International Conferences on Systems, Man, and Cybernetics (SMC 2017) (pp.1505-1510). Piscataway, NJ: IEEE.

Politis, I., Brewster, S., & Pollick, F. (2015, April). To beep or not to beep?: Comparing abstract versus language-based multimodal driver displays. *Proceedings of the ACM Conference on Human Factors in Computing Systems*, *33*, 3971-3980.

Radlmayr, J., Gold, C., Lorenz, L., Farid, M., & Bengler, K. (2014). How traffic situations and non-driving related tasks affect the take-over quality in highly automated driving. In *Proceedings of the human factors and ergonomics society annual meeting*, *58*, 2063-2067.

Sawyer, S. F. (2009). Analysis of variance: the fundamental concepts. *Journal of Manual & Manipulative Therapy*, *17*, 27E-38E.

Shen, S., & Neyens, D. M. (2017). Assessing drivers' response during automated driver support system failures with non-driving tasks. *Journal of Safety Research*, *61*, 149-155.

Schoettle, B., & Sivak, M. (2014). Public opinion about self-driving vehicles in China, India, Japan, the US, the UK, and Australia. (Transportation Research Institute No. *UMTRI-2014-30*.) Retrieved from University of Michigan Library website: https://deepblue.lib.umich.edu/bitstream/handle/2027.42/109433/103139.pdf.

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

Taxonomy, S. A. E. (2018). Definitions for terms related to on-road motor vehicle automated driving systems. Technical report, SAE International.

Wetton, M. A., Horswill, M. S., Hatherly, C., Wood, J. M., Pachana, N. A., & Anstey, K. J. (2010). The development and validation of two complementary measures of drivers' hazard perception ability. *Accident Analysis & Prevention*, *42*, 1232-1239.

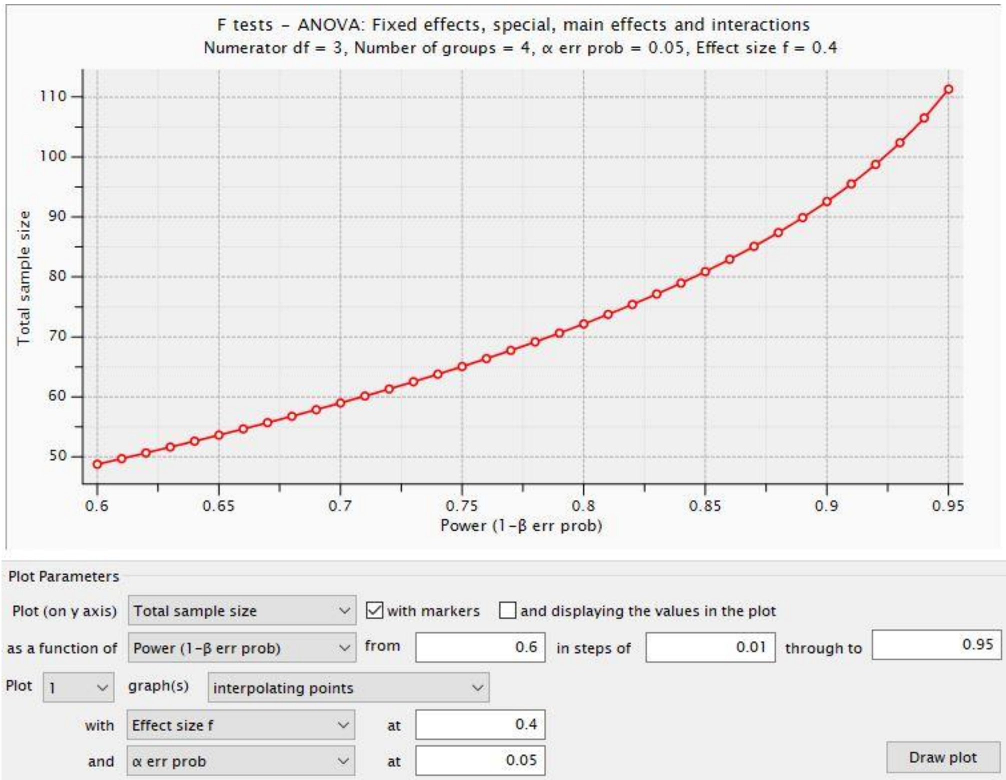
Wickens, C. D., & Kessel, C. (1979). The effects of participatory mode and task workload on the detection of dynamic system failures. *IEEE Transactions on Systems, Man, and Cybernetics*, *9*, 24-34.

Wickens, C. D., & Kessel, C. (1981). Failure detection in dynamic systems. *Human Detection and Diagnosis of System Failures*, *15*, 155-169.

Yamani, Y., & Horrey, W. J. (2018). A theoretical model of human-automation interaction grounded in resource allocation policy during automated driving. *International Journal of Human Factors and Ergonomics*, *5*, 225-239.

APPENDIX A

POWER ANALYSIS RESULTS



The power analysis (shown in the graph above) indicates that to achieve a statistical power of .80, 72 participants were needed with 18 in each group.

APPENDIX B

STIMULI AND APPARATUS SETUP

Appendix C

DEMOGRAPHICS FORM

				Gender		Ethnicity								
Date	Exp	Subject No.	Age	Male	Female	Hispanic/ Latino	American Indian/ Alaska Native	Asian	African/ African American	Caucasian	Native Hawaiian/ Pacific Islander	More than one race	Other/ Unknown	Dominant Hand

APPENDIX D

DRIVER EXPERIENCE QUESTIONNAIRE (ADAPTED FROM KYRIAKIDIS, HAPPEE, & DE WINTER, 2015)

1. What is your primary mode of transportation?

    1 Private Vehicle

    2 Public Transportation

    3 Motorcycle

    4 Walking/cycling

    5 Other

2. At what age did you receive your first driver's license?
3. On average, how frequently did you drive a vehicle in the last 12 months?

    1 Never

    2 Rarely

    3 Sometimes

    4 Often

    5 Everyday

4. Do you drive a vehicle to school or work?

    Yes

    No

5. About how many miles did you drive in the last 12 months? Please tell us in the box below.

6. How many accidents were you involved in in the last three years? Please tell us in the box below.

APPENDIX E

PARTICIPANT EXPERIMENTAL OPINION SURVEY

1. You were engaged in monitoring the automated driving system for the entire ride.

   1 Strongly Disagree
   2 Disagree
   3 Neither Agree or Disagree
   4 Agree
   5 Strongly Agree

2. Did you notice a warning during the drive?
   Yes
   No

3. You perceived the warning you received to be urgent.

   1 Strongly Disagree
   2 Disagree
   3 Neither Agree or Disagree
   4 Agree
   5 Strongly Agree

4. What did the warning you received mean? Please explain in the text box provided.
5. What did the warning you received represent?
   A. A system malfunction
   B. An automation failure
   C. A vehicle hacking attempt
   D. Low fuel level

6. What do you think the purpose of the study was? Please explain in the text box provided.

7. Were you aware of the potential for vehicle hacking before the experiment?
   Yes
   No

8. Are you concerned with automated vehicle hacking more or less than automation failure due to technical failure in the surface transportation domain?
   I am more concerned with automated vehicle hacking.
   I am more concerned with automation failure.
   Not concerned with either

9. What are your major concerns with automation in cars? Please list them in the text box provided.

VITA

Wyatt D. McManus

Old Dominion University
Department of Psychology
Norfolk, VA 23529

Education

Master's of Science/PhD in Human Factors Engineering at Old Dominion University, 2017- Present.

Advisor: Dr. Jing Chen

Proposed Master's Thesis: The Effects of Vehicle Automation Level and Warning Type on Responses to Vehicle Hacking.

Bachelor's of Science in Psychology from Purdue University (West Lafayette), 2016

Advisors: Dr. Robert Proctor, Dr. James Nairne, & Dr. Jeffrey Karpicke

Research Experience

Graduate Student/Researcher, Old Dominion University

Working in the HAC Lab with Dr. Jing Chen I am designing, creating, and analyzing experiments to investigate human automation interaction in the surface transportation domain.

Post Graduate Research Assistant, Purdue University 2016-2017

Working for Dr. Jeffrey Karpicke I was able to participate in learning and memory research, more specifically retrieval process effects, concept mapping, complex learning, and testing effects.

Working for Dr. James Nairne I was able to participate in evolutionary memory research, more specifically survival processing effects and adaptive memory research.

Working with Dr. Robert Proctor and Dr. Nicole Murchinson I was able to participate in researching how hand posture affects outside interference during Eriksen flanker tasks.

Research Interests

My current research is in the autonomous vehicle domain. Specifically, I am interested in the way that humans interact with autonomous vehicle systems during a cyberattack and how that is affected by driving modality. Additionally, I am interested in cyber security research. Specifically, I am interested in the effect of password strength feedback on password creation.

Publications/Presentations

McManus, W., Chen, J., & Mishler, S. (2018). Automated Vehicle Hacking: An Examination of Public Understanding and Priming. Poster presented at 2018 American Psychological Association Technology Mind and Society Conference, Washington D.C.