Driver's Arousal and Workload Under Partial Vehicle Automation

September 2020

607 14th Street, NW, Suite 201 | Washington, DC 20005 | 202-638-5944

Title

Driver's Arousal and Workload under Partial Vehicle Automation

(March 2020)

Authors

David L. Strayer, Joel M. Cooper, David M. Sanbonmatsu, Gus G. Erickson, Trent G. Simmons, Amy S. McDonnell, Amanda E. Carriero, Kaedyn W. Crabtree, and Monika Lohoni

University of Utah

Foreword

As more active driving systems are integrated into vehicles, the role and responsibilities of drivers using these technologies will change fundamentally from conventional vehicles. Vehicle automation now offers extended control of vehicle's speed, headway and lane position, but calling upon drivers to continuously monitor the road and traffic environment. This report summarizes an on-road study of drivers' workload, arousal and attentiveness when driving vehicles equipped with Level 2 automation.

Results summarized in this report should help researchers, automobile industry and government entities better understand driver performance, behavior and interactions in vehicles with advanced technologies.

C. Y. David Yang, Ph.D.

Executive Director
AAA Foundation for Traffic Safety

About the Sponsor

AAA Foundation for Traffic Safety
607 14th Street, NW, Suite 201
Washington, D.C. 20005
202-638-5944
www.aaafoundation.org

Founded in 1947, the AAA Foundation for Traffic Safety in Washington, D.C., is a nonprofit, publicly supported charitable research and education organization dedicated to saving lives by preventing traffic crashes and reducing injuries when crashes occur. Funding for this report was provided by voluntary contributions from AAA/CAA and their affiliated motor clubs, individual members, AAA-affiliated insurance companies, and other organizations or sources.

Table of Contents

Abbreviated Terms

ACC	Adaptive cruise control
BAC	Blood alcohol concentration
BPM	Beats per minute
DRT	Detection Response Task
ECG	Electrocardiography
EEG	Electroencephalography
EOC	Electrooculogram
FISMA	Federal Information Security Management Act
HIPAA	Health Insurance Portability and Accountability Act
IRB	Institutional Review Board
ISO	International Organization for Standardization
LED	Light-Emitting Diode
NHTSA	National Highway Traffic Safety Administration
NTSA	National Transportation Safety Board
PANAS	Positive and Negative Affect Schedule
PVT	Psychomotor vigilance task
RMSSD	Root mean square of successive differences
SAE	Society of Automotive Engineers

Executive Summary

The Society of Automotive Engineers defines six levels of vehicle automation, ranging from Level-0, no automation, to Level-5, full automation (SAE, 2016). Currently, vehicles with Level-2 automation are equipped with systems that maintain lateral and longitudinal control. One requirement for Level-2 automation is for drivers to remain engaged with the driving task and monitor the environment at all times. In the case of automated control system failure or limitation, the driver must be able to safely regain control of the vehicle. The current research compared and contrasted the driver's experience when operating a vehicle under Level-2 automation with that of operating the same vehicle when automation was not engaged (i.e., Level-0).

This study involved an on-road evaluation of drivers between the ages of 21-64 operating up to four different vehicles that supported Level-2 automation on both urban and more rural roadways. The research design was 4 (Vehicle: Cadillac CT6, Nissan Rogue, Tesla Model 3, Volvo XC90) x 2 (Age: 21-42 vs. 43-64) x 2 (Automation: Level-0 vs. Level-2) x 2 (Interstate: I-80 vs. I-15) factorial with 24 younger and 24 older participants tested in each of the vehicles.

A variety of dependent measures were collected to assess driver's arousal and workload as they operated these vehicles. These included measures of reaction time and hit rate in the Detection Response Task (DRT), power in the parietal alpha spectra of electroencephalographic (EEG) brain activity, electrocardiographic (ECG) measures of heart rate and heart rate variability, and survey measures assessing nervousness, inattention, and excitement. Data from the DRT, EEG, ECG, and survey results were analyzed using linear mixed-effects models.

Consistent with role and responsibility of drivers described in the SAE framework, the data provide evidence for slightly *enhanced* engagement with the driving task and monitoring of the driving environment when participants operated the vehicle under Level-2 automation. Under those conditions, parietal alpha and DRT hit rates were lower and DRT reaction was longer compared to manual driving conditions, suggesting that participants paid more attention to the driving environment. Participants also reported driving under partial (Level-2) automation to be more exciting and resulting in more nervousness.

Though statistically significant, the magnitude of the differences between Level-0 and Level-2 automation was small accounting for — at most — 2.7% of the variance across measures. Even so, all the dependent measures trended in the direction of more attention directed to the driving task under Level-2 automation. There were no meaningful differences in arousal or workload between Level-0 and Level-2 automation. If there were meaningful differences, the study was sufficiently powered to detect them.

The current research protocol required that a researcher was present in the vehicle, ensuring that the data were accurately recorded, that the participant completed all specified conditions, and that safety standards were met. It is possible that the participant's behavior may change in the absence of a researcher or data monitoring equipment. Future research should examine the driver's behavior over much longer intervals to determine if this pattern is altered with greater experience with Level-2 automation (i.e., do driver transition from a "Novelty" ownership phase to "Post Novelty" and then to "Experienced" phases as suggested by Dunn, Dingus, & Soccolich, 2019). If the same pattern of data is

observed across longer intervals of exposure, it would suggest that drivers are able to sustain their attention to the driving task and monitor the driving environment. However, it is possible that additional familiarity with Level-2 automation could decrease driver excitement and nervousness and increase levels of trust and overconfidence.

Introduction

Several manufacturers offer systems that can automate the control of vehicle speed, following distance, and lane position (e.g., Tesla-*Autopilot*, Nissan-*ProPilot*, Volvo-*Pilot Assist*, Cadillac-*Super Cruise*, and Audi-*Traffic Jam Pilot*). By automating lateral and longitudinal vehicle control, these systems have the potential to reduce or eliminate single-vehicle and rear-end highway crashes. However, the technology is not perfect and requires drivers to maintain vigilance towards the traffic environment and readiness to take over vehicle control at all times. Ironically, the changing role of drivers from active controller to passive monitor may gradually erode driver attention, arousal, and readiness to respond which could potentially undermine possible safety benefits. Several research articles have been published on automation and its effects on arousal and sustained attention. However, very little research has looked at how the real-world use of partial vehicle automation affects driver workload and physiological arousal. Results presented in this report suggest that drivers may be better than expected at maintaining adequate attention to the driving task and that their arousal levels during partial automation are similar to manual driving conditions.

Levels of Automation

Commercially available vehicle automation generally encompasses technologies that are designed to automate lateral and longitudinal vehicle control. These technologies are often, though not always, marketed for use in highway environments, including Interstate use and stop-and-go traffic. SAE J3016 (2016) specifies six levels of vehicle automation ranging from 0 (no automation) to 5 (full automation) (See Figure 1). Vehicles with both adaptive cruise control and lane-centering technology that can be simultaneously engaged fall under SAE Level-2 automation. There are currently several Level-2 vehicles for sale in the United States and Tesla has announced that near-term versions of Autopilot, to be released in 2020, will be “Feature Complete” and support “Full Self Driving.” Whether these updates constitute Level-3 performance remains to be seen. Thus far, there are no true Level-3 vehicles on the market. However, Tesla’s Navigate on Autopilot and Audi’s Traffic Jam Assist come the closest.

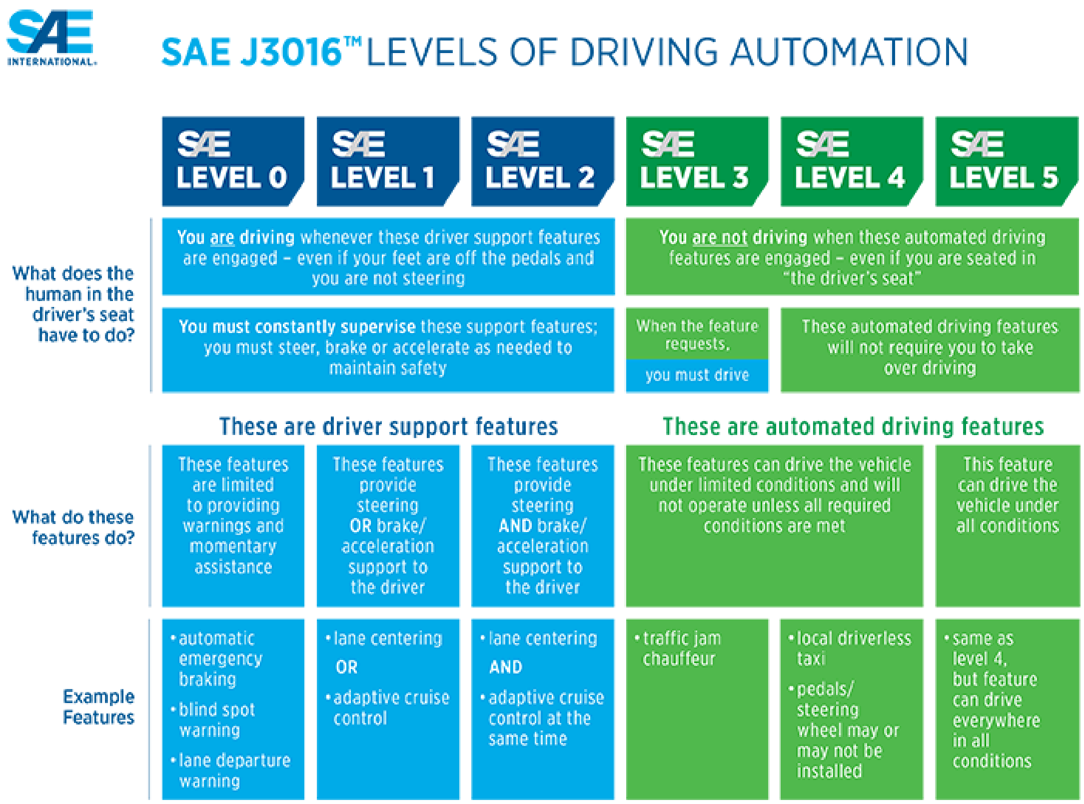

Figure 1. Levels of automation as specified by the Society of Automotive Engineers (SAE, 2016)

A requirement of Level-2 automation is that the operator is always monitor the driving task. To ensure that a minimum level of driver engagement and attention is present, vehicle manufacturers employ a variety of driver monitoring approaches that range from camera-based face and eye tracking (e.g., GM Super Cruise) to steering torque sensors (e.g., Tesla and Volvo). Vehicles also use different sensors to monitor the driving environment (e.g., LiDAR, Radar, Sonar, Machine Vision, GPS, etc.) with varying integration strategies. This results in wide variability in vehicle performance for sensing and responding to the driving environment. The impact of these differences on driver workload, arousal, and performance is not well understood.

A further differentiating factor of Level-2 vehicles is the way the systems are activated, how they communicate their state to the driver, and their engagement/disengagement strategies. Together, these design differences result in varying experiences for the driver with potential consequences for workload, arousal, and performance.

Attention Management

The Yerkes-Dodson model suggests that the relationship between attention, arousal and performance can be specified by an inverted-U shaped function (Yerkes & Dodson, 1908), where a theoretically optimal level of performance is attained with a moderate level of arousal (Cohen, 2011). Within this framework, attention directly moderates arousal whereby too much attention can ratchet up arousal — leading to increased stress and anxiety with poorer performance outcomes and disorganized behavior (Derakshan & Eysenck, 2009).

Driving situations that demand high and sustained levels of attention can lead to rapid performance declines, increased stress, and anxiety (Langner & Eickhoff, 2013). Performance under these conditions may be hindered and distractibility heightened, especially in the presence of task-irrelevant or threatening information (Derakshan & Eysenck, 2009). Adverse effects of over-arousal in the monitoring of partial vehicle automation have been shown to increase the onset of fatigue and hasten vigilance decrement (Greenlee, DeLucia, Newton, 2019). Conversely, lack of driver stimulation may also lead to fatigue, absent-mindedness, disengagement, and complacency (Manly, Robertson, Galloway, & Hawkins, 1999). In extreme cases, drivers may even fall asleep at the wheel, a major cause of driving accidents and fatalities (Fischer, 2016; Tefft, 2012).

As noted, within SAE automation Level-2, drivers are required to be ready and able to take over vehicle control at any time. However, with increasing automation, the role of drivers changes from active controller to passive monitor, which may make it more difficult for drivers to maintain optimal arousal. The extent to which driving under Level-2 automation may lead to over or under-arousal is not well understood. However, several high profile crashes, online videos, and preliminary research suggests a heightened potential for distraction and fatigue in incidents involving vehicle automation (National Transportation Safety Board [NTSB], 2017, 2018; see also Gaspar & Carney, 2019).

Safety

A 2017 report by the National Highway Traffic Safety Administration (NHTSA) suggests that Tesla vehicles equipped with automatic steering are 40% less likely to crash than those not equipped with the feature (Habib, 2017). In October of 2018, Tesla began voluntarily releasing quarterly crash statistics that compile data taken from their advanced vehicle telematics recording systems. Plotted in Figure 2, these data are provided for cases where Autopilot is engaged with other Active Safety Systems, where Autopilot is disengaged but other Active Safety Systems are engaged, and where neither Autopilot nor the Active Safety Systems are engaged. Tesla also reported the NHTSA released statistics on police-reported crashes for comparison ("Tesla Vehicle Safety Report", 2019). Notably, Tesla data suggests that drivers were more than two times less likely to crash when Autopilot is engaged, that crash rates with Autopilot are decreasing, and that Tesla drivers are less likely to crash, in general, than the US national average.

Several agencies have pointed out limitations on the numbers that Tesla has released, including that Tesla drivers are not representative of US drivers and that Tesla's data fails to account for roadway type (Krisher, 2018). A reanalysis of the NHTSA 2017 Tesla findings by the Quality Control Systems Corporation (2019) revealed several methodological issues with the analysis. A correction of these issues reversed the claims made by NHTSA and suggested a relative increase in crash risk with automatic steering. Whether partial automation increases or decreases crash rates in an open and evolving issue. Given constant improvements to computing power, crash avoidance technology and Artificial Intelligence, it is clear that hard crash numbers will remain elusive as technology continues to advance.

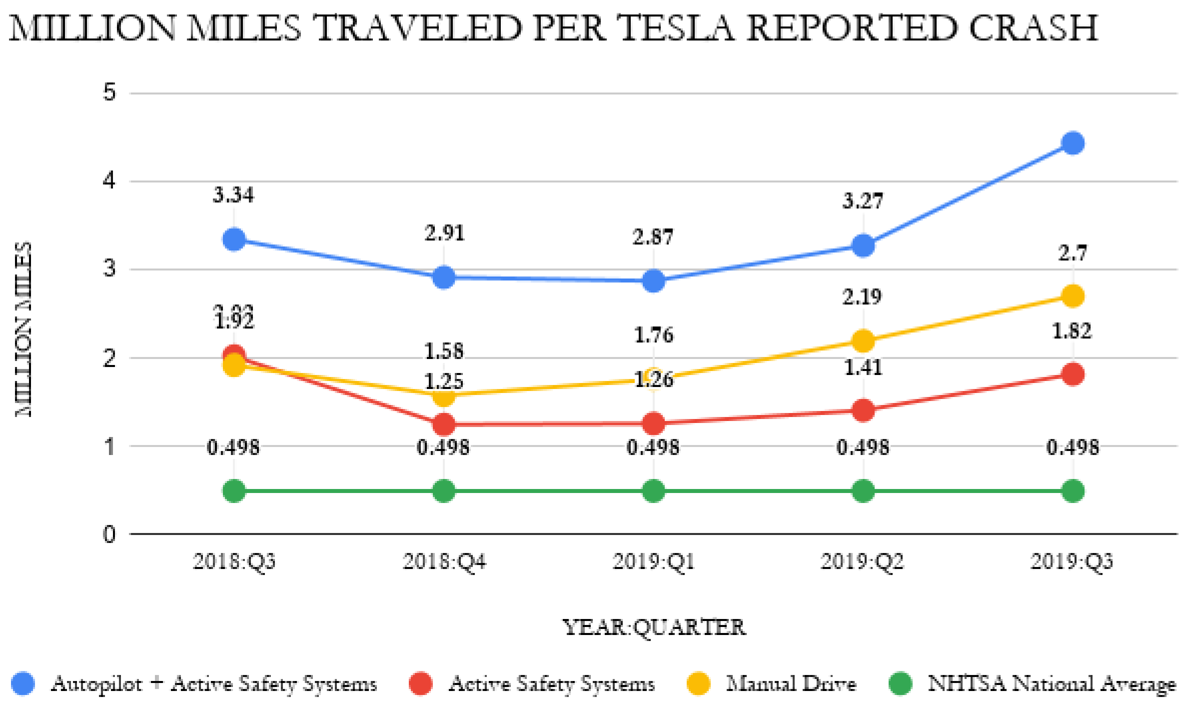

Figure 2. Million miles traveled per crash for Tesla vehicles using combinations of Autopilot and Active Safety Systems. Data are provided by Tesla quarterly. https://www.tesla.com/VehicleSafetyReport

A related issue is the changing role of the driver with increasing automation and the way in which drivers engage with automation. Video instrumentation of privately owned vehicles suggests that drivers engage Level-2 systems 34.8% of the time and that, contrary to general findings on vigilance decrement in automation (e.g., Grier et al., 2003; Raz & Buhle, 2006), drivers continue to actively monitor the roadway even when Level-2 automation is engaged (Fridman et al., 2019; Hatfield et al., 2019). A similar set of findings has been published by the National Advanced Driving Simulator researchers at the University of Iowa. Their research suggests that drivers monitor the roadway similarly in both Level-0 and Level-2 driving modes (Gaspar & Carney, 2019) but that drivers tended to increase the frequency of long glances away from the roadway with partial automation. Furthermore, Gaspar and Carney found that Level-2 systems were used during almost 80% of highway driving. Surveys provided to drivers that experience Level-2 systems suggest that trust increases with use and that familiarity can decrease stress and increase feelings of security (Endsley, 2017; Gaspar & Carney, 2019).

A study by the Virginia Tech Transportation Institute in 2019 (Dunn, Dingus, & Soccolich, 2019) suggests that drivers may transition through several phases of interaction with vehicle automation, moving from an initial “Novelty” phase to a “Post-Novelty” phase and finally to an “Experienced” phase. Dunn et al. suggest that each of these phases can be characterized by a different set of driving behaviors that gradually transition from high system experimentation, high arousal and high vigilance to over-reliance, low arousal and low vigilance. Initial support for the theory is provided through a complementary set of naturalistic driving studies that looked at behavior in a group of mostly Level-1 and some Level-2 automated vehicles. In these studies, drivers either agreed to have their own vehicle instrumented with a video recording system, or borrowed a vehicle provided by the research team. In both cases, all vehicles supported Adaptive Cruise Control and most supported some form of Lane Keeping Assist (i.e., Level-1 vehicle automation). Some of the vehicles also provided Lane Centering (i.e., Level-2 automation when coupled with ACC). Baseline driving epochs, where automation was not engaged, were selected to match participant and weather conditions. Results suggest that drivers that were lent vehicles for a month maintained vigilance during Level-1/Level-2 engagement but that drivers who owned their vehicle were more likely to engage in secondary task distraction events when automation was active. While these results are suggestive, drivers are likely selective about when to engage automation and the performance and attention requirements of driving under Level-1 and Level-2 automation are likely quite different.

Current Study

The aim of the current study is to compare and contrast driver's arousal and workload under no automation (Level-0) and partial automation (Level-2). In order to maximize the generalizability of these findings, the study involved an on-road evaluation of drivers between the ages of 21-64 operating up to four different vehicles that supported Level-2 automation on both urban and more rural roadways. Prior research found that older adults often struggle with advanced electronics in the vehicle (Cooper et al., 2019), yet if such technology is implemented correctly and users are given adequate training, this cohort of drivers may benefit the most from its usage. To provide a comprehensive assessment of driver's arousal and workload, we use a diverse set of physiological, performance, and subjective measures. Below, each of these measures and their respective response to over and under-arousal is presented.

The Detection Response Task (DRT) is a standard research tool that has been developed and validated to measure the cognitive workload of drivers (International Organization of Standardization [ISO] 17488, 2016). The DRT is a simple stimulus-response task that requires drivers to press a microswitch attached to their finger each time they feel a small vibration or detect the illumination of an LED. The DRT is sensitive to driving difficulty (Bruyas & Dumont, 2013), visual demands of secondary tasks (Cooper, Castro, & Strayer, 2016), and general cognitive load (Bengler, Kohlmann, & Lange, 2012). The Psychomotor Vigilance Task (PVT), an older task with a structure like the DRT, has been validated as a measure of vigilance (Gartenberg et al., 2012; Temple et al., 2000; Yarnell, Lopresti, Lanlokun, & Balkin, 1982), fatigue (Drummond et al., 2005), and arousal (Gunzelmann et al., 2007). Stimulus-response tasks like the DRT and PVT often provide two outcome measures, reaction time and hit rate. As workload increases, reaction time goes up and hit rate goes down (Young, Hseih, & Seaman, 2013). Similarly, increased fatigue, sleep deprivation, and reduced vigilance also lead to increased reaction time and reduced hit rate in the PVT. The DRT is sensitive to measuring both over and under-arousal but cannot effectively discriminate between the two states.

Electrocardiography (ECG) is a method for measuring the electrical activity of the heart. Evaluations of heart rate and heart rate variability are perhaps the most common approaches to characterize the systematic ECG waveform (Agrafioti, Hatzinakos, &

Anderson, 2012). Several studies have established that both heart rate and heart rate variability are sensitive to over and under-arousal in driving (Jiao, Li, Chen, Wang, & Qi, 2004; Mehler, Reimer, & Wang, 2011; Tozman, Magdas, MacDougall, & Vollmeyer, 2015). A comprehensive review by Lohani, Payne, and Strayer (2019) found that heart rate decreases with drowsiness, as vigilance decreases, and with the use of partial automation. Additionally, heart rate variability increases with drowsiness, fatigue, and disengagement. By contrast, heart rate increases with workload and stress and heart rate variability decreases with workload, stress, and heightened vigilance.

Electroencephalography (EEG) is the measurement of the summated electrical activity of the brain, recorded non-invasively from electrodes on the scalp. EEG provides high temporal resolution (within milliseconds) and therefore a direct record of neural activity in real-time (see Lohani, Payne, & Strayer, 2019). Oscillatory components of EEG can be decomposed into canonical spectral frequencies (e.g., Delta ~0.5-4 Hz, Theta ~4-8 Hz, Alpha ~8-12 Hz, and Beta ~12-30 Hz) using Fourier analysis (Cohen, 2014). These frequency bands have been studied in relation to various neurocognitive functions. For example, research has shown that fatigue increases alpha power (Chuang et al., 2018; Käthner, Wriessnegger, Müller-Putz, Kubler, & Halder, 2014), while an increase in arousal and cognitive workload suppresses alpha power (Mun, Whang, Park, & Park, 2017). Therefore, the power in the alpha frequency band can be measured to index changes in arousal and cognitive demand while driving. Alpha power in the parietal regions of the brain has also been studied in relation to selective visual attention. Higher alpha power suggests lower visual engagement with the environment whereas lower alpha power suggests higher visual engagement (Bowman et al., 2017; Foxe & Snyder, 2011). Parietal alpha power is highest when an individual's eyes are closed (Goldman, Stern, Engel, & Cohen, 2002; see Appendix A for additional information).

Several survey questions help to supplement, support, and disambiguate DRT, EEG, and ECG measures. Survey measures in the current study centered around constructs of nervousness, inattention, and excitement help to understand the driver's experience as they operate vehicles under Level-0 and Level-2 automation.

If driving under Level-2 vehicle automation decreases arousal through fatigue, drowsiness, and disengagement, we would have expected to see increased reaction time and reduced hit rate (due to lapses of attention), reduced heart rate, increased heart rate variability, and increased EEG Alpha power, coupled with lower self-reported nervousness, increased inattention, and decreased excitement. Conversely, if Level-2 automation leads to increased arousal and attention toward the driving task, we would have expected to see increased DRT reaction time and reduced hit rate (due to increased attention directed towards driving). Other expected outcomes would have been increased heart rate and reduced heart rate variability, reduced EEG parietal Alpha power, higher self-reported nervousness, decreased inattention, and increased excitement (see Table 1).

Table 1. *Expected direction of Level-2 automation (compared to Level-0 automation) for each dependent measure used in the current research for over and under-arousal.*		
	Under Arousal	Over Arousal
Parietal Alpha	↑	↓
Heart Rate	↓	↑
RMSSD*	↑	↓
DRT-Reaction time	↑	↑
DRT-Hit Rate	↓	↓
Nervousness	↓	↑
Inattention	↑	↓
Excitement	↓	↑
Note: *Root Mean Squared of Successive Differences		

Method

Participants

Seventy-four participants (27 female) were recruited via University of Utah Institutional Review Board (IRB) approved flyers and by referral from previous participants. Seventy-one participants were included in the final analysis; data from three were lost due to equipment malfunctions.

The younger cohort, ages 21-42 ($M_{age} = 28.82$, $SD_{age} = 6.41$), was made up of 39 participants (13 females) and the older cohort, ages 43-64 ($M_{age} = 52.72$ y, $SD_{age} = 6.33$), was made up of 32 participants (12 female). This broad range of age groups was recruited in order to ensure a representative sample of participants across the age range. All participants had a valid driver's license, no at-fault accidents in the past two years, had driven at least an average of ten hours per month, had sustained no injuries that caused a loss of consciousness in the last two years, had no history of neurological disorders, had no heart conditions, and were not pregnant or planning to become pregnant. Upon arrival at the University of Utah, participants confirmed that they had normal or corrected to normal vision, six or more hours of sleep the previous night, a valid driver's license on their person, and a BAC level at 0.00% (verified using a BACtrac® breathalyzer).

Following University of Utah policy, participants were required to pass a 20-minute online defensive driving course and certification test. To ensure participants held a clean driving record and were eligible to participate in the study, a motor vehicle record report was obtained by the University of Utah's Division of Risk Management prior to enrollment. Compensation was prorated at $20 per hour. Participants were awarded a $10 bonus after completing their second session, a $15 bonus after completing their third session, and a $25 bonus after completing their fourth session.

Four vehicles were selected to compare and contrast their Level-2 automation capabilities, resulting in a 4 (Vehicle) x 2 (Age Cohort) x 2 (Level of Automation) x 2 (Interstate) factorial design. In total, 71 participants (39 from the younger cohort and 32 from the older cohort) were tested. Some participants were tested in more than one vehicle (on different days). As shown in Table 2, 24 of the participants drove in all four vehicles, 19 drove three out of the four vehicles, 11 drove in two of the four vehicles, and 17 drove in only one out of the four vehicles. This planned missing data design (e.g., Graham, Taylor, Olchowski, & Cumsille, 2006; Little & Rhemtulla, 2013) was used to facilitate participant recruitment and was appropriately handled by the data analysis.

Table 2. *Number of vehicles driven by a participant broken down by age cohort.*

Age Cohort	*Tested in 1 Vehicle*	*Tested in 2 Vehicles*	*Tested in 3 Vehicles*	*Tested in 4 Vehicles*
Younger	11	8	11	9
Older	6	3	8	15

Note: e.g., 9 younger and 15 older participants were tested in all 4 of the vehicles.

Stimuli and Apparatus

Vehicles

A 2018 Cadillac CT6 Premium Luxury with Super Cruise, 2019 Nissan Rogue SL Premium with ProPILOT Assist, 2018 Tesla Model 3 AWD/Long Range with Autopilot, and a 2018 Volvo XC90 Momentum with Pilot Assist were used in this study. Each of these vehicles were equipped with the partial vehicle automation that centered the vehicle within the lane (e.g., Lane Centering) and maintained following distance and speed (e.g., Adaptive Cruise Control). Activated together, these features fulfilled the SAE requirements for Level-2 vehicle automation.

Equipment

Detection Response Task (DRT). A vibrotactile variant of DRT was used to probe driver attention. A small vibration motor was taped to the participant's left bicep and a microswitch was attached to either the index or middle finger of the left hand so that it could be depressed against the steering wheel when the participant noticed the onset of the stimulus. The left bicep, rather than the standard left collarbone placement, was selected in order to reduce potential interference with the ECG recordings. Millisecond resolution response time to the vibrotactile onset was recorded via an embedded microcontroller and stored on the host computer (See ISO 17488, 2016).

Following the ISO 17488 guidelines (2016), the vibrotactor emitted a small vibration stimulus, similar to a vibrating cell phone. This stimulus was presented pseudo-randomly every 3-5 seconds. Participants were instructed to respond as quickly as possible to the stimulus by pressing the microswitch against the steering wheel. If no response was made within 1 second, the vibration stimulus turned off.

Electrocardiography. Electrical activity generated by the heart was recorded using Biopac® Smart Center systems at a 2000 Hz sampling rate. Electrodes were placed on the bottom of the left and right ribs, and on the right collarbone (3-Lead system). The electrode placement areas were sterilized with an alcohol swab and Signa Gel was applied to the electrode before being placed on the participant.

Electroencephalography. Electrical activity generated by the brain was recorded using Biopac® Smart Center systems at a 2000 Hz sampling rate. Three passive, NATUS Neurology Grass reusable EEG electrodes were placed along the midline—frontal (Fz), central (Cz), and parietal (Pz)—as well as a ground electrode on the center of the forehead and a reference electrode on the right mastoid bone. Two electrooculogram (EOG) electrodes were placed above and below the right eye (vertical). These electrodes captured muscle movement of the eye to record eye blinks for later data processing. Scalp electrode placement was based on the International 10-20 system (Jasper, 1958).

Eye electrode placement was determined by aligning the electrodes with the middle of the pupil. The electrode placement areas were sterilized with an alcohol swab and then prepped with NuPrep Skin Prep Gel and cotton swabs to clear away dead skin cells at each electrode site. Ten20 adhesive gel and medical tape were used to secure the electrodes to the scalp, forehead, mastoid, and face. Impedance levels were kept below 10kΩ, as verified using BIOPAC's EL-CHECK® electrode impedance checker. The participant's field of view and range of motion were not impeded when wearing the electrodes.

Other Equipment. An iPad mini 4 was used to capture all survey responses which were hosted on REDCap, a 21 CFR Part 11, FISMA, and HIPAA-compliant data capture repository (see Appendix B). A Garmin Virb XE action camera was mounted on the dashboard of each vehicle and used to collect video of the road. Each vehicle was also outfitted with a Logitech Pro Webcam on the dashboard to collect video of the participant's face during data collection. Dell XPS 13 laptops were used to record incoming physiological measures and administer the DRT. A BACtrac® breathalyzer was used to measure BAC levels.

Procedure

Experimental Design

The experimental design was a 4 (Vehicle: Cadillac CT6, Nissan Rogue, Tesla Model 3, Volvo XC90) x 2 (Age: 21-42 vs. 43-64) x 2 (Automation: Level-0 vs. Level-2) x 2 (Interstate: I-80 vs. I-15) factorial with seventy-one participants included in the analyses. Each participant was tested in between 1 and 4 vehicles (see Table 2). Linear mixed models are specifically designed to accommodate this form of missing data (i.e., they are appropriate for unbalanced designs with different numbers of observations for different participants).

Surveys and Physiological Set-up

Prior to arrival, participants reviewed a training document for the specific vehicle they were expected to drive and watched a short video about the vehicle's partial automation. Upon arrival, participants were re-screened for eligibility, shown the driving routes, and reviewed the training document and video. Participants then completed a series of seven questionnaires on a tablet (see Appendix B). They first indicated their demographic characteristics, health, and state of fatigue on the Demographic Questionnaire and Brief Compound Fatigue Instrument. This was followed by the nine-item Beliefs about Automated Driving Systems Survey which measured their evaluations and expectations of the attributes and performance of advanced driving assistance systems. Participants finished by filling out the 12-item Proneness to Boredom Scale (Vodanovich, Wallace, & Kass, 2005), the 15-item short form of the Barratt Impulsivity Scale (Spinella, 2007), and the Big Five Inventory-2 (Soto & John, 2017). These personality scales were administered to explore individual differences predicting arousal and attention during Level-0 and Level-2 driving. While filling out these questionnaires, participants were simultaneously fitted with EEG, ECG, and a respiration belt around their chest.

Training

After all of the surveys were completed and the physiological equipment was attached, the participant and research administrator walked to the testing vehicle where the participant would familiarize themselves with the vehicle by adjusting the mirrors, seat, steering wheel, etc. Meanwhile, the research administrator ensured the safety of the vehicle by wiping any debris from external cameras and sensors, confirming that the horn, wipers, and blinkers functioned properly, and ensuring that tire pressure levels were adequate. Before driving the vehicle, participants sat in the driver's seat and were trained on how to activate the partial automation of the specific vehicle. Before leaving, 4 minutes of eyes-closed baseline EEG data were recorded. Participants were then instructed to drive toward the training route.

As shown in Figure 3, two routes were chosen for data collection. One route (I-80) was a low trafficked, curvy Interstate, shown in blue. The other route (I-15) was a more heavily trafficked, relatively straight Interstate, shown in green. The training route (I-215), shown in red, was driven prior to the experimental session. Participants traveled along all routes at the speed limit or 5 mph slower with a research administrator present in the passenger seat to monitor for safety and data collection.

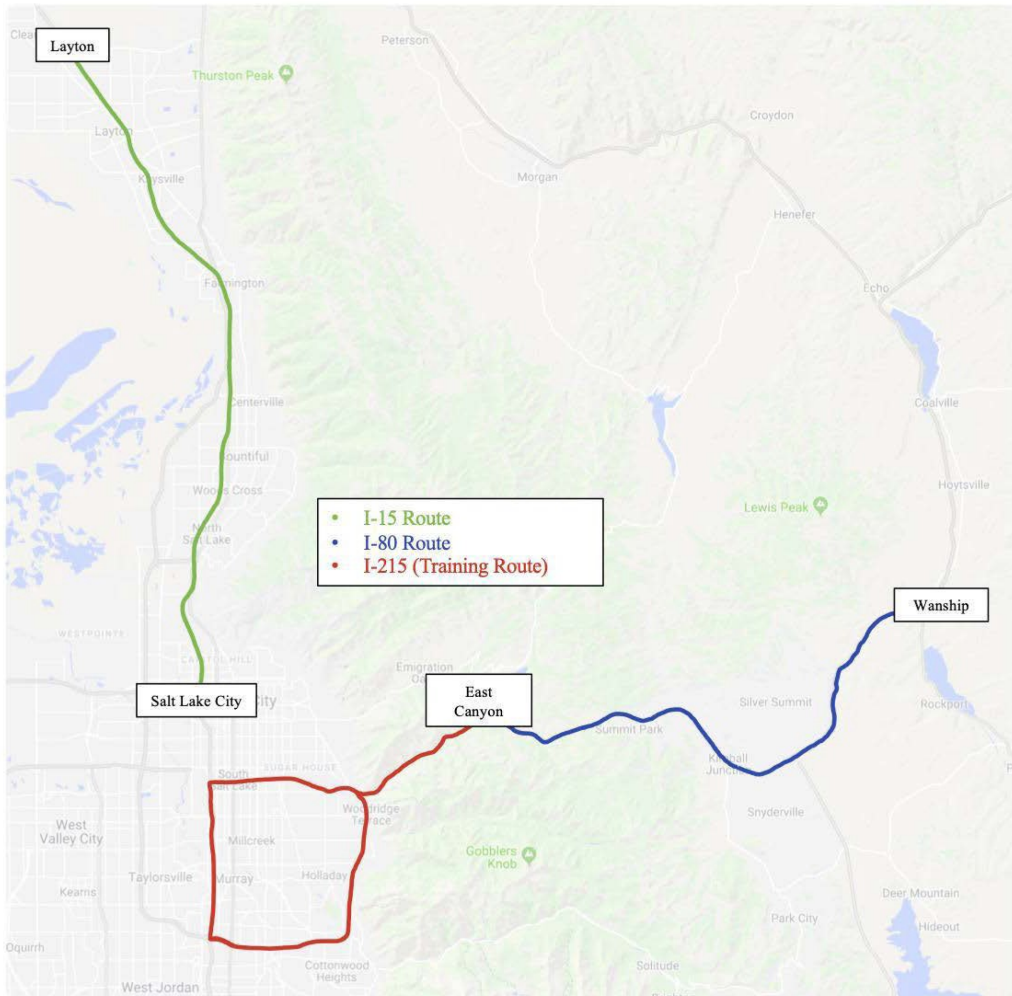

Figure 3. A map of the driving routes used in the study

Participants drove southbound from the University of Utah to I-215 where the training route began. Once on the freeway and traveling the speed limit, the research administrator instructed participants to engage Level-2 partial automation. Once engaged, the participant would practice multiple ways to disengage, change lanes, increase and decrease speed, and experience system warnings. The research administrator verbally verified with the participant that they were confident interacting with the system's features and were comfortable engaging the system on their own. The participant used the remaining time on the training route to ask questions and gain further exposure to the partial automation in the vehicle. Participants were randomly assigned to a counterbalance that determined which route, I-15 or I-80, would be driven first, as well as the order of the driving condition; either under partial automation or under no automation (i.e., Level-2 and Level-0 automation, respectively).

Both routes, I-15 and I-80, were split into two experimental blocks. For each route, the first experimental drive consisted of driving away from Salt Lake City, while the second consisted of driving back toward Salt Lake City. Participants were instructed to drive the vehicle either with partial automation engaged or without partial automation during each experimental drive. Each experimental drive began once the participant was up to speed, had passed any construction zone or object on the shoulder of the road, and felt safe and ready to begin testing. Each experimental drive was approximately 20 minutes in length. The first two minutes consisted of a baseline period containing only the driving task, which was followed with the addition of the DRT that continued until the end of the experimental drive.

Experimental Blocks

Immediately prior to the first experimental drive, participants were fitted with the DRT microswitch on either their index or middle finger of their left hand and the vibrotactor taped to their left bicep as per ISO 17488 specifications (2016). To train the participant on how to respond to the vibrotactor, the research administrator monitored the participant while they performed 10 practice trials with response times under 500 milliseconds while the vehicle was parked. After exhibiting a full understanding of the DRT and the operation of the vehicle, the participant drove toward the freeway to begin data collection.

Participants drove in two separate conditions: with partial automation engaged (Level-2), or without automation (Level-0). When in the partial automation condition, participants were instructed to engage the vehicle's partial automation when it was safe to do so. While driving in the no automation condition (Level-0), participants were instructed not to engage with any of the vehicle's partial automation capabilities (e.g. Adaptive Cruise Control or Lane Keep Assist).

I-15 Route. The first experimental drive on this route started near a gas station in Salt Lake City (Exit 307 on I-15) and began when the participant was driving northbound on I-15 toward Layton, Utah, approximately 27 miles away. Data collection ended ½ mile from exit 337 for 700 S. A break was then given at a nearby gas station. The second experimental drive on this route began once the participant was driving southbound on I-15. Data collection ended ½ mile from exit 307 for 400 S, where the research administrator and participant would stop again for a break. If the I-80 route had already been completed, the research administrator would drive the participant back to the University of Utah. If the I-80 route had not been completed, the participant and research administrator would have lunch before continuing to the second route.

I-80 Route. A 5-minute section of the road was used to familiarize the participant with how the partial automation in the vehicle behaved on curvy sections of the road. This occurred before data collection on I-80. The first experimental drive on this route started near East Canyon (Exit 134 on I-80), and began when the participant was driving eastbound on I-80 towards Wanship, Utah, approximately 21 miles away. Data collection ended ½ mile from exit 155 for Wanship. A break was then given at a nearby gas station. The second experimental drive on this route began once the participant was driving westbound on I-80. Data collection ended ½ mile from exit 134 for East Canyon. If the I-15 route had already been completed, the research administrator would drive the participant back to the University of Utah. If the I-15 route had not been completed, the research administrator would drive from a golf course to a gas station in Salt Lake City for a break where the participant and research administrator would have lunch before continuing to the second route.

At the end of each experimental drive, participants completed a modified version of the PANAS (Watson, Clark, & Tellegen, 1988) that assessed both their affective states and attentional processes during driving (see Appendix B). They were then given a ten-minute break. After all experimental drives were complete, participants filled out the Driving Experiences and Intentions Measure and Evaluations of Vehicle Automated Driving Systems questionnaires while the research administrator drove the participant back to the University of Utah.

Each experimental drive was collected in safe and normal driving conditions. Participants were instructed to abandon the DRT if disruptions to the natural driving environment occurred. Participants resumed the DRT when they felt it was safe to do so. The research administrator flagged scenarios "Do not use" to exclude from data processing due to miscellaneous reasons (e.g. debris on the road, heavy traffic, equipment malfunctions, etc.). Data collection ceased during any construction zones or when other hazards occurred on the road.

Data Handling and Dependent Measures

Detection Response Task (DRT)

DRT data were processed following guidelines outlined in ISO 17488 (2016). All response times under 100 milliseconds or over 2500 milliseconds were eliminated from the overall calculation for reaction time. Non-responses or responses that were made after 2.5 seconds from the stimulus onset were coded as misses. The DRT-related dependent measures used in the study are:

- Reaction time: Defined as the sum of all valid reaction times to the DRT divided by the number of valid reaction times.
- Hit rate: Defined as the number of correct and valid responses divided by the total number of stimuli presented during each condition.

Electrocardiography (ECG)

Heart rate is typically measured in the number of heart beats per minute. Heart rate data were cleaned using BIOPAC Acqknowledge (5.0) scripts, in conjunction with a visual inspection from research assistants. This process began with a band-pass filter with a low-frequency cutoff of 1 Hz and a high-frequency cutoff of 35 Hz. Each heart beat was marked at its highest peak (known as QRS peaks), and were used for analysis. QRS peaks were then identified, marked and kept for analysis. To ensure clean data, the heart rate channel was visually inspected for QRS anomalies. In the event of a missing peak, one was interpolated so as to be in the middle of the two surrounding QRS peaks. The two dependent measures extracted from the ECG data are:

- Heart rate: Defined as the mean number of beats per minute, calculated using sequential 60-second epochs.
- Heart rate variability: Defined as the root mean square of successive differences (RMSSD)

Electroencephalography (EEG)

EEG data were cleaned and processed using MATLAB and the EEGLAB (Delorme & Makeig, 2004) toolbox. EEG recordings were downsampled to 250 Hz and band-pass filtered with a low-frequency cutoff of 0.1 Hz and a high-frequency cutoff of 30 Hz. Data were then segmented into 1-second epochs. Ocular artifacts were corrected for using Gratton's method of eye movement correction in order to retain epochs that were contaminated by artifacts (Gratton, Coles, & Donchin, 1983). The average power at each frequency between 1 and 30 Hz was extracted using a Fast Fourier Transform (see Appendix A for details). Power in the alpha band (between 8-12 Hz) was then averaged across epochs, resulting in a separate estimate for each participant in each Vehicle, Level of Automation, and Interstate.

- Parietal alpha power: Defined as the average power in the alpha band between 8-12 Hz at the parietal (Pz) electrode.

Surveys

The 18 items from the Affect and Attention checklist were factor analyzed using principal component analysis (see Appendix B). The Kaiser-Meyer-Olkin measure of sampling adequacy was .72 and Bartlett's test of sphericity was significant, $\chi^2(153) = 518.26$, $p < .001$, indicating that the data were suitable for structure detection. An initial factor analysis indicated that 3 factors accounted for a total of 52.9% of the variance for the entire set of variables and provided the most interpretable solution. The eigenvalues leveled off after the 3 factors.

An oblimin rotation was performed because the factors were expected to be correlated. A total of 5 items were not included in the 3-factor solution. Irritable, alert, and unrelated activities were eliminated because they did not have a primary factor loading of over 0.40. Comfortable and bored were not included because they had loadings over 0.40 on two factors.

The first factor accounted for 24.9% of the variance and was labeled "Nervousness." It consisted of 5 items (loadings in parentheses) – scared (-.843), nervous (-.827), distressed (-.789), confident (.644), and relaxed (.640). The second factor accounted for 17.2% of the variance and was labeled "inattention" because all of the items assessed directly or indirectly the level of inattention to driving. "Inattention" also consisted of 5 items – daydreamed (.887), thoughts wandered (.866), thoughts about problems (.674), focused on road (-.611), and sleepy (.459). Finally, factor three explained 10.9% of the variance and was labeled "Excitement". This factor was based on excited (.786), enthusiastic (.771), and cheerful (.535).

The three dependent measures from the factor analysis are:

- Nervousness: the latent variable in the factor analysis loading on the ratings of "scared," "nervous," "distressed," "confident," and "relaxed" obtained in the survey.
- Inattention**:** the latent variable in the factor analysis loading on the items that assessed directly or indirectly the level of inattention to driving in the survey.
- Excitement**:** the latent variable in the factor analysis loading on the ratings of "excited," "enthusiastic," and "cheerful" obtained in the survey.

Results

Data from the DRT, ECG, EEG, and Survey results were analyzed using linear mixed-effects models in R (R Core Team, 2019; version 3.6.0). Models were run using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015; version 1.1-21). The repeated subjects factor was input into each model as a random intercept with other factors of interest alternately entered as fixed effects. Interactions between conditions were analyzed through the comparison of models where factors were specified as interactive to models where factors were specified as additive. A model comparison approach using a Likelihood Ratio Test was then used to assess the goodness of each model fit. These evaluations returned p-values which compared the full linear effects model to a partial model (see Winter, 2013). This statistical approach has the advantage of analyzing all available data while adjusting fixed effects, random effects, and likelihood ratio test estimates for missing data. Summary characteristics of the main effects and interactions are presented below.

Electroencephalography

Power in the alpha frequency band over the parietal regions of the brain was extracted from the EEG data in order to assess visual engagement and arousal. We expected that if driving in Level-2 automation increases visual engagement and arousal, parietal alpha power would decrease. On the other hand, if driving in Level-2 automation decreases visual engagement and arousal due to fatigue or drowsiness, parietal alpha power would increase (Bowman et al., 2017; Foxe & Snyder, 2011).

All EEG data files were manually inspected offline in order to assess the quality of the data. The inspector was blind to the condition of each file. Some individual files were lost due to either an electrode falling off, electrical interference between the wireless EEG transmitters and the BioPac Smart Center, or a human error in data collection. After offline inspection, the usable dataset consisted of 182 eyes-closed files, 168 I-15 Level-0 automation files, 167 I-15 Level-2 automation files, 172 I-80 Level-0 automation files, and 171 I-80 Level-2 automation files. The linear mixed-effects models utilized in the analysis accounted for the missing data.

Figures 4-7 present the main effects of Vehicle, Age, Automation, and Interstate on parietal alpha. Table 3 presents the results of the multi-level modeling. There were significant effects of Automation and Interstate.

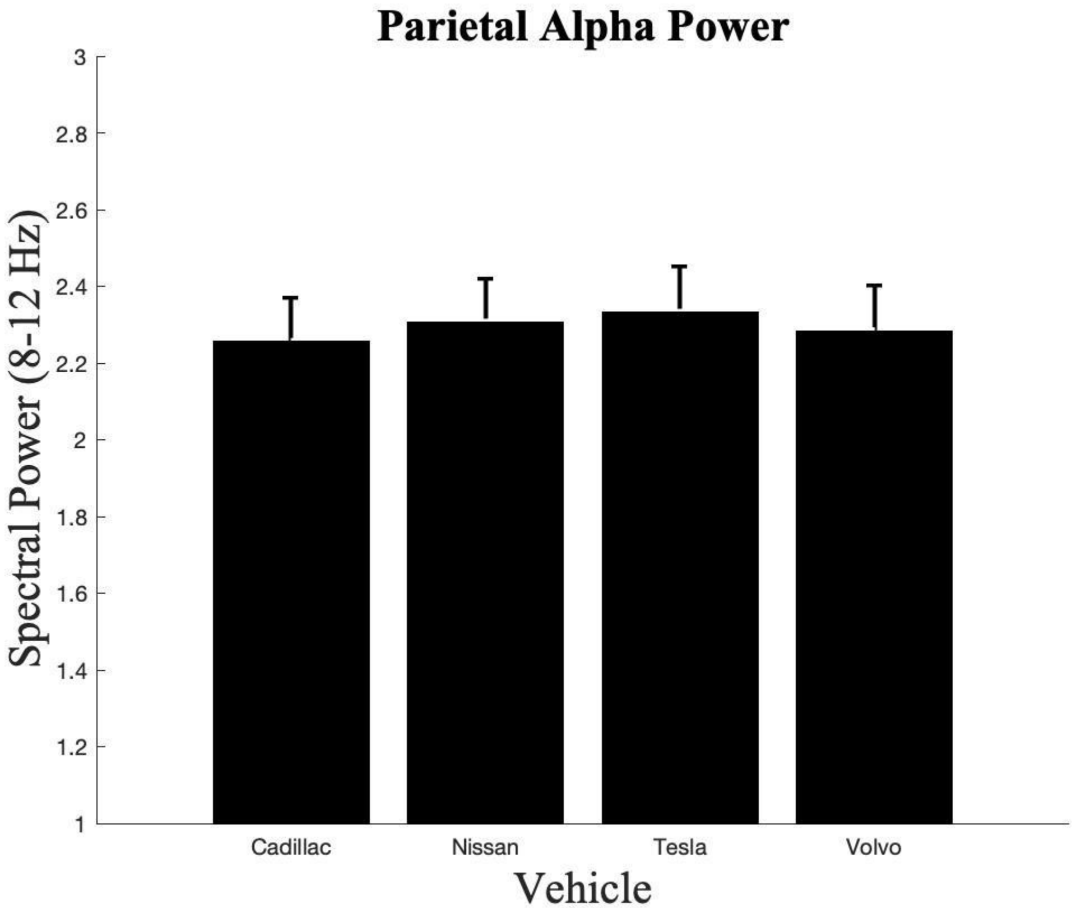

Figure 4. Parietal alpha plotted as a function of the vehicle. Error bars reflect the standard error of the mean. The effect of Vehicle (Cadillac: M = 2.26, SD = 1.45; Nissan: M = 2.31, SD = 1.45; Tesla: M = 2.33, SD = 1.50; Volvo: M = 2.28, SD = 1.50) was not significant (χ^2 (3) = 2.58, p = 0.46; R^2 = 0.001).

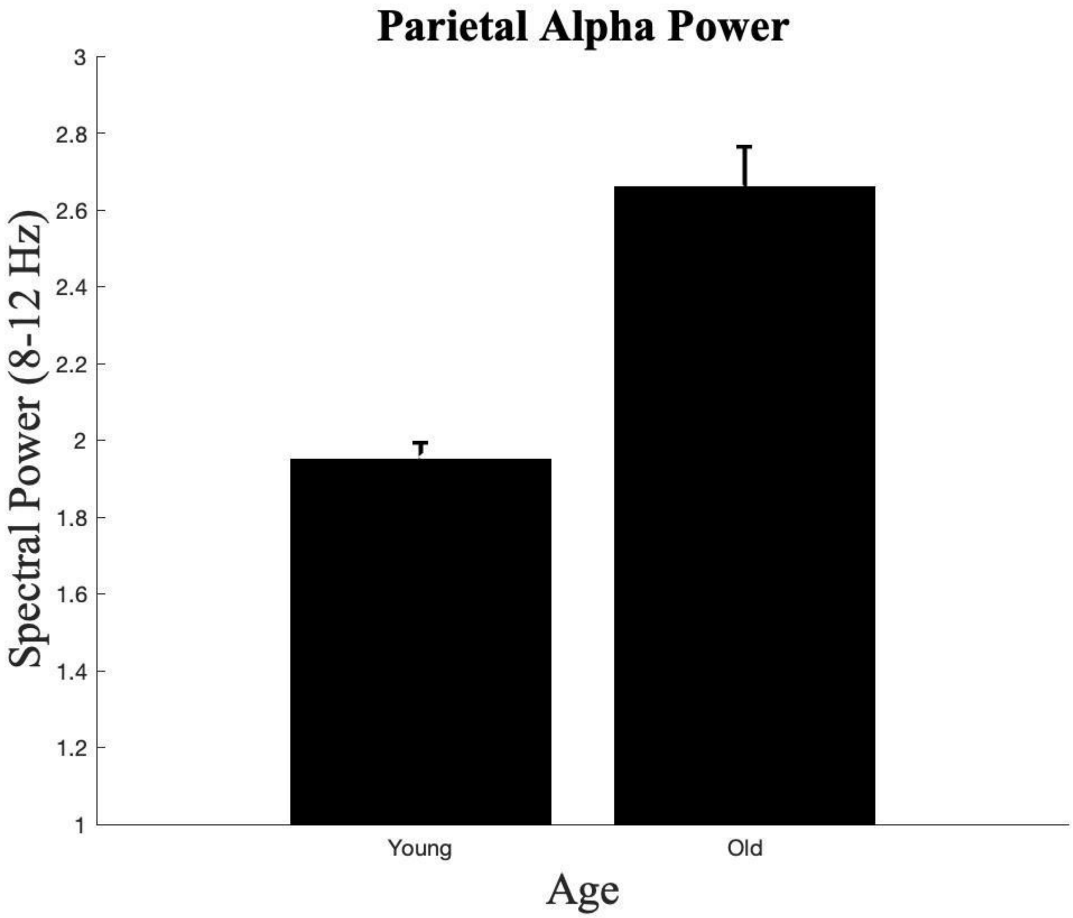

Figure 5. Parietal alpha plotted as a function of the participant's age. Error bars reflect the standard error of the mean. The effect of Age (Young: M = 1.95, SD = 0.80; Old: M = 2.66, SD = 1.88) was not significant (χ^2 (1) = 1.73, p = 0.19; R^2 = 0.02). The lack of significance was the result of entering the participant as a random intercept.

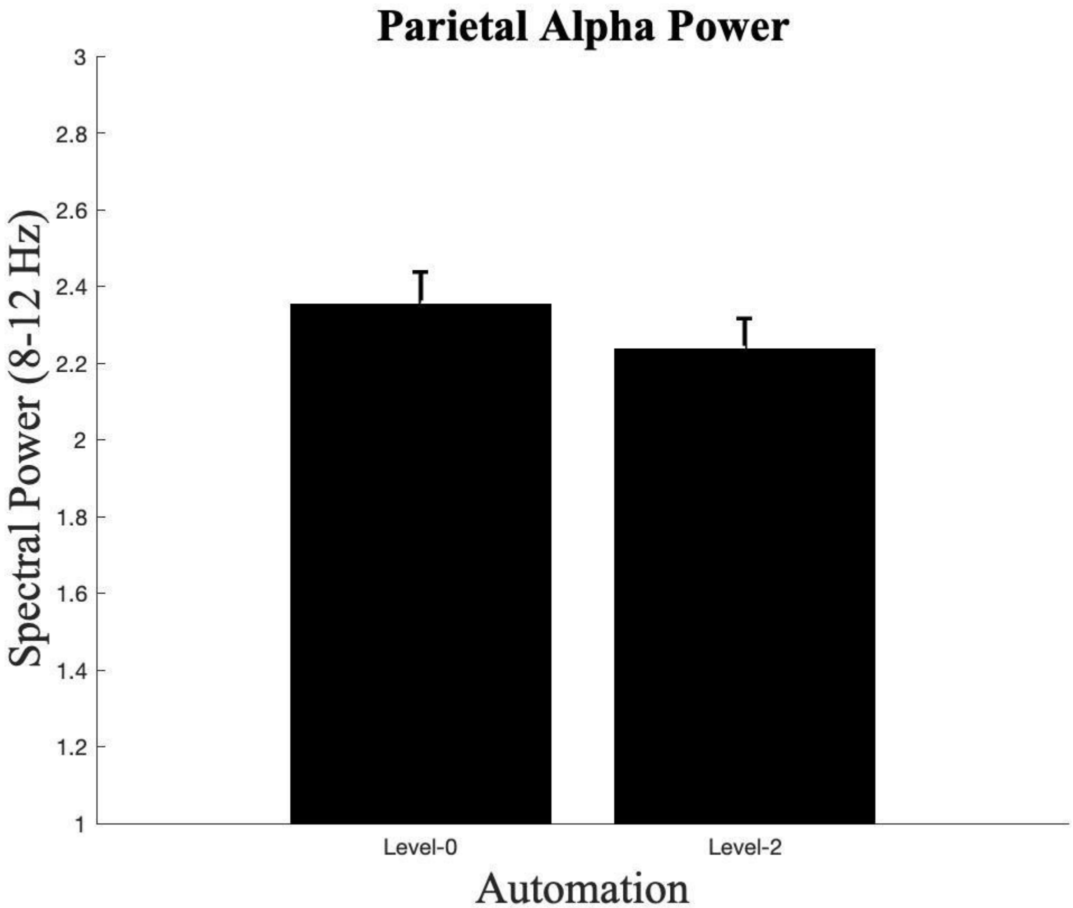

Figure 6. Parietal alpha plotted as a function of the level of automation. Error bars reflect the standard error of the mean. The effect of Automation (Level-0: M = 2.35, SD = 1.50; Level-2: M = 2.24, SD = 1.44) was significant (χ^2 (1) = 3.87, p = 0.049; R^2 = 0.001).

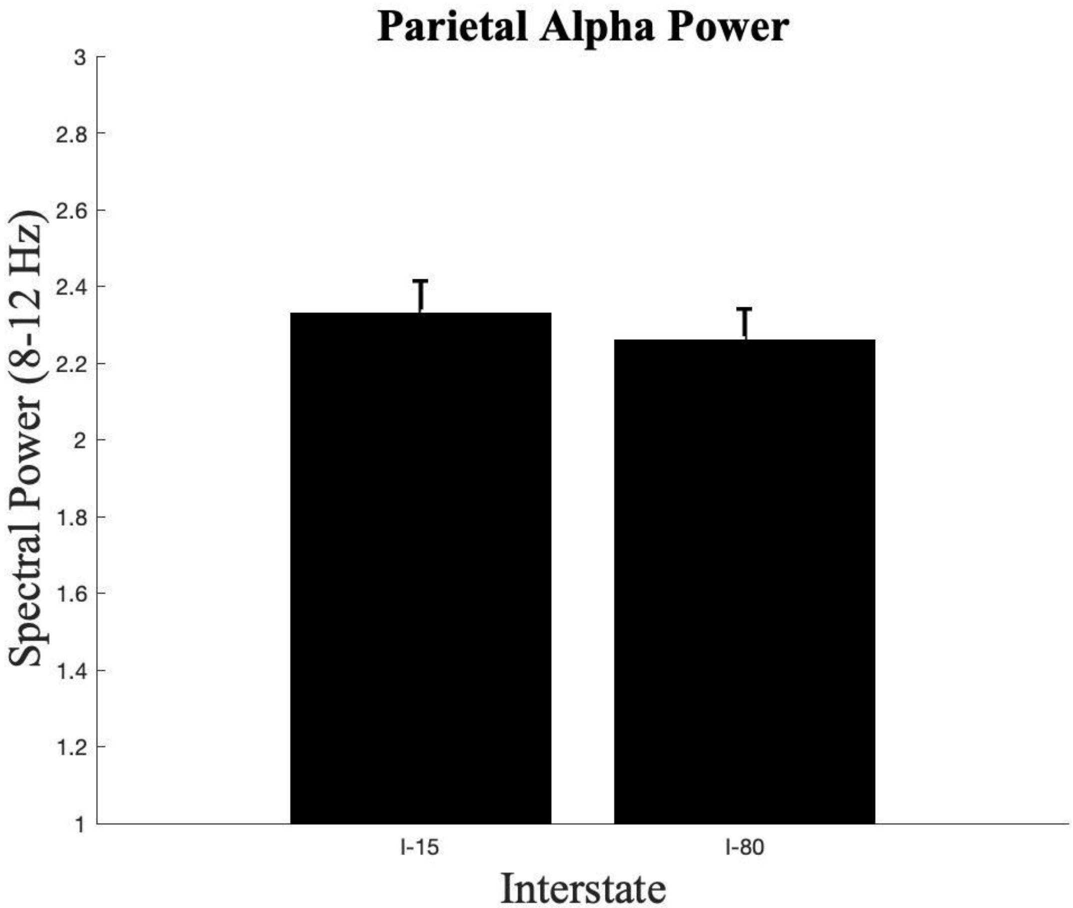

Figure 7. Parietal alpha plotted as a function of the Interstate. Error bars reflect the standard error of the mean. The effect of Interstate (I-15: M = 2.33, SD = 1.49; I-80: M = 2.26, SD = 1.46) was significant ($\chi^2(1) = 5.14$, $p = 0.02$; $R^2 = 0.001$).

Table 3. *Parietal Alpha*

Variable	*Chi-square*	*df*	*p*	R^2
Vehicle	2.578	3	0.461	0.001
Age	1.734	1	0.188	0.022
Automation	**3.874**	**1**	**0.049***	**0.001**
Interstate	**5.144**	**1**	**0.023***	**0.001**
Vehicle x Age	5.789	3	0.122	0.001
Vehicle x Automation	0.701	3	0.873	0.000
Vehicle x Interstate	4.458	3	0.216	0.000
Age x Automation	1.059	1	0.303	0.001
Age x Interstate	0.895	1	0.344	0.000
Automation x Interstate	0.016	1	0.900	0.000
Vehicle x Age x Automation	8.783	10	0.553	0.002
Vehicle x Age x Interstate	12.370	10	0.261	0.002
Vehicle x Automation x Interstate	6.440	10	0.777	0.001
Age x Automation x Interstate	2.005	4	0.735	0.001
Vehicle x Age x Automation x Interstate	19.170	25	0.789	0.003

Note: R Squared values represent the Marginal R^2 difference score from the comparison model.
Significance Code: `*` < .05.

Electrocardiography

Two types of data were extracted from the electrocardiograph (ECG) record: heart rate and heart rate variability. A total of 965,362 point estimates of heart rate (sampled once per second) were generated from the Biopac software. A first pass removal of outliers trimmed results that were less than 40 bpm or greater than 120 bpm. This cleaning removed 3,104 (0.3%) of the point estimates from the full data set. A second pass removed data that were +/- 3 Standard Deviations from the mean for each subject. This removed a further 7,160 (0.7%) of data. After this cleaning procedure, cells for each condition (Vehicle x Age x Automation x Interstate) contained an average of 1,237 (SD = 278) heart rate point estimates.

Heart rate variability, measured by Root Mean Squared of Successive Differences (RMSSD), was calculated in blocks of 60 seconds. A total of 12,573 blocks of data were generated for analysis. Results that were +/- 3 Standard Deviations from a subject mean were removed. This removed 58 (0.4%) 1-minute summary blocks. Remaining blocks were aggregated across each condition (Vehicle x Age x Automation x Interstate). An average of 16 (SD = 2.7) blocks were aggregated per condition to generate the cell means for analysis.

Heart Rate

Figures 8-11 present the main effects of Vehicle, Age, Automation, and Interstate on heart rate. Table 4 presents the results of the multi-level modeling. There were significant effects of Vehicle and Age.

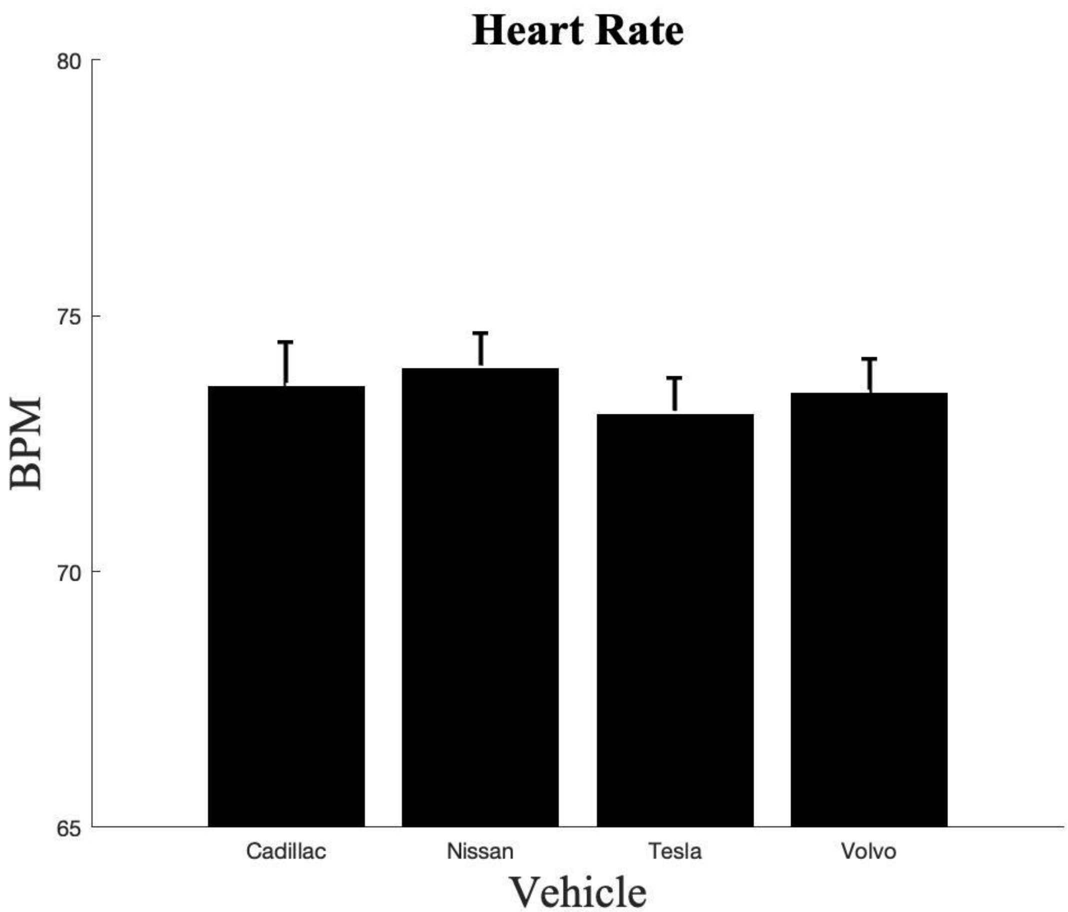

Figure 8. Heart rate, expressed as beats per minute (BPM), plotted as a function of the vehicle. Error bars reflect the standard error of the mean. The effect of Vehicle (Cadillac: M = 73.62, SD = 11.61; Nissan: M = 73.97, SD = 9.62; Tesla: M = 73.07, SD = 9.52; Volvo: M = 73.49, SD = 8.92) was significant (χ^2 (3) = 10.29, p = 0.016; R^2 = 0.004).

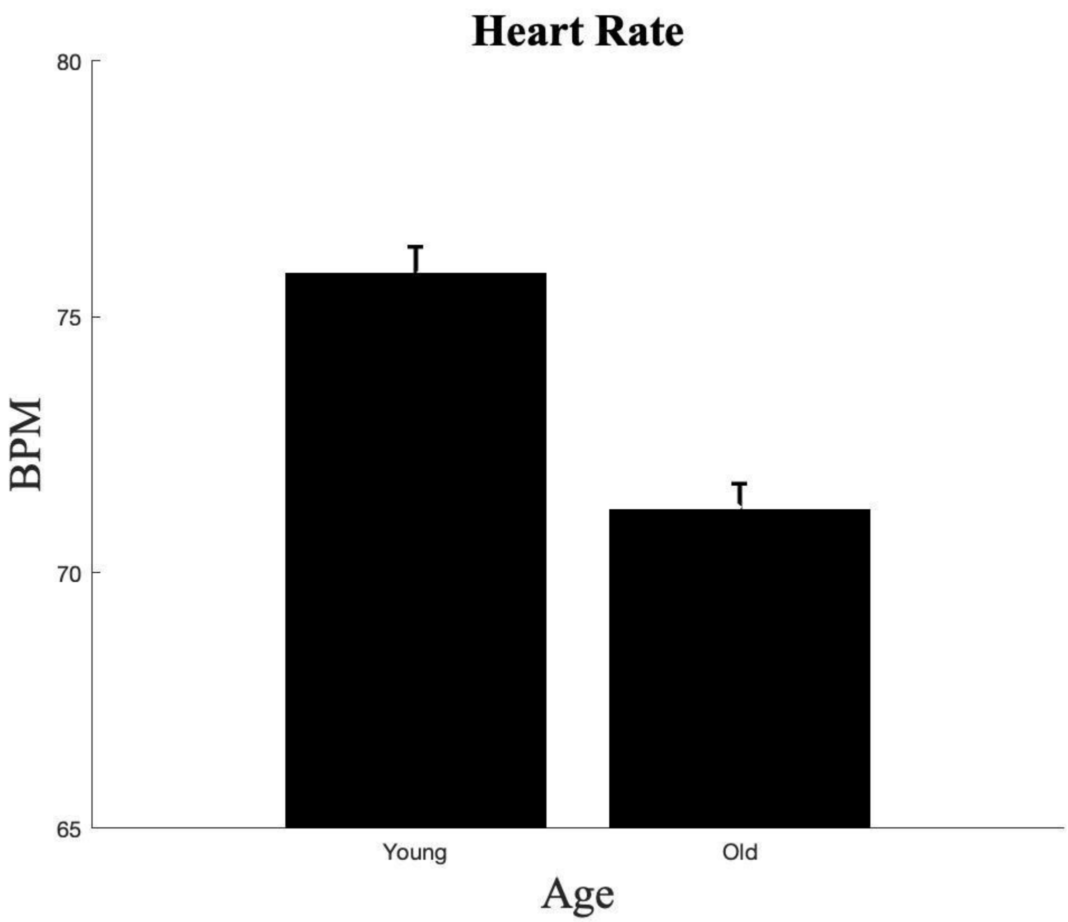

Figure 9. Heart rate, expressed as beats per minute (BPM), plotted as a function of the participant's age. Error bars reflect the standard error of the mean. The effect of Age (Young: M = 75.85, SD = 9.89; Old: M = 71.24, SD = 9.52) was significant ($\chi^2(1) = 3.89$, $p = 0.048$; $R^2 = 0.043$).

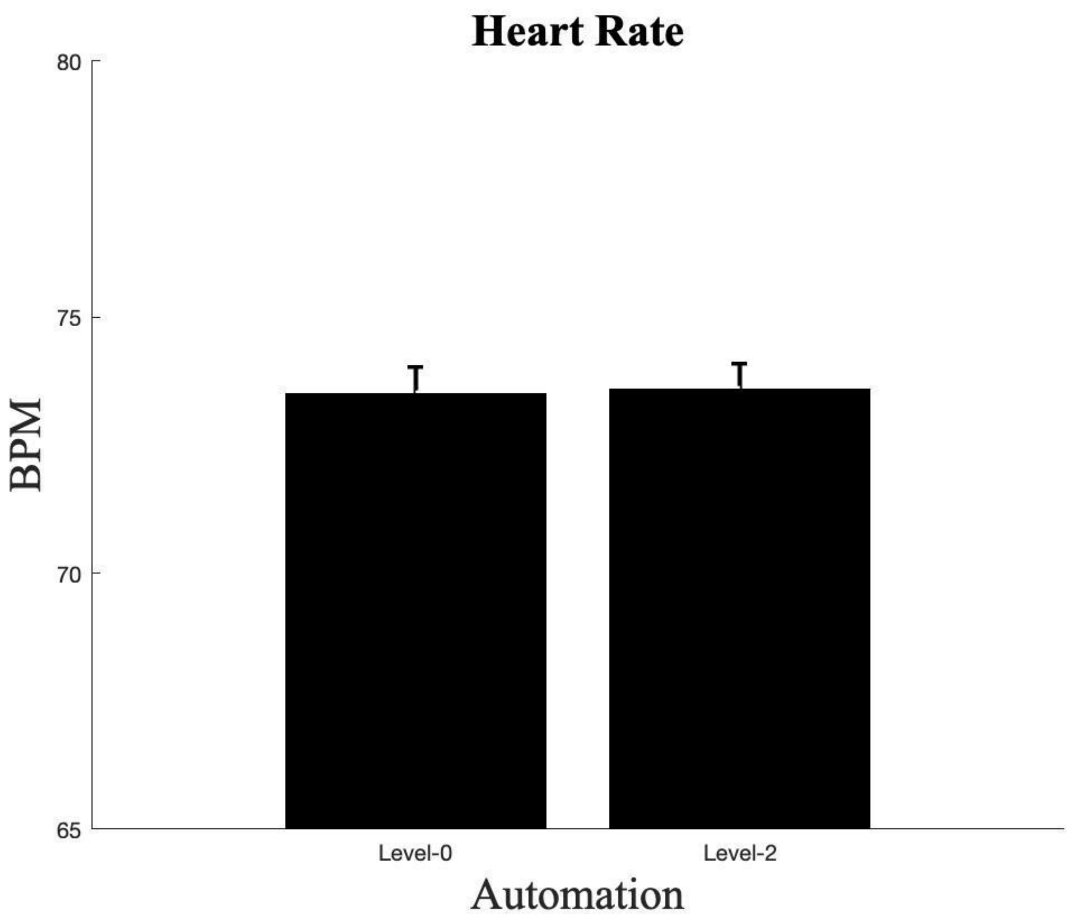

Figure 10. Heart rate, expressed as beats per minute (BPM), plotted as a function of the level of automation. Error bars reflect the standard error of the mean. The effect of Automation (Level-0: M = 73.59, SD = 9.76; Level-2: M = 73.50, SD = 10.18) was not significant (χ^2 (1) = 0.18, p = 0.68; $R^2 < 0.001$).

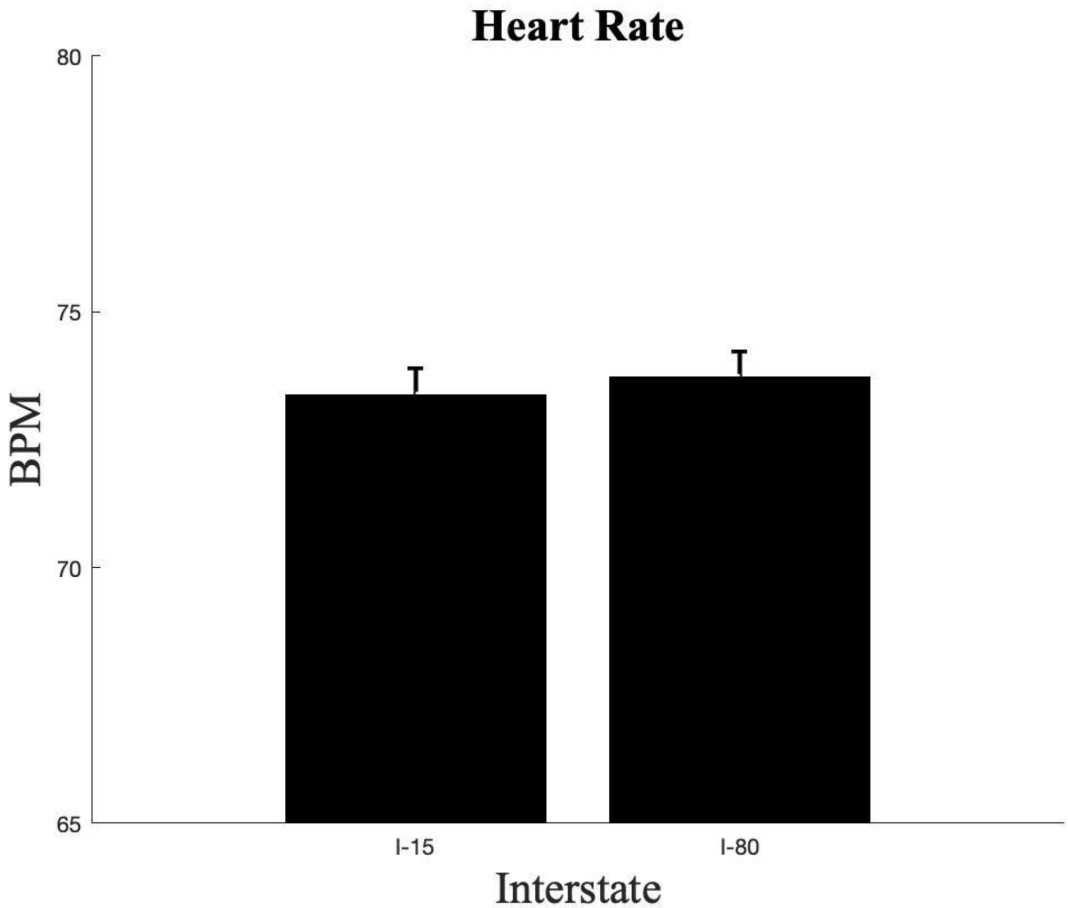

Figure 11. Heart rate, expressed as beats per minute (BPM), plotted as a function of the Interstate. Error bars reflect the standard error of the mean. The effect of Interstate (I-15: M = 73.37, SD = 10.16; I-80: M = 73.72, SD = 9.79) was not significant (χ^2 (1) = 0.81, *p* = 0.37; $R^2 < 0.001$).

Table 4. *Heart Rate (BPM)*

Variable	*Chi-square*	*df*	*p*	R^2
Vehicle	**10.293**	**3**	**0.016***	**0.004**
Age	**3.894**	**1**	**0.048***	**0.043**
Automation	0.176	1	0.675	0.000
Interstate	0.814	1	0.367	0.000
Vehicle x Age	7.091	3	0.069	0.002
Vehicle x Automation	0.619	3	0.892	0.000
Vehicle x Interstate	1.550	3	0.671	0.000
Age x Automation	1.560	1	0.212	0.000
Age x Interstate	0.529	1	0.467	0.000
Automation x Interstate	0.002	1	0.966	0.000
Vehicle x Age x Automation	9.879	10	0.451	0.003
Vehicle x Age x Interstate	9.506	10	0.485	0.003
Vehicle x Automation x Interstate	2.931	10	0.983	0.001
Age x Automation x Interstate	2.192	4	0.701	0.001
Vehicle x Age x Automation x Interstate	15.204	25	0.937	0.005

Note: R Squared values represent the Marginal R^2 difference score from the comparison model.
Significance Code: `*` < .05.

Heart Rate Variability: Root Mean Squared of Successive Differences (RMSSD)

Figures 12-15 present the main effects of Vehicle, Age, Automation, and Interstate on heart rate variability. Table 5 presents the results of the multi-level modeling. There were significant effects of Vehicle and Age. The Vehicle x Age interaction was also significant but accounted for a very small percentage of the overall variance (0.007) and were not indicative of a systematic trend.

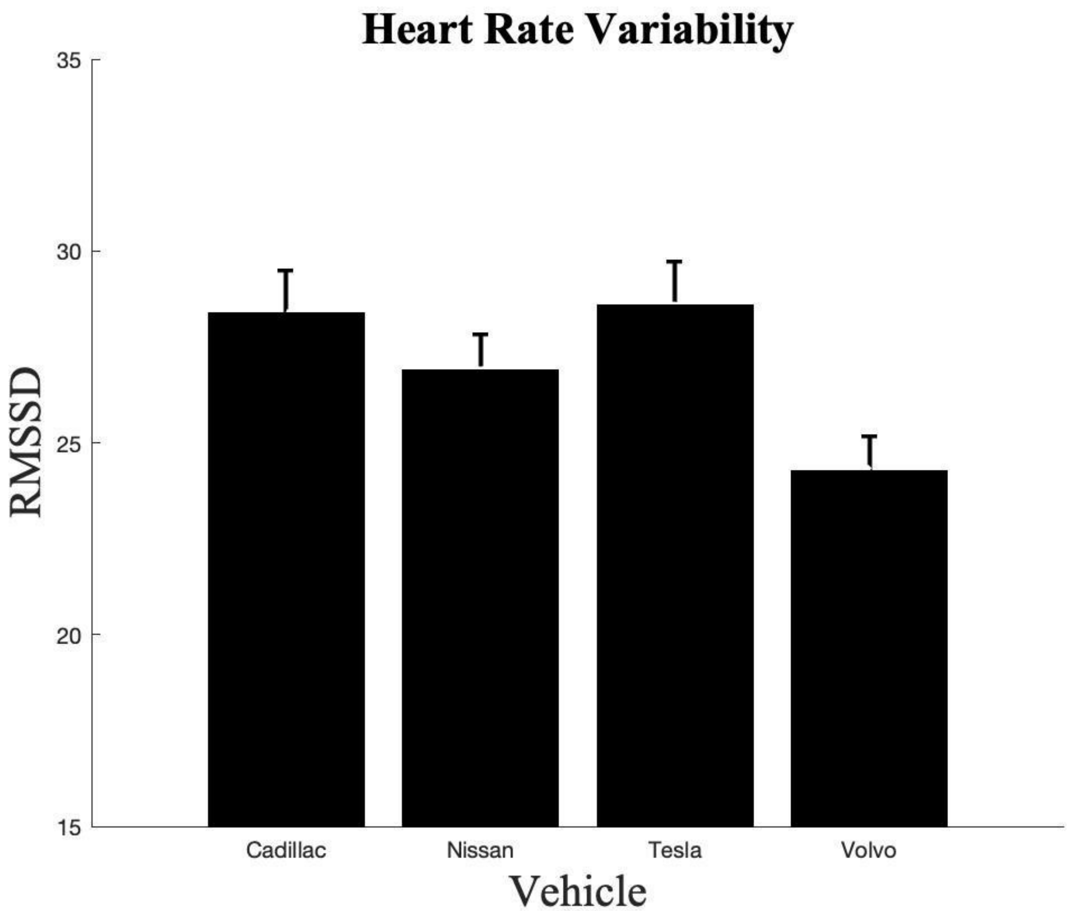

Figure 12. Heart rate variability, expressed as the Root Mean Squared of Successive Differences (RMSSD), plotted as a function of the vehicle. Error bars reflect the standard error of the mean. The effect of Vehicle (Cadillac: M = 28.40, SD = 14.76; Nissan: M = 26.89, SD = 13.05; Tesla: M = 28.60, SD = 15.11; Volvo: M = 24.27, SD = 11.82) was significant (χ^2 (3) = 22.71, p = <.001; R^2 = 0.011).

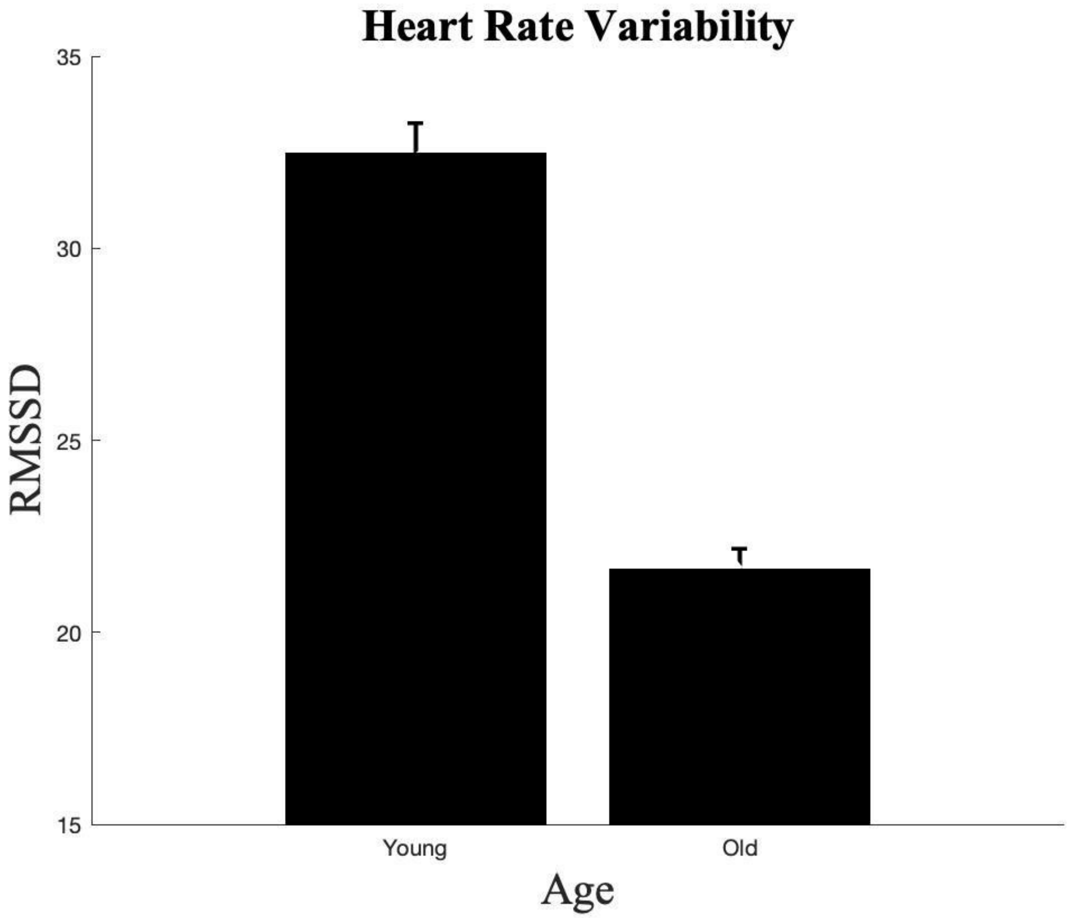

Figure 13. Heart rate variability, expressed as the Root Mean Squared of Successive Differences (RMSSD), plotted as a function of the participant's age. Error bars reflect the standard error of the mean. The effect of Age (Young: M = 32.48, SD = 15.00; Old: M = 21.64, SD = 9.96) was significant ($\chi^2(1) = 13.09$, p = <.001; $R^2 = 0.12$).

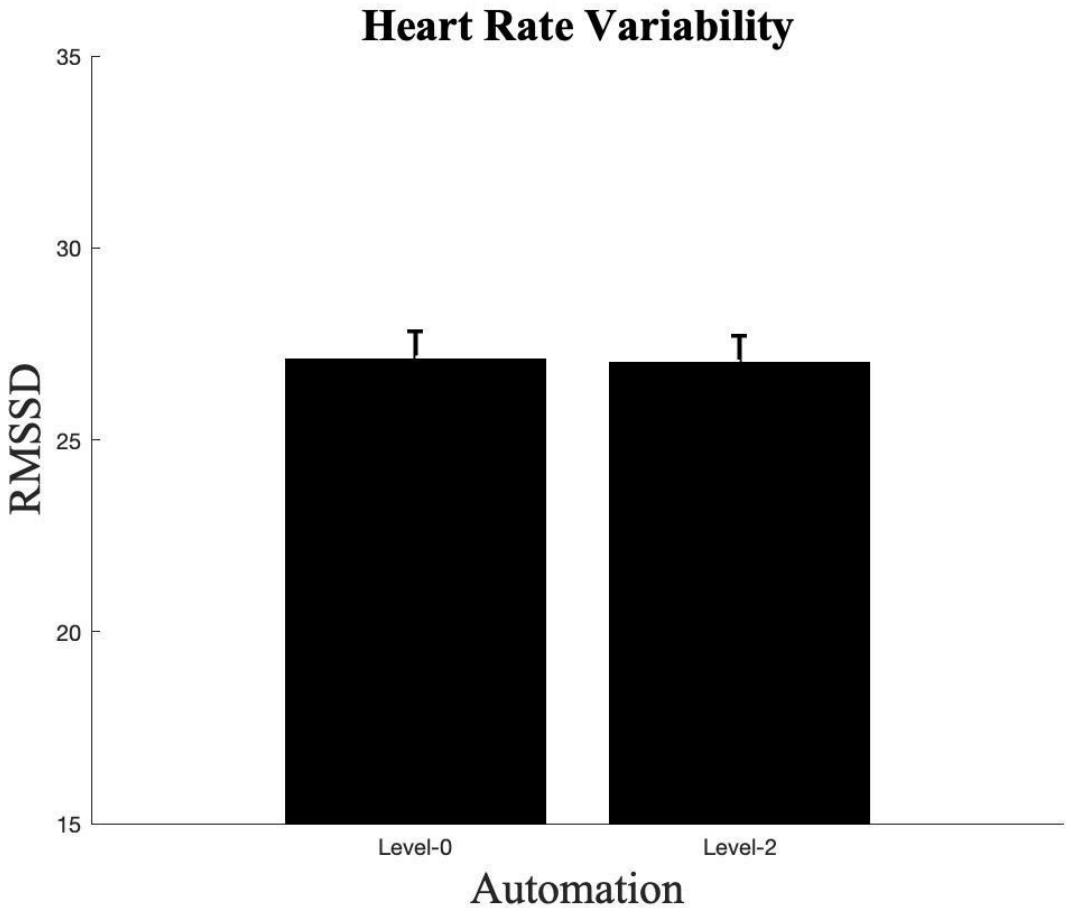

Figure 14. Heart rate variability, expressed as the Root Mean Squared of Successive Differences (RMSSD), plotted as a function of the level of automation. Error bars reflect the standard error of the mean. The effect of Automation (Level-0: M = 27.00, SD = 13.70; Level-2: M = 27.11, SD = 13.98) was not significant (χ^2 (1) = 0.04, p = 0.85; $R^2 < 0.001$).

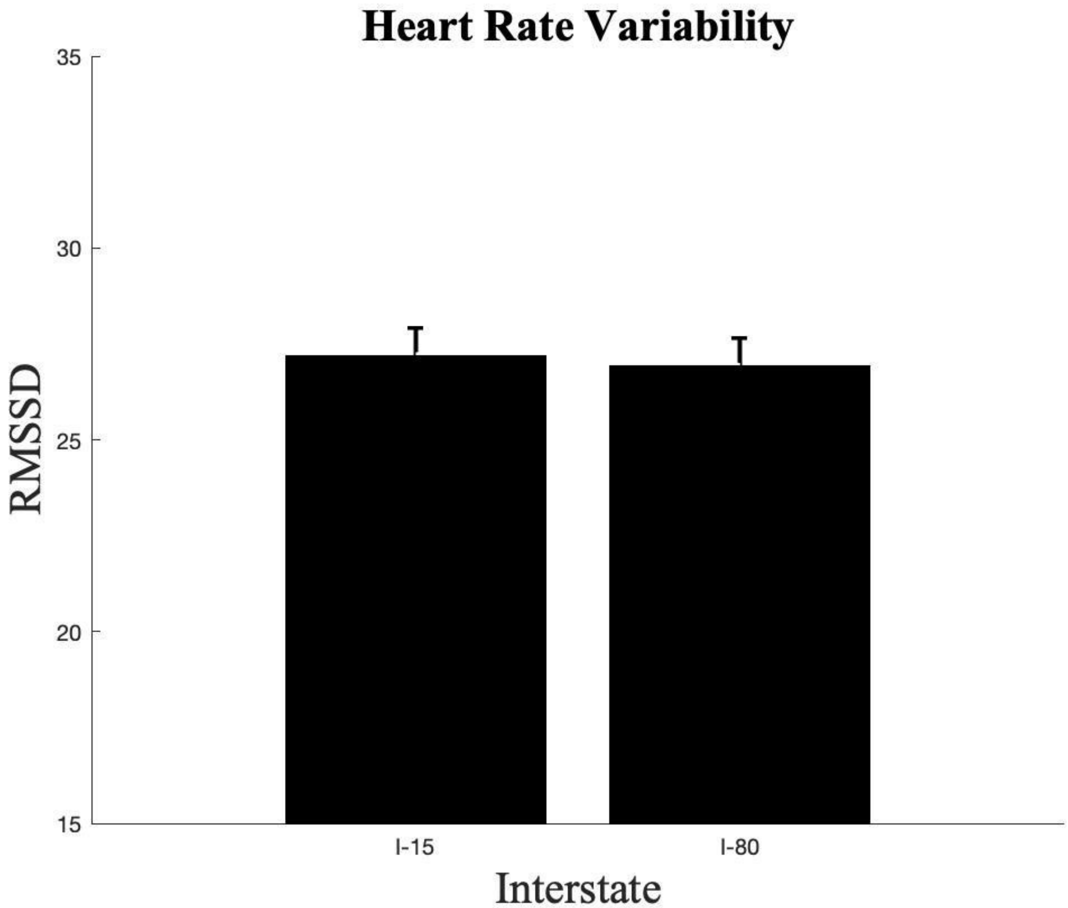

Figure 15. Heart rate variability, expressed as the Root Mean Squared of Successive Differences (RMSSD), plotted as a function of the Interstate. Error bars reflect the standard error of the mean. The effect of Interstate (I-15: M = 27.19, SD = 13.64; I-80: M = 26.91, SD = 14.04) was not significant ($\chi^2(1) = 0.16$, $p = 0.69$; $R^2 < 0.001$).

Table 5. *Heart Rate Variability (RMSSD)*				
Variable	*Chi-square*	*df*	*p*	R^2
Vehicle	**22.706**	**3**	**<.001*** **	**0.011**
Age	**13.086**	**1**	**<.001*** **	**0.123**
Automation	0.036	1	0.850	0.000
Interstate	0.164	1	0.685	0.000
Vehicle x Age	**14.649**	**3**	**0.002** **	**0.007**
Vehicle x Automation	0.308	3	0.959	0.001
Vehicle x Interstate	1.298	3	0.730	0.001
Age x Automation	1.291	1	0.256	0.001
Age x Interstate	0.376	1	0.540	0.000
Automation x Interstate	0.006	1	0.937	0.000
Vehicle x Age x Automation	17.207	10	0.070	0.007
Vehicle x Age x Interstate	17.583	10	0.062	0.007
Vehicle x Automation x Interstate	1.660	10	0.998	0.001
Age x Automation x Interstate	2.300	4	0.681	0.001
Vehicle x Age x Automation x Interstate	22.117	25	0.629	0.007

Note: R Squared values represent the Marginal R^2 difference score from the comparison model. Significance Codes: `**` < .01, `***` < .001.

Detection Response Task (DRT)

A total of 185,947 DRT stimuli were available for analysis. Of these, 9,013 were coded as misses (e.g., no response within 2,500ms of stimulus onset) and 176,808 stimuli were coded as hits (e.g., within 100-2500ms of stimulus onset). An additional 3,579 responses (~2%) were removed that were outside +/- 3 standard deviations of the mean of each participant. These cleaning procedures left an average of 233 (SD = 40) valid reaction times for each participant in each aggregated cell (Vehicle x Age x Automation x Interstate).

Hit rate results averaged across the total of valid hits and misses. This expanded data set included an average of 244 (SD = 38.7) potential responses (hits + misses) for each participant per cell (Vehicle x Age x Automation x Interstate).

Reaction Time

Figures 16-19 present the main effects of Vehicle, Age, Automation, and Interstate on DRT reaction time. Table 6 presents the results of the multi-level modeling. There were significant effects of Vehicle, Automation, and Interstate.

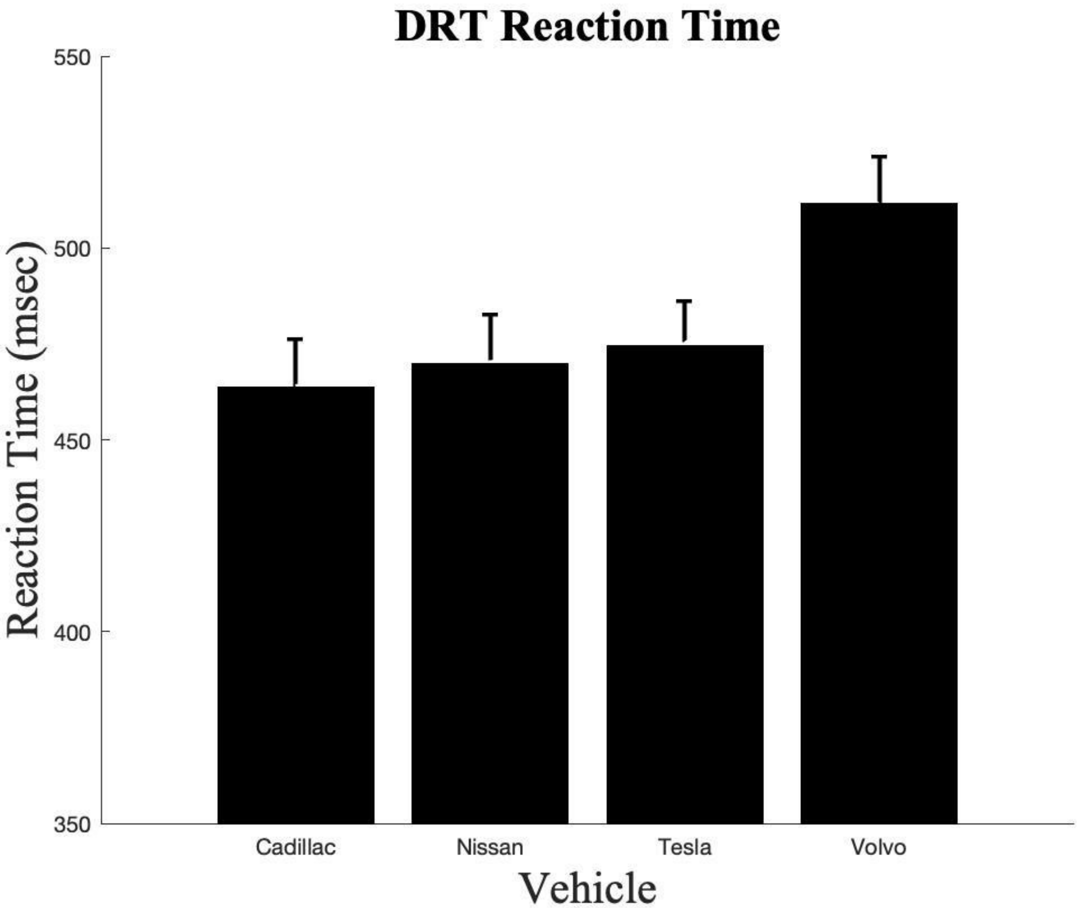

Figure 16. DRT reaction time plotted as a function of vehicle. Error bars reflect the standard error of the mean. The effect of Vehicle (Cadillac: M = 463.60, SD = 173.60; Nissan: M = 469.87, SD = 174.59; Tesla: M = 474.60, SD = 158.45; Volvo: M = 511.53, SD = 168.51) was significant (χ^2 (3) = 21.62, p = <.001; R^2 = 0.009).

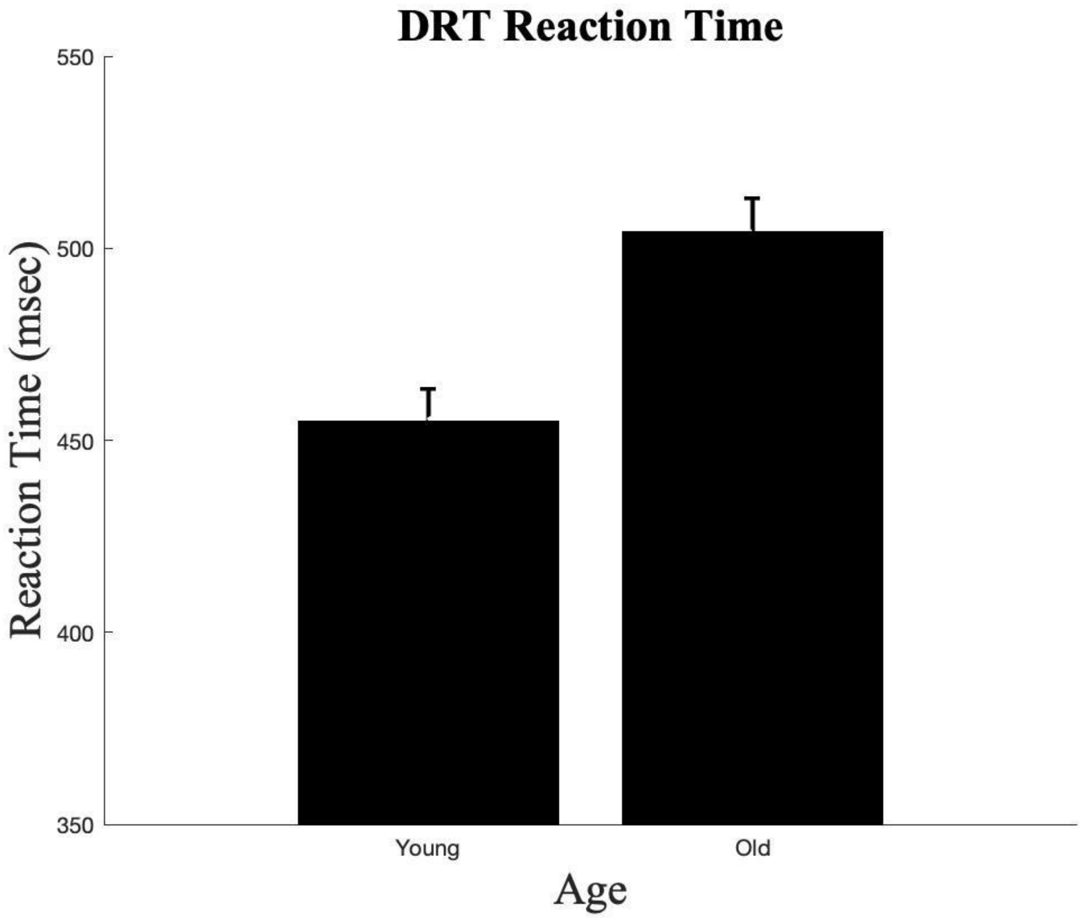

Figure 17. DRT reaction time plotted as a function of the participant's age. Error bars reflect the standard error of the mean. The effect of Age (Young: M = 455.00, SD = 163.59; Old: M = 504.20, SD = 171.93) was not significant (χ^2 (1) = 1.19, p = 0.28; R^2 = 0.013). The lack of significance was the result of entering the participant as a random intercept, effectively partialling out any variance between participants across the age spectrum.

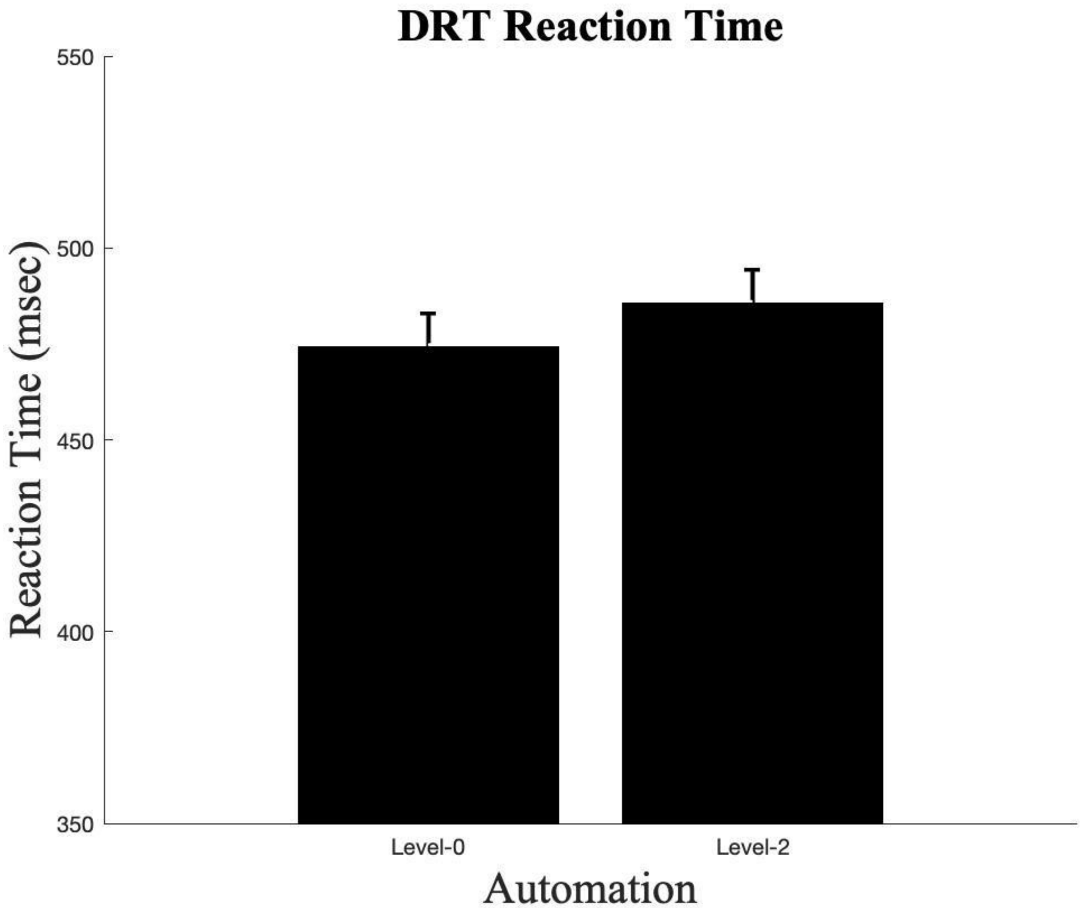

Figure 18. DRT reaction time plotted as a function of the level of automation. Error bars reflect the standard error of the mean. The effect of Automation (Level-0: M = 474.24, SD = 170.76; Level-2: M = 485.59, SD = 168.32) was significant ($\chi^2(1) = 4.29$, $p = 0.038$; $R^2 = 0.001$).

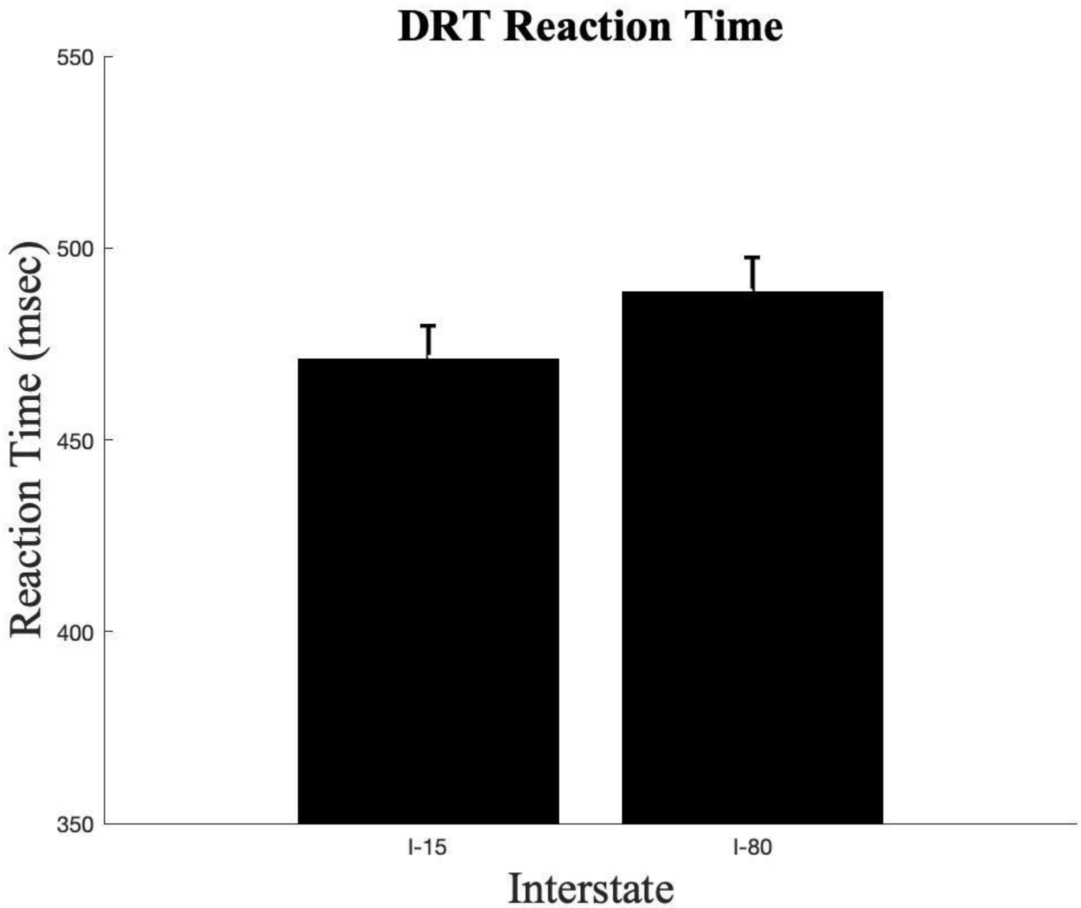

Figure 19. DRT reaction time plotted as a function of the Interstate. Error bars reflect the standard error of the mean. The effect of Interstate (I-15: M = 471.16, SD = 166.99; I-80: M = 488.57, SD = 171.83) was significant (χ^2 (1) = 8.91, p = 0.003; R^2 = 0.003).

Table 6. *DRT reaction time*

Variable	Chi-square	df	p	R^2
Vehicle	**21.617**	**3**	**<.001***	**0.009**
Age	1.193	1	0.275	0.013
Automation	**4.286**	**1**	**0.038***	**0.001**
Interstate	**8.911**	**1**	**0.003****	**0.003**
Vehicle x Age	4.404	3	0.221	0.002
Vehicle x Automation	2.142	3	0.544	0.001
Vehicle x Interstate	3.894	3	0.273	0.002
Age x Automation	0.155	1	0.694	0.000
Age x Interstate	0.404	1	0.525	0.000
Automation x Interstate	0.039	1	0.844	0.000
Vehicle x Age x Automation	7.489	10	0.679	0.004
Vehicle x Age x Interstate	9.971	10	0.443	0.004
Vehicle x Automation x Interstate	6.494	10	0.772	0.002
Age x Automation x Interstate	0.980	4	0.913	0.001
Vehicle x Age x Automation x Interstate	14.527	25	0.952	0.006

Note: R Squared values represent the Marginal R^2 difference score from the comparison model. Significance Codes: \`*\` < .05, \`**\` < .01, \`***\` < .001.

Hit Rate

Figures 20-23 present the main effects of Vehicle, Age, Automation, and Interstate on DRT hit rate. Table 7 presents the results of the multi-level modeling. There were significant main effects of Vehicle, Age, Automation, and Interstate.

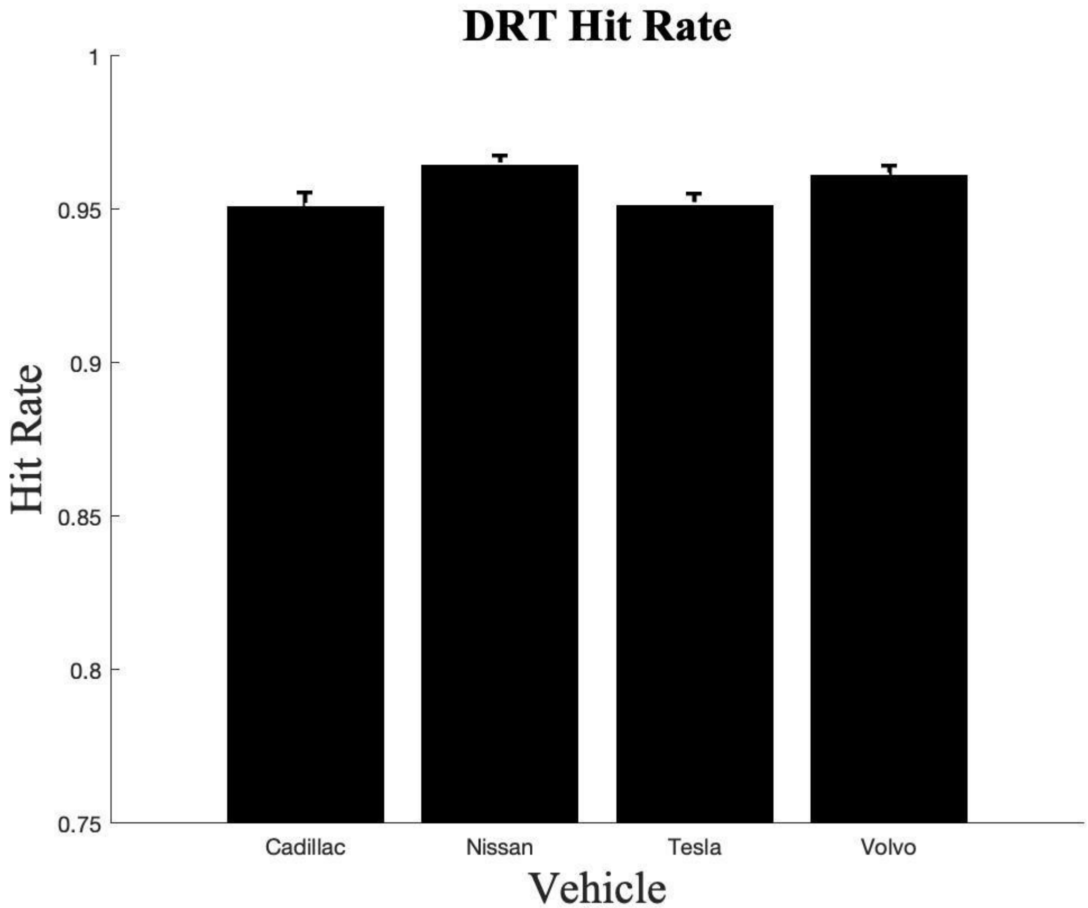

Figure 20. DRT hit rate plotted as a function of vehicle. Error bars reflect the standard error of the mean. The effect of Vehicle (Cadillac: M = 0.95, SD = 0.06; Nissan: M = 0.96, SD = 0.05; Tesla: M = 0.95, SD = 0.05; Volvo: M = 0.96, SD = 0.05) was significant (χ^2 (3) = 14.58, p = 0.002; R^2 = 0.011).

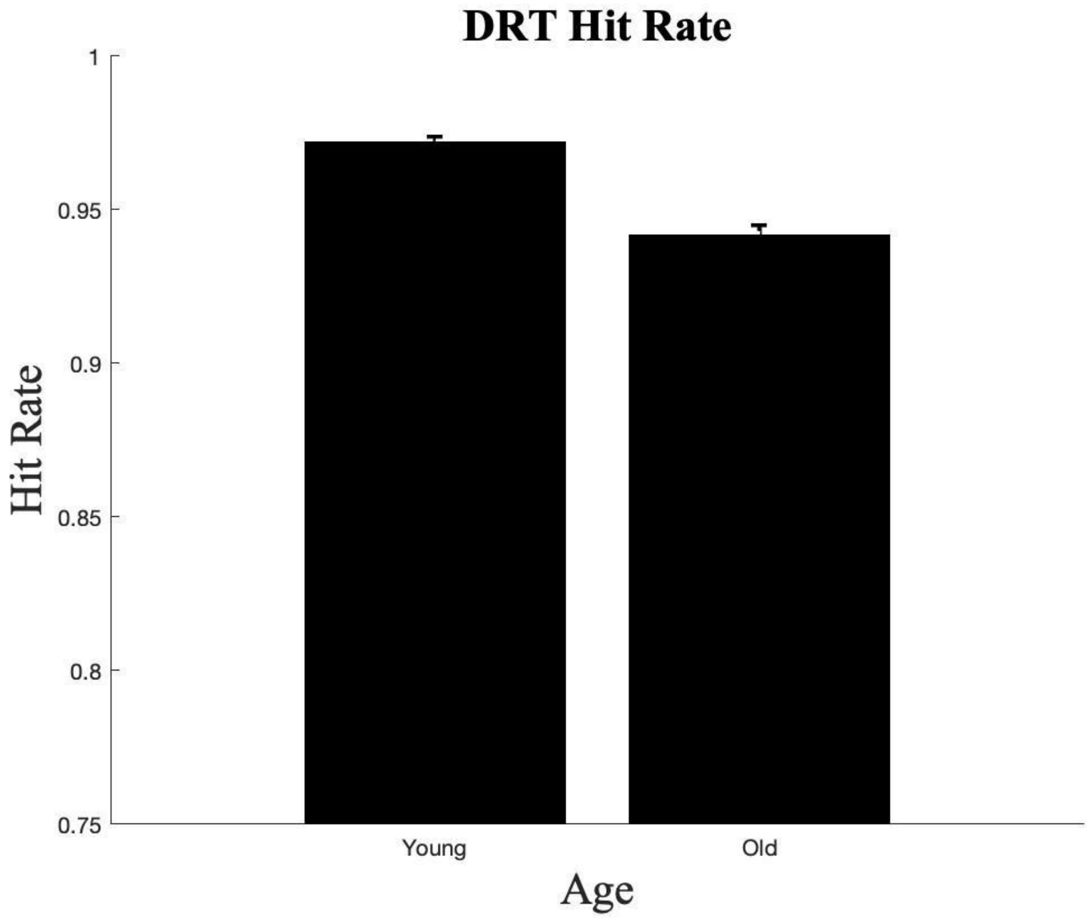

Figure 21. DRT hit rate plotted as a function of the participant's age. Error bars reflect the standard error of the mean. The effect of Age (Young: M = 0.97, SD = 0.03; Old: M = 0.94, SD = 0.06) was significant ($\chi^2(1) = 7.27$, $p = 0.007$; $R^2 = 0.057$).

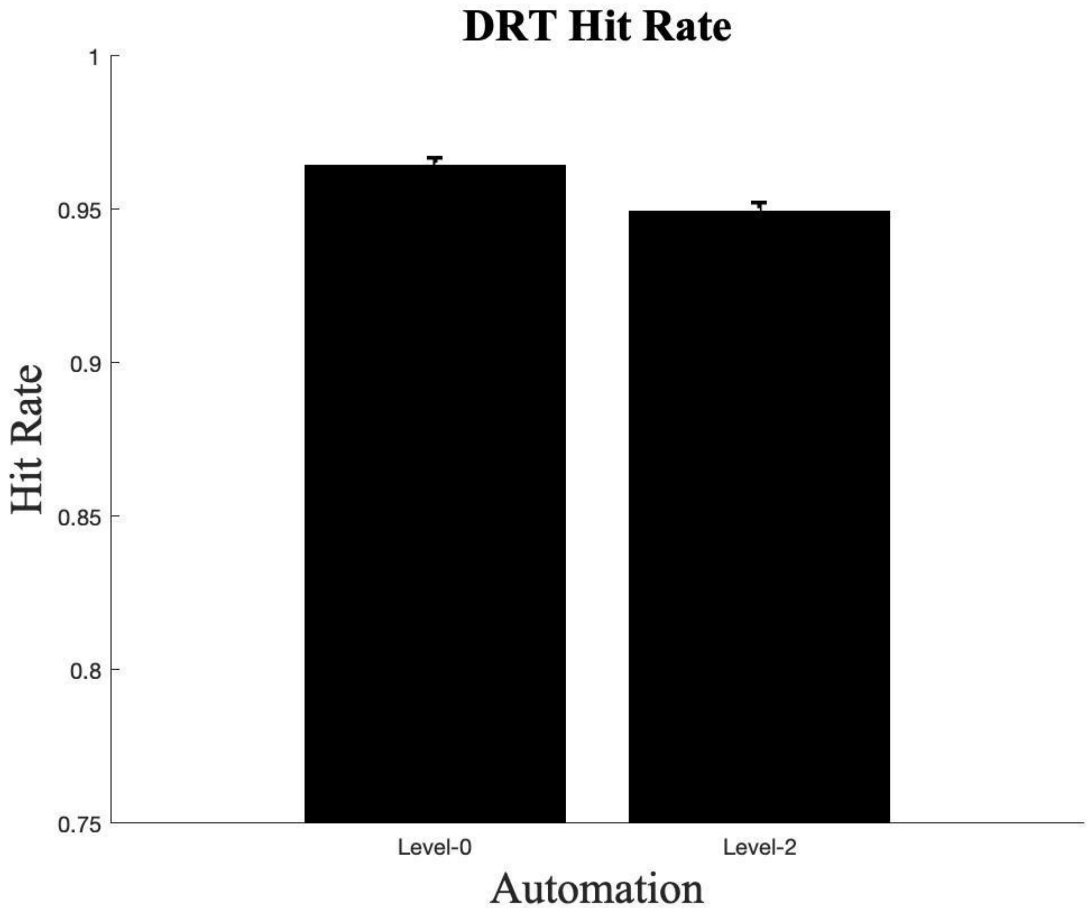

Figure 22. DRT hit rate plotted as a function of the level of automation. Error bars reflect the standard error of the mean. The effect of Automation (Level-0: M = 0.96, SD = 0.05; Level-2: M = 0.95, SD = 0.06) was significant ($\chi^2 (1) = 33.65$, $p < 0.001$; $R^2 = 0.021$).

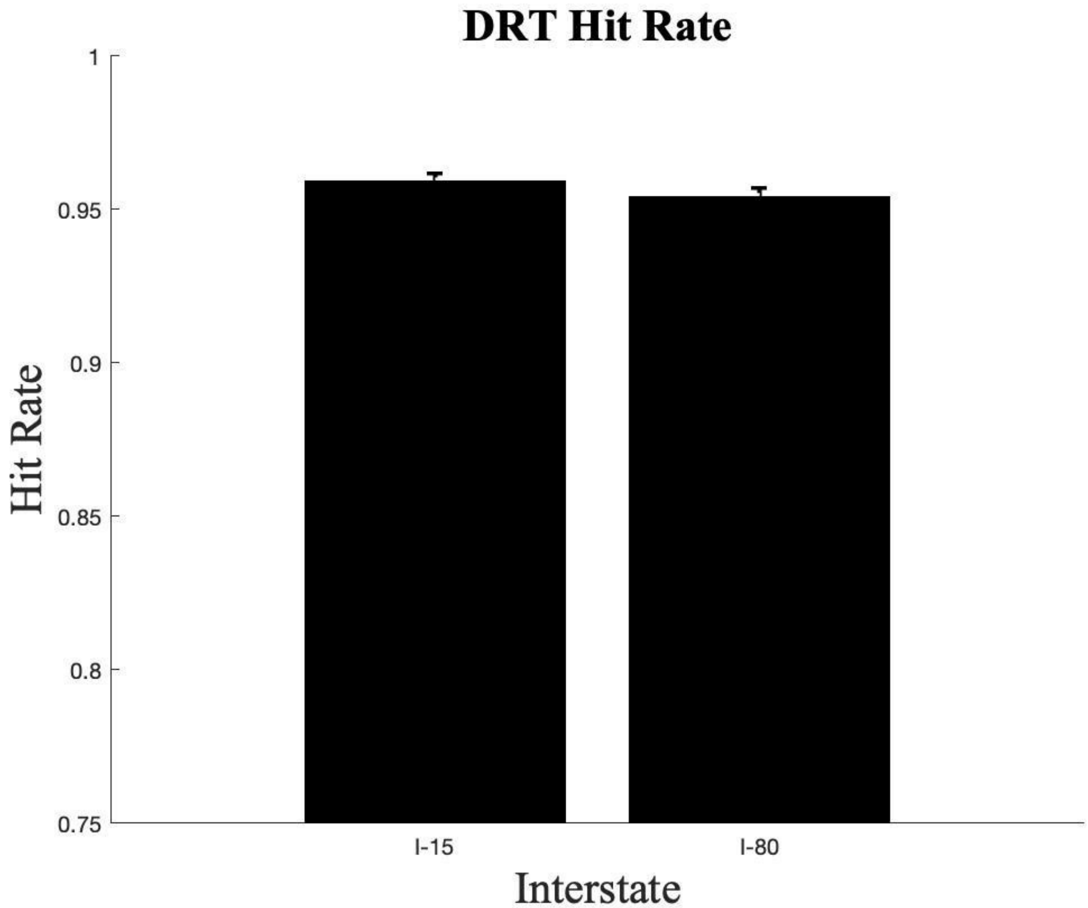

Figure 23. DRT hit rate plotted as a function of the Interstate. Error bars reflect the standard error of the mean. The effect of Interstate (I-15: M = 0.96, SD = 0.05; I-80: M = 0.95, SD = 0.06) was significant (χ^2 (1) = 4.06, *p* = 0.044; R^2 = 0.003).

Table 7. *DRT Hit Rate*

Variable	*Chi-square*	*df*	*p*	R^2
Vehicle	**14.584**	**3**	**0.002****	**0.011**
Age	**7.269**	**1**	**0.007****	**0.057**
Automation	**33.648**	**1**	**<.001*****	**0.021**
Interstate	**4.058**	**1**	**0.044***	**0.003**
Vehicle x Age	**8.976**	**3**	**0.030***	**0.006**
Vehicle x Automation	2.870	3	0.412	0.001
Vehicle x Interstate	1.072	3	0.784	0.001
Age x Automation	**4.872**	**1**	**0.027***	**0.003**
Age x Interstate	0.286	1	0.593	0.000
Automation x Interstate	0.355	1	0.551	0.001
Vehicle x Age x Automation	**18.734**	**10**	**0.044***	**0.012**
Vehicle x Age x Interstate	11.645	10	0.310	0.007
Vehicle x Automation x Interstate	8.079	10	0.621	0.005
Age x Automation x Interstate	5.776	4	0.217	0.003
Vehicle x Age x Automation x Interstate	27.901	25	0.312	0.016

Note: R Squared values represent the Marginal R^2 difference score from the comparison model. Significance Codes: `*` < .05, `**` < .01.

The Vehicle x Age, Age x Automation and Vehicle x Age x Automation interactions were also significant. These higher order interactions accounted for relatively small proportions of the variance (e.g., $R^2 <= 0.012$) and were not indicative of a systematic trend.

Survey Data

Nervousness is the latent variable in the factor analysis loading on the ratings of "scared", "nervous", "distressed", "confident", and "relaxed" obtained in the survey. Subjective ratings of nervousness ranged from 1 (*very slightly or not at all*) to 5 (*extremely*).

Figures 24-27 present the main effects of Vehicle, Age, Automation, and Interstate on Nervousness. Table 8 presents the results of the multi-level modeling. There were significant main effects of Vehicle and Automation.

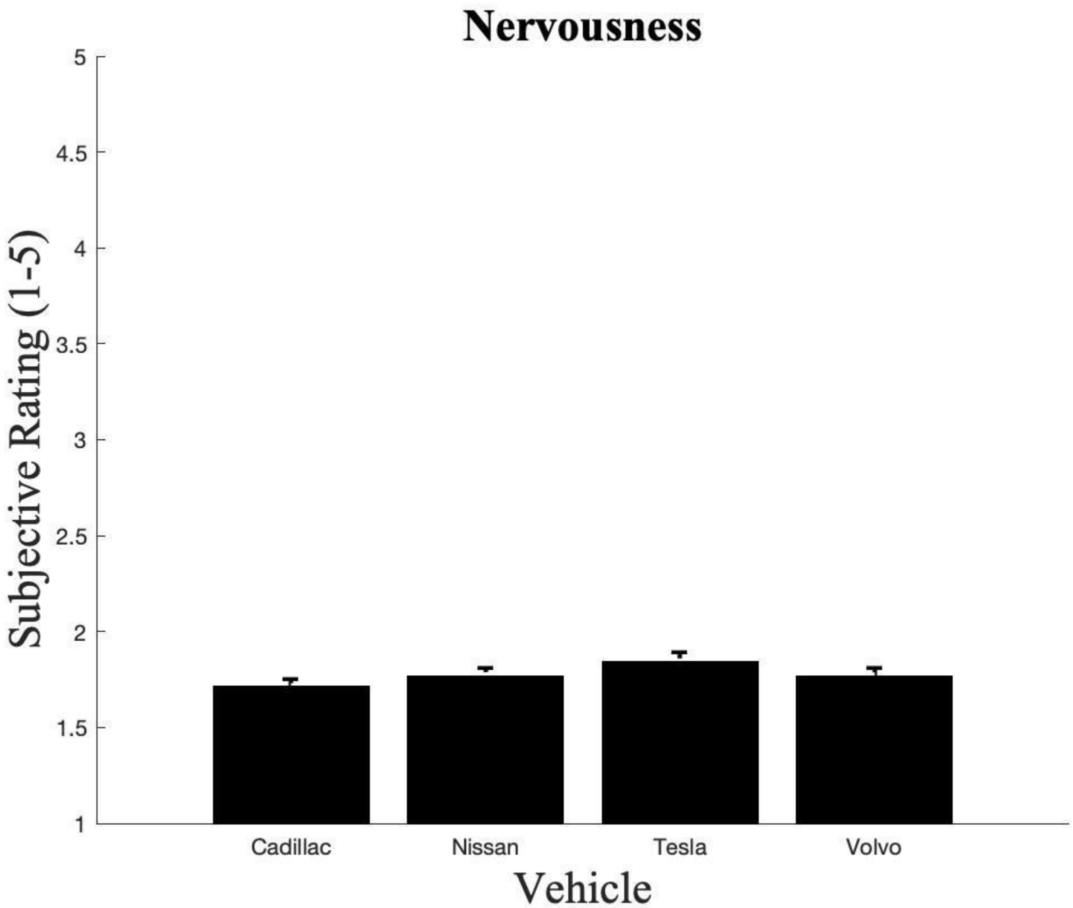

Figure 24. Nervousness (1 = *very slightly or not at all*, 5 = extremely) plotted as a function of vehicle. Error bars reflect the standard error of the mean. The effect of Vehicle (Cadillac: M = 1.71, SD = 0.53; Nissan: M = 1.77, SD = 0.53; Tesla: M = 1.84, SD = 0.64; Volvo: M = 1.77, SD = 0.54) was significant (χ^2 (3) = 12.63, p = 0.006; R^2 = 0.010).

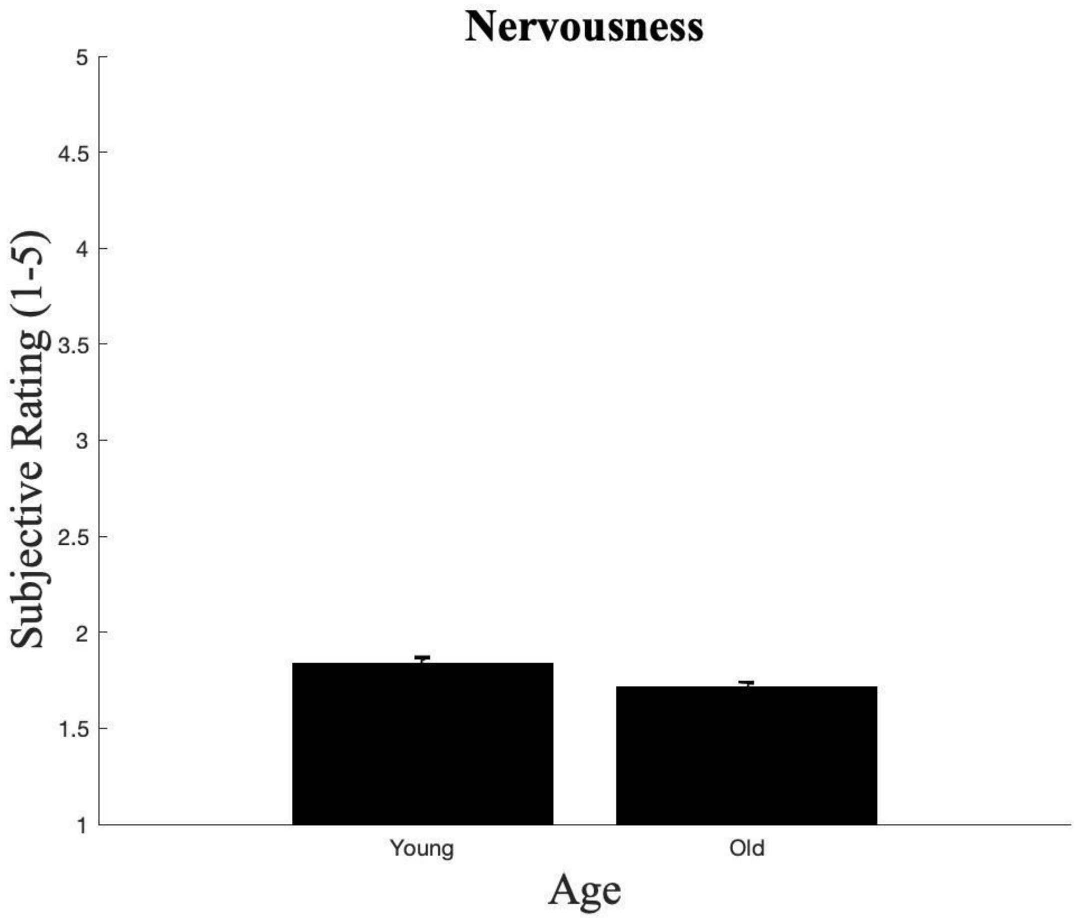

Figure 25. Nervousness (1 = *very slightly or not at all*, 5 = *extremely*) plotted as a function of the participant's age. Error bars reflect the standard error of the mean. The effect of Age (Young: M = 1.84, SD = 0.58; Old: M = 1.71, SD = 0.54) was not significant (χ^2 (1) = 1.99, p = 0.16; R^2 = 0.014). The lack of significance was the result of entering the participant as a random intercept, effectively partialling out any variance between participants across the age spectrum.

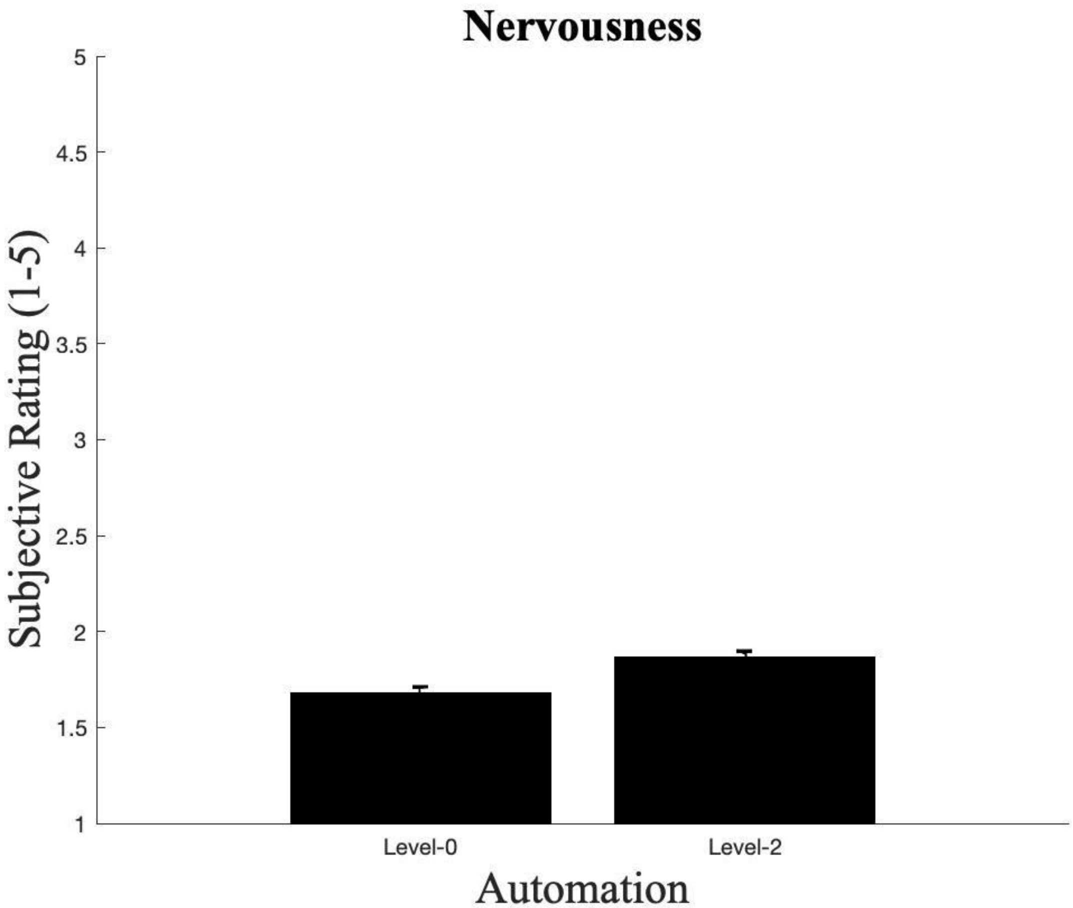

Figure 26. Nervousness (1 = *very slightly or not at all*, 5 = *extremely*) plotted as a function of the level of automation. Error bars reflect the standard error of the mean. The effect of Automation (Level-0: M = 1.68, SD = 0.49; Level-2: M = 1.86, SD = 0.61) was significant (χ^2 (1) = 38.67, *p* = <.001; R^2 = 0.027).

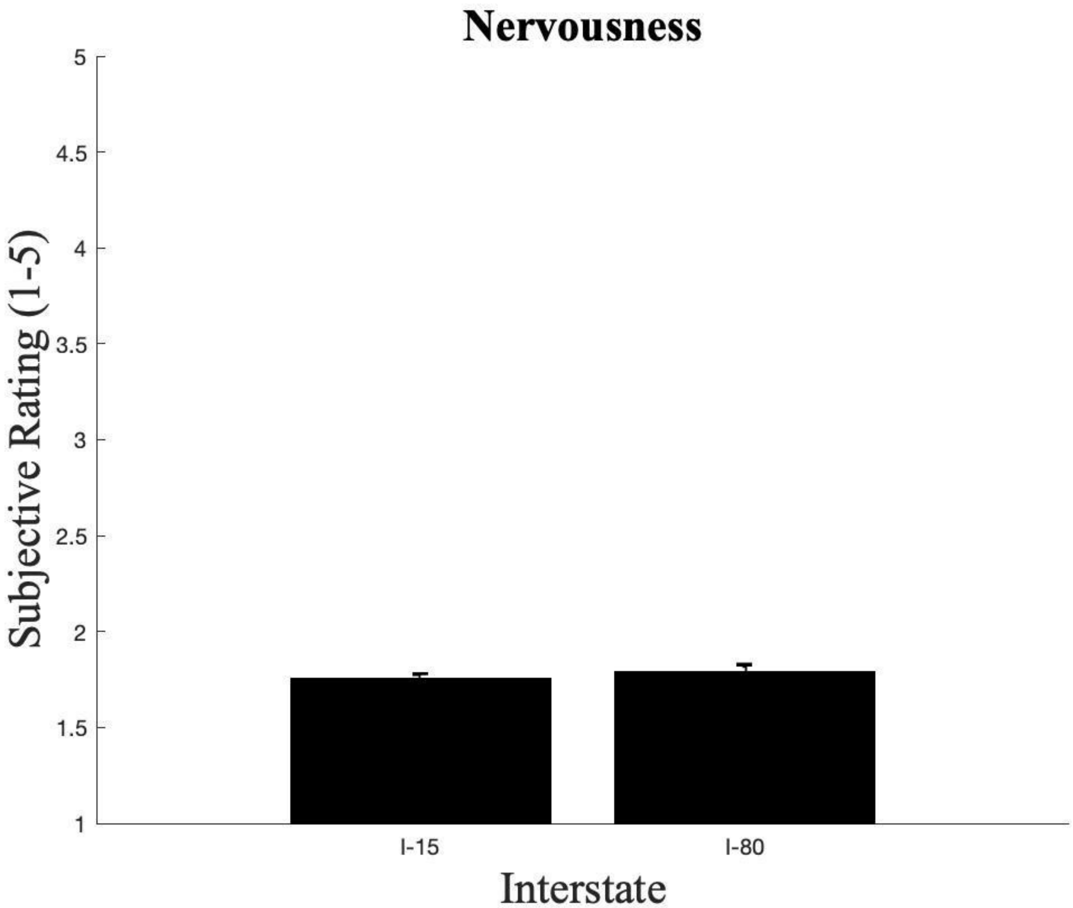

Figure 27. Nervousness (1 = *very slightly or not at all*, 5 = *extremely*) plotted as a function of the Interstate. Error bars reflect the standard error of the mean. The effect of Interstate (I-15: M = 1.75, SD = 0.53; I-80: M = 1.79, SD = 0.59) was not significant (χ^2 (1) = 1.45, p = 0.23; R^2 = 0.001).

Table 8. *Nervousness*

Variable	Chi-square	df	p	R^2
Vehicle	**12.627**	**3**	**0.006****	**0.010**
Age	1.986	1	0.159	0.014
Automation	**38.671**	**1**	**<.001*****	**0.027**
Interstate	1.445	1	0.229	0.001
Vehicle x Age	4.631	3	0.201	0.004
Vehicle x Automation	**8.768**	**3**	**0.033***	**0.006**
Vehicle x Interstate	0.916	3	0.822	0.001
Age x Automation	3.188	1	0.074	0.002
Age x Interstate	**6.625**	**1**	**0.010***	**0.005**
Automation x Interstate	**4.160**	**1**	**0.041***	**0.003**
Vehicle x Age x Automation	**20.289**	**10**	**0.027***	**0.013**
Vehicle x Age x Interstate	16.284	10	0.092	0.011
Vehicle x Automation x Interstate	**19.267**	**10**	**0.037***	**0.013**
Age x Automation x Interstate	**14.575**	**4**	**0.006****	**0.010**
Vehicle x Age x Automation x Interstate	**43.662**	**25**	**0.012***	**0.027**

Note: R Squared values represent the Marginal R^2 difference score from the comparison model. Significance Codes: \`*\` < .05, \`**\` < .01, \`***\` < .001.

Several interactions were also significant (see Table 8). All higher order interactions are provided for completeness, however, given the relatively small effect sizes and lack of a systematic trend, the data are difficult to interpret.

Inattention is the latent variable in the factor analysis loading on the items that assessed directly or indirectly the level of inattention to driving in the survey. Subjective ratings of inattention ranged from 1 (*very slightly or not at all*) to 5 (*extremely*).

Figures 28-31 present the main effects of Vehicle, Age, Automation, and Interstate on Inattention. Table 9 presents the results of the multi-level modeling. There were significant main effects of Age and Interstate.

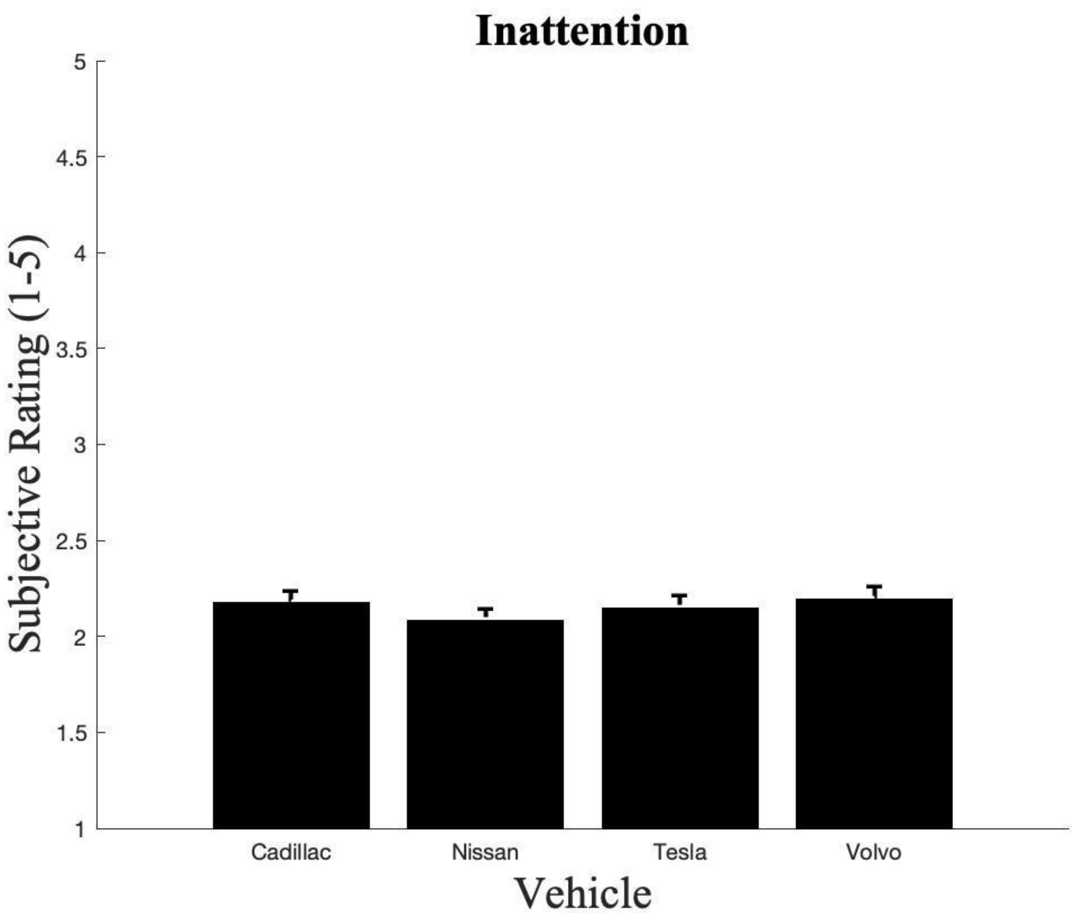

Figure 28. Inattention (1 = *very slightly or not at all*, 5 = *extremely*) plotted as a function of vehicle. Error bars reflect the standard error of the mean. The effect of Vehicle (Cadillac: M = 2.18, SD = 0.83; Nissan: M = 2.08, SD = 0.83; Tesla: M = 2.15, SD = 0.86; Volvo: M = 2.19, SD = 0.85) was not significant ($\chi^2(3) = 5.67$, $p = 0.13$; $R^2 = 0.003$).

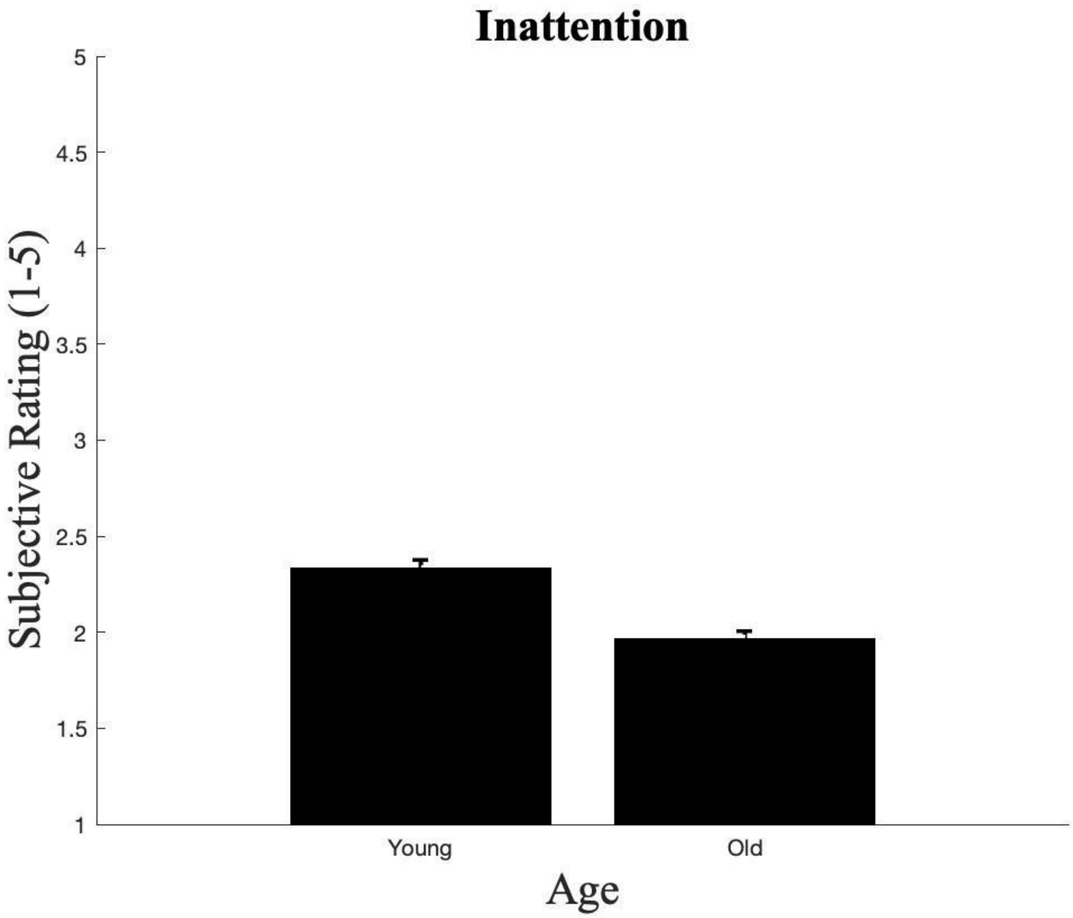

Figure 29. Inattention (1 = *very slightly or not at all*, 5 = *extremely*) plotted as a function of the participant's age. Error bars reflect the standard error of the mean. The effect of Age (Young: M = 2.33, SD = 0.83; Old: M = 1.97, SD = 0.81) was significant (χ^2 (1) = 9.30, p = 0.002; R^2 = 0.081).

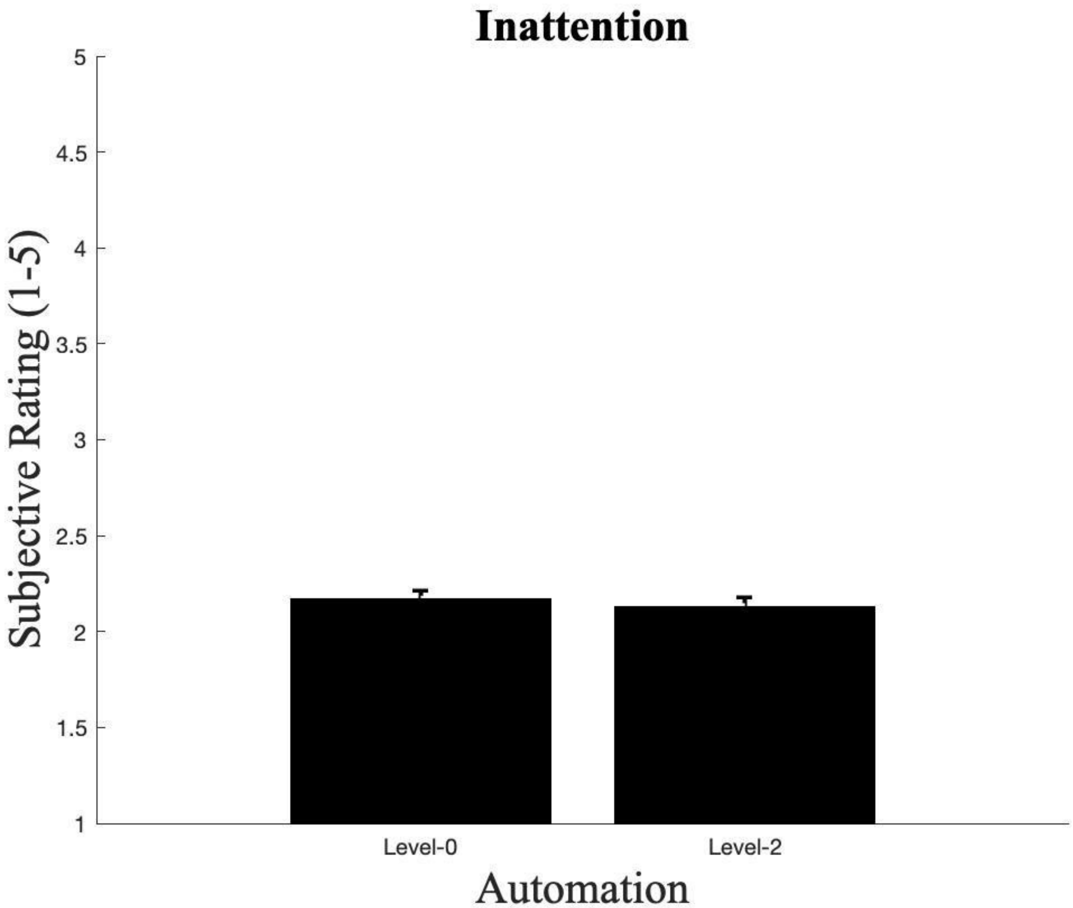

Figure 30. Inattention (1 = *very slightly or not at all*, 5 = *extremely*) plotted as a function of the level of automation. Error bars reflect the standard error of the mean. The effect of Automation (Level-0: M = 2.17, SD = 0.81; Level-2: M = 2.13, SD = 0.87) was not significant (χ^2 (1) = 1.02, p = 0.31; $R^2 < 0.001$).

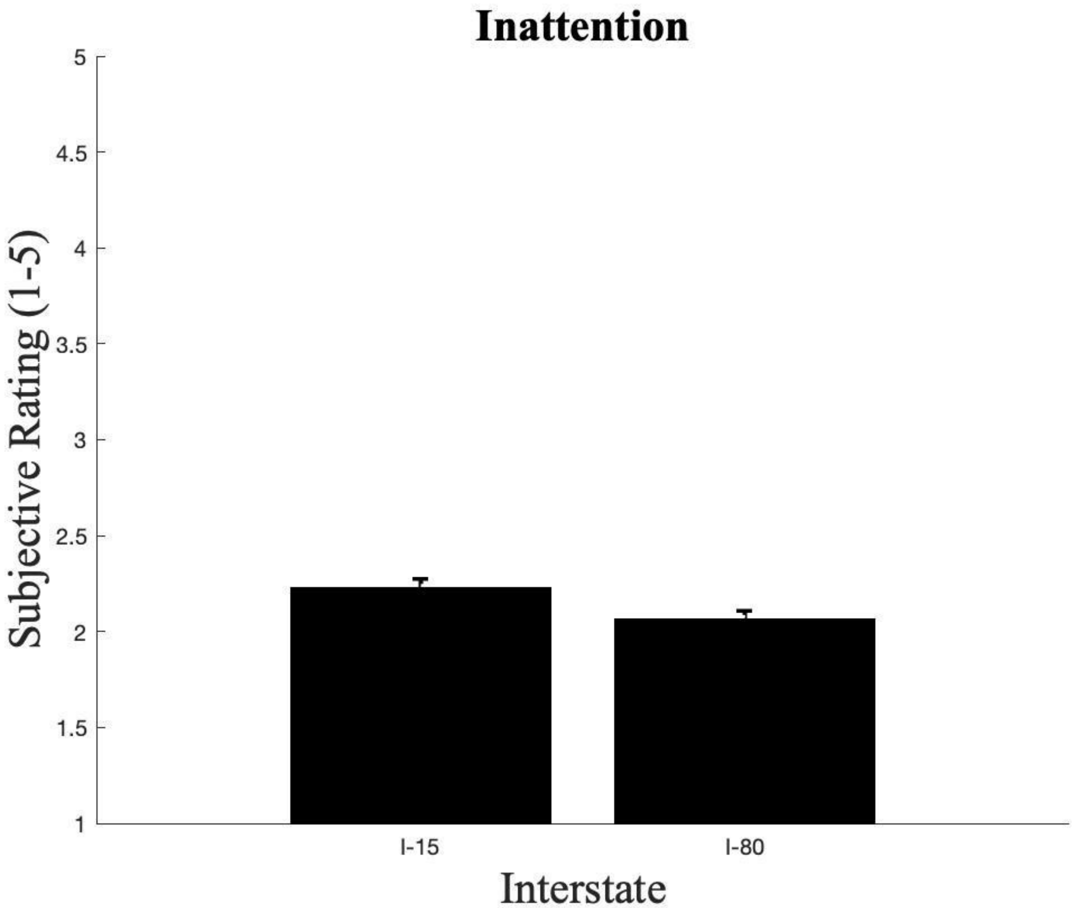

Figure 31. Inattention (1 = *very slightly or not at all*, 5 = *extremely*) plotted as a function of the Interstate. Error bars reflect the standard error of the mean. The effect of Interstate (I-15: M = 2.23, SD = 0.87; I-80: M = 2.07, SD = 0.80) was significant (χ^2 (1) = 18.27, p < 0.001; R^2 = 0.009).

Table 9. *Inattention*				
Variable	*Chi-square*	*df*	*p*	R^2
Vehicle	5.665	3	0.129	0.003
Age	**9.298**	**1**	**0.002****	**0.081**
Automation	1.022	1	0.312	0.000
Interstate	**18.268**	**1**	**<.001*****	**0.009**
Vehicle x Age	**8.639**	**3**	**0.035***	**0.005**
Vehicle x Automation	5.714	3	0.126	0.002
Vehicle x Interstate	2.035	3	0.565	0.001
Age x Automation	0.040	1	0.842	0.000
Age x Interstate	1.951	1	0.163	0.001
Automation x Interstate	2.937	1	0.087	0.002
Vehicle x Age x Automation	15.462	10	0.116	0.009
Vehicle x Age x Interstate	13.594	10	0.192	0.008
Vehicle x Automation x Interstate	11.330	10	0.332	0.006
Age x Automation x Interstate	5.671	4	0.225	0.003
Vehicle x Age x Automation x Interstate	25.962	25	0.410	0.012

Note: R Squared values represent the Marginal R^2 difference score from the comparison model. Significance Codes: `*` < .05, ` **` < .01, `***` < .001.

Excitement is the latent variable in the factor analysis loading on the ratings of "excited," "enthusiastic," and "cheerful" obtained in the survey. Subjective ratings of excitement ranged from 1 (*very slightly or not at all*) to 5 (*extremely*).

Figures 32-35 present the main effects of Vehicle, Age, Automation, and Interstate on Excitement. Table 10 presents the results of the multi-level modeling. There were significant main effects of Vehicle, Automation, and Interstate.

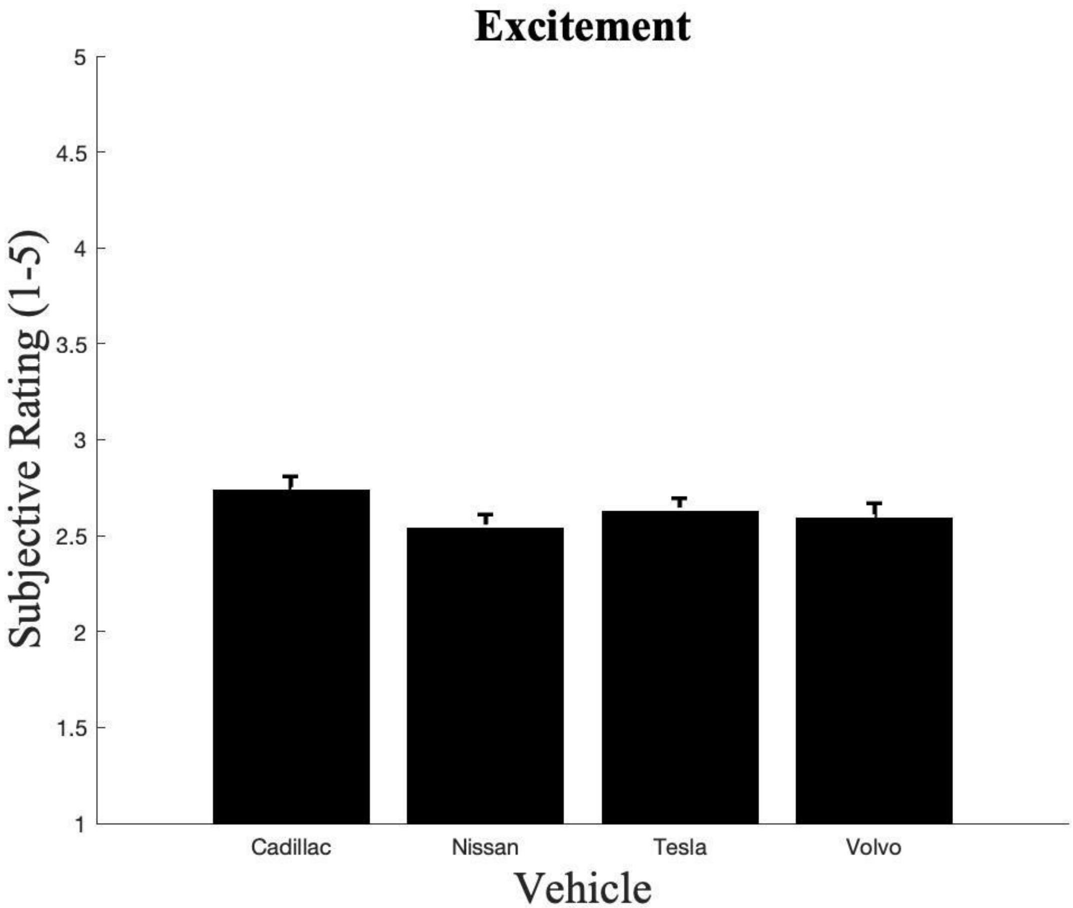

Figure 32. Excitement (1 = *very slightly or not at all*, 5 = *extremely*) plotted as a function of vehicle. Error bars reflect the standard error of the mean. The effect of Vehicle (Cadillac: M = 2.74, SD = 0.96; Nissan: M = 2.54, SD = 0.93; Tesla: M = 2.63, SD = 0.90; Volvo: M = 2.59, SD = 1.00) was significant ($\chi^2(3) = 42.56$, $p < 0.001$; $R^2 = 0.019$).

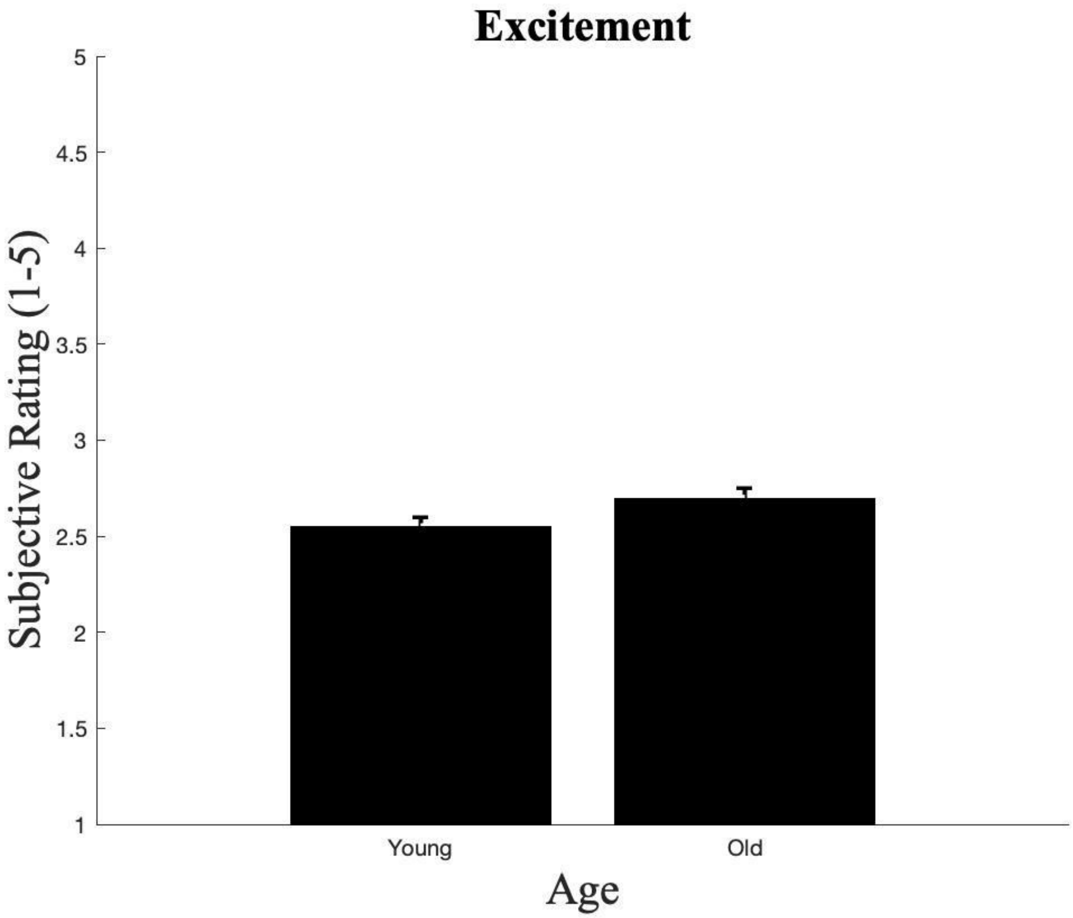

Figure 33. Excitement (1 = *very slightly or not at all*, 5 = *extremely*) plotted as a function of the participant's age. Error bars reflect the standard error of the mean. The effect of Age (Young: M = 2.55, SD = 0.91; Old: M = 2.70, SD = 0.98) was not significant ($\chi^2(1) = 1.05$, $p = 0.31$; $R^2 = 0.011$). The lack of significance was the result of entering the participant as a random intercept, effectively partialling out any variance between participants across the age spectrum.

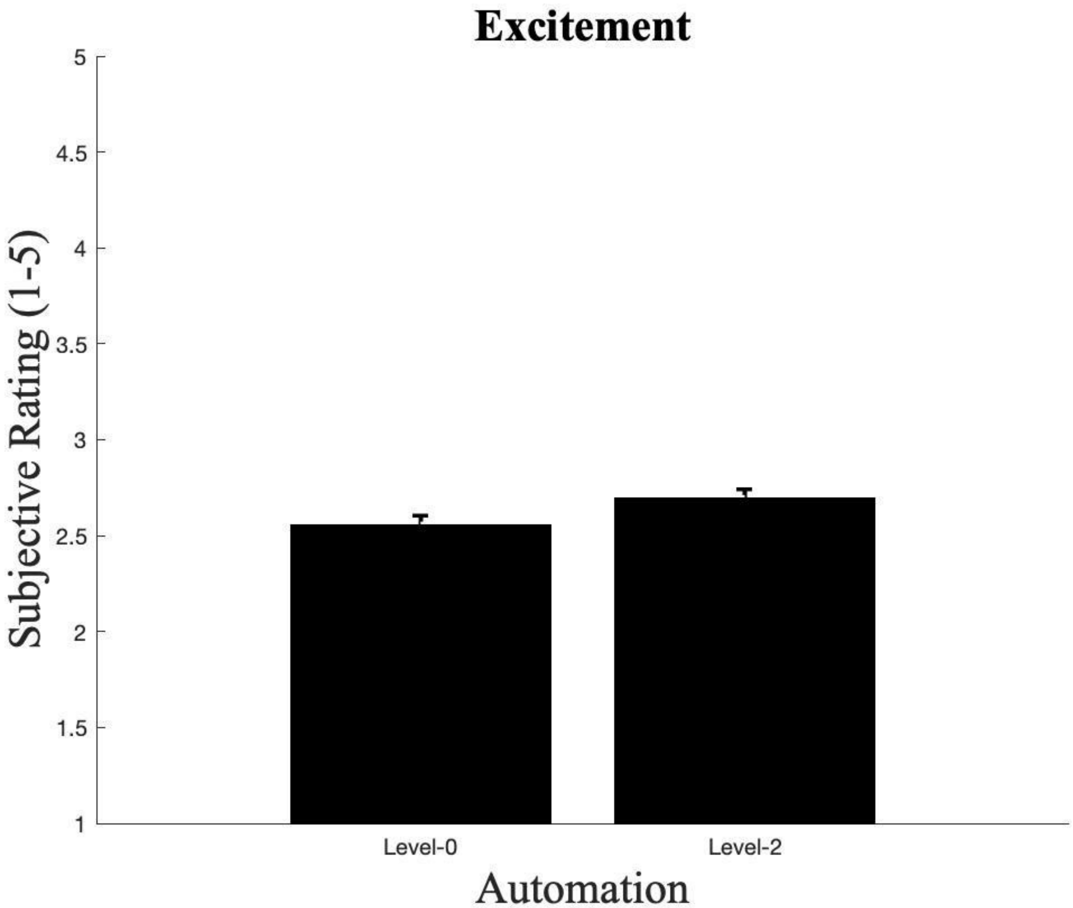

Figure 34. Excitement (1 = *very slightly or not at all*, 5 = *extremely*) plotted as a function of the level of automation. Error bars reflect the standard error of the mean. The effect of Automation (Level-0: M = 2.55, SD = 0.96; Level-2: M = 2.69, SD = 0.93) was significant (χ^2 (1) = 12.64, $p < 0.001$; $R^2 = 0.005$).

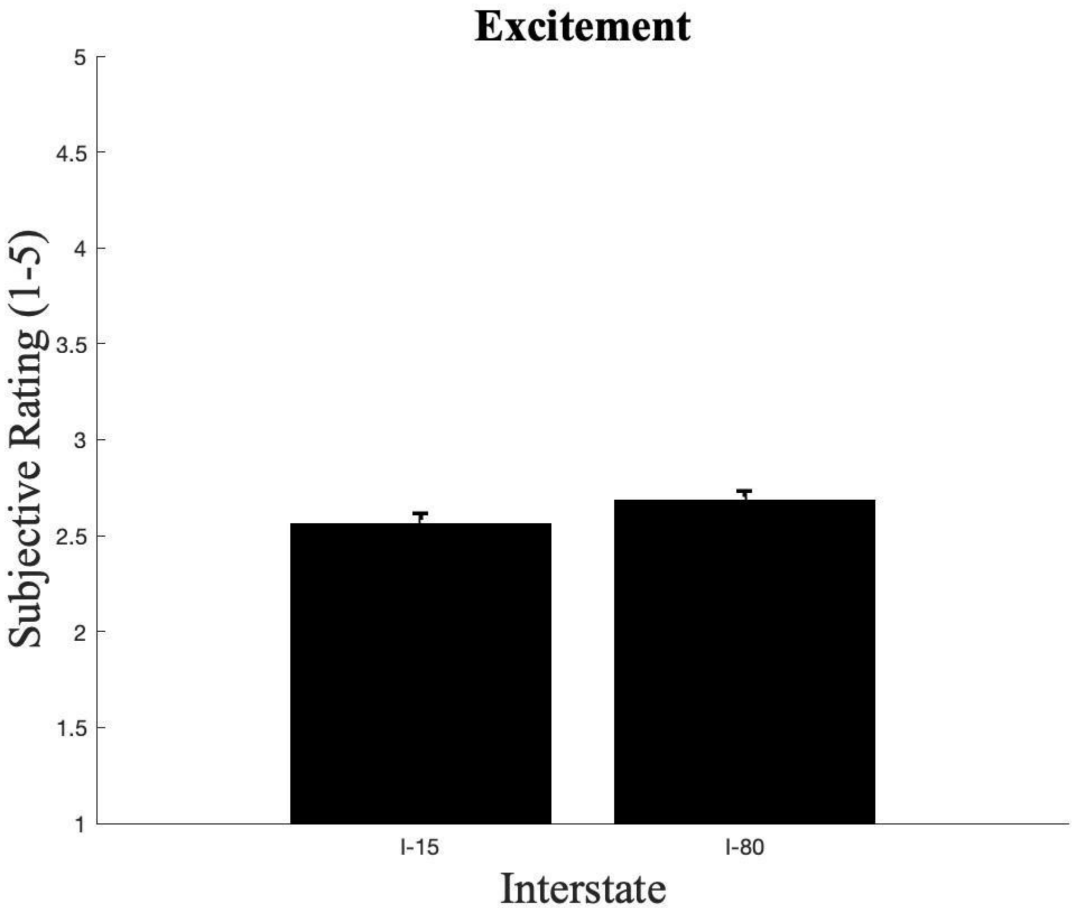

Figure 35. Excitement (1 = *very slightly* or not at all, 5 = *extremely*) plotted as a function of the Interstate. Error bars reflect the standard error of the mean. The effect of Interstate (I-15: M = 2.56, SD = 0.94; I-80: M = 2.69, SD = 0.95) was significant (χ^2 (1) = 9.39, *p* = 0.002; R^2 = 0.004).

Table 10. *Excitement*				
Variable	*Chi-square*	*df*	*p*	R^2
Vehicle	**42.564**	**3**	**<.001*****	**0.019**
Age	1.053	1	0.305	0.011
Automation	**12.642**	**1**	**<.001*****	**0.005**
Interstate	**9.386**	**1**	**0.002****	**0.004**
Vehicle x Age	2.908	3	0.406	0.002
Vehicle x Automation	3.407	3	0.333	0.001
Vehicle x Interstate	1.064	3	0.786	0.000
Age x Automation	3.210	1	0.073	0.001
Age x Interstate	0.826	1	0.364	0.001
Automation x Interstate	0.014	1	0.905	0.000
Vehicle x Age x Automation	10.408	10	0.405	0.004
Vehicle x Age x Interstate	6.759	10	0.748	0.002
Vehicle x Automation x Interstate	5.917	10	0.822	0.002
Age x Automation x Interstate	4.094	4	0.393	0.002
Vehicle x Age x Automation x Interstate	16.213	25	0.908	0.005

Note: R Squared values represent the Marginal R^2 difference score from the comparison model. Significance Codes: `**` < .01, `***’ < .001.

Discussion

The primary focus of this research was to compare the driver’s experience under partial automation (i.e., Level-2) with that of no automation (i.e., Level-0). The study evaluated participants between the ages of 21-64 operating four vehicles supporting Level-2 automation on relatively straight Interstate (I-15) and more curvy Interstate (I-80) roadways. Under Level-2 automation (compared to Level-0 automation), parietal alpha and DRT hit rates were lower and DRT reaction time was longer, suggesting that participants paid more attention to the driving environment (see Table 11). Participants also reported driving under partial automation to be more exciting and resulting in more nervousness. The magnitude of these effects was small.

Table 11. *Summary of Significant Main Effects from the Multilevel Modeling*				
	Vehicle	Age (Young vs Older)	Automation (Level-0 vs Level-2)	Interstate (I-15 vs I-80)
Parietal Alpha			*↓	*↓
Heart Rate	*	* ↓		
RMSSD	*	* ↓		
DRT-Reaction Time	*		* ↑	* ↑
DRT-Hit Rate	*	* ↓	* ↓	* ↓
Nervousness	*		* ↑	
Inattention		* ↓		* ↓
Excitement	*		* ↑	* ↑

Note. (* = $p < .05$). '↑' reflects a significant increase from one level of a variable to another, whereas '↓' reflects a significant decrease from one level of a variable to another. Red symbols denote an R^2 difference accounting for at least 1% of the variance.

Taken together, the data establish that participants in the current study remained engaged with the driving task when Level-2 automation was enabled. Under Level-2 automation, the vehicle has combined functions like acceleration and steering, *but the driver must remain engaged with the driving task and monitor the environment at all times* (SAE, 2016). The pattern of data observed in the current on-road evaluation (as well as a pilot study reported in Appendix D) indicate that the participants in the current study met these requirements for Level-2 automation. If anything, the data suggest slightly enhanced engagement with the driving task and enhanced monitoring of the driving environment when participants operated the vehicle under Level-2 automation.

The differences we observed between Level-0 and Level-2 automation were small. In fact, the main effect of Automation accounted for between 0.0% and 2.7% of the variance for the eight dependent measures reported in the results section. Even so, the experimental design, with 192 5-hour participant testing sessions, was sufficiently powered to detect an effect accounting for 0.5% of the variance 95% of the time. Even effects that accounted for less than 0.5% of the variance trended in the direction of more attention allocated to driving under Level-2 automation.

In this context, the informal comment from one participant after testing all four vehicles was illuminating. This participant noted that the partial automation seemed to work well, but that "you had to keep an eye on the vehicle to make sure it was doing what it was supposed to do." This participant's comment is consistent with the empirical data reported above and is in line with the requirements for Level-2 automation.

The study also found that driver engagement was modulated by the driving environment. Specifically, parietal alpha and DRT hit rates were lower and DRT reaction time was longer on I-80, a more winding section of roadway east of Salt Lake City, as compared to the straighter section of I-15 north of Salt Lake City. Likewise, the pilot data collected on the very straight, low-density traffic section of I-80 west of Salt Lake City found even lower levels of driver engagement (see Appendix D). That is, driver engagement as measured with EEG and DRT was systematically modulated by the level of difficulty of the driving environment. Interestingly, this pattern held for both Level-0 and Level-2 automation, suggesting that the cues that drivers use to allocate attention to the roadway are similar in both contexts. As driving becomes more demanding, it appears as though motorists allocate more attention to the roadway, and they do so even when Level-2 automation is enabled.

By and large, the pattern of driver behavior we observed was similar for younger and older drivers. In fact, those cases where there were significant interactions with age accounted for a very small proportion of variance and were not indicative of a systematic trend. This finding is noteworthy, because it suggests that drivers of all ages, at least across the range covered in the current study, were able to remain engaged with the driving task and monitor the environment. By contrast, compared to younger drivers, older drivers find interacting with the complex in-vehicle infotainment systems that are found in these vehicles to be much more difficult and mentally demanding (Cooper et al., 2019).

The assessment of driver engagement was informed by a combination of behavioral, psychophysiological, and subjective measures. Together, these measures provided an assessment of the driver's engagement with the driving task, even under Level-2 automation where the requirements for visual monitoring increase and the requirements for manual control decrease. The behavioral measures obtained with the DRT and the subjective measures are similar to those used in prior research to evaluate in-vehicle infotainment systems (e.g., Strayer, et al., 2019, Cooper, et al., 2019).

One complementary measure used in this study was a spectral analysis of the parietal EEG. Higher parietal alpha power provides an index of lower visual engagement whereas lower parietal alpha power provides evidence of higher visual engagement with the driving environment (Bowman et al., 2017; Foxe & Snyder, 2011). As shown in Appendix A, parietal alpha power was highest when the participant's eyes were closed (when the vehicle was parked) and was lowest under Level-2 automation. This provides direct evidence that participants in this study did not disengage from the driving task when operating under Level-2 automation. Although this brain-based EEG measure appears to be very diagnostic in assessing driver engagement in an experimental setting, it is currently difficult to incorporate in everyday driving contexts. Nevertheless, this brain-based measure has the potential to provide a sensitive on-line measure of driver engagement and arousal that is better than current methods used to make this assessment.

Limitations and Future Directions

This research provides a snapshot of driver arousal and workload when operating a vehicle under partial automation. Research participants for the study were experienced with manually operating a vehicle (i.e., Level-0 automation) but most were naïve to the partial automation in the Level-2 research vehicles that they drove (confirmed in a demographics survey). The relative inexperience of drivers with Level-2 automation may have affected several of the outcome measures that were collected. Indeed the finding that participants allocated more attention to the driving task when automation was engaged is consistent with the performance expectations laid out by Dunn, Dingus, & Soccoloch (2019) for drivers in the Novelty phase of use. Accordingly, it is possible that additional familiarity with Level-2 automation would decrease driver excitement and nervousness and increase trust and acceptance. Future research should evaluate the effect of novelty and experience on driver arousal with Level-2 automation across an extended time period.

The research protocol required that a researcher was present in the vehicle during all data collection. Researchers played a critical role in ensuring that data were accurately recorded, that the participant completed all specified conditions, and that safety standards were met. It is unclear if and how the participant's behavior may change in the absence of a researcher or data monitoring equipment. For example, drivers were not allowed to use their phones, play music, unroll the window, or engage in any behavior that could have been a distraction or otherwise moderate arousal. Given the reported prevalence of these behaviors in the naturalistic driving environment (Dingus et al., 2016), it is possible that drivers would have behaved differently had they been given the opportunity. It is also possible that the inclusion of the DRT measurements may have increased the arousal level of the participants.

Future research should evaluate driver performance using both a combination of naturalistic and controlled experimental methods to compare and contrast the observed driver behavior in both levels of automation. Care needs to be taken in the naturalistic portion of the study to equate the driving demands (e.g., using a random assignment of days when automation can be enabled). Some driving behaviors may not be observable in controlled environments while many aspects of driver arousal and workload may go undetected in a naturalistic environment. It will be important to perform this hybrid approach over longer intervals (e.g., 2 months) to see how trust, acceptance, and driving behavior change over time. If we observe the same pattern of data in the hybrid study, then it would suggest that drivers are able to sustain their attention to the driving task and monitor the driving environment. However, it is possible that the pattern of data may change over longer intervals and/or that the behavior observed in the naturalistic portion of the study does not align with what is observed in the experimental portion of the study.

Summary

This research compared the driver's experience when operating a vehicle under partial automation and the same vehicle without automation. Under Level-2 automation, drivers must remain engaged with the driving task and monitor the environment at all times. We performed an on-road evaluation of drivers between the ages of 21-64 operating up to four different vehicles that supported Level-2 automation on both urban and more rural roadways.

Consistent with the performance requirements established in the SAE framework for Level-2 automation, the data provide evidence for slightly enhanced engagement with the driving task and enhanced monitoring of the driving environment when participants operated the vehicle under partial automation. Future research should examine driver's behavior over longer intervals to determine if this pattern is altered with greater experience.

Acknowledgement

Equipment comprising the Detection Response Time (DRT) task was provided by Red Scientific, which was founded by the co-investigator, JMC.

References

Agrafioti, F., Hatzinakos, D., & Anderson, A. K. (2012). ECG pattern analysis for emotion detection. *IEEE Transactions on Affective Computing*, *3*(1), 102-115. https://doi.org/10.1109/T-AFFC.2011.28

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48. http://dx.doi.org/10.18637/jss.v067.i01

Bengler, K., Kohlmann, M., & Lange, C. (2012). Assessment of cognitive workload of in-vehicle systems using a visual peripheral and tactile detection task setting. *Work*, *41*(Supplement 1), 4919-4923. https://doi.org/10.3233/WOR-2012-0786-4919

Bowman, A. D., Griffis, J. C., Visscher, K. M., Dobbins, A. C., Gawne, T. J., DiFrancesco, M. W., & Szaflarski, J. P. (2017). Relationship between alpha rhythm and the default mode network: An EEG-fMRI study. *Journal of Clinical Neurophysiology, 34*(6), 527-533.

Bruyas, M. P. & Dumont, L. (2013). Sensitivity of Detection Response Task (DRT) to the driving demand and task difficulty. *7th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*. Symposium presented at Bolton Landing, N.Y., June 17–20, 2013.

Chuang, C. H., Cao, Z., King, J. T., Wu, B. S., Wang, Y. K., & Lin, C. T. (2018). Brain electrodynamic and hemodynamic signatures against fatigue during driving. *Frontiers in Neuroscience, 12,* 181.

Cohen, R. A. (2011). Yerkes–Dodson Law. In J. S. Kreutzer, J. DeLuca, & B. Caplan (Eds.), *Encyclopedia of Clinical Neuropsychology,* 2737–2738. https://doi.org/10.1007/978-0-387-79948-3_1340

Cohen, M. X. (2014). *Analyzing Neural Time Series Data: Theory and Practice.* MIT press.

Cooper, J. M., Castro, S. C., & Strayer, D. L. (2016). Extending the Detection Response Task to simultaneously measure cognitive and visual task demands. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 60*(1), 1962-1966.

Cooper, J. M., Wheatley, C. L., McCarty, M. M., Motzkus, C. J., Lopes, C. L., Erickson, G. G., … & Strayer, D. L. (2019). *Age-Related Differences in the Cognitive, Visual and Temporal Demands of In-Vehicle Information Systems* (Technical Report). Washington, D.C.: AAA Foundation for Traffic Safety.

Delorme, A. & Makeig, S. (2004). EEGLAB: An open-source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods, 134*, 9–21. doi:10.1016/j.jneumeth.2003.10.009

Derakshan, N. & Eysenck, M. W. (2009). Anxiety, processing efficiency, and cognitive performance: New developments from attentional control theory. *European Psychologist, 14*(2), 168–176. https://doi.org/10.1027/1016-9040.14.2.168

Dingus, T. A., Guo, F., Lee, S., Antin, J. F., Perez, M., Buchanan-King, M., & Hankey, J. (2016). Driver crash risk factors and prevalence evaluation using naturalistic driving data. *Proceedings of the National Academy of Sciences, 113*(10), 2636-2641.

Drummond, S. P., Bischoff-Grethe, A., Dinges, D. F., Ayalon, L., Mednick, S. C., & Meloy, M. J. (2005). The neural basis of the psychomotor vigilance task. *Sleep*, *28*(9), 1059-1068.

Dunn, N., Dingus, T., & Soccolich, S. (2019). *Understanding the Impact of Technology: Do Driver Assistance and Semi-Automated Vehicle Systems Lead to Improper Driving Behavior?* (Technical Report). Washington, D.C.: AAA Foundation for Traffic Safety.

Endsley, M. R. (2017). Autonomous driving systems: A preliminary naturalistic study of the Tesla Model S. *Journal of Cognitive Engineering and Decision Making*, *11*(3), 225-238.

Fischer, P. (2016). *Wake Up Call! Understanding Drowsy Driving and What States Can Do*. Washington, D.C.: Governors Highway Safety Association.

Foxe, J. J. & Snyder, A. C. (2011). The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention. *Frontiers in Psychology*, 2, 154.

Fridman, L., Brown, D. E., Glazer, M., Angell, W., Dodd, S., Jenik, B., … Reimer, B. (2019). MIT Advanced Vehicle Technology Study: Large-scale naturalistic driving study of driver behavior and interaction with automation. *IEEE Access*, *7*, 102021–102038. https://doi.org/10.1109/ACCESS.2019.2926040

Gartenberg, D., McGarry, R., Pfannenstiel, D., Cisler, D., Shaw, T., & Parasuraman, R. (2012). Development of a neuroergonomic application to evaluate arousal. In K. M. Stanney & K. S. Hale (Eds.), *Advances in Cognitive Engineering and Neuroergonomics,* 239-238. https://doi.org/10.1201/b12313

Gaspar, J. & Carney, C. (2019). The effect of partial automation on driver attention: A naturalistic driving study. *Human Factors, 61*(8), 1261–1276. https://doi.org/10.1177/0018720819836310

Goldman, R. I., Stern, J. M., Engel Jr, J., & Cohen, M. S. (2002). Simultaneous EEG and fMRI of the alpha rhythm. *Neuroreport, 13*(18), 2487.

Graham, J. W., Taylor, B. J., Olchowski, A. E., & Cumsille, P. E. (2006). Planned missing data designs in psychological research. *Psychological Methods, 11*(4), 323.

Gratton, G., Coles, M. G., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and clinical neurophysiology, 55*(4), 468-484.

Greenlee, E. T., DeLucia, P. R., & Newton, D. C. (2019). Driver vigilance in automated vehicles: Effects of demands on hazard detection performance. *Human Factors, 61*(3), 474–487. https://doi.org/10.1177/0018720818802095

Grier, R. A., Warm, J. S., Dember, W. N., Matthews, G., Galinsky, T. L., Szalma, J. L., & Parasuraman, R. (2003). The vigilance decrement reflects limitations in effortful attention, not mindlessness. *Human Factors*, *45*(3), 349–359. https://doi.org/10.1518/hfes.45.3.349.27253

Gunzelmann, G., Gluck, K. A., Price, S., Van Dongen, H. P., & Dinges, D. F. (2007). Decreased arousal as a result of sleep deprivation. *Integrated Models of Cognitive Systems,* 243-253.

Habib, K., (2017). Automatic Vehicle Control Systems. *National Highway Traffic Safety Administration - ODI Resume.* PE 16-007. Retrieved from: https://static.nhtsa.gov/odi/inv/2016/INCLA-PE16007-7876.pdf

Hatfield, N., Yamani, Y., Palmer, D. B., Yahoodik, S., Vasquez, V., Horrey, W. J., & Samuel, S. (2019). Analysis of visual scanning patterns comparing drivers of simulated L2 and L0 systems. *Transportation Research Record*, *2673*(10), 755–761. https://doi.org/10.1177/0361198119852339

International Organization for Standardization (2016). *Road Vehicles – Transport Information and Control Systems – Detection Response Task (DRT) for Assessing Attentional Effects of Cognitive Load in Driving* (Rep. ISO 17488) Geneva, Switzerland: International Organization for Standardization. ISO/TC 22/SC 39.

Jasper, H. H. (1958). The ten-twenty electrode system of the International Federation. *Electroencephalogr. Clin. Neurophysiol., 10,* 370-375.

Jiao, K., Li, Z., Chen, M., Wang, C., & Qi, S. (2004). Effect of different vibration frequencies on heart rate variability and driving fatigue in healthy drivers. *International Archives of Occupational and Environmental Health, 77*(3), 205–212. https://doi.org/10.1007/s00420-003-0493-y

Käthner, I., Wriessnegger, S. C., Müller-Putz, G. R., Kübler, A., & Halder, S. (2014). Effects of mental workload and fatigue on the P300, alpha and theta band power during operation of an ERP (P300) brain-computer interface. *Biological Psychology, 102,* 118-129.

Krisher, T. (2018, May 18). AP FACT CHECK: Tesla safety claims aren't quite right. Retrieved from: https://apnews.com/be561d4323034316695f64b5c158dbf7c

Langner, R. & Eickhoff, S. B. (2013). Sustaining attention to simple tasks: A meta-analytic review of the neural mechanisms of vigilant attention. *Psychological Bulletin, 139*(4), 870–900. https://doi.org/10.1037/a0030694

Little, T. D. & Rhemtulla, M. (2013). Planned missing data designs for developmental researchers. *Child Development Perspectives, 7*(4), 199-204.

Lohani, M., Payne, B. R., & Strayer, D. L. (2019). A review of psychophysiological measures to assess cognitive states in real-world driving. *Frontiers in Human Neuroscience,* 13.

Manly, T., Robertson, I. H., Galloway, M., & Hawkins, K. (1999). The absent mind: Further investigations of sustained attention to response. *Neuropsychologia, 37*(6), 661–670. https://doi.org/10.1016/S0028-3932(98)00127-4

Mehler, B., Reimer, B., & Wang, Y. (2011). A comparison of heart rate and heart rate variability indices in distinguishing single-task driving and driving under secondary cognitive workload. *Driving Assessment Conference,* 590–597.

Mun, S., Whang, M., Park, S., & Park, M.C. (2017). Effects of mental workload on involuntary attention: A somatosensory ERP study. *Neuropsychologia, 106,* 7–20. https://doi.org/10.1016/j.neuropsychologia.2017.08.021

National Transportation Safety Board (2017). Collision between a car operating with automated vehicle control systems and a tractor-semitrailer truck near Williston, Florida, May 7, 2016 (Highway Accident Report NTSB/HAR-17/02). Washington, D.C.: NTSB.

National Transportation Safety Board (2018).Vehicle Automation Report. HWY18MH010. Washington, D.C.: NTSB.

Quality Control Systems Corporation (2019). NHTSA's Implausible Safety Claim for Tesla's Autosteer Driver Assistance System. Crownsville, Maryland. Retrieved from: http://www.safetyresearch.net/Library/NHTSA_Autosteer_Safety_Claim.pdf

Raz, A. & Buhle, J. (2006). Typologies of attentional networks. *Nature Reviews Neuroscience*, *7*(5), 367–379. https://doi.org/10.1038/nrn1903

R Core Team (2019). R: A language and environment for statistical computing. R Project for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/.

REDCap Grant Citation: (8UL1TR000105 (formerly UL1RR025764) NCATS/NIH)

SAE (2016). Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles (Surface Vehicle Recommended Practice: Superseding J3016 Jan 2014), *SAE International*. Retrieved from https://www.sae.org/standards/content/j3016_201806/

Soto, C. J. & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology, 113*(1), 117–143. https://doi.org/10.1037/pspp0000096

Spinella, M. (2007). Normative data and a short form of the Barratt Impulsiveness Scale. *International Journal of Neuroscience*, *117*(3), 359–368. https://doi.org/10.1080/00207450600588881

Strayer. D. L., Cooper, J. M., Goethe, R. M., McCarty, M. M., Getty, D., & Biondi, F. (2019). Assessing the visual and cognitive demands of using in-vehicle information systems. *Cognitive Research: Principles & Implication,* 4:18. https://doi.org/10.1186/s31235-019-0166-3

Tefft, B. C. (2012). Prevalence of motor vehicle crashes involving drowsy drivers, United States, 1999–2008. *Accident Analysis & Prevention, 45,* 180–186. https://doi.org/10.1016/j.aap.2011.05.028

Temple, J. G., Warm, J. S., Dember, W. N., Jones, K. S., LaGrange, C. M., & Matthews, G. (2000). The effects of signal salience and caffeine on performance, workload, and stress in an abbreviated vigilance task. *Human Factors, 42*(2), 183–194. https://doi.org/10.1518/001872000779656480

Tesla Vehicle Safety Report. (2019, January 8). Retrieved from: https://www.tesla.com/VehicleSafetyReport

Tozman, T., Magdas, E. S., MacDougall, H. G., & Vollmeyer, R. (2015). Understanding the psychophysiology of flow: A driving simulator experiment to investigate the relationship between flow and heart rate variability. *Computers in Human Behavior, 52,* 408–418. https://doi.org/10.1016/j.chb.2015.06.023

Vodanovich, S. J., Wallace, J. C., & Kass, S. J. (2005). A confirmatory approach to the factor structure of the boredom proneness scale: Evidence for a two-factor short form. *Journal of Personality Assessment, 85,* 295-303. doi:10.1207/s15327752jpa8503_05

Watson, D., Clark, L. A. & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063–1070. doi:10.1037/0022-3514.54.6.1063.

Winter, B. (2013). Linear models and linear mixed-effects models in R with linguistic applications. *ArXiv:1308.5499 [Cs],* 42.

Yarnell, A. M., Lopresti, M. L., Lanlokun, O. O., & Balkin, T. J. (1982). Sleep, rest, and recovery. *Human Neurobiology, 1*(3), 195-204.

Yerkes, R. M. & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology of Psychology, 18,* 459-482.

Young, R. A., Hsieh, L., & Seaman, S. (2013). The tactile detection response task: Preliminary validation for measuring the attentional effects of cognitive load. *Proceedings of the Seventh International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*. 71-77. https://doi.org/10.17077/drivingassessment.1469

Appendix A

Parietal Alpha Logic

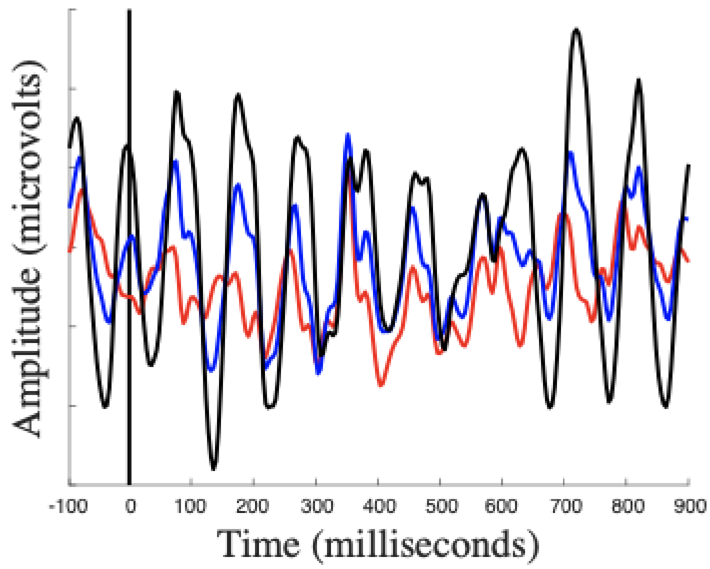

Figure A1. Sample EEG recorded in an older participant from Fz (red line), Cz (blue line), and Pz (black line) in the eyes-closed condition. In the figure, successive 1000 msec epochs are averaged together to create the grand average. Large alpha oscillations are clearly visible at Pz.

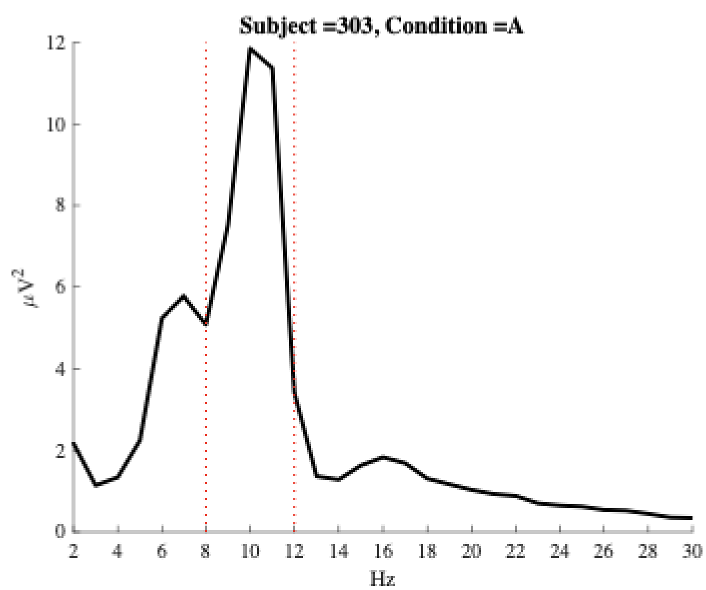

Figure A2. The power as a function of spectral frequency for parietal EEG (Pz) presented in Figure A1. Note the spike in power between 8-12 Hz, reflecting the alpha band.

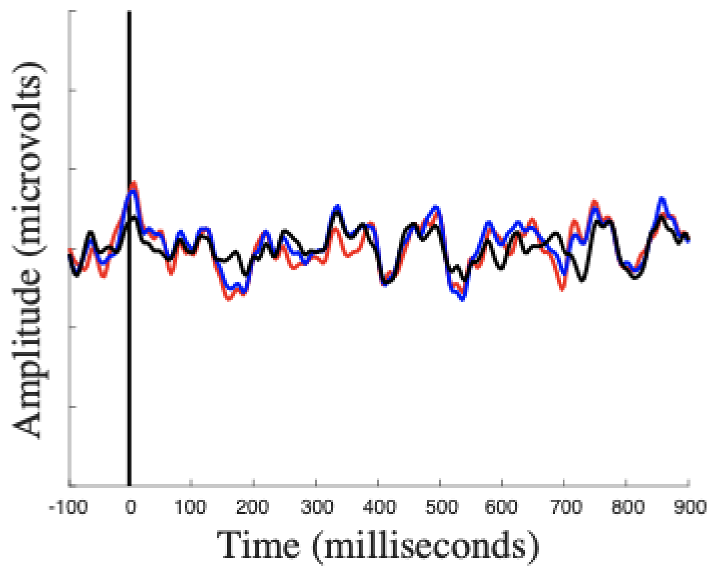

Figure A3. Sample EEG recorded in an older participant from Fz (red line), Cz (blue line), and Pz (black line) in the partially automated I-15 driving condition. In the figure, successive 1000 msec epochs are averaged together to create the grand average. Compared to Figure A1, the alpha oscillations at Pz are much less prominent.

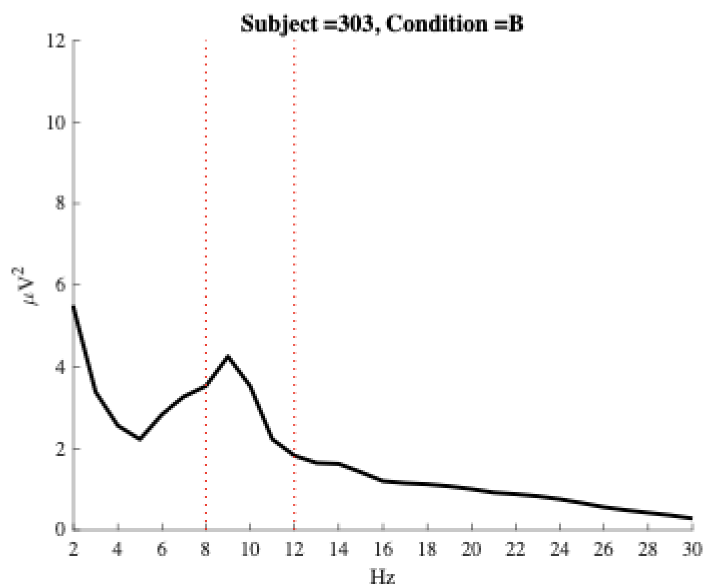

Figure A4. The power as a function of spectral frequency for parietal EEG (Pz) presented in Figure A2. Note the reduced power between 8-12 Hz, reflecting the alpha band compared to Figure A2.

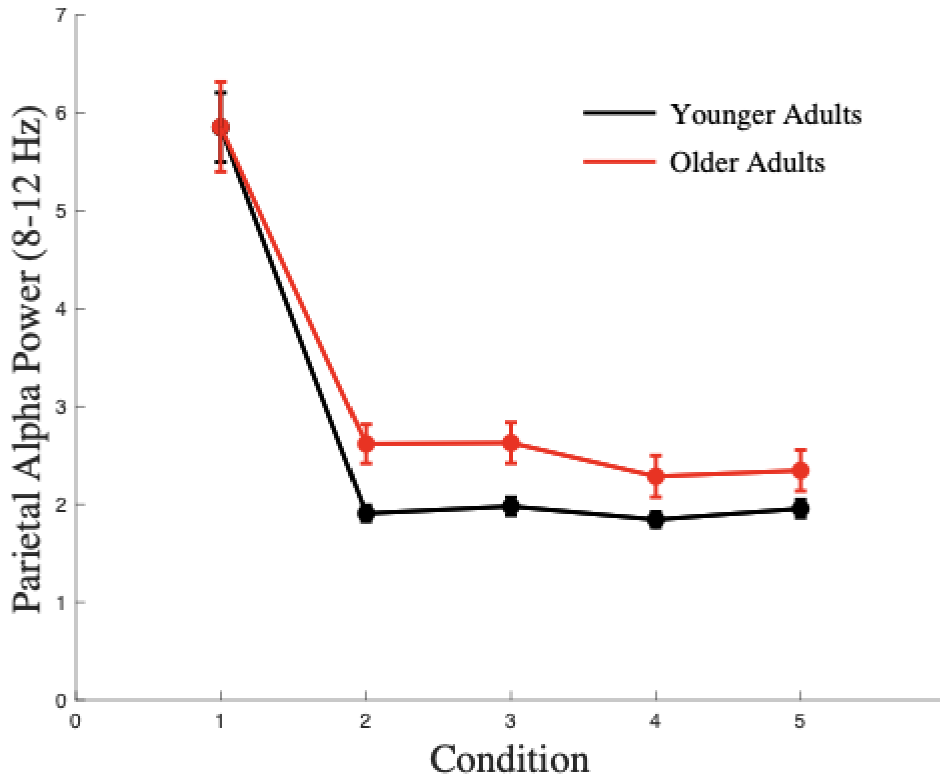

Figure A5. Parietal alpha power plotted as a function of experimental condition for younger adults (black line) and older adults (red line). Error bars reflect the standard error of the mean. Condition 1 is eyes-closed, condition 2 is driving under partial automation on I-15, condition 3 is driving under no automation on I-15, condition 4 is driving under partial automation on I-80, and condition 5 is driving under no automation on I-80.

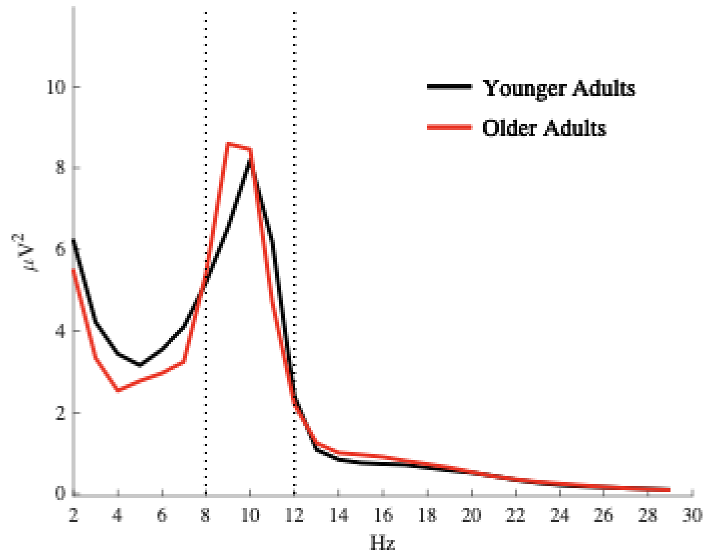

Figure A6. Parietal power in the eyes-closed condition for younger adults (black line) and older adults (red line).

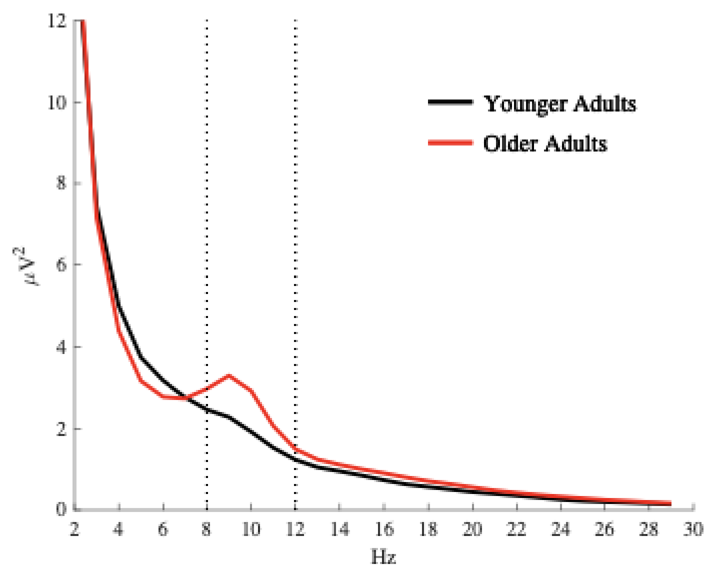

Figure A7. Parietal power in the Level-2 partial automation driving condition on I-15 for younger adults (black line) and older adults (red line).

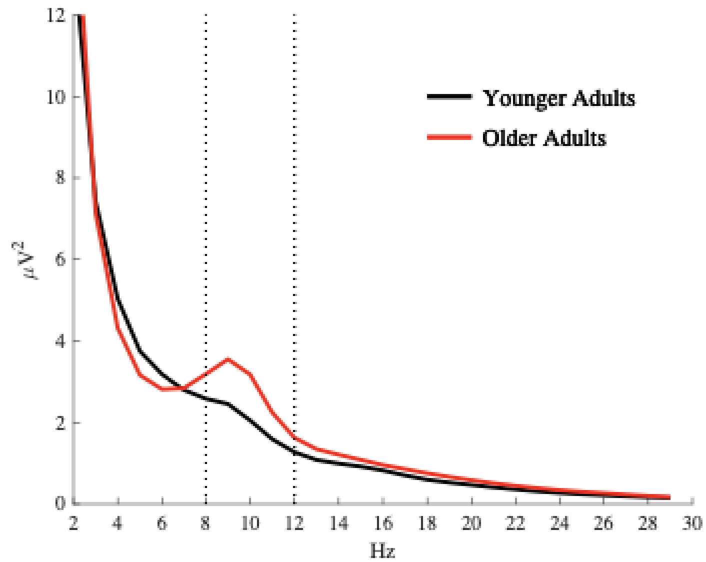

Figure A8. Parietal power in the manual driving condition on I-15 for younger adults (black line) and older adults (red line).

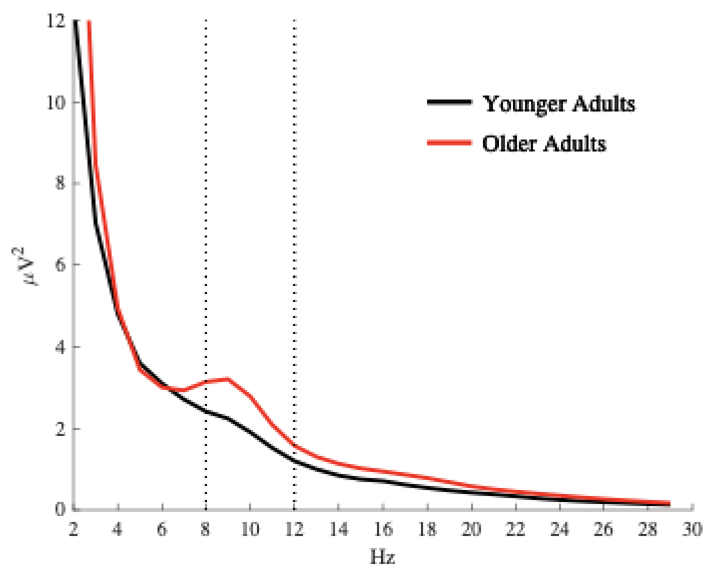

Figure A9. Parietal power in the partially automated driving condition on I-80 for younger adults (black line) and older adults (red line).

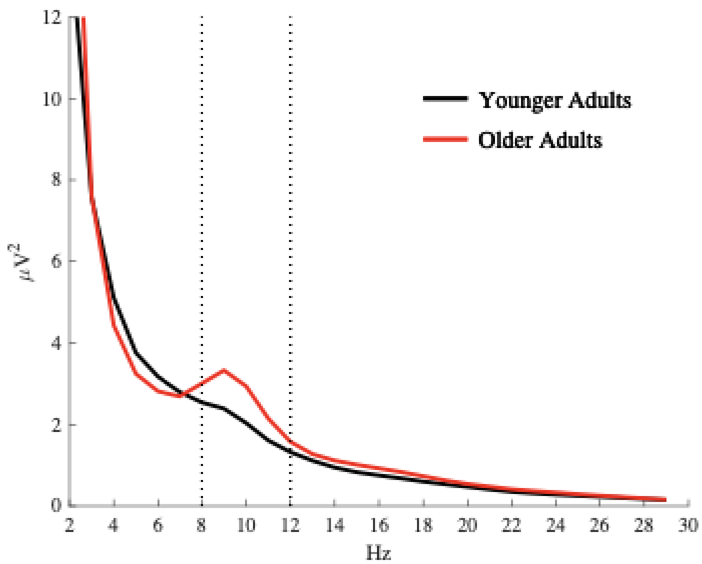

Figure A10. Parietal power in the manual driving condition on I-80 for younger adults (black line) and older adults (red line).

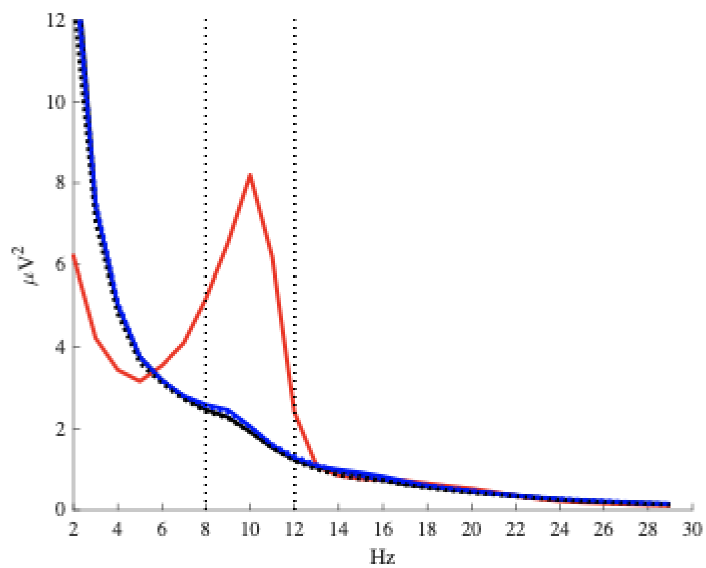

Figure A11. Parietal power plotted as a function of experimental condition for younger adults. Condition 1 (solid red line) is eyes-closed, condition 2 (black solid line) is driving under partial automation on I-15, condition 3 (blue solid line) is driving under no automation on I-15, condition 4 (black dashed line) is driving under partial automation on I-80, and condition 5 (blue dashed line) is driving under no automation on I-80.

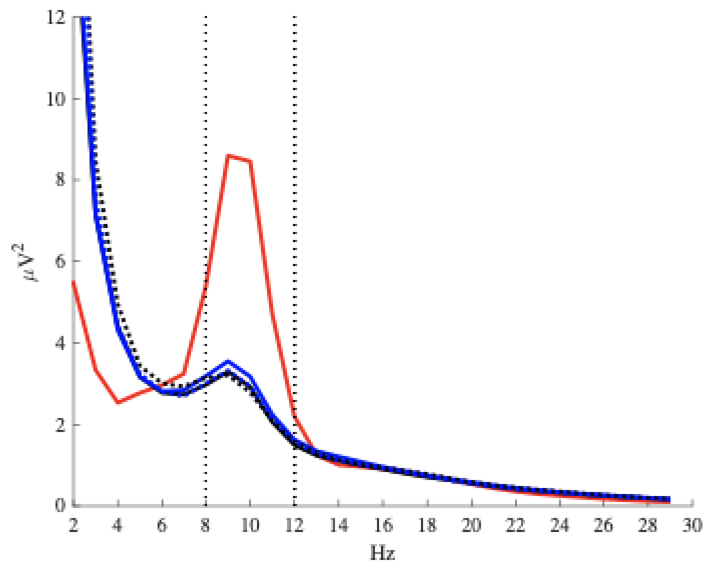

Figure A12. Parietal power plotted as a function of experimental condition for older adults. Condition 1 (solid red line) is eyes-closed, condition 2 (black solid line) is driving under partial automation on I-15, condition 3 (blue solid line) is driving under no automation on I-15, condition 4 (black dashed line) is driving under partial automation on I-80, and condition 5 (blue dashed line) is driving under no automation on I-80.

Appendix B

Questionnaires

Demographics and Driving History

This questionnaire provided basic demographic information and served as a final screening process to ensure participant eligibility on the day of testing.

Brief Compound Fatigue Instrument (BCFI)

The purpose of the BCFI was to evaluate self-reported fatigue. Participants answered:

- On a 10-point scale, anchored by severity from *absent* and *severe*
    - To what degree have you experienced fatigue today?
- On a 10-point scale, anchored by severity (1 = *none* to 10 = *severe*)
    - How severe is the fatigue which you have been experiencing?
    - To what degree has mental fatigue caused you distress this week?
    - To what degree do you feel fatigued currently when you walk or exercise?
    - To what degree do you feel fatigued currently when you focus/think/read?
- On a 4-point scale (4 = *Almost asleep,* 3 = *Very tired,* 2 = *Slightly tired,* 1 = *Completely alert/awake*), participants indicated their agreement regarding:
    - How drowsy do you think you will become while driving today?

Beliefs about Automated Driving Systems

Participants indicated their beliefs on a 6-point scale, with the sixth option listed as *no basis for judgment*. Participants answered:

- Anchored by *not at all dependable* and *highly dependable*
    - How dependable are automated driving systems?
- Anchored by *do not work well at all* and *work extremely well*
    - How well do automated driving systems work?
- Anchored by *not at all safe* and *highly safe*
    - How safe are automated driving systems?
- Anchored by *do not trust at all* and *trust completely*
    - How much do you trust automated driving systems?
- Anchored by *not at all useful* and *highly useful*
    - How useful are automated driving systems to you?
- Anchored by *highly negative* and *highly positive*
    - What is your overall evaluation of automated driving systems?
- Anchored by *disagree completely* and *agree completely*
    - Automated driving systems are susceptible to malfunctioning.
    - Automated driving systems handle a car well in terms of steering, acceleration, and braking.
    - Automated driving systems are prone to turning off unexpectedly.

Participants were then provided the following definitions:

Automation Levels:

No automation: full-time performance by the human driver of all aspects of the dynamic driving task, even when enhanced by warning or intervention systems.

Driver assistance automation: the driving mode-specific execution by a driver assistance system of either steering or acceleration/deceleration using information about the driving environment and with the expectation that the human driver performs all remaining aspects of the dynamic driving task Human driver and system.

Partial automation: the driving mode-specific execution by one or more driver assistance systems of both steering and acceleration/deceleration using information about the driving environment and with the expectation that the human driver performs all remaining aspects of the dynamic driving task.

Conditional automation: the driving mode-specific performance by an automated driving system of all aspects of the dynamic driving task with the expectation that the human driver will respond appropriately to a request to intervene.

High automation: the driving mode-specific performance by an automated driving system of all aspects of the dynamic driving task, even if a human driver does not respond appropriately to a request to intervene.

Full automation: the full-time performance by an automated driving system of all aspects of the dynamic driving task under all roadway and environmental conditions that can be managed by a human driver.

- Participants then rated the overall level of automation using a 6-point scale stating the levels of automation as it related to this question: (*No automation, Driver assistance automation, Partial automation, Conditional automation, High automation,* or *Full automation*):
  - How would you rate the overall level of automation in this car you will drive (or drove)?

Proneness to Boredom (Vodanovich, Wallace, & Kass, 2005)

The Proneness to Boredom questionnaire consisted of 12 statements.

- Participants indicated their boredom on a 7-point scale anchored by *highly disagree* and *highly agree*. Participants answered:
  - It is easy for me to concentrate on my activities.
  - I find it easy to entertain myself.
  - I get a kick out of most things I do.
  - In any situation, I can usually find something to do or see to keep me interested.
  - Many people would say that I am a creative or imaginative person.
  - Among my friends, I am the one who keeps doing something the longest.
  - Having to look at someone's home movies or travel slides bores me tremendously.

- Many things I have to do are repetitive and monotonous.
- It would be very hard for me to find a job that is exciting enough.
- Unless I am doing something exciting, even dangerous, I feel half-dead and dull.
- It seems that the same old things are on television or the movies all the time; it's getting old.
- When I was young, I was often in monotonous and tiresome situations.

Barratt Impulsivity Scales (Spinella, 2007)

The Barratt Impulsivity Scale questionnaire consisted of 15 statements.

- Participants indicated their impulsivity on a 4-point scale and anchored by *rarely/never* and *almost always/always*. Participants answered:
    - I plan tasks carefully.
    - I do things without thinking.
    - I don't ‘pay attention’.
    - I concentrate easily.
    - I save regularly.
    - I ‘squirm’ at plays or lectures.
    - I am a careful thinker.
    - I plan for job security.
    - I say things without thinking.
    - I act on impulse.
    - I get easily bored when solving through problems.
    - I act on the spur of the moment.
    - I buy things on impulse.
    - I am restless at lectures or talks.
    - I plan for the future.

Big Five Inventory-2 (Soto & John, 2017)

The Big Five Inventory-2 questionnaire consisted of 60 statements.

- Participants responded to questions on a 5-point scale anchored by *disagree strongly* and *agree strongly*. Participants answered:
    - Is outgoing, sociable.
    - Is compassionate, has a soft heart.
    - Tends to be disorganized.
    - Is relaxed, handles stress well.
    - Has few artistic interests.
    - Has an assertive personality.
    - Is respectful, treats others with respect.
    - Tends to be lazy.
    - Stays optimistic after experiencing a setback.
    - Is curious about many different things.
    - Rarely feels excited or eager.
    - Tends to find fault with others.

- Is dependable, steady.
- Is moody, has up and down mood swings.
- Is inventive, finds clever ways to do things.
- Tends to be quiet.
- Feels little sympathy for others.
- Is systematic, likes to keep things in order.
- Can be tense.
- Is fascinated by art, music, or literature.
- Is dominant, acts as a leader.
- Starts arguments with others.
- Has difficulty getting started on tasks.
- Feels secure, comfortable with self.
- Avoids intellectual, philosophical discussions.
- Is less active than other people.
- Has a forgiving nature.
- Can be somewhat careless.
- Is emotionally stable, not easily upset.
- Has little creativity.
- Is sometimes shy, introverted.
- Is helpful and unselfish with others.
- Keeps things neat and tidy.
- Worries a lot.
- Values art and beauty.
- Finds it hard to influence people.
- Is sometimes rude to others.
- Is efficient, gets things done.
- Often feels sad.
- Is complex, a deep thinker.
- Is full of energy.
- Is suspicious of others' intentions.
- Is reliable, can always be counted on.
- Keeps their emotions under control.
- Has difficulty imagining things.
- Is talkative.
- Can be cold and uncaring.
- Leaves a mess, doesn't clean up.
- Rarely feels anxious or afraid.
- Thinks poetry and plays are boring.
- Prefers to have others take charge.
- Is polite, courteous to others.
- Is persistent, works until the task is finished.
- Tends to feel depressed, blue.
- Has little interest in abstract ideas.
- Shows a lot of enthusiasm.
- Assumes the best about people.
- Sometimes behaves irresponsibly.
- Is temperamental, gets emotional easily.
- Is original, comes up with new ideas.

Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988)

The PANAS consisted of 18 questions.

- Participants responded to the first 13 items on a 5-point scale anchored by *very slightly or not at all* and *extremely*. Participants were asked to rate the extent to which they were:
    - Distressed
    - Excited
    - Confident
    - Scared
    - Enthusiastic
    - Irritable
    - Alert
    - Relaxed
    - Cheerful
    - Nervous
    - Sleepy
    - Bored
    - Comfortable
- Participants responded to the last 5 statements on a 5-point scale anchored by *disagree completely* and *agree completely*. Participants were asked to indicate their agreement with the following statements:
    - I engaged in activities unrelated to driving.
    - I focused on the road and driving safely during the entire drive.
    - I allowed my thoughts to wander during the drive.
    - I daydreamed during portions of the drive.
    - I thought about everyday problems and things I need to do during the drive.

Driving Experiences

The Driving Experiences questionnaire consisted of 12 statements on a 5-point scale anchored by *disagree completely* and *agree completely*.

- Participants indicated their agreement to the following statements:
    - I was able to relax when the automated driving systems were on.
    - I was able to engage in more activities unrelated to driving when the automated driving systems were on.
    - The automated driving system made traveling boring for me.
    - I was concerned that the automated driving systems would shut off unexpectedly.
    - The automated driving systems made traveling safer.
    - I was anxious and nervous when the automated driving systems were on.
    - The automated driving system made traveling more enjoyable.
    - The automated driving system reduced the stress of driving.
    - The automated driving system took the fun out of driving.
    - The automated driving system allowed me to think and daydream.
    - I was uncomfortable relinquishing control of the vehicle to the automated driving system on curvy and hilly roads.

    - I was uncomfortable relinquishing control of the vehicle to the automated driving systems in heavier traffic.

Intentions Measure

Participants responded to 6 statements on a 5-point scale anchored by *disagree completely* and *agree completely*.

- Participants were given the following prompt and then asked to indicate their agreement on the following statements: You may be driving vehicles equipped with automated driving systems such as adaptive cruise control and lane-keeping currently or in the near future. Please indicate your agreement or disagreement with the following statements about your willingness to use automated driving systems.
    - I would not feel comfortable using automated driving systems on most roads.
    - If I was tired or distracted, I would rely heavily on automated driving systems.
    - I would utilize the automated driving systems in a vehicle as much as possible.
    - I would not feel comfortable using the automated driving systems in a vehicle without monitoring it closely.
    - I am going to make sure that the next car I buy or lease has automated driving systems.
    - If I can afford it, I am going to buy or lease a car with automated driving systems.

Evaluations of Vehicle Automated Driving Systems (EVADS)

The EVADS questionnaire consisted of 14 items.

- The first 13 items of the EVADS questionnaire on a 5-point scale anchored by *disagree completely* and *agree completely*. Participants indicated their agreement to the following statements:
    - I trust the lane-keeping assistance system in the vehicle.
    - I think the lane-keeping assistance system in the vehicle is useful.
    - The lane-keeping assistance system in the vehicle is easy to learn and use.
    - The lane-keeping assistance system in the vehicle does not work well on curvy and hilly roads.
    - The lane-keeping assistance system in the vehicle does not work well in heavier traffic.
    - I trust the adaptive cruise control system in the vehicle.
    - I think the adaptive cruise control system in the vehicle is useful.
    - The adaptive cruise control system in the vehicle is easy to learn and use.
    - The adaptive cruise control system in the vehicle does not work well on curvy and hilly roads.
    - The adaptive cruise control system in the vehicle does not work well in heavier traffic.
    - The automated driving systems in the vehicle are safe.

    - The automated driving systems in the vehicle are susceptible to malfunctioning.
    - The automated driving systems in the vehicle handle a car well in terms of steering, acceleration, and braking.
- In the final item, participants evaluated the vehicles using a 6-point scale anchored by *highly negative* and *highly positive* as well as a *no basis for judgment* option. Participants were asked:
    - What is your overall evaluation of the automated driving systems in the vehicle?

Appendix C

Beliefs about Automation

Beliefs about Automated Systems before Driving the Test Vehicle

Participants indicated their beliefs about and expectations of automated systems prior to driving in the study (see Table C1). Responses of “no basis for judgment” were excluded from the analyses. The mean ratings of the dependability, safety, usefulness, handling, and workability of automated driving systems, and overall evaluations were significantly higher than the midpoint of 3 on a 5-point scale where higher scores indicated greater favorability. Thus, participants’ beliefs about automated systems were generally favorable prior to driving.

At the same time, a sizable portion of the participants expressed reservations about the automation. 79.0% of participants agreed (somewhat or completely) that “automated driving systems are prone to malfunctioning” and 63.0% of participants agreed that “automated driving systems are prone to turning off unexpectedly.” Moreover, only 27.1% of participants reported that they trust automated driving systems a lot or completely, 52.5% indicated that they trust them somewhat and 20.3% indicated that they trust them a little or not at all. Similarly, only 33.9% of participants reported that automated driving systems are reliably dependable or highly dependable, 50.9% indicated that they are somewhat dependable and 15.1% indicated that they are slightly or not at all dependable.

Finally, 47.2 % of participants reported that they thought automated driving systems are safe or highly safe, 39.6% indicated that they are somewhat safe and 13.2% indicated that they are slightly or not at all safe.

What is clear from the pre-driving survey is that participants were split about halfway on their beliefs of the trustworthiness, dependability and safety of the automated systems.

Table C1. *Mean beliefs about automated systems before driving (N = 71).*

	Mean	*SD*
How dependable are automated driving systems? 1 = *not at all*, 5 = *highly dependable*	3.25*	0.85
How well do automated driving systems work? 1 = *do not work well at all*, 5 = *work extremely well*	3.53***	0.83
How safe are automated driving systems? 1 = *not at all safe*, 5 = *highly safe*	3.38**	0.77
How much do you trust automated driving systems? 1 = *do not trust at all*, 5 = *trust completely*	3.08	0.79
How useful are automated driving systems to you? 1 = *not at all useful*, 5 = *highly useful*	3.32*	1.18
What is your overall evaluation of automated driving systems? 1 = *highly negative*, 5 = *highly positive*	3.77***	0.89
Automated driving systems are susceptible to malfunctioning. 1 = *Disagree completely*, 5 = *Agree completely*	4.02***	0.94
Automated driving systems handle a car well in terms of steering, acceleration, and braking. 1 = *Disagree completely*, 5 = *Agree completely*	3.90***	0.75
Automated driving systems are prone to turning off unexpectedly. 1 = *Disagree completely*, 5 = *Agree completely*	3.09	1.01

Note: SD = standard deviation. Significantly different than midpoint. * $p < 0.05$, ** $p < 0.01$, *** $p < .001$.

Driving Experiences with the Automated Systems

Participants reported their experiences with and future intentions toward automated driving systems after completing their last drive in the study. Table C2 presents participants' mean level of agreement with statements about their experiences with the automation. The means suggest that the automated systems generally made driving and traveling more favorable for participants. They generally agreed that the automation made driving safer, more enjoyable, and less stressful, and that they were able to relax when the automated driving systems were on. Most participants did not agree that the automation made driving boring or took the fun out of driving, or that they were more anxious and nervous when the automated systems were on.

At the same time, 64.8% of the drivers agreed (somewhat or completely) that they were not comfortable relinquishing control of the vehicle to the automated systems on curvy and hilly roads. Similarly, 53.5% reported that they were not comfortable relinquishing control of the vehicle to the automated systems in heavy traffic. Finally, 52.1% were concerned that the automation would shut off unexpectedly. Thus, most participants were wary about the automation in difficult driving conditions and many expressed concerns about the automated systems shutting off.

Table C2. *Mean experiences with automated driving systems (N = 71).*

	Mean	*SD*
I was able to relax when the automated driving systems were on.	3.70***	1.01
I was able to engage in more activities unrelated to driving when the automated driving systems were on.	2.63*	1.32
The automated driving system made traveling boring for me.	2.27***	1.18
I was concerned that the automated driving systems would shut off unexpectedly.	3.04	1.26
The automated driving system made traveling safer.	3.46***	1.00
I was anxious and nervous when the automated driving systems were on.	2.37***	1.25
The automated driving system made traveling more enjoyable.	3.59***	1.09
The automated driving system reduced the stress of driving.	3.31*	1.18
The automated driving system took the fun out of driving.	2.41***	1.33
The automated driving system allowed me to think and daydream.	2.82	1.32
I was uncomfortable relinquishing control of the vehicle to the automated driving system on curvy and hilly roads.	3.49***	1.25
I was uncomfortable relinquishing control of the vehicle to the automated driving system in heavier traffic.	3.15	1.35

Note: Scales anchored by 1 = *disagree completely* and 5 = *agree completely*. SD = standard deviation. Significantly different than midpoint. * $p < 0.05$, *** $p < .001$.

Intentions to Use and Purchase Automated Driving Systems

Table C3 presents participants' mean level of agreement with statements about their intentions toward automated driving systems. Most participants conveyed a willingness to purchase and use automated driving systems in the future. In fact, 67.6% of participants agreed (somewhat or completely) that they would utilize the automated systems in a vehicle as much as possible while driving. Similarly, 64.8% indicated that they would not feel uncomfortable using automated driving systems on most roads. Only 38% of participants indicated that the next car they buy or lease will have automated driving systems. However, 64.8% reported that they would buy or lease a vehicle with automated systems if they could afford it.

While most participants conveyed a willingness to use automated driving systems, the vast majority (80.3%) reported that they did not feel comfortable using the automation without monitoring the vehicle closely.

Table C3. *Mean intentions toward automated driving systems (N = 71).*

	Mean	*SD*
I would not feel comfortable using automated driving systems on most roads.	2.37***	1.23
If I was tired or distracted, I would rely heavily on automated driving systems.	2.93	1.40
I would utilize the automated driving systems in a vehicle as much as possible.	3.72***	1.15
I would not feel comfortable using the automated driving systems in a vehicle without monitoring it closely.	4.04***	1.02
I am going to make sure that the next car I buy or lease has automated driving systems.	3.17	1.21
If I can afford it, I am going to buy or lease a car with automated driving systems.	3.79***	1.22

Note: Scales anchored by 1 = *disagree completely* and 5 = *agree completely*. SD = standard deviation. Significantly different than midpoint. *** $p < .001$.

Appendix D

Pilot Evaluation

Methods

The procedure was identical to the main study with the following exceptions.

Participants. Thirty-eight participants (9 female, M_{age}= 28.71 years of age, SD_{age} = 5.39) were recruited via University of Utah IRB approved flyers and by referral from previous participants.

Driving route. The driving route (see Figure D-1) was a 44-mile drive between Tooele and Aragonite, Utah on I-80. In a counterbalanced order, participants were instructed to drive with Level-0 automation or with Level-2 automation. Each drive was initiated once the participant was driving 75 mph on the Interstate and they felt safe and ready to begin testing. The first five minutes of each drive consisted of a baseline period containing only the driving task and was immediately followed by the addition of the DRT, which lasted until the end of the condition.

For the first run, participants drove on the westbound section of I-80. At the end of the run, participants completed the PANAS and two DRT questions, followed by a ten-minute break. For the second run, participants drove on the eastbound section of on I-80. At the end of the run, participants completed the PANAS with two DRT questions and then the research administrator drove the participant back to the University of Utah.

Experimental Design. The experimental design was a 3 (Vehicle: Cadillac CT6, Tesla Model S, Volvo XC90) x 2 (Automation: Level-0 vs. Level-2) factorial with 24 participants tested in each vehicle.

Figure D-1. A low trafficked Interstate with an 80 mph speed limit was used for the on-road driving study.

Results

Data from the DRT, ECG, and EEG were analyzed using linear mixed-effects models in R (R Core Team, 2019; version 3.6.0). Models were run using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015; version 1.1-21). The repeated subjects factor was input into each model as a random intercept with other factors of interest alternately entered as fixed effects. Interactions between conditions were analyzed through the comparison of models where factors were specified as interactive to models where factors were specified as additive. A model comparison approach using a Likelihood Ratio Test was then used to assess the goodness of each model fit. These evaluations returned p-values which compared the full linear effects model to a partial model (see Winter, 2013). This statistical approach has the advantage of analyzing all available data while adjusting fixed effects, random effects, and likelihood ratio test estimates for missing data.

The means and standard deviations from the 2 X 3 factorial design are reported in Table D-1 for each of the dependent measures.

Table D-1. *Pilot Study Results*

Automation	Cadillac CT6	Tesla Model S	Volvo XC90
Reaction Time (in ms)			
Level-0	523 (189)	445 (150)	456 (153)
Level-2	502 (178)	441 (169)	478 (204)
Hit Rate			
Level-0	0.96 (0.06)	0.96 (0.05)	0.98 (0.02)
Level-2	0.96 (0.04)	0.96 (0.04)	0.98 (0.03)
ECG: Beats Per Minute			
Level-0	77.9 (13.7)	73.9 (13.6)	74.0 (12.0)
Level-2	73.7 (13.4)	73.8 (13.3)	75.2 (11.1)
ECG: RMSSD			
Level-0	38.3 (26.5)	43.2 (31.8)	39.6 (20.0)
Level-2	36.3 (16.3)	46.1 (35.1)	38.1 (17.3)
EEG: Parietal Alpha			
Level-0	3.14 (3.17)	2.91 (2.55)	3.11 (3.36)
Level-2	2.84 (2.88)	2.84 (2.61)	2.97 (3.66)

As shown in Table D-2, there was only one significant effect the analyses: The main effect of Vehicle on DRT Hit Rate. None of the other main effects or interactions were significant.

Table D-2. *Pilot Study Results*

Reaction Time	*Chi-square*	*df*	*p*	R^2
Vehicle	0.139	2	0.932	0.000
Automation	0.006	1	0.813	0.000
Vehicle x Automation	2.685	2	0.261	0.003
Hit Rate				
Vehicle	**12.71**	**2**	**0.001*****	**0.072**
Automation	0.750	1	0.386	0.004
Vehicle x Automation	0.010	2	0.995	0.074
ECG: Beats Per Minute				
Vehicle	2.525	2	0.282	0.004
Automation	0.541	1	0.462	0.001
Vehicle x Automation	0.425	2	0.809	0.005
ECG: RMSSD				
Vehicle	0.007	2	0.997	0.000
Automation	0.105	1	0.745	0.000
Vehicle x Automation	1.442	2	0.486	0.004
EEG: Parietal Alpha				
Vehicle	0.327	2	0.849	0.001
Automation	0.007	1	0.929	0.000
Vehicle x Automation	0.069	2	0.966	0.001

Note. R Squared values represent the Marginal R^2 difference score from the comparison model. Significance Code: `***` <.001.