Transfer from Highly Automated to Manual Control: Performance & Trust

SAFETY RESEARCH USING SIMULATION

UNIVERSITY TRANSPORTATION CENTER

Chris Schwarz, Ph.D.
Associate Research Engineer
National Advanced Driving Simulator
The University of Iowa

Timothy L. Brown, Ph.D.
Associate Research Scientist
National Advanced Driving Simulator
The University of Iowa

Transfer from Highly Automated to Manual Control: Performance & Trust

Chris Schwarz, Ph.D.
Associate Research Engineer
National Advanced Driving Simulator
The University of Iowa

Clara Keum
Graduate Research Assistant
Department of Epidemiology
The University of Iowa

Timothy L. Brown, Ph.D.
Associate Research Scientist
National Advanced Driving Simulator
The University of Iowa

John Gaspar, Ph.D.
Assistant Research Scientist
National Advanced Driving Simulator
The University of Iowa

A Report on Research Sponsored by Safer-Sim University Transportation Center
with matching funds provided by Honda Research Institute, Toyota Motor Corporation, and the
National Advanced Driving Simulator

August 2016

Acknowledgments

The authors wish to thank the SaferSim staff for their assistance in helping the project reach a successful conclusion. Thanks to Omar Ahmad for agreeing to contribute NADS simulator hours as a match for the project. Finally, thanks to Tyron Louw at the University of Leeds for helpful discussions on studying automated vehicles with driving simulators.

Table of Contents

List of Figures

List of Tables

Abstract

The development of automated vehicles is ongoing at a breakneck pace. The human factors challenges of designing safe automation systems are critical as the first several generations of automated vehicles are expected to be semi-autonomous, requiring frequent transfers of control between the driver and vehicle. A driving simulator study was performed with 20 participants to study transfers of control in highly automated vehicles. We observed driver performance and measured comfort as an indicator of the development of trust in the system. One study drive used an automation system that was able to respond to most events by slowing or changing lanes on its own. The other study drive issued takeover requests (TORs) in all cases. Thus there was a change in reliability over the course of the study drives; some participants experienced the more-capable system first followed by the other, and others had the opposite experience. We observed three types of people with respect to their comfort profiles over the course of their three drives. Some started out very comfortable, while others took a long time to become comfortable. Takeovers were split into physical takeover, visual attention, and vehicle stabilization. Response time and performance measures showed that there was a 15- to 25-second period between the physical takeover and a return to normal driving performance. This confirms some observations in previous studies on transfer of control.

1 Introduction

1.1 Background

Automated vehicles are under active development by many auto manufacturers as well as other companies, such as Google. The projected benefits of automated vehicles are many and varied, but so are the concerns over their technical limitations, legal barriers, and human factors challenges. The National Highway Traffic Safety Administration (NHTSA) has become interested in automated vehicles and has published a position paper recommending that, for now, states only allow testing of high-level automation [1]. However, they are expected to issue further guidance as soon as 2017. NHTSA and the Society of Automotive Engineers (SAE) have published levels-of-automation taxonomies that provide a common language for automation systems [2].

This study was primarily concerned with automation level 3, termed *conditional automation* by SAE, in which the vehicle takes both longitudinal and lateral control. Whereas level 2 automation requires the operator to supervise the automation and scan the roadway for hazards, level 3 allows the operator to engage in other tasks, provided they can become available to take over again should the system request it. Level 4 SAE systems would provide a fallback strategy so that the vehicle might slow down and pull over if the operator does not take over in a timely manner. Both level 2 and level 3 raise concerns about how much the driver is out of the loop and how quickly they can regain situational awareness (SA) to effectively drive again.

Transfer of control is a complex topic given the number of possible scenarios. Consider a breakdown of control transfer by direction of transfer and expectation of transfer. Then a matrix of conditions can be constructed like the one in Table 1.1. Expected transfers to higher automation have commonly been initiated by a simple button press. Expected transfers to lower automation are often initiated by the driver turning the steering wheel or pressing a pedal, just like disengaging from cruise control. Unexpected transfers to higher levels of automation are exemplified by the intervention of collision avoidance systems. Finally, unexpected transfers to lower automation may occur because the automation failed in some way, or a situation was encountered that was beyond the automation's performance limitations. While the optimal user interface for expected transfers may not be known, the greatest unknown for manufacturers and regulators is this lower right quadrant. This project was focused on a combination of expected and unexpected transfers from conditional automation to manual control. We did not consider automation failures in the sense that the vehicle fails to request a takeover. Rather, our study implemented TORs in several types of events as a study condition.

Table 1.1 – Transfer of control examples under various conditions.

	Expected	Unexpected
Lower to Higher	Button press	Collision avoidance
Higher to Lower	Grab wheel	Automation failure

Bainbridge pointed out that humans are challenged when performing under time pressure and that when automation takes over the easy tasks from an operator, difficult tasks may become even more difficult [3]. Stanton and Marsden highlighted several potential problems that could plague automated vehicles, specifically while reclaiming control from automation. These include over-reliance, misuse, confusion, reliability problems, skills maintenance, error-inducing designs, and shortfalls in expected benefits [4, 5]. The lack of situational awareness that occurs when a driver has dropped out of the control loop has been studied for some time in several different contexts [6, 7, 8].

More recently, it has been shown that drivers had significantly longer reaction times in responding to a critical event when they were in automation and required to intercede, compared to when they were driving manually [9]. A slightly more nuanced result showed that performance in responding to a critical event was similar in the absence of a secondary task, but worse when automated and distracted [10]. More recent data suggest that drivers may take around 15 seconds to regain control from a high level of automation and up to 40 seconds to completely stabilize the vehicle control [11].

Takeover requests are issued by the automation to let the operator know that they should take back manual control of the dynamic driving task (DDT). The appropriate timing of such TORs has been a topic of research. Takeover request timings of five and seven seconds ahead of encountering an obstacle in the road were tested in a driving simulator [12]. While it was possible for drivers to take over in only a couple of seconds in both conditions, there were more braking responses and less time to check their blind spots in the five-second timing condition. Some of the extra time in the seven-second condition was used for decision-making and was valuable for avoiding sudden braking responses.

A NHTSA-funded test track study used both imminent and staged TORs, where the imminent TOR was issued once with an external threat and once without [13]. The staged alert had four phases as follows: 1) a tone followed by an informational message, 2) a verbal alert with a cautionary message, 3) a repeated tone in addition to an orange visual alert, and 4) a repeated imminent tone with a red alert. The visual components were text messages with associated colors to indicate urgency. The four messages were the following:

1) Prepare for manual control
2) Please turn off autodrive
3) Turn off autodrive now (orange)
4) Turn off autodrive now (red)

The average response time to an imminent alert was 2.3 seconds without an external threat and 2.1 seconds with it. The average response time to the staged alert was 17 seconds, which may have been partly due to a countdown that accompanied the informational warning.

A driver's trust in automation greatly influences whether that automation is used appropriately, misused, or disused. Trust should be calibrated appropriately so that a driver does not over- or under-trust an automated system [14]. Lee and See [14] proposed a closed-loop conceptual model of a dynamic process that governs trust, recognizing that trust might be considered as a function over time that can rise and fall.

Drivers start out with beliefs that can greatly influence their trust; however, it is their attitudes and how they translate into behaviors that should be considered [14].

Unfortunately, when observing behaviors, it is possible to confuse the effect of trust with effects from other causal factors like SA or self-confidence. In one study, drivers did show a willingness to disengage from the supervision task and engage more with entertainment devices, showing a degree of system trust [15]. However, they also paid a greater amount of attention to the road in heavy traffic, showing that their trust was not absolute.

Trust and comfort are correlated constructs that are both important for human-robot interaction [16, 17]. Indeed, it is hard to imagine the development of trust without some degree of comfort being present. Sanders et al. identified four factors of trust: performance, reliance, individual differences, and collaboration. Another breakdown of trust included the following factors: predictability, dependability, faith, and overall trust [18, 19].

A series of driving simulator studies on adaptive cruise control done with and without motion showed similar results, and the authors concluded that motion may therefore not be necessary [20]. However, most recent driving simulation studies in vehicle automation have used higher-fidelity systems with motion bases. It seems reasonable that the 'feel' of the car from a simulator's motion cues is critical to a driver who may be completely visually disengaged from the driving task, as is the case in high automation levels. This study used the NADS-1 high-fidelity motion base simulator.

1.2 Objectives

The study was designed to address the following research questions:

1. To what degree do drivers trust the automation?
2. Does less-capable automation decrease trust, and how does reliability influence trust in automation?
3. When do drivers choose to begin an expected transfer of control, and how long does it take?
4. After manual takeovers, how long does it take for the driver to return control to the automation?
5. How long does an unexpected transfer of control take, including vehicle stabilization?
6. Does the act of transferring control have any performance decrements associated with it?
7. Are there differences between gender and age groups in trust or driving performance?

Our aim was to validate previous results on the length of time required for a transfer to manual control and vehicle stabilization. Additionally, results from unexpected transfers would shed light on drivers' capacities to quickly and accurately take over manual control. It was expected that there would be decrements to the quality of the transfer due to the need to regain SA while at the same time assuming vehicle control. Finally, it was also expected that automation failures would damage the driver's trust in the system and that the effects of that reduced trust might be observed in subsequent driving and takeover choices.

2 Methodology

2.1 Simulator and Apparatus

The National Advanced Driving Simulator (NADS) is located at the University of Iowa's Research Park. The main simulator, the NADS-1, consists of a 24-foot dome in which an entire car cab is mounted. All participants drove the same vehicle—a 1996 Malibu sedan. The motion system, on which the dome sits, provides 400 square meters of horizontal and longitudinal travel and ±330 degrees of rotation. The driver feels acceleration, braking, and steering cues much as if he or she were actually driving a real vehicle. High-frequency road vibrations up to 40 Hz are reproduced from vibration actuators placed in each wheel well of the cab. A picture of the NADS-1 simulator and an image from the interior of the dome are shown in Figure 2.1.

The NADS-1 displays graphics by using 16 high-definition (1920x1200) LED (light-emitting diode) projectors. These projectors provide a 360-degree horizontal, 40-degree field of view. The visual system also features a custom-built Image Generator (IG) system that is capable of generating graphics for 20 channels (16 for the dome and an additional 4 for task-specific displays). The IG performs warping and blending of the image to remove seams between projector images and displays scenery properly on the interior wall of the dome. The NADS produces a thorough record of vehicle state (e.g., lane position) and driver inputs (e.g., steering wheel position), sampled at 240 Hz.

The cab is equipped with a Face Lab™ 5.0 eye-tracking system that is mounted on the dash in front of the driver's seat above the steering wheel. In the best-case scenario, where the head is motionless and both eyes are visible, a fixated gaze may be measured with an error of about 2º. With the worst-case head pose, accuracy is estimated to be about 5º. The eye tracker samples at a rate of 60 Hz.

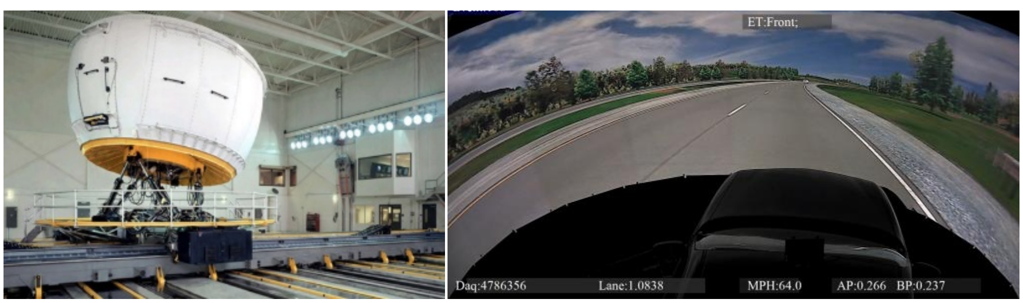


Figure 2.1 – NADS-1 driving simulator (left) with a driving scene inside the dome (right).

2.2 Driving Scenarios

Two driving scenarios, requiring approximately 30 minutes each to complete, were designed with the same set of events, and a seven-minute practice drive was created. The study drives involved typical vehicle control in a variety of situations. Once the driver achieved highway speed, he or she was instructed to engage the automation by pressing a button on the steering wheel. One scenario implemented automation that was capable of handling most events on its own. The other scenario included the same events but with automation-initiated TORs where the driver needed to intervene to maintain safe control of the vehicle. Thus, each participant experienced a change in system reliability during his or her visit. Some drove the more-capable automation and then the less-capable automation, and others did the opposite.

The practice drive scenario served to adapt participants to driving in the simulator, as well as exposing them to automation control transfers and TORs. All five events existed in both study drives, but in different orders and with different automation capabilities. Moreover, the locations of the events as well as the starting and ending locations of the drives were also varied to minimize predictability. Towards the end of each drive, a normal, manual takeover request took place before a scheduled exit off the highway. The five main events are summarized in Table 2.1.

Table 2.1 – Scenario events in drives A and B with varying takeover request (TOR) timing.

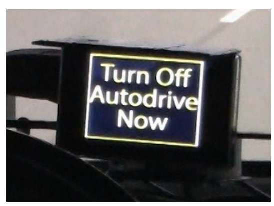
Event	More Capable	Less Capable	Notes
#1 Work zone	No TOR	TOR with 10-second warning	Warning occurs about 15 seconds ahead of the work zone. Traffic in left lane.
#2 Missing lane lines	No TOR	TOR with 10-second warning	Warning occurs when lane lines are lost.
#3 Sharp curve	No TOR	TOR with 10-second warning	Elevated ramp with walls.
#4 Slow lead vehicle	TOR with 10-second warning	TOR with 5-second warning	Lead vehicle driving at 25 mph with hazards on. Traffic in left lane.
#5 Exit highway	TOR with 30-second warning	TOR with 30-second warning	Always the last event of the drive. No difference between A and B.

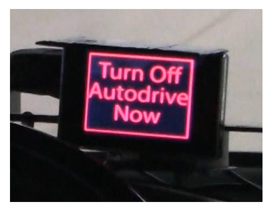
Finally, there were extra events interspersed between the study events. In these extra events, a lead vehicle would slow from the speed limit to 55 mph for a period of time, forcing the automation to slow the participant's vehicle as well. After a short time, the lead vehicle sped back up to the speed limit. These events served as brief disturbances that would draw the operator's attention and provide experiences in which the automation behaved as desired with no loss of capability. It was expected that these extra events would help to build trust.

2.3 Automation Icon and Takeover Requests

Automated driving was indicated by a visual icon on a high heads-up display. Takeover requests were composed of both visual and audio cues. Visual cues appeared on the same display. When the driver needed to transfer control, a chime sound played with the appearance of a visual sign saying to either turn on or off the automation. Depending on each event and scenario, a TOR took place either 5, 10, or 30 seconds prior to the event. If the driver did not transfer control from automated to manual in some set interval after the TOR fired, the automation system slowed the vehicle down and pulled over to the side of the highway. This fallback strategy is characteristic of SAE Level 4 automation, though participants were not trained on it ahead of time. All four possible display icons are shown in Figure 2.2.

(a)





(b) (c) (d)

Figure 2.2 – Automation interface in high heads-up display location: (a) automated-mode icon in blue, (b) informational warning in white, (c) cautionary alert in yellow, (d) imminent alert in red.

2.4 <u>Secondary Tasks</u>

Participants were asked to work on trivia questions from Trivia Plaza (www.triviaplaza.com) as a secondary task while the vehicle was under automated control during both drives to mimic distraction associated with engagement with non-driving tasks that may occur when the driver is in a supervisory control mode. Trivia Plaza is a website that offers numerous sets of questions in nine major categories (see Figure 2.3). Within each category, there are many subcategories (e.g., subcategories of "Movie" include various time periods, genres, production companies, etc.). An iPad was given to each participant for the duration of the drives to allow access to the website. In order to encourage participants to be actively involved in trivia, they were told to pick any topic(s) that they were interested in and that any participant who reached a cumulative score of 100 or higher would receive a bonus compensation of $15.

Participants could play multiple times to reach the given score. In reality, the bonus compensation was a deception, and all subjects received the $15 extra.

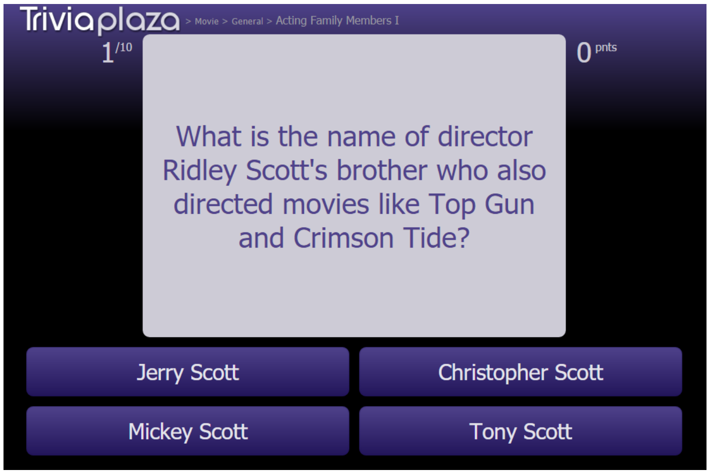


Figure 2.3 – Example screen from Trivia Plaza (www.triviaplaza.com).

2.5 Experimental Design

A 2 (drive) x 2 (age) x 2 (gender) mixed design was used for this study. The within-subject independent variable was the automation capability (more capable, less capable). The between-subject independent variables were gender (male, female) and age (18 - 25, 25 - 55). The age variable was blocked by using the minimization method to balance out the number of participants in each group.

2.6 Participants

A total of 20 participants that were on the NADS IRB-approved registry were recruited and enrolled in this study. They were contacted by email or phone and were provided a general overview of the study and screened to verify eligibility. The inclusion/exclusion criteria included questions about driving qualifications, health history, current health status, and medications. Answers from the NADS screening procedures were not recorded. Participants were licensed adults between the ages of 18 and 55 with at least three years of driving experience with a minimum of 2000 miles and were in good general health. Participants were paid $65 for completing both drives and were asked to avoid consuming alcohol or other drugs not prescribed by a physician in the 24 hours preceding their visit.

2.7 Procedure

When the participants arrived, members of the research team explained and reviewed the study with them. The participants were asked to provide informed consent to their participation. Participants were then provided with a general survey on their trust of technology and a general demographics survey followed by a presentation on the automated vehicle they would be driving and general simulator procedures. Participants were then escorted to the simulator where they were provided a brief overview of the cab layout and then allowed to adjust the seat, steering wheel, and mirrors. The eye tracker was then calibrated for the participant. Each participant was given approximately seven minutes of practice driving in both manual and automated modes to have an opportunity to become familiar with the control of the vehicle.

The participants were randomly assigned to two groups. One group drove the scenario with more-capable automation followed by the other scenario, while the other group drove the scenarios in the reverse order. Participants were asked to work on the same unlimited levels of trivia questions from Trivia Plaza while under automated control, but they were responsible for the overall safety of the drive. Participants could stop playing trivia whenever the control transferred to manual, as well as whenever they did not feel safe or comfortable. Participants were also asked to answer a series of seven-point Likert scale questions during each drive. The participants then completed the two half-hour drives with a short break in between. After completing both drives, a post-drive survey on their overall driving experience and level of trust were completed by the participants along with a wellness survey to assess signs of simulator sickness in a private room.

3 Experimental Results

3.1 Reduced Data

Data was collected from three main sources. Simulator data files contained many variables, including driver inputs, vehicle signals, and other cab signals such as the Likert survey input. Eye tracker data was recorded to log files from the FaceLab system. Lastly, post-drive surveys were administered to collect additional data on comfort and attitudes towards automated vehicles. The simulator and eye tracker data were processed using a data reduction script in Matlab to obtain several dependent measures used in the analysis.

Two types of measures were calculated. The first set was calculated once per event and is listed in Table 3.1. These measures included the values of independent variables, the event number and name, and key dependent measurements of the event. The percent road center (PRC) gaze [21] measured the percentage of time that the driver's gaze was directed at the front scene, computed in a running 17-second window [22].

A second type of dependent measure was recorded at regular intervals either after the beginning of manual driving mode, or after the end of manual driving mode in the event. A fixed interval spacing of five seconds was used, and up to 12 segments, or one minute, were computed. These measures created a type of longitudinal data that could be analyzed for trends. The approach was adapted from the methodology used at the University of Leeds [11]. The longitudinal dependent measures are summarized in Table 3.2.

Table 3.1 – Measures that were calculated once per event.

Measure	Description
Subject	Participant number
Gender	M or F
Age	Age of the participant in years
Scenario	More- or less-capable automation
Order	Was the drive in the first or second order?
FirstScenario	Which drive was first (A or B)?
Event	Event number (0-8). Events 6-8 were the extra events.
EventName	Name of the event
StartFrame	DAQ frame at the start of the event
StartTime	Time at the start of the event
Duration	The duration of the event
PctAuto	Percentage of event time spent in automated mode
TakeOverRT	Response time to take over from automation after warning
GiveBackRT	Response time to give back control to automation after cue
MeanPrc17Auto	Average PRC gaze while in automated mode
MedPrc17Auto	Median PRC gaze while in automated mode
MeanPrc17Manual	Average PRC gaze while in manual mode
MedPrc17Manual	Median PRC gaze while in manual mode
DurationManual	The time that was spent in manual mode
Manual	Did the driver take back control from the automation?

Table 3.2 – Dependent measures that were calculated in multiple segments during the event. Each measure was calculated in up to 12 consecutive segments.

Measure	Description
MinSpeed	The minimum speed in each manual segment (mph)
MeanSpeed	The average speed in each manual segment (mph)
SR	The steering reversal rate in each manual segment, calculated in a 15-second running window (rev/sec)
SDLP	Average value of standard deviation of lane position in each manual segment, calculated in a 15-second running window (ft)
HFSteer	High-frequency steering content in each manual segment
PRC	Percent road center gaze in each manual segment, calculated in a 17-second running window (%)
PRCpost	Percent road center gaze in each segment after return to automated mode, calculated in a 17-second running window (%)

The steering reversals and high-frequency steering (HFSteer) measures were also adapted from the Leeds methodology. Steering reversals count the number of one-degree reversals in a time period. The steering reversal rate per second (Leeds used per minute) was then calculated by dividing by the number of seconds in the segment. The HFSteer measure is based on a high-frequency control of steering computation that is defined as the ratio between the power of a high-frequency band of steering activity to the power of a lower-frequency band [23, 24].

3.2 Data Analysis

Several methods were used to analyze the data. The SAS statistical software package was used to analyze the survey data, while the R statistical software language [25] was used to analyze the simulator and eye tracker measures. ANOVAs were used to compare dependent variables across various study conditions, and box plots are the graph type we preferred for visualizing significant differences that were observed. Box Cox transformations were applied to the dependent measure, where appropriate, to optimize the normality of the residual error. Normality was tested by observing the Q-Q plot of the residuals as well as by running a Shapiro-Wilk test to see if the null hypothesis of normality should be rejected. Additionally, a cluster analysis was used to identify three distinct profiles of longitudinal comfort that were observed among the participants.

For the retrospective trust survey data, the restricted range and ordinal scale of the data associated with Likert-type survey responses requires that care be taken in that analysis. Although there is significant debate over the acceptability of various analysis approaches and where the data can be considered as interval scale and be used with ANOVA, the authors concur with Sullivan and Artino [26] that an ANOVA is an appropriate technique. Accordingly, the SAS general linear model (GLM) procedure was used to conduct an ANOVA on the data.

For the analyses of the simulator and remaining survey data, linear mixed-effects regression (LMER) was used to analyze the longitudinal measures with the lme4 package in R [27]. The specific application of LMER that includes time as a factor is called growth curve analysis [28]. Moreover, the time can be transformed to allow for quadratic, polynomial, or piecewise linear time segments. We used the following variations on time: raw, piecewise with two pieces, and orthogonal polynomial. Transformations of the time vector can make the intercept and slope coefficients of the model harder to interpret, but it is still possible to predict values with the model.

Linear mixed-effects regression models can be evaluated on both relative and absolute measures of accuracy. Two relative measures that are generated in lme4 are the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) [29]. These metrics can only be interpreted by taking the difference between two models as a comparison. Burnham and Anderson [29] noted that a difference of at least two in the AIC can be used as a rule of thumb in detecting substantial differences between models.

Absolute accuracy is often reported as a p-value, but lme4 does not generate p-values for LMER models because of the difficulties inherent in getting reliable estimates. However, there are other options available to the user of such models. We have adopted a simple approach to estimating R^2 values [30, 31]. This approach yields two values, a marginal R^2 value as well as a conditional R^2 value. The former describes the variance explained by the fixed effects in the model, while the latter describes the variance explained by both the fixed and random effects in the model. When reporting on accuracy of LMER models, we provide the AIC, BIC, marginal and conditional R^2, and the degrees of freedom. Since many people are not familiar with LMER models and growth curve analysis, figures are also provided to visualize model significance.

3.3 Descriptive Statistics

A total of 20 people participated in the study. Gender was equally balanced, while age was balanced using the minimization method, which resulted in two equally balanced groups. However, the allocation of age to the order condition ended up being unbalanced by one participant. Baseline demographics of the participants is shown in Table 3.3.

Table 3.3 – Baseline demographics of participants.

	Order 1 (A -> B)	Order 2 (B-> A)	Total (%)*
N	10	10	20 (100)
Age (yrs)			
Mean	29.7	28.4	29.1**
18 - 25	6	4	10 (50.0)
26 - 55	4	6	10 (50.0)
Gender			
Male	5	5	10 (50.0)
Female	5	5	10 (50.0)

* Column percentages ** Mean age of all participants (across all groups)

A large portion of the analysis deals with determining the effect of order, age, and gender on the dependent measures. Even the analysis of the longitudinal measures with LMER models are able to include order, age, and gender as conditional factors to judge their effects on the model fit. The scenario (more or less capable (A or B)) is the within-subjects analogue of the between-subjects variable order. Half the participants drove the more-capable automation first, and the other half drove the less-capable automation first. We would expect to see differences in the dependent measures between scenarios, but one of the interesting questions here is whether order also has an effect on those measures. The following two tables summarize the statistics of several dependent measures split out by age, gender, and event.

Table 3.4 describes the scenario with more capability, while Table 3.5 describes the scenario with less capability.

Table 3.4 – Sample means and standard deviations of Likert, Likert reaction time, duration and mean Prc17Auto for each event, split out by gender and age (more-capable automation).

	Male (18 - 25)		Female (18 - 25)		Male (26 - 55)		Female (26 - 55)	
	Mean	St.Dev	Mean	St.Dev	Mean	St.Dev	Mean	St.Dev
Work Zone								
Likert	2.6	1.4	2.0	0.6	1.4	0.8	2.8	1.7
Likert Reaction Time	4.7	2.6	5.1	1.4	5.4	1.5	4.1	0.7
Duration	50.1	1.2	49.6	2.7	49.9	2.0	49.2	2.8
Mean Prc 17Auto	0.4	0.2	0.2	0.3	0.4	0.3	0.4	0.2
Lane Lines								
Likert	2.6	1.4	2.2	1.0	1.2	0.4	2.6	0.8
Likert Reaction Time	4.6	0.5	4.1	0.8	4.3	1.4	4.0	1.4
Duration	43.1	0.0	43.0	0.0	43.0	0.0	43.0	0.0
Mean Prc 17Auto	0.2	0.1	0.2	0.2	0.3	0.3	0.2	0.2
Curve								
Likert	2.8	1.0	1.8	0.8	1.4	0.8	2.2	1.0
Likert Reaction Time	5.9	3.3	3.8	0.4	5.1	0.7	3.4	1.0
Duration	84.4	0.0	84.4	0.0	84.4	0.0	84.4	0.0
Mean Prc 17Auto	0.3	0.2	0.1	0.1	0.2	0.2	0.2	0.1
Slow Vehicle								
Likert	2.0	1.1	1.6	0.8	1.2	0.4	1.8	0.8
Likert Reaction Time	4.5	1.8	3.1	0.5	4.6	0.8	3.4	0.5
Duration	24.9	2.5	29.2	2.2	30.7	4.7	32.8	3.6
Mean Prc 17Auto	0.3	0.1	0.3	0.1	0.4	0.1	0.4	0.2
Exit								

	Male (18 - 25)		Female (18 - 25)		Male (26 - 55)		Female (26 - 55)	
	Mean	St.Dev	Mean	St.Dev	Mean	St.Dev	Mean	St.Dev
Likert	2.2	1.2	1.4	0.8	1.6	0.8	2.4	0.5
Likert Reaction Time	4.5	1.2	3.5	0.7	5.6	1.8	3.4	0.7
Duration	71.5	4.4	77.3	5.6	76.3	2.4	78.1	2.4
Mean Prc 17Auto	0.4	0.2	0.5	0.2	0.4	0.2	0.6	0.1
Filler 1								
Likert	1.2	0.4	1.8	0.8	1.4	0.8	1.2	0.4
Likert Reaction Time	4.8	1.8	3.6	0.6	4.0	1.8	3.4	0.6
Duration	37.8	0.6	38.4	1.1	38.8	0.7	37.8	1.2
Mean Prc 17Auto	0.2	0.2	0.2	0.3	0.3	0.3	0.3	0.2
Fillller 2								
Likert	1.8	1.2	1.6	0.8	1.2	0.4	2.0	0.6
Likert Reaction Time	4.7	2.1	3.6	0.7	4.0	1.3	3.9	1.0
Duration	41.6	4.8	39.1	6.1	44.1	0.1	38.8	5.6
Mean Prc 17Auto	0.1	0.1	0.1	0.2	0.1	0.1	0.2	0.2
Fillller 3								
Likert	1.6	0.8	1.2	0.4	1.6	0.8	2.0	0.6
Likert Reaction Time	6.1	4.0	3.2	0.7	5.0	0.5	3.5	0.6
Duration	38.0	0.3	38.0	0.4	36.3	1.7	37.6	1.4
Mean Prc 17Auto	0.1	0.1	0.2	0.2	0.3	0.2	0.2	0.2

Table 3.5 Sample means and standard deviations for Likert, Likert reaction time, duration, and meanPrc17Auto for each event, split out by gender and age (less-capable automation).

	Male (18 - 25)		Female (18 - 25)		Male (26 - 55)		Female (26 - 55)	
	Mean	St.Dev	Mean	St.Dev	Mean	St.Dev	Mean	St.Dev
Work Zone								
Likert	2.4	1.5	2.0	1.3	1.6	0.8	2.6	1.7
Likert Reaction Time	6.0	2.2	4.8	1.6	5.3	1.1	4.5	0.4
Duration	34.6	3.6	39.4	5.0	36.6	3.3	38.2	3.0
Mean Prc 17Auto	0.4	0.1	0.3	0.1	0.4	0.1	0.5	0.1
Lane Lines								
Likert	1.4	0.5	2.4	1.0	1.4	0.8	2.0	1.1
Likert Reaction Time	5.1	2.0	3.5	0.3	5.1	1.4	3.6	0.5
Duration	47.8	9.0	45.9	1.6	46.9	3.5	45.6	0.8
Mean Prc 17Auto	0.4	0.1	0.4	0.2	0.4	0.1	0.3	0.1
Curve								
Likert	2.2	1.0	1.8	1.0	1.6	0.8	2.8	1.2
Likert Reaction Time	5.8	1.6	3.2	0.8	4.2	0.9	3.8	0.9
Duration	89.3	4.8	106.4	11.2	99.1	4.0	93.3	6.6
Mean Prc 17Auto	0.4	0.2	0.4	0.1	0.4	0.1	0.4	0.2
Slow Vehicle								
Likert	1.4	0.5	1.6	0.8	1.2	0.4	1.8	0.8
Likert Reaction Time	6.0	2.0	3.2	0.9	4.3	0.9	3.3	0.4
Duration	25.2	6.3	31.6	3.5	24.4	1.9	31.2	4.0
Mean Prc 17Auto	0.3	0.1	0.3	0.1	0.3	0.1	0.3	0.1
Exit								

	Male (18 - 25)		Female (18 - 25)		Male (26 - 55)		Female (26 - 55)	
	Mean	St.Dev	Mean	St.Dev	Mean	St.Dev	Mean	St.Dev
Likert	1.8	0.8	1.6	0.8	1.4	0.5	1.8	0.4
Likert Reaction Time	4.3	1.8	4.0	0.7	4.6	1.4	3.6	0.5
Duration	75.4	6.2	86.2	21.8	79.1	2.4	78.4	7.2
Mean Prc 17Auto	0.4	0.2	0.6	0.1	0.5	0.1	0.7	0.1
Filler 1								
Likert	1.6	0.8	2.2	1.0	1.2	0.4	1.8	0.4
Likert Reaction Time	6.2	4.2	3.2	0.6	6.3	1.8	3.2	0.6
Duration	46.1	0.6	45.6	0.9	45.1	0.8	45.8	0.9
Mean Prc 17Auto	0.2	0.1	0.1	0.1	0.2	0.1	0.1	0.1
Fillller 2								
Likert	1.4	0.5	1.8	0.8	1.2	0.4	2.2	1.5
Likert Reaction Time	3.3	0.2	3.9	1.3	4.4	0.8	3.2	0.2
Duration	48.0	9.1	46.6	0.1	46.5	0.7	4.9	0.9
Mean Prc 17Auto	0.2	0.2	0.1	0.1	0.3	0.2	0.2	0.2
Fillller 3								
Likert	1.6	0.5	1.8	0.8	1.2	0.4	2.0	0.9
Likert Reaction Time	4.3	1.5	3.5	0.4	4.3	0.9	3.9	0.6
Duration	62.3	0.1	62.3	0.0	62.3	0.0	62.3	0.0
Mean Prc 17Auto	0.2	0.1	0.1	0.1	0.2	0.1	0.2	0.2

3.4 Results on Operator Trust

How much did operators trust the automation?

The amount of comfort an operator had in the automation during their drives was probed at semi-regular intervals using a Likert survey that appeared on a display located in front of the cab's center console. The single question asked the operator to rate his or her level of comfort at that moment on a scale of 1 (Very Comfortable) to 7 (Very Uncomfortable). The wording of comfort was selected as an overall approximation of the more complex concept of trust and was thought to estimate the participants' nascent level of trust in a system that was new to them.

Two such surveys were administered in the practice drive. There were eight additional surveys in each main drive, for a total of 18 comfort measurements. They were spaced in between events, and nothing related to any event was happening at the time the surveys were administered. Sometimes the survey occurred after one of the four main events, but sometimes it occurred after an extra, 'filler' event during which a lead vehicle slowed momentarily.

The log of the Likert score was used as the main trust measure. All 18 measurements in a drive constituted a longitudinal comfort profile that evolved in ways unique to each individual. Each participant's longitudinal comfort profile is plotted individually in Figure 3.1. The scenario is coded both by color and by marker shape. Linear mixed-effects regression models were fit to the comfort (log of Likert responses). The survey instance (0-17) was used as time and included as a factor; thus, the resulting models were growth curve models.

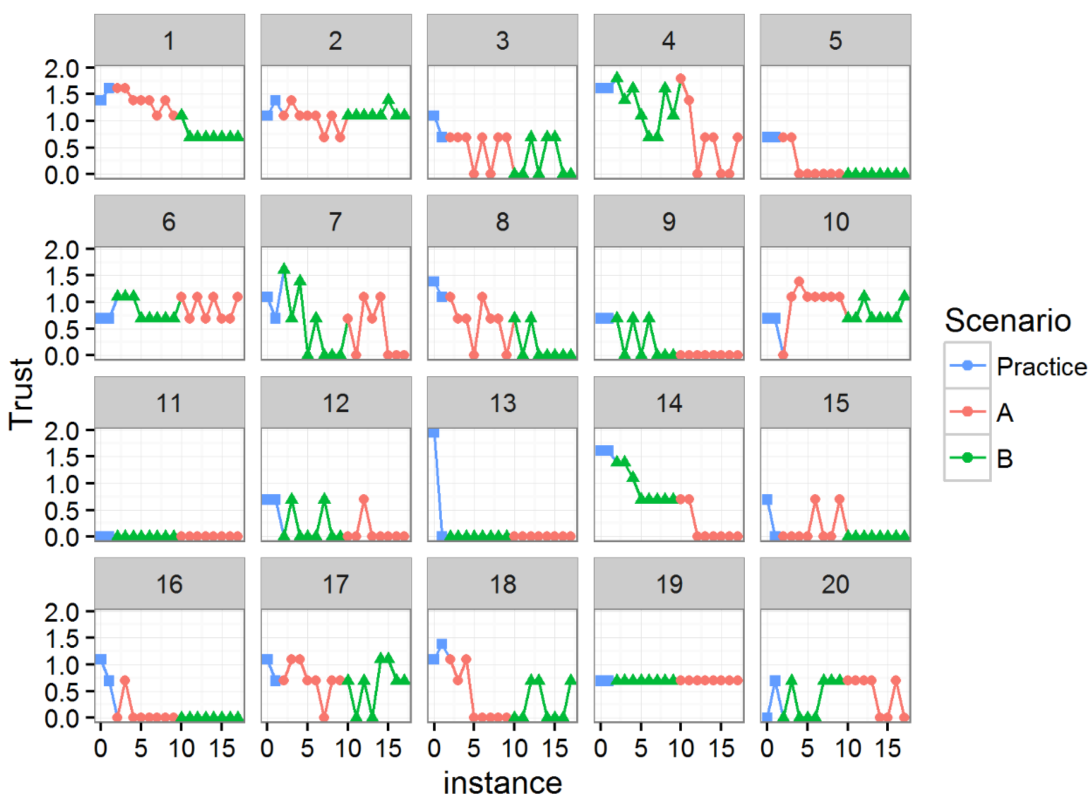


Figure 3.1 – Longitudinal comfort (log of 18 Likert responses) for 20 subjects across a practice drive and two study drives (A: more-capable automation, B: less-capable automation).

The surveys were administered at semi-regular intervals in between other events. Thus, the actual time between surveys varied some, and the times of instances would be different in the two orders (A first and B first). A cumulative time variable was calculated for each participant so that the absolute time of each Likert survey could be used instead. Some experimentation was done with fitting growth curve models using absolute time, but it was found that they did not yield additional information over the models that used instance number. Since using instances simplified comparisons between drives significantly, they were used as the time factor for all comfort models.

A piecewise linear formulation for time was selected, based on trends that were observed in Figure 3.1. Three competing models were made in which the break in the piecewise lines was placed after instances 4, 5, and 6. The three models were compared, and the results are shown in Table 3.6. The middle model, which used five points for the first line and 13 for the second, was selected as the best fit. Figure 3.2 shows the longitudinal comfort profiles for all participants in light gray lines. Overlaid on these profiles are the group means for each instance in red dots. Finally, the 95% confidence interval region of the growth model fixed effects is shown as a grey band overlaid on the plot.

Additional models were made by including age, gender, and order (A first or B first) as factors in the fixed and/or random effects of the model. However, no significant improvements in model fit were found, indicating that those factors may not have had a significant effect on comfort.

Table 3.6 – Three piecewise linear model fits where the pieces are split into 5 and 13 points, 4 and 14 points, and 6 and 12 points, respectively.

Model	DF	AIC	BIC	Marginal R2	Conditional R2
Piecewise 5/13	10	248.38	287.25	0.13	0.69
Piecewise 4/14	10	252.13	290.99	0.12	0.68
Piecewise 6/12	10	252.79	291.65	0.13	0.69

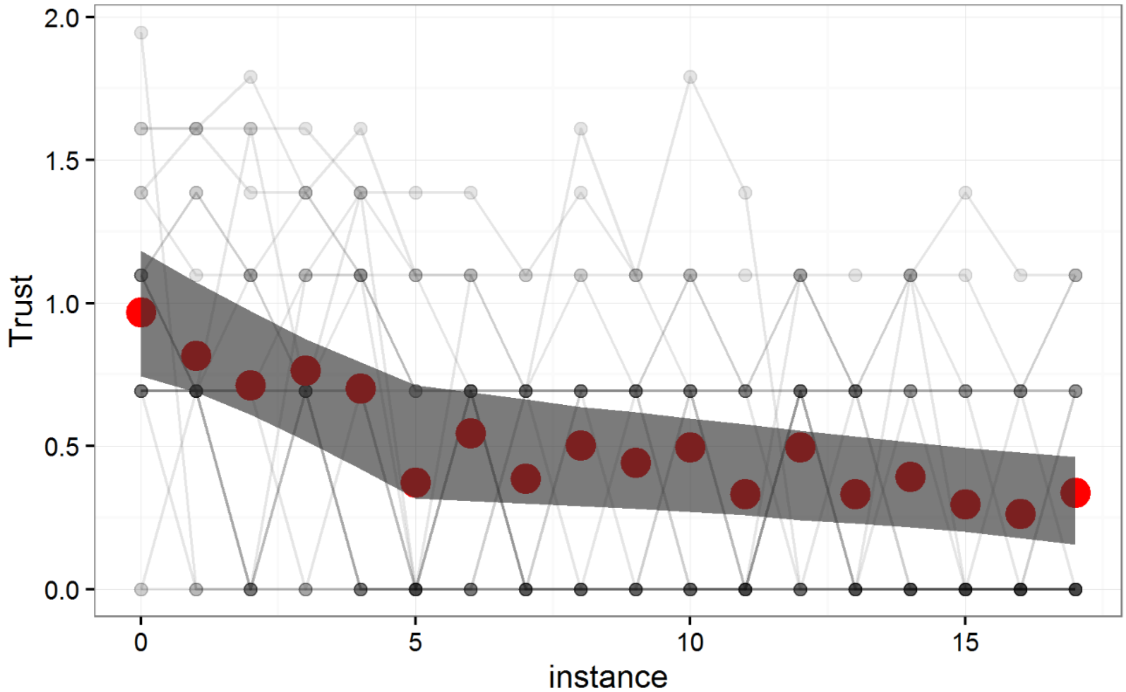


Figure 3.2 – Longitudinal comfort (log of 18 Likert responses) for 20 participants. Individual profiles are shown as grey lines, while the group means for each instance are red dots. The grey ribbon shows the 95% confidence interval of the fixed effects model.

A hierarchical clustering analysis was conducted using the random intercept and two random slopes from the growth curve model. Three clusters were selected from the analysis (see Figure 3.3), and participants were assigned to one of the three. Figure 3.4 shows the longitudinal comfort profiles once again, this time with 95% confidence intervals from the random effects overlaid on each plot. Additionally, the cluster for each participant is color-coded in the figure.

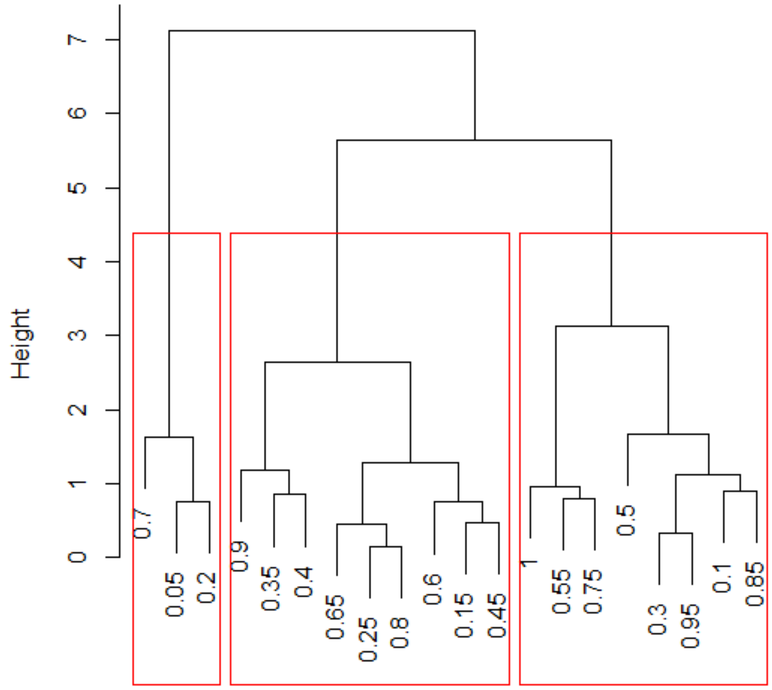


Figure 3.3 – Cluster dendogram showing Euclidian distance between clusters. A three-cluster fit was selected. Dendogram leaves are labeled as participant number divided by 20.

The three clusters may be easily described on inspection of Figure 3.4. The participants in cluster 1 gradually increased in comfort (the log of the Likert response is inversely proportional to comfort) over the course of the practice drive and two main drives. Participants in cluster 2 started with about the level of comfort that they maintained throughout their three drives. Finally, participants in cluster 3 started with less comfort, but their scores improved over a fixed amount of time and then leveled off for the remainder of the drives. Participant 13 may be an outlier if the first large Likert response was just an aberration. Participant 4 was unusual in that the responses indicated a loss of comfort near the end of the first drive (identified as Drive B, or the less-capable automation system, from Figure 3.1).

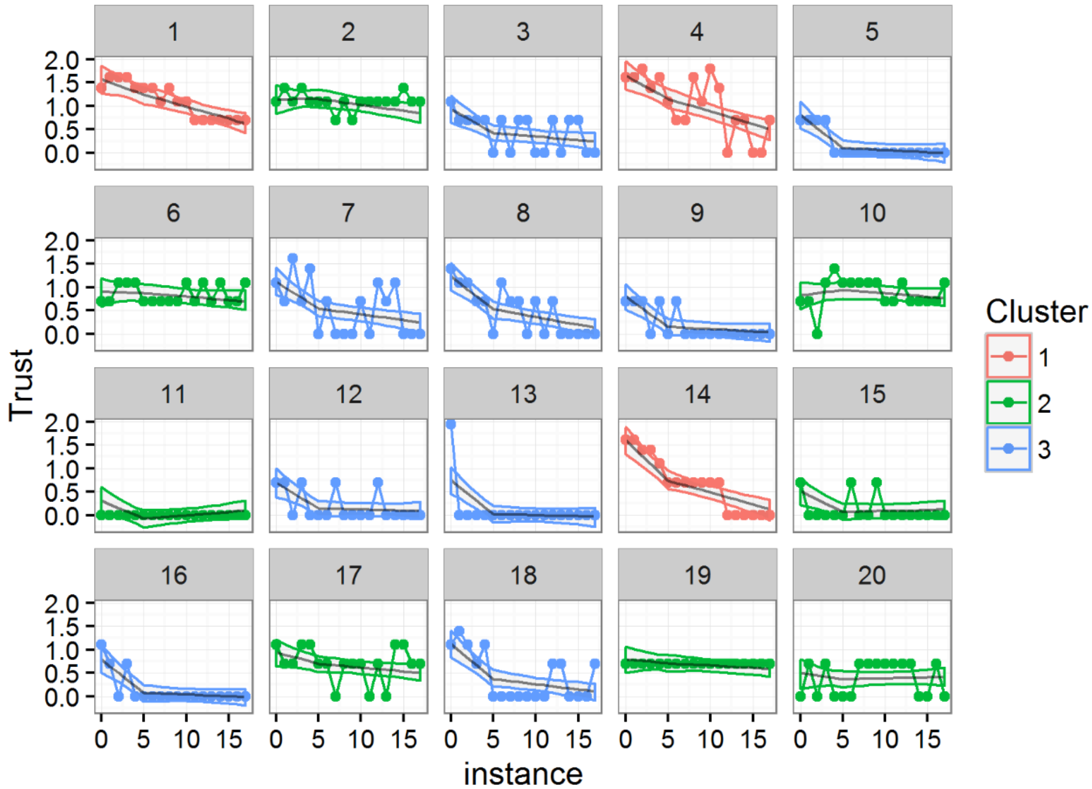


Figure 3.4 – Longitudinal comfort (log of 18 Likert responses) for 20 subjects across a practice drive and two study drives. Ribbon overlays show the 95% confidence interval of the random effects model fit. Subjects are clustered and color-coded into three identified profiles of longitudinal trust.

A cluster group was added to the comfort data set, and three chi-squared tests were run using the frequencies of each order, gender, and age group in each of the three clusters, respectively. None of the chi-squared tests were significant, so no main effects of order, gender, or age on cluster membership could be deduced.

Response Time of Likert Responses

A three-way ANOVA was conducted to analyze the variance of response time to the Likert surveys, depending on age, gender, and order. A Box Cox transformation with lambda equal to -0.6 was applied to the response time variable. Significant main effects of order and gender were observed, as was a two-way interaction between order and age, and a three-way interaction between order, age, and gender. The results of the ANOVA are shown in Table 3.7, and box plots of the response times are displayed in Figure 3.5. The figure shows that, on average, men took longer to respond to the probe surveys, and young men took the longest.

Table 3.7 – Three-way ANOVA on the effects of order, age, and gender on response time to the in-cab comfort survey.

	Df	Sum Sq	Mean Sq	F value	p-value
Order	1	0.1046	0.1046	21.128	6.25e-06
Age	1	0.0000	0.0000	0.005	0.9432
Gender	1	0.3094	0.3094	62.499	4.64e-14
Order:Age	1	0.0351	0.0351	7.085	0.0082
Order:Gender	1	0.0038	0.0038	0.7740	0.3796
Age:Gender	1	0.0009	0.0009	0.1760	0.6754
Order:Age:Gender	1	0.0638	0.0638	12.883	0.0004
Residuals	312	1.5446	0.0050		

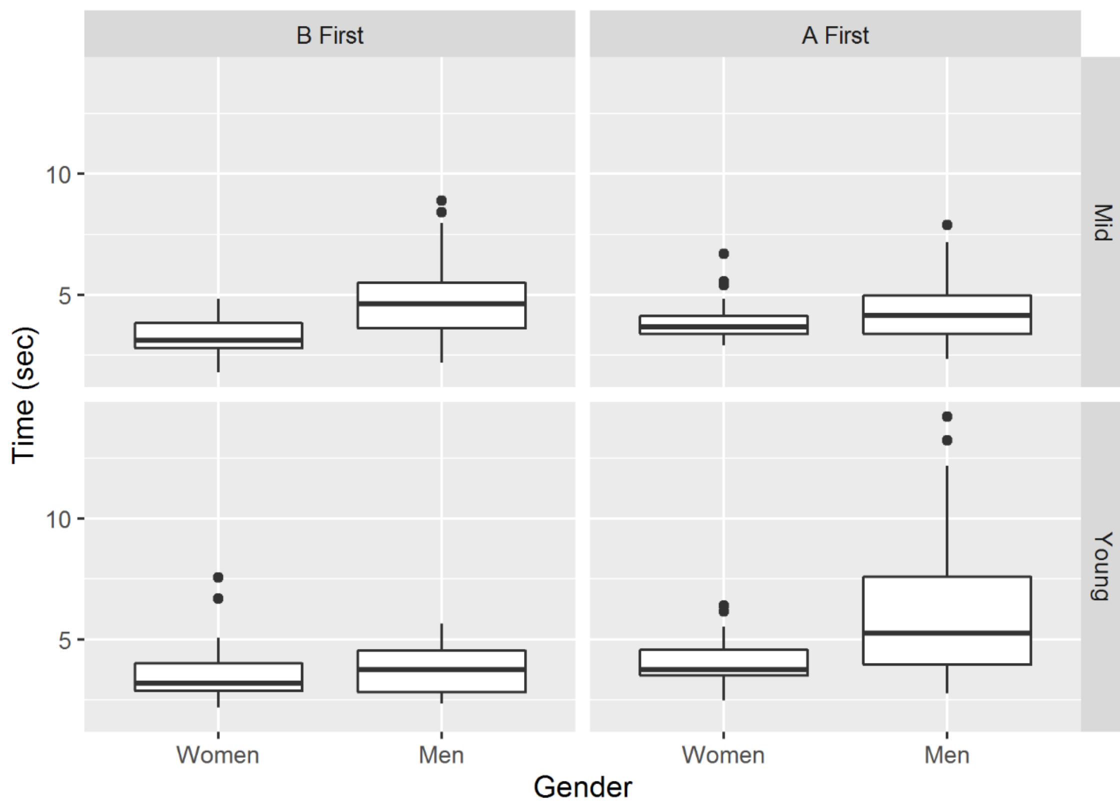


Figure 3.5 – Average response time to the in-cab trust survey, split out by gender, age, and order.

How Did Participants Rate Their Trust Retrospectively?

Participants were asked via survey after each of the two drives to rate their perceived comfort during different portions of the drive. To determine whether the automation capability and drive order influenced driver comfort, we conducted repeated-measures ANOVAs with scenario (more or less capable (A or B)) and order (first or second drive (1 or 2)) as within-subjects factors for each of four questions where participants provided Likert responses.

The first question asked participants to indicate how comfortable they felt when transferring into automated mode (Figure 3.6). Overall, participants felt quite comfortable, and there were no significant effects or interactions involving either scenario or order ($p > 0.05$).

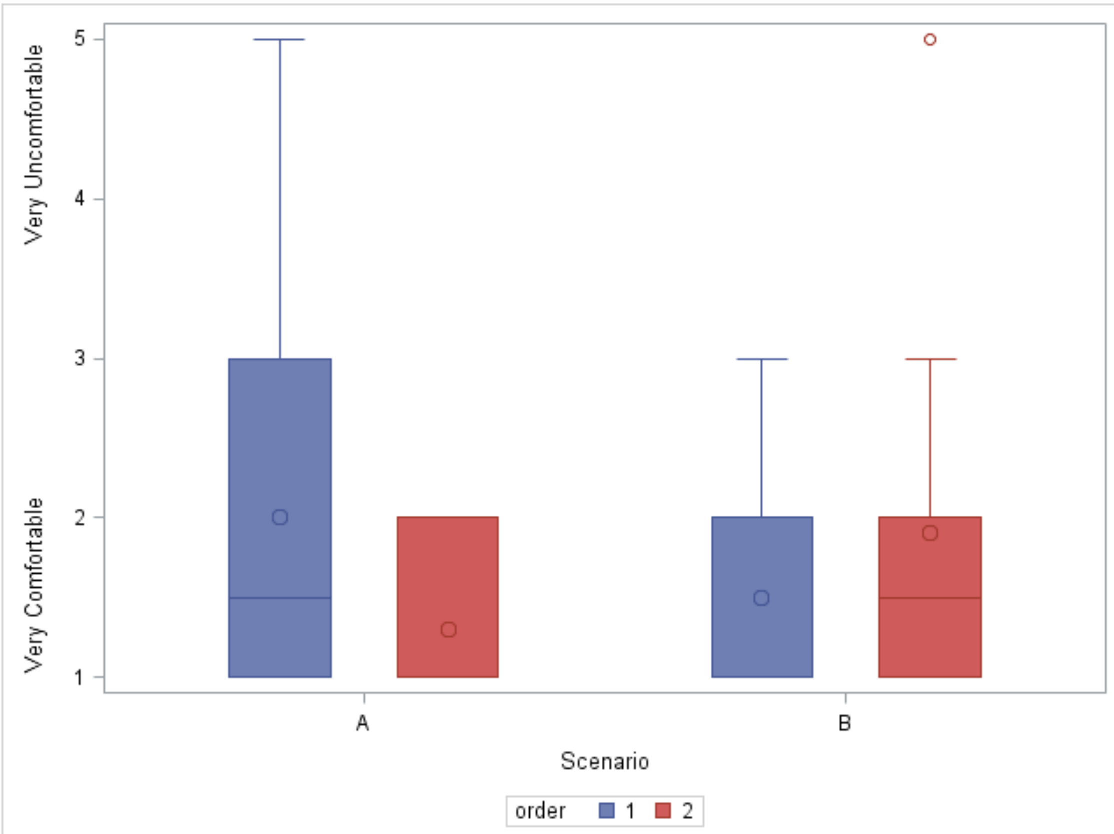


Figure 3.6 – "How comfortable did you feel when transferring into automated mode?"

The second question asked participants how comfortable they felt when resuming manual control back from the automation (Figure 3.7). The main effect of order was marginally significant (p = 0.09), suggesting that drivers tended to be less comfortable in their first drives (1) relative to their second drives (2). This is to be expected as drivers grew more familiar with the automation and transferring control. The main effect of scenario was not significant, nor was the interaction between order and scenario (p > 0.05).

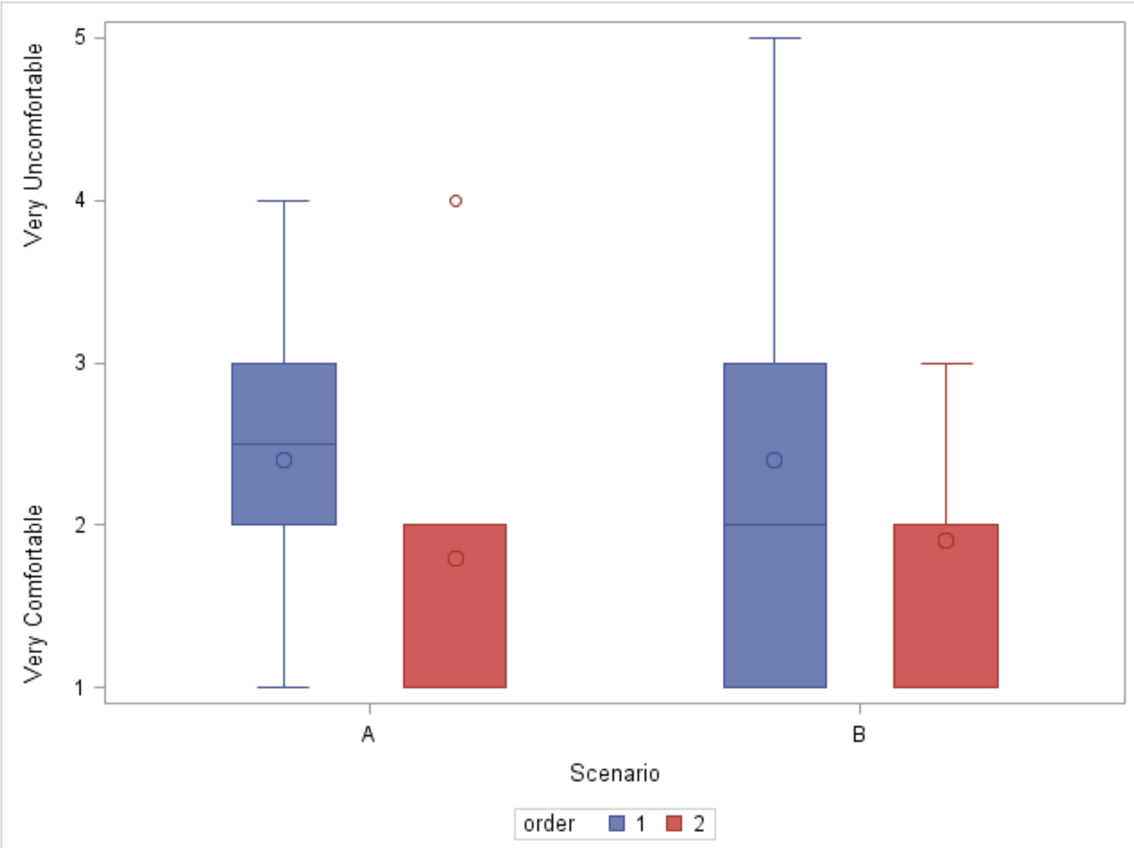


Figure 3.7 – "How comfortable did you feel resuming manual control from the automation?"

The third question asked drivers how comfortable they felt when the automation failed and they had to regain control (Figure 3.8). Neither the main effect of order nor scenario reached significance, nor did the order by scenario interaction ($p > 0.05$).

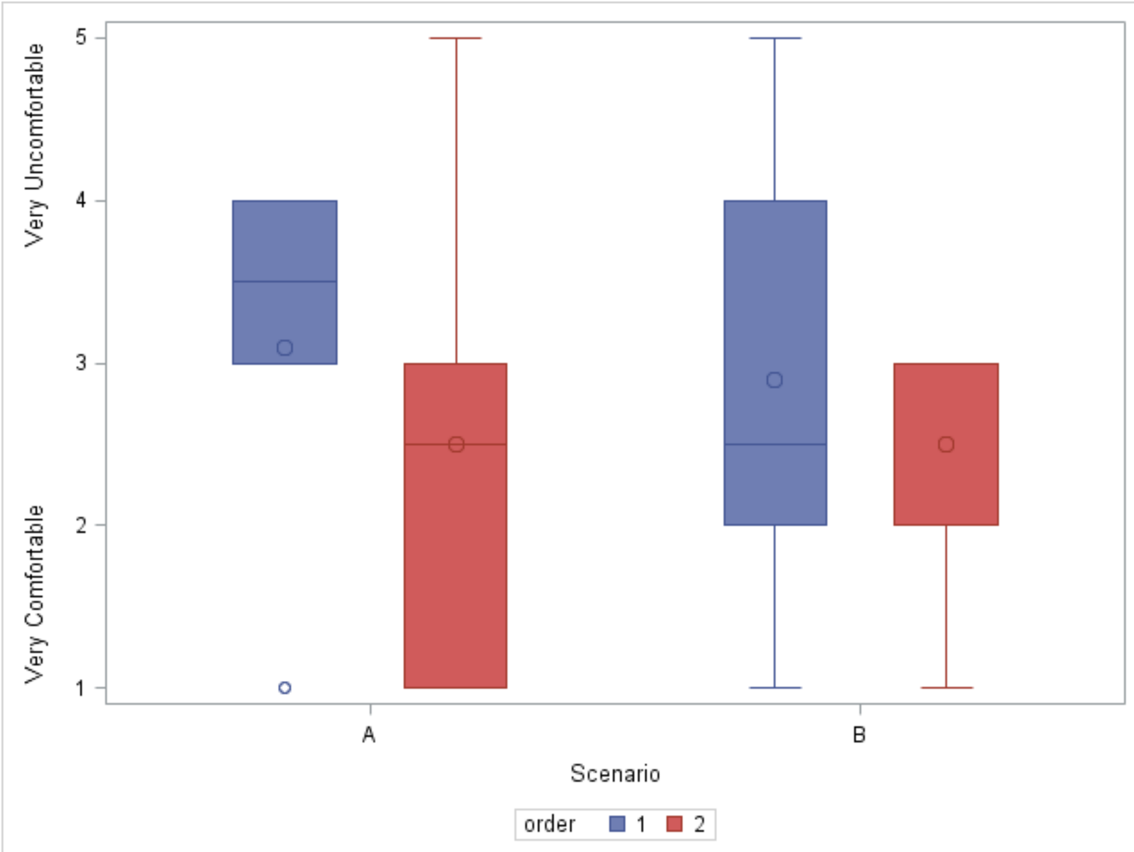


Figure 3.8 – "How comfortable did you feel when the automation failed and you had to regain control?"

The final Likert scale question asked participants how comfortable they felt when driving in automated mode (Figure 3.9). Again, the main effects of order and scenario and the order by scenario interaction did not reach significance ($p > 0.05$).

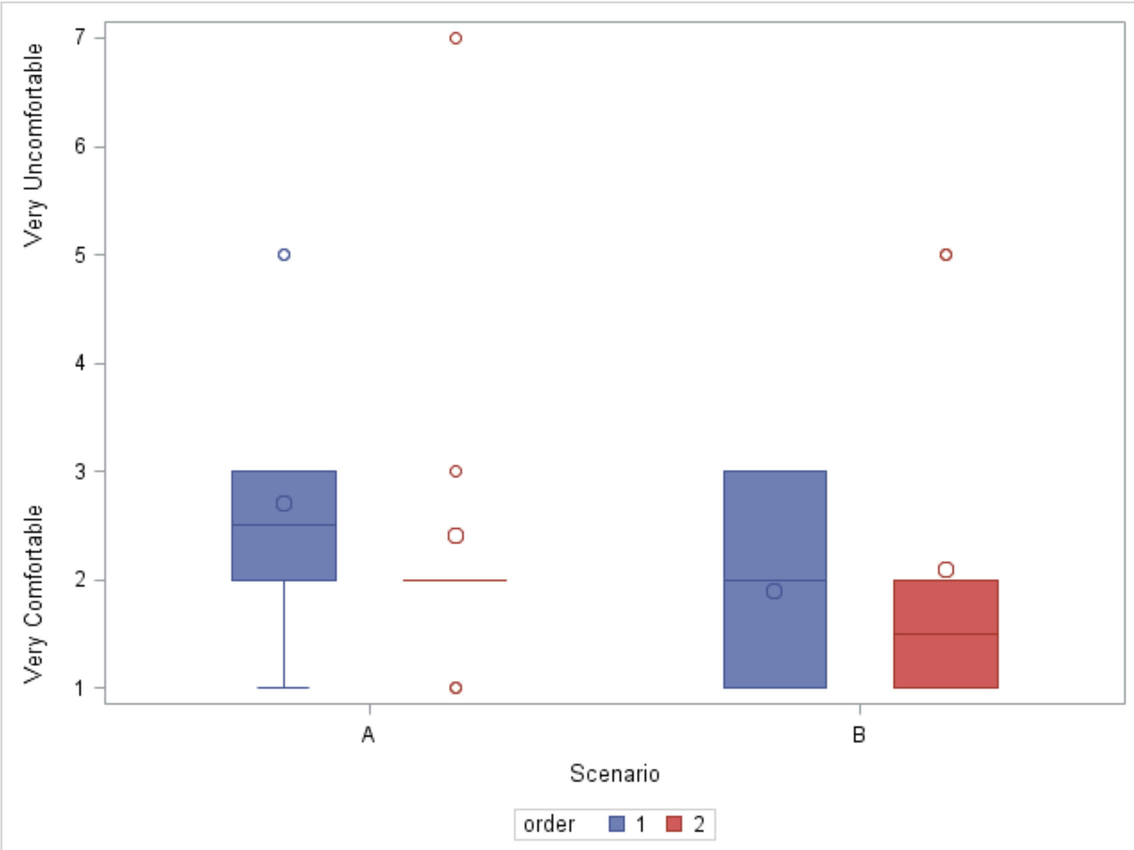


Figure 3.9 – "How comfortable did you feel driving in automated mode?"

These results generally suggest that the capability of the automation (scenario) and the order in which drivers experienced the different conditions had a limited effect on drivers' retrospective perceptions of comfort in interacting with the automation.

3.5 Results on Simulator Measures

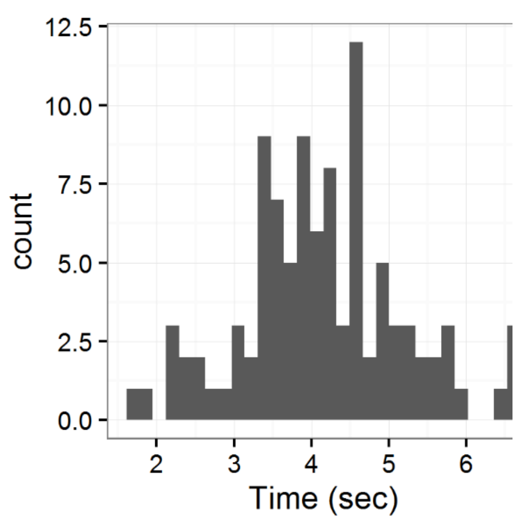
How long do transfers of control take?

Transfers of control from automated to manual operation have several phases that should be considered individually, though some are more difficult to study than others. Situational awareness, for example, is a difficult concept to define, much less measure, and we do not attempt it here, though visual attention is likely a good minimum bound on the time required to regain it. Four phases of takeover from automation are presented in Table 3.8.

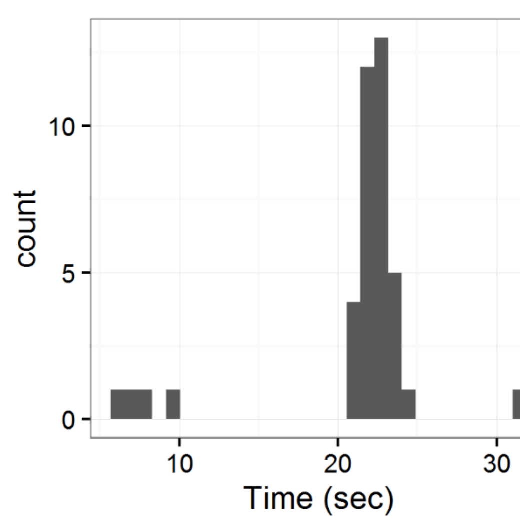
Table 3.8 – Phases of takeover from automated to manual mode.

Takeover Phase	Dependent Measure
Physically taking control by pressing the transfer button or the brake pedal	Takeover response time from cautionary TOR
Physically stabilizing control of the vehicle after taking control	Longitudinal dependent measures for steering and lane keeping
Visually attending to the dynamic driving task	Longitudinal dependent measure for PRC gaze during manual mode
Regaining full situational awareness	None

The first phase may be characterized by the drivers' response times in taking over after being given a TOR. Events 1 through 4 used cautionary TORs. The average response time was 4.13 seconds with a standard deviation of 1.04 seconds (see Figure 3.10a). The exit event, event 5, first issued an information TOR, followed by a cautionary TOR and an imminent TOR, each lasting for 10 seconds. Observe in Figure 3.10b that the distribution of response times for event 5 is tri-modal. Some people responded after the first TOR and some after the third one. One person responded after 30 seconds, which should have been when the vehicle was slowing down and preparing to pull over. The first group had a mean time of 7.60 seconds with standard deviation of 1.28 seconds. The middle, largest, group had a mean response time of 22.37 seconds with standard deviation of 0.85 seconds. The participant in the third group responded at 31.57 seconds. Three-way ANOVAs were run on takeover response time for each event using order, gender, and age. No significant effects of these conditions were found.



(a)



(b)

Figure 3.10 – Distribution of response time to take back manual control after a cautionary TOR for (a) events 1 through 4, and (b) event 5.

The second phase of manual takeover includes the time required to stabilize physical control of the vehicle. The high-frequency control of steering, captured in the HFSteer measure, is thought to be sensitive to distraction. A larger amount of variance was observed in the HFSteer measure in the first six time segments, and less variance was observed in the last six time segments. This is shown in Figure 3.11.

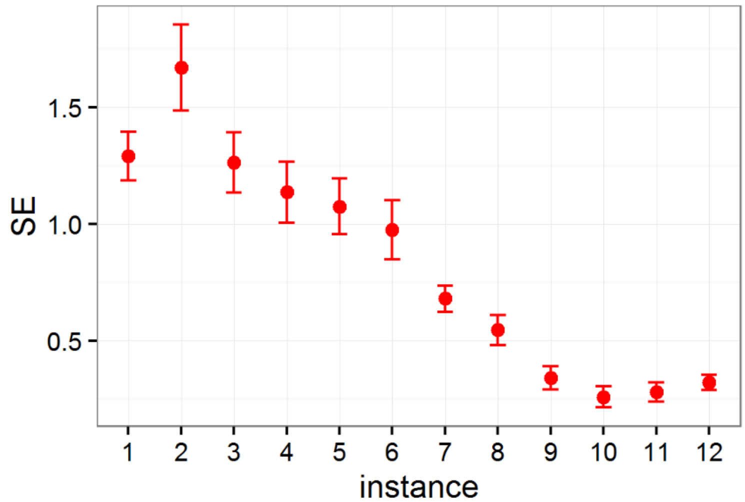


Figure 3.11 – Standard error of HFSteer measure across all participants and all events for each time segment after a manual takeover.

The third phase of manual takeovers considers the time required for the driver to become fully visually engaged in the dynamic driving task. We used the percent road center (PRC) gaze measure recorded using the eye tracker to indicate visual attention. Percent road center has been used not only as a measure of visual distraction, but also to detect cognitive distraction. Simply put, PRC has a normal range, and values that are too low or too high indicate a lack of proper attention.

After manual takeovers, PRC gaze increased as drivers returned their gaze to the road until achieving normal gaze patterns once more. The PRC gaze was calculated on a 17-second running window, which has been used for the detection of distraction [22]. The increasing piece of the PRC gaze trend, up until it peaked, was fit to a linear model, and linear interpolation (or extrapolation, as appropriate) was used to estimate the time at which the PRC would reach 0.7. The distribution of these times is shown in Figure 3.12. In actuality, the PRC never reached 0.7 in some events for some participants. Such cases caused the increasing trend to have a very shallow slope, resulting in very large estimates for the 0.7 intercept time. However, the estimate is useful as a way to compare events and participants against one another.

A three-way ANOVA was run on the estimated 0.7 intercept time for all events in the less-capable condition using gender, age, and order. A significant main effect of age was observed ($F=9.653$, $p=0.003$), as well as a three-way interaction between age, gender, and order ($F=4.733$, $p=0.033$). Boxplots show the differences between groups in Figure 3.13. Generally, it took less time for younger drivers to return their gaze to the road. Moreover, younger male drivers exhibited a pattern wherein their time to return their gaze to the road dropped significantly from their first drive to their second, showing a possible jump in trust.

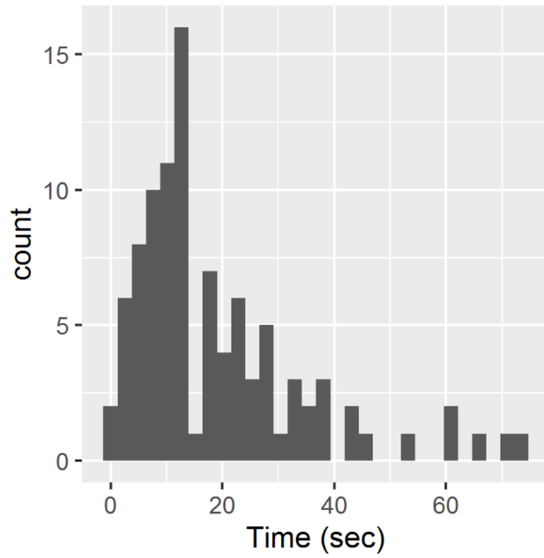


Figure 3.12 – Distribution of times projected for PRC to reach 0.7 after transfer to manual mode.

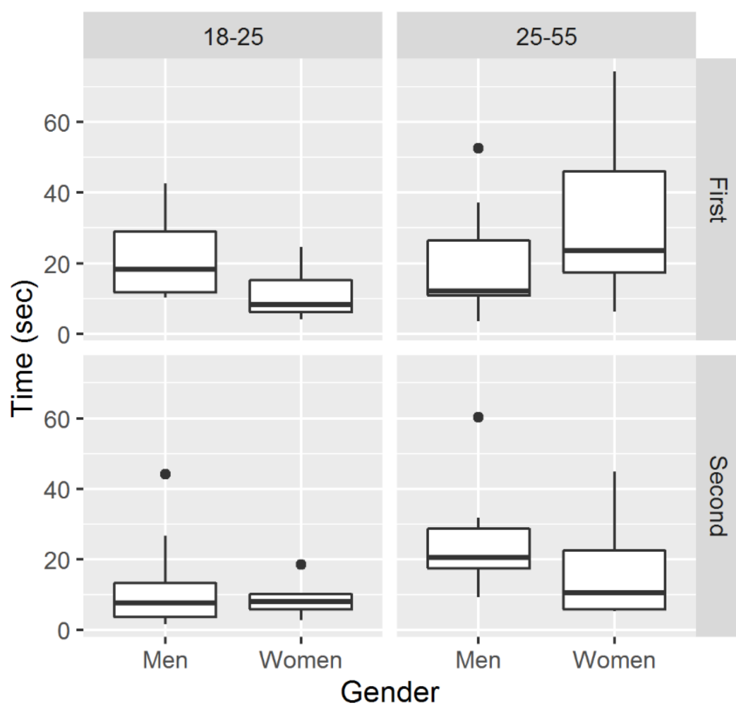


Figure 3.13 – Projected time for PRC gaze to reach 0.7. Younger drivers took significantly less time to return gaze to the road.

Transfers of control from manual to automated mode are simpler in that stabilization and situational awareness are not factors after the transfer. Rather, analyzing transfers to automated mode may tell us about the degree of trust the operator has in the automation. After each event, an audio/visual cue was given to the driver that they could once again transfer control to the automation. The response time was measured from the time this cue was issued. The distribution of response times for the driver to hand back control to the automation is shown in Figure 3.14. After removing the times larger than 20 seconds as outliers, the mean response time was calculated to be 5.31 seconds with standard deviation of 3.15 seconds.

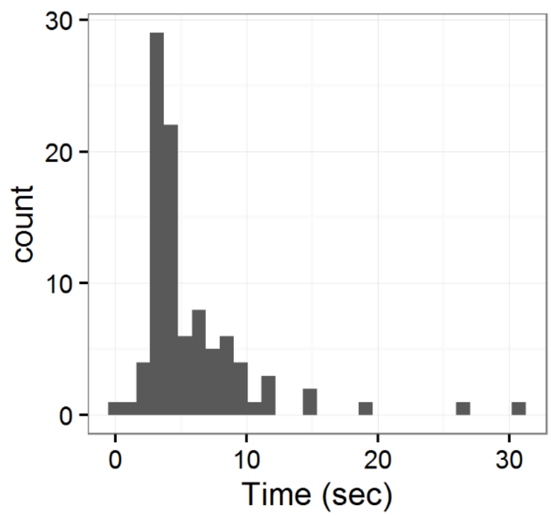


Figure 3.14 – Distribution of response time to give back control to the automation after a reminder cue in events 1 through 4.

A three-way ANOVA was run on the manual-to-automation transfer response time. A significant effect of order (F=8.007, p-value=0.0164), as well as a significant three-way interaction (F=5.772, p-value=0.0351) were found in the slow lead vehicle event with less-capable automation. Participants took longer to respond to the reminder cue and transfer control when they encountered the slow lead vehicle with less-capable automation in their first drive, as compared to when they encountered it in their second drive. This was exaggerated even more for young male participants. The response times split by order, age, and gender are shown in Figure 3.15.

After control was returned to the automation, the PRC gaze dropped until the driver engaged once more with the trivia task. The PRC gaze trend was fit to a linear model, and the time was estimated at which the PRC would reach 0.1. A distribution of these times is shown in Figure 3.16.

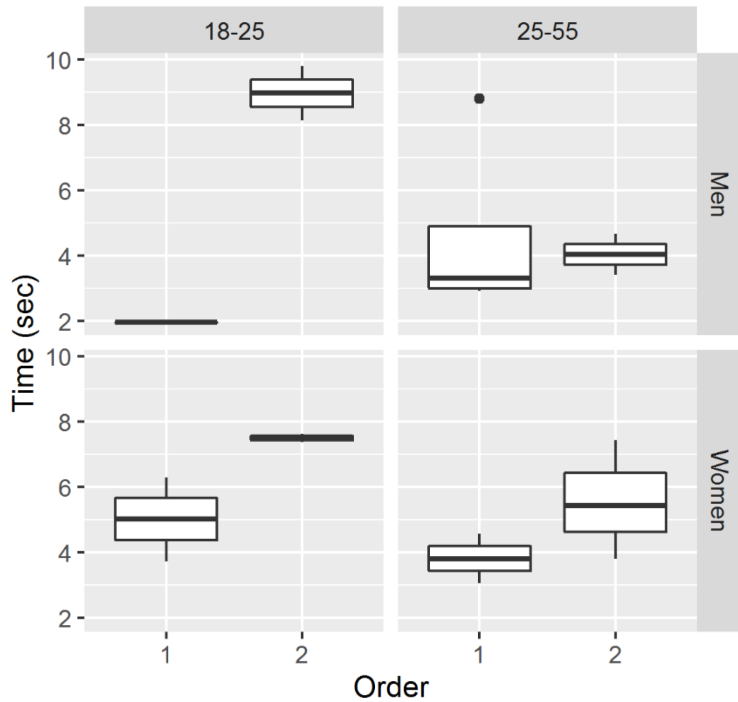


Figure 3.15 – Response time to give back control to the automation after a reminder cue in slow lead vehicle event (event 4), less-capable automation scenario.

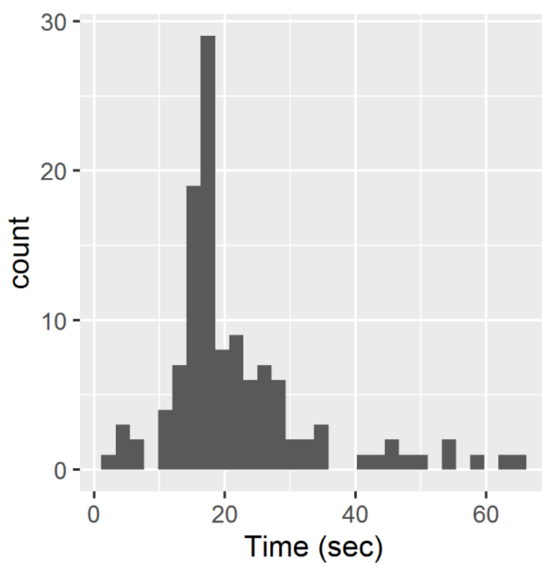


Figure 3.16 – Distribution of times for PRC to reach 0.1 after transfer to automated model.

A three-way ANOVA was run on this time for all events in the less-capable condition using gender, age, and order. A significant main effect of age was observed (F=4.621, p=0.035), as well as a two-way interaction between age and order (F=8.378, p=0.005). Boxplots show the differences between groups in Figure 3.17. Generally, it took less time for younger drivers to take their gaze away from the road. However, older women clearly took longer to reallocate their attention after engaging the automation in the first drive, though they took much less time on their second drive.

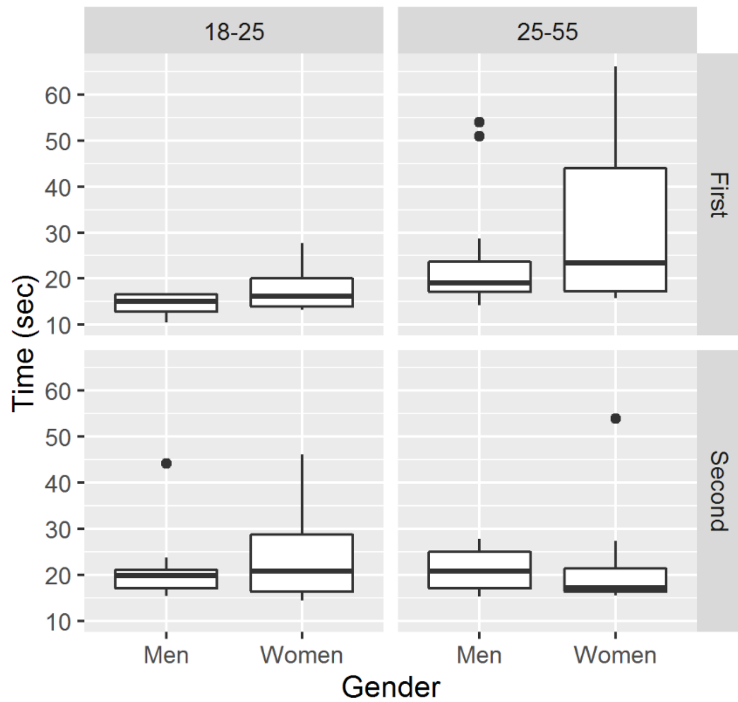


Figure 3.17 – Projected time for PRC gaze to reach 0.1 with less-capable automation. Younger drivers took less time to take their gaze off the road.

How Long Did Drivers Spend in Manual Mode?

The duration that each driver spent in manual mode during the events that required a takeover was measured as part of the data reduction. That duration was heavily dependent on the details of the event. A distribution of the time durations in manual mode, grouped by event, is shown in Figure 3.18. The mean and standard deviation of each group is listed in Table 3.9.

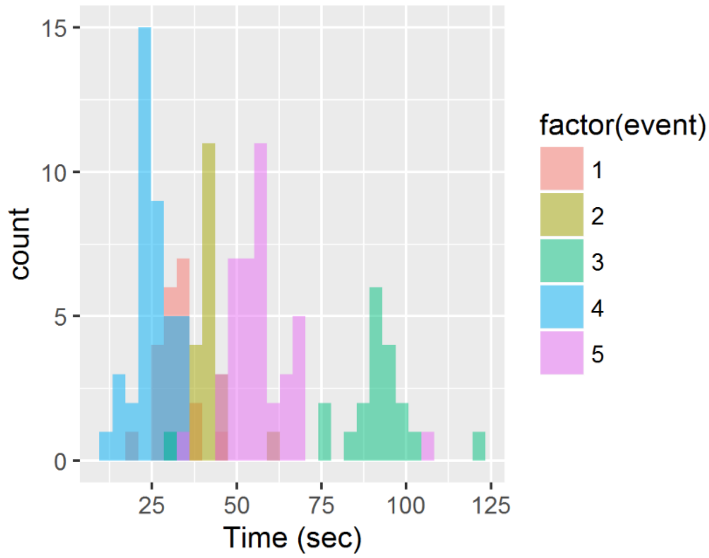


Figure 3.18 – Distribution of time spent in manual mode, grouped by event.

Table 3.9 – Mean and standard deviation of time spend in manual mode, by event.

Event	Event Name	Mean Duration	STD Duration
1	Work Zone	32.06	5.20
2	Missing Lane Lines	41.82	5.73
3	Curve	89.74	17.02
4	Slow Lead Vehicle	24.72	5.21
5	Exit	56.67	10.79

A three-way ANOVA was run on this duration for each event using gender, age, and order. A significant interaction between age and gender was found in the curve event ($F=10.274$, $p=0.009$). Young men spent significantly less time in manual mode than young women (see Figure 3.19). Additionally, a significant effect of age was observed in the slow lead vehicle event with the more-capable automation ($F=5.383$, $p=0.039$), and there was a significant effect of gender in the same event with less-capable automation ($F=6.70$, $p=0.024$). Younger drivers spent less time in manual mode with more-capable automation, while men spent less time in manual mode than women with less-capable automation (see Figure 3.20).

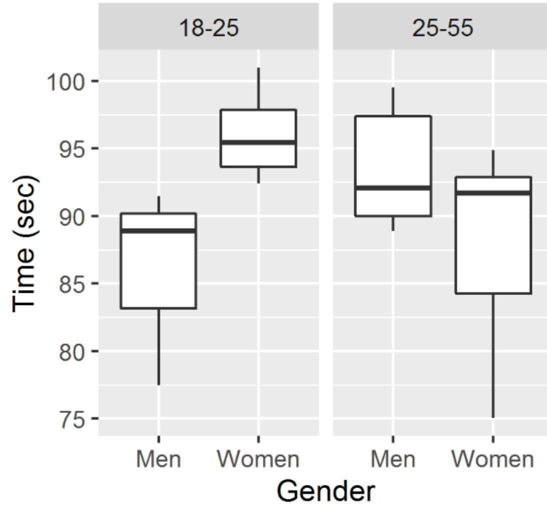
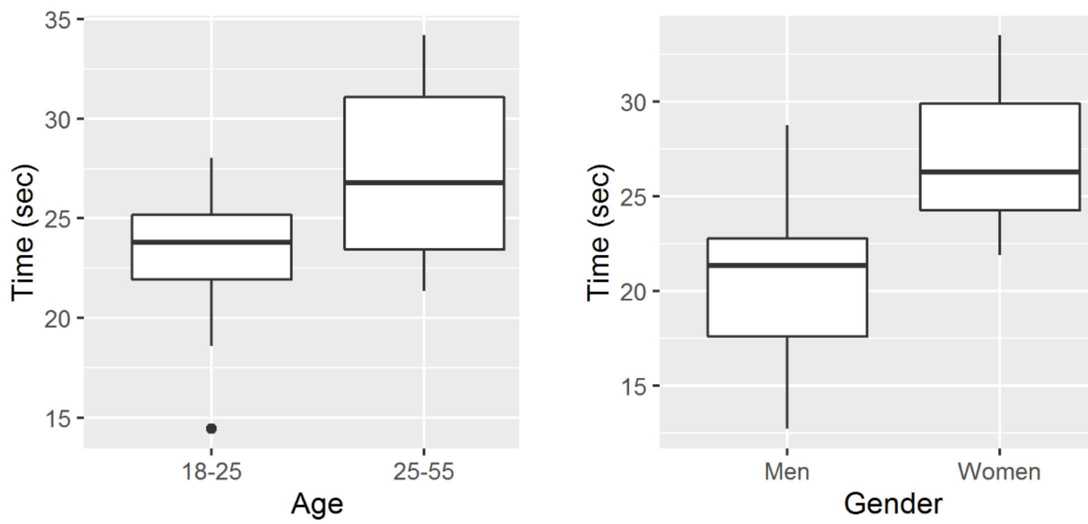


Figure 3.19 – Duration spent in manual mode for curve event. Young males spent significantly less time in manual mode.



(a) (b)

Figure 3.20 – Duration spent in manual mode for lead vehicle event: (a) with more-capable automation, younger drivers spent significantly less time in manual mode, and (b) with less-capable automation, men spent significantly less time in manual mode than women.

3.6 Results on Longitudinal Measures

Several measures were computed in a sequential series of five-second segments after the driver took back manual control from the automation. Moreover, one such measure, PRC gaze, was also computed in similar segments after the driver returned control to the automation. Up to 12 segments, or one minute, were recorded. The duration of manual driving for several events was less than one minute in length, which resulted in fewer than 12 segments for these events. This segmented data provided additional longitudinal variables that could be analyzed using growth curve models. The set of longitudinal variables that were computed for each event is summarized in Table 3.2.

A growth curve analysis was performed on each longitudinal measure in each event to see if there were significant dependencies on gender, age, or order conditions. In most cases, orthogonal polynomials up to second order were computed from the time index (1-12) to fit the growth curve. If an alternate method was used, it is noted in the sections that follow.

In each case, three unconditional growth models were tested first: one that included time as a random effect on the intercept, one that included time as a random effect on the slope, and one that included both intercept and slope in the random effects of the model. The best fit of the three was selected as the base for adding in conditional factors to the model.

Minimum Speed

A gender effect was observed in the slow lead vehicle event (event 4) with less-capable automation. Orthogonal polynomials on the segment index were created for the linear (ot_1) and quadratic (ot_2) terms. Of the three unconditional models that were made, the random slope model was selected as the best fit (the random intercept was removed). This was designated as Model A. The equations governing Model A are written as

Level 1:
$$Y_{ij} = \beta_{0j} + \beta_{1j} ot_{1ij} + \beta_{2j} ot_{2ij} + R_{ij} \quad \textbf{(3.1)}$$
$$R_{ij} \sim \mathcal{N}(0, \sigma^2)$$

Level 2:
$$\beta_{0j} = \gamma_{00}$$
$$\beta_{1j} = \gamma_{10} + U_{1j}$$
$$\beta_{2j} = \gamma_{20} + U_{2j}$$
$$\begin{pmatrix} U_{1j} \\ U_{2j} \end{pmatrix} \sim \mathcal{N}\begin{pmatrix} 0 & & \tau_{10}^2 & \tau_{12} \\ 0 & , & \tau_{12} & \tau_{20}^2 \end{pmatrix}$$

Level 1 refers to the repeated measure, minimum speed (Y_{ij}), and level 2 refers to the subject.

Other factors (age, gender, order) were added as conditions to the fixed and random effects, and the changes in AIC were observed. A model that conditioned the fixed slope on gender was designated as Model B, again with the random intercept dropped, and its governing equations are given by

Level 1:
$$Y_{ij} = \beta_{0j} + \beta_{1j} ot_{1ij} + \beta_{2j} ot_{2ij} + R_{ij} \quad \textbf{(3.2)}$$
$$R_{ij} \sim \mathcal{N}(0, \sigma^2)$$

Level 2:
$$\beta_{0j} = \gamma_{00} + \gamma_{01} Gender_j$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11} Gender_j + U_{1j}$$
$$\beta_{2j} = \gamma_{20} + \gamma_{21} \mathrm{Gender}_j + U_{2j}$$
$$\begin{pmatrix} U_{1j} \\ U_{2j} \end{pmatrix} \sim \mathcal{N}\begin{pmatrix} 0 & & \tau_{10}^2 & \tau_{12} \\ 0 & , & \tau_{12} & \tau_{20}^2 \end{pmatrix}$$

A comparison of Model A and Model B is summarized in Table 3.10. The AIC dropped by about 7, serving as good evidence of improved fit and as an indicator of the significance of gender in Model B. Moreover, the Marginal R^2 value increased from 0.296 to 0.501.

Table 3.10 – Minimum speed in lead vehicle event shows a fixed effect of gender with less-capable automation.

Model	DF	AIC	BIC	Marginal R2	Conditional R2
A	7	654.92	672.95	0.296	0.884
B	10	647.97	673.72	0.501	0.875

The fixed coefficients of Model B are graphed in a caterpillar plot with 95% confidence intervals in Figure 3.21. Intervals that do not include zero may be interpreted as coefficients that are significant to the model. From the figure, it can be seen that there is a significant interaction between ot_1 and gender, as well as a marginally significant interaction between ot_2 and gender.

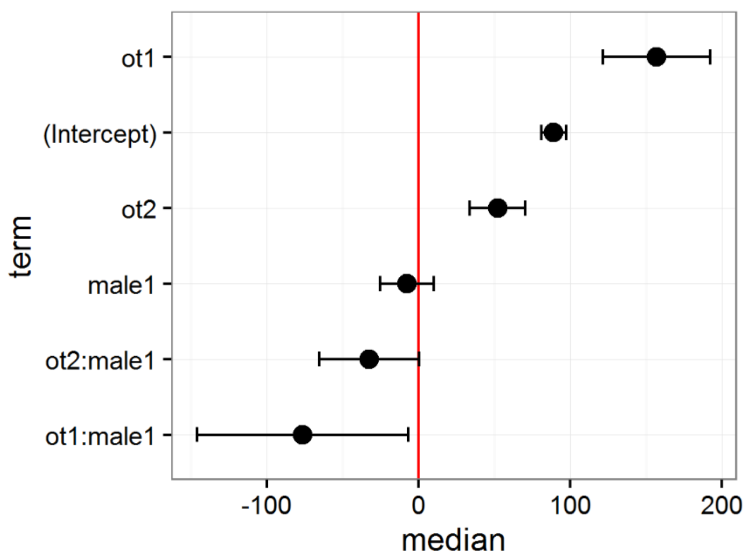


Figure 3.21 – Fixed effect sizes for minimum speed in lead vehicle event, less-capable automation.

The actual effect may be observed by viewing the longitudinal measure of minimum speed for men and women in Figure 3.22. Women systematically achieved lower minimum speeds than men did in the slow lead vehicle event with the less-capable automation.

Gender and age were both significant factors in the exit event (event 5). Orthogonal polynomials on the segment index were created for the linear (ot_1) and quadratic (ot_2) terms. Of the three unconditional models that were made, the random slope model was selected as the best fit (the random intercept was removed). This was designated as Model A. The equations governing Model A are given in equation set (3.1).

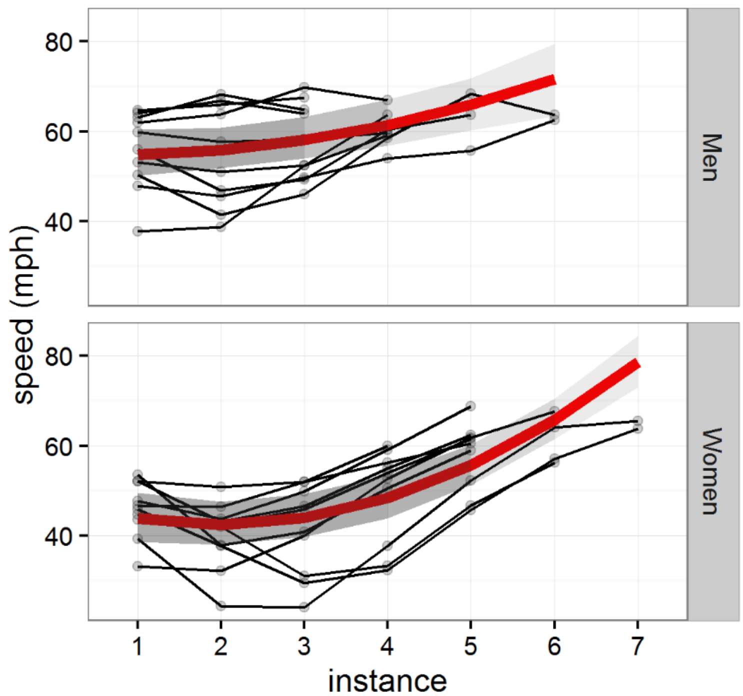


Figure 3.22 – Longitudinal measure of minimum speed grouped by gender.

Other factors (age, gender, order) were added as conditions to the fixed and random effects, and the changes in AIC were observed. A model that conditioned the fixed slope on age and the fixed intercept on gender was designated as Model B, again with the random intercept dropped, and its governing equations are given by

Level 1: $$Y_{ij} = \beta_{0j} + \beta_{1j} ot_{1ij} + \beta_{2j} ot_{2ij} + \beta_{2j} Gender_{2ij} + R_{ij} \quad \textbf{(3.3)}$$

$$R_{ij} \sim \mathcal{N}(0, \sigma^2)$$

Level 2: $$\beta_{0j} = \gamma_{00} + \gamma_{01} Age_j$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} Age_j + U_{1j}$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21} Age_j + U_{2j}$$

$$\beta_{3j} = \gamma_{30}$$

$$\begin{pmatrix} U_{1j} \\ U_{2j} \end{pmatrix} \sim \mathcal{N}\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{matrix} \tau_{10}^2 & \tau_{12} \\ \tau_{12} & \tau_{20}^2 \end{matrix} \Bigg)$$

A comparison of Model A and Model B is summarized in Table 3.11. The AIC dropped by about 20, serving as good evidence of improved fit and as an indicator of the significance of gender and age in Model B. The marginal R^2 value also increased from 0.006 to 0.138.

Table 3.11 – Minimum speed in exit event shows fixed effects of gender (fixed intercept) and age (fixed slope).

Model	DF	AIC	BIC	Marginal R2	Conditional R2
A	7	740.62	762.14	0.006	0.524
B	11	720.44	754.26	0.138	0.563

The fixed coefficients of Model B are graphed in a caterpillar plot with 95% confidence intervals in Figure 3.23. Intervals that do not include zero may be interpreted as coefficients that are significant to the model. From the figure, it can be seen that age is significant, as is the interaction between age and both time terms. Gender may be marginally significant.

The actual effect may be observed by viewing the longitudinal measure of minimum speed grouped by gender and age in Figure 3.24. Women systematically achieved lower minimum speeds than men did in the exit event. Younger drivers tended to slow down in the first 20 seconds after taking over, while the older group had more instances of speeding up as they approached their exit.

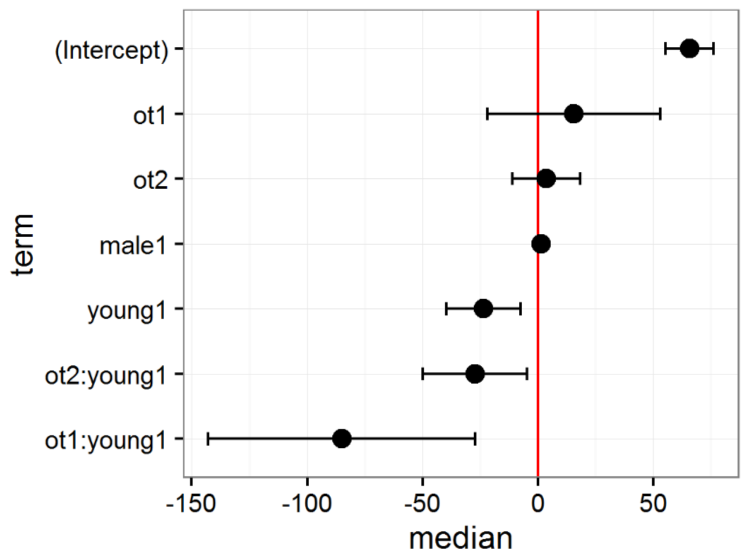


Figure 3.23 – Fixed effect sizes for minimum speed in first 20 seconds of exit event, both scenarios.

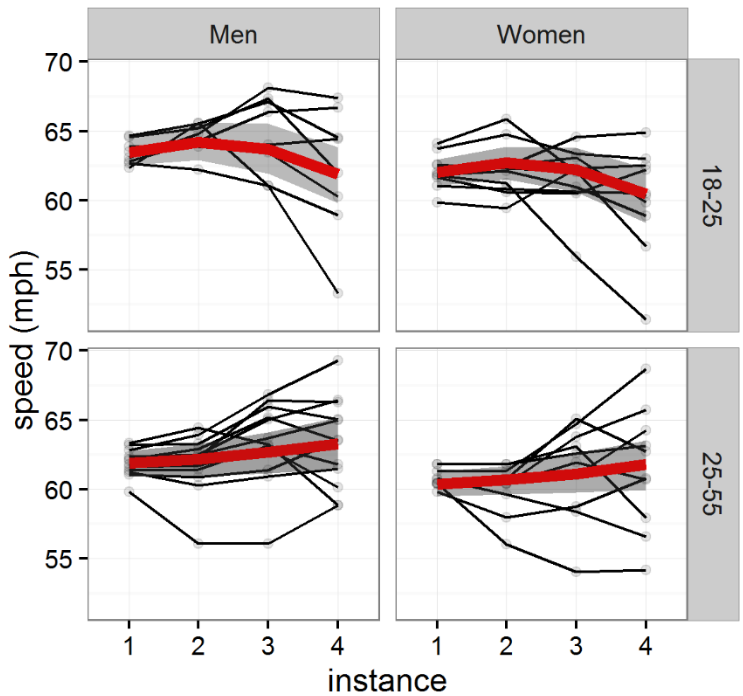


Figure 3.24 – Minimum speed in first 20 seconds of exit event, both scenarios. Fixed effects of age and gender are observed.

Steering Reversal Rate

Age was a significant factor in the work zone event (event 1) with the less-capable automation. Orthogonal polynomials on the segment index were created for the linear (ot_1) and quadratic (ot_2) terms. Of the three unconditional models that were made, the random slope model was selected as the best fit (the random intercept was removed). This was designated as Model A. The equations governing Model A are given in equation set (3.1).

Other factors (age, gender, order) were added as conditions to the fixed and random effects, and the changes in AIC were observed. A model that conditioned the fixed intercept on age was designated as Model B, again with the random intercept dropped, and its governing equations are given by

Level 1: $$Y_{ij} = \beta_{0j} + \beta_{1j}ot_{1ij} + \beta_{2j}ot_{2ij} + \beta_{2j}Age_{2ij} + R_{ij} \quad \textbf{(3.4)}$$

$$R_{ij} \sim \mathcal{N}(0, \sigma^2)$$

Level 2: $$\beta_{0j} = \gamma_{00}$$

$$\beta_{1j} = \gamma_{10} + U_{1j}$$
$$\beta_{2j} = \gamma_{20} + U_{2j}$$
$$\beta_{3j} = \gamma_{30}$$
$$\begin{pmatrix} U_{1j} \\ U_{2j} \end{pmatrix} \sim \mathcal{N}\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} \tau_{10}^2 & \tau_{12} \\ \tau_{12} & \tau_{20}^2 \end{pmatrix}$$

A comparison of Model A and Model B is summarized in Table 3.12. The AIC dropped by about 2.5, indicating the significance of age in Model B. The marginal R^2 value also increased from 0.157 to 0.253.

Table 3.12 – Steering reversal rate in work zone event shows fixed effect of age (fixed intercept).

Model	DF	AIC	BIC	Marginal R2	Conditional R2
A	7	-216.37	-196.19	0.157	0.706
B	8	-219.92	-196.85	0.253	0.693

The fixed coefficients of Model B are graphed in a caterpillar plot with 95% confidence intervals in Figure 3.25. Intervals that do not include zero may be interpreted as coefficients that are significant to the model. From the figure, it can be seen that age is significant.

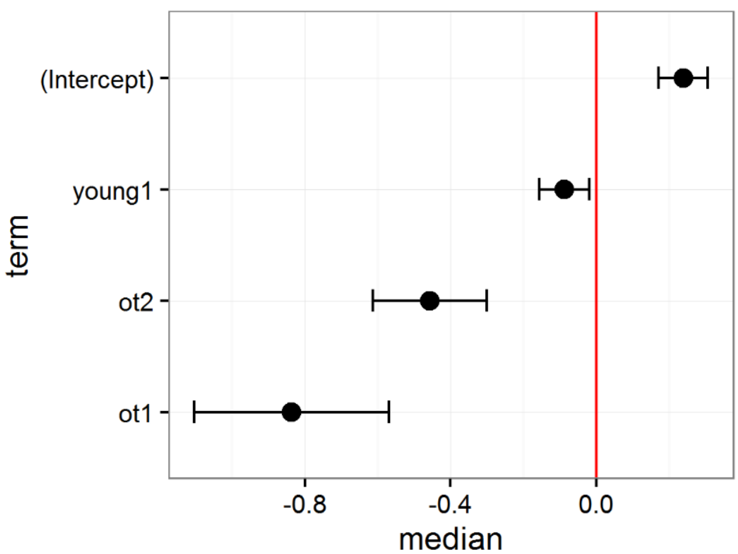


Figure 3.25 – Fixed effect sizes for steering reversal rate, work zone event, less-capable automation.

The actual effect may be observed by viewing the longitudinal measure of steering reversal rate (SRR) grouped by age in Figure 3.26. Younger drivers generally had a lower steering reversal rate than drivers in the older group. Observe that the older group had more variability in SRR as well.

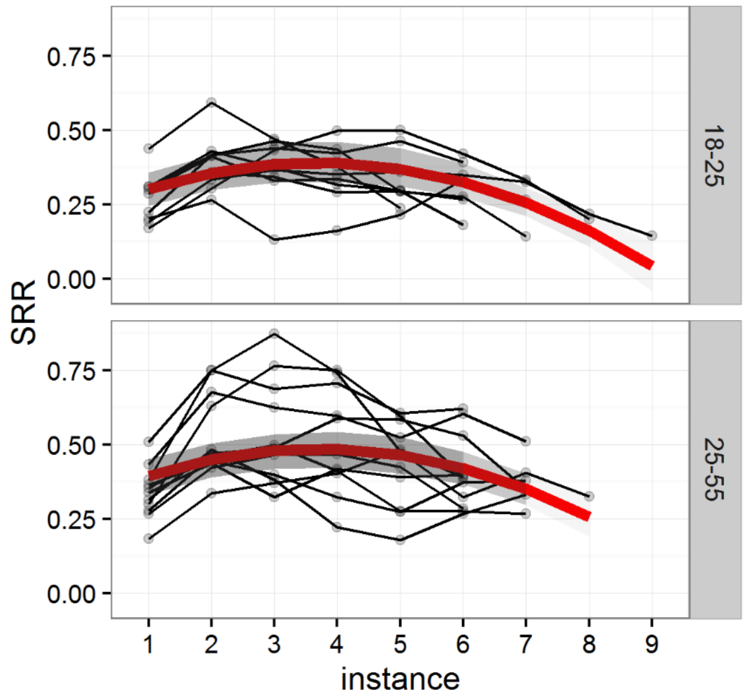


Figure 3.26 – Steering reversal rate in the work zone event, less-capable automation.

Order was a significant factor in the missing lane lines event (event 2) with the less-capable automation. Orthogonal polynomials on the segment index were created for the linear (ot_1) and quadratic (ot_2) terms. Of the three unconditional models that were made, the one with both random slope and intercept was selected as the best fit. This was designated as Model A. The equations governing Model A are given by

Level 1: $$Y_{ij} = \beta_{0j} + \beta_{1j}ot_{1ij} + \beta_{2j}ot_{2ij} + R_{ij} \quad \textbf{(3.5)}$$

$$R_{ij} \sim \mathcal{N}(0, \sigma^2)$$

Level 2: $$\beta_{0j} = \gamma_{00} + U_{0j}$$

$$\beta_{1j} = \gamma_{10} + U_{1j}$$
$$\beta_{2j} = \gamma_{20} + U_{2j}$$
$$\begin{pmatrix} U_{0j} \\ U_{1j} \\ U_{2j} \end{pmatrix} \sim \mathcal{N}\begin{pmatrix} 0 & \tau_{00}^2 & \tau_{01} & \tau_{02} \\ 0 & ,\tau_{01} & \tau_{10}^2 & \tau_{12} \\ 0 & \tau_{02} & \tau_{12} & \tau_{20}^2 \end{pmatrix}$$

Other factors (age, gender, order) were added as conditions to the fixed and random effects, and the changes in AIC were observed. A model that conditioned the fixed slope on order was designated as Model B, and its governing equations are given by

Level 1: $$Y_{ij} = \beta_{0j} + \beta_{1j}ot_{1ij} + \beta_{2j}ot_{2ij} + R_{ij} \quad \textbf{(3.6)}$$

$$R_{ij} \sim \mathcal{N}(0, \sigma^2)$$

Level 2:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}Order_j + U_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}Order_j + U_{1j}$$
$$\beta_{2j} = \gamma_{20} + \gamma_{21}Order_j + U_{2j}$$
$$\begin{pmatrix} U_{0j} \\ U_{1j} \\ U_{2j} \end{pmatrix} \sim \mathcal{N}\left(\begin{matrix} 0 \\ 0 \\ 0 \end{matrix}, \begin{matrix} \tau_{00}^2 & \tau_{01} & \tau_{02} \\ \tau_{01} & \tau_{10}^2 & \tau_{12} \\ \tau_{02} & \tau_{12} & \tau_{20}^2 \end{matrix}\right)$$

A comparison of Model A and Model B is summarized in Table 3.13. The AIC dropped by about 8, indicating the significance of order in Model B. The marginal R^2 value also increased from 0.102 to 0.382.

Table 3.13 – Steering reversal rate in lane lines event shows fixed effect of order.

Model	DF	AIC	BIC	Marginal R2	Conditional R2
A	10	-331.27	-300.15	0.102	0.877
B	13	-339.66	-299.20	0.382	0.881

The fixed coefficients of Model B are graphed in a caterpillar plot with 95% confidence intervals in Figure 3.27. Intervals that do not include zero may be interpreted as coefficients that are significant to the model. From the figure, it can be seen that order is significant, as is its interaction with ot_2.

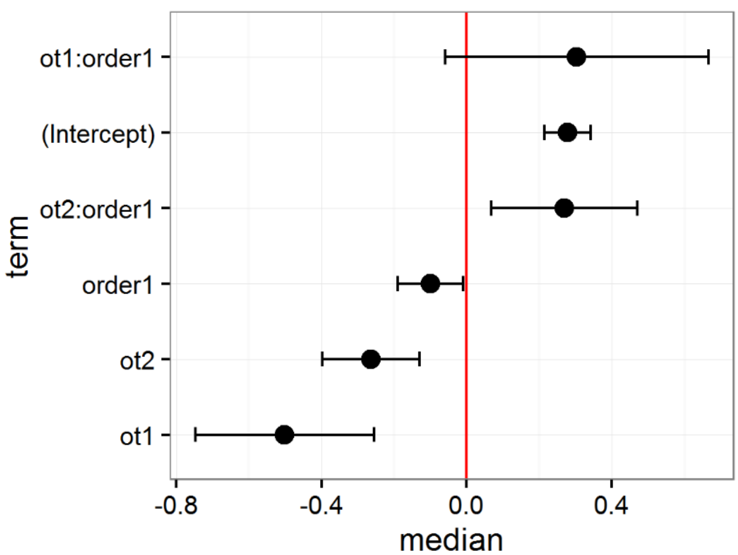


Figure 3.27 – Fixed effect sizes for steering reversal rate, lane lines event, less-capable automation.

The actual effect may be observed by viewing the longitudinal measure of SRR grouped by order in Figure 3.28. Drivers exhibited higher SRR after taking control the first time this event was encountered than they did in the second.

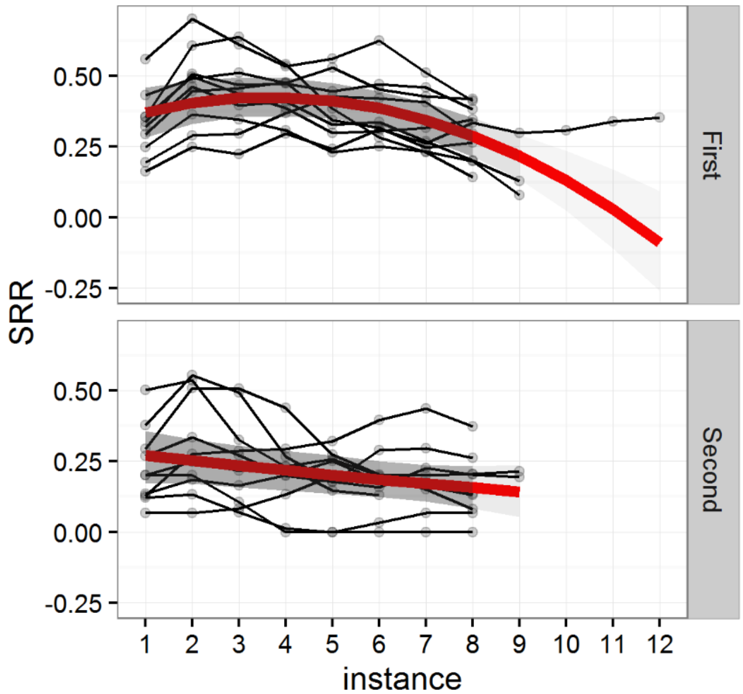


Figure 3.28 – Steering reversal rate in the missing lane lines event, less-capable automation.

Age was a significant factor in the slow lead vehicle event (event 4) with less-capable automation. Orthogonal polynomials on the segment index were created for the linear (ot_1) and quadratic (ot_2) terms. Of the three unconditional models that were made, the one with both random slope and intercept was selected as the best fit. This was designated as Model A. The equations governing Model A are given by the equation set (3.5).

Other factors (age, gender, order) were added as conditions to the fixed and random effects, and the changes in AIC were observed. A model that conditioned the fixed intercept on age was designated as Model B, and its governing equations are given by

Level 1: $$Y_{ij} = \beta_{0j} + \beta_{1j} ot_{1ij} + \beta_{2j} ot_{2ij} + \beta_{2j} Age_{2ij} + R_{ij} \quad \textbf{(3.7)}$$

$$R_{ij} \sim \mathcal{N}(0, \sigma^2)$$

Level 2: $$\beta_{0j} = \gamma_{00} + U_{0j}$$

$$\beta_{1j} = \gamma_{10} + U_{1j}$$
$$\beta_{2j} = \gamma_{20} + U_{2j}$$
$$\beta_{3j} = \gamma_{30}$$
$$\begin{pmatrix} U_{0j} \\ U_{1j} \\ U_{2j} \end{pmatrix} \sim \mathcal{N}\begin{pmatrix} 0 & \tau_{00}^2 & \tau_{01} & \tau_{02} \\ 0 & ,\tau_{01} & \tau_{10}^2 & \tau_{12} \\ 0 & \tau_{02} & \tau_{12} & \tau_{20}^2 \end{pmatrix}$$

A comparison of Model A and Model B is summarized in Table 3.14. The AIC dropped by about 5, indicating the significance of age in Model B. The marginal R^2 value also increased from 0.266 to 0.364.

Table 3.14 – Steering reversal rate in slow lead vehicle event, less-capable automation, shows fixed effect of age.

Model	DF	AIC	BIC	Marginal R2	Conditional R2
A	10	-165.74	-139.99	0.266	0.919
B	14	-170.80	-134.75	0.364	0.932

The fixed coefficients of Model B are graphed in a caterpillar plot with 95% confidence intervals in Figure 3.29. Intervals that do not include zero may be interpreted as coefficients that are significant to the model. From the figure, it can be seen that age appears to be significant.

The actual effect may be observed by viewing the longitudinal measure of SRR grouped by age in Figure 3.30. Younger drivers had slightly lower SRR values.

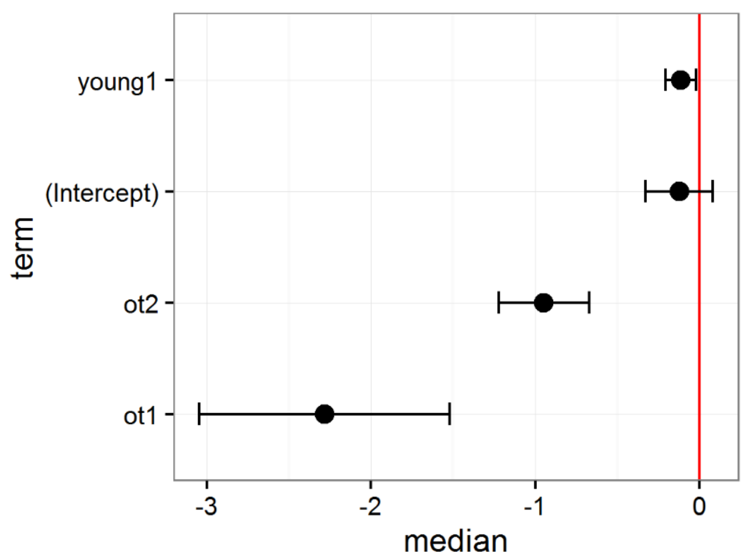


Figure 3.29 – Fixed effect sizes for steering reversal rate, slow lead vehicle event, less-capable automation.

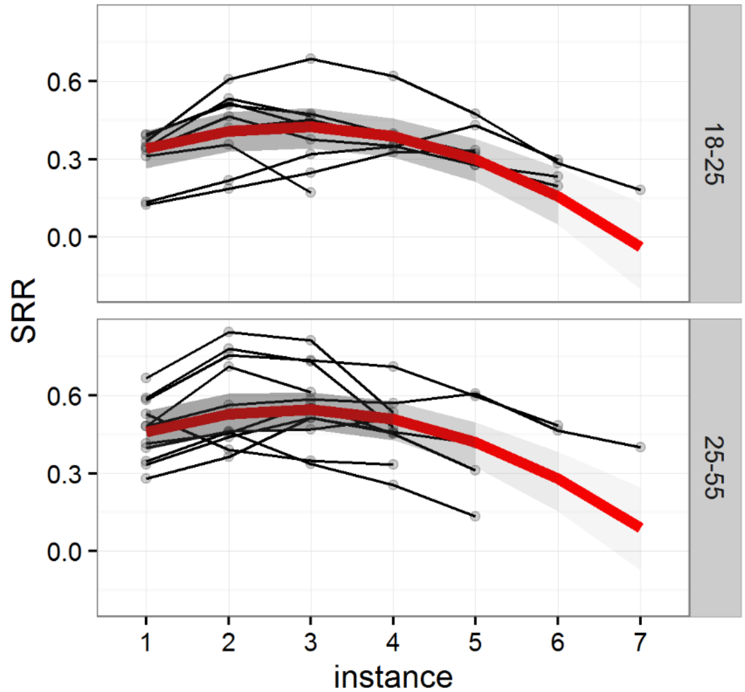


Figure 3.30 – Steering reversal rate in the slow lead vehicle event, less-capable automation.

Standard Deviation of Lane Position

Order was a significant factor in the missing lane lines event (event 2) with less-capable automation. Orthogonal polynomials on the segment index were created for the linear (ot_1) and quadratic (ot_2) terms. Of the three unconditional models that were made, the random slope model was selected as the best fit. This was designated as Model A. The equations governing Model A are given by the equation set (3.1).

Other factors (age, gender, order) were added as conditions to the fixed and random effects, and the changes in AIC were observed. A model that conditioned the fixed intercept on order was designated as Model B, and its governing equations are given by

Level 1: $$Y_{ij} = \beta_{0j} + \beta_{1j}ot_{1ij} + \beta_{2j}ot_{2ij} + \beta_{2j}Order_{2ij} + R_{ij} \quad \textbf{(3.8)}$$

$$R_{ij} \sim \mathcal{N}(0, \sigma^2)$$

Level 2: $$\beta_{0j} = \gamma_{00}$$

$$\beta_{1j} = \gamma_{10} + U_{1j}$$
$$\beta_{2j} = \gamma_{20} + U_{2j}$$
$$\beta_{3j} = \gamma_{30}$$
$$\begin{pmatrix} U_{1j} \\ U_{2j} \end{pmatrix} \sim \mathcal{N}\begin{pmatrix} 0 & , & \tau_{10}^2 & \tau_{12} \\ 0 & & \tau_{12} & \tau_{20}^2 \end{pmatrix}$$

A comparison of Model A and Model B is summarized in Table 3.15. The AIC dropped by about 4, indicating the significance of order in Model B. The marginal R^2 value also increased from 0.028 to 0.068.

Table 3.15 – SDLP in lane lines event, less-capable automation, shows fixed effect of order.

Model	DF	AIC	BIC	Marginal R2	Conditional R2
A	7	-48.35	-26.82	0.028	0.735
B	8	-52.14	-27.54	0.068	0.727

The fixed coefficients of Model B are graphed in a caterpillar plot with 95% confidence intervals in Figure 3.31. Intervals that do not include zero may be interpreted as coefficients that are significant to the model. From the figure, it can be seen that order is significant.

The actual effect may be observed by viewing the longitudinal measure of SDLP grouped by order in Figure 3.32. Drivers had larger, and more varied, SDLP values the first time they encountered the event than they did the second time.

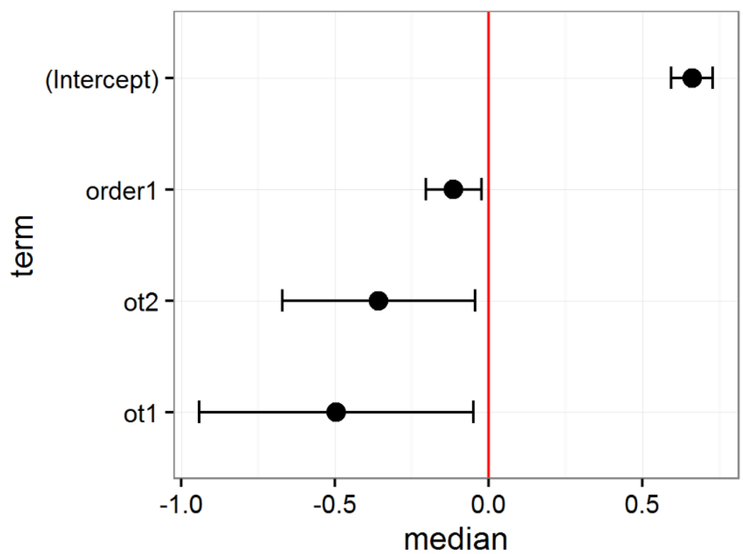


Figure 3.31 – Fixed effect sizes for SDLP, lane lines event, less-capable automation.

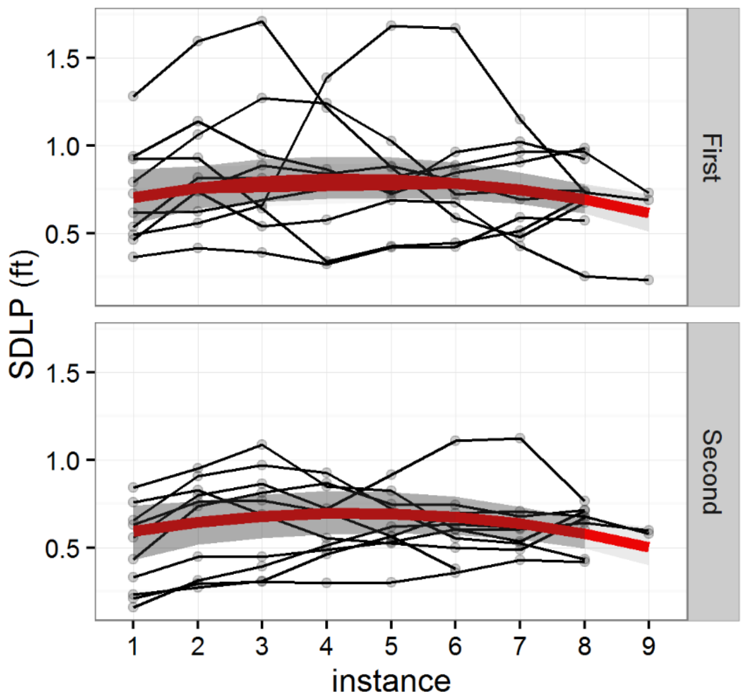


Figure 3.32 – Standard deviation of lane position in the missing lane lines event, less-capable automation.

Age was a significant factor in the slow lead vehicle event (event 4) with the less-capable automation. Orthogonal polynomials on the segment index were created for the linear (ot_1) and quadratic (ot_2) terms. Of the three unconditional models that were made, the random slope model was selected as the best fit. This was designated as Model A. The equations governing Model A are given by the equation set (3.1).

Other factors (age, gender, order) were added as conditions to the fixed and random effects, and the changes in AIC were observed. A model that conditioned the fixed intercept on age was designated as Model B, and its governing equations are given by

Level 1: $$Y_{ij} = \beta_{0j} + \beta_{1j} ot_{1ij} + \beta_{2j} ot_{2ij} + \beta_{2j} Age_{2ij} + R_{ij} \quad \textbf{(3.9)}$$

$$R_{ij} \sim \mathcal{N}(0, \sigma^2)$$

Level 2: $$\beta_{0j} = \gamma_{00}$$

$$\beta_{1j} = \gamma_{10} + U_{1j}$$
$$\beta_{2j} = \gamma_{20} + U_{2j}$$
$$\beta_{3j} = \gamma_{30}$$
$$\begin{pmatrix} U_{1j} \\ U_{2j} \end{pmatrix} \sim \mathcal{N}\begin{pmatrix} 0 & , & \tau_{10}^2 & \tau_{12} \\ 0 & & \tau_{12} & \tau_{20}^2 \end{pmatrix}$$

A comparison of Model A and Model B is summarized in Table 3.16. The AIC dropped by about 4, indicating the significance of age in Model B. The marginal R^2 value also increased from 0.227 to 0.411.

Table 3.16 – SDLP in slow lead vehicle event, less-capable automation, shows fixed effect of age.

Model	DF	AIC	BIC	Marginal R2	Conditional R2
A	**7**	**-42.60**	**-27.82**	**0.227**	**0.945**
B	**8**	**-46.06**	**-29.18**	**0.411**	**0.932**

The fixed coefficients of Model B are graphed in a caterpillar plot with 95% confidence intervals in Figure 3.33. Intervals that do not include zero may be interpreted as coefficients that are significant to the model. Age is seen to be significant from the figure.

The actual effect may be observed by viewing the longitudinal measure of SDLP grouped by order in Figure 3.34. Younger drivers exhibited larger SDLP during the slow lead vehicle event with the less-capable automation.

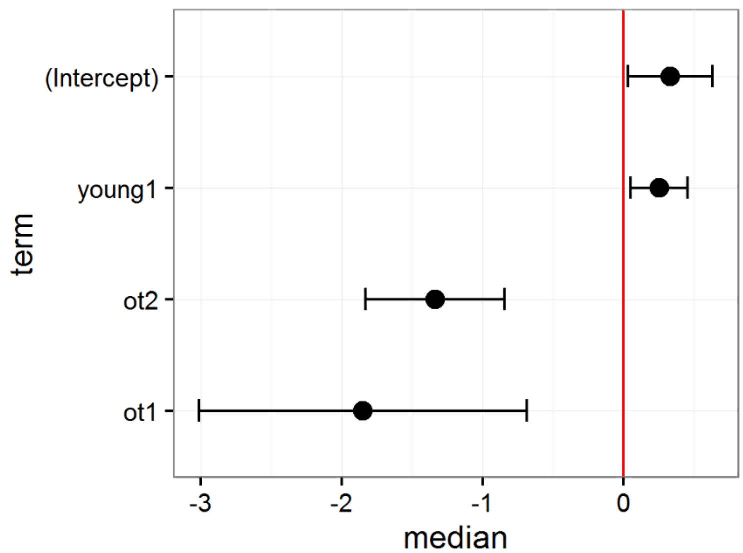


Figure 3.33 – Fixed effect sizes for SDLP, slow lead vehicle event, less-capable automation.

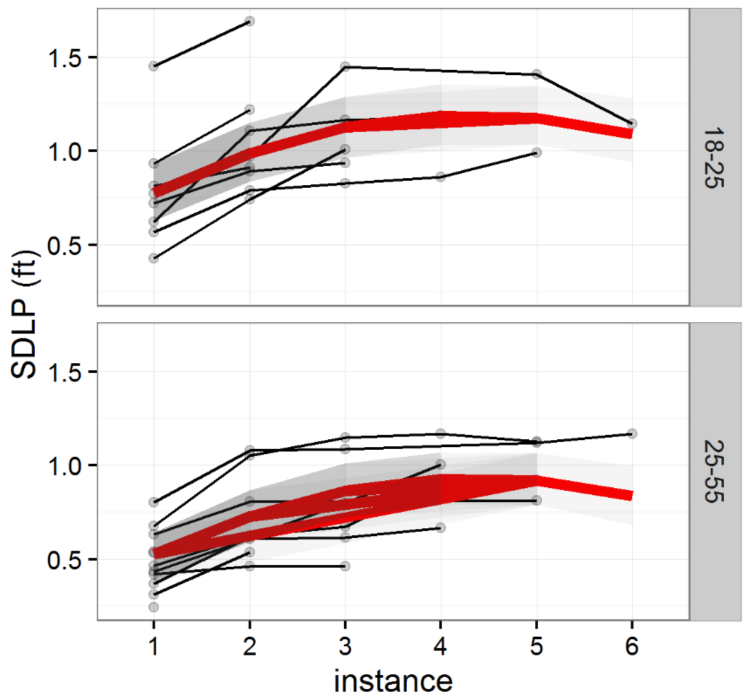


Figure 3.34 – Standard deviation of lane position in the slow lead vehicle event, less-capable automation.

High-Frequency Steering

Order was a significant factor in the slow lead vehicle event (event 4) with less-capable automation. Orthogonal polynomials on the segment index were created for the linear (ot_1) and quadratic (ot_2) terms. Of the three unconditional models that were made, the random slope model was selected as the best fit. This was designated as Model A. The equations governing Model A are given by the equation set (3.1).

Other factors (age, gender, order) were added as conditions to the fixed and random effects, and the changes in AIC were observed. A model that conditioned the fixed slope on order was designated as Model B, and its governing equations are given by

Level 1: $$Y_{ij} = \beta_{0j} + \beta_{1j}ot_{1ij} + \beta_{2j}ot_{2ij} + \mathrm{R}_{ij} \qquad \textbf{(3.10)}$$

$$R_{ij} \sim \mathcal{N}(0, \sigma^2)$$

Level 2: $$\beta_{0j} = \gamma_{00}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}Order_j + U_{1j}$$
$$\beta_{2j} = \gamma_{20} + \gamma_{21}Order_j + U_{2j}$$
$$\begin{pmatrix} U_{1j} \\ U_{2j} \end{pmatrix} \sim \mathcal{N}\begin{pmatrix} 0 & & \tau_{10}^2 & \tau_{12} \\ 0 & , & \tau_{12} & \tau_{20}^2 \end{pmatrix}$$

A comparison of Model A and Model B is summarized in Table 3.17. The AIC dropped by about 9, indicating the significance of order in Model B. The marginal R^2 value also increased from 0.377 to 0.415.

Table 3.17 – High-frequency steering in slow lead vehicle event, less-capable automation, shows fixed effect of order.

Model	DF	AIC	BIC	Marginal R2	Conditional R2
A	7	-8.54	9.48	0.377	0.710
B	10	-17.79	7.95	0.415	0.764

The fixed coefficients of Model B are graphed in a caterpillar plot with 95% confidence intervals in Figure 3.35. Intervals that do not include zero may be interpreted as coefficients that are significant to the model. Order is seen to be a significant factor, as are the interactions of order with ot_1 and ot_2.

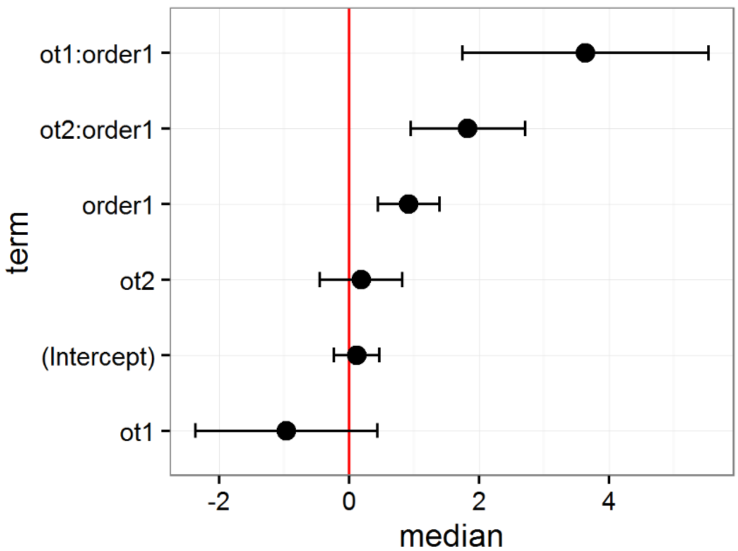


Figure 3.35 – Fixed effect sizes for high-frequency steering, slow lead vehicle event, less-capable automation.

The actual effect may be observed by viewing the longitudinal measure of high-frequency steering grouped by order in Figure 3.36. The fixed effect of polynomial time takes on a different shape depending on whether it was in the first or second drive. Were it not for two cases in particular in the first order group, the fixed portion of the model would have been more similar to the second.

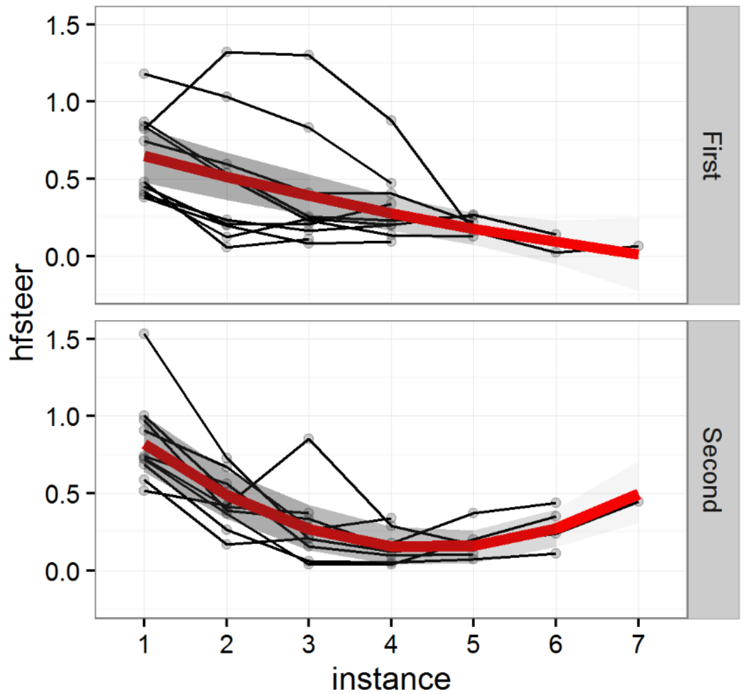


Figure 3.36 – High-frequency steering in the slow lead vehicle event, less-capable automation.

4 Discussion and Conclusions

Twenty participants took part in an automated driving study using the NADS-1 motion base driving simulator. The automation was described generally as SAE Level 3 (conditional automation), however it was implemented as SAE Level 4 with a fail-safe mode to mitigate the risk that an automated vehicle would actually collide with a lead vehicle or drive through a work zone. Those negative outcomes did not happen, and the fail-safe mode was not needed in any of the events, except perhaps during one participant's exit event at the end of a drive.

Comfort was measured using a probe survey that was administered twice during the practice drive and eight times during each main drive. The probes did not take place during events, but a few minutes later while all was normal. Also, a post-drive survey was administered after each main drive; it asked the participants to retrospectively consider their comfort with the automation. We surmised that asking about comfort would be an effective way to capture the nascent trust of an operator just becoming familiar with an automation system. Future work could delve deeper into multiple facets of trust, including predictability/performance, dependability/reliance, faith, and collaboration.

A cluster analysis revealed three distinct longitudinal comfort profiles from the probe surveys. One cluster started with a high level of comfort and stayed that way. Another started with a lower level of comfort, but it gradually increased after a few surveys and then stayed level. A third cluster started with low comfort and gradually increased over the course of the practice and two main drives. Apart from single instances of reduced comfort, only participant 4 showed a temporary trend of decreasing comfort. We could not associate the clustering with age, gender, or order. It may be that it is associated with some latent variable such as sensation seeking or a personality trait. The longitudinal comfort profiles support the notion that trust can be modeled as a function of time, especially in the sense that instantaneous levels of trust depend on their previously measured levels [32].

Although it is interesting, it is uncertain whether the significant effects of order, gender, and various interactions on the response time to the probe surveys speak directly to differences in trust. The probe survey could conceivably be thought of as a distraction task; however, the primary task at the time was not the dynamic driving task (DDT), but the trivia game. We surmise that it does speak to how engrossed the occupant was in their task and how willing they were to redirect their attention from it. On average, men took more time to respond than women. Interestingly, the response time was slightly higher when the more-capable automation was driven first than it was with the opposite ordering. The largest response time of any group was from young men experiencing more-capable automation first. Since the scenario with greater automation capability did not issue any TORs until the fourth and fifth events, it may be an indicator of how much people allowed themselves to be taken out of the loop, and this may be reflective of an increased amount of trust.

The physical response times to TORs and automation reminders were both under 10 seconds (4.13 sec +- 1.04 sec and 5.31 sec +- 3.15 sec, respectively). The group that differed from the means the most was young men giving back control. There was a trend of taking longer to give back control in the second drive. Young men, especially, had the fastest responses in the first drive (~2 sec) and the longest responses in the second drive (~9 sec). As drivers learned to trust the automation, they also learned its capabilities and limitations. The increased response time in the second drive could be an indication that they were calibrating their trust in the automation by waiting until a safe time to give it control.

We measured visual attention to the driving task through the percent road center gaze, calculated over a 17-second running window. Longitudinal measures of PRC after manual takeovers as well as PRC after return to automation were calculated, and their trends were identified through linear regression fits. The analysis showed that young men were the quickest to return their eyes to the road. Generally, younger drivers took the least amount of time to take their eyes back off the road after returning to automation. The older female group took the most time to take their eyes off the road, but only in their first drive. Their behavior in the second drive was more normal. The results indicate that the younger group has a greater ability to quickly switch contexts between the DDT and the trivia game. This may be partly due to greater trust, and it may also be due to greater mental agility or greater comfort with technology.

Consideration of the response times for physical takeovers, stabilization, and visual attention leads to concern for the driver's safety after taking control. Drivers are capable of physically taking over control in less than five seconds. However, PRC gaze showed that it could take 20 seconds or more to return their full attention to the roadway. Additionally, the variation in high-frequency steering offers evidence that drivers do not return to their normal driving control for up to 30 seconds. This leaves a 15- to 25-second gap during which the driver may be vulnerable to missing a response to a safety-critical event at an inopportune moment.

It is remarkable that so many significant effects were observed for the slow lead vehicle event in the scenario with less-capable automation. This event is the closest thing to a safety-critical event in the study, but it was by no means as severe as a forward collision event might be. Women were seen to achieve lower minimum speeds than men. Men spent more time in manual mode than did women. Younger drivers had a lower SRR and larger SDLP than did the older group. Finally, when drivers experienced this event in their first drive, they tended to have larger amounts of high-frequency steering than when they experienced it in their second drive. No other event exposed differences between the study groups as well as this one did.

We were able to see the development of comfort in a vehicle automation system over the course of a practice drive and two main drives. As exposure to the system and its functions increased, so did comfort levels. This, however, is only the beginning stage of trust, and it most likely conflated with other factors such as an increase in self-confidence in using the system. A longer process was that of learning about the capabilities and limitations of the automation, which they did through the five main study events and the extra events that never required intervention. This learning process is associated with the proper calibration of trust, and we were able to observe it in differences between the first and second drives. Drivers were observed to adjust the amount of time before handing back control to the automation, and they were better able to manage their lane-keeping ability in the second drive.

The main limitations of this study were that it used a fairly small sample size (20 participants), and that it was not able to fully explore the different dimensions of trust. Future research should address both of those limitations. Additionally, the inclusion of safety-critical events would allow a better judgement of whether the driver has regained SA and whether the takeover times observed have an adverse effect on safety. We modeled our driver-vehicle interface (DVI) largely on previous research. However, there is still much that could be done to test different modalities and timing for DVI design. Finally, we conjecture that the best DVI would be one that is capable of monitoring the driver and adapting elements of the DVI, transfers of control, and other aspects of the automation to the perceived state of the operator.

References

1. NHTSA. (2013). *NHTSA Preliminary Statement of Policy Concerning Automated Vehicles*. Retrieved August 25, 2016, from http://www.nhtsa.gov/staticfiles/rulemaking/pdf/Automated_Vehicles_Policy.pdf.
2. SAE. (2014). Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems. Standard J3016. SAE.
3. Bainbridge, L. (1983). Ironies of Automation. *Automatica* 19 (6): 775–79. doi:10.1016/0005-1098(83)90046-8.
4. Stanton, N.A., & Marsden, P. (1996). From fly-by-wire to drive-by-wire: Safety implications of automation in vehicles. *Safety Science* 24 (1): 35–49. doi:10.1016/S0925-7535(96)00067-7.
5. Stanton, N. A., Young, M., & McCaulder, B. (1997). Drive-by-wire: The case of driver workload and reclaiming control with adaptive cruise control. *Safety Science* 27(2-3), 149-159. doi: 10.1016/S0925-7535(97)00054-4
6. Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 37 (2): 381–94. doi:10.1518/001872095779064555.
7. Endsley, M. R. (1996). Automation and situation awareness. In *Automation and Human Performance: Theory and Applications*, xx:163–81. Human Factors in Transportation. Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.
8. Kaber, D. B., & M. R. Endsley. (1997). Out-of-the-loop performance problems and the use of intermediate levels of automation for improved control system functioning and safety. *Process Safety Progress* 16 (3): 126–131. doi:10.1002/prs.680160304.
9. Merat, N., & Jamson, A. H. (2009). How do drivers behave in a highly automated car? In Proceedings of the 5th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design.
10. Merat, N., Jamson, A. H., Lai, F.C.H., & Carsten, O. (2012). Highly automated driving, secondary task performance, and driver state. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 54 (5): 762–71. doi:10.1177/0018720812442087.
11. Merat, N., Jamson, A. H., Lai, F. C. H., Daly, M., & Carsten, O. M. J. (2014). Transition to manual: driver behaviour when resuming control from a highly automated vehicle. *Transportation Research Part F: Traffic Psychology and Behaviour*. Retrieved November 30, 2015. doi:10.1016/j.trf.2014.09.005.
12. Gold, C., Damböck, D., Lorenz, L. & Bengler, K. (2013). Take over!' How long does it take to get the driver back into the loop? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 57 (1): 1938–42. doi:10.1177/1541931213571433.
13. Blanco, M., Atwood, J., Vasquez, H. M., Trimble, T. E., Fitchett, V. L., Radlbeck, J., Fitch, G. M. et al. (2015). *Human Factors Evaluation of Level 2 and Level 3 Automated Driving Concepts*. Final Report DOT HS 812 182. Washington, D.C.: NHTSA.

14. Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 46 (1): 50–80. doi:10.1518/hfes.46.1.50_30392.
15. Jamson, A. H., Merat, N., Carsten, O. M. J., & Lai, F. C. H. (2013). Behavioural changes in drivers experiencing highly-automated vehicle control in varying traffic conditions. *Transportation Research Part C: Emerging Technologies* 30 (May): 116–25. doi:10.1016/j.trc.2013.02.008.
16. Evers, V., Maldonado, H., Brodecki, T., & Hinds, P. (2008). Relational vs. group self-construal: Untangling the role of national culture in HRI. In *2008 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 255–62. doi:10.1145/1349822.1349856.
17. Sanders, T., Oleson, K. E., Billings, D. R., Chen, J. Y. C., & Hancock, P. A. (2011). A model of human-robot trust theoretical model development. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 55 (1): 1432–36. doi:10.1177/1071181311551298.
18. Price, M. A., Venkatraman, V., Gibson, M., Lee, J.D., & Mutlu, B. (2016). "Psychophysics of trust in vehicle control algorithms." In *Proceedings of the SAE World Congress*. Detroit, MI. http://papers.sae.org/2016-01-0144/.
19. Muir, B. M., & Moray, N. (1996). Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics* 39 (3): 429–60. doi:10.1080/00140139608964474.
20. Stanton, N. A., & Young, M. S. (1998). Vehicle automation and driving performance. *Ergonomics* 41 (7): 1014–28. doi:10.1080/001401398186568.
21. Victor, T. W., Harbluk, J. L., & Engström, J. A. (2005). Sensitivity of eye-movement measures to in-vehicle task difficulty. *Transportation Research Part F: Traffic Psychology and Behaviour*, 8 (2): 167–90. doi:10.1016/j.trf.2005.04.014.
22. Lee, J. D., Moeckli, J., Brown, T. L., Roberts, S. C., Schwarz, C., Yekhshatyan, L., Nadler, E. et al. (2013). *Distraction Detection and Mitigation Through Driver Feedback*. Final Report DOT HS 811 547A. http://trid.trb.org/view.aspx?id=1251342.
23. Mclean, J.R., & Hoffmann, E.R. (1971). Analysis of drivers' control movements. *Human Factors: The Journal of Human Factors and Ergonomics Society* 13 (5): 407–18. doi:10.1177/001872087101300503.
24. McLean, J. R., & Hoffmann, E. R. (1973). The effects of restricted preview on driver steering control and performance. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 15 (4): 421–30. doi:10.1177/001872087301500413.
25. R Development Core Team. (2009). *{R: A Language and Environment for Statistical Computing}*. Vienna, Austria: R Foundation for Statistical Computing.
26. Sullivan, G. M., & Artino, A. R. (2013). Analyzing and interpreting data from Likert-type scales. *Journal of Graduate Medical Education* 5(4), 541-542. doi: 10.4300/JGME-5-4-18.
27. Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67 (1). doi:10.18637/jss.v067.i01.
28. Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences.

Journal of Memory and Language, Special Issue: Emerging Data Analysis, 59 (4): 475–94. doi:10.1016/j.jml.2007.11.006.

29. Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference understanding AIC and BIC in model selection. *Sociological Methods & Research* 33 (2): 261–304. doi:10.1177/0049124104268644.
30. Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution* 4 (2): 133–42. doi:10.1111/j.2041-210x.2012.00261.x.
31. Johnson, P. C.D. (2014). Extension of Nakagawa & Schielzeth's R2GLMM to random slopes models. *Methods in Ecology and Evolution* 5 (9): 944–46. doi:10.1111/2041-210X.12225.
32. Lee, J. D., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics* 35 (10): 1243–70. doi:10.1080/00140139208967392.