U.S. Department of Transportation
National Highway Traffic Safety Administration

NHTSA
www.nhtsa.gov

DOT HS 812 182

August 2015

# Human Factors Evaluation of Level 2 and Level 3 Automated Driving Concepts

DISCLAIMER



Suggested APA Format Citation:

Blanco, M., Atwood, J., Vasquez, H. M., Trimble, T. E., Fitchett, V. L., Radlbeck, J., ... & Morgan, J. F. (2015, August). *Human factors evaluation of level 2 and level 3 automated driving concepts*. (Report No. DOT HS 812 182). Washington, DC: National Highway Traffic Safety Administration.

**Technical Report Documentation Page**

| 1. Report No. DOT HS 812 182 | 2. Government Accession No. | 3. Recipient's Catalog No. |
|---|---|---|
| 4. Title and Subtitle Human Factors Evaluation of Level 2 and Level 3 Automated Driving Concepts | | 5. Report Date August 2015 |
| | | 6. Performing Organization Code |
| 7. Authors Myra Blanco, Jon Atwood, Holland M. Vasquez, Tammy E. Trimble, Vikki L. Fitchett, Joshua Radlbeck, Gregory M. Fitch, & Sheldon M. Russell (VTTI); Charles A. Green (General Motors); Brian Cullinane (Google Self-Driving Car Project); Justin F. Morgan (Battelle) | | 8. Performing Organization Report No. |
| 9. Performing Organization Name And Address Virginia Tech Transportation Institute<br>Center for Automated Vehicle Systems<br>3500 Transportation Research Plaza<br>Blacksburg, Virginia 24061 | | 10. Work Unit No. (TRAIS) |
| | | 11. Contract or Grant No. DTNH22-11-D-00236, Task Order #11 |
| 12. Sponsoring Agency Name and Address National Highway Traffic Safety Administration<br>Electronic Systems Safety Division, NVS-333<br>1200 New Jersey Avenue SE.<br>Washington, DC 20590 | | 13. Type of Report and Period Covered Final Report, July 2013 – February 2015 |
| | | 14. Sponsoring Agency Code |

**15. Supplementary Notes**
The Contracting Officer's Technical Representative was Dr. Paul Rau.

**16. Abstract**
The purpose of this study was to investigate user interactions with Level 2 and Level 3 partially automated vehicles. L2 and L3 were of interest because it is at this point where the driver's role transitions and longitudinal and lateral control are ceded in varying degrees to the vehicle. At this point, the driver becomes an intermittent operator. For L2 and L3, the level of involvement by the human might vary. Therefore, we use the term operator as opposed to driver. The study focused on how these intermittent operators transition between automated and non-automated vehicle operation, and how this interaction is affected by the human-machine interface (HMI). Three experiments were performed with prototype partially automated vehicles on controlled test tracks in mixed traffic. The first experiment investigated which HMI characteristics are most effective at issuing a take-over request (TOR) during the operation of an L2 automated vehicle. The second experiment investigated how to prompt operators to attend to the road when distracted during the operation of an L2 automated vehicle, and whether these prompts are effective over time. The third experiment investigated which HMI characteristics are most effective at issuing a TOR during the operation of an L3 automated vehicle. The findings suggest that the most effective hand-off strategies were those that incorporated nonvisual components. The driver engagement patterns observed in this study provide data and evidence that could support the future development of human factors design principles for L2 and L3 partially automated vehicles.

| 17. Key Words Automated driving technology, human factors, human-machine interface, levels of vehicle automation, self-driving vehicles, vehicle automation | 18. Distribution Statement Document is available to the public from the National Technical Information Service www.ntis.gov | | |
|---|---|---|---|
| 19. Security Classif. (of this report) | 20. Security Classif. (of this page) | 21. No. of Pages 320 | 22. Price |

# Acknowledgements

The authors would like to extend their gratitude to the members of the stakeholder group for this project: John Campbell from Battelle, Richard Bishop from Bishop Consulting, Darren Liccardo from BMW, Hauke Schmidt and Jan Becker from Bosch, David Agnew from Continental, Charles Green and Jeremy Salinger from General Motors, Brian Cullinane from Google, Luca Delgrossi, Viviane Eide, and Charles Kawashima from Mercedes Benz, Steve Dellenback and Jerry Towler from Southwest Research Institute, Andrew Culhane from TORC Robotics, James Foley and Chuck Gulash from Toyota, and Daniel Brown from Travelers. Thank you for your time and feedback during the research plan and final report review process.

Thanks also to Missy Cummings from Duke University, Thomas B. Sheridan from the Massachusetts Institute of Technology, Peter Hancock from the University of Central Florida, Nadine Sarter from the University of Michigan, and John D. Lee from the University of Wisconsin–Madison for their papers related to the human factors aspects of automated driving.

This research would not have been possible without the assistance of many people that provided their time and expertise, particularly all the personnel at the GM Milford Proving Ground: Sharon Greene, W. D. Smith, Carlos Franca, Vance McCabe, Manohar Reddy, Shilpa Prabhuswamy, and Vinaya Patil. Google personnel were an integral part of this study, including the assistance of Nicolas Raetz Munley, David Kram, Alan Grady, Joseph Blas, and Brian Cullinane.

Thanks also to Kelly Rickli, who assisted with running participants at Milford Proving Ground, Linda Harris, who managed a substantial portion of the recruitment and scheduling, the VTTI Center for Technology Development, which assisted with the instrumentation needed to obtain the level of resolution required for the analyses presented in this document, and to Dr. Eddy Llaneras and Dr. Rich Hanowski, who served as internal reviewers.

# Table of Contents


**Acknowledgements** .......... ii
**Executive Summary** .......... 1

*Experiment 1 – Alerting Operators to Regain Control of an L2 Automated Vehicle* .......... 2
Approach .......... 2
Findings .......... 3

*Experiment 2 – Prompting Operators to Monitor the Road when Operating an L2 Automated Vehicle* .......... 3
Approach .......... 3
Findings .......... 4

*Experiment 3 – Alerting Operators to Regain Control of an L3 Automated Vehicle* .......... 5
Approach .......... 6
Findings .......... 6

*Research Question 1 – How Do Drivers Interact with and Operate Vehicles that Offer Level 2 and Level 3 Automation?* .......... 8

*Research Question 2 – What Are the System Performance Risks from Operator Involvement with, and Interruption from, Secondary Tasks?* .......... 8

*Research Question 3 – What Are the Most Effective Hand-Off Strategies between the System and the Operator?* .......... 9

*Research Question 4 – How Do Operators Engage, Disengage, and Reengage with the Driving Task?* .......... 9

*Research Question 5 – How Do Operators Perform Under Various Operational Concepts within Level 2 and Level 3 Automation?* .......... 10

*Research Question 6 – What Are the Most Effective Human-Machine Interface Concepts which Optimize the Safe Operation of Level 2 and Level 3 Systems?* .......... 10

*Summary* .......... 11

**Chapter 1 Introduction** .......... 12

*Objective* .......... 12

*Rationale* .......... 12

*Background* .......... 12
Levels of Automation .......... 16

*Project Research Questions* .......... 18

**Chapter 2 Level 2 Automated Vehicle, Experiment 1: Alerting Operators to Regain Control of an L2 Automated Vehicle** .......... 21

*Purpose* .......... 21

*Method* ... *21*

*Experimental Design* ... *22*

Independent Variables ... 22
Dependent Variables ... 23
Participants' Subjective Assessment ... 24
Facility ... 24
Vehicle ... 25
Tablet Computer ... 26
Participants ... 27
Instruction and Training ... 27
Study Session ... 28

*Results* ... *32*

Performance ... 32
Time to React to Alert ... 32
Time to Regain Control ... 33
Time to Activate the Automation ... 34
Time to Release Control of Steering ... 35
Time to Resume a Non-Driving Task ... 35
Operator Behavior Analysis Summary ... 35
Eye-glance Behavior ... 36
Trust Scales ... 38

*Experiment 1 Summary* ... *42*

**Chapter 3 Level 2 Automated Vehicle, Experiment 2: System Prompt Effectiveness over Time ... 47**

*Purpose* ... *47*

*Method* ... *47*

*Experimental Design* ... *48*

Independent Variables ... 48
Dependent Variables ... 49
Participants' Subjective Assessment ... 50

*Facility* ... *50*

*Vehicle* ... *51*

Tablet Computer ... 52
Participants ... 52
Instruction and Training ... 52
Study Session ... 53

*Results for Lane Drift Analysis* ... *58*

Operator Behavior ... 58
Operator Behavior Analysis Summary ... 61
Eye-glance Behavior ... 62

*Results for Prompt Analysis* ... *64*

Operator Behavior ... 64

Operator Behavior Analysis Summary ........ 65
Eye-glance Behavior ........ 65

*Results for Trust Scales Analysis ........ 68*

*Experiment 2 Summary ........ 74*

**Chapter 4 Level 3 Automated Vehicle, Experiment 3: Human-Automation System Performance Over Time ........ 78**

*Purpose ........ 78*

*Method ........ 78*

*Experimental Design ........ 78*

Independent Variables ........ 79
Dependent Variables ........ 80

*Participants' Subjective Assessment ........ 82*

*Facility ........ 82*

*Vehicle ........ 83*

Tablet Computer ........ 84
Participants ........ 85
Instruction and Training ........ 85
Study Session ........ 86

*Results ........ 91*

Operator Behavior ........ 91
Eye-glance Behavior ........ 99
Trust Scales ........ 100

*Experiment 3 Summary ........ 104*

**Chapter 5 Discussion ........ 106**

*Main Findings ........ 109*

Research Question 1: How Do Drivers Interact with and Operate Vehicles that Offer Level 2 and Level 3 Automation? ........ 109
Research Question #2: What Are the System Performance Risks from Operator Involvement in Secondary Tasks? ........ 110
Research Question #3: What Are the Most Effective Hand-Off Strategies between the System and the Operator? ........ 110
Research Question #4: How Do Operators Engage, Disengage, and Reengage with the Driving Task in Response to the Various States of Level 2 and Level 3 Automation? ........ 111
Research Question #5: How Do Operators Perform under Various Operational Concepts within Level 2 and Level 3 Automation? ........ 113
Research Question #6: What Are the Most Effective Human-Machine Interface Concepts? ........ 114

*Vehicle Automation Theories ........ 116*

Primary Task Reversal ........ 116
Alert Annoyance Habituation ........ 117

*Conclusion ........ 118*

**References ........................................................................................................ 121**
**Appendix A. List of Acronyms ............................................................ 123**
**Appendix B. Experiment 1: Expanded Methods, Results, and Analysis....... 124**

*Experimental Design ........................................................................................ 124*

*Procedure ........................................................................................................ 124*

Screening ........................................................................................................ 124
Greeting ........................................................................................................ 125
Participants ........................................................................................................ 125

*Results ........................................................................................................ 127*

Operator Behavior ........................................................................................................ 128
Trust ........................................................................................................ 149

**Appendix C. Experiment 2: Expanded Methods, Results, and Analysis....... 154**

*Experimental Design ........................................................................................ 154*

*Procedure ........................................................................................................ 154*

Screening ........................................................................................................ 154
Greeting ........................................................................................................ 155
Participants ........................................................................................................ 155

*Results for Lane Drifts ........................................................................................ 157*

Operator Behavior ........................................................................................................ 158
Eye-glance Behavior ........................................................................................................ 180

*Results for Prompts ........................................................................................ 195*

Operator Behavior ........................................................................................................ 195
Eye-glance Behavior ........................................................................................................ 199
Trust ........................................................................................................ 208

**Appendix D. Experiment 3: Expanded Methods, Results, and Analysis....... 213**

*Experimental Design ........................................................................................ 213*

*Procedure ........................................................................................................ 213*

Screening ........................................................................................................ 213
Greeting ........................................................................................................ 214
Participants ........................................................................................................ 214

*Results ........................................................................................................ 216*

Operator Behavior ........................................................................................................ 217
Operator Behavior Analysis Summary ........................................................................ 245
Eye-glance Behavior ........................................................................................................ 246
Trust ........................................................................................................ 252

**Appendix E. Trust and Satisfaction Scales ........................................................ 258**
**Appendix F. Subjective Data ........................................................................ 261**

*After-experience Interview Analysis ........................................................................ 261*

Purpose ..... 261
Data Analysis Method ..... 261
Survey Instrument ..... 263
*Experiment 1 Analysis Findings ..... 263*
Theme 1: Comfort with System ..... 264
Theme 2: Alert Reactions ..... 269
Theme 3: Design Suggestions ..... 270
Theme 4: Additional Thoughts ..... 272
*Experiment 2 Analysis Findings ..... 272*
Theme 1: Comfort with System ..... 273
Theme 2: Alert Reactions ..... 280
Theme 3: Design Suggestions ..... 283
Theme 4: Additional Thoughts ..... 285
*Experiment 3 Analysis Findings ..... 286*
Theme 1: Comfort with System ..... 286
Theme 2: Alert Reactions ..... 292
Theme 3: Design Suggestions ..... 294
Theme 4: Additional Thoughts ..... 296

## List of Tables


Table ES-1. Mean Times for Responses to Different Alerts and Prompts (units: seconds) ............... 10
Table 1-1. Down-selected List of Automation-relevant Parameters (Marinik, Bishop, Fitchett, Morgan, Trimble, and Blanco, 2014) and Corresponding Representation Within the Experimental Design ........ 14
Table 1-2. Selected Automation Concepts for HMI Evaluation (Marinik, Bishop, Fitchett, Morgan, Trimble, and Blanco, 2014) and Corresponding Representation Within the Experimental Design ........ 15
Table 1-3. Overview of the Three Experiments ........ 20
Table 2-1. Summary Table for Operator Behavior Analysis for Experiment 1 ........ 36
Table 2-2. Mean Values for Cautionary Multimodal, Compared to Mean Values for Other Relevant Alert Type/Alert Modality Combinations in Experiment 1 ........ 43
Table 2-3. Mean Values for Cautionary Unimodal, Compared to Mean Values for Other Relevant Alert Type/Alert Modality Combinations in Experiment 1 ........ 43
Table 2-4. Mean Values for Imminent Multimodal, Compared to Mean Values for Other Relevant Alert Type/Alert Modality Combinations in Experiment 1 ........ 44
Table 2-5. Mean Values for Imminent Unimodal, Compared to Mean Values for Other Relevant Alert Type/Alert Modality Combinations in Experiment 1 ........ 44
Table 2-6. Mean Values for Staged Multimodal, Compared to Mean Values for Other Relevant Alert Type/Alert Modality Combinations in Experiment 1 ........ 45
Table 2-7. Mean Values for Staged Unimodal, Compared to Mean Values for Other Relevant Alert Type/Alert Modality Combinations in Experiment 1 ........ 45
Table 2-8. Mean Values for Time to Activate the Automation, Compared Between Relevant Alert Type/Alert Modality Combinations in Experiment 1 ........ 46
Table 3-1. Summary Table for Operator Behavior Analysis for Experiment 2 ........ 62
Table 3-2. Summary Table for Eye-glance Behavior Analysis for Experiment 2 ........ 66
Table 3-3. Mean Values for Alert with the 2-second Prompt Condition, Compared to Mean Values for Other Relevant Event Type/Prompt Condition Combinations in Experiment 2 ........ 74
Table 3-4. Mean Values for Alert with the 7-second Prompt Condition, Compared to Mean Values for Other Relevant Event Type/Prompt Condition Combinations in Experiment 2 ........ 75
Table 3-5. Mean Values for Alert with the No Prompts Condition, Compared to Mean Values for Other Relevant Event Type/Prompt Condition Combinations in Experiment 2 ........ 75
Table 3-6. Mean Values for No Alert with the 2-second Prompt Condition, Compared to Mean Values for Other Relevant Event Type/Prompt Condition Combinations in Experiment 2 ........ 76
Table 3-7. Mean Values for No Alert with the 7-second Prompt Condition, Compared to Mean Values for Other Relevant Event Type/Prompt Condition Combinations in Experiment 2 ........ 77
Table 3-8. Mean Values for No Alert with the No Prompts Condition, Compared to Mean Values for Other Relevant Event Type/Prompt Condition Combinations in Experiment 2 ........ 77
Table 4-1. Operational Definitions for Experiment 3 Eye-glance-related Terms ........ 82
Table 4-2. Summary Table for Operator Behavior Analysis for Experiment 3 ........ 99
Table 4-3. Mean Values for the Three Alerts for the Dependent Variables in Experiment 3 ........ 105
Table 5-1. Main Findings ........ 107

Table 5-2. Mean Times for Responses to Different Alerts and Prompts (units: seconds) ............... 112
Table 5-3. Selected Automation Concepts for Human-Machine Interface Evaluation (Marinik, Bishop, Fitchett, Morgan, Trimble, and Blanco, 2014) and Corresponding Representation Within the Experimental Design ........ 113
Table B-1. Time to React, Analysis of Variance (ANOVA) Table for Experiment 1 ........ 131
Table B-2. Time to React Differences in Least Squares Means, Modality for Experiment 1 ........ 132
Table B-3. Time to Regain Control, ANOVA Table for Experiment 1 ........ 135
Table B-4. Significant Differences of the Log of Time to Regain Control Within Alert Modality and Alert Type for Experiment 1 ........ 136
Table B-5. Difference in Least Squares Means of the Log of Time to Regain Control Between Multimodal and Unimodal Alerts, Performed Within Cautionary and Staged Alerts, Respectively ........ 137
Table B-6. Significant Differences in Least Squares Means of the Log of Time to Regain Control Between Alert Types Within Unimodal Alerts ........ 137
Table B-7. Time to Activate the Automation, ANOVA Table for Experiment 1 ........ 139
Table B-8. Time to Activate the Automation Differences in Least Squares Means, Alert Type by Alert Modality for Experiment 1 ........ 140
Table B-9. Time to Release Control of Steering, ANOVA Table for Experiment 1 ........ 143
Table B-10. Significant Differences of the Log of Time to Release Control Within Alert Modalities and Alert Types for Experiment 1 ........ 144
Table B-11. Difference in Least Squares Means Between Multimodal and Unimodal Alerts, Within Staged Alerts, For Experiment 1 ........ 145
Table B-12. Significant Differences in Least Squares Means Between Alert Types, Within Multimodal Alerts, For Experiment 1 ........ 145
Table B-13. Time to Resume a Non-driving Task, ANOVA Table for Experiment 1 ........ 147
Table B-14. Time to Resume Non-driving Task Differences in Least Squares Means, Alert Type for Experiment 1 ........ 148
Table B-15. Summary of Trust Analysis for Experiment 1 ........ 150
Table B-16. Trust Model, Experiment 1 ........ 152
Table B-17. Descriptive Statistics for Responses to the After-experience Trust Scales for Experiment 1 ........ 153
Table C-1. Time to React, ANOVA Table for Experiment 2 ........ 162
Table C-2. Comparison of Least Squares Means of the Log of Time to React Between Event Types, Averaged Across Levels of Prompt Condition and Session ........ 164
Table C-3. Difference in Least Squares Means of the Log of Time to React Between Alerts and No Alerts ........ 164
Table C-4. Tests for Significance of Prompt Condition of the Log of Time to React Within Combinations of Sessions 1 and 2 and Event Types ........ 165
Table C-5. Comparison of Least Squares Means of the Log of Time to React Between Prompt Conditions Within Session 1, No Alert Lane Drifts ........ 165
Table C-6. Probability of Regaining Control Without Instruction for Experiment 2 ........ 169
Table C-7. Analysis for the Log of Time to Regain Control, ANOVA Table ........ 169
Table C-8. Analysis for the Log of Time to Activate the Automation, ANOVA Table for Experiment 2 ........ 172
Table C-9. Significant Differences of Time to Activate the Automation Within Session and Event Type for Experiment 2 ........ 174

Table C-10. Difference in the Least Squares Means of the Log of Time to Activate the Automation Between Alert and No Alert, Within Session 1 ... 174
Table C-11. Time to Release Control of Steering, Tests for Significance for Experiment 2 ... 177
Table C-12. Analysis of Time to Resume a Non-driving Task, ANOVA Table, For Experiment 2 ... 179
Table C-13. Analysis of Monitoring Rate, ANOVA Table, for Experiment 2 ... 185
Table C-14. Tests for Significance of the Change in Log Odds of Monitoring the Roadway from Before to After the Lane Drift, For Alerts and No Alerts ... 186
Table C-15. Differences in Least Squares Means of the Log Odds of Driving-related Glancing Before an Alert to After an Alert ... 187
Table C-16. Differences Between Prompt Conditions in the Change in the Log Odds of Driving-related Glancing from Before to After the Alert ... 188
Table C-17. Non-driving-related Glance Analysis, Lane Drifts, ANOVA Table, for Experiment 2 ... 191
Table C-18. Tests for Significant Effects of Time Within Event Type Levels, and Event Type Within Time Levels, for Experiment 2 ... 192
Table C-19. Differences in Least Squares Means of the Log Rate of Non-driving-related Glances, Event Type*Time ... 193
Table C-20. Tests for Significance Within Different Levels of Prompt Condition/Session ... 194
Table C-21. Differences in Least Squares Means of the Log Rate of Non-driving-related Glances, Prompt Condition*Session ... 195
Table C-22. Analysis of the Time to React to Prompts, ANOVA Table, for Experiment 2 ... 197
Table C-23. Significant Differences in Least Squares Means of the Log of Time to React to Prompts Across Sessions ... 198
Table C-24. Significant Differences in Least Squares Means of the Log of Time to React to Prompt Between Sessions ... 199
Table C-25. Analysis of Time to Regain Control, Prompts, ANOVA Table, for Experiment 2 ... 199
Table C-26. Analysis of Monitoring Rate, ANOVA Table, for Experiment 2 ... 201
Table C-27. Tests for Significance Within Different Levels of Prompt Condition for Experiment 2 ... 201
Table C-28. Significant Differences in Least Squares Means for the Log Odds of Monitoring the Roadway, Between Prompt Conditions, Significant at Bonferroni Adjusted *P* Values for Experiment 2 ... 202
Table C-29. Significant Differences in Least Squares Means, for Prompt Condition and Time, Significant at Bonferroni Adjusted *P* Values for Experiment 2 ... 203
Table C-30. Analysis of Non-driving-related Glances Around Prompts, ANOVA Table, For Experiment 2 ... 205
Table C-31. Tests for Significance Within Different Levels of Prompt Condition and Time for Experiment 2 ... 206
Table C-32. Significant Differences in Least Squares Means, Before and After the 7-second Condition Prompt, Significant at Bonferroni Adjusted *P* Values for Experiment 2 ... 207
Table C-33. Significant Differences in Least Squares Means Between Prompt Conditions, Both Before and After the Prompt, Significant at Bonferroni Adjusted *P* Values for Experiment 2 ... 207
Table C-34. Analysis of Trust over Time, ANOVA Table, for Experiment 2 ... 209

Table C-35. Significant Differences in Least Squares Means of Trust Change Across Sessions ... 209
Table C-36. Significant Differences in Trust Across Conditions ... 209
Table C-37. Descriptive Statistics for Responses to the After-experience Trust Scales for Experiment 2 ... 210
Table C-38. Descriptive Statistics for Responses to the Satisfaction Scale Survey for Experiment 2 ... 211
Table D-1. Time to React, ANOVA Table for Experiment 3 ... 224
Table D-2. Significant Differences of Time to React Within Alert Type and Looking Forward for Experiment 3 ... 225
Table D-3. Difference in Time to React Between Forward-Looking and Not-Forward-Looking Group, Within Staged Alerts, For Experiment 3 ... 225
Table D-4. Significant Differences in Reaction Times Between Alert Types, Within Forward-Looking Instances, for Experiment 3 ... 226
Table D-5. Time to Regain Control, ANOVA Table for Experiment 3 ... 229
Table D-6. Time to Regain Control Differences in Least Squares Means, Alert Type, Significant at Bonferroni Adjusted *P* Values for Experiment 3 ... 230
Table D-7. Analysis for the Log of Time to Activate the Automation, ANOVA Table for Experiment 3 ... 236
Table D-8. Analysis for the Log of Time to Release Control of Steering, ANOVA Table ... 239
Table D-9. Log of Time to Release Control of Steering Differences in Least Squares Means, Alert Type, Significant at Bonferroni Adjusted *P* Values, for Experiment 3 ... 241
Table D-10. Missing Times to Resume Non-driving Task for Experiment 3 ... 245
Table D-11. Analysis of Time to Resume Non-driving Task, ANOVA Table for Experiment 3 ... 245
Table D-12. Summary Table for Eye-glance Behavior Analysis for Experiment 3 ... 246
Table D-13. Friedman Test Results, Differences in Time Point Mean Percentage Driving-related Glance Time, by Alert Type for Experiment 3 ... 248
Table D-14. Non-driving-related Glance Model, Tests for Significance for Experiment 3 ... 250
Table D-15. Differences in Least Squares Means, Bonferroni Adjustment, Alert Type*Time for Experiment 3 ... 251
Table D-16. Differences in Least Squares Means, Bonferroni Adjustment, Non-driving Task for Experiment 3 ... 251
Table D-17. Summary Table for Trust Analysis for Experiment 3 ... 253
Table D-18. Analysis of "7" Rating Change over Time, Coefficient Estimates for Experiment 3 ... 255
Table D-19. Analysis of "7" Rating, Tests for Effect Significance for Experiment 3 ... 255
Table D-20. Differences in Least Squares Means of Log Odds of "7" Rating for Alert Types for Experiment 3 ... 256
Table D-21. Differences in Least Squares Means of Log Odds of "7" Rating for Sessions in Experiment 3 ... 256
Table D-22. Descriptive Statistics for Responses to the After-experience Trust Scales for Experiment 3 ... 256
Table F-1. Categories Within Subtheme 1.1 General Reactions ... 265
Table F-2. Experiment 1 Participants' Self-reported Comfort Levels ... 265
Table F-3. Categories Within Experiment 2 Subtheme 1.1 General Reactions ... 274
Table F-4. Experiment 2 Participants' Self-reported Comfort Levels ... 276

Table F-5. Examples of Comments Associated with Experiment 2 Participant's Self-reported Levels of Comfort with the Automated System....276
Table F-6. Categories Within Subtheme 1.1 General Reactions....287
Table F-7. Example Self-reported Levels of Comfort with the Automated System for Experiment 3 Participants....287
Table F-8. Categories Within Subtheme 1.1 General Reactions for Experiment 3....293

## List of Figures

Figure ES-1. Participants' Monitoring Rate Before and After Receiving 2-second Prompts, 7-second Prompts, or No Prompts ..... 5
Figure ES-2. Time to Regain Control of an L3 Automated Vehicle when Staged Alerts Are Presented ..... 7
Figure 1-1. The NHTSA Automation Levels ..... 16
Figure 2-1. Sequence of the Experiment 1 Dependent Variables Pertaining to Operator Behavior ..... 24
Figure 2-2. Aerial View of GM's Milford Proving Ground ..... 25
Figure 2-3. 2009 Chevrolet Malibu with a Prototype L2 System ..... 25
Figure 2-4. GM's Data Acquisition System (DAS) Camera Views for Experiment 1 ..... 26
Figure 2-5. Asus Nexus 7 Tablet Computer ..... 27
Figure 2-6. Cautionary, Imminent, and Staged Take Control Alert Timelines for Experiment 1 ..... 29
Figure 2-7. Experimental Session Timeline for Experiment 1 ..... 31
Figure 2-8. Mean and Standard Error Bar Plots of Time to React by Modality for Experiment 1 ..... 33
Figure 2-9. Mean and Standard Error Bar Plots of Time to Regain Control, Stratified by Alert Modality and Alert Type for Experiment 1 ..... 34
Figure 2-10. Mean and Standard Error Plots of Time to Look Forward by Modality for Experiment 1 ..... 37
Figure 2-11. Mean and Standard Error Bar Plots for Operator Variables After Lane Drift Events in Experiment 1 ..... 38
Figure 2-12. Change in Mean Trust Rating Over Time for Experiment 1 ..... 39
Figure 2-13. Mean Rating for After-experience Trust Scales for Experiment 1 ..... 40
Figure 2-14. Experiment 1 Participants' Self-reported Level of Comfort with the Automated System ..... 41
Figure 2-15. Experiment 1 Participants' Time to Achieve Comfort with the System by Age ..... 42
Figure 3-1. Sequence of the Experiment 2 Dependent Variables Pertaining to Operator Behavior ..... 50
Figure 3-2. 2010 Cadillac SRX with a Prototype L2 System Used in Experiment 2 ..... 51
Figure 3-3. VTTI's DAS Camera Views for Experiment 2 ..... 52
Figure 3-4. Alert Timelines for 2-second & 7-second Prompt Conditions in Experiment 2 ..... 55
Figure 3-5. Experimental Session Timeline for Experiment 2 ..... 57
Figure 3-6. Mean and Standard Error Bar Plots, Time to React, by Event Type for Experiment 2 ..... 59
Figure 3-7. Bar Plots of Performance by Alert Type for Experiment 2 ..... 60
Figure 3-8. Mean and Error Bar Plots of Monitoring Rate Immediately Before and After Lane Drifts with Alerts, Stratified by Prompt Condition ..... 63
Figure 3-9. Frequency of Regaining Control After Prompt, by Prompt Stage ..... 65
Figure 3-10. Percentage of Driving-related Glance Time for Participants Engaged in a Non-driving Task at Prompt Start or Selected Time Point ..... 67
Figure 3-11. Time Trend of Trust, Stratified by Prompt Conditions ..... 68
Figure 3-12. Mean and Standard Error Bar Plots of Trust Rating Before and After the Lane Drift in Experiment 2, Stratified by Alert Type ..... 69
Figure 3-13. Mean Post-session Trust Scale Rating for 2-second Prompt Condition ..... 70
Figure 3-14. Mean After-session Trust Scale Rating for 7-second Prompt Condition ..... 70

Figure 3-15. Mean Rating for Satisfaction Survey for Experiment 2 .......... 71
Figure 3-16. Experiment 2 Participants' Self-reported Level of Comfort with the Automated System .......... 72
Figure 3-17. Experiment 2 Participants' Time to Achieve Comfort with the System by Age .......... 73
Figure 3-18. Experiment 2 Participants' Time to Achieve Comfort with System by Prompt Condition .......... 73
Figure 4-1. Sequence of the Dependent Variables Pertaining to Operator Behavior for Experiment 3 .......... 81
Figure 4-2. Aerial View of the Virginia Smart Road Test Track at VTTI .......... 83
Figure 4-3. The 2012 Lexus RX450h Used in Experiment 3 .......... 83
Figure 4-4. VTTI's DAS Camera Views from Experiment 3 .......... 84
Figure 4-5. Asus Nexus 7 Tablet Computer .......... 85
Figure 4-6. Staged and Imminent Take Control Alert Timelines for Experiment 3 .......... 87
Figure 4-7. Take Control Alert Locations on the Virginia Smart Road Test Track .......... 88
Figure 4-8. Example Traffic Configurations from Experiment 3 .......... 89
Figure 4-9. Experimental Session Timeline for Experiment 3 .......... 90
Figure 4-10. Time to React to Alert, Stratified by Alert Type, After the Onset of the Alert .......... 92
Figure 4-11. Time to React to Alert, Stratified by Alert Type, Relative to the Onset of the Imminent Phase .......... 93
Figure 4-12. Time to Regain Control, Stratified by Alert Type, After the Onset of the Alert .......... 94
Figure 4-13. Time to Regain Control, Stratified by Alert Type, Relative to the Onset of the Imminent Phase .......... 95
Figure 4-14. Mean and Error Bar Plots, Operator Behavior Variables, Within Imminent Alerts .......... 98
Figure 4-15. Mean and Error Bar Plots, Operator Behavior Variables, Within Staged Alerts .......... 98
Figure 4-16. Mean and Standard Error Bar Plot of Percentage of Driving-related Glance Time, Stratified by Alert Type and Time Interval Before/After the Alert for Experiment 3 .......... 100
Figure 4-17. Proportion of "7" Ratings by Time for Experiment 3 .......... 101
Figure 4-18. Mean and Standard Error Bar Plot of Experiment 3 Mean After-experience Trust Ratings by Statement .......... 102
Figure 4-19. Self-reported Levels of Comfort with the Automated System for Experiment 3 .......... 103
Figure 4-20. Time to Achieve Comfort with the System by Age for Experiment 3 .......... 103
Figure 5-1. Reaction Times for Multimodal Imminent Alerts for Level 2 & 3 Automation .......... 119
Figure 5-2. Reaction Time Percentile for Multimodal Imminent Alerts for Level 2 & 3 Automation .......... 119
Figure B-1. Distribution of Experiment 1 Participants Across the NHTSA Age Categories .......... 125
Figure B-2. Experiment 1 Participants' Experience with Smartphones .......... 126
Figure B-3. Experiment 1 Participants' Experience with and Awareness of Automated Technology for Vehicles .......... 127
Figure B-4. Mean and Standard Error Bar Plots of Time to React by Alert for Experiment 1 .......... 128
Figure B-5. Boxplots of Time to React by Modality, Alert for Experiment 1 .......... 129
Figure B-6. Bar Plot of Frequencies of Not Looking Forward/Looking Forward, by Alert Modality*Alert Type .......... 130
Figure B-7. Mean and Standard Error Bar Plot of Time to React, by Alert Modality*Alert Type Group and Looking Forward .......... 131
Figure B-8. Plot of Least Squares Means for Time to React with Standard Error Bars, Modality for Experiment 1 .......... 132

Figure B-9. Mean and Standard Error Bar Plots of Time to Regain Control by Modality for Experiment 1 ........ 133
Figure B-10. Mean and Standard Error Bar Plots of Time to Regain Control by Alert for Experiment 1 ........ 134
Figure B-11. Boxplots of Time to Regain Control by Modality, Alert for Experiment 1 ........ 135
Figure B-12. Plot of Least Squares Means for Time to Regain Control with Standard Error Bars, Alert Type by Alert Modality for Experiment 1 ........ 136
Figure B-13. Mean and Standard Error Bar Plots of Time to Activate the Automation, Alert/Modality for Experiment 1 ........ 138
Figure B-14. Boxplots of Time to Activate the Automation by Modality, Alert for Experiment 1 ........ 139
Figure B-15. Plot of Least Squares Means for Time to Activate the Automation with Standard Error Bars, Alert/Modality for Experiment 1 ........ 140
Figure B-16. Mean and Standard Error Bar Plots of Time to Release Control of the Steering, by Alert Modality for Experiment 1 ........ 141
Figure B-17. Mean and Standard Error Bar Plots of Time to Release Control of the Steering, by Alert Type for Experiment 1 ........ 142
Figure B-18. Boxplots of Time to Release Control of Steering by Modality, Alert for Experiment 1 ........ 143
Figure B-19. Plot of Least Squares Means for Time to Release Control of Steering with Standard Error Bars, Alert/Modality for Experiment 1 ........ 144
Figure B-20. Mean and Standard Error Bar Plots of Time to Resume Non-driving Task, Modality for Experiment 1 ........ 146
Figure B-21. Mean and Standard Error Bar Plots of Time to Resume Non-driving Task, Alert for Experiment 1 ........ 146
Figure B-22. Boxplots of Time to Resume a Non-driving Task, Modality and Alert for Experiment 1 ........ 147
Figure B-23. Plot of Least Squares Means for Time to Resume Non-driving Task with Standard Error Bars, Alert Type for Experiment 1 ........ 148
Figure B-24. Bar Plot of Experiment 1 Trust Ratings Across Full Experiment ........ 151
Figure B-25. Subject Time Trends in Trust for Experiment 1 ........ 152
Figure C-1. Distribution of Experiment 2 Participants Across the NHTSA Age Categories ........ 155
Figure C-2. Experiment 2 Participants' Experience with Smartphones ........ 156
Figure C-3. Experiment 2 Participants' Experience with and Awareness of Automated Technology for Vehicles ........ 157
Figure C-4. Mean and Standard Error Bar Plots, Time to React, by Prompt Condition for Experiment 2 ........ 159
Figure C-5. Boxplots of Time to React, Stratified by Event Type and Prompt Condition for Experiment 2 ........ 160
Figure C-6. Mean and Standard Error Bar Plots, Time to React, Not Looking Forward, by Alert Type ........ 161
Figure C-7. Boxplots of Time to React, Stratified by Event Type, Prompt Condition, and Session for Experiment 2 ........ 163
Figure C-8. Stacked Bar Plot for Correct Responses (Regained Control Without Instruction) Stratified by Session and Prompt Condition ........ 166
Figure C-9. Stacked Bar Plot for Incorrect Responses (Did Not Regain Control Without Instruction from the Experimenter) Stratified by Session and Prompt Condition ........ 167

Figure C-10. Mean and Standard Error Bar Plots, Time to Regain Control During Alerts, by Prompt Condition for Experiment 2 .......... 168
Figure C-11. Boxplots of Time to Regain Control During the Alert Event Type, Stratified by Driving Session and Prompt Condition for Experiment 2 .......... 169
Figure C-12. Mean and Standard Error Bar Plots, Time to Activate Automation, by Alert Type for Experiment 2 .......... 170
Figure C-13. Mean and Standard Error Bar Plots, Time to Activate Automation, by Prompt Condition for Experiment 2 .......... 171
Figure C-14. Boxplots of Time to Activate the Automation, Stratified by Event Type and Prompt Condition for Experiment 2 .......... 172
Figure C-15. Plot of Least Squares Means for Time to Activate the Automation for Experiment 2 .......... 173
Figure C-16. Mean and Standard Error Bar Plots, Time to Release Control of Steering by Event Type for Experiment 2 .......... 175
Figure C-17. Mean and Standard Error Bar Plots, Time to Release Control of Steering by Prompt Condition for Experiment 2 .......... 175
Figure C-18. Boxplots of Time to Release Control of Steering, Stratified by Event Type and Prompt Condition for Experiment 2 .......... 176
Figure C-19. Mean and Standard Error Bar Plots, Time to Resume Non-driving Task by Event Type for Experiment 2 .......... 178
Figure C-20. Mean and Standard Error Bar Plots, Time to Resume Non-driving Task by Prompt Condition for Experiment 2 .......... 178
Figure C-21. Boxplots of Time to Resume a Non-driving Task Stratified by Event Type and Prompt Condition for Experiment 2 .......... 179
Figure C-22. Eye-glance Data Calculation Periods for Experiment 2 .......... 181
Figure C-23. Mean and Standard Error Bar Plot of Percentage of Driving-related Glance Time, Stratified by Event Type and Time Interval Before/After the Lane Drift for Experiment 2 .......... 182
Figure C-24. Mean and Standard Error Bar Plot Percentage of Driving-related Glance Time, Stratified by Prompt Condition and Time Interval Before/After the Alert for Experiment 2 .......... 183
Figure C-25. Boxplots of Monitoring, Stratified by Event Type, Condition, and Time (Before/After the Lane Drift) .......... 184
Figure C-26. Least Squares Means of the Log Odds of Driving-related Glancing, Instances Following an Alert, Stratified by Time and Prompt Condition .......... 186
Figure C-27. Least Squares Means of the Log Odds of Driving-related Glancing, Instances Without an Alert, Stratified by Time and Prompt Condition .......... 187
Figure C-28. Boxplots of the Rate of Non-driving-related Glances Before and After the Lane Drift, Stratified by Alert Type and Prompt Condition .......... 190
Figure C-29. Least Squares Means of the Log of the Rate of Non-driving-related Glances Before and After the Lane Drift, Stratified by Event Type .......... 192
Figure C-30. Least Squares Means of the Log of Non-driving-related Glances Across Sessions, Stratified by Prompt Condition .......... 194
Figure C-31. Mean and Error Bar Plots of Time to React to Prompt, by Prompt Condition .......... 196
Figure C-32. Mean and Error Bar Plots of Time to React to Prompts, by Session .......... 197
Figure C-33. Boxplots of Time to React to the Prompts, Stratified by Session and Prompt Condition .......... 197

Figure C-34. Least Squares Means of the Log of Time to React to Prompts Across Sessions ...... 198
Figure C-35. Boxplots of Monitoring Rate Around Prompts, Stratified by Prompt Condition, and Time ........ 200
Figure C-36. Least Squares Means of the Log Odds of Monitoring the Roadway, Stratified by Prompt Condition and Before/After the Prompt ........ 201
Figure C-37. Boxplots of Non-driving-related Glances, Stratified by Prompt Condition and Time (Before/After the Prompt or Randomly Selected Time Point) ........ 204
Figure C-38. Least Squares Means of the Log of Non-driving-related Glances Across Sessions ........ 205
Figure C-39. Least Squares Means of the Log of Non-driving-related Glances Before and After the Prompt, Across Prompt Conditions ........ 206
Figure C-40. Frequency of Rating Levels for Experiment 2 ........ 208
Figure D-1. Distribution of Experiment 3 Participants Across the NHTSA Age Categories ........ 214
Figure D-2. Experiment 3 Participants' Experience with Smartphones ........ 215
Figure D-3. Experiment 3 Participants' Experience with and Awareness of Automated Technology for Vehicles ........ 216
Figure D-4. Mean and Standard Error Bar Plots, Time to React, by Session for Experiment 3 ...... 218
Figure D-5. Boxplots of Time to React Stratified by Session, and Alert Type for Experiment 3 ....... 219
Figure D-6. Non-driving Task Type Observed at the Alert Onset for Experiment 3 ........ 220
Figure D-7. Mean and Standard Error Bar Plots, Time to React, by Non-driving Task Engagement for Experiment 3 ........ 221
Figure D-8. Mean and Standard Error Bar Plots, Time to React, Not Looking Forward, by Alert Type for Experiment 3 ........ 222
Figure D-9. Action Sequence for Participants Who Were Looking Forward at the Time of the Alert for Experiment 3 ........ 223
Figure D-10. Action Sequence for Participants Who Were Not Looking Forward at the Time of the Alert for Experiment 3 ........ 223
Figure D-11. Plots of Least Squares Means for Reaction Time with Standard Error Bars, Combinations of Alert Type and Looking Forward Status for Experiment 3 ........ 224
Figure D-12. Mean and Standard Error Bar Plots, Time to Regain Control, by Session for Experiment 3 ........ 227
Figure D-13. Boxplots of Time to Regain Control, Stratified by Session and Alert Type for Experiment 3 ........ 228
Figure D-14. Plot of Least Squares Means for Time to Regain Control with 95 Percent Confidence Bands, Alert Type for Experiment 3 ........ 229
Figure D-15. Plot of Times to Regain Control Through the Phases of a Staged Alert for Experiment 3 ........ 231
Figure D-16. Method Used to Regain Control During the Staged Alert in Experiment 3 ........ 232
Figure D-17. Method Used to Regain Control During the Imminent Warning due to a Simulated Internal System Failure (i.e., Imminent–No External Threat; n = 25) for Experiment 3 ........ 233
Figure D-18. Method Used to Regain Control During the Imminent Warning due to an Object Present in the Roadway (i.e., Imminent–External Threat) for Experiment 3 ........ 233
Figure D-19. Mean and Standard Error Bar Plots, Time to Activate the Automation, by Alert Type for Experiment 3 ........ 234

Figure D-20. Mean and Standard Error Bar Plots, Time to Activate the Automation, by Session for Experiment 3 ........ 235
Figure D-21. Boxplots of Time to Activate the Automation, Stratified by Session and Alert Type for Experiment 3 ........ 236
Figure D-22. Mean and Standard Error Bar Plot, Time to Release Control of Steering, by Alert Type for Experiment 3 ........ 237
Figure D-23. Mean and Standard Error Bar Plots, Time to Release Control of Steering, by Session for Experiment 3 ........ 238
Figure D-24. Boxplots of Time to Release Control of Steering, Stratified by Session and Alert Type for Experiment 3 ........ 239
Figure D-25. Least Squares Means of the Log of Time to Release Control of Steering with 95 Percent Confidence Bands, Alert Type for Experiment 3 ........ 240
Figure D-26. Non-driving-task Type Observed After Experiencing the Alert for Experiment 3 ........ 242
Figure D-27. Mean and Standard Error Bar Plots, Time to Resume Non-driving Task, by Alert Type for Experiment 3 ........ 243
Figure D-28. Mean and Standard Error Bar Plots for Time to Resume Non-driving Task, Session for Experiment 3 ........ 244
Figure D-29. Boxplots of Time to Resume Non-driving Task Stratified by Session and Alert Type for Experiment 3 ........ 244
Figure D-30. Eye-glance Data Calculation Periods for Experiment 3 ........ 247
Figure D-31. Boxplots of Percentage of Driving-related Glance Time Stratified by Alert and Time Interval for Experiment 3 ........ 248
Figure D-32. Boxplots of Rates of Non-driving-related Glances per Second, Alert Type and Time Interval for Experiment 3 ........ 249
Figure D-33. Frequency of Rating Levels for Experiment 3 ........ 253
Figure D-34. Time Trends in Trust Rating for Individuals for Experiment 3 ........ 254
Figure F-1. Illustration of Framework Analysis. Adapted from "The Framework Approach to Qualitative Data Analysis," by The National Centre for Social Research (2012) ........ 263
Figure F-2. Experiment 1 Participants' First Thoughts and Actions when Provided a Message ........ 270
Figure F-3. Experiment 2 Participants' First Thoughts and Actions when Provided a Message ........ 283
Figure F-4. Experiment 3 Participants' First Thoughts and Actions when Provided a Message ........ 294

# Executive Summary

Vehicles with increasing levels of automation have the increasing ability to manage safety-critical functions such as steering, accelerator controls, and braking once the driver decides to cede control. It could be hypothesized that at the point automation takes control of the vehicle that the driver becomes an intermittent operator (i.e., the driver is an operator that must be constantly involved with both the longitudinal and lateral control of the vehicle). For Level 2 and Level 3 automation, the driver's role transitions and longitudinal and lateral control is ceded in varying degrees to the vehicle. At this point, the driver becomes an intermittent operator (i.e., the driver is an operator that only sometimes assumes longitudinal and lateral control of the vehicle). Therefore, we use the term operator as opposed to driver. At the commencement of this project, a wide variety of automated vehicle technologies and concepts were beginning to emerge. The National Highway Traffic Safety Administration describes these technologies using five levels of automation. These levels describe a continuum of vehicle control that ranges from vehicles that do not have any of their control systems automated (Level 0) through fully automated vehicles (Level 4). The purpose of this study was to investigate how operators interact with partial automation under Levels 2 and 3. Levels 2 and 3 were of interest because it is at this point where the driver's role transitions from one of driving to one of operating (i.e., the driver shares control with the vehicle to the largest degree). The definitions for Levels 2 and 3 (from NHTSA, 2013) are presented below.

***Level 2 (L2) Combined Function Automation***

Level 2 "involves automation of at least two primary control functions designed to work in unison to relieve the driver of control of those functions. Vehicles at this level of automation can utilize shared authority when the driver cedes active primary control in certain limited driving situations. The driver is still responsible for monitoring the roadway and safe operation and is expected to be available for control at all times and on short notice. The system can relinquish control with no advance warning and the driver must be ready to control the vehicle safely. An example of combined functions enabling a Level 2 system is adaptive cruise control in combination with lane centering. The major distinction between level 1 and level 2 is that, at level 2 in the specific operating conditions for which the system is designed, an automated operating mode is enabled such that the driver is disengaged from physically operating the vehicle by having his or her hands off the steering wheel AND foot off pedal at the same time."

***Level 3 (L3) Limited Self-Driving Automation***

Vehicles with Level 3 automation "enable the driver to cede full control of all safety-critical functions under certain traffic or environmental conditions and in those conditions to rely heavily on the vehicle to monitor for changes in the conditions requiring transition back to driver control. The driver is expected to be available for occasional control, but with sufficiently comfortable transition time. The vehicle is designed to ensure safe operation during the automated driving mode. An example would be an automated or self-driving car that can determine when the system is no longer able to support automation, such as from an oncoming construction area, and then signals to the driver to reengage in the driving task, providing the driver with an appropriate amount of transition time to safely regain manual control. The major distinction between level 2 and level 3 is that at level 3, the vehicle is designed so that the driver is not expected to constantly monitor the roadway while driving."

This study focused on how drivers, or rather operators, transition between automated and non-automated vehicle operation, and how this interaction is affected by the human-machine interface (HMI). Three human factors experiments were performed with prototype systems for partially automated vehicles on controlled test tracks to generate answers to six research questions posed by NHTSA. The three experiments are described below and are followed by the study's primary findings.

Each study included a familiarization process during which participants were given basic instructions for interacting with the system. As part of those instructions, participants were informed as to how they would know the system was ready for the automation to be activated. All three prototype systems clearly depicted when the system was ready for activation of the automation.

# Experiment 1 – Alerting Operators to Regain Control of an L2 Automated Vehicle

There are numerous ways to notify operators that they need to regain control of a partially automated vehicle. The purpose of this experiment was to investigate what HMI characteristics are most effective at issuing a Take-Over Request (TOR) and to identify the transition times between the operator and the automated functions in regard to the driving tasks.

## Approach

Twenty-five participants drove an L2 automated vehicle on a controlled test track for 90 minutes. Participants were asked to perform several visually intensive non-driving tasks at the command of an in-vehicle experimenter. (Non-driving tasks are those tasks that do not involve the vehicle's controls. Participants were handed cards that provided tasks such as: send an email to a specific address and ask a particular question, enter a specific number-street-city-state address as noted on the card, or access a certain Internet site and answer a question based on the information from that site.) They were also alerted to regain control of the automated vehicle throughout the experiment. Six types of alerts were generated that varied in modality and severity. The modality could be either a visual alert (unimodal) or a combined visual and haptic seat alert (multimodal). The severity could be either a Cautionary alert, an Imminent alert, or a Staged alert. (Staged alerts comprised two phases. In this experiment, Staged alerts were ones in which a cautionary alert phase was followed by an imminent alert phase; thus, the alert increased in severity as time went on.) The six types of alerts were presented three times each for a total of 18 times (i.e., three repetitions of the three different types of alerts, with each alert being presented in two modalities, visual only and visual plus haptic). The alerts were generated in the absence of any unexpected events. At the end of the study, however, the in-vehicle experimenter triggered the vehicle to unexpectedly drift out of the lane. Participants received a visual and haptic Imminent alert as the lane drift began. This study recorded characteristics of their responses to the alert (e.g., time to react, time to regain control).

## Findings

The alert modality significantly affected the time taken to regain control of the vehicle. Participants regained control of the vehicle significantly faster when the visual alert included a haptic component (mean time of 1.3 s [S.E. = 0.1 s]) compared to when the alert was just visual (mean time of 4.8 s [S.E. = 0.5 s]). This was likely because the nature of the haptic component enabled participants to perceive the alert when involved in a non-driving task. This is further supported by the finding that participants were just as fast to regain control of the vehicle when alerted by Cautionary alerts as they were with Imminent alerts if both alerts included a haptic component.

Participants regained control sooner when alerted by an Imminent visual alert comprising a red light-emitting diode (LED; mean time of 2.9 s [S.E. = 0.4s]) compared to a Cautionary visual alert comprising a yellow LED (mean time of 6.3 s [S.E. = 1.0 s]). It is possible that the yellow Cautionary visual alert was less noticeable than the red Imminent visual alert and made it difficult for operators to detect its onset. This is further supported by the finding that there were times when participants completely missed the Cautionary visual alert. Cautionary alerts should thus either include a nonvisual component, or potentially place visual feedback closer to the source of the non-driving task.

Participants regained control of the vehicle just as fast to the Imminent visual and haptic alert when it corresponded with the unexpected lane drift (mean time of 1.2 s [S.E. = 0.1 s]) as when it was presented without the unexpected lane drift (mean time of 1.3 s [S.E. = 0.1 s]). However, it is important to note that because the alert was generated as the lane drift began, most participants reacted to the alert before the vehicle had moved significantly in the lane, making the experience very similar to an alert generated in the absence of the lane drift. However, an alternative interpretation is that Imminent visual and haptic alerts can prevent an extensive lane departure from occurring if they are generated at the moment a lane keeping performance issue occurs.

Overall, participants reported that they greatly trusted the partial automation before, during, and after experiencing it. Some participants (9 of 25) reported that they had heard of automated vehicles prior to participating in the study.

# Experiment 2 – Prompting Operators to Monitor the Road when Operating an L2 Automated Vehicle

The second experiment investigated how to prompt operators to attend to the road when distracted during the operation of an L2 automated vehicle. A secondary purpose was to investigate the effectiveness of the prompts over time.

## Approach

Fifty-six participants drove an L2 automated vehicle on a controlled test track for three 1-hour sessions. Participants were asked to perform several visually intensive non-driving tasks at the command of the experimenter. (Non-driving tasks are those tasks that do not involve the vehicle's controls. Participants were handed cards that provided tasks such as: send an email to a specific address and ask a particular question, enter a specific number-street-city-state address as noted on the card, or access a certain Internet site and answer a question based on the information from that site.) Through a between-subjects design, participants experienced one of three types of attention prompts. They were either prompted when they stopped monitoring the driving environment for 2 consecutive seconds, when they stopped monitoring the driving environment for 7 consecutive seconds, or they were not prompted at all. The prompts were composed of staged intervals of visual alerts, visual with haptic seat alerts, and a combination of visual, haptic, and auditory alerts. Each stage of the prompt lasted 5 s or until the participant responded.

In each session, participants also experienced either normal operation of the automation, an unexpected lane drift without an alert, or an unexpected lane drift with an Imminent visual and haptic alert. (As in Experiment 1, the in-vehicle experimenter triggered the lane drifts to occur when operators were involved in a non-driving task and the adjacent lanes were free of other vehicles.) Note that while alerts were fundamentally different than prompts in this study with regard to their cause, their differences were not explained to participants. To the participants with the 2-second and 7-second prompt conditions, the alert that they received along with the lane drift was indistinguishable from the prompts that they had been receiving based on their attention to the driving environment. The prompts were triggered by the participants' attention state whereas the lane drift alerts were initiated by the in-vehicle experimenter at the appropriate time relative to the vehicle drifting out of the lane.

## Findings

Overall, prompting was successful in getting participants to monitor the road. However, the prompts worked in different ways. The 7-second prompts increased participants' attention to the road after they were presented (Figure ES-1). This increase was also sustained over time. However, because they were only issued when participants exhibited 7 consecutive seconds of inattention, they were presented only when participants were extremely inattentive to the driving environment. This is reflected in the low monitoring rate observed before the prompts. In contrast, the 2-second prompts were not found to increase participants' attention to the driving environment after they were presented (Figure ES-1). However, they did lead to the highest amount of attentiveness over the course of the experiment. This may have occurred because the 2-second prompts were presented more frequently than the 7-second prompts, and were presented before participants became extremely inattentive to the driving environment. Overall, the results suggest that a 2-second prompt encourages operators to monitor the driving environment more than a 7-second prompt when operating a vehicle equipped with an L2 system. However, it is important to note that participants were not informed why they were receiving a prompt. Rather, they were only told (at the beginning of the study) what to do when they received a prompt. As such, operators' attention to the driving environment after receiving a prompt may, in part, be to diagnose why they received an alert in the first place. It must also be noted that there were participants who ignored the first stage of the prompt (visual alert only) and only reacted once they received the haptic alert, which was added 5 s into their inattention. As such, future research should explore alternative HMI approaches for the 2-second prompt to identify prompts that are effective at evoking an immediate response from operators without being annoying.

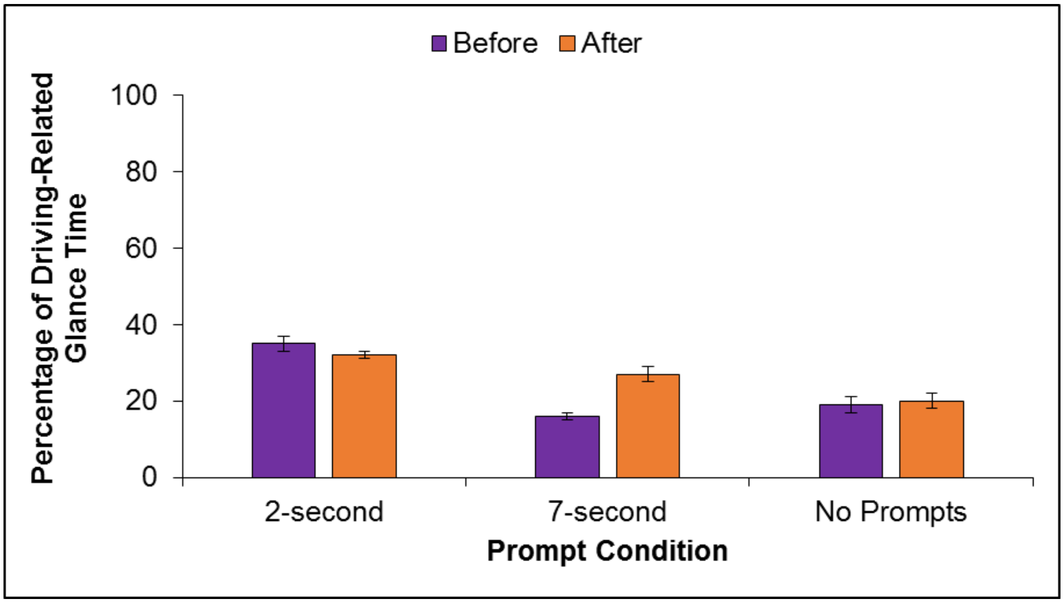



**Figure ES-1. Participants' Monitoring Rate Before and After Receiving 2-second Prompts, 7-second Prompts, or No Prompts**

The visual plus haptic Imminent alert was effective at getting participants to regain control of the vehicle when an unexpected lane drift occurred. Participants took 2.4 s to regain control of the vehicle in response to an unexpected lane drift if they received an alert. (Note that the alert and the lane drift began simultaneously.) Many participants had to be instructed to regain control of the vehicle when they did not receive an alert (23 of 49 participants without an alert versus 1 participant out of 54 with an alert). Those regaining control without an alert or an instruction from the experimenter to regain control took 4.4 s to regain control. This highlights the value of an alert in notifying operators to a lane keeping performance issue.

Similar to Experiment 1, participants reported that they greatly trusted the partial automation before, during, and after experiencing it. Some participants (16 of 56) reported that they had heard of automated vehicles prior to participating in this study. However, participants that experienced a lane keeping performance issue without an alert lost some trust in the automation. This suggests that operators may have somewhat calibrated their trust to the capabilities of the automation.

## Experiment 3 – Alerting Operators to Regain Control of an L3 Automated Vehicle

The purpose of the third experiment was to investigate what HMI characteristics are effective at alerting operators to regain control of an L3 automated vehicle and to identify the transition times between the operator and the automated functions in regard to the driving tasks.

## Approach

Twenty-five participants drove an L2 vehicle that can simulate L3 driving (with the prototype "autodrive" system) on a controlled test track for 90 minutes. They were given a tablet computer and permitted to use it of their own free will. In-vehicle time was divided into three 30-minute sessions. Participants received one TOR alert per session. They received a Staged alert in the absence of an external threat, an Imminent alert in the absence of an external threat, and an Imminent alert in response to an external threat (a revealed box on the road). In this experiment, Staged alerts comprised various phases, with alerts of increasing severity as time goes on. In this experiment, the Staged alert was composed of four phases: 1) a short tone followed by an informational message asking operators to prepare for manual control (including a countdown timer); 2) a cautionary verbal alert played in addition to an animated HMI display with the instruction to "please turn off autodrive" presented for 10 s; 3) a repeated cautionary tone played in addition to an orange visual alert with the instruction to "turn off autodrive now" presented for 10 s; and 4) a repeated imminent tone played in addition to a red visual alert with the instruction to "turn off autodrive now" presented for 10 s combined with the automation beginning to apply the brakes.

The Imminent alert (whether there was or was not an external threat) was composed of a red visual alert with the instruction to "turn off autodrive now" presented for 10 s along with with the automation applying the brakes. Participants' responses to these alerts were then investigated (e.g., Time to React, Time to Regain Control).

## Findings

In the absence of an external threat, participants were faster at regaining control of the vehicle when they received an Imminent alert (mean time of 2.3 s [S.E. = 0.2 s]) compared to when they received a Staged alert (mean time of 17.0 s [S.E. = 1.2 s]). This was likely because the Imminent alert presented direct instructions to take control now, while the information phase of the Staged alert presented a message instructing operators to prepare for manual control along with a countdown timer. The long response time may not be attributable to participants ignoring the informational phase of the Staged alert, but rather that they were following the instructions given by the HMI. It is also possible that because participants had seen a video demonstrating all four phases of the Staged alert, they might not have felt an urgency to take control. Importantly, all participants receiving the Staged alert regained control before the alert reached its third phase (Figure ES-2).

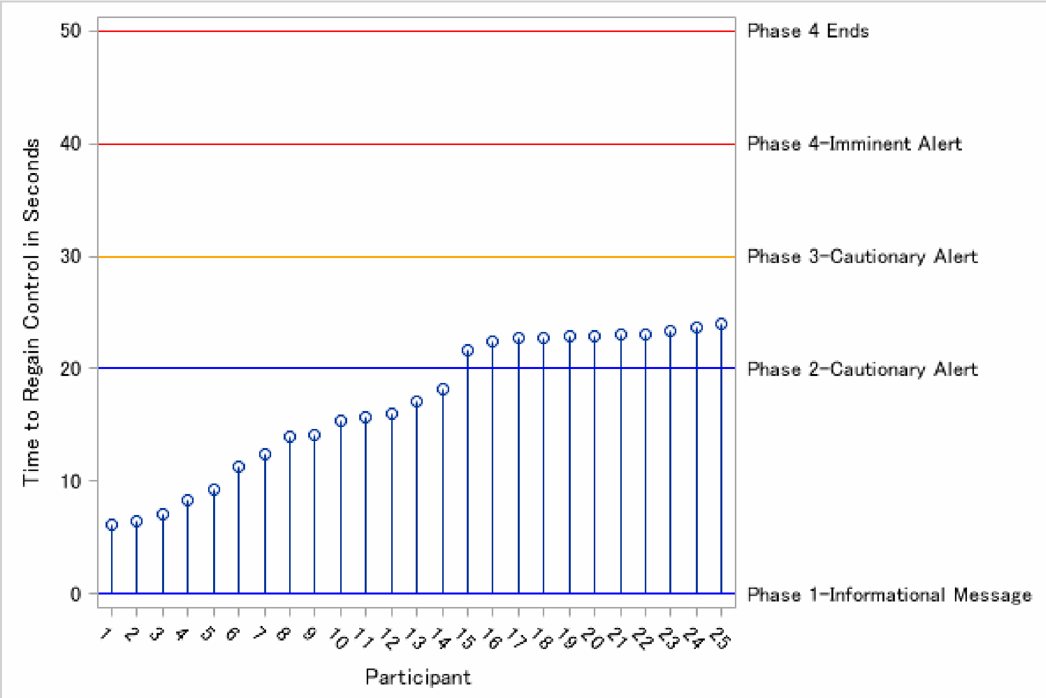



**Figure ES-2. Time to Regain Control of an L3 Automated Vehicle when Staged Alerts Are Presented**

Participants were equally as fast at regaining control of the L3 automated vehicle after an Imminent alert in the presence of an external threat (mean time of 2.1 s [S.E. = 0.1 s]) as they were in the absence of an external threat (mean time of 2.3 s [S.E. = 0.2 s]). This may be because the Imminent alert was generated at a 10-second TOR time. (Note: for both the Imminent–External Threat and Imminent–No External Threat alerts, the time buffer to react was also increased because the vehicle automatically decelerated when the TOR was issued.) There was thus a considerable amount of time provided to participants to allow them to comfortably take control before the box became threatening.

Similar to Experiments 1 and 2, trust was rated very high prior to the experiment. In addition, trust significantly increased over time. However, it should be noted that this could be because 11 of 25 participants had heard of self-driving vehicles prior to participating. Of the 11 participants who were previously aware of self-driving vehicles, 10 specifically mentioned the Google self-driving car.

After an alert, participants were observed to activate the automation and release control of the driving task. The time taken to release control of the driving task once the automation was engaged was found to be affected by the severity of the preceding alert and scenario. Participants took 1.4 s (S.E. = 0.2 s) to release control when there was a Staged alert, they took 1.8 s (S.E. = 0.3 s) to release control when there was an Imminent alert without an external threat present, and 2.7 s (S.E = 0.5 s) to release control when there was an Imminent alert in response to an unexpected box in the road. This suggests that participants may have become more skeptical of releasing control back to the automation as the severity of the scenario increased. However, it is important to note that participants still reported that they had high trust.

## Research Question 1 – How Do Drivers Interact with and Operate Vehicles that Offer Level 2 and Level 3 Automation?

The research team assessed the operation of automated vehicles by considering the time participants took to regain control when presented with an unexpected event, whether they missed or showed confusion to the TOR alerts, and whether they continued to monitor the road when it was required. Under the most effective alerting approaches tested, participants regained control of the partially automated vehicles in a relatively small amount of time when alerted to an unexpected event (i.e., unexpected lane drift when operating a vehicle equipped with an L2 system, and an unexpected box in the road when operating a vehicle equipped with an L3 system). Participants took a mean of 1.2 s (S.E. = 0.1 s) to regain control of the vehicle equipped with an L2 system when presented with an Imminent visual and haptic alert, and they took a mean of 2.1 s (S.E. = 0.1 s) to regain control of the vehicle equipped with an L3 system when presented with an Imminent visual plus auditory alert and automation-supplied braking. It should be noted that the severity of the unexpected events was relatively low at the time the Imminent alerts were presented; the lane drift had yet to occur at the time of the alert, and the TOR was issued at a 10-second time-to-collision to the box.

However, there were alerts that were missed by participants. This highlights the importance of the HMI in supporting safe human-automation interaction. The data show that operators can have a delayed response to, or completely miss, strictly visual Cautionary alerts. However, combining visual alerts with nonvisual modalities (e.g., haptic, auditory) is not a simple solution, as doing so increases the likelihood for operator annoyance.

Participant performance with the automation changed over time in certain aspects. In Experiment 2, participants took longer to respond to the prompts as their time in the study progressed over the 3-hour session. Their mean time to respond in the first session was 2.4 s (S.E. = 0.2 s), it was 3.0 s (S.E. = 0.3 s) in the second session, and 3.0 s (S.E. = 0.2 s) in the third session. This may be because the 2-second and 7-second prompts were not generated in response to an automation performance issue that required a corrective action.

## Research Question 2 – What Are the System Performance Risks from Operator Involvement with, and Interruption from, Secondary Tasks?

It was found that some operators of the vehicle equipped with an L2 system ignored the first stage of the attention prompts (a 5-second visual alert) when performing a non-driving task. There were also some operators of the vehicle equipped with an L3 system who did not regain control of the vehicle until the second stage of the TOR alert was issued (after 20 s of the first phase expired). The data indicate that some operators will exhibit a complete reversal in priority from driving-related tasks to non-driving-related tasks. When operators shift their priorities to non-driving tasks, their readiness to respond to driving-related prompts and alerts can be delayed by a feeling of obligation to complete the non-driving task first.

## Research Question 3 – What Are the Most Effective Hand-Off Strategies between the System and the Operator?

The most effective hand-off strategies in this study were those that involved nonvisual components. With the L2 automated vehicles in Experiments 1 and 2, participants were fastest at regaining control of the vehicle when they received a combined visual and haptic seat alert. In Experiment 2, some participants even missed the visual prompts and only responded when the prompts included a haptic component. Including a nonvisual component in the hand-off strategy may lead to the best outcome given the potential for operators to exhibit a primary task reversal when performing non-driving tasks. Other types of visual alerts may also be effective.

Interestingly, Experiment 3 with the L3 automated vehicle showed that operators will substantially delay how long they take to regain control of the vehicle when presented with an informational alert that combines visual and auditory elements. This delay, however, is possibly a consequence of the HMI notifying operators to “prepare” to take control (rather than stating to take control now) as well as showing them just how much time they have to react (a countdown timer was presented in this case). As such, if the HMI conveys to operators that they have time to prepare to regain control, many may take that time to finish their non-driving activities.

## Research Question 4 – How Do Operators Engage, Disengage, and Reengage with the Driving Task?

When the automation issued a TOR alert, operators’ engagement in the driving task was characterized as the time taken for them to react and regain control, as well as the sequence of actions they took to regain control. The time taken to reengage the automation and release control was also assessed (Table ES-1). Participants reacted (e.g., looked in the forward direction) to the TOR alerts within 1.2 s, on average, depending on the study. Participants preferred to grab the steering wheel first to regain control of the L2 automated vehicle. This took, on average, 1.3 s to 2.4 s. Participants preferred to press the off button on the steering wheel to regain control of the L3 automated vehicle if they received an informational alert. However, participants took most of the allotted time to perform this action, with a mean time of 17 s. In contrast, participants operating an L3 automated vehicle preferred to immediately press the brake pedal if they received an Imminent alert. This took, on average, less than 2.3 s. Perhaps operators pressed the brake pedal here because of the urgency conveyed by the Imminent alert.

With regard to reengaging the automation after a brief moment of manual control, participants took, on average, less than 4.6 s to release control once the automation became available. This time fluctuated somewhat and may have depended on various factors potentially including whether participants believed the automation could go into a more severe warning state (as in the case of Staged alerts) once they had received the first Staged alerts.

**Table ES-1. Mean Times for Responses to Different Alerts and Prompts (units: seconds)**

| Dependent Variable | Experiment 1 – Level 2: Imminent Multimodal | Experiment 1 – Level 2: Staged Multimodal | Experiment 2 – Level 2: Lane Drift with Alert | Experiment 2 – Level 2: Lane Drift without Alert | Experiment 2 – Level 2: 2-sec Prompt [b] | Experiment 2 – Level 2: 7-sec Prompt [b] | Experiment 3 – Level 3: Staged | Experiment 3 – Level 3: Imminent –No External Threat | Experiment 3 – Level 3: Imminent –External Threat |
|---|---|---|---|---|---|---|---|---|---|
| Time to React | 0.7 | 0.6 | 1.0 | 3.6 | 2.7 | 2.9 | -38.8 [c]<br>1.2 [d] | 0.7 | 0.7 |
| Time to Regain Control | 1.3 | 1.4 | 2.4 | 5.7 [a] | 12.1 [b] (n=6) | 32.3 [b] (n=6) | -23.0 [c]<br>17.0 [d] | 2.3 | 2.1 |
| Method to Regain Control | Steering Wheel | Steering Wheel | Steering Wheel | Steering Wheel | Steering Wheel [b] | Steering Wheel [b] | Off-Button (Steering Wheel) | Brake | Brake |
| Time to Release Control | 2.4 | 4.6 | 3.3 | 4.2 | 2.9 [b] (n=6) | 2.7 [b] (n=6) | 1.4 | 1.8 | 2.7 |

[a] Not all operators were able to regain control; this value only includes those that were able to do so for the event.
[b] Not all operators had to regain control; only the last phase of the prompt required the operator to regain control.
[c] Time calculated relative to the point in time when the Imminent alert would have been presented.
[d] Time calculated from the point in time when the informational phase of the Staged alert began.

## Research Question 5 – How Do Operators Perform Under Various Operational Concepts within Level 2 and Level 3 Automation?

The three experiments were performed at test-track facilities simulating mixed-traffic conditions at highway speeds through the use of multiple confederate vehicles. Under these conditions, participants interacted with the automation and performed reasonably well when an effective HMI was utilized. As mentioned earlier, however, operator performance can substantially degrade when the HMI message is not perceived by the operator, such as in cases when the operator is engaged in a non-driving task. It should be noted that under L2 automation, operators are expected to continue monitoring the road. Their performance in regaining control of the automated vehicle when performing a visually distracting task represents a severe, although realistic, use case.

It is interesting that participants quickly released control to the automated vehicle when traveling in proximity to other vehicles on the road. This further highlights the trust in the partial automation that was sustained by participants over their time in the study.

## Research Question 6 – What Are the Most Effective Human-Machine Interface Concepts which Optimize the Safe Operation of Level 2 and Level 3 Systems?

Of HMI concepts tested, the most effective were those that involved nonvisual alerts in addition to the visual alerts. When engaged in a non-driving task, some participants exhibited a primary task reversal and chose to prioritize the completion of the non-driving task over the operation of the partially automated vehicle. This phenomenon (overreliance on vehicle automation) has the potential to counteract many of the safety benefits of automated vehicle technologies. The HMI concepts that included auditory and haptic components were more successful in grabbing the attention of participants when they were engaged in a visually demanding task and were much more effective at eliciting a safe response compared to the HMI concepts that only relied on a visual alert. It is important to note that there are many different types of visual, auditory, and haptic alerts that could possibly yield different results from this study.

One reason why this occurred could be that the omnidirectional properties of the haptic and auditory alerts allowed operators to perceive the alerts when engaged in a visually demanding non-driving task. However, another reason could be that the haptic and auditory alerts conveyed a higher urgency than the strictly visual alerts.

It is important to note here that the strictly visual alert concepts were considered in order to decrease the likelihood of annoying the vehicle operators. There is a concern that an HMI that continually generates auditory or haptic alerts will be unacceptably annoying to operators who could, in turn, choose not to purchase vehicles with the safety benefits of automated technologies. It is therefore important to carefully balance an alerting approach’s conspicuity, urgency, and annoyance. Future research would help to better understand how to optimize the HMI for partially automated vehicles.

# Summary

This study investigated how well operators interact with L2 and L3 partially automated vehicles on a controlled test track in mixed traffic. The study shows that different HMI elements can have a large impact on how operators interact with these vehicles. Overall, participants greatly trusted the capabilities of the automated systems. Although this trust is essential for widespread adoption, participants were also observed prioritizing non-driving activities over the operation of the vehicle and disregarding TORs when they were presented. Future research could explore these issues to help optimize the automation’s HMI. The driver engagement patterns observed in this study provide data and evidence that could support the future development of human factors design principles for L2 and L3 partially automated vehicles.

# Chapter 1 Introduction

## Objective

The purpose of this study was to examine the interaction between human users and automated vehicle systems. Specifically, how do human users interact with vehicles that have L2 and L3 automated systems, can these users take over control of the driving task when required, and can they determine the acceptable balance between controlling the vehicle when necessary and letting the automated system function as designed to perform the driving task when appropriate? The ultimate goal of this research was to ascertain how operators interact with automated vehicles and determine how automated vehicle technology can best support safe driving.

## Rationale

As automated driving technology advances, the “driver’s” role is shifting from active vehicle control to passive supervision of the automated system and/or the environment. The current study focused on the human factors issues that arise when vehicles equipped with automation technologies shift the human from the role of driver to that of operator. Automated vehicle systems must be designed to instruct and prompt the operator to act, if and when needed, in a timely and appropriate way in order to ensure safe operation of the vehicle.

## Background

Technological advancements over the past decade have led to the emergence of advanced driver assistance systems. Current features such as Adaptive Cruise Control (ACC), collision warning, automatic braking, and lane-keeping assist (LKA) systems are becoming commonplace in modern automobiles. Further, automated systems that combine limited lateral and longitudinal control over a vehicle are becoming commercially available. Some of these systems incorporate various methods to ensure driver participation.

While automated systems offer the potential for increased safety and reduced human error, their use may create issues that could benefit from further investigation. These issues may include negative adaptations based on misunderstanding, misuse, over-reliance on the automated systems, and potential distraction from the driving task due to interaction with the automated system. These issues should be examined in order to address any potential for unforeseen consequences of increased automation.

Of specific interest is how an automated system will impact operators’ willingness to engage in non-driving-related tasks. As noted in previous work, (e.g., Llaneras, Salinger, & Green, 2013), the current generation of automated systems is designed to support, rather than replace, the driver. The presence of automated systems may create the perception to free operators’ attention, which may then be directed at non-driving tasks. Redirection of attention to non-driving tasks may also impact an operator's situational awareness, including the ability to perceive critical factors in the environment or to detect issues with the automated systems (e.g., system state changes or failures).

This research effort was informed by two previous documents: *Past Research, State of Automation Technology, and Emerging System Concepts*, DOT HS 812 043 (Trimble, Bishop, Morgan, and Blanco, 2014) and the *Concepts of Operation for Levels 2 and 3*, DOT HS 812 044 (Marinik, Bishop, Fitchett, Morgan, Trimble, and Blanco, 2014). In the former document, Trimble et al. (2014) conducted a review of key human factors studies related to automated vehicle operations within the context of automation Levels 2 and 3. The review expanded and updated the results from a prior literature review that was performed for the U.S. Department of Transportation (DOT). The content reflected the latest research and the activity of original equipment manufacturers (OEMs) as of June 2013, and studies directly addressing both automated driving and those relevant to automated driving concepts were included. Additionally, documents beyond the academic literature—such as articles, summaries, and presentations from OEMs and suppliers—were researched, and information from both U.S. and international projects and researchers was included. The document also identified relevant automated-driving databases in support of future research efforts.

A parallel effort focused on the preparation of a concepts of operation (ConOps) document (Marinik et al., 2014).This document evaluated the functional framework of operations for Level 2 and Level 3 automated vehicle systems. This task was achieved by defining the varying levels of automation, the operator-vehicle interactions, and system components, and then further assessing the automation-relevant parameters from a scenario-based analysis standpoint. Within the ConOps, both a broad set and a down-selected set of parameters were identified based on industry activity in research and development. Table 1-1 provides a list of the down-selected parameters and high-ranking modalities, as well as whether these parameters were addressed within the three experiments performed in the current study.

**Table 1-1. Down-selected List of Automation-relevant Parameters (Marinik, Bishop, Fitchett, Morgan, Trimble, and Blanco, 2014) and Corresponding Representation Within the Experimental Design**

| Parameter | High-ranking Modality | Experiment 1 | Experiment 2 | Experiment 3 |
|---|---|---|---|---|
| **Road Facility Type** | Limited-access highway | Simulated limited-access highway on a test track | Simulated limited-access highway on a test track | Simulated limited-access highway on a test track |
| **Automated Vehicle Segregation** | None (mixed traffic) | None | None | None |
| **Infrastructure Adaptation** | None | None | None | None |
| **Connected Automated Operation** | Cellular/GPS only | N/A | N/A | GPS only |
| **Inter-Vehicle Coordination** | Platooning (research) | N/A | N/A | N/A |
| **Inter-Vehicle Coordination** | No coordination (individual vehicle; industry) | No coordination | No coordination | No coordination |
| **Speed of Travel** | 0-30 mph (earliest marketing timing) | N/A | N/A | N/A |
| **Speed of Travel** | 30-75 mph (near term) | 50 mph | 60 mph | 45 mph |
| **Traffic Density based on the FHWA Levels of Service (LOS)**[1] | All FHWA LOS | LOS C | LOS C | LOS B |
| **Awareness of and Operation Relative to Traffic Control Devices** | Speed signs | N/A | N/A | N/A |
| **Awareness of Other Vehicle Indications** | None | None | None | None |
| **Situational Awareness** | Vehicles, motorcycles, pedestrians | Vehicles | Vehicles | Vehicles |
| **Vehicle Maneuvers Under Automated Control** | Stay in original lane | Stay in original lane | Stay in original lane | Stay in original lane |
| **Weather Conditions** | Dry, clear, rain | Dry | Dry | Dry |
| **Roadway Surface Conditions** | Dry, wet | Dry | Dry | Dry |
| **Driver Ability in Manual Driving** | Novice, experienced | Experienced | Experienced | Experienced |
| **Driver Monitoring** | Both monitored and unmonitored | Unmonitored | Monitored | Unmonitored |
| **Driver Task Requirement to Maintain Engagement** | Yes and no | No | Yes | No |

| Parameter | High-ranking Modality | Experiment 1 | Experiment 2 | Experiment 3 |
|---|---|---|---|---|
| **Intended Duration of Automation[2]** | All | Medium | All | Extended |
| **Engage/Disengage Method** | Initiated by driver/system/both | Both | Both | Both |
| **Driver Engagement Timing[3]** | All | Short, Medium, Long | Short, Medium, Long | Medium, Extended |
| **Driver Training[4]** | None | None | None | Minimal |

[1] The Federal Highway Administration (FHWA) LOS are as follows: LOS A (free flow), LOS B (flow with some restrictions), LOS C (stable flow: maneuverability and speed more restricted), LOS D (unstable flow, temporary restrictions momentarily slow vehicle), LOS E (unstable flow, vehicles unable to pass, temporary stoppages), and LOS F (forced flow condition with congestion and low speed).
[2] Intended durations of automation are categorized as follows: short (less than 1 minute), medium (1–10 minutes), extended (10–30 minutes), and long (greater than 30 minutes)
[3] Driver Engagement Timing, i.e., the amount of time the system allows the driver to reengage operations after a period of automated driving is defined as follows: short (less than 3 s), medium (3–10 s), long (10–30 s), and extended (greater than 30 s).
[4] Driver Training alternatives are defined as follows: none (no driver training; the driver discovers system operation and forms his/her mental model of system operation), minimal (the driver is provided information, e.g., a video shown at the dealership, before driving the vehicle for the first time), and substantial (before driving an automated vehicle, a special driving certification based on training must be obtained).

This down-selected set of parameters was further reduced to a list of selected concepts for human-machine interface (HMI) evaluation. This list includes those high-ranking modalities identified as common to all automation concepts: highway, mixed traffic, infrastructure as-is, only cellular and Global Positioning System (GPS) connectivity, no vehicle coordination (individual vehicle operation), awareness of vehicles, motorcycles, and pedestrians, dry and wet weather/road conditions, novice and experienced drivers, and engage/disengage by the operator or system. Table 1-2 provides a list of these common concepts as well as whether these parameters were addressed within the three experiments.

**Table 1-2. Selected Automation Concepts for HMI Evaluation (Marinik, Bishop, Fitchett, Morgan, Trimble, and Blanco, 2014) and Corresponding Representation Within the Experimental Design**

| Parameter | High-ranking Modality | Experiment 1 | Experiment 2 | Experiment 3 |
|---|---|---|---|---|
| **Road Facility Type** | Highway | Limited-access test track | Limited-access test track | Closed track |
| **Automated Vehicle Segregation** | None (mixed traffic) | None | None | None |
| **Infrastructure Adaptation** | Infrastructure as-is | None | None | None |
| **Connected Automated Operation** | Cellular/GPS only | N/A | N/A | GPS only |

| | | | | |
|---|---|---|---|---|
| **Inter-Vehicle Coordination** | No coordination (individual vehicle; industry) | No coordination | No coordination | No coordination |
| **Situational Awareness** | Vehicles, motorcycles, pedestrians | Vehicles | Vehicles | Vehicles |
| **Weather Conditions** | Dry, clear, rain | Dry | Dry | Dry |
| **Roadway Surface Conditions** | Dry, wet | Dry | Dry | Dry |
| **Driver Ability in Manual Driving** | Novice, experienced | Experienced | Experienced | Experienced |
| **Driver Monitoring** | Both monitored and unmonitored | Unmonitored | Monitored | Unmonitored |
| **Engage/Disengage Method** | Initiated by driver/system/both | Both | Both | Both |

## Levels of Automation

The National Highway Traffic Safety Administration (NHTSA) has defined five levels of vehicle automation (NHTSA, 2013). These definitions are based on the level of user versus vehicle control. In this taxonomy, as the level of automation increases, the role of the operator shifts from primary control to that of supervisory control (Figure 1-1). This taxonomy was used to guide all research under the current study, which focused on Level 2 and Level 3 vehicle automation. Two experiments utilized Level 2 vehicles, and one experiment utilized a Level 3 vehicle. (It should be noted that, while the Society of Automotive Engineers has also developed standards for automation levels, their nomenclature will not be addressed in this report. [SAE, 2014]).

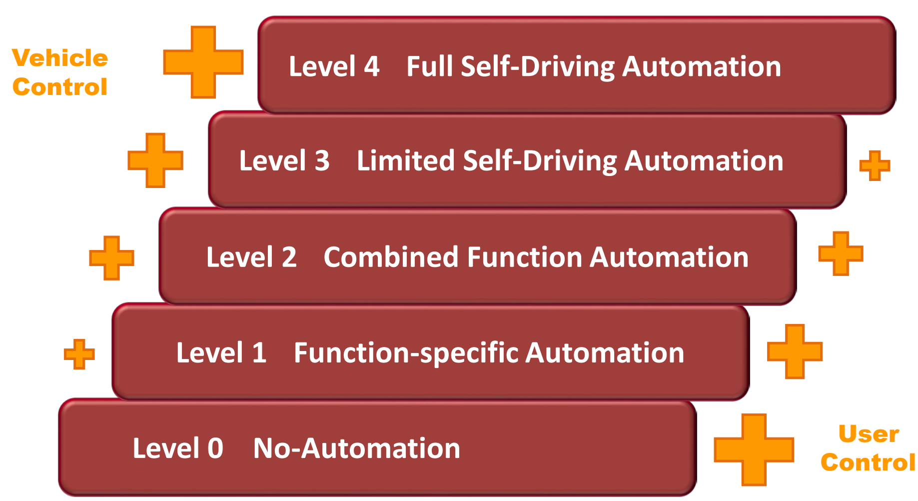



**Figure 1-1. The NHTSA Automation Levels**

The operational definitions below were extracted from NHTSA's Preliminary Statement of Policy Concerning Automated Vehicles (NHTSA, 2013).

***Level 0 (L0) No-Automation***

In Level 0 automation (or no automation), "[t]he driver is in complete and sole control of the primary vehicle controls (brake, steering, throttle, and motive power) at all times, and is solely responsible for monitoring the roadway and for safe operation of all vehicle controls. Vehicles that have certain driver support/convenience systems but do not have control authority over steering, braking, or throttle would still be considered 'Level 0' vehicles. Examples include systems that provide only warnings (e.g., forward collision warning, lane departure warning, blind spot monitoring) as well as systems providing automated secondary controls such as wipers, headlights, turn signals, hazard lights, etc. Although a vehicle with V2V [vehicle-to-vehicle] warning technology alone would be at this level, that technology could significantly augment, and could be necessary to fully implement many of the technologies described below, and is capable of providing warnings in several scenarios where sensors and cameras cannot (e.g., vehicles approaching each other at intersections)."

***Level 1 (L1) Function-specific Automation***

Level 1 automation "involves one or more specific control functions; if multiple functions are automated, they operate independently from each other. The driver has overall control, and is solely responsible for safe operation, but can choose to cede limited authority over a primary control (as in adaptive cruise control), the vehicle can automatically assume limited authority over a primary control (as in electronic stability control); or the automated system can provide added control to aid the driver in certain normal driving or crash-imminent situations (e.g., dynamic brake support in emergencies). The vehicle may have multiple capabilities combining individual driver support and crash avoidance technologies, but does not replace driver vigilance and does not assume driving responsibility from the driver. The vehicle's automated system may assist or augment the driver in operating one of the primary controls – either steering or braking/throttle controls (but not both). As a result, there is no combination of vehicle control systems working in unison that enables the driver to be disengaged from physically operating the vehicle by having his or her hands off the steering wheel AND feet off the pedals at the same time. Examples of function-specific automation systems include: cruise control, automatic braking, and lane keeping."

***Level 2 (L2) Combined Function Automation***

Level 2 "involves automation of at least two primary control functions designed to work in unison to relieve the driver of control of those functions. Vehicles at this level of automation can utilize shared authority when the driver cedes active primary control in certain limited driving situations. The driver is still responsible for monitoring the roadway and safe operation and is expected to be available for control at all times and on short notice. The system can relinquish control with no advance warning and the driver must be ready to control the vehicle safely. An example of combined functions enabling a Level 2 system is adaptive cruise control in combination with lane centering. The major distinction between level 1 and level 2 is that, at level 2 in the specific operating conditions for which the system is designed, an automated operating mode is enabled such that the driver is disengaged from physically operating the vehicle by having his or her hands off the steering wheel AND foot off pedal at the same time."

### *Level 3 (L3) Limited Self-Driving Automation*

Vehicles with Level 3 automation “enable the driver to cede full control of all safety-critical functions under certain traffic or environmental conditions and in those conditions to rely heavily on the vehicle to monitor for changes in the conditions requiring transition back to driver control. The driver is expected to be available for occasional control, but with sufficiently comfortable transition time. The vehicle is designed to ensure safe operation during the automated driving mode. An example would be an automated or self-driving car that can determine when the system is no longer able to support automation, such as from an oncoming construction area, and then signals to the driver to reengage in the driving task, providing the driver with an appropriate amount of transition time to safely regain manual control. The major distinction between level 2 and level 3 is that at level 3, the vehicle is designed so that the driver is not expected to constantly monitor the roadway while driving.”

### *Level 4 (L4) Full Self-Driving Automation*

The Level 4 vehicle “is designed to perform all safety-critical driving functions and monitor roadway conditions for an entire trip. Such a design anticipates that the driver will provide destination or navigation input, but is not expected to be available for control at any time during the trip. This includes both occupied and unoccupied vehicles. By design, safe operation rests solely on the automated vehicle system.”

NHTSA’s Preliminary Statement of Policy Concerning Automated Vehicles provides a footnote for the word “driver” where it appears in the Level 4 description, explaining that:

> Several State automated vehicle laws consider the person who activates the automated vehicle system to be the “driver” of the vehicle even if that person is not physically present in the vehicle. NHTSA, however, is not aware of any prototype automated vehicle systems that are capable of operating on public roads without the presence of a driver in the driver’s seat who is ready to control the vehicle. (NHTSA, 2013, p. 5)

## Project Research Questions

This study centered on six key research questions developed by NHTSA. The focus of this research was to address each question based on sound empirical research findings. The research questions were:

1. How do drivers interact with and operate vehicles that offer Level 2 and Level 3 automation; e.g., what is the driver performance profile over length of time in continuous or sustained automation?
2. What are the system performance risks from driver involvement with, and interruption from, secondary tasks (such as portable electronic device use) that could arise when operating Level 2 or Level 3 automated vehicle systems?
3. What are the most effective hand-off strategies between the system and the driver, including response to faults/failures?

4. How do drivers engage, disengage, and reengage with the driving task in response to the various states of Level 2 and Level 3 automation?

5. How do drivers perform under various operational concepts within Level 2 and Level 3 automation, such as systems intended for everyday driving on open roadways in mixed traffic or systems intended for dedicated roadway-vehicle applications (e.g., automated lanes, remote highways)?

6. What are the most effective human-machine interface concepts, guided by human factors best practices, which optimize the safe operation of Level 2 and Level 3 systems?

The six aforementioned research questions were addressed in the three experiments. Experiment 1 examined how best to alert operators to regain control of the vehicle, Experiment 2 examined the system prompt effectiveness over time, and Experiment 3 examined human-automation system performance over time. Experiments 1 and 2 were conducted using a vehicle equipped with an L2 system, while Experiment 3 was conducted with a vehicle equipped with an L2 system that can simulate L3 driving on a test track. A summary of the three experiments is provided in Table 1-3. The details of the experiments and their findings are presented in the next three chapters.

**Table 1-3. Overview of the Three Experiments**

| Human Factors Evaluation of Level 2 and Level 3 Automated Driving Concepts | | | | |
|---|---|---|---|---|
| | | Experiment 1 | Experiment 2 | Experiment 3 |
| Experiment Goal | | Alerting Operators to Regain Control of Vehicle | System Prompt Effectiveness Over Time | Human-Automation System Performance Over Time |
| NHTSA Automation Level | | Level 2 – Combined Function Automation | | Level 3 – Limited Self-Driving Automation |
| Test Scenario | Test Type | Test Track simulating a mixed traffic highway driving condition | | |
| | Venue | Milford Proving Ground circle track in Milford, MI | | Virginia Smart Road test track in Blacksburg, VA |
| | Speed | 60 mph | 50 mph | 45 mph |
| | Duration | One 90-minute driving session, ~175 minutes for the entire experimental session | Three 1-hour driving sessions, ~310 minutes for the entire experimental session | Three 30-minute driving sessions, ~ 200 minutes for the entire experimental session |
| Experiment | Design | 3 x 2 within-subject design | 3 x 3 x 3 mixed factorial design | 3 x 3 within-subject design |
| | Alerts | 19 alerts | 3 event types and 3 prompt conditions | 3 alert types |
| | Non-Driving Tasks | Tablet tasks (navigation, email, and Web browsing) provided by the in-vehicle experimenter | Tablet tasks (navigation, email, and Web browsing) provided by the in-vehicle experimenter | Free exposure to interact with tablet and smartphone |
| Participants | Age | 44.3 years (S.D. = 19.2 years) | 41.0 years (S.D. = 16.3 years) | 38.8 years (S.D. = 13.8 years) |
| | Total | 25 Participants | 56 participants | 25 participants |
| Independent Variables | | • Take Control Alert Type: Cautionary, Imminent, and Staged<br>• Alert Modality: Unimodal and Multimodal | • Prompt Conditions: 2-second, 7-second, or No Prompts<br>• Event Type: Alert, No Alert, No Lane Drift<br>• Driving Session: 1st, 2nd, and 3rd | • Take Control Alert Type: Staged, Imminent–External Threat, and Imminent–No External Threat<br>• Driving Session: 1st, 2nd, and 3rd |
| Dependent Variables | Operator Behavior | • Time to react to alert<br>• Time to regain control<br>• Performance<br>• Time to release control<br>  ▪ Time to activate automation<br>  ▪ Time to release control of steering<br>• Time to resume non-driving task | • Time to react to alert<br>  ▪ Time to react to event<br>  ▪ Time to react to prompt<br>• Time to regain control<br>• Performance<br>• Time to activate automation<br>• Time to release control of steering<br>• Time to resume non-driving task<br>• Monitoring rate and non-driving-related glances | • Time to react to alert<br>• Time to regain control<br>• Performance<br>• Method used to regain control/cancel automation<br>• Time to release control<br>  ▪ Time to activate automation<br>  ▪ Time to release control of steering<br>• Time to resume non-driving task<br>• Monitoring rate and non-driving-related glances |
| | Subjective Analysis | • 10 in-vehicle trust scales<br>• After-experience trust scale<br>• Open-ended interview | • 21 in-vehicle trust scales<br>• 3 post-session trust scales and 3 post-session satisfaction scales<br>• Open-ended interview | • 12 in-vehicle trust scales<br>• After-experience trust scale<br>• Open-ended interview |

# Chapter 2 Level 2 Automated Vehicle, Experiment 1: Alerting Operators to Regain Control of an L2 Automated Vehicle

## Purpose

The goal of Experiment 1 was to examine how operators in a mixed-traffic (involving surrounding legacy vehicles, Level 0) simulated highway driving condition respond to and assess the utility of alternative notification strategies; i.e., advance notification of potential problems where operators will possibly need to intervene versus urgent imminent notifications requiring immediate action. For this study, two aspects of the Level 2 operational definition (NHTSA, 2013) are of importance and set the tone for the experimental design and scenarios of interest:

- "The driver is still responsible for monitoring the roadway and safe operation and is expected to be available for control at all times and on short notice."
- "The system can relinquish control with no advance warning and the driver must be ready to control the vehicle safely."

Based on these operational definitions, Experiment 1 was designed to explore which type of alert will be more effective in assisting operators when they have to take control of the vehicle on short notice.

## Method

For Experiment 1, a single, long-exposure, experiment was conducted. Participants were provided with a thorough familiarization of the vehicle and its operation, followed by a single, approximately 90-minute, exposure to the vehicle in L2 automated driving. During the driving session, participants were instructed to perform non-driving-related tasks (e.g., e-mail, Web browsing) and were, at times, presented with alerts stating that they must take control of the vehicle. Three forms of alerts were presented: Cautionary, Imminent, and Staged. The Cautionary alerts provided information to the participants that a potential problem was detected. The Imminent alerts provided the participants with a message that an active fault was detected. The Staged alerts transitioned from a cautionary alert phase to an imminent alert phase. Participants' reactions to these messages, both in duration and method of response, were among the variables examined in this experiment.

# Experimental Design

The study was performed as a within-subject design. All participants completed one driving session during which they received a total of 19 system alerts. Specifically, each participant experienced six Cautionary alerts, six Staged alerts, and six Imminent alerts. For each of those alert types, participants experienced three unimodal alerts (visual only) and three multimodal alerts (visual + haptic). After receiving these 18 alerts, each participant received an Imminent multimodal alert coupled with an experimenter-triggered lane drift, resulting in a total of 19 alerts. The study was designed to mimic worst-case scenarios when conditions for monitoring the roadway were decreased.

Eighteen of the nineteen alerts were presented to participants in six different orders. Each order consisted of the six different combinations of alert type and alert modality, and this was repeated three times within the experimental session. Using all six possible alert type and alert modality combinations, a Latin square was developed to create six different orders of alert presentation. The order was repeated three times within the 90-minute driving session, resulting in 18 alerts. The alerts were presented at random times, between 2 and 8 minutes; thus, participants were less likely to be able to anticipate when they would occur.

## Independent Variables

The research design included two independent variables, detailed below.

### *Take Control Alert Type*

This is a within-subject variable with three levels: Cautionary, Imminent, and Staged. According to NHTSA's Level 2 automation definition, the operator is expected to be available for control at all times and on short notice. Thus, the alerts were timed to provide limited time for the participant to react and regain control of the vehicle. (Depending upon the situation, the in-vehicle experimenter would instruct the participant to regain control if the participant failed to do so in the allotted time.) Note that these timings are not values suggested for future design; rather, they were deemed appropriate for this study to start the exploration of the different reaction times of interest and participants' behavior.

- Cautionary. This alert involved an approximately 30-second period for the participant to take control of the vehicle before the system issue resolved by itself (i.e., the alert turned off but the automation did not turn off). The participant was always prompted by a visual alert on the HMI with the instruction to "take steering," along with a yellow light-emitting diode (LED).
- Imminent. This alert allowed the participant approximately 30 seconds (s) to take control of the vehicle before the automation turned off. The participant was prompted by a visual alert on the HMI with the instruction to "take steering," along with a flashing red LED. Upon taking control of the steering wheel, the HMI informed the participant that the automation was no longer available. If the participant failed to regain control within 30 s, the automation turned off, at which point neither the system nor the participant was in control of the vehicle. Thus, the in-vehicle experimenter instructed the participant to take control of the vehicle to ensure safety.
- Staged. This alert involved an approximately 30-second period of increasing urgency for the participant to take control of the vehicle before the automation turned off. During the first 10 s of

the alert, the participant was prompted by a visual alert on the HMI providing the instruction to "take steering," along with a yellow LED. If the participant regained control within these 10 s, the situation resolved by itself (i.e., the alert turned off but the automation did not turn off). However, if the participant failed to regain control during this time, the alert progressed to an imminent phase. During this phase, the participant was prompted by the visual alert on the HMI providing the instruction to "take steering," along with a flashing red LED. If the participant failed to regain control of the vehicle within 20 s of the onset of the imminent phase of the Staged alert, the automation turned off, at which point neither the system nor the participant was in control of the vehicle. Thus, the in-vehicle experimenter instructed the participant to take control of the vehicle to ensure safety.

#### *Alert Modality*

This is a within-subject variable with two levels: unimodal and multimodal.

- Unimodal. This alert condition consisted of only the visual alert (i.e., the "take steering" instruction displayed on the HMI along with a yellow or flashing red LED).
- Multimodal. This alert condition consisted of the visual alert (i.e., the "take steering" instruction displayed on the HMI along with a yellow or flashing red LED) in addition to a haptic seat vibration.

## Dependent Variables

The research design included seven dependent variables, detailed below.

#### *Operator Behavior*

- **Time to react to alert** (in seconds)
  This was defined as the time, after the alert is first presented, at which point the participant performs an action that could be considered a reaction to the alert (e.g., look forward, move foot or hand).
- **Time to regain control** (in seconds)
  This was defined as the time which begins when the alert is first presented until the participant attempts to regain manual control by taking control of the steering wheel.
- **Performance**
  This was defined as the correct action taken by the participant in response to an event of interest (e.g., does the participant regain control upon receiving the alert?).
- **Time to release control**
  This variable consisted of two components.
    - <u>Time to activate the automation</u> (in seconds)
      This was defined as the time from autonomy availability as indicated by the system until the participant attempts to activate the automation. Note that during the Cautionary alerts, the system issue resolved by itself (i.e., the alert turned off but the automation did not turn off); thus, most participants did not have time to activate the automation for this alert type.

- Time to release control of steering (in seconds)
  This was defined as the time from successfully activating the automation until the participant removes both hands from the steering wheel.

- **Time to resume non-driving task** (in seconds)
  This was defined as the time from releasing control of the steering wheel until the participant resumes interaction with a non-driving-related task.

Figure 2-1 illustrates the sequence in which the dependent variables pertaining to operator behavior were measured.

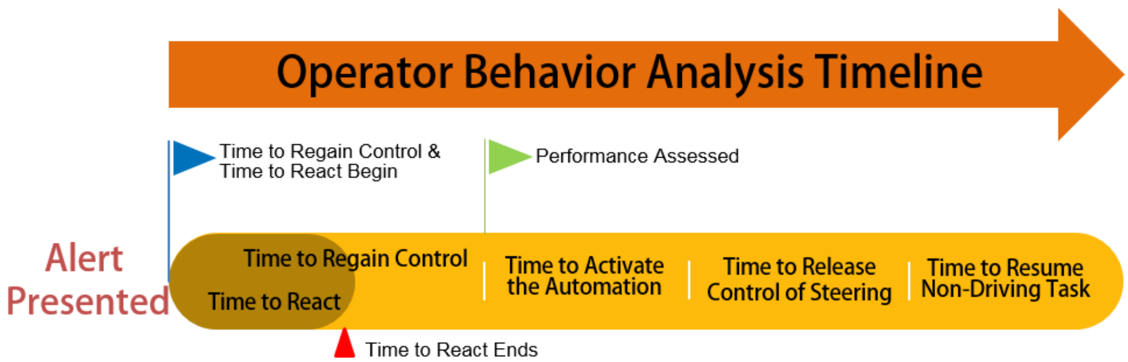



**Figure 2-1. Sequence of the Experiment 1 Dependent Variables Pertaining to Operator Behavior**

## Participants' Subjective Assessment

A trust scale (Appendix E) was administered 10 times throughout the experimental session at approximately 9-minute intervals. Participants were asked to rate their trust in the ability of the automation to function properly while they engaged in non-driving tasks using a 7-point Likert-type scale, in which 1 corresponded with "strongly disagree" and 7 corresponded with "strongly agree."

In addition to the 10 trust ratings collected throughout the experimental session, participants were asked to complete the after-experience trust scales (Appendix E) and participate in an open-ended interview (Appendix F) upon completing the driving session.

## Facility

This experiment was conducted at the Milford Proving Ground (MPG) circle track in Milford, Michigan (Figure 2-2). This facility is owned and maintained by General Motors (GM) and includes a 7.2-km (4.5-mi) banked circle track with five travel lanes. The travel speed for each lane falls within a designated speed range, with the innermost lane allowing for stop-and-go traffic and the outermost lane being restricted to speeds of 100 mph and above. Experiment 1 was conducted in Lane 3, which allowed speeds of between 50 and 70 mph. This track has been used to simulate freeway conditions in previous studies on automated driving. The circle track is restricted to approved studies and approved trained drivers; therefore, other non-study traffic was present during the experiment.

**Figure 2-2. Aerial View of GM's Milford Proving Ground**

## Vehicle

A 2009 Chevrolet Malibu equipped with a prototype L2 automated driving system was used as the experimental vehicle for Experiment 1 (Figure 2-3). As part of the automated driving system, several HMI components were installed. This vehicle was modified to include ACC and lane centering, along with a flexible driver interface and researcher's control console. The purpose of the researcher's console was to allow the in-vehicle experimenter to trigger various displays and to change the operation of the automated driving system, which included simulating erroneous behavior and equipment failures.

**Figure 2-3. 2009 Chevrolet Malibu with a Prototype L2 System**

The vehicle was also instrumented with a data collection and recording device. The data recorder was connected to the automated driving system and the vehicle Controller Area Network (CAN).

Key variables collected included status of the automation (e.g., off, on and actively controlling, failure mode), vehicle speed, lane position, and flags indicating the presentation of messages and system failures. In addition, the following video views were collected (Figure 2-4):

- Operator's face
- Over-the-shoulder (OTS) view
- Forward roadway
- Rear
- Foot (pedal area)
- HMI
- Exterior left rear
- Exterior right rear

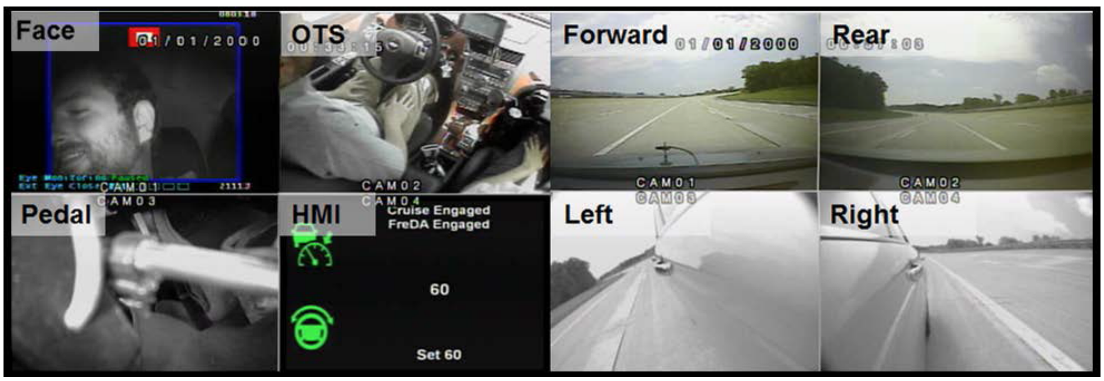



**Figure 2-4. GM's Data Acquisition System (DAS) Camera Views for Experiment 1**

## Tablet Computer

An Asus Nexus 7 tablet computer was used for this study in order to provide opportunities for the participants to perform non-driving tasks while the vehicle was in Level 2 automation (Figure 2-5). The tablet has exterior dimensions of 198.5 mm by 120 mm, providing a diagonal screen size of approximately 178 mm (7 inches). The tablet has a screen resolution of 1,280 by 800 pixels (216 pixels per inch) and was equipped with a standard QWERTY touchscreen keyboard. The tablet was on a stalk located near the center console and was able to be used in various positions. Internet connection for the tablet was available via Wi-Fi to a portable hotspot. Throughout the driving session, the in-vehicle experimenter provided the participant with a series of navigation, email, and Web-browsing distraction tasks to complete. (Note: Labels and arrows for the task icons have been added to this image to enhance the reader's understanding.)

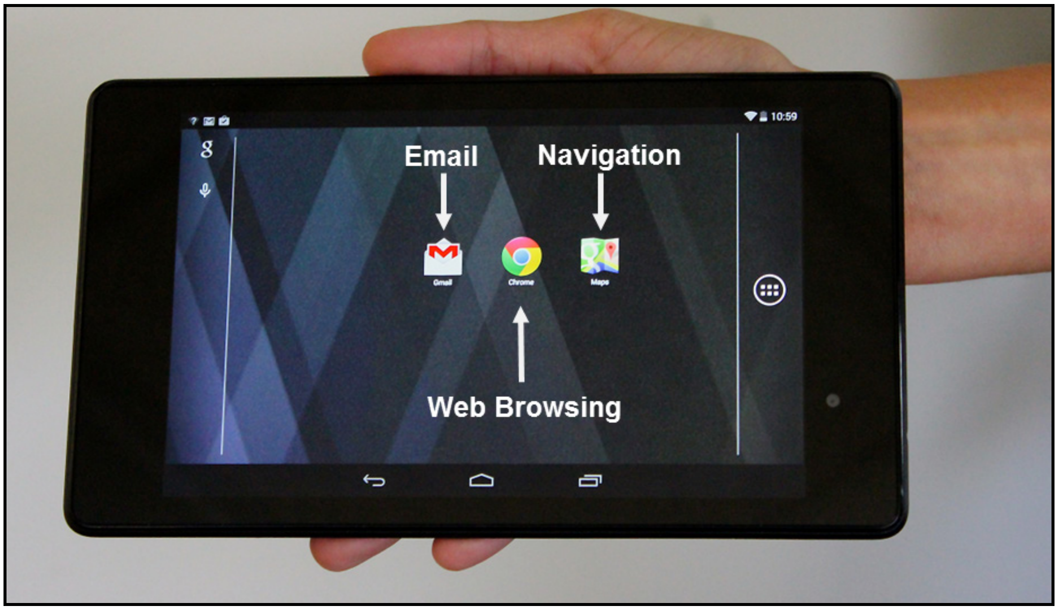



**Figure 2-5. Asus Nexus 7 Tablet Computer**

## Participants

Data were collected from 35 participants; however, 10 participants were considered invalid (i.e., session cancellation due to adverse weather, track closures, or technical issues associated with the prototype vehicle). The analysis presented in this chapter represents data from 25 participants (16 males, 9 females). The mean age of participants was 44.3 years old (standard deviation [S.D.] = 19.24), with ages ranging from 18 to 72 years old. Participants were categorized into four different age groups in order to ensure all age groups were represented: 18–24, 25–39, 40–54, and 55+; of which, five were between the ages of 18 and 24 (3 males, 2 females), six were between the ages of 25 and 39 (5 males, 1 female), four were between the ages of 40 and 54 (1 male, 3 females) and 10 were 55 and older (7 males, 3 females). However, age group was not a dependent variable for this study. Additional information regarding Experiment 1 participants can be found in Appendix B.

## Instruction and Training

Prior to proceeding to the test vehicle, the participants were provided with the tablet that they would be using in the vehicle. They were given a brief introduction to the types of tasks they would be performing, and were then allowed time for any questions, and to practice one of each of the different types of tasks, if needed. Participants were provided with a static orientation to the experimental vehicle, which included basic controls and the L2 automation features. Following this, participants received an on-track orientation consisting of four laps on the test track. During this training, participants were not given an explanation of the levels of automation nor were they given the impression that the system should allow them to divert attention from the roadway.

The purpose of this on-track orientation was to allow participants to acclimate to the vehicle and the test-track environment. Participants were asked to drive to the circle track entrance, enter the third travel lane, and maintain a speed of 60 mph. The first lap was completed under manual driving. During the second lap, ACC was activated, which allowed participants to release longitudinal control to the automated system while they maintained lateral control of the vehicle. The third and fourth laps were completed using the L2 automation (ACC plus lane centering), which allowed participants to release both longitudinal and lateral control of the vehicle. The prototype system clearly depicted when the system was ready for activation of the automation, and participants were told that they could activate the automation at their own discretion (i.e., when they felt comfortable). Following this, participants were provided with additional instructions pertaining to the experimental session and were given the opportunity to ask any additional questions. Note: Experimenters followed a set script and protocol in order to ensure consistency between experimenters.

## Study Session

Participants completed one driving session of approximately 90 minutes. Participants were instructed to begin interacting with a variety of non-driving tasks during the driving session upon activating the L2 automation.

Participants were presented with three types of non-driving tasks to complete using the tablet computer: navigation, email, and Web browsing. Participants were given 30 tasks in each category, described below, for a total of 90 different distraction tasks.

- Navigation. The navigation tasks required the participant to open an application of the device, choose the new destination option, and enter the address provided on the notecard in a printed number-street-city-state format.

- Email. The email tasks required the participant to compose an email using the tablet. Participants were asked to complete an email task from a list of 30 potential tasks. These tasks were the same for Experiments 1 and 2. A sample task which reflects average difficulty is the following: email clayton.ellis@lunar.com, ask when Jane's birthday is, and confirm that she will be six years old.

- Web browsing. The Web-browsing tasks required the participant to open the Web browser on the tablet. Participants were asked to determine the answer to a specific question or task that required searching or interacting with the Internet.

Each task was typed on a notecard. The in-vehicle experimenter provided the participant with a notecard and, upon completing the task, the participant passed the notecard back to the in-vehicle experimenter, who then provided the participant with a new notecard containing a different task. The tasks, which were presented in a random order, were similar in terms of the visual/manual demand required. The pace of these tasks was not forced but at the participants' leisure.

At approximately 5-minute intervals (in random values ranging from 2 to 8 minutes), participants were provided with unimodal (visual only) or multimodal (visual + haptic seat vibration) alerts (Cautionary, Imminent, or Staged) instructing them to take control of the vehicle. Participants were told that they could activate the automated system at their own discretion (e.g., when they felt comfortable doing so). Each participant experienced the Cautionary, Imminent, and Staged messages as both unimodal and multimodal alerts. All of the instances categorized as failures at the end of the imminent portion of the alert were prescribed and injected into the condition of interest using the experimenter console. They simulated conditions where the system captures a potential malfunction and warns the vehicle operator to regain control. Figure 2-6 provides a summary for the different alert types.

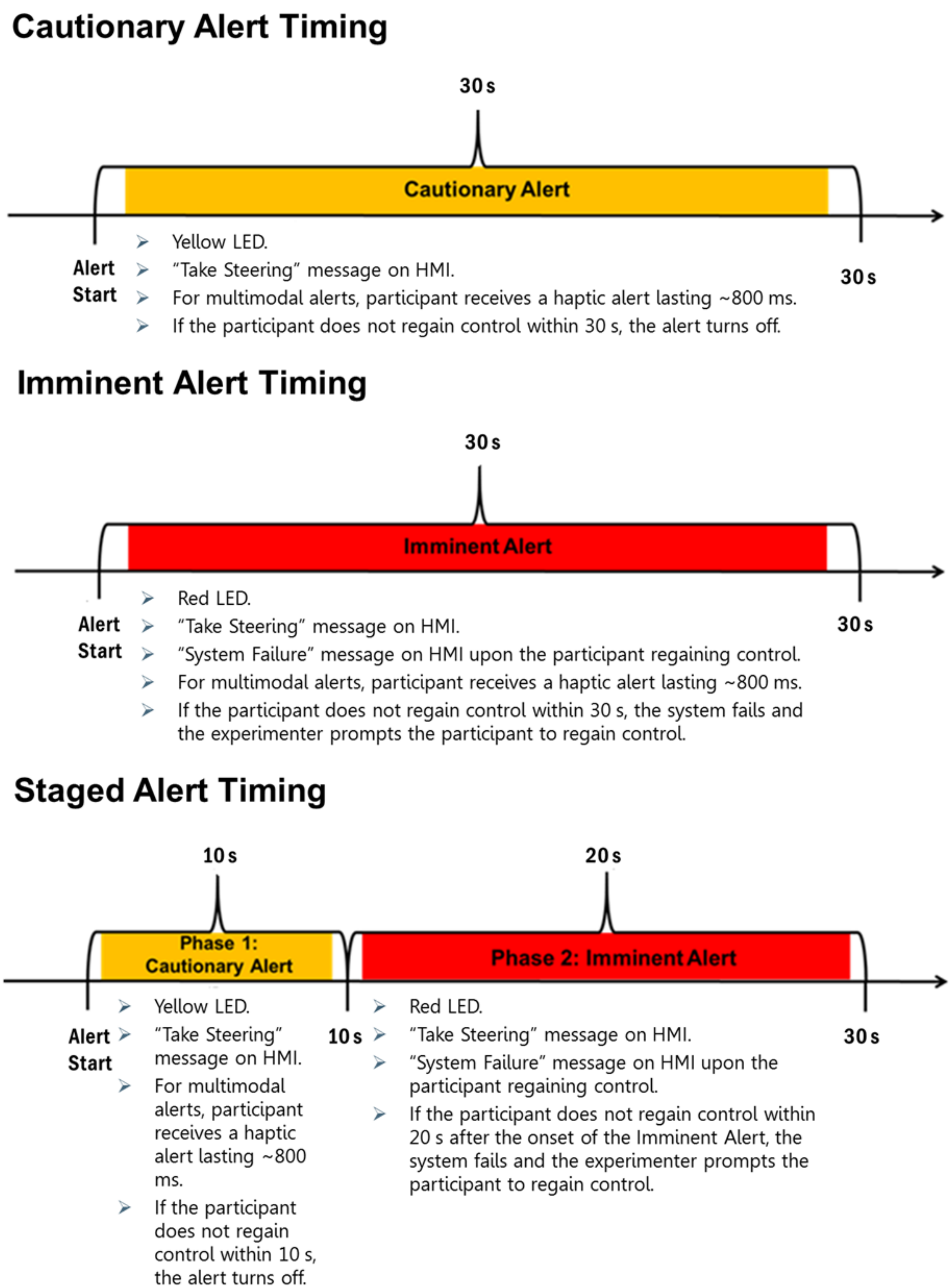



**Figure 2-6. Cautionary, Imminent, and Staged Take Control Alert Timelines for Experiment 1**

Participant responses to these system messages were measured in terms of response times using visual evidence from the DAS. As depicted in the figure above, the Imminent alerts and the Staged alerts that progressed to the imminent phase both resulted in a system failure that required the participants to regain control. However, the Cautionary messages resolved (i.e., the alert turned off) without requiring the participant to intervene.

There were six total combinations of alert type (Cautionary, Imminent, Staged) and alert modality (unimodal and multimodal) and six total orders of these combinations, which formed a Latin square. These orders, which were randomly assigned to participants, were replicated three times for each participant. Participants received a total of 18 system messages: six Cautionary alerts (three unimodal, three multimodal), six Imminent alerts (three unimodal, three multimodal), and six Staged alerts (three unimodal, three multimodal).

***Unexpected Lane Drift***

After receiving these 18 alerts, each participant received an Imminent multimodal alert coupled with an experimenter-triggered lane drift. This failure was injected when safe to do so (no other cars present in the surrounding lanes), simulating a failure of the lane centering system. This event was always presented as an Imminent multimodal alert. Upon triggering the alert, the vehicle would begin to drift across the left adjacent lane and would exit the lane in 3 to 5 s. If the car had completely entered the adjacent lane, the experimenter would have informed the participant to take control of the vehicle. However, all participants successfully gained control prior to entering the adjacent left lane. Following this event, the participant was instructed to exit the circle track and return to the preparation area. The participant was then interviewed, asked to complete the after-experience trust scales, and then provided with a debriefing as to the purpose of the study. Participants were given compensation for participation in the study. Figure 2-7 presents the timeline of events for Experiment 1.

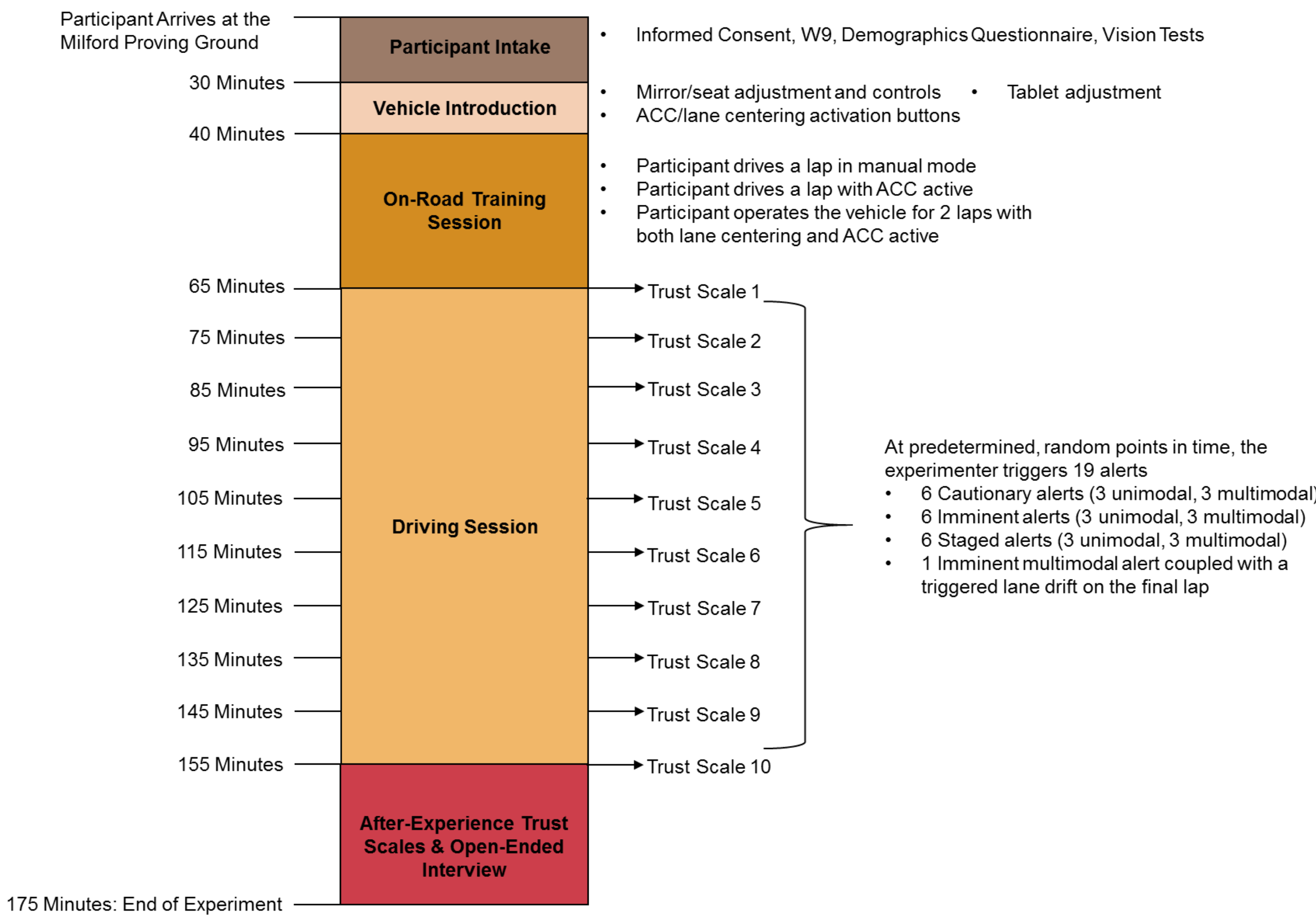



**Figure 2-7. Experimental Session Timeline for Experiment 1**

Unplanned events also occurred during the study. As the system used a camera to perform lane centering, there were instances in which environmental conditions (i.e., low sun angles) resulted in the system being unable to function properly; as a result, participants were sometimes instructed to take control of the vehicle.

# Results

It should be noted that in the analysis of this experiment, "unimodal" refers to an alert with a visual component, while "multimodal" refers to an alert with a visual component and a haptic component. Therefore, any significant differences between unimodal and multimodal can only be interpreted as differences between a visual alert and a visual + haptic alert. Interpretations cannot be generalized to any unimodal alert versus any multimodal alert.

There were a total of 505 alerts during the entire experiment, 408 valid and 97 invalid. Fifty-two of the invalid events were unplanned events, which (as explained above) were used as covariates in the analysis (if two unplanned events took place between event 2 and event 3 for a particular participant, for example, then the value of unplanned events for event 3 would be "2"). The remaining 45 invalid events were excluded from the analysis. Of these events, 22 were Staged alerts (6 multimodal, 16 unimodal), 18 Cautionary (10 multimodal and 8 unimodal), and 5 Imminent (all unimodal). The invalid Staged alerts occurred when the prototype automation system froze (3) or failed before the participant regained control (15). The invalid Cautionary alerts occurred when the system froze (2) or failed (16). All five Imminent invalid events occurred because the system failed before the participant regained control. The remaining 408 events were analyzed visually, descriptively, and statistically to determine if there were any significant effects of alert modality or alert type.

## Performance

A focus of this experiment was to investigate alert effectiveness at getting operators to regain control of the L2 automated vehicle. Performance was defined as the correct action taken by the participant in response to an event of interest. For the 408 valid instances when the participants were presented with an alert, only 18 (4.4 percent) were deemed as incorrect performance. All 18 of these are under the Cautionary alert, and 15 out of these 18 (83.3 percent) were unimodal Cautionary alerts.

During these 18 instances in which performance was incorrect, the participant reacted to the alert (e.g., looked forward) but did not perform any other action to regain control of the vehicle. Hence, for these instances there are missing values for time to regain control, time to activate the automation, time to release control, and time to resume a non-driving task. In addition, there were two instances in which a participant reacted to the alert, but performed only one other action. During one of these instances, the participant resumed a non-driving task, while in the other the participant regained control.

## Time to React to Alert

The alert's modality was found to have a significant effect on participants' time to react, $F(1, 223) = 162.87$, $p < .0001$. The time to react to an alert was, on average, 4.6 times significantly faster when a visual and haptic component was used (mean = 0.66 s, standard error [S.E.] = 0.03 s, n =

209, min = 0.04 s, max = 4.34 s) as compared to when the alerts had just a visual component (mean = 3.04 s, S.E. = 0.33 s, n = 197, min= 0.15 s, max = 31.04 s).

The alert's severity was not found to have a significant effect on participants' time to react, $p > .05$). Participants' mean time to react to Imminent alerts was 1.31 s (S.E. = 0.15 s, n = 139, min = 0.04 s, max = 13.81 s), their mean time to react to Staged alerts was 1.89 s (S.E. = 0.30 s, n = 130, min = 0.08 s, max = 16.19 s), and their mean time to react to Cautionary alerts was 2.26 s, S.E. = 0.38 s, n = 137, min = 0.15 s, max = 31.04 s). There was also no significant interaction of alert type and alert modality, $p > .05$. See Figure 2-8 for a graphical depiction and Appendix B, Analysis for Time to React to Alert, Statistical Analysis, for additional details on the statistical analysis.

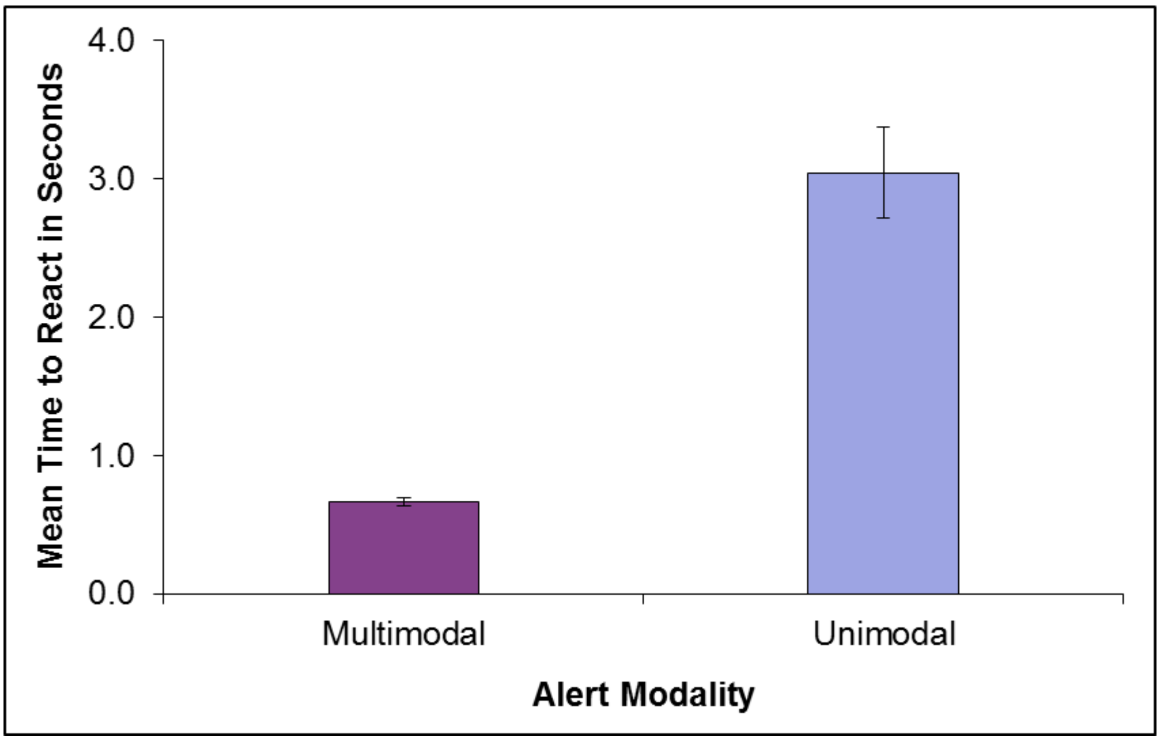



**Figure 2-8. Mean and Standard Error Bar Plots of Time to React by Modality for Experiment 1**

## Time to Regain Control

A significant interaction was found between alert type and alert modality, $F(2,132) = 4.33$, $p = .0151$. This indicates that the effect of alert type on time to regain control varied between visual alerts and visual + haptic alerts.

Within all three alert types, participants regained control significantly faster after an alert with a visual + haptic component than they did after an alert with just a visual component. The differences in means was smallest for Imminent alerts (M = 1.30 s, S.E. = 0.08 s, n = 73, min = 0.38 s, max = 4.18 s for visual + haptic vs. M = 2.90 s, S.E. = 0.36 s, n = 66, min = 0.67 s, max = 14.64 s for only visual, $t(91.2) = -7.78$, $p < .0001$) and largest for Cautionary alerts (M =

1.14 s, S.E. = 0.05 s, n = 63, min = 0.50 s, max = 2.51 s for visual + haptic vs. M = 6.29, S.E. = 1.04 s, n = 58, min = 0.33 s, max = 30.53 s for only visual, $t(64.4) = -7.61$, $p < .0001$).

Alert type did not have a significant effect if the alert had a visual and haptic component, $p > .05$. However, for alerts with only a visual component, participants regained control 2.2 times faster after an Imminent alert (M = 2.90 s, S.E. = 0.36, n = 66, min = 0.67 s, max = 14.64 s) as they did after a Cautionary alert (M = 6.29 s, S.E. = 1.04 s, n = 58 s, min = 0.33 s, max = 30.529 s). This result was statistically significant, $t(86.2) = 2.86$, $p = .0162$. (For additional details on the statistical analysis, see Figure 2-9 for a graphical depiction and Appendix B, Analysis for Time to Regain Control, Statistical Analysis.)

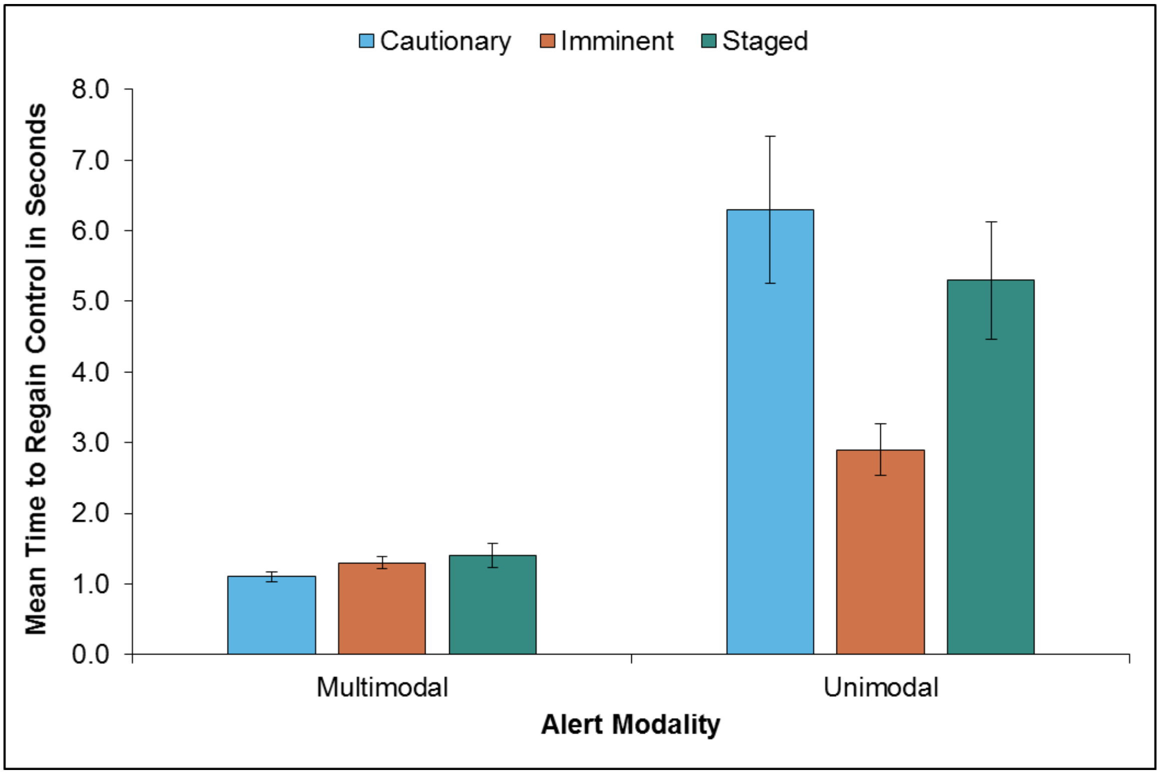



**Figure 2-9. Mean and Standard Error Bar Plots of Time to Regain Control, Stratified by Alert Modality and Alert Type for Experiment 1**

## Time to Activate the Automation

For time to activate the automation, only Imminent alerts (both with a visual + haptic component and with only a visual component) and Staged alerts that progressed to the imminent phase could be compared.

A significant difference between these three groups was found, $F(2, 116) = 5.76$, $p = .0041$. Specifically, when participants experienced an alert that was Imminent but had only a visual component, they took significantly longer, on average, to activate the automation (M = 4.04 s, S.E. = 0.30 s, n = 66, min = 0.84 s, max = 10.68 s) than they did if the alert was Imminent but had a visual and haptic component (M = 3.36 s, S.E. = 0.29 s, n = 73, min = 1.00 s, max = 13.18 s, $t(112) = 2.60$, $p = .0321$. Additionally, participants took longer after an Imminent visual alert than if the alert was Staged with only a visual component (M = 3.40 s, S.E. = 0.88 s, n = 16, min = 0.67 s, max = 15.18 s), $t(121) = 2.92$, $p = .0125$. (For additional details on the statistical analysis, see Appendix B, Analysis for Time to Activate the Automation, Statistical Analysis.)

## Time to Release Control of Steering

A significant interaction was found between alert type and alert modality, $F(2,346) = 4.44$, $p = .0124$). This indicates that the effect of alert type on time to release control varied between visual alerts and visual + haptic alerts.

Post hoc tests found that if the alert had a visual and haptic component, then the alert types differed significantly in their time to activate the automation. Specifically, if an alert had a visual and haptic component, then participants released control significantly faster when the alert was Imminent (M = 2.44 s, S.E. = 0.17 s, n = 73, min = 0.67 s, max = 11.52 s) than they did when the alert was Staged (M = 4.56 s, S.E. = 0.69 s, n = 69, min = 0.83 s, max = 44.22 s), $t(345) = -5.42$, $p < .0001$. Additionally, Imminent multimodal alerts resulted in significantly faster times to release control than if the alert was Cautionary (M = 3.99 s, S.E. = 0.43 s, n = 63, min = 1.34 s, max = 25.19 s), $t(356) = -4.72$, $p < .0001$. (For additional details on the statistical analysis, see Appendix B, Analysis for Time to Release Control of Steering, Statistical Analysis.)

## Time to Resume a Non-Driving Task

The alert's severity was found to have a significant effect on the time to resume a non-driving task $F(2,353) = 4.53$, $p = .0114$. The means of time to resume a non-driving task are similar between Cautionary alerts (M = 2.74 s, S.E. = 0.31 s, n = 121, min = 0.08 s, max = 21.69 s), Imminent alerts (M = 2.76 s, S.E. = 0.31 s, n = 138 s, min = 0.14 s, max = 35.04 s), and Staged alerts (M = 2.84 s, S.E. = 0.44, n = 129, min = 0.16 s, max = 34.7 s). Post hoc tests revealed that when participants had Imminent alerts, it took them longer to resume a non-driving task than when they experienced Staged alerts, $t(346) = 2.85$, $p = .0141$.

There was no significant effect of modality, $p > .05$. Participants' mean time to resume a non-driving task for visual + haptic alerts was 2.93 s (S.E. = 0.27 s, n = 205, min = 0.08 s, max = 21.69 s). For visual alerts, the mean was 2.62 s (S.E. = 0.32 s, n = 183, min = 0.14 s, max = 35.04 s). (For additional details, see Appendix B, Analysis for Time to Resume a Non-Driving Task, Statistical Analysis.)

## Operator Behavior Analysis Summary

After experiencing a multimodal alert, participants reacted and regained control faster than they did after a unimodal alert. Additionally, the alert type and alert modality may play a role in how quickly operators will cede control back to the system and resume a non-driving task. Participants took more time to activate the automation after a unimodal Imminent alert than they did after a multimodal Imminent alert and a unimodal Staged alert. After experiencing a multimodal alert, participants released control of the vehicle significantly faster if the alert was Imminent than they did if the alert was Cautionary or Staged. Meanwhile, they were slower to release control after a multimodal alert compared to a unimodal alert if the alert was Staged. Finally, participants took significantly longer to resume a non-driving task after an Imminent alert compared to a Staged alert.

Table 2-1 summarizes the results of the operator behavior analysis. Additional details regarding the analysis and results can be found in Appendix B.

**Table 2-1. Summary Table for Operator Behavior Analysis for Experiment 1**

| Variable | Alert Modality | Alert Type | Takeaways |
|---|---|---|---|
| **Time to React** | **Significant** | Not Significant | Operators may react faster after a visual + haptic alert than after a visual alert. |
| **Time to Regain Control** | **Significant** | **Significant** | Operators may regain control faster after a visual + haptic alert than they would after a visual alert. Additionally, when operators experience a visual alert, they may regain control faster after an Imminent alert than after a Cautionary alert. |
| **Time to Activate the Automation** | **Significant** | **Significant** | When operators experience an Imminent alert with a visual component, they may take more time to activate the automation than they would if the Imminent alert had visual and haptic components. |
| **Time to Release Control of Steering** | **Significant** | **Significant** | If an alert has a visual and a haptic component, operators may release control faster if the alert is also Imminent. |
| **Time to Resume Non-Driving Task** | Not Significant | **Significant** | Operators may take more time to resume a non-driving task after experiencing an Imminent alert than they would after experiencing a Staged alert, but not necessarily a Cautionary alert. |
| **Performance** | N/A | N/A | Most operators may regain control of the vehicle before 30 s have passed. However, there may be times when they will fail to do so if the alert is Cautionary, especially if the alert only has a visual component. |

## Eye-glance Behavior

This analysis helped to determine if the alert type affected eye-glance behavior. The time it took for participants to look forward upon receiving the alert was calculated. As shown in Figure 2-10, it was apparent that participants were faster to look forward during the multimodal conditions compared to the unimodal conditions, and this trend held across all three alert types. During the Cautionary, Imminent, and Staged multimodal alerts, participants took a mean of 0.84 s, 0.88 s, and 0.78 s, respectively, to look forward. In comparison, during the Cautionary, Imminent, and Staged unimodal alerts, participants took a mean of 4.06 s, 2.23 s, and 3.70 s, respectively, to look forward.

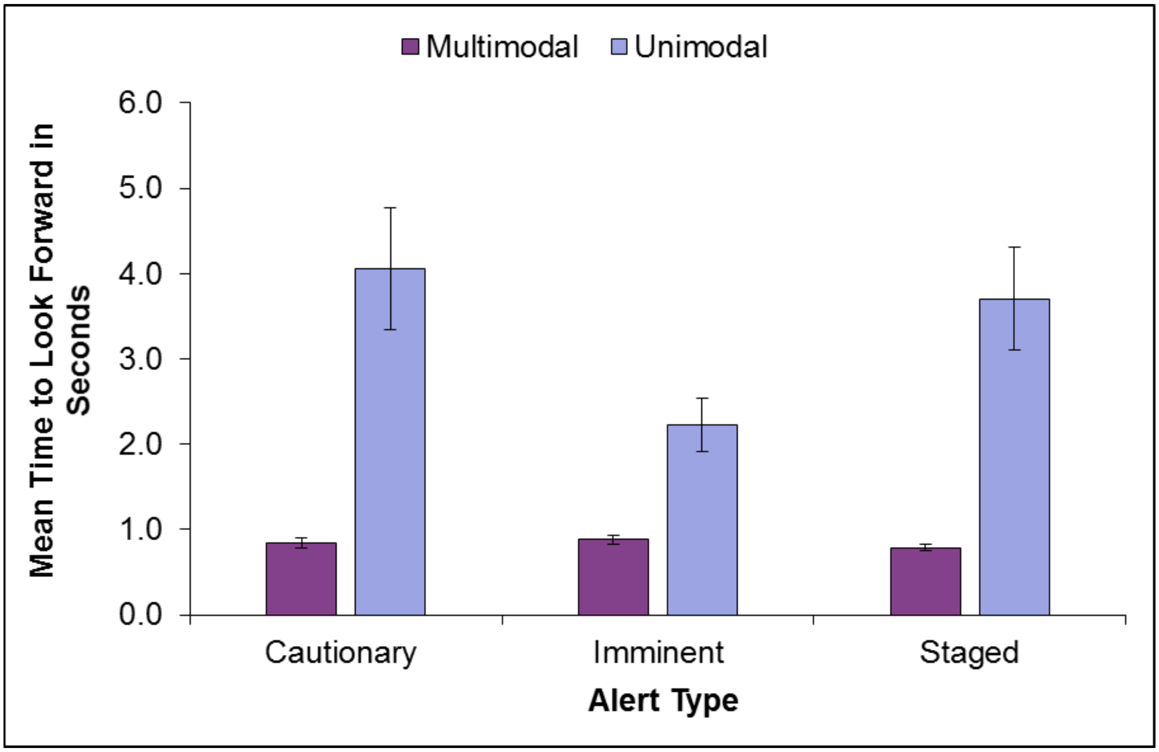



**Figure 2-10. Mean and Standard Error Plots of Time to Look Forward by Modality for Experiment 1**

***Unexpected Lane Drift***

At the end of each experimental run, a lane drift was enacted that moved the vehicle from the center of the lane. At the same time, a multimodal, Imminent alert was presented. This resulted in 22 participants experiencing valid lane-drift events. During this final event in the experiment, all participants were engaged in a non-driving task. As such, none of them were looking forward at the time of the event, none had their hands on the wheel, and only one had his/her foot on the pedal. Each participant would then perform all of the normal functions (e.g., reacting, regaining control, etc.) and obtained values for all five operator variables (one participant did not resume a non-driving task). Because the lane drift occurred at the end of every experimental run, and was associated with a multimodal, Imminent alert, any comparison with previous events of the lane-drift effect would be confounded by the alert type, alert modality, and time effect. Therefore, no statistical comparison was made between the lane-drift events and the previous events. However, the lane-drift events were analyzed descriptively.

As demonstrated in Figure 2-11, participants took little time to react to the alert when there was a lane drift present. The mean time to react was 0.5 s (S.E. = 0.03 s), about 0.2 s lower than the mean time to react for multimodal, Imminent alerts without the lane drift (mean = 0.7 s, S.E. = 0.03 s). Similarly, the participants successfully regained control soon after the alert, with a mean time to regain control of 1.2 s (S.E. = 0.1 s), which was about 0.1 s less than the mean time to regain control for multimodal, Imminent alerts without the lane drift (mean = 1.3 s, S.E. = 0.1 s).

The mean time to activate the automation was 2.7 s (S.E. = 0.3 s), which was approximately 0.7 s less than the mean time to activate the automation for Imminent multimodal alerts without a lane drift (mean = 3.4 s, S.E. = 0.3 s). For time to release control, the mean time was 2.7 s (S.E. = 0.3 s), which was approximately 0.3 s higher than the mean time to release control for multimodal, Imminent alerts without a lane drift (mean = 2.4 s, S.E. = 0.2 s). Finally, the mean time to resume a non-driving task of 2.2 s (S.E. = 0.4 s) for lane-drift events was approximately 0.4 s lower than for other multimodal, Imminent alerts (mean = 2.6 s, S.E. = 0.3 s). The differences do not appear to be substantial between multimodal, Imminent alerts without a lane drift and those with a lane drift. However, it must be noted that all lane-drift events occurred at the end of the experiment, which was not the case for non-lane-drift events.

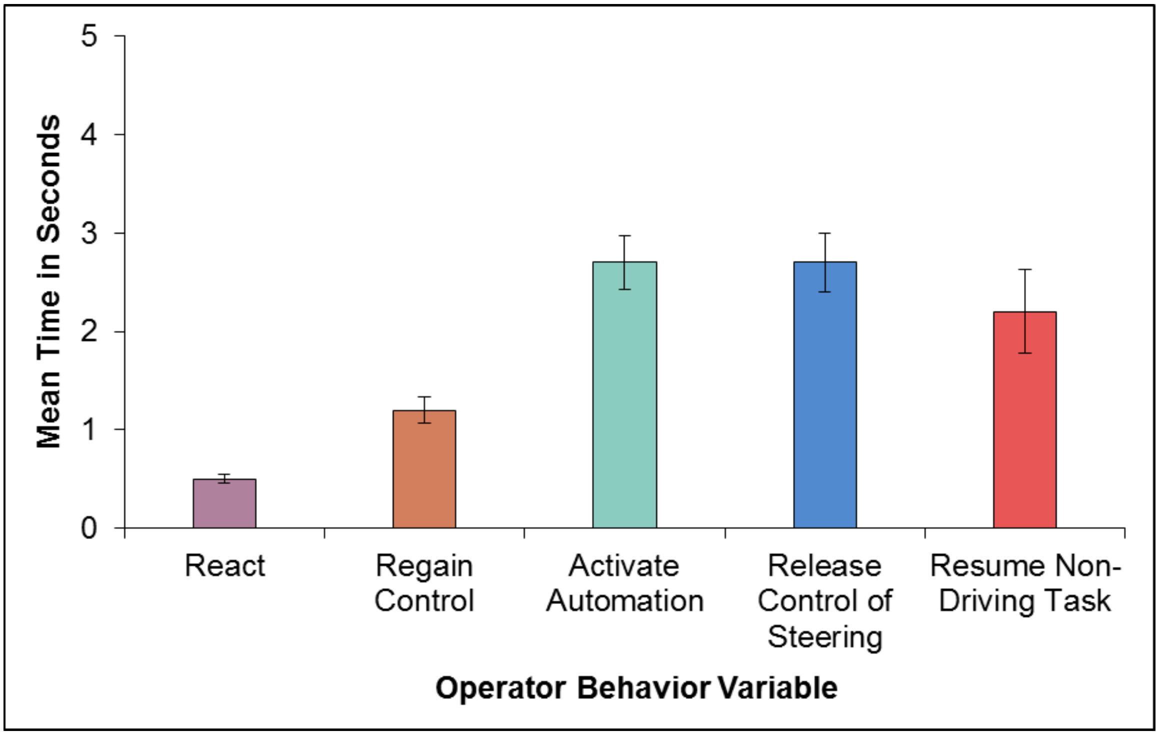



**Figure 2-11. Mean and Standard Error Bar Plots for Operator Variables After Lane Drift Events in Experiment 1**

## Trust Scales

Participants reported that they trusted the system more often than not during the experiment. In 202 instances (81.1 percent), participants stated that they either “moderately agreed” or “strongly agreed” that they trusted the system. In only 17 instances did participants disagree that they trusted the system. However, trust did not significantly change over time, $p > .05$.

Figure 2-12, which displays the change in mean rating per participant over time, does not indicate any strong change in rating one way or the other. The highest mean rating is at the 5$^{th}$ time point, with a mean rating of 6.24, while the first and fourth time points have the lowest mean rating of 6. The mean rating at the 10$^{th}$ time point was about 6.04, indicating that across the experiment, there were was a mean improvement in rating of only 0.04. Thus, there does not appear to be a strong relationship between time and rating. (For additional details on the statistical analysis, see Appendix B, Trust over Time, Statistical Analysis.)

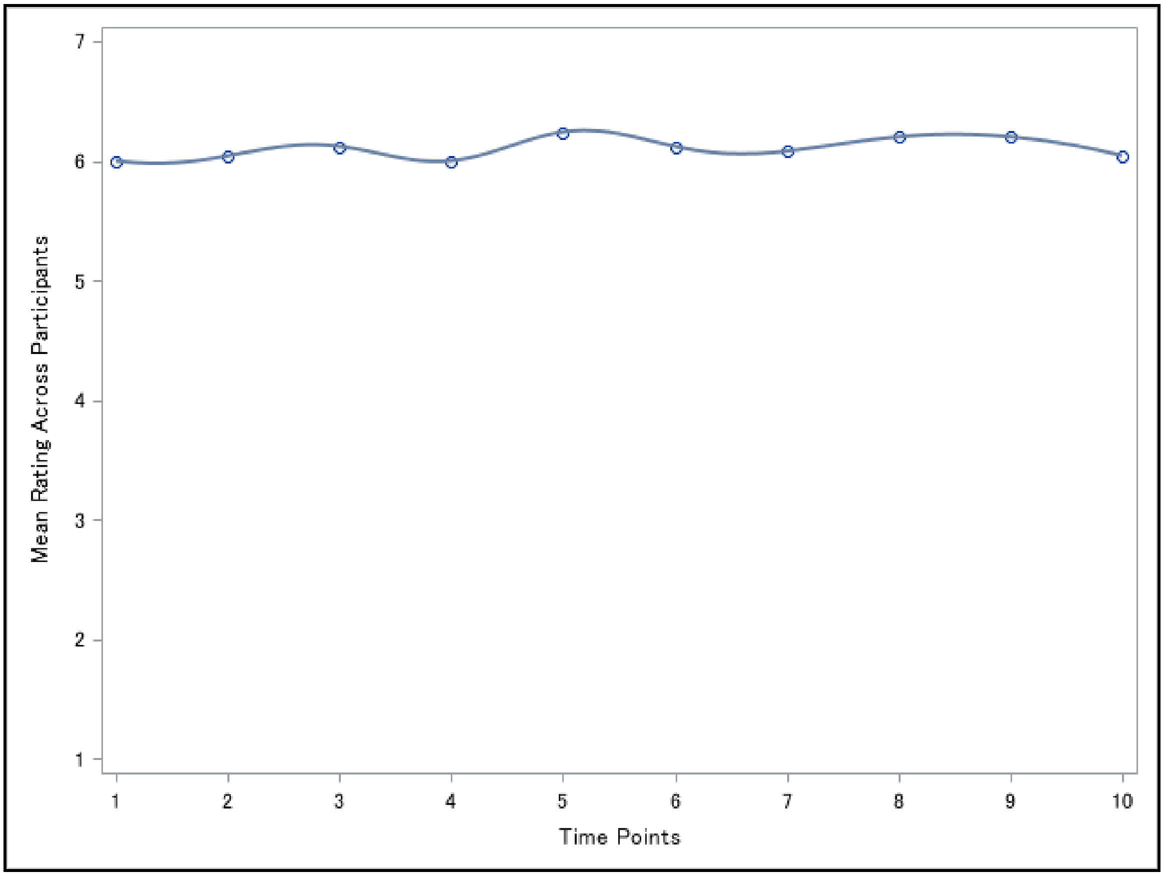



**Figure 2-12. Change in Mean Trust Rating Over Time for Experiment 1**

***After-experience Trust Scales***

A survey of six statements relating to how much the participant trusted the system was given at the end of the experiment (i.e., not at the end of each session, only when all three sessions had concluded). The statements were:

TS1. I can rely on the automated system to function properly while I am doing something else.

TS2. The automated system provided the alerts when needed.

TS3. The automated system gave false alerts.

TS4. The automated system is dependable.

TS5. I am familiar with the automated system.

TS6. I trust the automated system.

All responses were based on a 7-point Likert-type scale with options ranging from "1" for strongly disagree to "7" for strongly agree. All but one of the trust statements had positive wording, so a higher score indicates a higher level of trust. Statement 3, "The automated system gave false alerts," had negative wording, so the responses to this statement were recoded such that "1" =

“7”, “2” = “6”, etc. A mean and standard error bar plot for all responses is displayed in Figure 2-13.

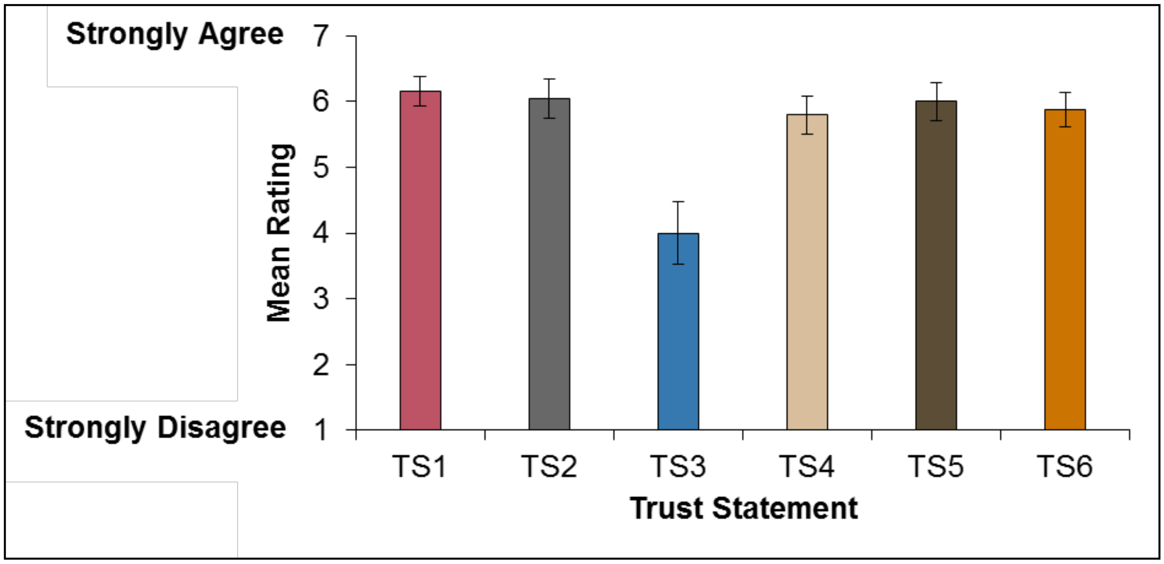



**Figure 2-13. Mean Rating for After-experience Trust Scales for Experiment 1**

***After-experience Interview***

Following the in-vehicle experience, participants were asked a series of six questions. The interviews were audio recorded as a backup to researcher notes. Digital audio recordings were made into transcripts, but no individual names were attached to any comments. Responses for 25 participants were recorded and transcribed. The transcripts and researcher notes served as the basis for a qualitative content analysis which was completed using a framework analysis method. The results of this analysis were used to help researchers to potentially understand participant behavior that varies from the norm. A full discussion of the methods used to complete the framework analysis as well as detailed findings can be found in Appendix F.

Generally, participants indicated that they had a positive experience and seemed impressed by and confident in the system. The majority of participants reported that they were comfortable with the system (Figure 2-14). Those who did not express a specific comfort level provided general descriptions of their comfort. These general expressions were grouped together as “other.” Note, participants were not asked to describe their comfort levels using a Likert scale.

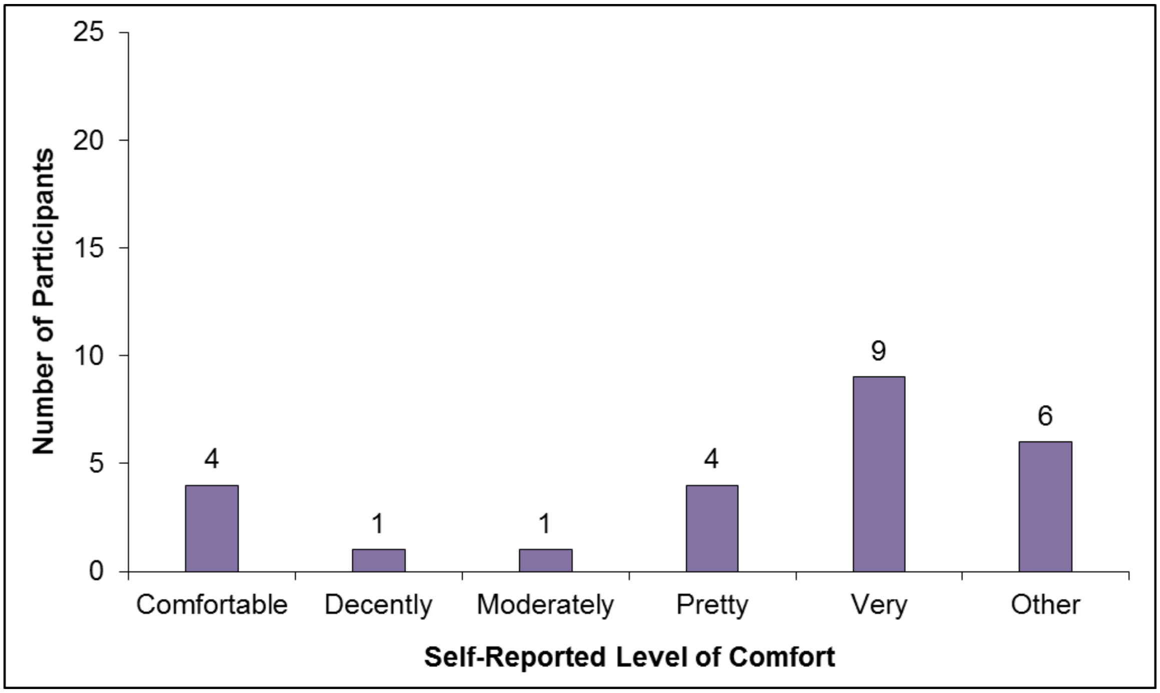



**Figure 2-14. Experiment 1 Participants' Self-reported Level of Comfort with the Automated System**

When asked how quickly (specific or relative time) they achieved their level of comfort, the majority of participants reported that their ultimate level of comfort was achieved in 15 minutes or less (Figure 2-15). Of the seven participants who provided a temporally vague response to the question, two replied that they became comfortable with the system very quickly, three replied pretty quickly, and one replied that after witnessing the system react to other vehicles a couple of times, he felt very comfortable with the system. The final participant indicated discomfort associated with the alerts (which were part of the test condition) and with the presence of other cars on the test track.

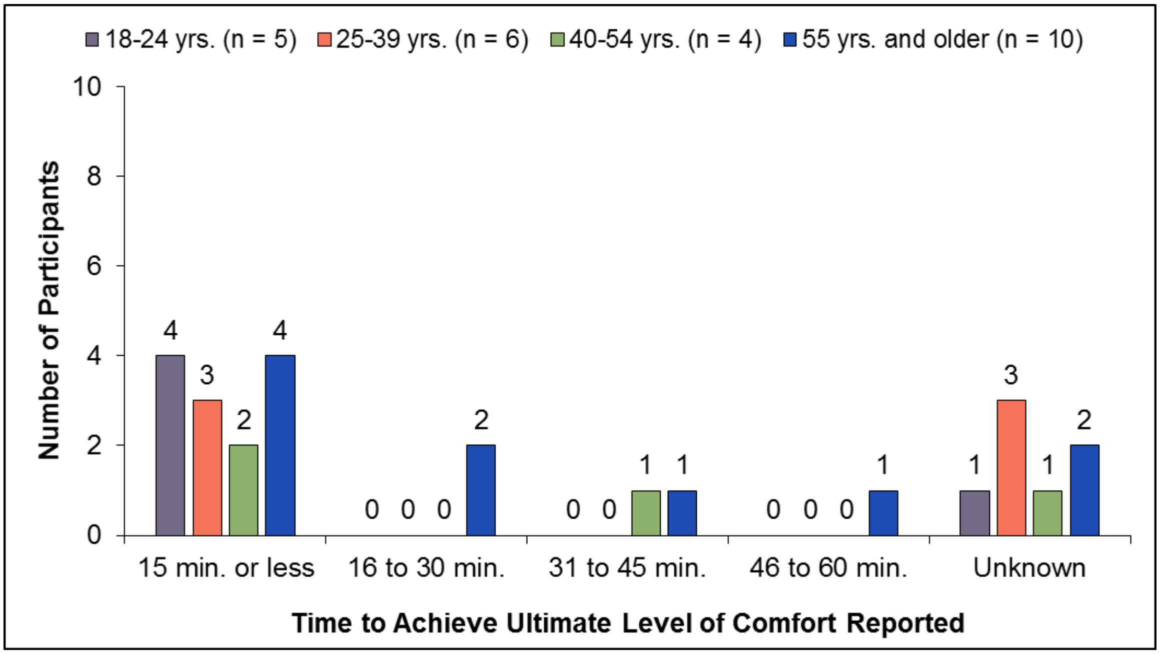



**Figure 2-15. Experiment 1 Participants' Time to Achieve Comfort with the System by Age**

In terms of the alerts presented, responses suggest that participants would prefer a multimodal alert system. For example, when asked what factors contributed to their comfort levels, 15 participants specifically commented on the alerts. Participants especially noted that they liked the haptic (9 responses) and visual alerts (10 responses). Of those participants, seven expressed that they wanted and/or expected the haptic alerts to be presented consistently in combination with auditory or visual alerts or when a system fault occurs. Several participants (5 responses) indicated that they wanted and/or expected consistent visual alerts. Participants also recommended the inclusion of an auditory alert (4 responses). Similar suggestions were made in the context of design suggestions.

# Experiment 1 Summary

The following tables (Table 2-2 through Table 2-7) summarize the significant differences between relevant combinations of alert types and modalities for each dependent variable. In each table, a particular combination of alert type and alert modality is compared to all other relevant combinations. For example, in Table 2-2, Cautionary Multimodal is the combination of interest, and it is directly compared to Imminent Multimodal and Staged Multimodal because they share the same modality; it is also compared to Cautionary Unimodal because they share the same alert type. Similar tables are supplied for each combination of alert type and alert modality. When there is no significant difference between the combination of interest and the other relevant combinations, this is designated as "Not Significant." "Significant" differences are also designated as such.

**Table 2-2. Mean Values for Cautionary Multimodal, Compared to Mean Values for Other Relevant Alert Type/Alert Modality Combinations in Experiment 1**

| Dependent Variable | Cautionary Multimodal | Imminent Multimodal | Staged Multimodal | Cautionary Unimodal |
|---|---|---|---|---|
| **Time to React** | 0.7 s | 0.7 s<br>Not Significant | 0.6 s<br>Not Significant | 3.7 s<br>Significant |
| **Time to Regain Control** | 1.1 s | 1.3 s<br>Not Significant | 1.4 s<br>Not Significant | 6.3 s<br>Significant |
| **Time to Release Control** | 4.0 s | 2.4 s<br>Significant | 4.6 s<br>Not Significant | 3.6 s<br>Not Significant |
| **Time to Resume a Non-Driving Task** | 3.1 s | 2.6 s<br>Not Significant | 3.1 s<br>Not Significant | 2.3 s<br>Not Significant |

**Table 2-3. Mean Values for Cautionary Unimodal, Compared to Mean Values for Other Relevant Alert Type/Alert Modality Combinations in Experiment 1**

| Dependent Variable | Cautionary Unimodal | Imminent Unimodal | Staged Unimodal | Cautionary Multimodal |
|---|---|---|---|---|
| **Time to React** | 3.7 s | 2.0 s<br>Not Significant | 3.4 s<br>Not Significant | 0.7 s<br>Significant |
| **Time to Regain Control** | 6.3 s | 2.9 s<br>Significant | 5.3 s<br>Not Significant | 1.1 s<br>Significant |
| **Time to Release Control** | 3.6 s | 2.8 s<br>Not Significant | 3.1 s<br>Not Significant | 4.0 s<br>Not Significant |
| **Time to Resume a Non-Driving Task** | 2.3 s | 2.9 s<br>Not Significant | 2.6 s<br>Not Significant | 3.1 s<br>Not Significant |

**Table 2-4. Mean Values for Imminent Multimodal, Compared to Mean Values for Other Relevant Alert Type/Alert Modality Combinations in Experiment 1**

| Dependent Variable | Imminent Multimodal | Cautionary Multimodal | Staged Multimodal | Imminent Unimodal |
|---|---|---|---|---|
| **Time to React** | 0.7 s | 0.7 s<br>Not Significant | 0.6 s<br>Not Significant | 2.0 s<br>Significant |
| **Time to Regain Control** | 1.3 s | 1.1 s<br>Not Significant | 1.4 s<br>Not Significant | 2.9 s<br>Significant |
| **Time to Release Control** | 2.4 s | 4.0 s<br>Significant | 4.6 s<br>Significant | 2.8 s<br>Not Significant |
| **Time to Resume a Non-Driving Task** | 2.6 s | 3.1 s<br>Not Significant | 3.1 s<br>Significant | 2.9 s<br>Not Significant |

**Table 2-5. Mean Values for Imminent Unimodal, Compared to Mean Values for Other Relevant Alert Type/Alert Modality Combinations in Experiment 1**

| Dependent Variable | Imminent Unimodal | Cautionary Unimodal | Staged Unimodal | Imminent Multimodal |
|---|---|---|---|---|
| **Time to React** | 2.0 s | 3.7 s<br>Not Significant | 3.4 s<br>Not Significant | 0.7 s<br>Significant |
| **Time to Regain Control** | 2.9 s | 6.3 s<br>Significant | 5.3 s<br>Not Significant | 1.3 s<br>Significant |
| **Time to Release Control** | 2.8 s | 3.6 s<br>Not Significant | 3.1 s<br>Significant | 2.4 s<br>Not Significant |
| **Time to Resume a Non-Driving Task** | 2.9 s | 2.3 s<br>Not Significant | 2.6 s<br>Significant | 2.6 s<br>Not Significant |

**Table 2-6. Mean Values for Staged Multimodal, Compared to Mean Values for Other Relevant Alert Type/Alert Modality Combinations in Experiment 1**

| Dependent Variable | Staged Multimodal | Cautionary Multimodal | Imminent Multimodal | Staged Unimodal |
|---|---|---|---|---|
| **Time to React** | 0.6 s | 0.7 s<br>Not Significant | 0.7 s<br>Not Significant | 3.4 s<br>Significant |
| **Time to Regain Control** | 1.4 s | 1.1 s<br>Significant | 1.3 s<br>Not Significant | 5.3 s<br>Significant |
| **Time to Release Control** | 4.6 s | 4.0 s<br>Not Significant | 2.4 s<br>Significant | 3.1 s<br>Significant |
| **Time to Resume a Non-Driving Task** | 3.1 s | 3.1 s<br>Not Significant | 2.6 s<br>Significant | 2.6 s<br>Not Significant |

**Table 2-7. Mean Values for Staged Unimodal, Compared to Mean Values for Other Relevant Alert Type/Alert Modality Combinations in Experiment 1**

| Dependent Variable | Staged Unimodal | Cautionary Unimodal | Imminent Unimodal | Staged Multimodal |
|---|---|---|---|---|
| **Time to React** | 3.4 s | 3.7 s<br>Not Significant | 2.0 s<br>Not Significant | 0.6 s<br>Significant |
| **Time to Regain Control** | 5.3 s | 6.3 s<br>Not Significant | 2.9 s<br>Not Significant | 1.4 s<br>Significant |
| **Time to Release Control** | 3.1 s | 3.6 s<br>Not Significant | 2.8s<br>Not Significant | 4.6 s<br>Significant |
| **Time to Resume a Non-Driving Task** | 2.6 s | 2.3 s<br>Not Significant | 2.9 s<br>Significant | 3.1 s<br>Not Significant |

Table 2-8 describes comparisons made for time to activate the automation. Because participants only needed to activate the automation when the alert was in its imminent phase, only three combinations of alert type and alert modality (Imminent Unimodal, Imminent Multimodal, and Staged Multimodal) presented valid samples for time to activate the automation. All three of these combinations are compared with each other in the table below. If there is no significant difference between two particular alert types, they will share a common letter. Since Imminent Unimodal was significantly different from both Imminent Multimodal and Staged Multimodal, Imminent Unimodal is assigned the letter "A", while Imminent Multimodal and Staged Multimodal were assigned "B".

**Table 2-8. Mean Values for Time to Activate the Automation, Compared Between Relevant Alert Type/Alert Modality Combinations in Experiment 1**

| Dependent Variable | Imminent Unimodal | Imminent Multimodal | Staged Multimodal |
|---|---|---|---|
| Time to Activate the Automation | 4.0 s<br>A | 3.4 s<br>B | 3.5 s<br>B |

# Chapter 3 Level 2 Automated Vehicle, Experiment 2: System Prompt Effectiveness over Time

## Purpose

The goal of Experiment 2 was to examine how operators in a mixed-traffic (involving surrounding legacy vehicles, Level 0) simulated highway driving condition recognize and respond to attention state prompts over time. To that end, the effects of attention state prompts on the overall operator behavior over time were analyzed. As stated in Chapter 1, the ultimate goal was to ascertain how operators interact with automated vehicles and determine how automated vehicle technology can best support safe driving. For this study, two aspects of the Level 2 operational definition (NHTSA, 2013) are of importance and set the tone for the experimental design and scenarios of interest:

- "The driver is still responsible for monitoring the roadway and safe operation and is expected to be available for control at all times and on short notice."

- "The system can relinquish control with no advance warning and the driver must be ready to control the vehicle safely."

Based on these operational definitions, Experiment 2 was designed to look at the effectiveness of a mechanism to prompt operators to monitor the driving environment.

## Method

For Experiment 2, a single, long-exposure experiment was conducted. Participants were provided with a brief familiarization with the vehicle and its operation, followed by three 60-minute experimental sessions. During the sessions, participants were given tasks to be completed using a tablet computer. During these tasks, participants received prompts based on their predetermined prompt condition (either 2-second, 7-second, or No Prompts). For the 2- or 7-second prompt conditions, participants received prompts after periods of inattention for the corresponding amount of time. Participants given the No Prompts condition did not receive any prompts and they were free to behave as they thought was appropriate.

In addition to these prompts, at a random time during one predetermined session, the participant received an alert for a surprise left lane drift, consisting of a haptic seat alert and a flashing red LED. In a different predetermined session, the participant experienced a surprise lane drift with no alert, which consisted only of a left lane drift without any alert and with the prompting system disabled. The experimenter-injected lane drift was used to simulate a lane-keeping performance issue combined with a failure of the prompting system. Note that, to the participants with the 2-second and 7-second prompt conditions, the alert that they received along with the lane drift was indistinguishable from the prompts that they had been receiving based on their attention state. Participants' reactions to these prompts, alerts, and lane drifts, both in duration and method of response, were examined in this experiment.

# Experimental Design

The study was performed as a 3 x 3 x 3 mixed factorial design. Each participant completed three successive driving sessions, and each session included one of the following: a lane drift with an alert, a lane drift without an alert, or no lane drift. Participants experienced each of these conditions once during the experiment. In addition, there were also three different prompt conditions that were used with the driver monitoring system, and each participant experienced only one prompt condition, either 2-second, 7-second, or No Prompts. The prompt timing was based on previous distraction research (2-second prompts) (e.g., Klauer et al., 2006) and expert opinion (7-second prompts). Additionally, the study was designed to mimic worst-case scenarios when conditions for monitoring the roadway are decreased.

## Independent Variables

This research design used three independent variables, detailed below.

### *Prompt Condition*

This is a between-subjects variable with three levels: 2-second, 7-second, and No Prompts. According to the Level 2 automation definition, the operator is expected to be available for control at all times and on short notice. The intention of the timing of these alerts was to provide a prompt once the participant's attention had been off-road for the threshold determined by their prompt condition.

- 2-second. The participant was provided with a progressive experimental prompt after attention state had been off-road for 2 s while engaged in the non-driving-related tasks.
- 7-second. The participant was provided with a progressive experimental prompt after attention state had been off-road for 7 s while engaged in the non-driving-related tasks.
- No Prompts. No experimental prompts were presented.

### *Event Type*

This is a within-subject variable with three levels.

- Alert. The system provided a visual and haptic indication that it was not providing L2 automated driving. The vehicle began the surprise lane drift and entered the adjacent lane after providing the alert.
- No Alert. The system did not indicate that it had deactivated or any problem was present. The vehicle began the surprise lane drift (occurring in 3 to 5 s) and entered the adjacent lane with no alert provided. The prompting system was also disabled for this condition, simulating a failure.

- No lane drift was injected.

### *Session Number*

This is a within-subject variable with three levels:

- Session 1
- Session 2
- Session 3

## Dependent Variables

This research design included nine dependent variables, detailed below.

### *Operator Behavior*

- **Time to react** (in seconds)
  This variable was defined as the time from activation of the lane drift, or start of the prompt, at which point the participant performs an action that could be considered a reaction (e.g., look forward).

- **Time to regain control** (in seconds)
  This variable was defined as the time which begins at the activation of the lane drift, or start of the prompt, to the point at which the participant assumes manual control of the vehicle by grasping the steering wheel.

- **Performance**
  This variable was defined as the correct action taken by the participant in response to an event of interest (e.g., the participant assumes control of the vehicle when signaled).

- **Time to activate the automation** (in seconds)
  This variable was defined as the time from autonomy availability as indicated by the system until the participant activates the automation.

- **Time to release control of steering** (in seconds)
  This variable was defined as the time from successfully activating the automation until the participant removes both hands from the steering wheel.

- **Time to resume non-driving task** (in seconds)
  This variable was defined as the time from successfully activating the automation until the participant resumes interaction with a non-driving-related task.

Figure 3-1 illustrates the sequence in which the dependent variables pertaining to operator behavior were measured.

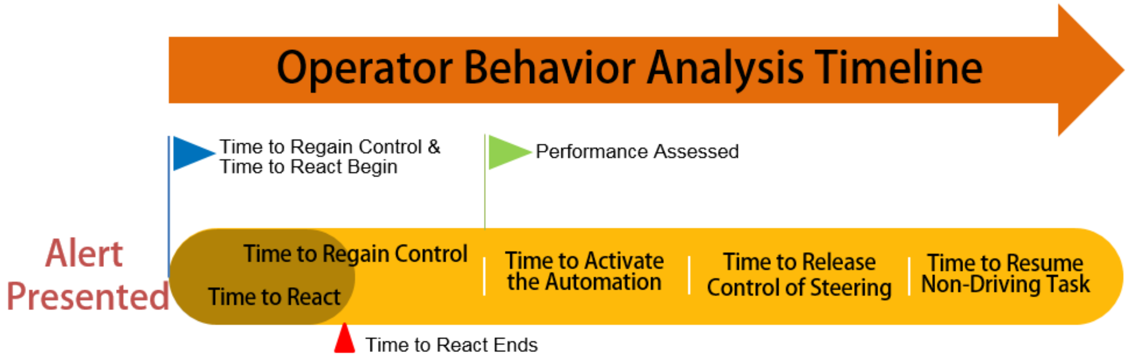



**Figure 3-1. Sequence of the Experiment 2 Dependent Variables Pertaining to Operator Behavior**

***Eye-glance Behavior***

Eye-glance behavior was sampled 10 s prior to and 60 s after the lane drift; 15 s prior to and after the randomly sampled prompts; and, for the No Prompts condition, 15 s prior to and after a selected point in time.

- Monitoring rate. This was defined as the percentage of driving-related glance time in a given time interval. (Driving-related glances pertain to glances made to the driving environment. These include the following glances: forward, left and right windshield, left and right windows, left and right side mirrors, rear-view mirror, HMI, and the instrument cluster.)
- Non-driving-related glances. This was defined as the number and duration of non-driving-related glances in a given time period. (Non-driving-related glances do not pertain to the driving environment and include glances to the following: smartphone, tablet computer, passenger, hygiene-related glances, etc.)

## Participants' Subjective Assessment

A trust scale (Appendix E) was administered 7 times in each experimental session at 10-minute intervals (7 times per 3 experiments, resulting in 21 trust ratings). Participants were asked to rate their trust in the ability of the automation to function properly while they engaged in non-driving tasks using a 7-point Likert-type scale, in which 1 corresponded with “strongly disagree” and 7 corresponded with “strongly agree.”

In addition to the 21 trust ratings collected throughout the experimental session, participants were asked to complete an after-experience trust scale and an after-experience satisfaction scale (Appendix E) after each of the three driving sessions, and an open-ended interview (Appendix F) after completing all the driving sessions.

# Facility

As was the case for Experiment 1, Experiment 2 was also conducted at GM's Milford Proving Ground circle track in Milford, Michigan (see Figure 2-2). Experiment 2 was conducted in Lane 2, which allowed speeds between 48 and 80 km/h (30 and 50 mph). As explained in the previous chapter, the circle track has been used in previous studies on automated driving to simulate freeway conditions and is restricted to approved studies and approved trained drivers. Therefore, other non-study traffic was present during the experiment.

## Vehicle

A 2010 model year Cadillac SRX equipped with a prototype L2 automated driving system was used as the experimental vehicle for Experiment 2 (Figure 3-2.). As part of the automated driving system, several HMI components were installed. These included an instrument panel binnacle-mounted screen providing information on the automated driving system, and two steering wheel buttons to control the automation: one ACC button and one button for the lane-centering system, a prototype automated vehicle system.

**Figure 3-2. 2010 Cadillac SRX with a Prototype L2 System Used in Experiment 2**

The vehicle was instrumented with Virginia Tech Transportation Institute's (VTTI) DAS. The DAS was connected to the prototype automated driving system via Ethernet. The variables collected by the DAS included status of the automation, vehicle speed, and lane position. In addition, the following video views were collected (shown in Figure 3-3):

- Operator's face
- OTS view
- Forward roadway
- Exterior left rear
- Foot (pedal area)
- HMI

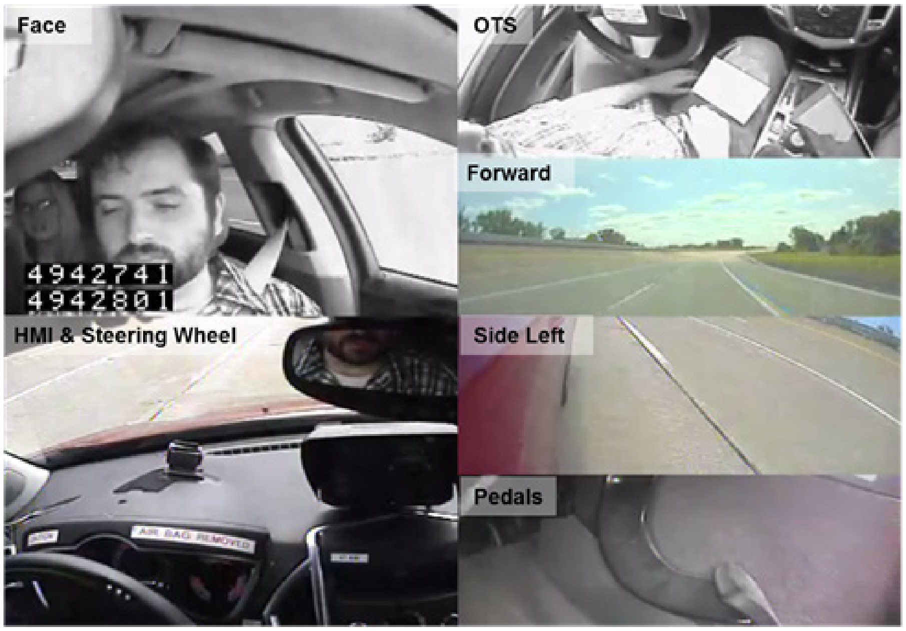



**Figure 3-3. VTTI's DAS Camera Views for Experiment 2**

## Tablet Computer

An Asus Nexus 7 tablet computer, identical to the one used in Experiment 1, was used for this study in order to provide opportunities for the participants to perform non-driving tasks while the vehicle was in Level 2 automation (see Figure 2-5). Internet connection for the tablet was available via Wi-Fi to a portable hotspot. Throughout the driving session, the in-vehicle experimenter provided the participant with a series of navigation, email, and Web-browsing distraction tasks to complete. Participants completed up to 90 different non-driving-related tasks, 30 in each category. The tasks were presented in a random order.

## Participants

Data were collected from 56 participants (28 males, 28 females) with a mean age of 41 years old (S.D. = 16.3), ranging from 18 to 72 years old. Participants were categorized into four different age groups in order to ensure all age groups were represented: 18–24, 25–39, 40–54, and 55+; of which, 14 were between the ages of 18 and 24 (7 males, 7 females), 12 were between the ages of 25 and 39 (6 males, 6 females), 16 were between the ages of 40 and 54 (8 males, 8 females), and 14 were 55 and older (7 males, 7 females). However, age group was not a dependent variable for this study. Additional information regarding Experiment 2 participants can be found in Appendix C.

## Instruction and Training

Prior to proceeding to the test vehicle, the participants were provided with the tablet that they would be using in the vehicle. They were given a brief introduction to the types of tasks they would be performing, and were allowed time for any questions and to practice one of each of the different types of tasks, if needed. Participants were then provided with a static orientation to the experimental vehicle, which included basic controls and the L2 automation features. Following this, participants received an on-track orientation consisting of four laps on the test track. During this training, participants were not given an explanation of the levels of automation nor were they given the impression that the system should allow them to divert attention away from the roadway.

The purpose of this on-track orientation was to allow participants to acclimate to the vehicle and the test-track environment. Participants were asked to drive to the circle track entrance, enter the second travel lane, and maintain a speed of 50 mph. The first lap was completed under manual driving. During the second lap, ACC was activated, which allowed participants to release longitudinal control to the automated system while they maintained lateral control of the vehicle. The third and fourth laps were completed using the L2 automation (ACC plus lane centering), which allowed participants to release both longitudinal and lateral control of the vehicle. The prototype system clearly depicted when the system was ready for activation of the automation. Following this, participants were provided with additional instructions pertaining to the experimental session and were given the opportunity to ask any additional questions. Note: Experimenters followed a set script and protocol in order to ensure consistency between experimenters.

## Study Session

Participants completed three driving sessions, with each lasting approximately 60 minutes. Participants were instructed to begin interacting with a variety of non-driving tasks during the driving session upon activating the L2 automation.

Participants were presented with three types of non-driving tasks: navigation, email, and Web browsing. A total of up to 90 different non-driving tasks, 30 in each category, were completed. These tasks were similar in terms of the visual/manual demand required and they were presented in a random order. All tasks were performed using the tablet computer, which was equipped with a standard QWERTY touchscreen keyboard. Each task was typed on a notecard. For each task, the in-vehicle experimenter provided the participant with the notecard and, upon completing the task, the participant passed the notecard back to the in-vehicle experimenter. The experimenter provided the participant with a new notecard containing a different task 30 to 60 s later. Note, the pace of these tasks was not forced but at the participants' leisure.

- Navigation. The navigation tasks required the participant to open an application of the device, choose the new destination option, and enter the address provided on the notecard in a printed number-street-city-state format.

- Email. The email tasks required the participant to compose an email using the tablet.

- Web browsing. The Web-browsing tasks required the participant to open the Web browser on the tablet. Participants were asked to determine the answer to a specific question or task that required searching or interacting with the Internet.

Each participant was assigned a prompt condition: either 2-second, 7-second, or No Prompts. The driver-monitoring system provided three stages of prompts based on the assigned prompt condition and the participants' attention state. If the participants' attention state was off-road, the system provided alerts based on the assigned prompt condition. For the 2-second prompt condition, the prompts began after the participant's attention state was off-road for 2 s. For the

7-second prompt condition, the prompts began after the participant's attention state was off-road for 7 s. Participants who were assigned to the No Prompts condition did not receive any prompts. The driver monitoring system provided three progressive stages of alerts. Details on the stages are provided below.

- Stage 1. This stage was activated once the participant's attention was off-road for the threshold set by the prompt condition.
  - This stage was characterized by a lower-urgency visual alert.
  - The prompt ended if the participant's attention state changed to on-road, and the system resumed normal operation.
  - If the participant's attention state did not change to on-road within 5 s, the prompt continued to Stage 2.
- Stage 2. This stage was activated once the Stage 1 prompt was active for 5 s without the participant's attention state changing to on-road.
  - This stage was characterized by a higher-urgency multimodal alert (i.e., visual and haptic).
  - The prompt ended if the participant's attention state changed to on-road, and the system resumed normal operation.
  - If the participants' attention state did not change to on-road within 5 s, the prompt continued to Stage 3.
- Stage 3. This stage was activated once the Stage 2 prompt was active for 5 s without the participant's attention state changing to on-road.
  - This stage was characterized by a higher-urgency multimodal alert (i.e., visual, haptic, and auditory).
  - This was the final stage of the prompt, and continued until participants took control of steering.
  - Once the participants took control of steering, the system turned off and needed to be reengaged manually.

Figure 3-4 depicts the alert timelines for the 2-second and 7-second task interval scenarios.

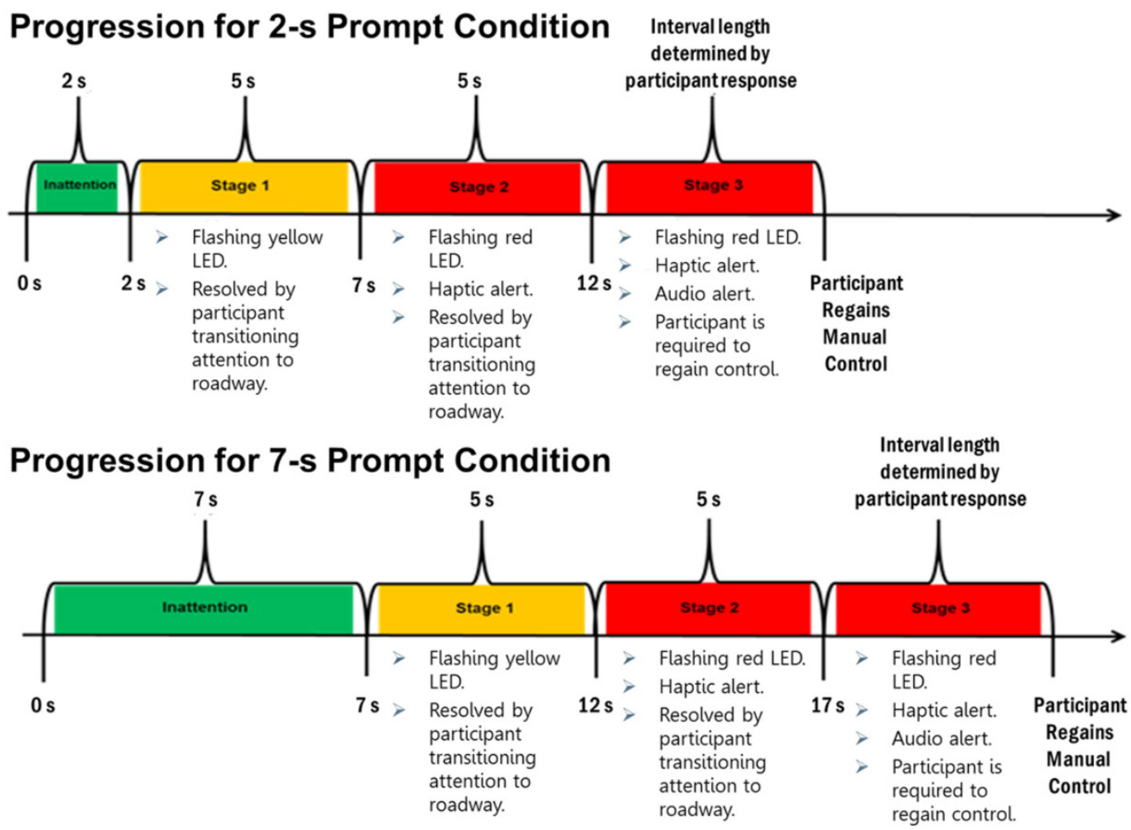



**Figure 3-4. Alert Timelines for 2-second & 7-second Prompt Conditions in Experiment 2**

During the experiment, each participant experienced both types of lane drift (with and without an alert) and no lane drift—one time each—in different driving sessions, and at random times. Details pertaining to each of the event types (lane drift with alert, lane drift without alert, no lane drift) are described below. All of the surprise lane drifts were prescribed and injected into the condition of interest using the experimenter console. The lane drifts with the alerts represent the condition of a lane keeping performance issue in which the system warns the vehicle operator to regain control. The situations with no alerts represent conditions where there is a lane keeping performance issue and a simultaneous failure of the prompt system, thus the system does not warn the vehicle operator.

- Alert. The alert event type consisted of a left lane drift with a multimodal alert (visual and haptic). This event was triggered at a random time during one of the three sessions and while participants were engaged in a non-driving task.

- No Alert. This event type consisted of a left lane drift with no warning or alert. This event was triggered at a random time during one of the three sessions and while participants were engaged in a non-driving task. The prompting system, if utilized for the experimental condition, was disabled for this event, simulating a simultaneous failure of that system.

- No lane drift. During one of the three sessions, there were no injected lane drifts.

Participant responses to the attention state prompts and experimenter-injected lane drifts were measured using visual evidence from the DAS. The participants received a total of two lane drifts, one with an alert and one without, and these occurred in different driving sessions. These lane drifts were triggered only when there was no surrounding traffic, and when participants were actively engaged in non-driving tasks. If participants did not respond to the lane drift once the subject vehicle was fully in the adjacent lane (about 3 to 5 s), the front-seat experimenter would then instruct them to take control of steering.

After completing three driving sessions, participants were instructed to exit the circle track and return to the preparation area. Participants were then interviewed, asked to complete the after-experience trust scales and interview, and provided a debriefing on the purpose of the study. Participants were given compensation for participation in the study. Figure 3-5 presents the timeline of events for this experiment.

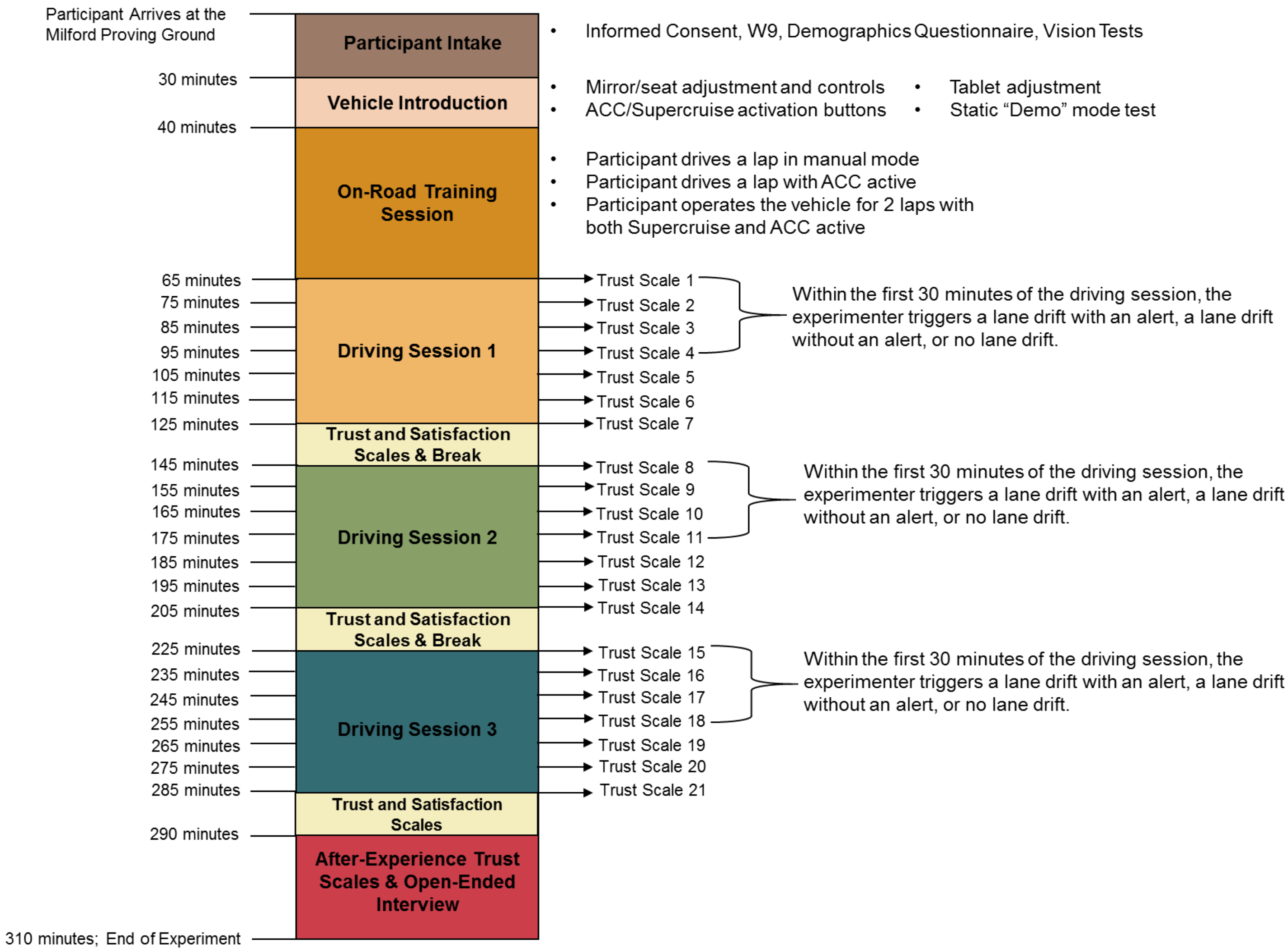



**Figure 3-5. Experimental Session Timeline for Experiment 2**

# Results for Lane Drift Analysis

## Operator Behavior

***Time to React***

Participants reacted to an unexpected lane drift 3.7 times faster if there was a visual plus haptic alert (M = 1.00 s, S.E. = 0.59 s, n = 54, min = 0.08 s, max = 2.81 s) than they did if there was no alert at all (M = 3.70 s, S.E. = 0.35 s, n = 49, min = 0.14 s, max = 8.72 s), and this result was statistically significant, $F(1,48.5) = 56.78$, $p < .0001$.

The only significant effect of prompt condition occurred within Session 1 for participants who received a lane drift without an alert. In this case, participants who were in the 2-second prompt condition reacted faster than those who did not receive prompts at all, $t(9.05) = -2.95$, $p = .048$. Those in the 7-second prompt condition also reacted faster than those in the No Prompts condition in Session 1 after lane drifts with no alerts, $t(12.3) = -5.83$, $p < .0001$. However, there was no significant difference between the 2-second and 7-second conditions, $p > .05$.

Overall, the mean time to react to the lane drift for participants who experienced the 2-second condition was 1.97 s (S.E. = 0.35 s, n = 32, min = 0.43 s, max = 7.68 s), 2.14 s for the 7-second condition (S.E. = 0.32 s, n = 39, min = 0.08 s, max = 7.29 s), and 2.77 s for those with no prompts (S.E. = 0.44 s, n = 32, min = 0.14 s, max = 8.72 s).

Mean and error bar plots of the time to react to the lane drift are displayed in Figure 3-6 by event type. (For additional details on the statistical analysis, see Appendix C, Results for Lane Drifts, Operator Behavior, Analysis for Time to React, Statistical Analysis.)

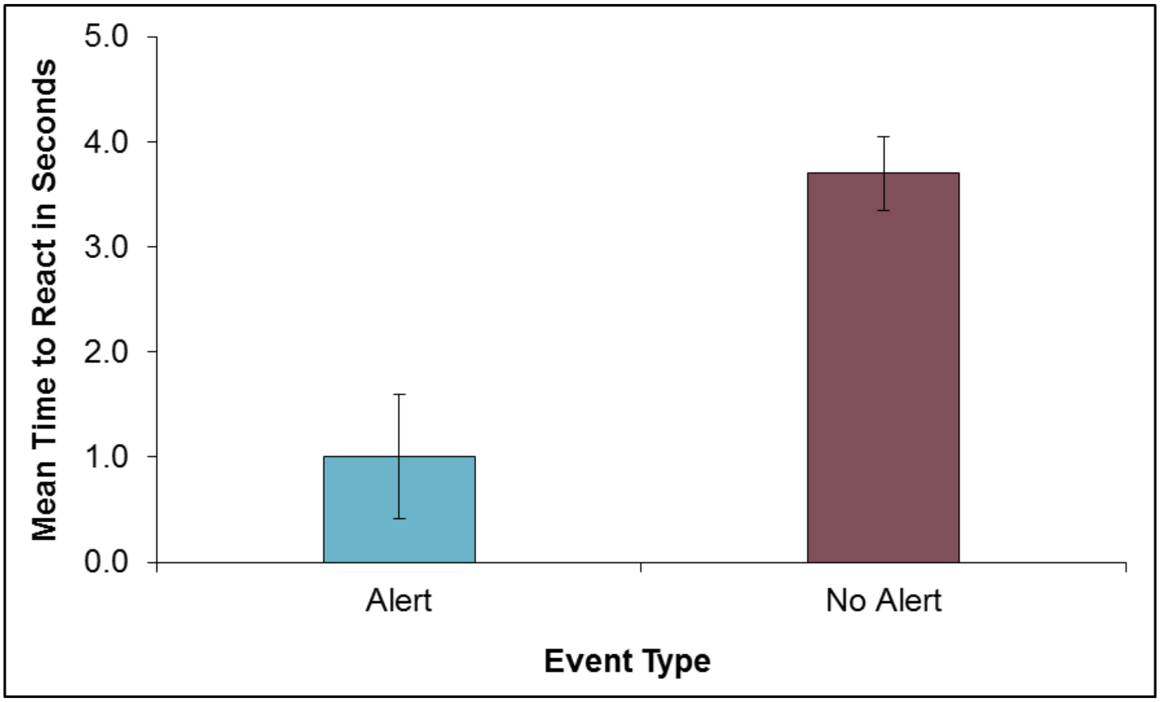



**Figure 3-6. Mean and Standard Error Bar Plots, Time to React, by Event Type for Experiment 2**

***Time to Regain Control***

When participants received a visual plus haptic alert at the time of a lane drift, they were much more likely to regain control of the vehicle before it had completely exited the lane (53 out of 54 instances, 98.2 percent) than if they had a lane drift but did not receive an alert (26 out of 49 instances, 53.1 percent), and this result was statistically significant ($\chi^2 = 15.09$, $p = .0001$). If the lane drift was announced and the participant regained control in time, then the mean time to regain control for participants was 2.02 s (S.E. = 0.14 s, n = 17, min = 1.18 s, max = 3.15 s) for participants who received prompts after looking away for 2 s, 2.50 s (S.E. = 0.48 s, n = 21, min = 0.89 s, max = 9.48 s) for participants receiving prompts after looking away for 7 s, and 2.57 s (S.E. = 0.29 s, n = 15, min = 1.22 s, max = 5.25 s) for those who did not receive prompts. These differences were not statistically significant. Bar plots of the instances in which participants did and did not regain control prior to exiting the lane, stratified by event type, are displayed in Figure 3-7. In this plot, “correct” means the participant regained control without instruction from the in-vehicle experimenter and “incorrect” means the participant required instruction from the in-vehicle experimenter to regain control. (For additional details, see Appendix C, Results for Lane Drifts, Operator Behavior, Analysis for Time to Regain Control, Statistical Analysis.)

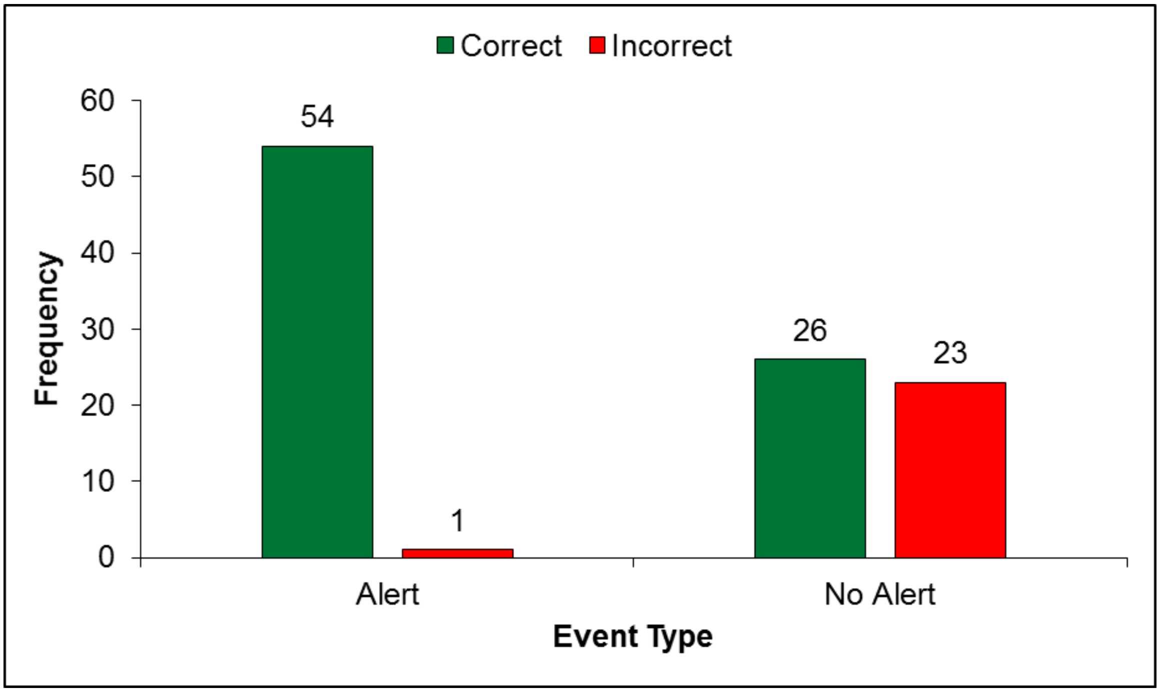



**Figure 3-7. Bar Plots of Performance by Alert Type for Experiment 2**

***Time to Activate the Automation***

The mean time to activate the automation after a lane drift was 7.27 s (S.E. = 0.50 s, n = 51, min = 0.96 s, max = 15.63 s) if the participant experienced an alert, and 5.02 s (S.E. = 0.59 s, n = 47, min = 0.12 s, max = 22.00 s). There was a significant interaction between alert type and session, F(2,56.2) = 4.78, *p* = .0121, indicating that the effect of an alert on the time to activate the automation after a lane drift varies across time. Post hoc tests indicated that in the first session, participants that experienced an alert took longer to activate the automation than participants who did not experience an alert, *t*(64.5) = 3.96, *p* = 0.0002. For different prompt conditions, mean time to activate the automation was 5.18 s (S.E. = 0.71, n = 30, min = 0.56 s, max = 11.80 s) for participants receiving prompts after 2 s of looking away, 7.10 s (S.E. = 0.74 s, n = 37, min = 1.29 s, max = 22.00 s) for participants who received prompts after 7 s of looking away, and 6.07 s (S.E. = 0.56 s, n = 31, min = 0.12 s, max = 12.51 s) for participants who never received prompts. No significant effect of prompt condition was observed. Additionally, there was no significant change in the time to activate the automation over time. (For additional details on the statistical analysis, see Appendix C, Results for Lane Drifts, Operator Behavior, Analysis for Time to Activate the Automation, Statistical Analysis.)

***Time to Release Control of Steering***

No significant effect, either of alert or prompts, was found on the time to release control, nor was there a significant change over time. When participants received an alert at the lane drift, the mean time to release control was 3.25 s (S.E. = 0.45 s, n = 50, min = 0.01 s, max = 16.47 s), while the mean when participants did not receive an alert was 4.23 s (S.E. = 0.72 s, n = 47, min = 0.03 s, max = 25.00 s). For prompt conditions, the mean times to release control were 4.67 s (S.E. = 1.09 s, n = 29, min = 0.31 s, max = 25.00 s) when participants received prompts 2 s after looking away, mean = 3.37 s (S.E. = 0.48 s, n = 37, min = 0.10 s, max = 16.47 s) when participants received prompts 7 s after looking away, and 3.25 s (S.E. = 0.58, n = 31, min = 0.01 s, max = 11.92 s) when participants did not receive prompts. (For details on the statistical analysis, see Appendix C, Results for Lane Drifts, Operator Behavior, Analysis of Time to Release Control, Statistical Analysis.)

#### *Time to Resume a Non-driving Task*

No significant effect, either of alert or prompts, was found on the time to resume a non-driving task, nor was there a significant change over time. When participants received an alert at the lane drift, the mean time to release control was 3.25 s (S.E. = 0.45 s, n = 50, min = 0.01 s, max = 16.47 s), while the mean when participants did not receive an alert was 4.23 s (S.E. = 0.72 s, n = 47, min = 0.03 s, max = 25.00 s). For prompt conditions, the mean times to release control were 4.67 s (S.E. = 1.09 s, n = 29, min = 0.31 s, max = 25.00 s) when participants received prompts 2 s after looking away, mean = 3.37 s (S.E. = 0.48 s, n = 37, min = 0.10 s, max = 16.47 s) when participants received prompts 7 s after looking away, and 3.25 s (S.E. = 0.58, n = 31, min = 0.01 s, max = 11.92 s) when participants did not receive prompts. (For additional details on the statistical analysis, see Appendix C, Results for Lane Drifts, Operator Behavior, Analysis of Time to Resume a Non-Driving Task, Statistical Analysis.)

## Operator Behavior Analysis Summary

The strongest results in the analysis of operator behavior in response to unanticipated events came in the effect of experiencing an alert at the time of the lane drift compared to not experiencing the alert. Specifically, when participants experienced an alert at the time of the lane drift, they reacted to the lane drift 66 percent faster than if there was no alert. Similarly, the participants were 26.1 times more likely to regain control after the lane drift without instruction if they experienced an alert at the time of the lane drift compared to if they did not. Additionally, those experiencing an alert with the lane drift may be more hesitant at first to activate the automation than those experiencing a lane drift without an alert, but this difference may dissipate over time. Table 3-1 summarizes the results of the operator behavior analysis for Experiment 2.

**Table 3-1. Summary Table for Operator Behavior Analysis for Experiment 2**

| Variable | Event Type | Prompt Condition | Performance Over Time (Represented by Session) | Takeaways |
|---|---|---|---|---|
| **Time to React** | **Significant** | Not Significant | Not Significant | After a lane drift, vehicle operators may react faster if there is an alert at the time of the lane drift. |
| **Time to Regain Control** | **Significant** | Not Significant | Not Significant | Vehicle operator may be more likely to regain control after a lane drift within a reasonable amount of time if the lane drift is accompanied by an alert. |
| **Time to Activate the Automation** | **Significant** | Not Significant | **Significant** | In early driving sessions, experiencing an alert along with a lane drift may result in increased time to activate the automation; however, this relationship may not last as time goes on. |
| **Time to Release Control of Steering** | Not Significant | Not Significant | Not Significant | There is no statistically significant evidence of an effect of event type or prompt condition on time to release control. |
| **Time To Resume Non-Driving Task** | Not Significant | Not Significant | Not Significant | There is no statistically significant evidence of an effect of event type or prompt condition on time to resume non-driving task. |

## Eye-glance Behavior

***Monitoring Rate***

Experiment 2 included an analysis of participants' monitoring rate. It was observed that participants monitored the driving environment significantly more often immediately after the lane drift compared to before the lane drift if there was a visual + haptic alert, $F(1,43) = 126.13$, $p < .0001$, but this relationship did not hold if the lane drift did not include an alert, $p > .05$. Further, the change was greater if the participants did not receive prompts compared to if they were in the 2-second prompt condition, $t(50) = -4.75$, $p < .0001$, or the 7-second prompt condition, $t(50) = -3.56$, $p = .0024$. This is mainly due to the fact that prior to the announced lane drift, the monitoring rate was significantly lower for those who did not receive prompts compared to those who did.

If the alert was announced, the mean percentage of driving-related glance time for those in the 2-second prompt condition was 22.12 percent before the lane drift (S.E. = 4.87 percent, n = 16, min = 4.73 percent, max = 75.37 percent) and 47.41 percent after the lane drift (S.E. = 4.01 percent, n

= 16, min = 20.41 percent, max = 76.03 percent). For participants in the 7-second condition, the mean was 20.67 percent before the lane drift (S.E. = 5.25 percent, n = 20, min = 0.00 percent, max = 100.00 percent) and 59.50 percent after the lane drift (S.E. = 6.27 percent, n = 20, min = 0.00 percent, max = 100.00 percent). For those who did not experience prompts, the mean was 3.86 percent before the lane drift (S.E. = 1.42 percent, n = 16, min = 0.00 percent, max = 18.57 percent) and 56.79 percent after the lane drift (S.E. = 3.74 percent, n = 16, min = 28.39 percent, max = 74.03 percent). Mean and error bar plots of monitoring rate immediately before and after announced lane drifts are displayed in Figure 3-8. (For additional details on the statistical analysis, see Appendix C, Results for Lane Drifts, Eye-glance Behavior, Analysis for Monitoring Rates, Statistical Analysis.)

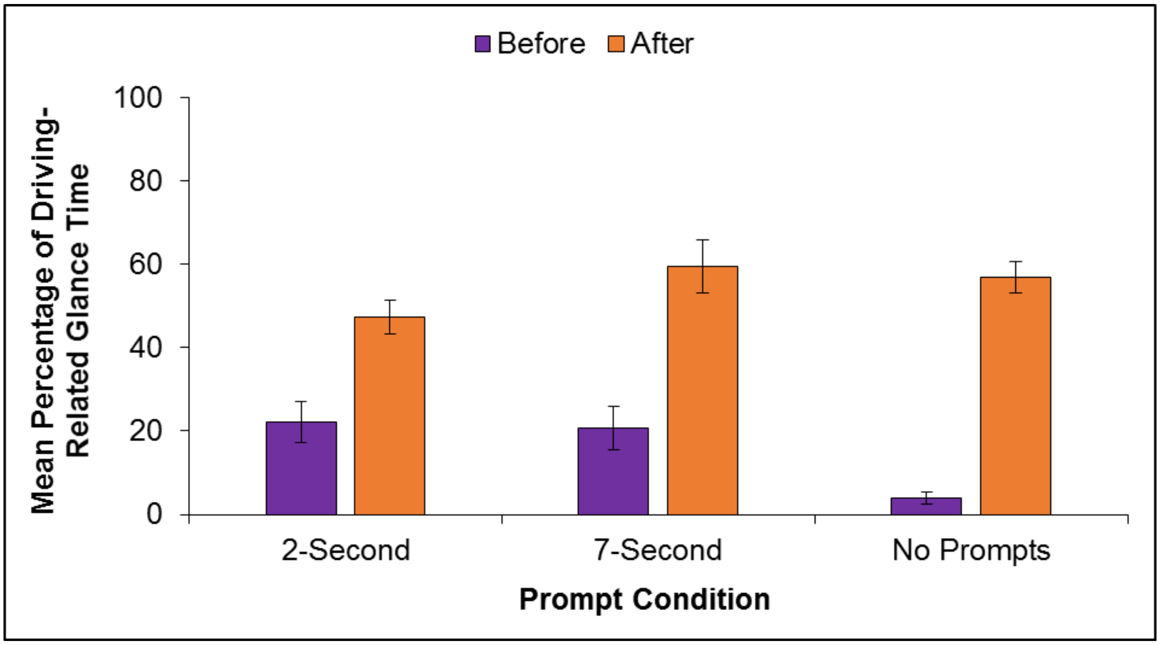



**Figure 3-8. Mean and Error Bar Plots of Monitoring Rate Immediately Before and After Lane Drifts with Alerts, Stratified by Prompt Condition**

***Non-driving-related Glances***

A significant interaction was found between prompt condition and session, $\chi^2(4) = 11.9$, $p = .0181$, indicating that the effect of prompt condition on the rate of non-driving-related glances changed over time. Specifically, the rate of non-driving-related glances for those who received no prompts was significantly lower in Session 1 than it was for those in the 2-second prompt condition, $z = -5.63$, $p < .0001$, and those in the 7-second prompt condition, $z = -4.09$, $p < .0001$, but this significant difference was absent for Sessions 2 and 3. In Session 1, the mean rate of non-driving-related glances within the group without prompts was 0.22 glances per second (S.E. = 0.03 glances/s, n = 22, min = 0.10 glances/s, max = 0.61 glances/s), compared to a mean of 0.45 glances/s for the 2-second condition (S.E. = 0.06 glances/s, n = 16, min = 0.14 glances/s, max = 0.87 glances/s) and a mean of 0.42 glances/s for the 7-second condition (S.E. = 0.05 glances/s, n = 26, min = 0.00 glances/s, max = 1.12 glances/s).

Additionally, there was a significant increase in the rate of non-driving-related glances from before to after the lane drift if there was an alert, but not if there was no alert, $\chi^2(1) = 53.69$, $p < .0001$.

(For additional details on the statistical analysis, see Appendix C, Results for Lane Drifts, Eye-glance Behavior, Analysis for Non-Driving-Related Glances, Statistical Analysis.)

# Results for Prompt Analysis

## Operator Behavior

### *Time to React*

A significant increase in time to react to the prompt was found from Session 1 to Session 3, $t(38.3) = 2.59$, $p = .0405$. Although a significant interaction was found between session and prompt condition, indicating that the effect of prompt condition on the time to react to prompts changed significantly over time, post hoc tests did not reveal significant findings when adjusting for multiple comparisons.

The mean time to react to prompts for the 2-second condition was 2.41 s in Session 1 (S.E. = 0.26 s, n = 88, min = 0.01 s, max = 11.42 s), 2.45 s in Session 2 (S.E. = 0.32 s, n = 90, min = 0.06 s, max = 18.57 s) and 3.26 s in Session 3 (S.E. = 0.32 s, n = 75, min = 0.13 s, max = 11.19 s). The mean time to react to prompts for the 7-second condition was 2.33 s in Session 1 (S.E. = 0.29 s, n = 72, min = 0.14 s, max = 13.95 s), 3.72 s in Session 2 (S.E. = 0.51 s, n = 70, min = 0.09 s, max = 30.86 s), and 3.25 s in Session 3 (S.E. = 0.32 s, n = 75, min = 0.13 s, max = 11.19 s). (For additional details on the statistical analysis, see Appendix C, Results for Prompts, Operator Behavior, Analysis for Time to React, Statistical Analysis.)

### *Time to Regain Control*

Although participants were required to regain control of the vehicle only after the onset of Stage 3 of the prompt (10 s after the prompt begins), during several instances where the prompt did not progress to Stage 3, participants chose to regain control anyway. Out of 337 prompts that progressed only to the first stage, participants regained control after 97 of these prompts (28.8 percent). Out of 117 prompts that progressed to Stage 2, the participant regained control after 63 of these prompts (53.9 percent). It should be noted that additional prompts may have occurred between their reaction to these sample prompts and when they regained control, so it may have been a different prompt that caused the participants to regain control in these instances. A bar plot of the proportion of participants regaining control in different stages, stratified by condition and session, is displayed in Figure 3-9. (For additional details on the statistical analysis, see Appendix C, Results for Prompts, Operator Behavior, Statistical Analysis.)

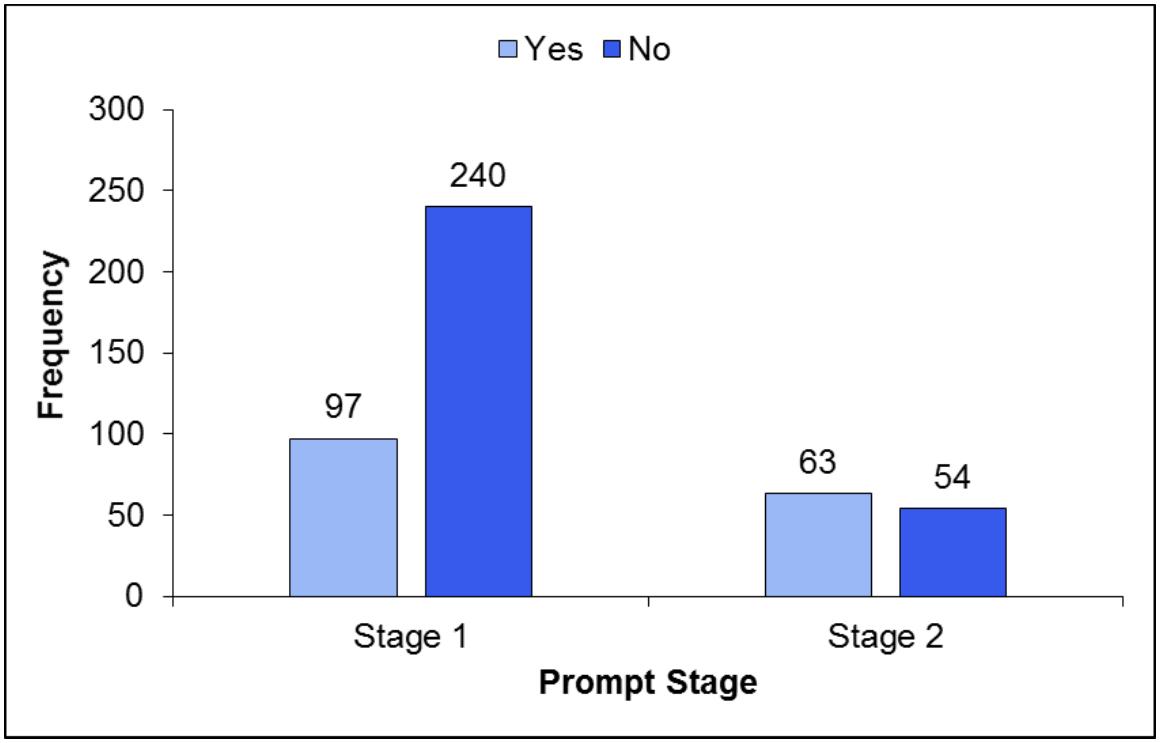



**Figure 3-9. Frequency of Regaining Control After Prompt, by Prompt Stage**

## Operator Behavior Analysis Summary

With only one exception, when the prompting system escalated to a request for the operator to take control, participants did so in all cases. However, time may play a role in how quickly vehicle operators react to these prompts. Specifically, participants' reaction time to prompts increased over time. However, the frequency with which the participants received prompts did not appear to have a significant effect on how quickly they reacted to prompts, nor how likely they were to regain control after prompts. The experiment did not examine any effect of prompting during the lane drift without alert; if prompting had been done during the lane drifts, the hypothesis is that it could possibly have helped.

## Eye-glance Behavior

In order to investigate the eye-glance behavior around prompts, a prompt was sampled in between each of the trust scales within each driving session for participants who experienced either the 2-second or 7-second prompts. To compare this behavior with participants who did not receive prompts, a random time point was sampled between the trust scales for the No Prompts group. Eye-glance behavior was then studied for the 15 s before and the 15 s after the prompt (or the random time point within the No Prompts group).

This analysis helped to determine if the alert type affected eye-glance behavior, and if this effect changed over time, which will aid in determining whether the effect over time should be considered in future research investigations. Table 3-2 provides a summary of the results and general takeaways for monitoring rate and non-driving-related glances. Event type and session were marked as "**Significant"** if they achieved statistical significance, and "Not Significant" if they did not.

**Table 3-2. Summary Table for Eye-glance Behavior Analysis for Experiment 2**

| Variable | Prompt Condition | Event Type | Performance Over Time (Represented by Intervals Within Sessions) | Takeaways |
|---|---|---|---|---|
| **Monitoring Rate** | **Significant** | Not Significant | **Significant** | If participants were not paying attention to the driving environment for a period of time, 7-second prompts significantly increased participants' monitoring rate. The overall monitoring rate was higher just before a prompt in the 2-second condition, compared to random times without prompts. |
| **Non-driving-related Glances** | **Significant** | **Significant** | **Significant** | The amount of non-driving-related glances decreased with time. Additionally, experiencing prompts increased the amount of glances participants made. |

***Monitoring Rate***

When participants were engaged in a non-driving task, the monitoring rate (i.e., the percentage of driving-related glance time in a given time interval) was significantly higher immediately prior to a 2-second prompt compared to what was seen for those in the No Prompts condition, $t(45) = 5.20$, $p < .0001$ (Figure 3-10), while there was no significant difference in monitoring rate immediately prior to the 7-second prompt compared to those in the No Prompts condition, $p > .05$. However, only the 7-second prompt condition experienced a significant increase in monitoring rate from immediately before to immediately after the prompt, $t(45) = 4.65$, $p < .0001$. This may be because the 7-second prompts occurred only when the participant was looking away for 7 seconds, so these were the more extreme cases of low monitoring rates within this group.

The mean percentage of driving-related glance time within the 2-second condition was 34. 51 percent before the prompt (S.E. = 1.44 percent, n = 254, min = 0.00 percent, max = 92.44 percent) and 31.50 percent after the prompt (S.E. = 1.35 percent, n = 254, min = 4.44 percent, max = 100.00 percent). The mean percentage of driving-related glance time within the 7-second condition was 17.20 percent before the prompt (S.E. = 0.99 percent, n = 205, min = 0.00 percent, max = 61.41 percent) and 27.88 percent after the prompt (S.E. = 1.54 percent, n = 205, min = 0.00 percent, max = 100.00 percent). In the No Prompts condition, the mean percentage of driving-related glance time in the 15 s prior to the randomly selected time point was 18.23 percent (S.E. = 1.89, n = 185, min = 0.00 percent, max = 100.00 percent). Mean and error bar plots for monitoring rate by prompt condition are displayed in Figure 3-10. (For additional details on the statistical analysis, see Appendix C, Results for Prompts, Eye-glance Behavior, Analysis of Monitoring Rate, Statistical Analysis.)

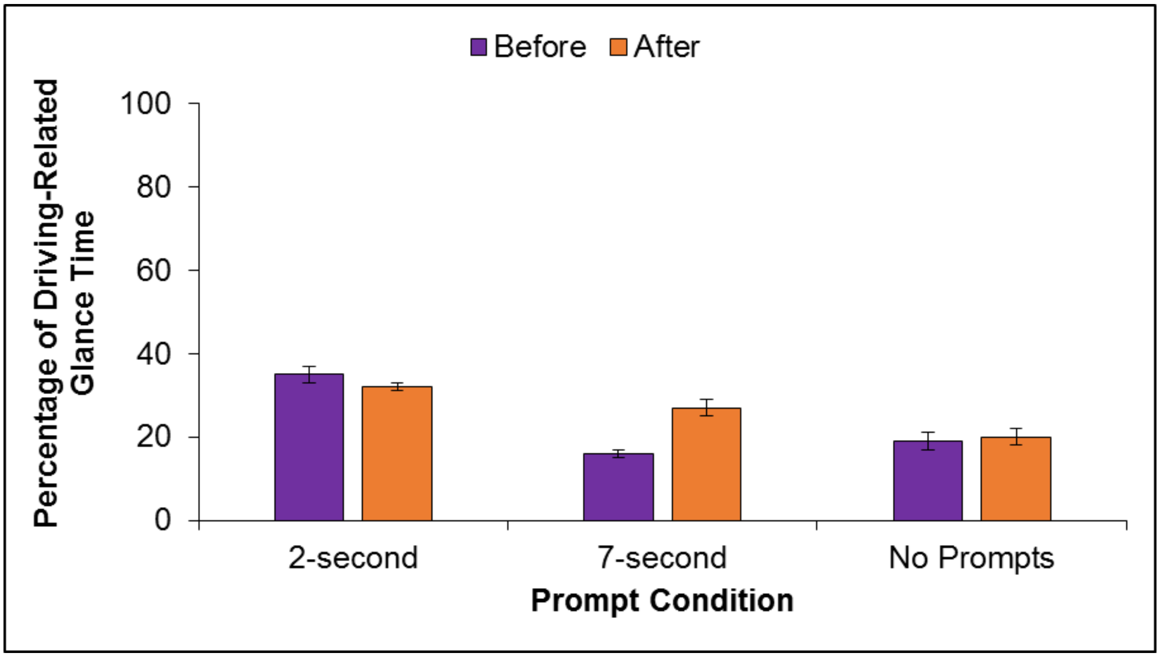



**Figure 3-10. Percentage of Driving-related Glance Time for Participants Engaged in a Non-driving Task at Prompt Start or Selected Time Point**

***Non-driving-related Glances***

When participants were engaged in a non-driving task, the number of non-driving-related glances changed significantly from immediately before to immediately after a prompt in the 2-second condition, $t(45) = 2.77$, $p = .0400$, and in the 7-second condition, $t(45) = 3.98$, $p = .0010$. Additionally, the number of non-driving-related glances was found to be significantly higher just prior to the 2-second prompt compared to just prior to the 7-second prompt, $t(45) = .3.28$, $p = .0100$, and also higher than in the No Prompts condition, $t(45) = 3.49$, $p = .0055$. However, there was no significant difference in non-driving-related glances between the No Prompts condition and just prior to the prompt in the 7-second condition.

The mean number of non-driving-related glances for the 2-second prompt was 3.26 before the prompt (S.E. = 0.09, n = 254, min = 1, max = 9) and 3.47 after the prompt (S.E. = 0.09, n = 254, min = 0, max = 7). The mean number of non-driving-related glances within the 7-second condition was 2.5 before the prompt (S.E. = 0.08, n = 205, min = 1, max = 6) and 3.07 after the prompt (S.E. = 0.08, n = 205, min = 0, max = 6). Finally, for the No Prompts condition, the mean number of non-driving-related glances for the first 15 seconds in a randomly selected time interval was 2.27 (S.E. = 0.11, n = 186, min = 0, max = 8). (For additional details on the statistical analysis, see Appendix C, Results for Prompts, Eye-glance Behavior, Analysis for Non-Driving-Related Glances, Statistical Analysis.)

# Results for Trust Scales Analysis

### *Trust Over Time*

Of interest is how much participants trusted the system they were experiencing and whether or not their ratings of trust changed over time. A plot of the mean rating change over the entire experiment (all three sessions) is displayed in Figure 3-11 for the three separate conditions. There is a slight increase in the trust from Session 1 to Session 2, $t(45) = 2.95$, $p = 0045$. However, this increase does not continue from Session 2 to Session 3, $p > .05$. (For additional details, see Appendix C, Subjective Data, Analysis for Trust over Time, Statistical Analysis.)

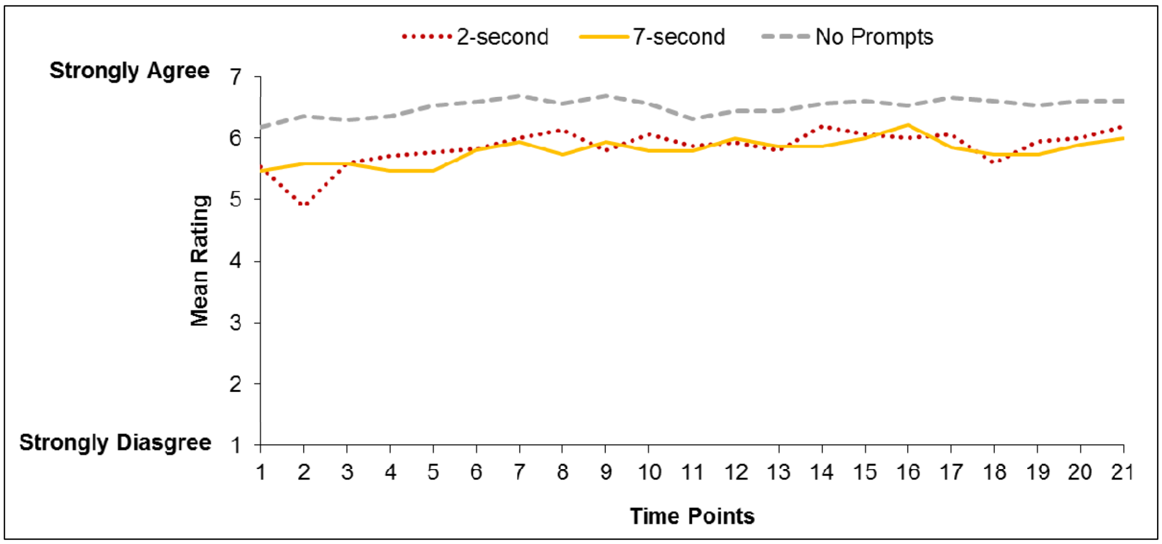



**Figure 3-11. Time Trend of Trust, Stratified by Prompt Conditions**

It should be noted that participants who experienced a lane keeping performance issue without an alert lost some trust in the automation. This suggests that operators may have somewhat calibrated their trust to the capabilities of the automation.

### *Trust Before/After the Lane Drift*

The change in trust from before to after the lane drift was different between announced and silent lane drifts, $F(1,88) = 14.48$, $p = .0003$. Specifically, there was a significant decrease in mean trust rating from before to after a silent lane drift, $t(88) = -5.11$, $p < .0001$, but there was no significant change if the lane drift was announced, $p > .05$. The mean trust rating for participants just before a lane drift was 6.06 (S.E. = 0.17, n = 50, min = 2, max = 7) if the lane drift was silent and 5.88 (S.E. = 0.19, n = 52, min = 1, max = 7) if the lane drift was announced. Soon after the lane drift, the mean rating fell to 5.26 (S.E. = 0.21, n = 50, min = 1, max = 7) if the lane drift was silent, but was 5.94 (S.E. = 0.17, n = 52, min = 1, max = 7) if the lane drift was announced. A mean and standard error bar plot of the ratings before and after the lane drift, stratified by event type, is displayed in Figure 3-12.

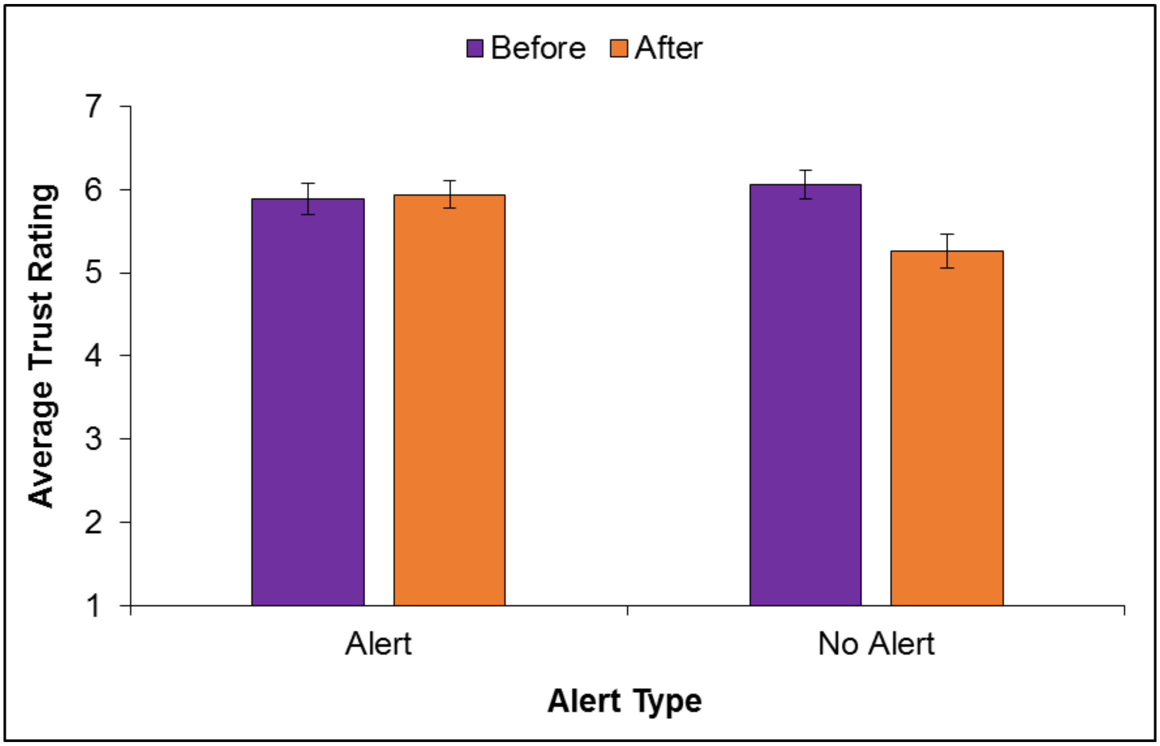



**Figure 3-12. Mean and Standard Error Bar Plots of Trust Rating Before and After the Lane Drift in Experiment 2, Stratified by Alert Type**

***After-experience Trust Scales***

A survey relating to how much the participant trusted the system was administered at the end of each of the three driving sessions. Participants were asked to respond to the following statements:

TS1. I can rely on the automated system to function properly while I am doing something else.

TS2. The automated system provided the alerts when needed.

TS3. The automated system gave false alerts.

TS4. The automated system is dependable.

TS5. I am familiar with the automated system.

TS6. I trust the automated system.

All responses were based on a 7-point Likert-type scale with options ranging from "1" for strongly disagree to "7" for strongly agree. All but one of the statements had positive wording, so a high score indicates a higher level of trust. Statement 3, "The automated system gave false alerts," had negative wording, so responses to this statement were recoded so that "1" = "7," "2" = "6," etc. Mean and standard error bar plots for responses to all statements are displayed in Figure 3-13 and Figure 3-14 for the 2-second and 7-second prompt conditions, respectively.

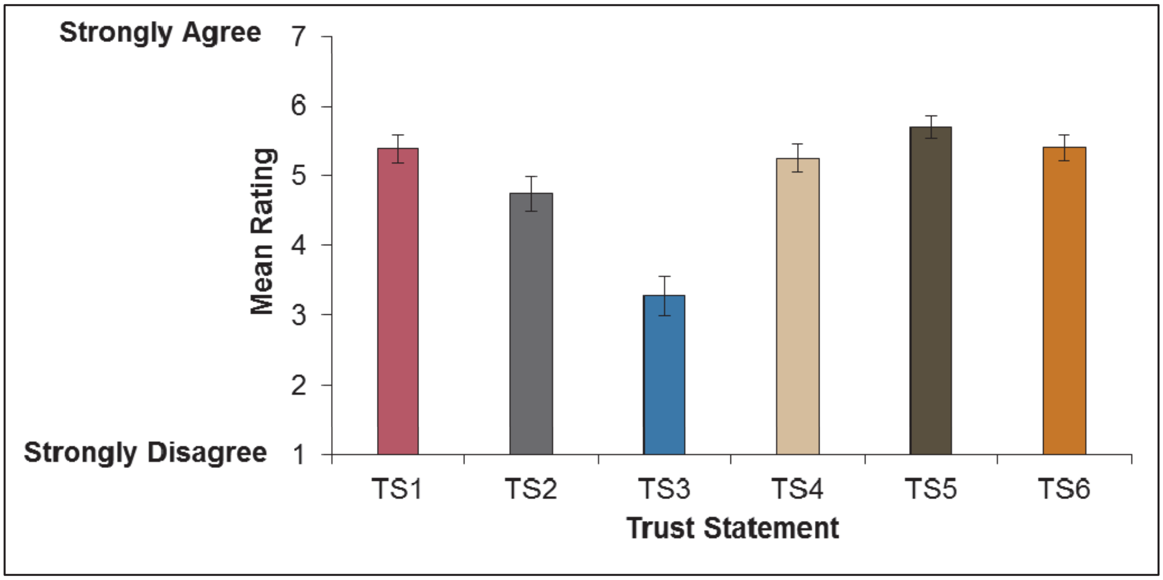



**Figure 3-13. Mean Post-session Trust Scale Rating for 2-second Prompt Condition**

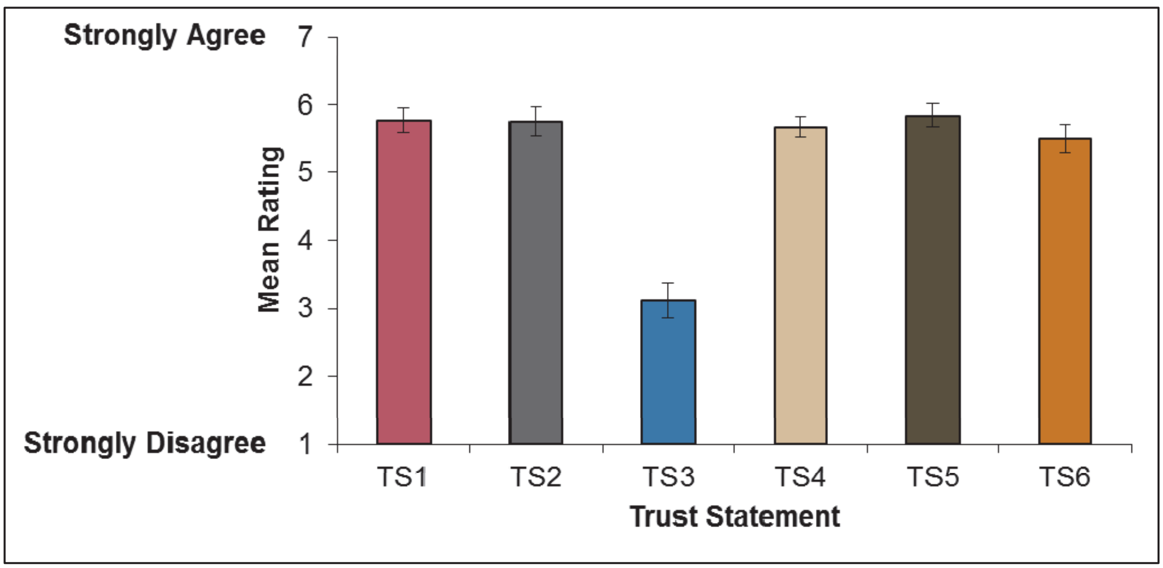



**Figure 3-14. Mean After-session Trust Scale Rating for 7-second Prompt Condition**

***After-experience Satisfaction Scales***

For Experiment 2 only, a survey comprising three questions and five statements (relating to the degree with which the participant was satisfied with the automated system) was administered at the end of each of the three driving sessions. Participants were asked to respond to the following:

1. Overall, how satisfied are you with the automated system?
2. How satisfied were you with the number of alerts provided?
3. How satisfied were you with the types of alerts provided?

4. The automated system's alerts provided sufficient time to make a decision.

5. The automated system's alerts provided sufficient information to make a decision.

6. I would use this type of automated system during my normal driving.

7. I would like to have this type of automated system as part of my current vehicle.

8. I would like to have this type of automated system as part of a future vehicle.

Note that the satisfaction questions 1, 2, and 3 were based on a 7-point Likert-type scale with options ranging from "1" for extremely dissatisfied to "7" for extremely satisfied. Satisfaction statements 4, 5, 6, 7, and 8 were based on a 7-point Likert-type scale with options ranging from "1" for strongly disagree to "7" for strongly agree. All of the questions and statements had positive wording, so a high score indicates a higher level of satisfaction. A mean and standard error bar plot for each satisfaction question and statement is displayed in Figure 3-15.

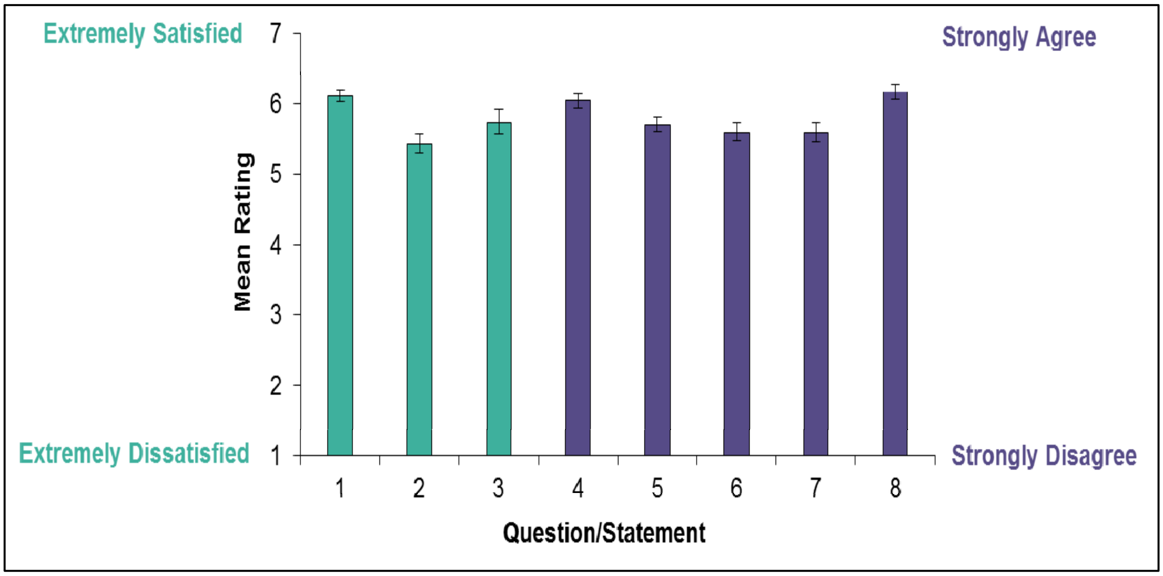



**Figure 3-15. Mean Rating for Satisfaction Survey for Experiment 2**

***After-experience Interview***

Following the in-vehicle experience, participants were asked a series of six questions. The interviews were audio recorded as a backup to researcher notes. Digital audio recordings were made into transcripts but no individual names were attached to any comments. Responses for all 56 participants were recorded and transcribed. The transcripts and researcher notes served as the basis for a qualitative content analysis which was completed using a framework analysis method. The results of this analysis were used to help researchers to potentially understand participant behavior that varies from the norm. A full discussion of the methods used to complete the framework analysis and detailed findings have been included in Appendix F.

Generally, participants indicated that they had a positive experience and seemed impressed by and confident in the system. The majority of participants reported that they were comfortable with the system (Figure 3-16).

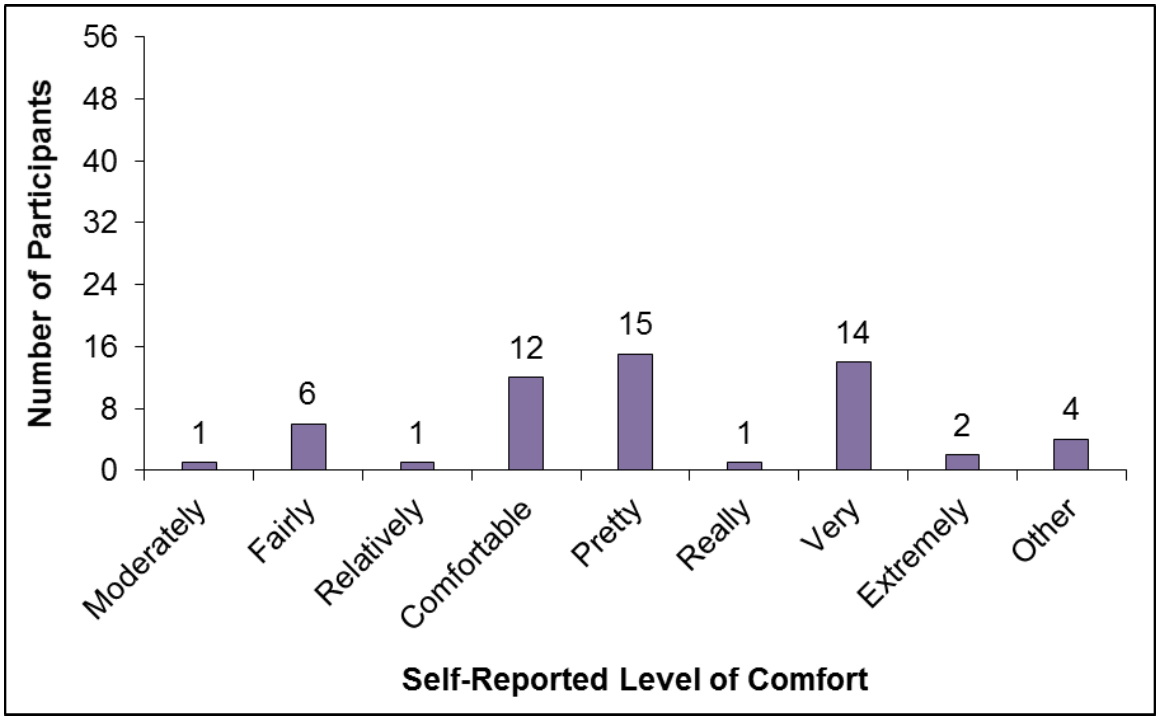



**Figure 3-16. Experiment 2 Participants' Self-reported Level of Comfort with the Automated System**

When asked how quickly (specific or relative time) they achieved their level of comfort, the majority of participants reported that their ultimate level of comfort was achieved in 30 minutes or less (Figure 3-17). The two participants who provided a temporally vague response to the question indicated that they achieved their ultimate level of comfort with the system "really quickly" and "very quickly." Additionally, nine participants noted that they did not reach their comfort level until the second half of the experiment (e.g., 90 to 180 minutes). The longest reported time to achieve comfort with the system was the "first quarter of the third hour" (i.e., 135 minutes), which was reported by a member of the 55+ age group.

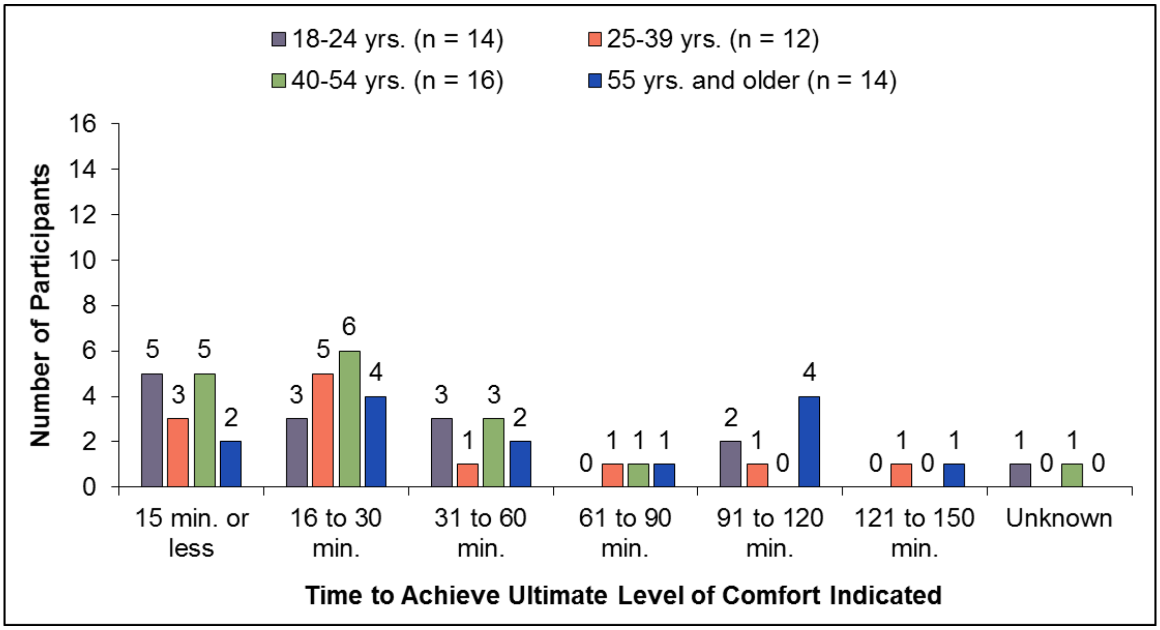



**Figure 3-17. Experiment 2 Participants' Time to Achieve Comfort with the System by Age**

Those requiring over 91 minutes to reach their ultimate level of comfort were more likely to fall in the 2-second or 7-second prompt conditions (Figure 3-18).

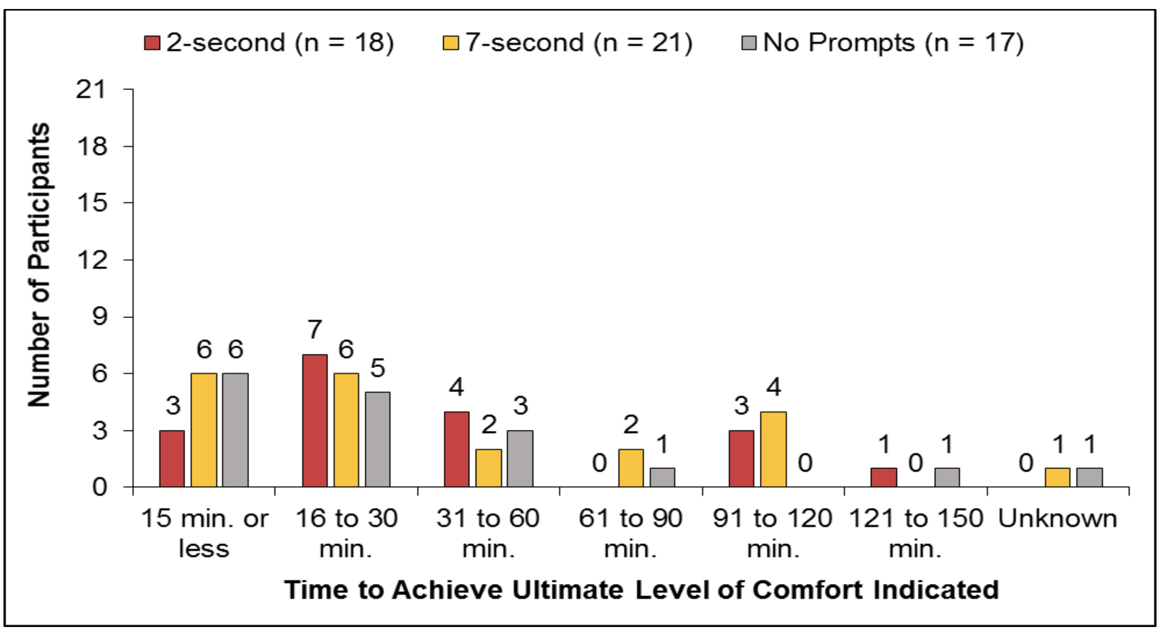



**Figure 3-18. Experiment 2 Participants' Time to Achieve Comfort with System by Prompt Condition**

# Experiment 2 Summary

The following tables summarize the significant differences between relevant combinations of event type and prompt conditions for each dependent variable. In each table, a particular combination of event type and prompt condition is compared to all other relevant combinations. For example, in Table 3-3, the combination of Alert with 2-second prompt condition is of interest, and it is directly compared to Alert with the 7-second prompt condition and Alert with the No Prompts condition because they share the same event type; it is also compared to No Alert with the 2-second prompt condition because they share the same prompt condition.

**Table 3-3. Mean Values for Alert with the 2-second Prompt Condition, Compared to Mean Values for Other Relevant Event Type/Prompt Condition Combinations in Experiment 2**

| Dependent Variable | Alert 2-Second | Alert 7-Second | Alert No Prompts | No Alert 2-Second |
|---|---|---|---|---|
| **Time to React** | 1.0 s | 1.0 s Not Significant | 1.1 s Not Significant | 2.8 s Significant |
| **Time to Regain Control** | 2.0 s | 2.5 s Not Significant | 2.6 s Not Significant | 4.6 s* Significant |
| **Time to Activate the Automation** | 5.5 s | 8.2 s Not Significant | 7.2 s Not Significant | 3.8 s** Significant |
| **Time to Release Control** | 3.4 s | 3.7 s Not Significant | 2.6 s Not Significant | 6.1 s Not Significant |
| **Time to Resume a Non-Driving Task** | 9.5 s | 8.8 s Not Significant | 6.2 s Not Significant | 13.8 s Not Significant |

* After a lane drift with no alert, 23 out of 49 participants did not regain control without instruction from the experimenter.
** This relationship held in Session 1 only.

**Table 3-4. Mean Values for Alert with the 7-second Prompt Condition, Compared to Mean Values for Other Relevant Event Type/Prompt Condition Combinations in Experiment 2**

| Dependent Variable | Alert 7-Second | Alert 2-Second | Alert No Prompts | No Alert 7-Second |
|---|---|---|---|---|
| Time to React | 1.0 s | 1.0 s Not Significant | 1.1 s Not Significant | 3.7 s Significant |
| Time to Regain Control | 2.5 s | 2.0 s Not Significant | 2.6 s Not Significant | 3.9 s* Significant |
| Time to Activate the Automation | 8.2 s | 5.5 s Not Significant | 7.2 s Not Significant | 5.6 s** Significant |
| Time to Release Control | 3.7 s | 3.4 s Not Significant | 2.6 s Not Significant | 3.0 s Not Significant |
| Time to Resume a Non-Driving Task | 8.8 s | 9.5 s Not Significant | 6.2 s Not Significant | 9.1 s Not Significant |

* After a lane drift with no alert, 23 out of 49 participants did not regain control without instruction from the experimenter.
** This relationship held in Session 1 only.

**Table 3-5. Mean Values for Alert with the No Prompts Condition, Compared to Mean Values for Other Relevant Event Type/Prompt Condition Combinations in Experiment 2**

| Dependent Variable | Alert No Prompts | Alert 2-Second | Alert 7-Second | No Alert No Prompts |
|---|---|---|---|---|
| Time to React | 1.1 s | 1.0 s Not Significant | 1.0 s Not Significant | 4.3 s Significant |
| Time to Regain Control | 2.6 s | 2.0 s Not Significant | 2.5 s Not Significant | 5.0 s* Significant |
| Time to Activate the Automation | 7.2 s | 5.5 s Not Significant | 8.2 s Not Significant | 4.9 s** Significant |
| Time to Release Control | 2.6 s | 3.4 s Not Significant | 3.7 s Not Significant | 3.9 s Not Significant |
| Time to Resume a Non-Driving Task | 6.2 s | 9.5 s Not Significant | 8.8 s Not Significant | 10.4 s Not Significant |

* After a lane drift with no alert, 23 out of 49 participants did not regain control without instruction from the experimenter.
** This relationship held in Session 1 only.

**Table 3-6. Mean Values for No Alert with the 2-second Prompt Condition, Compared to Mean Values for Other Relevant Event Type/Prompt Condition Combinations in Experiment 2**

| Dependent Variable | No Alert 2-Second | No Alert 7-Second | No Alert No Prompts | Alert 2-Second |
|---|---|---|---|---|
| **Time to React** | 2.8 s | 3.7 s<br>Not Significant | 4.3 s<br>Not Significant | 1.0 s<br>Significant |
| **Time to Regain Control** | 4.6 s | 3.9 s*<br>Not Significant | 5.0 s*<br>Not Significant | 2.0 s*<br>Significant |
| **Time to Activate the Automation** | 3.8 s | 5.6 s<br>Not Significant | 4.9 s<br>Not Significant | 5.5 s**<br>Significant |
| **Time to Release Control** | 6.1 s | 3.0 s<br>Not Significant | 3.9 s<br>Not Significant | 3.4 s<br>Not Significant |
| **Time to Resume a Non-Driving Task** | 13.8 s | 9.1 s<br>Not Significant | 10.4 s<br>Not Significant | 9.5 s<br>Not Significant |

* After a lane drift with no alert, 23 out of 49 participants did not regain control without instruction from the experimenter.
** This relationship held in Session 1 only.

**Table 3-7. Mean Values for No Alert with the 7-second Prompt Condition, Compared to Mean Values for Other Relevant Event Type/Prompt Condition Combinations in Experiment 2**

| Dependent Variable | No Alert 7-Second | No Alert 2-Second | No Alert No Prompts | Alert 7-Second |
|---|---|---|---|---|
| **Time to React** | 3.7 s | 2.8 s<br>Not Significant | 4.3 s<br>Not Significant | 1.0 s<br>Significant |
| **Time to Regain Control** | 3.9 s | 4.6 s*<br>Not Significant | 5.0 s*<br>Not Significant | 2.5 s*<br>Significant |
| **Time to Activate the Automation** | 5.6 s | 3.8 s<br>Not Significant | 4.9 s<br>Not Significant | 8.2 s**<br>Significant |
| **Time to Release Control** | 3.0 s | 6.1 s<br>Not Significant | 3.9 s<br>Not Significant | 3.7 s<br>Not Significant |
| **Time to Resume a Non-Driving Task** | 9.1 s | 13.8 s<br>Not Significant | 10.4 s<br>Not Significant | 8.8 s<br>Not Significant |

* After a lane drift with no alert, 23 out of 49 participants did not regain control without instruction from the experimenter.
** This relationship held in Session 1 only.

**Table 3-8. Mean Values for No Alert with the No Prompts Condition, Compared to Mean Values for Other Relevant Event Type/Prompt Condition Combinations in Experiment 2**

| Dependent Variable | No Alert No Prompts | No Alert 2-Second | No Alert 7-Second | Alert No Prompts |
|---|---|---|---|---|
| **Time to React** | 4.3 s | 2.8 s<br>Not Significant | 3.7 s<br>Not Significant | 1.1 s<br>Significant |
| **Time to Regain Control** | 5.0 s* | 4.6 s*<br>Not Significant | 3.9 s*<br>Not Significant | 2.6 s<br>Significant |
| **Time to Activate the Automation** | 4.9 s | 3.8 s<br>Not Significant | 5.6 s<br>Not Significant | 7.2 s**<br>Significant |
| **Time to Release Control** | 3.9 s | 6.1 s<br>Not Significant | 3.0 s<br>Not Significant | 2.6 s<br>Not Significant |
| **Time to Resume a Non-Driving Task** | 10.4 s | 13.8 s<br>Not Significant | 9.1 s<br>Not Significant | 6.2 s<br>Not Significant |

* After a lane drift with no alert, 23 out of 49 participants did not regain control without instruction from the experimenter.
** This relationship held in Session 1 only.

# Chapter 4 Level 3 Automated Vehicle, Experiment 3: Human-Automation System Performance Over Time

## Purpose

The goal of Experiment 3 was to examine how operators in a mixed-traffic (involving surrounding legacy vehicles, Level 0) simulated highway driving condition recognize and respond to hazards and how this is influenced by different visual allocation strategies. To that end, the effects of deactivation cues (i.e., take control alerts) on the overall operator behavior over time were analyzed. As stated in Chapter 1, the ultimate goal was to ascertain how operators interact with automated vehicles and determine how automated vehicle technology can best support safe driving. For this study, two aspects of the Level 3 operational definition (NHTSA, 2013) are of importance and set the tone for the experimental design and scenarios of interest:

- “The driver is expected to be available for occasional control, but with sufficiently comfortable transition time.”
- “… the vehicle is designed so that the driver is not expected to constantly monitor the roadway while driving.”

Based on these operational definitions, Experiment 3 was designed to study how operators respond to different alerts and driving scenarios, both by taking control from the automation and releasing control back to the automation.

## Method

For Experiment 3, a single-exposure experiment was performed. Participants were provided with a thorough familiarization with the vehicle and its operation, including use of the automated features. Training was followed by three 30-minute experimental sessions. During the sessions, participants had free exposure to a non-driving-related task (i.e., use of their own cell phone or the provided tablet as they felt was appropriate) and were presented with a message stating that they must take control of the vehicle. Participants’ reactions to these messages, both in duration and method of response, were examined in this experiment.

## Experimental Design

The study was performed as a within-subject design. Each participant completed three successive driving sessions, each session with one of the three alert types; all participants received all alert types exactly once. The three alert types were Staged, Imminent–External Threat, and Imminent–No External Threat.

## Independent Variables

The research design included two independent variables, detailed below.

### *Take Control Alert Type*

This is a within-subject variable with three levels. According to the Level 3 automation definition, the intention of the timing of these alerts was to provide "sufficiently comfortable" time for the participant to react in order to measure his/her reaction time to the alert onset. (If the participant failed to successfully regain control of the vehicle in the allotted time, the system would stop the vehicle.) Note that these timings are not values suggested for future design, but were deemed appropriate for this study to start the exploration of the different reaction times of interest and participants' behavior.

- Staged. Involved an approximately 50-second period of increasing urgency for the participant to take control of the automation before the vehicle would come to a stop. This was triggered by a falsified system failure message indicating that the participant needed to deactivate the automation and regain manual control. (The system failure message was experimenter-injected to ensure participant eyes were away from the forward roadway; however, some participants looked at the roadway as the alert was triggered.)

  The Staged alert represented a lower-urgency internal system failure (e.g., the system is losing confidence in its ability to detect the lane of travel due to developing fog). The system may still be able to function properly, but as visibility and the system's confidence decrease, the operator will need to eventually regain manual control. Based upon the training video, participants understood that the Staged alert progressed from a low-urgency informational message informing them to prepare for manual control to a medium-urgency message asking them to please turn off the automation, and ultimately to a high-urgency message instructing them to "turn off the automation now."

- Imminent–No External Threat. This alert, which consisted of autonomous braking in tandem with auditory and visual cues, involved an approximately 10-second period for the participant to take control from the automation before the vehicle would come to a stop. It began with the onset of the higher-urgency auditory alert along with a visual HMI alert instructing the participant to "turn off autodrive now" and it ended when the participant turned off the autodrive system or the system stopped the vehicle.

  This alert represented a higher-urgency internal system failure such as what may occur when the system's confidence in the vehicle's location rapidly degrades (i.e., it was intended to imply that there may be an internal system failure as opposed to an external physical threat). The system may no longer be able to maintain the vehicle's lane positioning for much longer, and the operator needs to regain manual control.

- Imminent–External Threat. This alert, which consisted of autonomous braking in tandem with auditory and visual cues, involved an approximately 10-second period for the participant to take control from the automation before the vehicle would come to a stop. It began with the onset of the higher urgency auditory alert along with a visual HMI alert instructing the participant to "turn off autodrive now" and it ended when the participant turned off the autodrive system or the system stopped the vehicle. This message was issued in conjunction with a box reveal that allowed the experimenter to trigger the 10-second time-to-collision (TTC) warning.

As evidenced by the name of this alert, it was issued to represent situations in which a physical threat external to the vehicle exists.

### *Session*

This is a within-subject variable with three levels; each participant's experience occurred in three consecutive sessions. Each participant experienced one alert type per session according to the order assigned to that participant. The three sessions assisted in representing the participants' performance over time.

## Dependent Variables

The research design included nine dependent variables, detailed below.

### *Operator Behavior*

- **Time to react to alert** (in seconds)
  This was defined as the time, after the alert is first presented, at which point the participant performs an action that could be considered a reaction to the alert (e.g., look forward, move foot or hand).

- **Time to regain control** (in seconds)
  This was defined as the time which begins when the alert is first presented until the participant attempts to regain manual control by taking control of the steering wheel.

- **Performance**
  This was defined as the correct action taken by the participant in response to an event of interest (i.e., does the participant regain control upon receiving the alert?).

- **Method used to regain control/cancel automation**
  This was defined as the method that the participant first uses to attempt to regain manual control of the vehicle (i.e., pressing the off button, steering input, or pedal input).

- **Time to release control**
  This variable consisted of two components.

  - <u>Time to activate the automation</u> (in seconds)
    This was defined as the time from the availability of autonomy as indicated by the system until the participant attempts to activate the automation. Note that during the Staged alerts, the system issue resolved by itself (i.e., the alert turned off but the automation did not turn off) and, thus, most participants did not have a time to activate the automation for this alert type.

  - <u>Time to release control of steering</u> (in seconds)
    This was defined as the time from successfully activating the automation until the participant removes both hands from the steering wheel.

- **Time to resume non-driving task** (in seconds)
  This was defined as the time from releasing control of the steering wheel until the participant resumes interaction with a non-driving-related task.

The time it took for participants to attempt to activate the automation, to release control of the steering wheel upon activating the automation, and the time to resume a non-driving task were of interest as the researchers speculated that these variables might be tied to trust in the automation. Specifically, the times noted for participants to perform the aforementioned tasks may be reflective of high or low levels of trust in the automated system.

Figure 4-1 illustrates the sequence in which the dependent variables were measured.

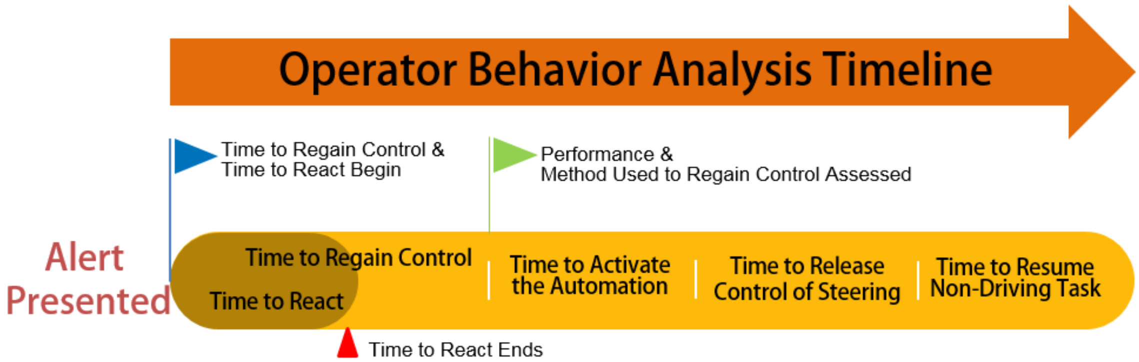



**Figure 4-1. Sequence of the Dependent Variables Pertaining to Operator Behavior for Experiment 3**

***Eye-glance Behavior***

Eye-glance behavior was sampled at three time points during each driving session. Monitoring rate and non-driving-related glances were measured during a 10-second interval starting at 65 s prior to the alert until 55 s prior to the alert, a 10-second interval starting at 10 s prior to the alert until the onset of the alert, and an interval starting at the onset of the alert until the operator successfully regained manual control of the vehicle. There were two components to eye-glance behavior:

- **Monitoring rate**
  This was defined as the percentage of driving-related glance time in a given time interval.

- **Non-driving-related glances**
  This was defined as the number and duration of non-driving-related glances in a given time period.

Various terms related to eye-glance behavior and their operational definitions are provided in Table 4-1.

**Table 4-1. Operational Definitions for Experiment 3 Eye-glance-related Terms**

| Eye-glance-related Term | Operational Definition |
|---|---|
| **Driving-Related Glance** | Glance pertaining to the driving scenario. Includes the following glances: forward, left and right windshield, left and right windows, left and right side mirrors, rear-view mirror, HMI, and the instrument cluster. |
| **Non-Driving-Related Glance** | Glance not pertaining to the driving scenario. Includes the following glances: smartphone, tablet computer, in-vehicle experimenter, hygiene-related glances, etc. |
| **Visually Engaged in a Non-Driving-Related Task** | The participant is allocating his/her visual attention to a non-driving-related task. For example, the participant is watching a movie on the tablet computer and is devoting his/her visual attention to the device. |
| **Holding Device—Not Visually Engaged in a Non-Driving-Related Task** | The participant is holding an electronic device (i.e., tablet or smartphone); however, he/she is not visually engaged with the device(s) and appears to be monitoring the driving scenario. |
| **Not Visually Engaged in a Non-Driving-Related Task** | The participant is not holding an electronic device (i.e., tablet or smartphone) and is not visually engaged in a non-driving-related task and he/she appears to be monitoring the driving scenario. |

# Participants' Subjective Assessment

A trust scale (Appendix E) was administered at 10-minute intervals throughout the experimental session. Participants were asked to rate their trust in the ability of the automation to function properly while they engaged in non-driving tasks using a 7-point Likert-type scale in which "1" corresponded with strongly disagree and "7" corresponded with strongly agree.

Twelve trust ratings were collected, which translates to four trust ratings in each driving session. In addition to the trust scales collected throughout the experimental sessions, upon completing all three driving sessions, participants were asked to complete the after-experience trust scales (Appendix E) and participate in an open-ended interview (Appendix F).

# Facility

This experiment was conducted on the Virginia Smart Road test track, which is located at VTTI in Blacksburg, VA (Figure 4-2). The test track is constructed to state and federal roadway standards and has a length of 3.5 km (2.2 mi), with looped turns at either end. The straight section of the track is approximately 0.8 km (0.5 mi) in length. Two lanes run the duration of the track, with the exception of the looped turns. Wireless Internet coverage is available on the track. The facility is closed to outside traffic and only study-related vehicles were present during the experiment.

**Figure 4-2. Aerial View of the Virginia Smart Road Test Track at VTTI**

# Vehicle

A 2012 Lexus RX450h was used as the experimental vehicle for Experiment 3 (Figure 4-3). This L2 vehicle was equipped with a prototype automated driving system that can simulate L3 driving on a test track. As part of the prototype system, several HMI components were installed. These included an instrument panel binnacle-mounted screen providing information on the automated driving system and two steering wheel buttons to control the automation: one on button on the left side of the wheel and one off button on the right side of the wheel.

**Figure 4-3. The 2012 Lexus RX450h Used in Experiment 3**

The vehicle was instrumented with VTTI's DAS. The DAS was connected to the automated driving system via Ethernet. The variables collected by the DAS included throttle/brake input and automation state. In addition, the following video views were collected (shown in Figure 4-4):

- Operator's face
- Over the shoulder
- Forward roadway
- Exterior left rear
- Foot (pedal area)
- HMI

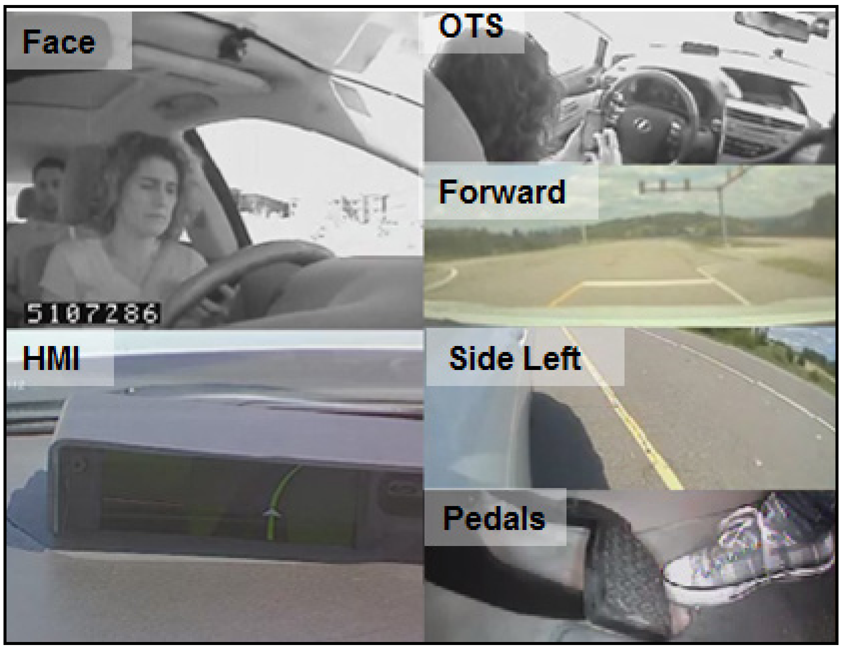



**Figure 4-4. VTTI's DAS Camera Views from Experiment 3**

## Tablet Computer

The same type of Asus Nexus 7 tablet computer that was used in Experiments 1 and 2 was also used for this study; this provided opportunities for the participants to engage in non-driving tasks while the vehicle was in Level 3 automation (Figure 4-5). However, for Experiment 3 additional programs were installed that had not been available during the previous two experiments. Because this experiment involved a vehicle with simulated L3 capabilities, games and movies were made available on the tablet, which was able to be used in any position that was comfortable for the participant. The tablet was connected to the Smart Road's 4G/LTE wireless modem and participants were allowed to perform different tasks with it at will (e.g., watch preloaded movies, play games, search the Internet, check email). In addition to the tablet computer that was provided, the participants were allowed to use their personal smartphones for email, texting, and other tasks they deemed appropriate during the study (but only when the vehicle was in Level 3 automation).

**Figure 4-5. Asus Nexus 7 Tablet Computer**

## Participants

Data were collected from 37 participants; however, 12 participants were considered invalid (i.e., session cancellation due to adverse weather, technical issues associated with the DAS or the prototype vehicle). The analysis presented in this chapter consists of data from 25 participants (17 males, 8 females). The mean age of the participants was 38.8 years old (S.D. = 13.77), with ages ranging from 18 to 69 years old. Participants were categorized into four different age groups to ensure that all age groups were represented: 18–24, 25–39, 40–54, and 55 and older; of which, four participants were between the ages of 18 and 24 (2 males, 2 females), nine were between the ages of 25 and 39 (6 males, 3 females), eight were between the ages of 40 and 54 (5 males, 3 females), and four were 55 and older (4 males, 0 females). However, age group was not a dependent variable for this study.

## Instruction and Training

Prior to receiving any hands-on training, the participant viewed a 10-minute video summarizing the vehicle's features with a specific focus on the automated components and operation of the vehicle. This video was a training requirement of the automated vehicle provider and is consistent with the recommendation in NHTSA's *Preliminary Statement of Policy Concerning Automated Vehicles* in the section entitled "Recommendations Concerning State Activities Related to Self-Driving Vehicles" (NHTSA, 2013; pp. 10-11). It is imperative to note that the objective of this research effort was not to develop a comprehensive training curriculum but to ensure that the participants understood how to operate a limited self-driving vehicle safely, as suggested by NHTSA for Level 3. This video was intended to detail the prototype system's operating capabilities and limitations. The participant was also shown the different types of the alerts (i.e., Staged, Imminent–External Threat [ET], and Imminent–No External Threat [NET]) during the video.

The participant was then escorted to the experimental vehicle. There were two in-vehicle experimenters: one front-seat experimenter and one back-seat experimenter. The front-seat experimenter drove the vehicle to the test track, and then completed one lap on the test track demonstrating how to activate and deactivate the automated system (termed “autodrive”). The prototype system clearly depicted when the system was read for activation of the automation. Additionally, the in-vehicle experimenter demonstrated the vehicle’s ability to accommodate a slower-moving vehicle. During this demonstration, the back-seat experimenter conspicuously notified via radio communication the driver of one of the two confederate vehicles used in this experiment to pull out in front of the subject vehicle. The lead vehicle dropped its speed to 25 mph, which required the automated system to adjust the subject vehicle’s speed. After this lap, the participant drove one lap around the track in manual mode, which allowed the participant to become familiar with both the vehicle and the test track environment and assisted the in-vehicle experimenters in gauging the participant’s driving ability. Note: Experimenters followed a set script and protocol in order to ensure consistency between experimenters.

During the second lap, the participant practiced activating the automation. The participant also practiced deactivating the automation using the following methods (in the order provided): pressing the off button on the steering wheel, steering wheel input, pressing the throttle, and pressing the brake. Additionally, during this lap, the back-seat experimenter triggered the lead vehicle to pull out in front of the subject vehicle and decrease its speed, which allowed the participant to experience the system accommodating the lead vehicle. Following the second lap, the participant completed two more laps using the automated system.

Next, the participant was asked to return the vehicle to the preparation area and park. At this point, the participant was given the opportunity to ask questions. The participant received instructions on the study procedure, including the instruction that the participant could disengage the automation and assume manual vehicle control at any point if he or she was not comfortable with the automated driving system. Participants were also informed that they were free to engage in non-driving tasks when using the automated system if they felt comfortable doing so. This orientation portion of the participant’s experience took approximately 30 minutes to complete.

## Study Session

The experiment consisted of three 30-minute driving sessions. Each session began when the participant was instructed to enter the test track. The participant entered the track and traveled to the preparation area. Once at this location during the first session, the participant was instructed to enable automated driving and engage it when ready. The participant was then asked to use the automated driving system for the remaining portion of the session. (For the successive sessions, the participant was instructed to engage the automated driving system and activate it upon entering the track.)

At a randomly selected point within each session, one of the three alert types was presented: Staged, Imminent–External Threat, or Imminent–No External Threat. The alerts happened at a predetermined location and participants experienced all three alert types during the experimental session. While each participant experienced all the alert types, the alert types were not always experienced in the same order. A 15-minute break was offered after each session to allow for participant comfort; however, some participants chose to forgo the breaks. The maximum speed for all sessions was 72 km/h (45 mph), with lower speeds used for the turns at both ends of the track.

An illustration depicting the alert timelines appears in Figure 4-6 below, followed by descriptions of the alerts. Participant responses to these system messages were measured in terms of response times using visual evidence from the DAS. A diagram of the Virginia Smart Road test track (Figure 4-7) shows the locations of the test events that are mentioned in the alert descriptions.

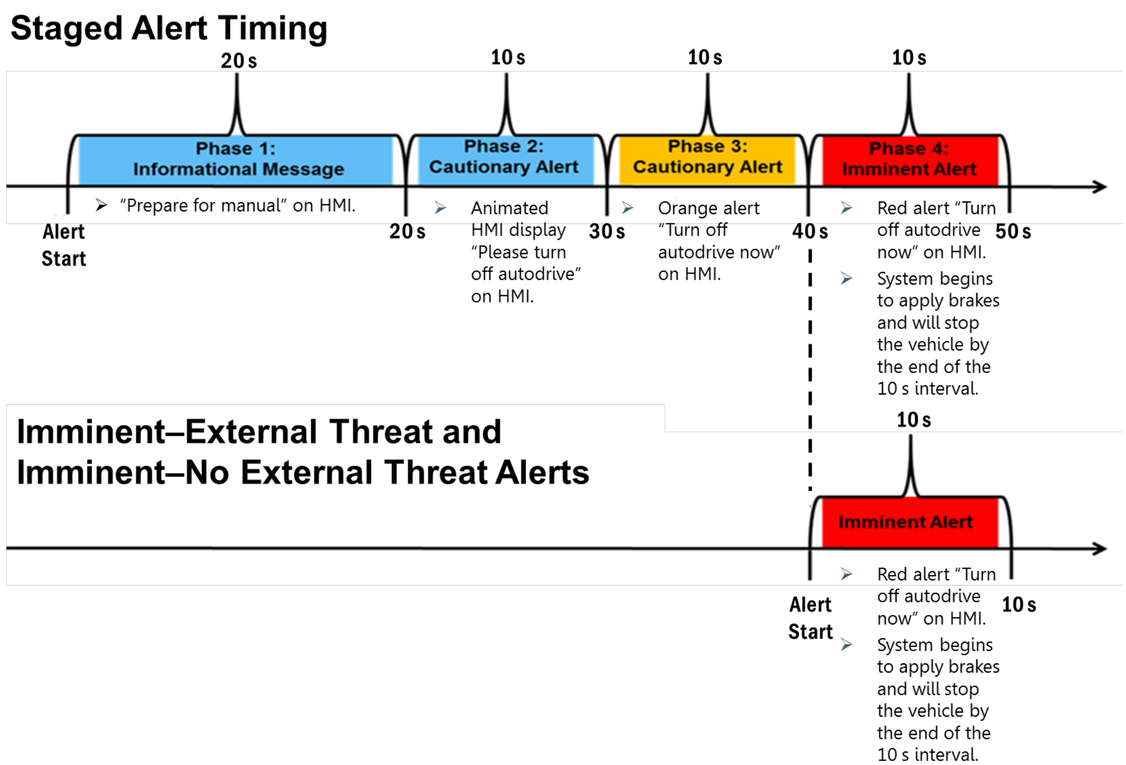



**Figure 4-6. Staged and Imminent Take Control Alert Timelines for Experiment 3**

***Staged Alert***

The Staged alert consisted of four phases of increasing urgency and contained both visual and auditory components. Figure 4-6 shows that the Staged alert was made up of an informational phase, followed by two cautionary alert phases, and then an Imminent alert phase. In contrast to both of the Imminent alerts, the Staged alert provided the participant with an informational message phase during the first 20 s, which progressed to two cautionary alert phases of 10 s each, which allowed a total of 40 s for a participant response before it progressed to an Imminent alert Figure 4-6). As shown in Figure 4-7, the Staged alert was issued when the vehicle was proceeding uphill, before the test track's top turnaround.

***Imminent–External Threat (ET) Alert***

A standard empty cardboard box (maximum dimensions: 75 cm by 50 cm by 20 cm, length/width/height) was placed one third of the way (approximately 257 m, or 0.16 miles) on the Smart Road's straightaway in the uphill direction. This was accomplished surreptitiously by the experimenter in the side vehicle, which, at that point, had pulled into a turnaround to park, while the lead and subject vehicles continued on the test track. At a TTC of approximately 10 s from the object (approximately 200 m, or 660 ft.) the lead vehicle would rapidly swerve into the adjacent lane, revealing the box that had been placed in the center of the participant's travel lane. (The box would not have been viewable by the participant until this point.) At this point, the Imminent–External Threat alert triggered. The participant could take control of the vehicle by any of the four disengage methods (braking, accelerating, pressing the off button, or turning the steering wheel [likely swerving into the adjacent lane following the lead vehicle's path]). A collision between the box and vehicle would not damage the vehicle. (However, no collisions happened during the study.)

In the downhill direction, upon entering the straightaway, the system was triggered to provide the Imminent–No External Threat alert (i.e., simulation of a warning due to a system problem). No details were provided to the participant about the magnitude or reason for the alert request by the system or experimenter. The participant could take control of the vehicle at that point by any of the four disengage methods (braking, accelerating, pressing the off button, or turning the steering wheel).

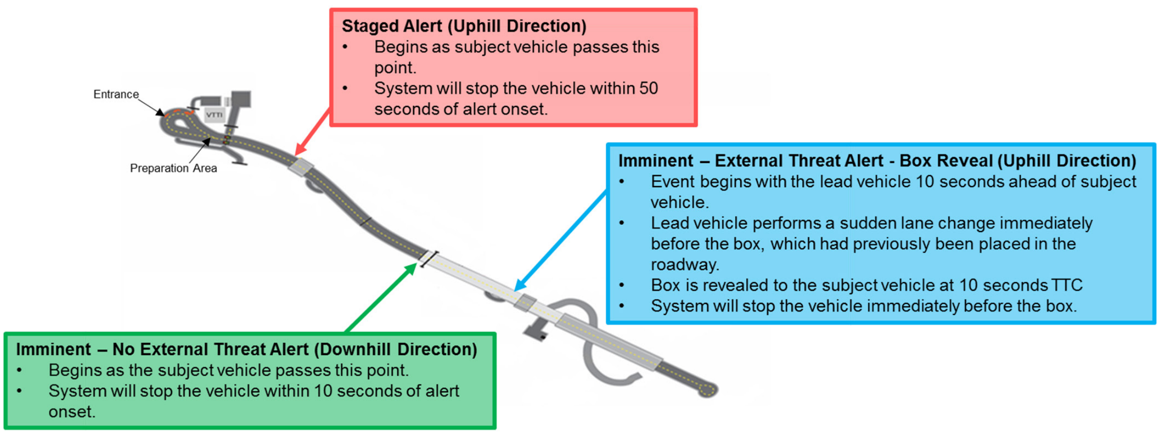



**Figure 4-7. Take Control Alert Locations on the Virginia Smart Road Test Track**

During the remainder of the trials, at any given point in time, either the lead vehicle, the side vehicle, or both confederate vehicles, were in close proximity to the participant's vehicle. The lead vehicle maintained a 10-second headway while the side vehicle remained two car lengths behind the subject vehicle. The side vehicle was only used on the two-lane sections of the track (i.e., the side vehicle was not present on turnarounds). The lead vehicle was used at multiple points along the course. Confederate vehicles entered and exited the participant's driving path using the pullouts and intersection present on the track. The possible combinations of confederate vehicles given the above constraints (i.e., lead, lead and side, side) were varied across laps and sessions. Figure 4-8 depicts two examples of the traffic configurations used.

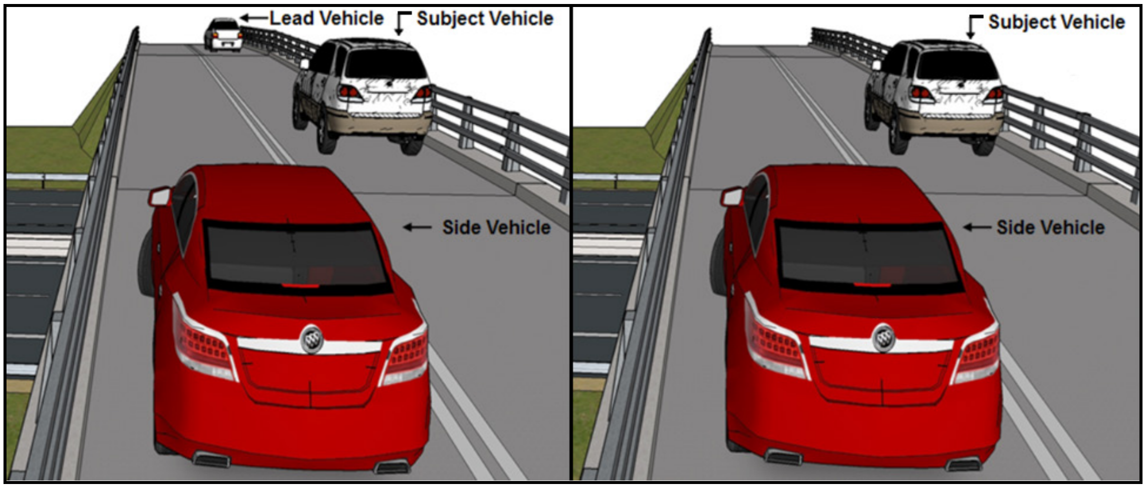



**Figure 4-8. Example Traffic Configurations from Experiment 3**

The trust scale (Appendix E) was presented at 10-minute intervals during each session (at the beginning of the session, followed by administrations after 10, 20, and 30 minutes). Upon completion of the third session, the participant was instructed to deactivate the vehicle's automation, assume manual control over the vehicle, exit the track, and return to the preparation area. An interview was performed at that point. At the end of the interview, compensation for participation in the study was provided. Figure 4-9 presents the timeline of events for Experiment 3.

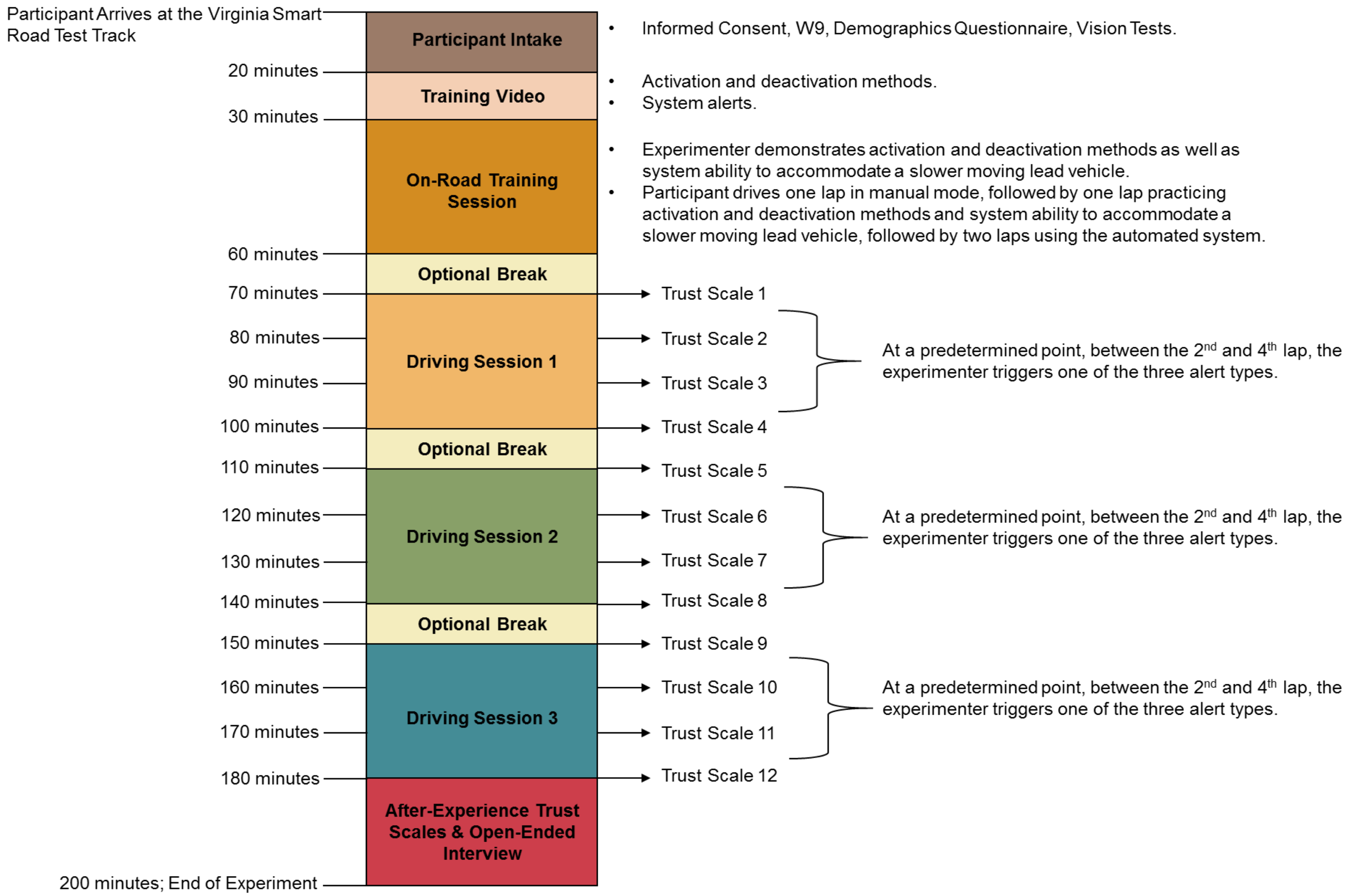



**Figure 4-9. Experimental Session Timeline for Experiment 3**

# Results

## Operator Behavior

In general, a key question is whether or not there are significant differences between Imminent alerts with and without an external threat in these measures. However, it may also be of interest to determine whether Staged alerts differed from Imminent alerts in some situations. For example, it would be expected that participants take more time to regain control after the onset of a Staged alert due to the alert being less urgent in nature. However, it would be a noteworthy result if participants, in fact, did not take significantly longer to regain control after a Staged alert. Therefore, for each operator behavior variable, Imminent and Staged alerts are briefly explored separately, and then briefly compared.

### *Time to React*

Imminent Alerts

There was no significant change in reaction time between Imminent alerts with and without an external threat, $p > .05$. Participants' mean reaction time to an Imminent alert was 0.69 s with an external threat (S.E. = 0.06 s, n = 25, min = 0.21 s, max = 1.68 s) and 0.66 s without an external threat (S.E. = 0.04 s, n = 25, min = 0.20 s, max = 1.01 s). Meanwhile, time to react did not significantly change over time.

Staged Alerts

There was a significant change of the time to react to Staged alerts depending on whether or not participants were looking forward at the time of the alert. Specifically, if participants experienced a Staged alert and were looking forward at the time of the alert (M = 2.45 s, S.E. = 0.82 s, n = 5, min = 1.01 s, max = 5.31 s), they reacted to the alert more slowly than if they were not (M = 0.61 s, S.E. = 0.1 s, n = 20, min = 0.27 s, max = 2.21 s). Interestingly, all five instances in which the participant was looking forward at the time of a Staged alert occurred during the third session. Meanwhile, reaction time did not significantly change over time.

Comparison of Imminent and Staged Alerts

In general, the participants looking forward at the time of the Staged alert had significantly higher reaction times than did participants who experienced an Imminent alert and were looking forward, whether there was an external threat, $t(22.9) = 3.85$, $p = .0024$, or not, $t(21.7) = 3.99$, $p = .0018$. However, participants who were not looking forward at the Staged alert did not react significantly slower than did participants who received an Imminent alert, regardless of whether they were looking forward or whether they experienced an external threat, $p > .05$.

In Figure 4-10 and Figure 4-11 mean and standard error bar plots are provided for time to react, with Figure 4-10 showing times after the onset of the alert, and Figure 4-11 showing times relative to the onset of the imminent phase. The purpose of including both graphs is to demonstrate that although operators may react slower to the Staged alert than the Imminent alert, they may still react with ample time before the Staged alert becomes imminent. Specifically, in Figure 4-10, the time to react to the Staged alert is shown to exceed that of Imminent alerts (though, as indicated in the statistical analysis, the difference was significant only if the participants were looking forward at the time of the Staged alert). However, in Figure 4-11, it is shown that when participants experienced the Staged alert, they reacted long before the alert reached its imminent phase. This reflects the usefulness of alerting operators to dangers with ample time to react, but also demonstrates that the alert given should reflect the severity of the driving situation. (For additional details on the statistical analysis, see Appendix D, Results, Operator Behavior, Analysis for Time to React, Statistical Analysis.)

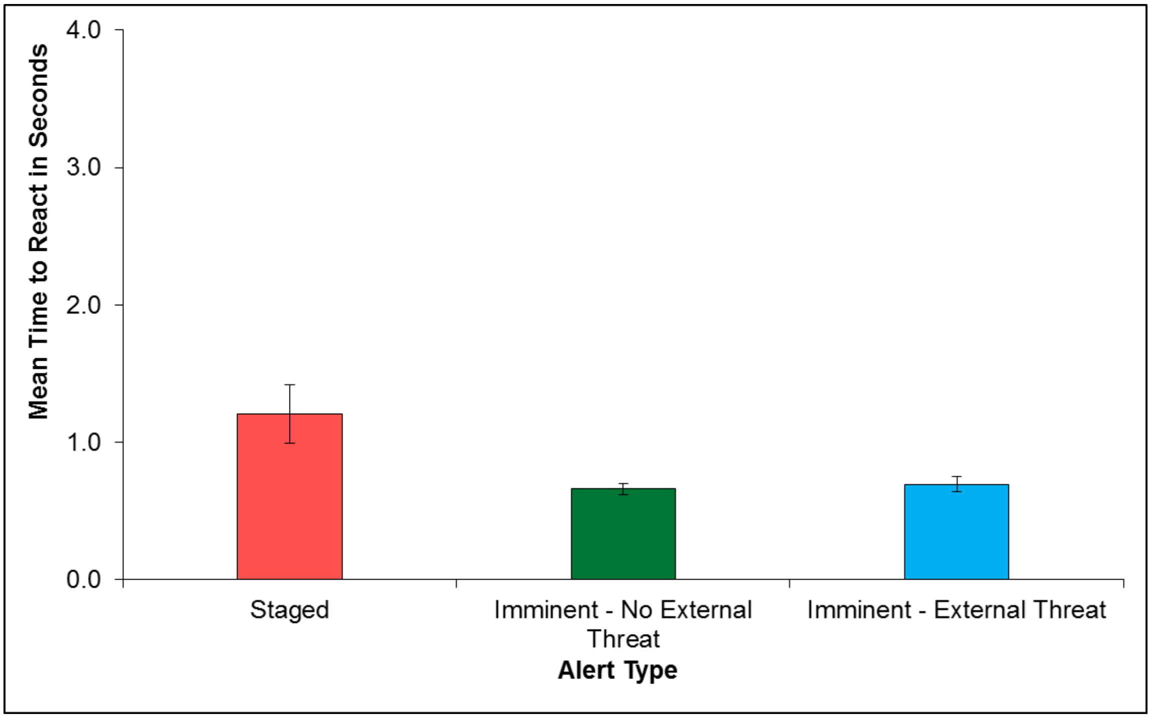



**Figure 4-10. Time to React to Alert, Stratified by Alert Type, After the Onset of the Alert**

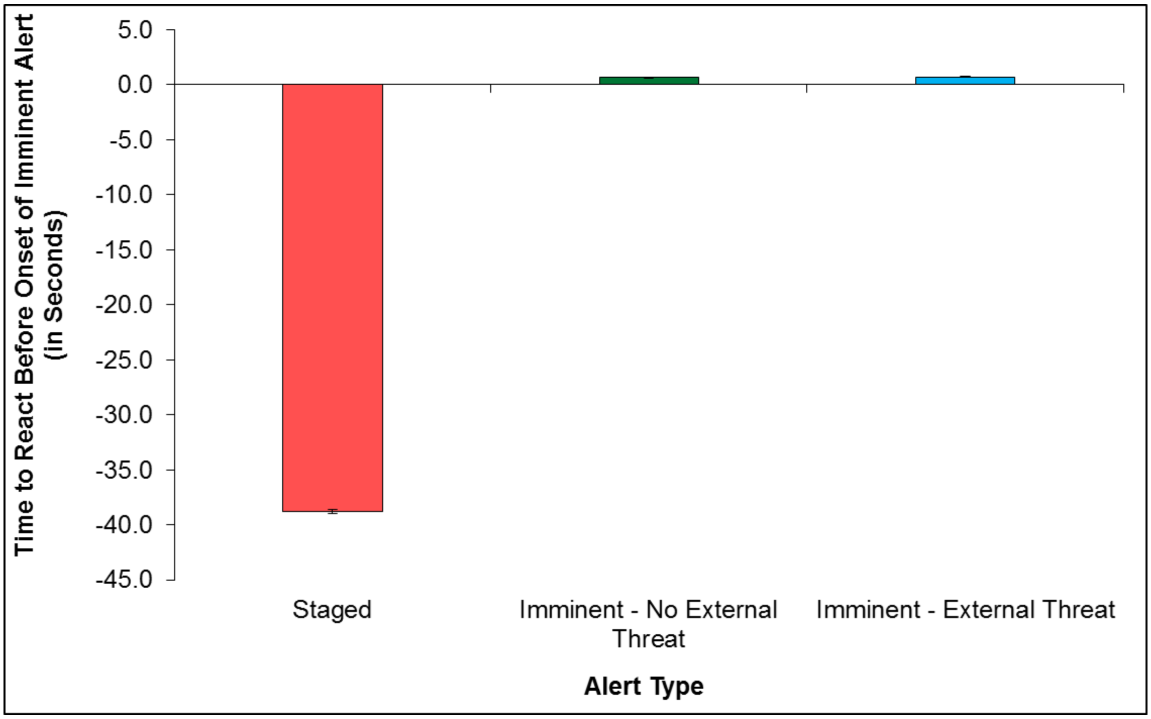



**Figure 4-11. Time to React to Alert, Stratified by Alert Type, Relative to the Onset of the Imminent Phase**

***Time to Regain Control***

<u>Imminent Alerts</u>

Imminent alerts with and without external threats did not significantly differ from each other in terms of time to regain control, $p > .05$. Further, there was no significant change over time. There was a significant effect of alert type on the time to regain control, $F(2,21) = 68.18$, $p < .0001$. On average, participants regained control after an Imminent alert in 2.07 s if accompanied by an external threat (S.E. = 0.14 s, n = 25, min = 0.69 s, max = 3.59 s) and 2.31 s if not accompanied by an external threat (S.E. = 0.18 s, n = 25, min = 1.10 s, max = 4.99 s).

<u>Staged Alerts</u>

Following a Staged alert, participants regained control after a mean of 16.96 s (S.E. = 1.25 s, n = 25, min = 6.21 s, max = 23.95 s). Fourteen of 25 participants (56 percent) regained control prior to the end of the informational phase of the Staged alert (20 s), while the remaining 11 participants (44 percent) regained control within the first cautionary phase (between 20 s and 30 s).

<u>Comparison of Imminent and Staged Alerts</u>

Participants took significantly longer to regain control after a Staged alert than they did after Imminent alerts, whether the Imminent alert had an external threat, $t(22) = 11.93$, $p < .0001$, or not, $t(22) = 11.75$, $p < .0001$. It should be noted that although participants look longer to regain control after the onset of the Staged alert, participants still regained control, on average, with ample time before the Staged alert progressed to the imminent phase.

In Figure 4-12 and Figure 4-13 mean and standard error bar plots are provided for time to regain control, with Figure 4-12 showing times after the onset of the alert, and Figure 4-13 showing the times relative to the onset of the imminent phase. The purpose of including both graphs is to demonstrate that although operators may regain control slower after the Staged alert than after the Imminent alert, they may still regain control with ample time before the Staged alert becomes imminent. Specifically, in Figure 4-12, the time to regain control after the Staged alert is shown to exceed that of Imminent alerts. However, in Figure 4-13, it is shown that when participants experienced the Staged alert, they regained control long before the alert reached its imminent phase. This reflects the usefulness of alerting operators to dangers with ample time to regain control, but also demonstrates that the alert given should reflect the severity of the driving situation. (For additional details on the statistical analysis, see Appendix D, Results, Operator Behavior, Analysis for Time to Regain Control, Statistical Analysis.)

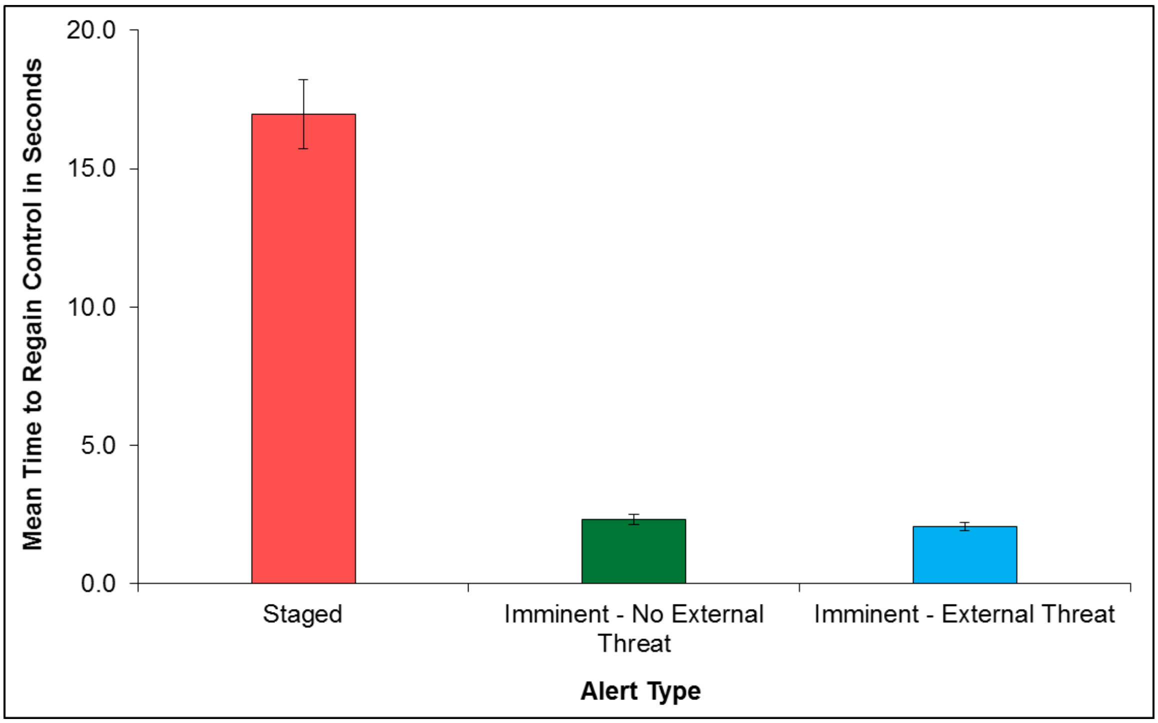



**Figure 4-12. Time to Regain Control, Stratified by Alert Type, After the Onset of the Alert**

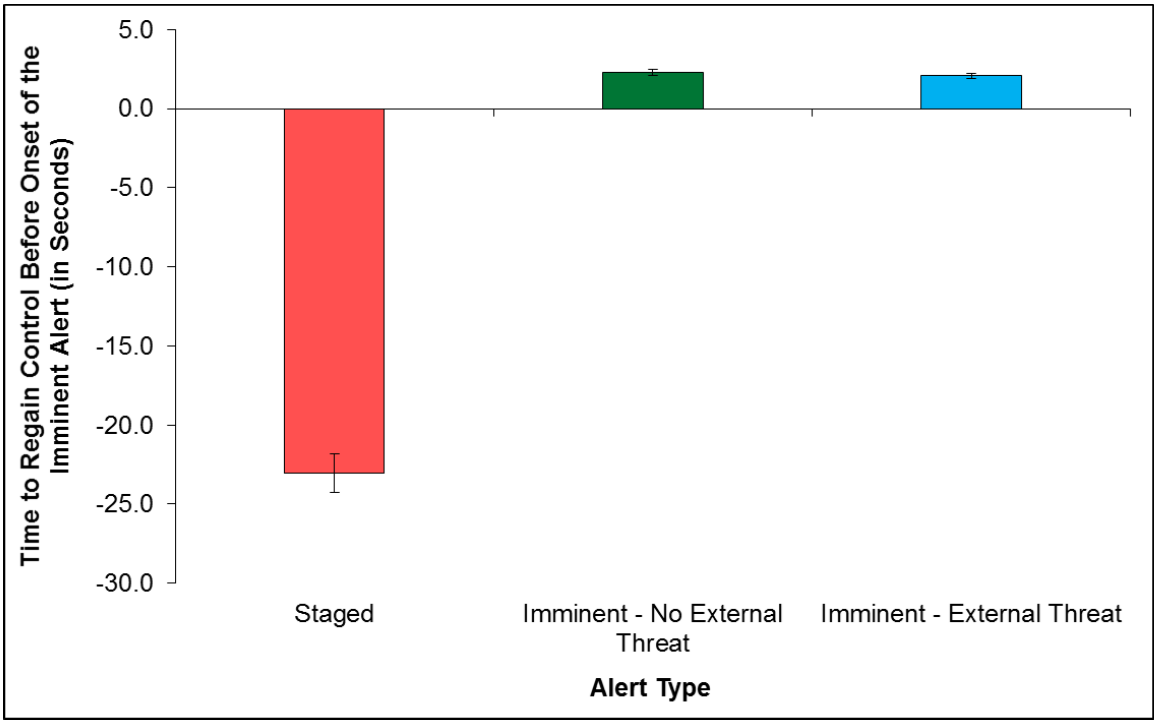



**Figure 4-13. Time to Regain Control, Stratified by Alert Type, Relative to the Onset of the Imminent Phase**

***Performance***

Characterizing the actions participants take is as important as their reaction times. Participants might react to an alert (e.g., look ahead) but not exhibit the appropriate response for a given scenario (e.g., take control of the vehicle as opposed to continuing a non-driving-related task). For this experiment, all participants attempted to regain control of the vehicle, which was the desired performance the alerts were trying to elicit. However, one of the participants did not perform this successfully (i.e., did not apply enough force to pedals to regain control of the vehicle). For this particular instance, the vehicle was able to stop as intended without the participant deactivating the automation. The analysis included the participant's performance for this particular case as appropriate, given that the attempt to regain control of the system was the correct response, even though the attempt was not successful. Thus, since all participants successfully attempted to regain control of the system when appropriate, all participants were considered to have performed correctly in all instances.

***Method Used to Regain Control/Cancel Automation***

For both types of Imminent alerts, the majority of participants elected to use the brake pedal in order to regain manual control. However, during the Staged alert, the majority of participants elected to use the off button on the steering wheel in order to regain manual control. This was likely due to the animated HMI display depicting an operator regaining control by using the off button, which was only presented during the Staged alert.

### ***Time to Activate the Automation***

Imminent Alerts

There was no significant difference in time to activate the automation between Imminent alerts with and without an external threat, nor was there a significant change over time, $p > .05$. The mean time to activate the automation was 7.97 s with an external threat (S.E. = 1.64 s, n = 25, min = 0.46 s, max = 28.28 s) and 6.25 s without an external threat (S.E. = 1.03 s, n = 25, min = 1.42 s, max = 20.83 s).

Staged Alerts

The time to activate the automation after staged alerts did not significantly change over time, $p > .05$. Participants' mean time to activate the automation for Staged alerts was 4.07 s (S.E. = 0.64 s, n = 25, min = 0.56 s, max = 11.96 s). See Appendix D for additional details.

Comparison of Imminent and Staged Alerts

There was no significant difference between Staged alerts and Imminent alerts, with or without an external threat, in terms of participants' time to activate the automation. (For additional details on the statistical analysis, see Appendix D, Results, Operator Behavior, Analysis for Time to Activate the Automation, Statistical Analysis.)

### ***Time to Release Control of Steering***

Imminent Alerts

There was no statistically significant difference in time to release control after an Imminent alert if there was or was not an external threat. The mean time to release control after an Imminent alert without an external threat was 1.83 s (S.E. = 0.31 s, n = 25, min = 0.44 s, max = 7.22 s), and with an external threat the mean was 2.67 s (S.E. = 0.49 s, n = 25, min = 0.63 s, max = 8.83 s). No significant change over time was found.

Staged Alerts

After a Staged alert, participants released control with a mean of 1.37 seconds (S.E. = 0.21 s, n = 25, min = 0.44 s, max = 5.78 s). There was no significant change from session to session in the time to release control after Staged alerts.

Comparison of Imminent and Staged Alerts

Participants took significantly longer to release control of steering if there was an Imminent alert with an external threat than they did if there was a Staged alert, $t(22) = 3.58$, $p = .0050$. However, there was no significant difference between the Staged alert and the Imminent alert without an external threat, $p > .05$. (For additional details on the statistical analysis, see Appendix D, Results, Operator Behavior, Analysis for Time to Release Control, Statistical Analysis.)

### ***Time to Resume a Non-driving Task***

Imminent Alerts

There was no statistically significant change in time to resume a non-driving task between Imminent alerts with and without a non-driving task, nor was there a significant change over time,

$p > .05$. The mean time to resume a non-driving task after an Imminent alert was 13.89 s without an external threat (S.E. = 4.18 s, n = 23, min = 0.70 s, max = 76.14 s) and 26.30 s with an external threat (S.E. = 9.55 s, n = 22, min = 0.90 s, max = 167.45 s).

Staged Alerts

After Staged alerts, there was no significant change between sessions of time to activate the automation, $p > .05$. Participants' mean time to resume a non-driving task after a Staged alert was 39.10 s (S.E. = 14.61 s, n = 24, min = 1.13 s, max = 258.30 s).

Comparison of Imminent and Staged Alerts

There was no significant difference between Staged alerts and Imminent alerts, with or without an external threat, in terms of their time to resume a non-driving task. (For additional details on the statistical analysis, see Appendix D, Results, Operator Behavior, Analysis for Time to Resume a Non-Driving Task, Statistical Analysis.)

***Summary of Operator Behavior Analysis***

This analysis helped to determine if the alert type affected operator behavior and if this effect changed over time, which will aid in determining whether the effect over time should be considered in future research investigations. Table 4-2 provides a summary of the results and general takeaways for time to react, time to regain control, time to activate the automation, time to release control of steering, and time to resume a non-driving task. Alert type and session were marked as "Significant" if they achieved statistical significance, and "Not Significant" if they did not. More detail on the analysis is available following the table. Each variable is explored both with graphs and descriptive statistics, and with statistical analysis to determine if differences are statistically significant. Mean and error bar plots of the dependent variables related to operator behavior within Imminent alerts are displayed in Figure 4-14, and for Staged alerts in Figure 4-15.

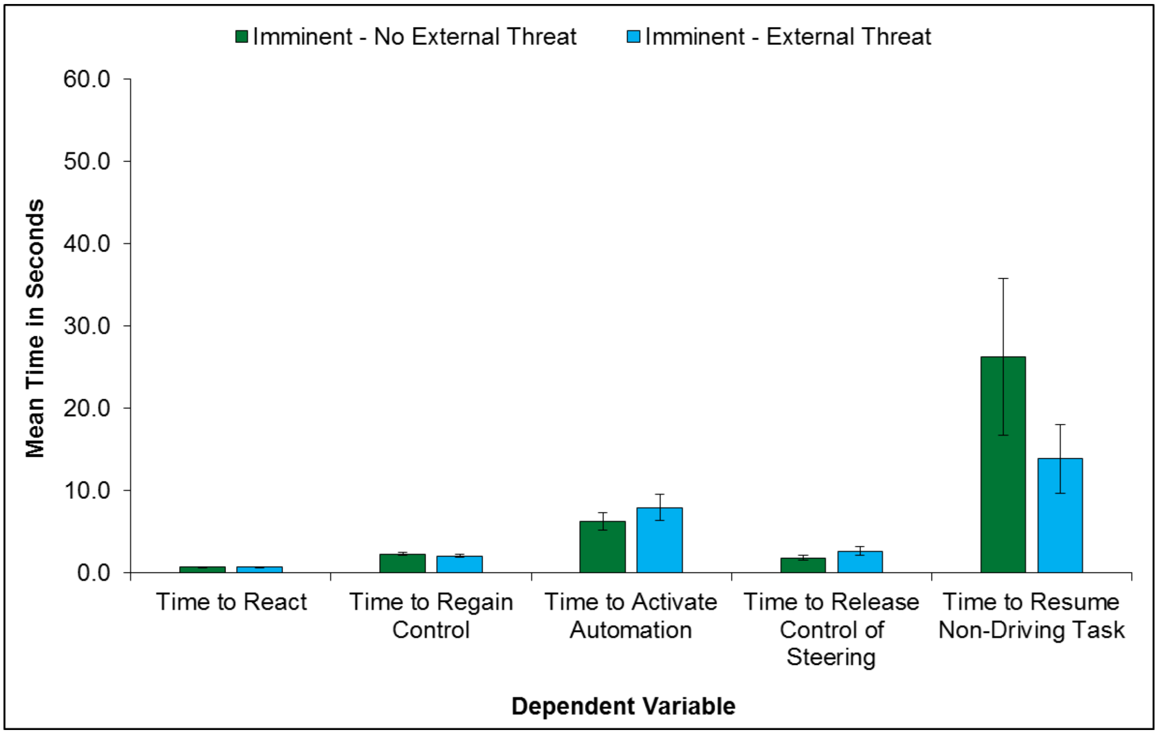



**Figure 4-14. Mean and Error Bar Plots, Operator Behavior Variables, Within Imminent Alerts**

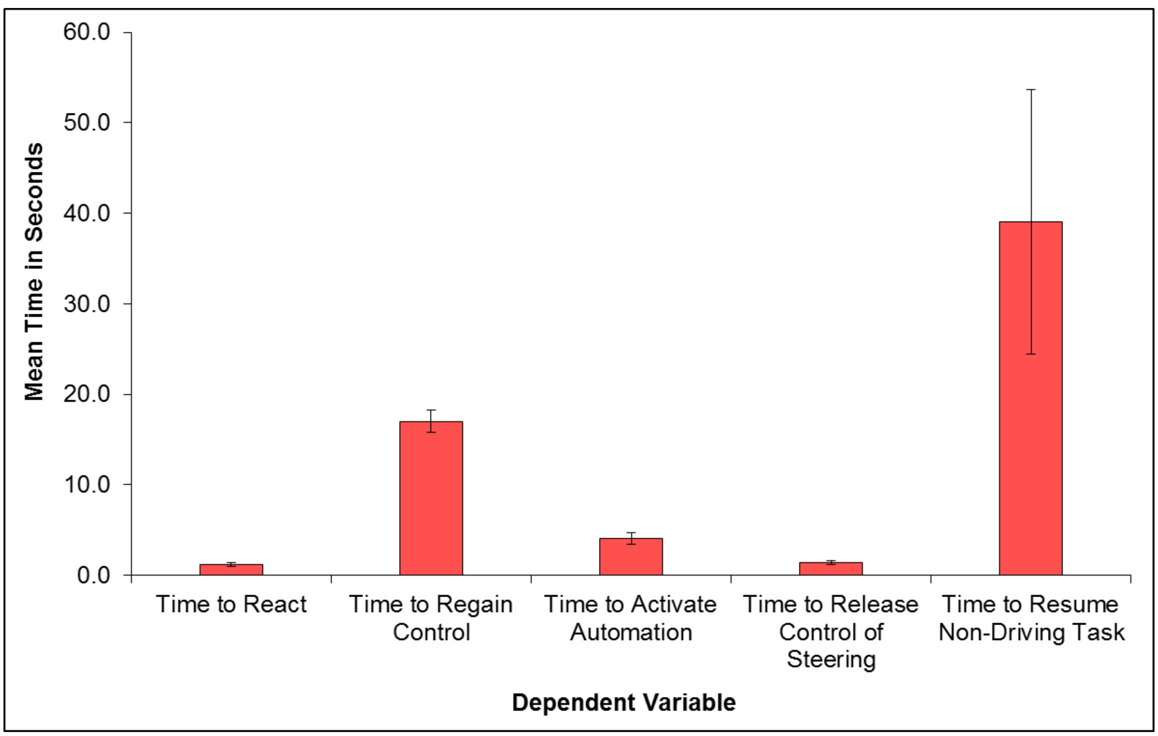



**Figure 4-15. Mean and Error Bar Plots, Operator Behavior Variables, Within Staged Alerts**

**Table 4-2. Summary Table for Operator Behavior Analysis for Experiment 3**

| Dependent Variable | Alert Type | Performance Over Time (Represented by Session) | Takeaways |
|---|---|---|---|
| **Time to React** | **Significant** | Not Significant | Those experiencing a Staged alert may take longer to react than they would to Imminent alerts, but this may only be if they are looking forward at the time of the alert. |
| **Time to Regain Control** | **Significant** | Not Significant | Those experiencing a Staged alert may take longer to regain control of the vehicle than after Imminent alerts. |
| **Time to Activate the Automation** | Not Significant | Not Significant | Different alert types may not affect how quickly operators activate the automation, nor does time to activate the automation seem to change over time. |
| **Time to Release Control of Steering** | **Significant** | Not Significant | Those experiencing a Staged alert may take less time to release control of the steering compared to when they experience an Imminent alert combined with an external threat, but not necessarily longer than an Imminent alert when an external threat is not present. |
| **Time to Resume Non-Driving Task** | Not Significant | Not Significant | Different alert types may not affect how quickly operators resume a non-driving task, nor does time to resume a non-driving task seem to change over time. |

## Eye-glance Behavior

### *Monitoring Rate*

Participants monitored the driving environment significantly more often immediately after the alert compared to immediately before the alert, and this was true for all three alert types, $p < .0001$. The mean percentage of time participants spent monitoring the roadway immediately after the alert was 87.61 percent (S.E. = 1.39 percent, n = 75, min = 53.22 percent, max = 100.00 percent), and the mean percentage before the alert was 37.55 percent (S.E. = 3.00 percent, n = 150, min = 0.00 percent, max = 100.00 percent; Figure 4-16). (For additional details on the statistical analysis, see Appendix D, Results, Eye-glance Behavior, Analysis for Monitoring Rate, Statistical Analysis.)

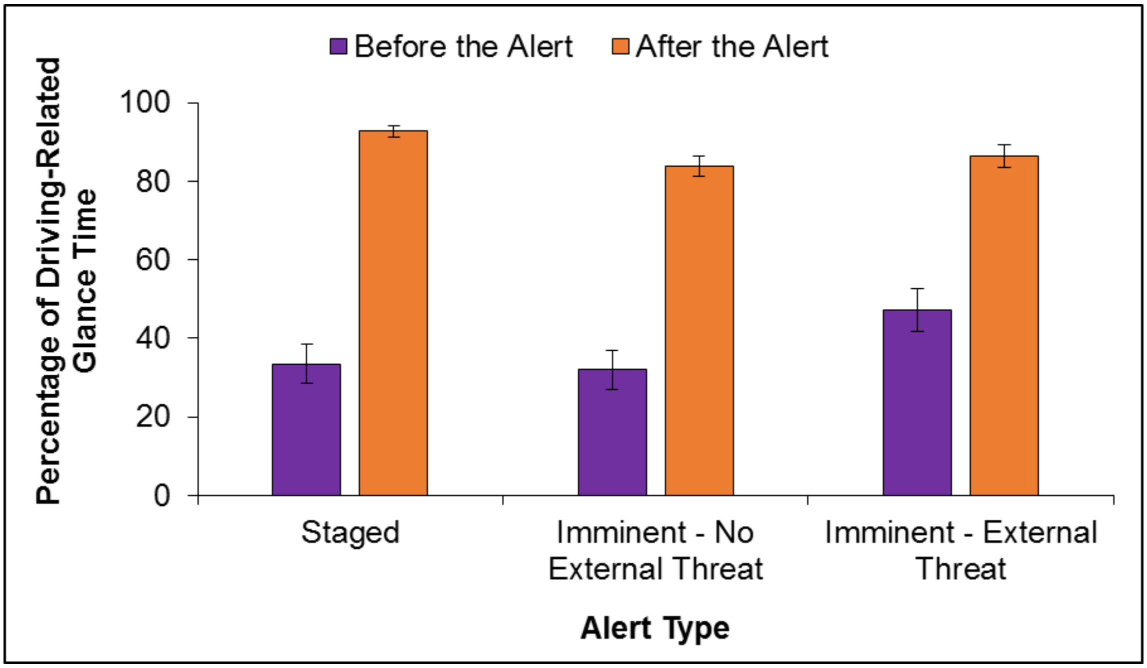



**Figure 4-16. Mean and Standard Error Bar Plot of Percentage of Driving-related Glance Time, Stratified by Alert Type and Time Interval Before/After the Alert for Experiment 3**

***Non-driving-related Glances***

The increase in the rate of non-driving-related glances per second from immediately before to immediately after the alert was significantly dependent on the type of alert $F(2,182) = 10.05$, $p < .0001$. Specifically, the rate of non-driving-related glances increased significantly immediately after an Imminent alert with no external threat, $t(204) = 3.10$, $p = .0069$, and decreased significantly after Staged alerts, $t(204) = -2.66$, $p = .0258$, but did not significantly change if the alert was Imminent and there was no external threat.

Participants' mean rate of non-driving-related glances per second before the Staged alert was 0.18 glances per second (S.E. = 0.02 glances/s, n = 50, min = 0.00 glances/s, max = 0.50 glances/s), and after the alert it was 0.14 glances/s (S.E. = 0.04 glances/s, n = 25, min = 0.00 glances/s, max = 0.92 glances/s). For Imminent alerts with an external threat, the means were 0.15 glances/s before the alert (S.E. = 0.02 glances/s, n = 50, min = 0.00 glances/s, max = 0.50 glances/s), and 0.31 glances/s after the alert (S.E. = 0.06, n = 24, min = 0.00 glances/s, max = 1.13 glances/s). If the Imminent alert did not have an external threat, the mean was 0.19 glances/s before the alert (S.E. = 0.02 glances/s, n = 50, min = 0.00 glances/s, max = 0.40 glances/s) and 0.42 glances/s after the alert (S.E. = 0.06, n = 25, min = 0.00, max = 0.97). (For additional details on the statistical analysis, see Appendix D, Results, Eye-glance Behavior, Analysis for Non-driving-related Glances, Statistical Analysis.)

## Trust Scales

Overall, participants reported strong levels of trust in the system. In 95.6 percent of instances, participants reported that they either "moderately agreed" or "strongly agreed" that they trusted the system. Further, trust also increased significantly over time, $z = 2.78$, $p = .0055$ (Figure 4-17). However, trust was not immediately affected by experiencing the alerts. (For additional details on the statistical analysis, see Appendix D, Results, Trust, Analysis for Trust over Time, Statistical Analysis.)

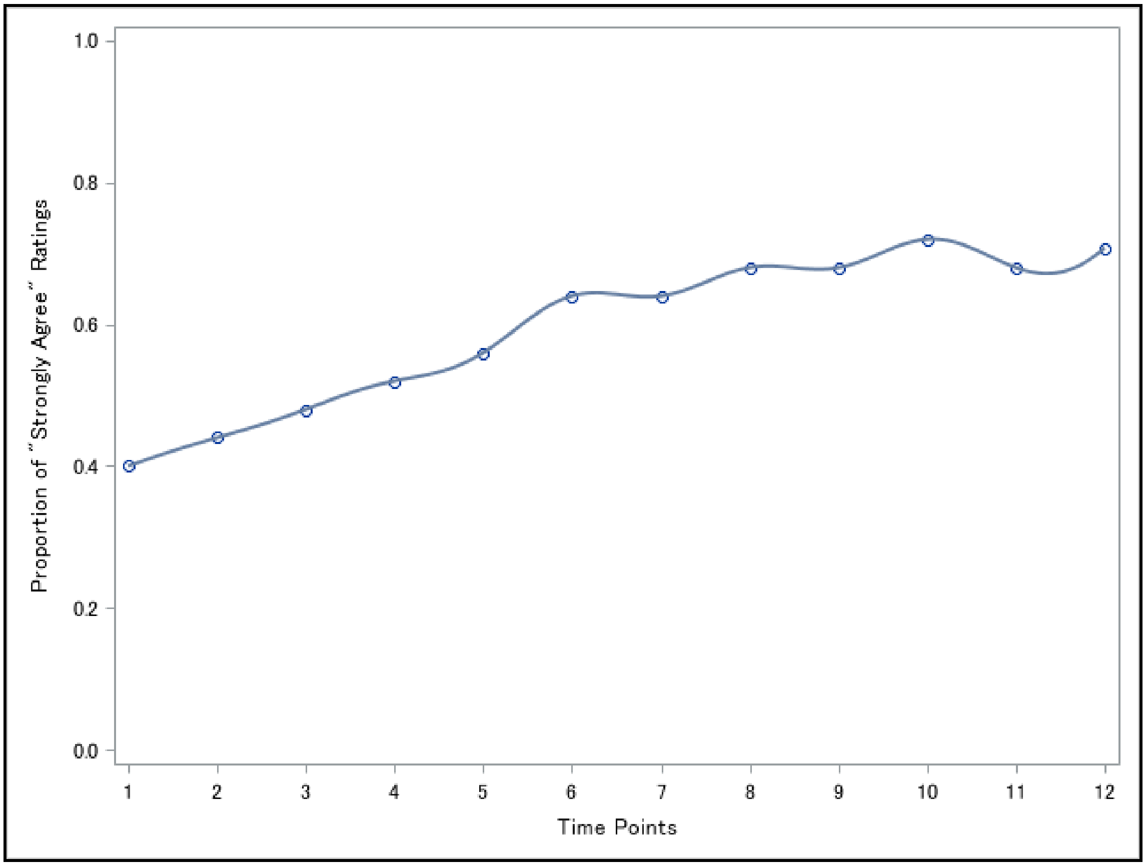



**Figure 4-17. Proportion of “7” Ratings by Time for Experiment 3**

The graph reflects an increasing time trend: as time increases, the proportion of “7” ratings increases as well, although the rate of increase becomes more sporadic after time point 6.

### *After-experience Trust Scales*

A survey of six statements relating to how much the participant trusted the system was given at the end of the experiment (i.e., not at the end of each session, only when all three sessions had concluded). The statements were:

TS1. I can rely on the automated system to function properly while I am doing something else.

TS2. The automated system provided the alerts when needed.

TS3. The automated system gave false alerts.

TS4. The automated system is dependable.

TS5. I am familiar with the automated system.

TS6. I trust the automated system.

All responses were based on a 7-point Likert-type scale with options ranging from “1” for strongly disagree to “7” for strongly agree. All but one of the statements had positive wording, so a higher score indicates a higher level of trust. Statement 3, “The automated system gave false alerts,” had negative wording, so responses to this statement were recoded so that “1” = “7,” “2” = “6,” etc. Mean and standard error bar plots for the responses are displayed in Figure 4-18 (with responses to statement 3 recoded).

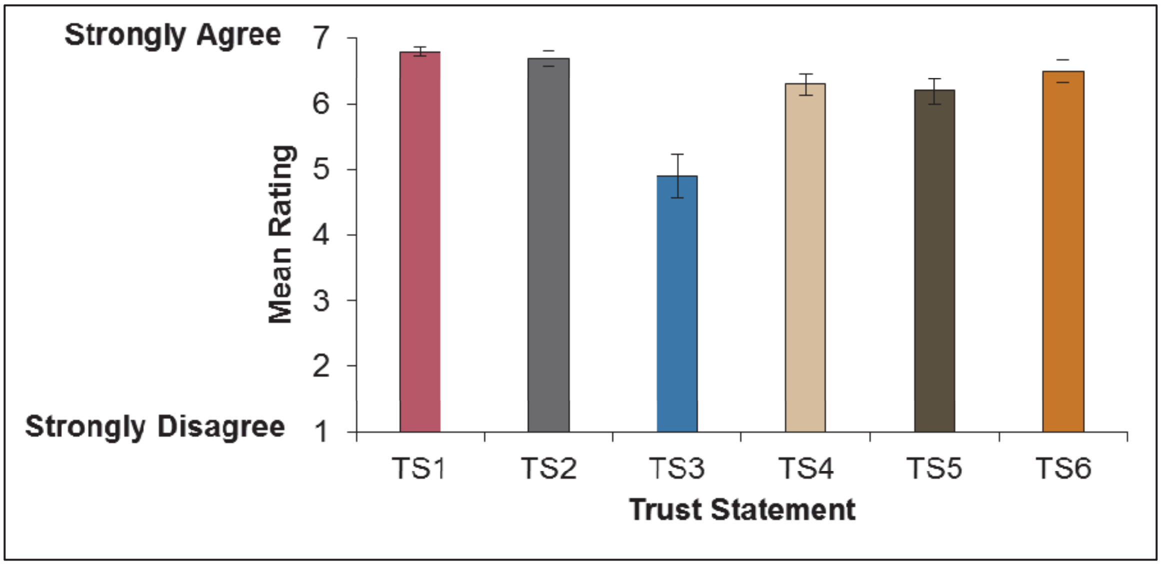



**Figure 4-18. Mean and Standard Error Bar Plot of Experiment 3 Mean After-experience Trust Ratings by Statement**

***After-experience Interview***

Following the in-vehicle experience, participants were asked a series of six questions. The interviews were audio recorded as a backup to researcher notes. Digital audio recordings were made into transcripts but no individual names were attached to any comments. Responses for 23 out of the 25 participants were recorded and transcribed. For the remaining two participants, the researchers’ written notes were consulted. The transcripts and notes served as the basis for a qualitative content analysis, which was completed using a framework analysis method. The results of this analysis were used to help researchers to potentially understand participant behavior that varies from the norm. (Some participants elected not to engage in a non-driving task, choosing instead to continue to monitor the roadway). A full discussion of the methods used to complete the framework analysis and detailed findings have been included as Appendix F.

Generally, participants indicated that they had a positive experience and seemed impressed by and confident in the system. The majority of participants reported that they were very comfortable with the system (Figure 4-19) and this comfort level was achieved in 15 minutes or less (Figure 4-20). Participants who indicated initial caution towards the automated system noted that by the end of the experimental test sessions they were comfortable using the system.

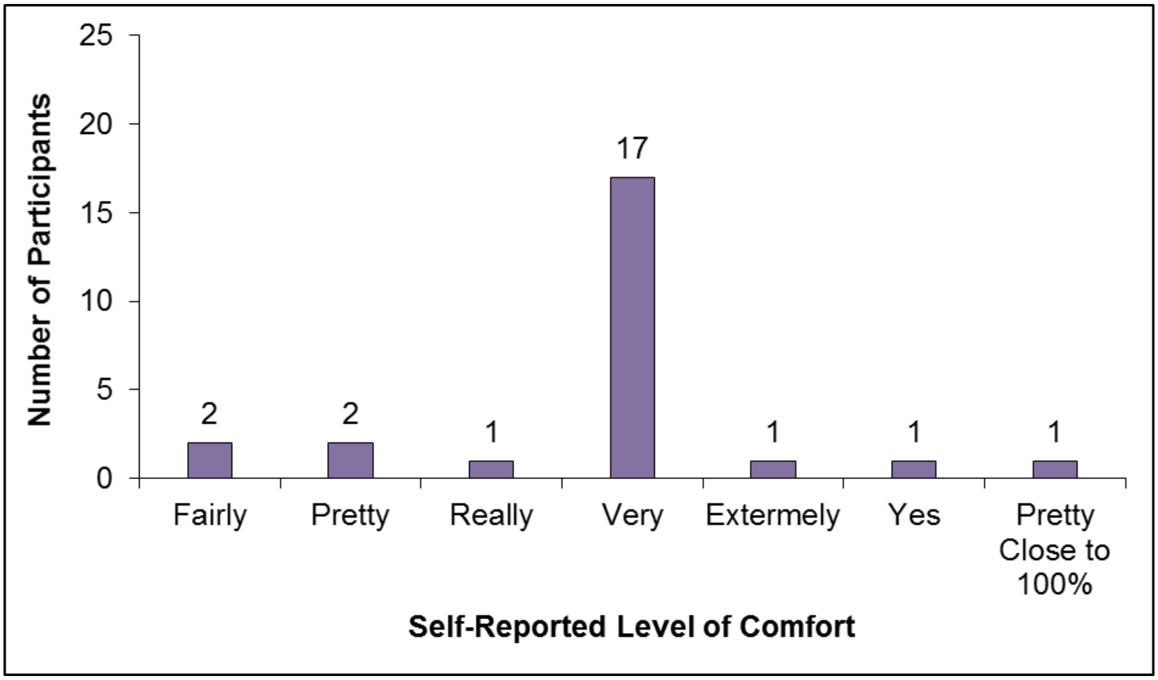



**Figure 4-19. Self-reported Levels of Comfort with the Automated System for Experiment 3**

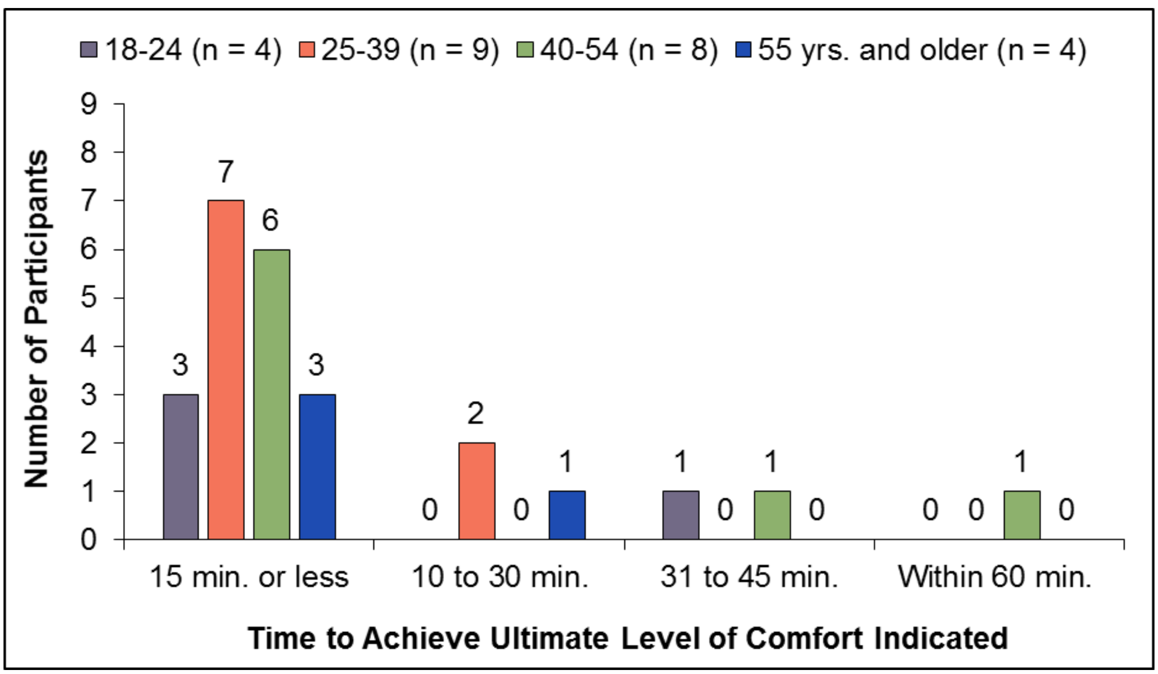



**Figure 4-20. Time to Achieve Comfort with the System by Age for Experiment 3**

# Experiment 3 Summary

This experiment examined the effects of deactivation cues (i.e., take control alerts) and the overall operator behavior over time. During three 30-minute driving sessions, participants had free exposure to a non-driving-related activity and were presented with a message stating that they must take control of the vehicle. Participants' reactions to these messages, both in duration and method of response, were some of the variables examined in this experiment.

Participants were able to differentiate between the types of alerts presented and acted accordingly. Furthermore, methods for activating the automated system and regaining control from the automated system were influenced by the type of alert and preconceived mental models (e.g., associating an activation button with one that is used to activate a cruise control system). Additionally, participants' level of comfort with and confidence in the system may have been improved with additional education regarding system capabilities. However, as this was a test situation, providing participants with additional education was not possible.

Staged alerts provide two important types of data. While the Imminent alerts had a distinct starting point, the Staged alerts encompassed four phases, which allowed for two interpretations, both of which provide salient insight into operator behavior. Two values were calculated: time to regain control after the Staged alert began, and time to regain control relative to the onset of the imminent phase of the Staged alert. The mean time to regain control was 17 s after the onset of the Staged alert, which was 23 s prior to the onset of the imminent phase. It should be noted that the majority of participants (56%, $n$ = 14) experiencing the Staged alerts regained control within the first phase, which was informational.

A summary depicting the mean values across the three alerts (Staged, Imminent–No External Threat, and Imminent–External Threat) for the dependent variables is provided below in Table 4-3. This table compares alert types for each dependent variable. If there is no significant difference between two particular alert types, they will share a common letter. For example, in time to release control, Staged and Imminent–External Threat alerts were significantly different, so they have different letters. However, neither were significantly different from Imminent–No External Threat alerts, so the Imminent–No External Threat alert has letters in common with both of the other alerts.

**Table 4-3. Mean Values for the Three Alerts for the Dependent Variables in Experiment 3**

| Dependent Variable | | Staged | Imminent–No External Threat | Imminent–External Threat |
|---|---|---|---|---|
| Time to React | | -38.8 s*<br>1.2 s**<br>A | 0.7 s<br>B | 0.7 s<br>B |
| Time to Regain Control | | -23 s*<br>17.0 s**<br>A | 2.3 s<br>B | 2.1 s<br>B |
| Time to Activate Automation* | | 4.1 s<br>A | 6.3 s<br>A | 8.0 s<br>A |
| Time to Release Control of Steering | | 1.4 s<br>A | 1.8 s<br>AB | 2.7 s<br>B |
| Time to Resume Non-Driving Task* | | 39.1 s<br>A | 13.9 s<br>A | 26.3 s<br>A |
| Monitoring Rate (Roadway) | Prior to Alert | 33.5%<br>A | 32.0%<br>A | 47.1%<br>A |
| | After Alert | 92.7%<br>A | 83.8%<br>A | 86.4%<br>A |
| Glances per Second (Non-Driving) | Prior to Alert | 0.2<br>A | 0.2<br>A | 0.2<br>A |
| | After Alert | 0.1<br>A | 0.4<br>B | 0.3<br>B |
| Trust Scales | | High Level of Trust (6.6 out of 7)<br>A | High Level of Trust (6.5 out of 7)<br>A | High Level of Trust (6.5 out of 7)<br>A |

* Time calculated relative to the point in time when the Imminent alert would have been presented.
** Time calculated from the point in time when the informational phase of the Staged alert began.

# Chapter 5 Discussion

The purpose of this study was to examine the interaction between operators and automated vehicle systems. This study focused on whether operators can safely transition between automated and non-automated vehicle operation, and how this interaction is affected by the HMI. Three experiments were performed to address six research questions. A summary of each experiment is provided in Table 5-1. The answers to the six research questions are discussed following the table.

Throughout this document, the vehicle user is referred to as the operator, not as the driver. Once the user activates the automation, he or she is not actively and uninterruptedly directing the course of the vehicle. For the Level 2 and Level 3 automation levels studied herein, the user can cede both the longitudinal and lateral control to the automation, sometimes for prolonged periods of time. However, vehicles at Level 2 and Level 3 could potentially look exactly the same, with the exception of the interface. Thus, designing systems with human factors principles in mind (i.e., user capabilities and limitations) can become a differentiator, including making the system's capabilities and limitations transparent to the operator. How the study results impact interface design are discussed in this section.

**Table 5-1. Main Findings**

<table>
<tr><th colspan="7">Human Factors Evaluation of Level 2 and Level 3 Automated Driving Concepts</th></tr>
<tr><th>Dependent Variable</th><th colspan="2">Experiment 1</th><th colspan="2">Experiment 2</th><th colspan="2">Experiment 3</th></tr>
<tr><td rowspan="2">Time to React to Alert</td><td colspan="2">Operators' mean time to react to multimodal alerts (0.7 s) was faster than their mean time to react to unimodal alerts (4.3 s)</td><td>Prompt Condition: 2-second, 7-second, and No Prompts</td><td>No significant difference in reaction time to events</td><td>Staged</td><td>Operators are likely to react during the information phase</td></tr>
<tr><td>Alert Types: Cautionary, Imminent, Staged</td><td>Alert type did not impact reaction time</td><td>Event Type: Alert and No Alert</td><td>Reaction time is much higher when there is no alert</td><td>Imminent</td><td>No difference in time to react between the two Imminent alerts</td></tr>
<tr><td rowspan="2">Time to Regain Control</td><td>Alert Modality: Unimodal, Multimodal</td><td>Time to regain control was faster for multimodal alerts</td><td colspan="2" rowspan="2">Operators regain control faster and more consistently with an alert. In 23 out of 49 cases without an alert (46.9 percent), participants did not regain control without instruction, while 53 out of 54 participants (98.1 percent) regained control with an alert.</td><td>Staged</td><td>Many operators are likely to regain control during information phase</td></tr>
<tr><td>Alert Types: Cautionary, Imminent, and Staged</td><td>Time to regain control was faster during the Imminent alerts</td><td>Imminent</td><td>No difference in time to regain control between the two Imminent alerts</td></tr>
<tr><td>Performance</td><td colspan="2">Operators were more likely to attempt to regain manual control during multimodal alerts</td><td colspan="2">Having an alert results in operators performing appropriately on a more consistent basis. As noted above, almost half the participants failed to regain control without an instruction when there was no alert, but almost all regained control without instruction when there was an alert.</td><td colspan="2">All operators attempted to regain control of the vehicle when prompted</td></tr>
<tr><td rowspan="2">Method Used to Regain Control</td><td colspan="4" rowspan="2">Always steering wheel (when applicable)</td><td>Staged</td><td>Off button on the steering wheel was the preferred method</td></tr>
<tr><td>Imminent</td><td>Brake was the preferred method for both Imminent alerts</td></tr>
<tr><td>Time to Activate Automation</td><td colspan="2">Unimodal Imminent alerts had a longer response time than multimodal Imminent alerts</td><td colspan="2">After a surprise lane drift, it took operators longer to activate automation when they received an alert compared to when they did not</td><td colspan="2">Time to activate automation is likely not impacted by alert type; also, this does not change over time</td></tr>
</table>

| | | | |
|---|---|---|---|
| Time to Resume Non-Driving Task | Operators took longer to resume non-driving tasks after Imminent alerts | No effect of prompt condition, event time, or session number on the time to resume a non-driving task | The time it takes operators to resume a non-driving task after experiencing a failure is not likely impacted by alert type; also, this does not change over time |
| Monitoring Rate | Operators tend to redirect attention to the forward roadway faster when they receive multimodal alerts | When operators are engaged in a non-driving task at the start of the measurement, monitoring rate increases with the frequency of prompts; most notably, the monitoring rate is higher when operators receive prompts versus no prompts. | Operators tend to spend a higher percentage of time monitoring the roadway after receiving an alert |
| Trust Ratings | Trust was generally high and did not change over time | Trust was generally high, with the No Prompts condition having a higher mean trust rating than the prompt conditions. Trust did not change over time. | Trust was generally high and this increased over time |
| After-experience Interview | Positive experience | Positive experience | Positive experience |
| Satisfaction Survey | N/A | Generally Satisfied | N/A |

# Main Findings

## Research Question 1: How Do Drivers Interact with and Operate Vehicles that Offer Level 2 and Level 3 Automation?

RQ1 specifically asked, “How do drivers interact with and operate vehicles that offer Level 2 and Level 3 automation; e.g., what is the driver performance profile over length of time in continuous or sustained automation?” The safe operation of the partially automated vehicles was assessed by considering the time operators took to regain control when presented with an unexpected event, whether operators missed or showed confusion to the TOR alerts, and whether operators continued to monitor the road when it was required. Under the most effective alerting approaches tested, operators regained control of the partially automated vehicles in a reasonable amount of time when alerted to an unexpected event (i.e., unexpected lane drift when operating a vehicle equipped with an L2 system, and an unexpected box in the road when operating a vehicle equipped with an L3 system). Operators took a mean of 1.3 s (S.E. = 0.08 s) to regain control of the vehicle equipped with an L2 system when presented with an Imminent visual and haptic alert. Operators took a mean of 2.1 s (S.D. = 0.7 s) to regain control of the vehicle equipped with an L3 system when presented with an Imminent visual plus auditory alert, which was accompanied by automation-supplied braking. It should be noted that the severity of the unexpected events was relatively low at the time the Imminent alerts were presented; the lane drift had yet to occur at the time of the alert, and the TOR was issued at a 10-s TTC to the box. However, there were alerts that were missed by operators. This highlights the importance of the HMI in supporting safe human-automation interaction. The data show that operators can have a delayed response to, or completely miss, strictly visual cautionary alerts. However, combining visual alerts with nonvisual modalities (e.g., haptic, auditory) is not a simple solution, as doing so increases the likelihood for operator annoyance.

Operator performance with the automation was observed to change over time. Participants took longer to respond to the prompts as their time in the study progressed. Their mean time to respond in the first session was 2.4 s (S.E. = 0.2 s), and increased to 3.0 s in Sessions 2 and 3 (S.E. = 0.3 s in the second session, S.E. = 0.2 s in the third session). This may be because the prompts were not generated in response to an automation performance issue that required a corrective action. Given that vehicle automation is not expected to exhibit frequent performance issues, it is possible that prompt effectiveness could degrade over time.

The exposure to the different prototype systems varied from 90 minutes to 3 hours. All three studies represent a reasonable exposure to the automation and much can be learned from studying the participants’ behavior. At the same time, it is important to consider that this study measured a one-time exposure of novice users (i.e., operators were not previously exposed to these levels of vehicle automation). As such, various unintended consequences that may occur with increased exposure to the automation, such as automation abuse, may not have been displayed. This is particularly true given the presence of an in-vehicle experimenter.

Another key factor that determines how operators interact with the automation is their trust in the automation. If the trust in automation is low, operators will likely not cede control to the system; or, if they decide to do so, they might monitor its performance closely (Lee & See, 2004; Stanton & Young 1998). If the trust is high, operators may feel more willing to cede control of the vehicle and redirect their attention to other tasks of interest (Hultgren et al., 2014). It was found that operators trusted the automation before, during, and after participating in the study. The after-experience interviews suggest that there were some interesting differences between what participants thought in the beginning of the study (before experience) and how they were able to comfortably cede control shortly after they experienced the system's capabilities. When asked how quickly (specific or relative time) they achieved their level of comfort, the majority of participants reported that their ultimate level of comfort was achieved in 15 minutes or less. However, participants that experienced a lane keeping performance issue with an L2 automated vehicle without an alert in Experiment 2 were found to lose some trust in the automation. This suggests that these operators may have somewhat calibrated their trust to the capabilities of the automation and became less likely to over-rely on the automation. However, whether their lowered trust is sustained over time should be investigated further.

It is worth discussing that in Experiment 2 no instructions were provided on how the system worked, or how the different features were intended to perform. The results could thus generalize to a "car rental scenario," where the only information given to the operator is what is shown at "key on" by the system (potentially no other information is provided, and there is no time to read the owner's manual or receive detailed dealership demonstrations). The results suggest that the HMI plays an important role in guiding how operators interact with the automation.

## Research Question #2: What Are the System Performance Risks from Operator Involvement in Secondary Tasks?

The second research question asked, "What are the system performance risks from operator involvement with, and interruption from, secondary tasks (such as portable electronic device use) that could arise when operating Level 2 or Level 3 automated vehicle systems?" All experiments performed as part of this effort requested and allowed participants to perform non-driving-related tasks while the vehicle automation was active. The automated vehicle prototypes tested did not prevent users from performing non-driving-related tasks and these tasks did not impact the automation from continuing to control the vehicle.

The main system performance risk found was that there were participants that ignored, or missed, alerts when engaged in a visually intensive non-driving task. For example, some operators of the vehicle equipped with an L2 system ignored the first stage of the attention prompts (a 5-second visual alert) and some operators of the vehicle equipped with an L3 system did not regain control of the vehicle until the second phase of the take-over request was issued (after 20 s of the first phase had expired). The data indicate that some operators will exhibit a complete reversal in priority from driving-related tasks to non-driving-related tasks. It may be that this primary task reversal phenomenon is the greatest risk facing operator involvement in secondary tasks when operating an automated vehicle. When operators shift their priorities to non-driving tasks, their readiness to respond to driving-related prompts and alerts can be delayed by a perceived obligation to complete the non-driving task first.

## Research Question #3: What Are the Most Effective Hand-Off Strategies between the System and the Operator?

The third research question asked, "What are the most effective hand-off strategies between the system and the driver, including response to faults/failures?" The most effective hand-off strategies in this study were those that involved nonvisual components. With the L2 automated vehicles in Experiments 1 and 2, operators were the fastest at regaining control of the vehicle when they received a combined visual and haptic seat alert. In Experiment 2, some operators missed the visual prompts and only responded when the prompts included a haptic component. Including a nonvisual component in the hand-off strategy may lead to the best outcome given the potential for operators to exhibit primary task reversals when performing non-driving tasks.

Interestingly, Experiment 3 with the L3 automated vehicle showed that operators will substantially delay how long they take to regain control of the vehicle when presented with an informational alert that combines visual and auditory elements. This delay, however, is likely a consequence of the HMI notifying participants to "prepare" to take control (rather than stating to take control now) as well as showing them just how much time they have to react (a countdown timer was presented in this case). As such, if the HMI conveys to operators that they have time to prepare to regain control, many will take that time to finish their non-driving activities. For instance, perhaps operators trust the timer and allow the automation to control the vehicle as long as possible before taking manual control.

## Research Question #4: How Do Operators Engage, Disengage, and Reengage with the Driving Task in Response to the Various States of Level 2 and Level 3 Automation?

The fourth research question asked, "How do operators engage, disengage, and reengage with the driving task in response to the various states of Level 2 and Level 3 automation?" Operators' engagement in the driving task when the automation issued a TOR was characterized as the time taken for them to react and regain control, as well as the sequence of actions they took to regain control. The time taken to re-engage the automation and release control was also assessed. Table 5-2 summarizes operator performance over this sequence of events. Operators reacted (i.e., looked forward) to the TORs within 1.2 s, depending on the context. Participants preferred to grab the steering wheel first to regain control of the L2 automated vehicle. This took 1 to 2.4 s, on average. Participants preferred to press the off button on the steering wheel to regain control of the L3 automated vehicle if they received an informational alert. However, participants took most of the allotted time to perform this action, with a mean time of 17 s. In contrast, participants operating an L3 automated vehicle preferred to immediately press the brake pedal if they received an Imminent alert. This took less than 2.3 s. One participant was not able to successfully press the brake pedal hard enough to regain control. However, other studies have shown that, during emergency events, not all drivers exert the appropriate force on the brake pedal to stop the vehicle (e.g., Fitch et al., 2009).

With regard to re-engaging the automation after a brief moment of manual control, participants took less than 4.6 s, on average, to release control once the automation became available. This time fluctuated somewhat and may have depended on whether operators believed the automation could go into a more severe warning state (as in the case of Staged alerts) once they experienced the Staged alerts. (Unimodal alerts were not included in Table 5-2 given they were not as effective as their multimodal counterpart.)

**Table 5-2. Mean Times for Responses to Different Alerts and Prompts (units: seconds)**

| Dependent Variable | Experiment 1 – Level 2 | | Experiment 2 – Level 2 | | | | Experiment 3 – Level 3 | | |
|---|---|---|---|---|---|---|---|---|---|
| | Imminent Multimodal | Staged Multimodal | Lane Drift with Alert | Lane Drift without Alert | 2-sec Prompt [b] | 7-sec Prompt [b] | Staged | Imminent –No External Threat | Imminent –External Threat |
| Time to React | 0.7 | 0.6 | 1.0 | 3.6 | 2.7 | 2.9 | -38.8 [c]<br>1.2 [d] | 0.7 | 0.7 |
| Time to Regain Control | 1.3 | 1.4 | 2.4 | 5.7 [a] | 12.1 [b] (n=6) | 32.3 [b] (n=6) | -23.0 [c]<br>17.0 [d] | 2.3 | 2.1 |
| Method to Regain Control | Steering Wheel | Steering Wheel | Steering Wheel | Steering Wheel | Steering Wheel [b] | Steering Wheel [b] | Off-Button (Steering Wheel) | Brake | Brake |
| Time to Release Control | 2.4 | 4.6 | 3.3 | 4.2 | 2.9 [b] (n=6) | 2.7 [b] (n=6) | 1.4 | 1.8 | 2.7 |

[a] Not all operators were able to regain control; this value only includes those that were able to do so for the event.
[b] Not all operators had to regain control; only the last phase of the prompt required the operator to regain control.
[c] Time calculated relative to the point in time when the Imminent alert would have been presented.
[d] Time calculated from the point in time when the informational phase of the Staged alert began.



It should be noted that the unexpected lane drift is not seen as a failure of the system, but something that is going to be part of earlier levels of automation since some of the technologies might not be fully matured (e.g., low sun angle impacting lane detection mechanisms impacted by saturation of the detector in low sun angle situations). This happened during the present study (low sun angle conditions impacted the lane centering prototype system) and could feasibly occur in typical driving settings; therefore, operators need to expect this and other noise in the system performance that will require higher situational awareness. System alerts requesting the operator to take control should use resources available from the operator that are not claimed by other tasks (e.g., haptic or auditory) given that most of the inattention characteristics impact the visual resources.

It should also be noted that Level 3 systems are expected to provide the operator with ample time to react. The L3 system tested provided Staged alerts to the operators starting with informational alerts that progress over time into cautionary and imminent phases as conditions evolve. The prototype system tested provided information to the user up to 50 s in advance. This was mainly for the informational phase. More urgent alerts were tested under Imminent alert conditions. It is important to highlight that Staged alerts with an informational component were able to allow operators to react to the alert, on average, about 39 s before it progressed to the imminent alert phase.

Typical human factors research, as it relates to surface transportation, looks into evaluating safety constructs as the potential for preventing fatalities or injuries (e.g., calculating prevention and exposure ratios). This study could provide a completely different perspective. It could inform designers and practitioners on how long it will take for operators to react and regain control in order to effectively design fail-safe and fail-operational states that will account for operators' most likely reactions. Due to more advanced technologies and higher levels of automation, it is very possible that some of those alerts might be due to conditions internal to the vehicle (e.g., deteriorated GPS signal, lower confidence captured from cameras due to faded pavement markings) and not related to an evolving event external to the vehicle. Therefore, understanding situations such as the ones presented in the Imminent–No External Threat is very important.

## Research Question #5: How Do Operators Perform under Various Operational Concepts within Level 2 and Level 3 Automation?

The fifth research question asked, “How do drivers perform under various operational concepts within Level 2 and Level 3 automation, such as systems intended for everyday driving on open roadways in mixed traffic or systems intended for dedicated roadway-vehicle applications (e.g., automated lanes, remote highways)?” This study conducted three experiments that were performed at test-track facilities simulating mixed-traffic highway conditions. Table 5-3 summarizes the operational conditions that were covered. An expanded list of test parameters is presented in Marinik et al. (2014). The list describes factors that are common to the testing of automation concepts: highway, mixed traffic, infrastructure as-is, only cellular and GPS connectivity, no vehicle coordination (individual vehicle operation), awareness of vehicles, motorcycles, and pedestrians, dry and wet weather/road conditions, novice and experienced drivers, and engage/disengage by the driver or system.

Under the selected conditions, operators interacted with the automation and performed reasonably well when an effective HMI was displayed. As mentioned earlier, however, operator performance can substantially degrade when the HMI elements fail to consider the potential state(s) of the operator; in this case, a state of high visual distraction. At the same time, it is interesting that participants quickly released control to the automated vehicle when traveling in proximity to other vehicles on the road. This further highlights the trust in the partial automation that was sustained by participants over their time in the study. The results suggest that this trust was adequately calibrated with the capabilities of the automation.

**Table 5-3. Selected Automation Concepts for Human-Machine Interface Evaluation (Marinik, Bishop, Fitchett, Morgan, Trimble, and Blanco, 2014) and Corresponding Representation Within the Experimental Design**

| Parameter | High-ranking Modality | Experiment 1 | Experiment 2 | Experiment 3 |
|---|---|---|---|---|
| **Road Facility Type** | Highway | Limited-access test track | Limited-access test track | Closed track |
| **Automated Vehicle Segregation** | None (mixed traffic) | None | None | None |
| **Infrastructure Adaptation** | Infrastructure as-is | None | None | None |
| **Connected Automated Operation** | Cellular/GPS only | N/A | N/A | GPS only |
| **Inter-Vehicle Coordination** | No coordination (individual vehicle; industry) | No coordination | No coordination | No coordination |

| | | | | |
|---|---|---|---|---|
| **Situational Awareness** | Vehicles, motorcycles, pedestrians | Vehicles | Vehicles | Vehicles |
| **Weather Conditions** | Dry, clear, rain | Dry | Dry | Dry |
| **Roadway Surface Conditions** | Dry, wet | Dry | Dry | Dry |
| **Driver Ability in Manual Driving** | Novice, experienced | Experienced | Experienced | Experienced |
| **Driver Monitoring** | Both monitored and unmonitored | Unmonitored | Monitored | Unmonitored |
| **Engage/Disengage Method** | Initiated by driver/system/both | Both | Both | Both |

It should be noted that there are marked differences between Level 2 and Level 3 automation. Transitioning from Level 2 to Level 3 automation shifts the paradigm from the operator needing to be ready to take complete control of the vehicle to the operator assuming a supervisory or increasingly passive role; this presents multiple challenges to designers and practitioners. Providing information in a timely manner to the operator in order to encourage appropriate vehicle control when needed supports the operator's safety and user acceptance. Moreover, the definitions of these two systems (as defined by NHTSA) rely on two completely different premises. For Level 2, the "[t]he driver is still responsible for monitoring the roadway and safe operation and is expected to be available for control at all times and on short notice." While for Level 3 "… the vehicle is designed so that the driver is not expected to constantly monitor the roadway while driving." These two drastically different characterizations provide completely different models of performance and requirements for how long the system needs to be able to sustain automation in order for the operator to successfully control the vehicle. Future research should continue to address this issue.

## Research Question #6: What Are the Most Effective Human-Machine Interface Concepts?

The final research question asked, "What are the most effective human-machine interface concepts, guided by human factors best practices, which optimize the safe operation of Level 2 and Level 3 systems?" This study showed that the HMI concepts that were the most effective were those that involved nonvisual alerts in addition to the visual alerts. When engaged in a non-driving task, some operators exhibited a primary task reversal and chose to prioritize the completion of the non-driving task over the operation of the partially automated vehicle. The HMI concepts that included auditory and haptic components were more successful in grabbing the attention of these participants when they were engaged in a visually demanding task and much more effective at eliciting a safe response compared to the HMI concepts that only relied on a visual alert. As mentioned earlier, it is important to note that there are many different types of visual, auditory, and haptic alerts that could possibly yield different results from this study.

Interestingly, the strictly visual alert concepts were considered in order to decrease the likelihood of annoying the vehicle operators. There is a concern that an HMI that continually generates auditory or haptic alerts will be unacceptably annoying to operators who could, in turn, choose not to purchase vehicles with the safety benefits of automated technologies. It is therefore important to carefully balance an alerting approach's conspicuity and annoyance. Future research would help to better understand how to optimize the HMI for partially automated vehicles.

A list of design implications extracted from this study is presented below.

***Multimodal Alerts***

- Operators took significantly less time to react to the alert after experiencing a multimodal alert than they did after experiencing a visual-only alert.
- Multimodal alerts resulted in significantly less time to regain control than did unimodal alerts.
- Most operators had correct performance during multimodal alerts.
- Based on the subjective assessment, if there is a possibility of multiple alerts in a short period of time, a haptic alert could be a suggested approach to decrease annoyance.
- For less frequent events where detailed instructions are needed, an auditory component (verbal instructions) could supplement the informational, visual portion of the alert.
- The visual component of the alerts should be considered informational for the following reasons:
  - Operators reacted faster to visual warnings when they accompanied by auditory and/or haptic components.
  - The operator may engage in non-driving-related tasks that might block the visual cues or delay them.

***Prompt Effectiveness over Time***

- The fact that Level 2 automation could encounter situations outside its performance capability, and require the operator to take over without advance notice (per NHTSA definition), could present some opportunities for human factors to address operator performance. The basic countermeasure evaluated for these situations was to try to prompt the operator to keep some level of awareness of the surroundings. However, this type of prompting is not a trivial design implication. Topics such as annoyance, urgency, positive reinforcement, and consequences come to mind as topics that could be explored in a study of optimizing the design of an interface.

- The initial prototype interface for the prompts used for Experiment 2 did not include the third stage (i.e., a basic consequence in the form or an auditory "attention minder"). This last stage was included after pilot testing revealed the first signs of Primary Task Reversal. The interface had some level of urgency, and correct visual and haptic components, but no consequences. At that point of the prototype interface design there were no consequences if the operator decided to continue the non-driving-related tasks or even finish them before paying attention to the roadway again. The prompt was then redesigned to include an "attention minder." This potential annoyance consequence, similar in purpose to a seatbelt minder but with the auditory characteristics of an urgent alert, was selected. However, the results suggest that, over time, even the annoyance factor could be overcome by the urgency of finishing another task. This provides even more support to the Primary Task Reversal Theory and suggests that stronger consequences or a better method to emphasize a positive reinforcement strategy could be investigated as a key countermeasure for Level 2 automation.

***Braking as a Regain Control Mechanism***

- Only one participant failed to depress the brake pedal hard enough to disengage the automated system in this study. Although this may be a concern, the phenomenon has been documented in previously completed braking studies. Fitch et al. (2009) show that braking can present a problem for some operators when certain patterns or force are required. This is particularly true for situations that demand an emergency braking maneuver. Although past brake assist technology only activated once the driver applied the brake pedal, it is important to note that the L3 automated vehicle tested in this study bypassed operator response time and applied the brakes before the conflict escalated.

# Vehicle Automation Theories

## Primary Task Reversal

Primary Task Reversal describes a full-priority shift from the driving-related task (e.g., monitoring the environment) to non-driving tasks (e.g., answering e-mail). The operator becomes fully focused on non-driving tasks, making this the primary task and demoting controlling the vehicle (driving) to a secondary task. This is very different from what has been previously observed during “driver distraction” or inattention research. This is a full reversal of priorities, where even alerts will not “distract” the user from the “new primary task” until he or she feels that the task has been accomplished successfully. This could be longer than a few seconds; potentially, the operator could perform the new primary task for minutes without being distracted by the secondary task of driving. Humans are task-oriented individuals, and users’ mental models evolve based on affordances. If the automated vehicle provides the sense (affordance) that it is able to accomplish the primary driving tasks (including maintaining headway to other vehicles, staying in the lane, and detecting and responding to objects and events), the user’s priority could potentially become the non-driving tasks.

The non-driving tasks used for the study presented clear secondary task goals through explicit study instructions, as an attempt to address “worst case” (and improper, for L2) use of automation. The non-driving tasks presented during these experiments were representative of the tasks many people perform during daily routines (e.g., email, Web browsing, navigation) and that keep us connected to our work, family, and friends. The study participants were all smartphone users that did not experience motion sickness; therefore, performing these tasks remotely or in motion did not deter their goal. When provided with the automation affordance and clear instructed goals (e.g., send email about a meeting that must be rescheduled), the participants frequently shifted to making the non-driving tasks their priority. As mentioned previously, it was found that the visual alerts presented during the pilot study conducted prior to Experiment 2 did not represent an urgency level that would assist in redirecting the user’s attention to the forward roadway. This finding prompted a quick reaction from the team to develop an additional stage as part of the prompt to attempt to provide the necessary feedback and add consequences for a prolonged Primary Task Reversal. However, after the design improvements, the “attention minder” attempted did not have a lasting effect throughout the sessions. Further research could be conducted to investigate other HMI elements that could break into this very strong “clear instructed goal” secondary task presentation.

Unintended consequences were not observed at this point. However, it is clear that under the instructions provided in the studies, non-driving-related tasks seem to have a higher priority than what they used to have during manual driving. Driver inattention has been a well-studied phenomenon for many years (Dingus et al., 2011; Fitch et al., 2013; Hickman and Hanowski, 2012; Klauer et al., 2006; Klauer et al., 2014; Llaneras et al., 2013; Olson et al., 2009). Vehicle automation introduces a new twist on this concept—Primary Task Reversal—with the caveat of the type of instructions provided in the current studies. That is to say, as roles change from driver to operator, formerly secondary tasks (e.g., using electronic devices) may become the primary task of interest to the operator.

## Alert Annoyance Habituation

The results mentioned in the Primary Task Reversal suggest another potential theory: Alert Annoyance Habituation. The system prompts in Experiment 2 of this study were designed to notify operators to monitor the driving environment when the system detected that they were not doing so. They were also designed to escalate in urgency if the operator failed to respond to them by monitoring the driving environment. It must be noted that the prompts were generated in the absence of any performance issue of the automated system; they were only issued when the operator was detected to not be monitoring the driving environment.

Early in participation, participants responded to the low-level visual prompts. However, as the study progressed, some participants became habituated to them and only monitored the driving environment when the alerts escalated in urgency. This behavior was suspected during initial pilot testing; thus, a continuous auditory alert was added to the final stage of the prompts. Despite this effort, some participants began to ignore even the higher order prompts as the study progressed, and some even ignored the final stage of the prompt with the continuous auditory alert in order to complete a non-driving task. It is believed that this may have occurred because they habituated to the prompts, particularly because they were generated in the absence of any automated system performance issue that could have had immediate ramifications.

There are two analogies that help explain this behavior.

- First, drivers often ignore their check engine light or low fuel indicator when there is no apparent vehicle issue or they know they have several minutes to respond. These single-stage notifications also do not escalate in severity.
- It is also interesting to consider the similarities of this behavior to that of new parents. New parents typically respond to any cry from their baby. However, as they habituate to this communication, they learn the different types of cries, and differentiate as to how and when a reaction is needed.

Based on this evidence, we hypothesize that vehicle operators could become habituated to the attention prompts over time, and acclimate to an increasing level of acceptance for the prompt urgency. How to prevent this habituation and ensure that all operators respond to the alerts—without being annoying to those that comply—is a crucial human factors dilemma. Consideration for alert annoyance habituation is therefore imperative in future interface design efforts.

# Conclusion

Several key documents came out of this research effort: *Past Research, State of Automation Technology, and Emerging System Concepts*, DOT HS 812 043 (Trimble, Bishop, Morgan, and Blanco, 2014) and the *Concepts of Operation for Levels 2 and 3*, DOT HS 812 044 (Marinik, Bishop, Fitchett, Morgan, Trimble, and Blanco, 2014). These documents guided the development of the experimental designs used for the studies presented in this report. This research effort also developed a comprehensive list of operational definitions as they pertain to vehicle automation research. Moreover, it developed theories that could assist in optimizing the interface designs of automated vehicles.

Vehicle automation comes with the promise of many improvements, such as quality of life, comfort and, potentially, safety. The studies presented herein suggest a high level of trust in automation and alternatives to present information to the operators in these vehicles that could effectively assist them to react and regain control when needed. Figure 5-1 below depicts average reaction times to multimodal imminent alerts from Level 2 and Level 3 automation. Participants in both conditions reacted in a similar timeframe when they received an imminent multimodal alert. Figure 5-2 shows the rank ordered response times observed for both Level 2 and Level 3 vehicles in the multimodal imminent alert condition. Also depicted is the 90$^{th}$ percentile for reaction time (1.08 seconds). This information may provide designers and practitioners with methods for alerting drivers as well as the time that take control decisions might take for a majority of users (e.g., 90th percentile) if the priorities are potentially shifted in favor of other convenience tasks (e.g., email, Web browsing). This will help to ensure that the systems are designed with the needed redundancies to provide enough time for the operator to regain control of the vehicle when necessary. Additional research with Level 2 and Level 3 capable vehicles would provide a more complete understanding of the reaction time patterns in a larger variety of settings.

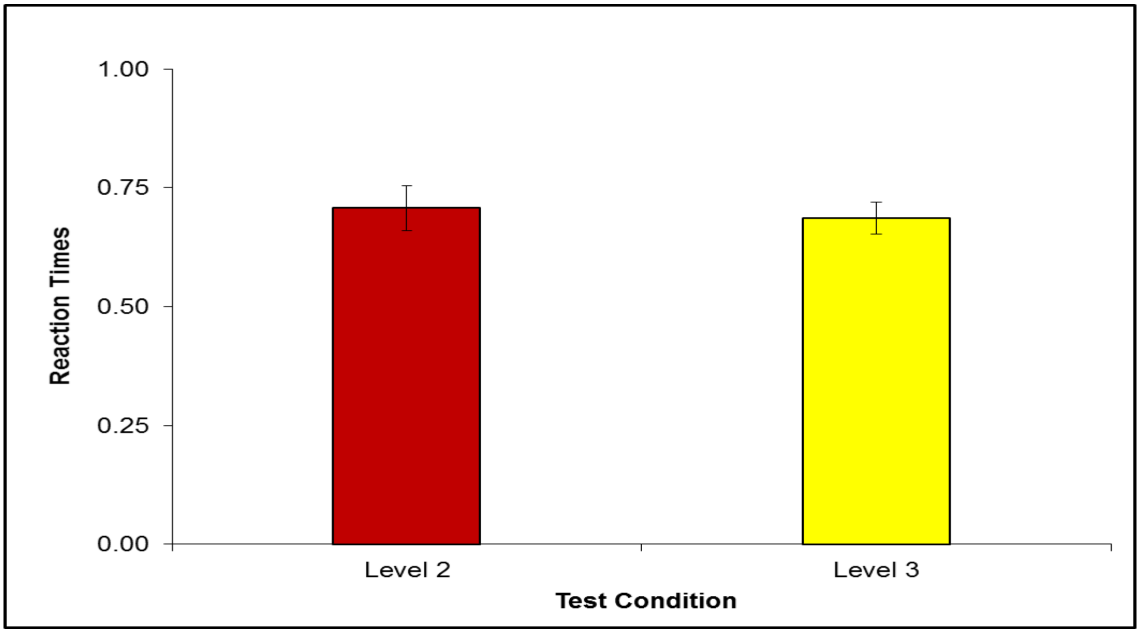



**Figure 5-1. Reaction Times for Multimodal Imminent Alerts for Level 2 & 3 Automation**

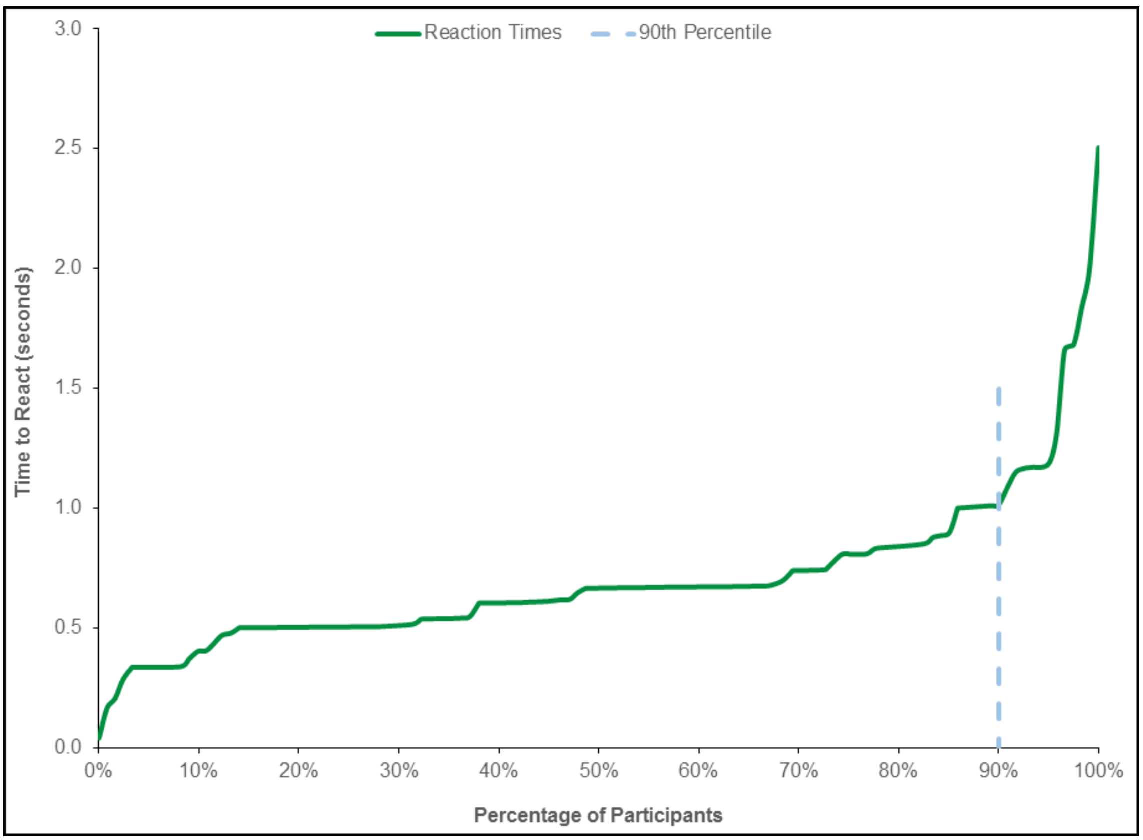



**Figure 5-2. Reaction Time Percentile for Multimodal Imminent Alerts for Level 2 & 3 Automation**

This set of studies represents an important research effort. It takes a closer look at multiple aspects that may impact the future of vehicle automation. Information presented herein represents the results of several comprehensive test-track studies with three different prototype vehicles (state-of-the-art). The methods that were implemented involve real vehicles in mixed-traffic conditions to ensure that the consequences of a potential crash were present at all times. This possibility served as an incentive for operators to behave as similarly as possible to how naïve drivers would behave when provided with an automated vehicle in real-world conditions. Future research should investigate operator interaction with automation under real-world driving conditions through a naturalistic driving study. This study should include a driving baseline condition in order to enable an investigation of driver performance without using the automated features of the vehicle. Such a baseline, for instance, could enable an analysis of drivers' ability to keep the vehicle in the lane and respond to hazards with and without support from the automation.

# References


78 Fed. Reg. 81 (April 26, 2013), pp. 24818-24890. Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices,

Dingus, T. A., Hanowski, R. J., & Klauer, S. G. (2011). Estimating crash risk. *Ergonomics in Design: The Quarterly of Human Factors Applications*, *19*(4), 8-12.

Fitch, G. M., Blanco, M., Morgan, J. F., Rice, J. C., Wharton, A. E., Wierwille, W. W., & Hanowski, R. J. (2009). *Human performance evaluation of light vehicle brake assist systems* (Report No. DOT HS 811 251). Washington, DC: National Highway Traffic Safety Administration.

Fitch, G. M., Soccolich, S. A., Guo, F., McClafferty, J., Fang, Y., Olson, R. L., & Dingus, T. A. (2013). *The impact of hand-held and hands-free cell phone use on driving performance and safety-critical event risk* (Report No. DOT HS 811 757). Washington, DC: National Highway Traffic Safety Administration.

Green, M. (2000). How long does it take to stop? Methodological analysis of driver perception-brake times. *Transportation Human Factors, 2*(3), 195-216. doi: 10.1207/STHF0203_1

Hickman, J. S., & Hanowski, R. J. (2012). An assessment of commercial motor vehicle driver distraction using naturalistic driving data. *Traffic Injury Prevention*, *13*(6), 566-574.

Hultgren, J. A., Kircher, K., & Larsson, A. (2014). Tactical driving behavior with different levels of automation. *IEEE Transactions on Intelligent Transportation Systems*, *15*(1), 158-167. doi: 10.1109/TITS.2013.2277725

Jian, J.-J., Bisantz, A. M., & Drudy, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. *Int. J. of Cognitive Ergonomics*, *4*(1), 53-71.

Kenward, M. G., & Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics, 53*, 983-997.

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

Klauer, S. G., Guo, F., Simons-Morton, B. G., Ouimet, M. C., Lee, S. E., & Dingus, T. A. (2014). Distracted driving and risk of road crashes among novice and experienced drivers. *New England Journal of Medicine*, *370*(1), 54-59.

Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). *Applied linear statistical models*. New York, NY: McGraw-Hill/Irwin.

Lee, J. D., & Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. *International Journal of Human-Computer Studies, 40*, 153-184.

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.

Llaneras, R. E., Salinger, J., & Green, C. A. (2013). Human factors issues associated with limited ability autonomous driving systems: Drivers' allocation of visual attention to the forward roadway. *Proceedings of the International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design, 7*, 92-98.

Luz, M. (2009). *Final report: Validation of a trust survey on example of MTCD in real time simulation with Irish controllers*. Eurocontrol. Retrieved from http://www.yumpu.com/en/document/view/19565510/025-mb-eurocontrol

Marinik, A., Bishop, R., Fitchett, V., Morgan, J. F., Trimble, T. E., & Blanco, M. (2014, July). *Human factors evaluation of level 2 and level 3 automated driving concepts: Concepts of operation.* (Report No. DOT HS 812 044). Washington, DC: National Highway Traffic Safety Administration.

National Centre for Social Research. (2012). *The framework approach to qualitative data analysis*. Retrieved from: http://www.surrey.ac.uk/sociology/research/researchcentres/caqdas/files/Session%201%20Introduction%20to%20Framework.pdf

National Highway Traffic Safety Administration. (2013). *Preliminary statement of policy concerning automated vehicles*. Retrieved from http://www.nhtsa.gov/staticfiles/rulemaking/pdf/Automated_Vehicles_Policy.pdf

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

Ritchie, J., Spencer, L., & O'Conner, W. (2003). Carrying out qualitative analysis. In J. Ritchie & J. Lewis (Eds.), *Qualitative research practice* (pp. 219-262). London: Sage.

SAE International. (2014). *Taxonomy and definitions for terms related to on-road automated motor vehicles*. J3016 (Approved 2014). Warrendale, PA: SAE International. Retrieved from http://standards.sae.org/j3016_201401/

Smithson, M., & Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods*, *11*(1), 54-71.

Stanton, N. A., & Young, M.S. (1998). Vehicle automation and driving performance. *Ergonomics*, *41*(7), 1014-1028.

Summala, H. (2000). Brake reaction times and driver behavior analysis. *Transportation Human Factors*, *2*(3), 217-226.

Trimble, T. E., Bishop, R., Morgan, J. F., & Blanco, M. (2014, July). *Human factors evaluation of level 2 and level 3 automated driving concepts: Past research, state of automation technology, and emerging system concepts.* (Report No. DOT HS 812 043). Washington, DC: National Highway Traffic Safety Administration.

# Appendix A. List of Acronyms

| | |
|---|---|
| ACC | Adaptive cruise control |
| ANOVA | Analysis of variance |
| CAN | Controller Area Network |
| CMH | Cochran-Mantel-Haenszel statistics |
| ConOps | Concepts of Operation document |
| COTS | Commercial-off-the-shelf |
| CWS | Crash warning system |
| DAS | Data acquisition system |
| Den DF | Denominator degrees of freedom |
| ET | External threat |
| GM | General Motors |
| GPS | Global Positioning System |
| HMI | Human-machine interface |
| IRB | Institutional Review Board |
| LCW | Lane change warning system |
| LED | Light-emitting diode |
| LKA | Lane keeping assist system |
| NET | No external threat |
| NHTSA | National Highway Traffic Safety Administration |
| Num DF | Numerator degrees of freedom |
| OEM | Original equipment manufacturer |
| OTS | Over the shoulder |
| Pr > F | *p* value |
| R.E.M.L. | Restricted maximum likelihood |
| s | Second(s) |
| S.D. | Standard deviation |
| S.E. | Standard error |
| TTC | Time-to-collision |
| V2V | Vehicle-to-vehicle |
| VTTI | Virginia Tech Transportation Institute |

# Appendix B. Experiment 1: Expanded Methods, Results, and Analysis

## Experimental Design

An a priori power analysis was conducted to determine the total number of participants required for this study in order to reach a power of 0.8 with $\alpha = 0.05$. The minimum sample size needed to detect a large difference in the dependent variables was calculated. Assuming a large effect size (0.8), 16 participants were required for the design; however, note that the experiment included 25 participants. In order to ensure that all age groups were represented, participants were recruited based on the four NHTSA age groups (18–24, 25–39, 40–54, and 55 and older; Visual-Manual NHTSA Driver Distraction Guidelines for In-vehicle Devices, 2013).

## Procedure

### Screening

The experimental protocols were reviewed and approved by the Virginia Tech Institutional Review Board (IRB). Participants were recruited from the greater Detroit, Michigan, area using newspaper and Craigslist advertisements, and through word of mouth. Participants were eligible for inclusion in the study if they met all of the following criteria:

- Valid driver's license, and were at least 18 years old;
- No disqualifying moving violations in the past 3 years;
- Had not previously participated in any GM study at the Milford Proving Ground;
- Owned a smartphone;
- Normal or corrected-to-normal vision and hearing;
- Able to drive without sunglasses or photochromatic lenses;
- Able to drive an automatic transmission vehicle without accommodation;
- If pregnant, discussed the risks of participating with a physician;
- Reported not taking medications or substances that interfere with driving ability;
- No lingering health effects or any recent heath events, including high propensity towards motion sickness;
- Eligible for employment in the United States; and
- Able to wear closed-toe shoes while driving.

## Greeting

Upon arriving at the test site, participants were greeted by the in-vehicle experimenter. The experimenter escorted the participant to the preparation area and provided the informed consent documentation. The participant's driver's license was examined to ensure validity. Following informed consent, the participant was asked to undergo a basic vision screening and complete the demographics questionnaire. Hearing was not formally assessed; however, the participant was asked to respond to basic questions in the greeting process that determined if normal hearing seemed to be present.

## Participants

Data were collected from 25 participants. The age groupings used in Figure B-1 reflect those proposed in the Visual-Manual NHTSA Driver Distraction Guidelines for In-vehicle Devices (2013).

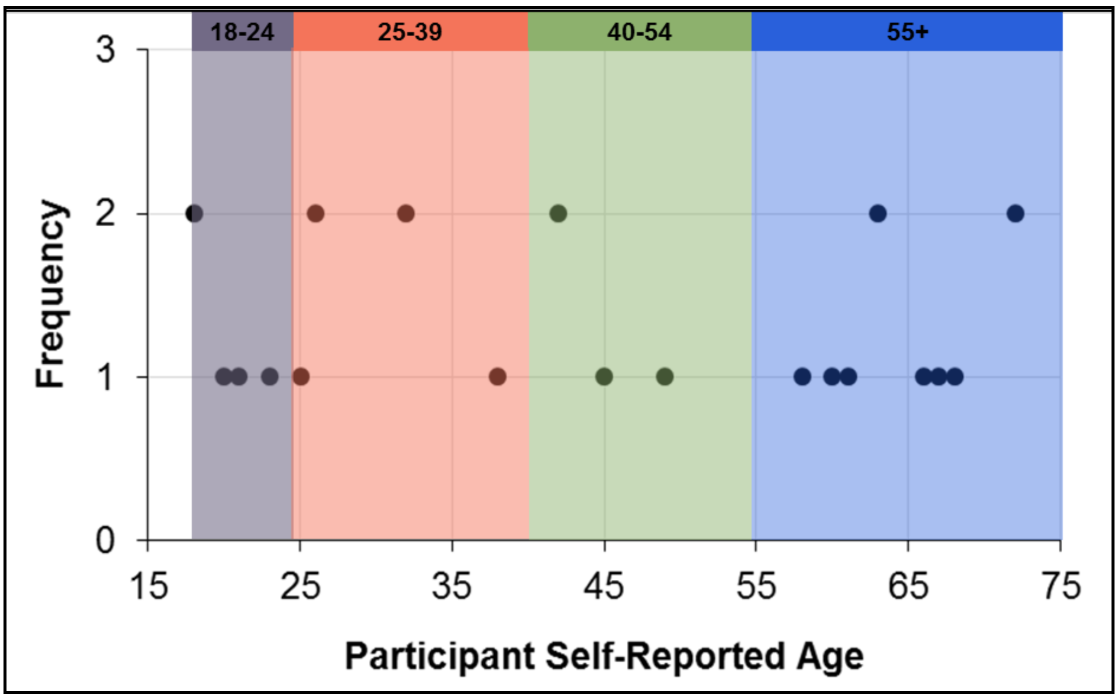



**Figure B-1. Distribution of Experiment 1 Participants Across the NHTSA Age Categories**

In addition to age, participants were asked to provide basic demographic information (e.g., gender, education level), driving experience (type of vehicles used and annualized miles driven), smartphone use and experience, and automotive technology experience. Participants were also asked about their awareness of automated vehicle technologies.

Owning a smartphone was required in order to participate in this study. This was to ensure that participants would be comfortable using the technology that would be utilized during the experiment. All participants indicated using their smartphone on a daily basis. Participants were asked to report how long they have used a smartphone: 36 percent (*n* = 9) indicated using a smartphone for less than 2 years; 24 percent (*n* = 6) indicated using a smartphone for 2 to 4 years; and 40 percent (*n* = 10) indicated using a smartphone for more than 4 years. These data are presented in Figure B-2 below.

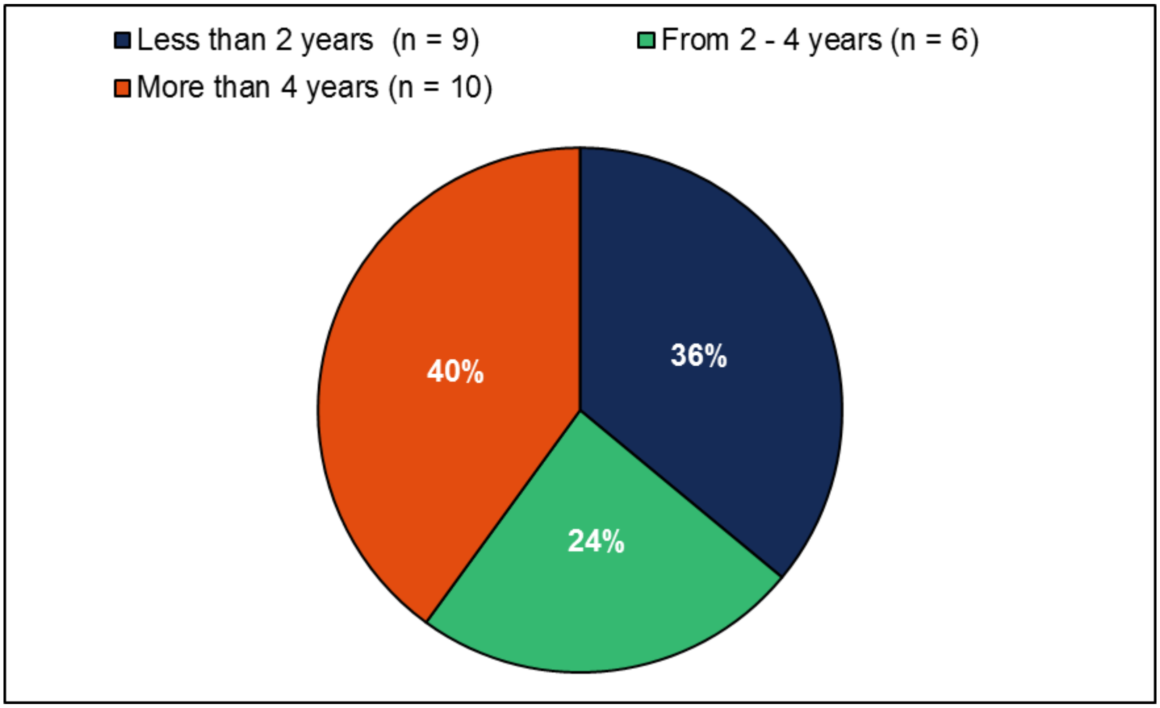



**Figure B-2. Experiment 1 Participants' Experience with Smartphones**

In addition, the participants' automotive technology experience and their familiarity with automated vehicle technologies were captured. Participants were asked to report any prior experience with Crash Warning Systems (CWS), ACC, and Lane Keeping Assist (LKA) systems (Figure B-3). Only 24 percent of participants (*n* = 6) reported having previously driven a vehicle equipped with CWS, and 8 percent of participants (*n* = 2) reported having previously driven a vehicle equipped with ACC. None of the participants reported having previously driven a vehicle equipped with an LKA system. Thirty-six percent of participants (*n* = 9) indicated being aware of self-driving/automated vehicles.

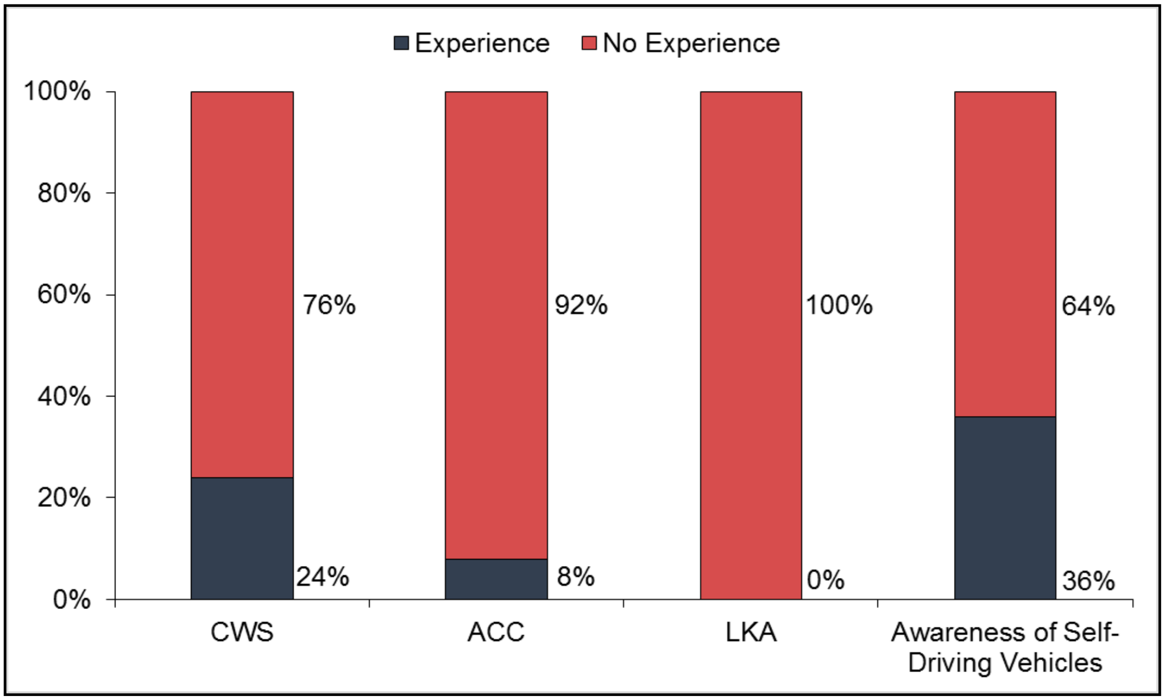



**Figure B-3. Experiment 1 Participants' Experience with and Awareness of Automated Technology for Vehicles**

# Results

It should be noted that in the analysis of this experiment, "unimodal" refers to an alert with a visual component, while "multimodal" refers to an alert with a visual component and a haptic component. Therefore, any significant differences between unimodal and multimodal can only interpreted as differences between a visual alert and a visual + haptic alert. Interpretations cannot be generalized to any unimodal alert versus any multimodal alert.

In modeling the operator behavior variables, the log transformation was applied for each variable to account for skewed data, and analysis was performed on the log-transformed times. For test statistics, denominator degrees of freedom were adjusted using the Kenward-Roger method (Kenward & Roger, 1997). Estimates for significant differences in the least squares means (adjusted for all variables) of alert type, alert modality, or their interaction were then back-transformed in order to make an interpretation in terms of the percentage change in the operator behavior variables from one group to another. To account for correlations in the data arising from multiple measurements of participants, a linear mixed-effect model was used with random intercepts for each participant, as well as random time slopes if the model allowed. Additionally, the variances of the times within different alert modality/alert type groups were allowed to differ in the model, if necessary.

## Operator Behavior

***Analysis for Time to React to Alert***

The time to react to the alert was explored for each alert modality and alert type. On average, the amount of time it took for a participant to react to the alert was 1.82 s (S.D. = 3.41 s). Interestingly, it took about 4.6 times longer to react to an alert when the alert was unimodal (M = 3.04 s, S.D. = 4.57 s) than for its multimodal counterpart (M = 0.66 s, S.D. = 0.42 s).

Though the differences between different alerts for time to react were not as great as the difference between the modalities, the participants did react more quickly after experiencing an Imminent alert than after a Staged alert or a Cautionary alert. The mean time to react for participants after experiencing an Imminent alert was 1.31 s (S.D. = 1.81 s), compared to Staged alerts (M = 1.89 s, S.D. = 3.40 s) and Cautionary alerts (M = 2.26 s, S.D. = 4.45 s). Out of 130 instances with a Staged alert, 120 (92.3 percent) reacted during the first phase of the alert, which represents the cautionary component of the alert (the first 10 s of the alert). Out of 139 Cautionary instances, 131 (94.2 percent) reacted prior to 10 s, 6 (4.3 percent) reacted after 10 s, and 2 (1.4 percent) did not react at all. Finally, during 136 of 139 Imminent alert instances (97.8 percent), the participant reacted within 10 s. All of these reaction times over 10 s occurred with only a visual component (unimodal alert). Mean and standard error bar plots for time to react stratified by alert are displayed in Figure B-4.

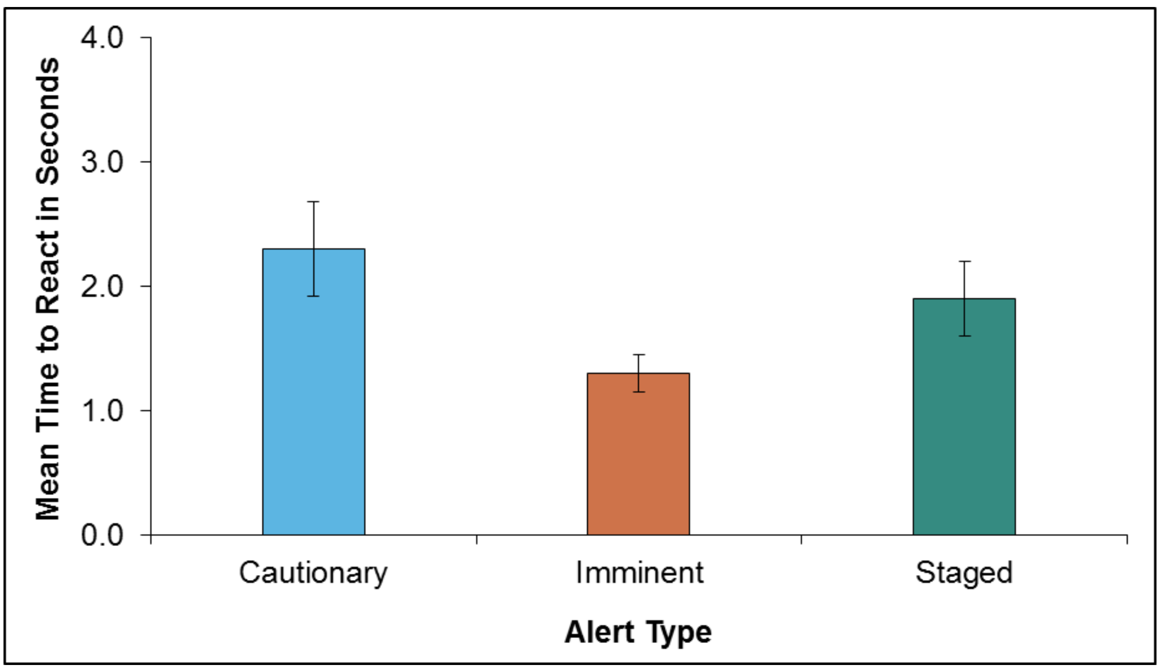



**Figure B-4. Mean and Standard Error Bar Plots of Time to React by Alert for Experiment 1**

Figure B-5 suggests that the distribution of time to react among the participants was highly skewed. All values that were identified as potential outliers were visually inspected for a second time to ensure that their validity classifications were accurate. Several outlying observations occur in all six combinations, with the highest observation at 31.0 s coming from a participant experiencing a unimodal/Cautionary alert about 58 minutes into the experiment. (The participant was engaged in a non-driving task at the onset of the visual alert and continued with the task for 31 seconds prior to reacting to the alert.) In fact, Figure B-5 indicates that there are higher skewed values within Cautionary alerts than for Imminent or Staged alerts, regardless of modality. This suggests that differences in means between the alerts observed in Figure B-4 may be inflated by these extreme values. However, in multimodal alerts, the mean time to react is faster than for unimodal alerts across all three alerts. The means within Cautionary alerts are 0.7 s for multimodal (S.D. = 0.6 s) compared to a mean of 3.7 s (S.D. = 5.8 s) for unimodal. For Imminent alerts, the mean time to react in multimodal alerts was 0.7 s (S.D. = 0.4 s), compared to 2.0 s (S.D. = 2.4 s) in unimodal alerts. Finally, within Staged alerts, the mean time to react was 0.60 s (S.D. = 4.57 s) in multimodal alerts compared to a 3.40-second mean (S.D. = 4.6 s) in unimodal alerts. The Staged and Cautionary alerts had similar results; this is due to the fact that in 120 out of 130 Staged alert instances (92.3 percent), participants responded to the cautionary portion of the Staged alert.

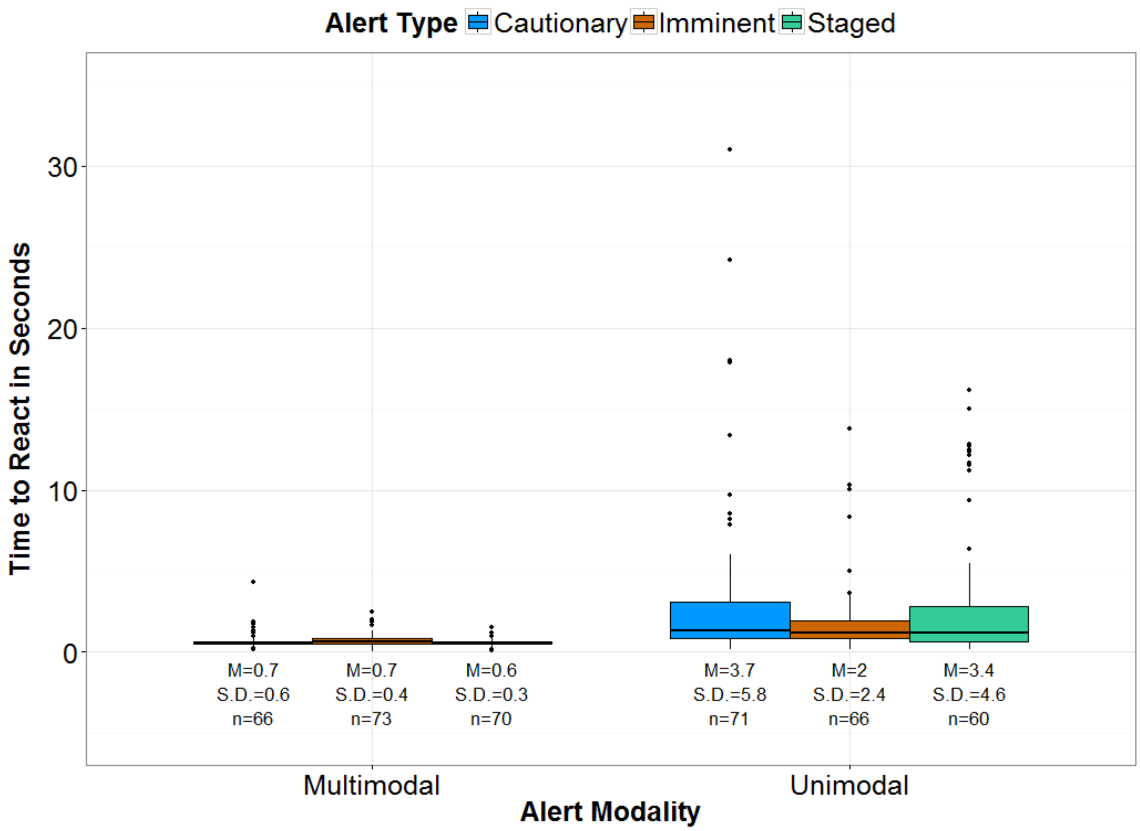



**Figure B-5. Boxplots of Time to React by Modality, Alert for Experiment 1**

One factor of interest that may influence reaction times is whether or not the participants are looking forward at the time of the event. If the participants are already looking forward, they may have adequate situational awareness of the driving situation. Hence, differing alert types and alert modalities may not make as big a difference as they would if they were not looking forward. It must be noted that whether or not a participant was looking forward at the time of the event was determined solely by examining their eyes at the exact time stamp of the event. However, they may or may not have been consistently monitoring the roadway before the event. There were 22 instances (out of the total 408) in which the participants were looking forward at the time of the event. A bar plot of the frequencies by alert modality*alert type group is displayed in Figure B-6 (note: the use of an asterisk [*] in this context indicates an interaction between two or more independent variables).

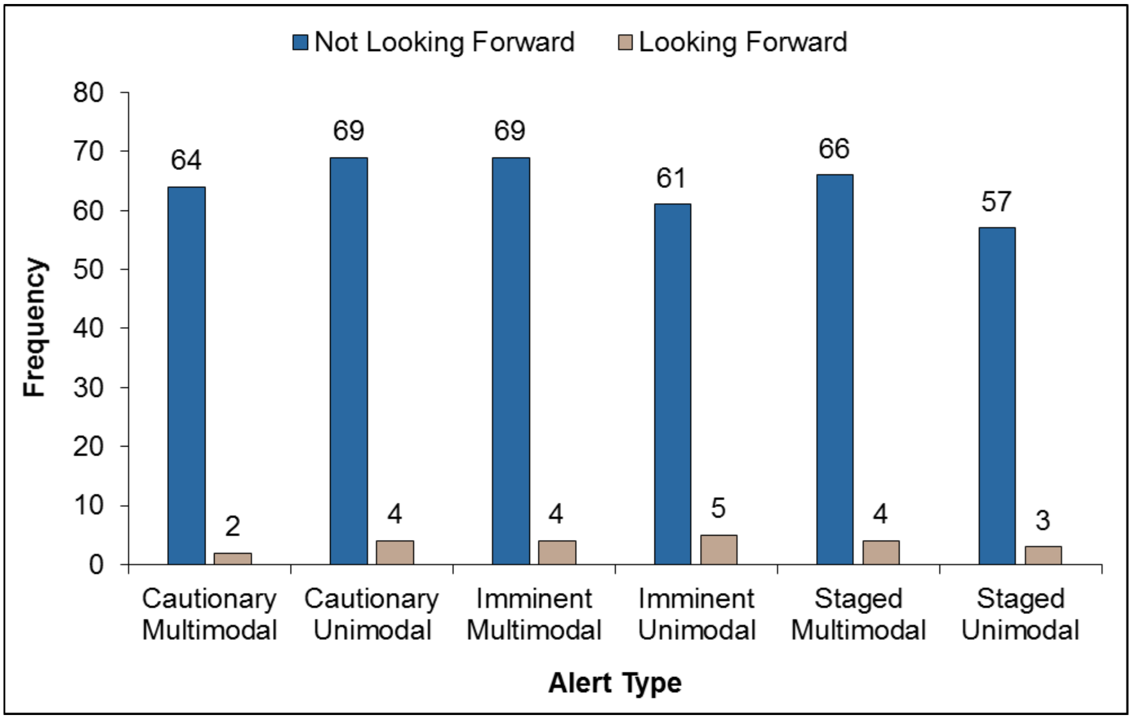
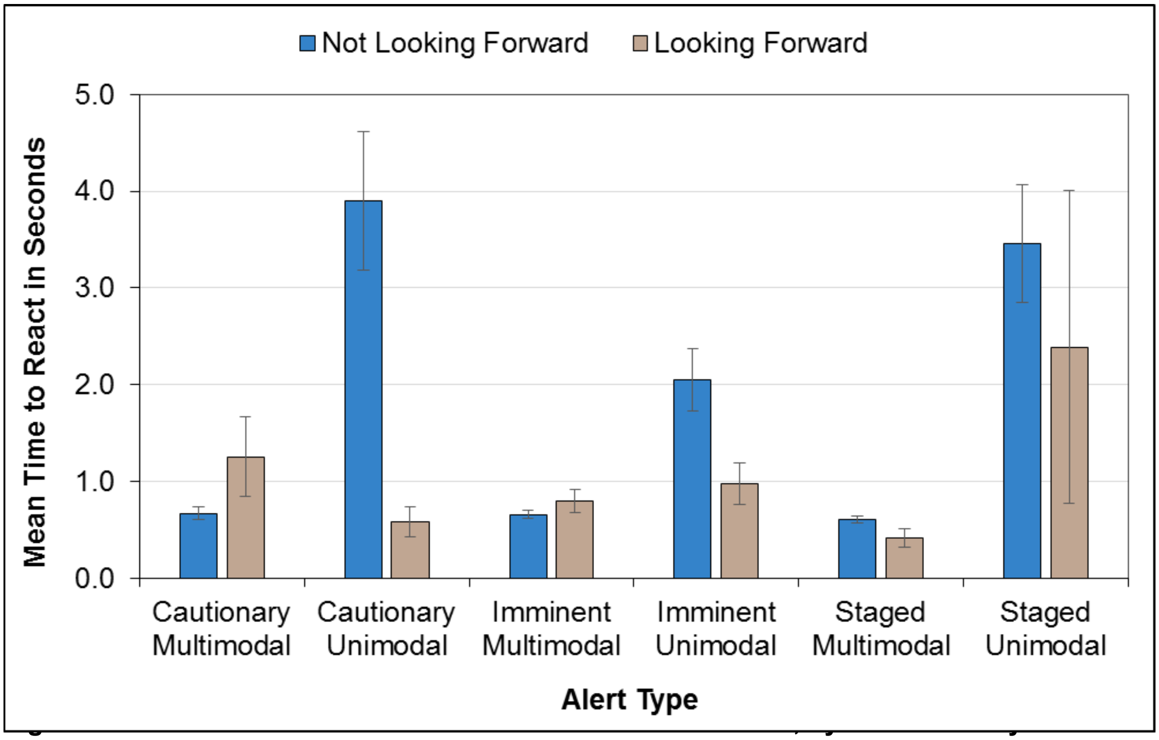



**Figure B-6. Bar Plot of Frequencies of Not Looking Forward/Looking Forward, by Alert Modality*Alert Type**

For each of the alert types (Cautionary, Imminent, and Staged), the mean time to react within unimodal alerts is higher when the participant was not looking forward compared to when the participant was looking forward. Interestingly, within the Cautionary unimodal group, in the four instances in which the participant was looking forward, the slowest reaction time was 1.0 s, faster than 51 of the 69 Cautionary unimodal participants (73.9 percent) who were not looking forward. The mean time to react after a Cautionary unimodal alert in which the participant was looking forward was 0.6 s (S.D. = 0.2 s), compared to the mean 3.9-second reaction time for those not looking forward (S.D. = 5.9 s).

Similarly, time to react within unimodal alerts was faster for both Staged and Imminent alerts if the participant was looking forward compared to when the participant was not looking forward. In Imminent unimodal alerts, the mean time to react for those looking forward was 1.0 s (S.D. = 0.5 s), compared with a mean of 2.1 s for those not looking forward (S.D. = 2.5 s). Within Staged unimodal alerts, the mean time to react for those looking forward was 2.4 s (S.D. = 3.4 s), compared to a mean of 3.5 s when they were not looking forward (S.D. = 4.6 s). A similar trend does not appear in multimodal alerts. However, due to the scant amount of instances where the participant was looking forward, analysis comparing times to react for instances looking forward to instances not looking forward may not be statistically significant. A mean and standard error bar plot for time to react stratified by the alert modality*alert type groups, and whether the participant was looking forward or not, is displayed in Figure B-7.




**Type Group and Looking Forward**

Statistical Analysis

To compensate for the skewed distribution of time to react, the analysis was conducted on the log-transformed time to react, and differences in means were back-transformed to allow for interpretations in terms of the untransformed time to react. For time to react, a random intercept was used to account for correlations for instances within the same participants, and the variances of each modality and alert combination were allowed to differ. The effect of looking forward was not significant in this analysis; thus, this variable was removed. The model results are displayed in Table B-1. In this table, and all other subsequent tables, Num DF = numerator degrees of freedom, Den DF = denominator degrees of freedom, Pr > F is the *p* value.

**Table B-1. Time to React, Analysis of Variance (ANOVA) Table for Experiment 1**

| Effect | Num DF | Den DF | F Value | Pr > F |
|---|---|---|---|---|
| **Alert Type** | 2 | 159 | 0.83 | 0.4367 |
| **Alert Modality** | **1** | **223** | **162.87** | **<0.0001** |
| **Alert Type * Alert Modality** | 2 | 158 | 1.58 | 0.2102 |

With $p < 0.0001$, alert modality is significant, indicating that there is a significant difference in the log of time to react between multimodal alerts and unimodal alerts. Alert type ($p = 0.4367$) and alert type by modality interaction ($p = 0.2102$) were not significant, indicating that differences in the log of time to react between different alerts were not statistically significant, and the difference in modalities did not vary across alerts.

To investigate further, the least squares means in the log of time to react (means adjusted for other variables) for multimodal and unimodal are displayed in Figure B-8, and the difference in the least squares means is displayed in Table B-2.


**Figure B-8. Plot of Least Squares Means for Time to React with Standard Error Bars, Modality for Experiment 1**

**Table B-2. Time to React Differences in Least Squares Means, Modality for Experiment 1**

| Modality | Comparison | Estimate | S.E. | *t* Value | *p* Value | Lower Confidence Limit | Upper Confidence Limit |
|---|---|---|---|---|---|---|---|
| **Multimodal** | **Unimodal** | -0.97 | 0.08 | -12.76 | <0.0001 | -1.12 | -0.82 |

The estimated difference of -0.97 indicates that, on average, the log of time to react was 0.97 log seconds less for multimodal alerts than for unimodal alerts (95 percent confidence interval -1.12 to -0.82). This estimate can be back-transformed using the anti-log transformation to make interpretations in terms of the actual time to react. When the back-transformation of the log of time occurs, the resulting back-transformed estimate is a multiplicative change in the actual time between one group and the other. For example, the back-transformation of -0.97 yields an estimate of 0.38. This implies that, on average, the time to react after experiencing a multimodal alert is 0.38 times (38 percent) as long, as the time to react after experiencing a unimodal alert. Another way to look at this is that on average, the amount of time operators would take to react may be about 62 percent faster after experiencing a multimodal alert than after experiencing a unimodal alert.

In summary, operators who experience a multimodal alert (visual and haptic components), may take significantly less time to react than they would after experiencing only a visual alert. However, the results seem to suggest that whether or not the alert is Cautionary, Imminent, or Staged may not significantly affect how long operators take to react to the alert.

***Analysis for Time to Regain Control***

As was the case for time to react, participants took longer, on average, to regain control of the vehicle after a unimodal alert compared to after a multimodal alert. The mean time to regain control after experiencing a multimodal alert was 1.3 s (S.E. = 0.07 s), compared to a mean of 4.8 s (S.E. = 0.45 s) for a unimodal alert. A mean and standard error bar plot for time to regain control is displayed in Figure B-9 by modality.

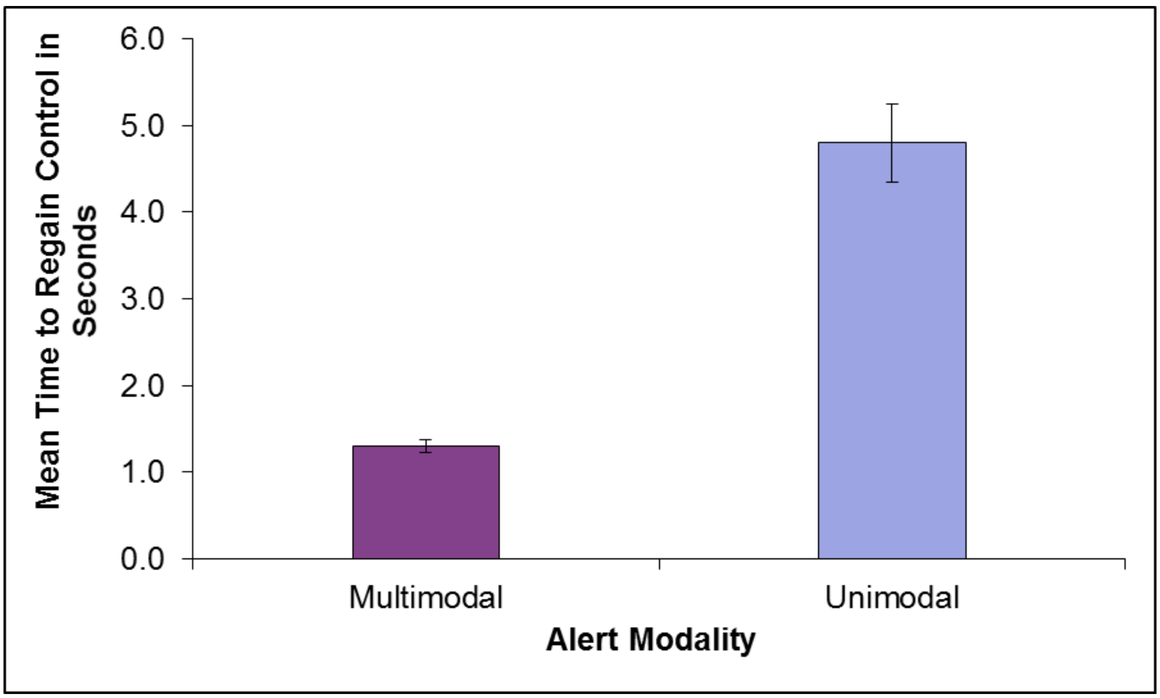



**Figure B-9. Mean and Standard Error Bar Plots of Time to Regain Control by Modality for Experiment 1**

Also similar to time to react, participants took a shorter time, on average, to regain control after experiencing an Imminent alert than they did after a Staged alert or a Cautionary alert. The mean time to regain control for Imminent alerts was 2.1 s (S.E. = 0.19 s), compared to a mean of 3.2 s for Staged alerts (S.E. = 0.43 s) and 3.6 s for Cautionary alerts (S.E. = 0.5 s). A mean and standard error bar plot for time to regain control is displayed in Figure B-10 by alert type.

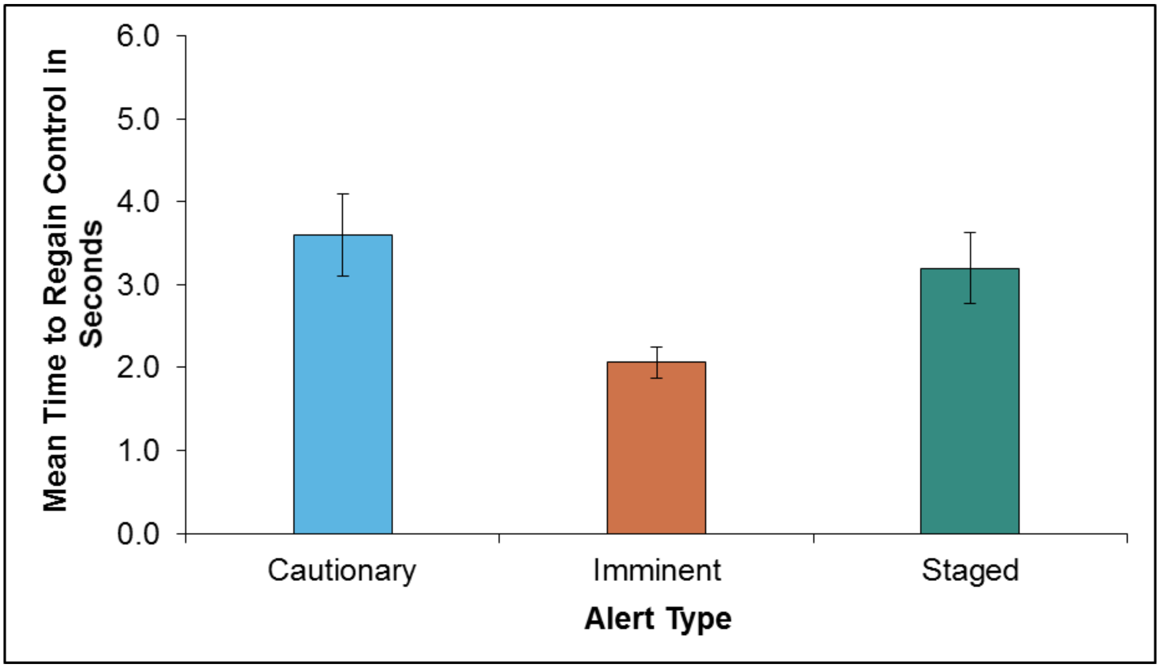



**Figure B-10. Mean and Standard Error Bar Plots of Time to Regain Control by Alert for Experiment 1**

The spread for time to regain control appears similar to that for time to react, as evidenced in the boxplots in Figure B-11. The difference is that, within unimodal alerts, there appears to be a greater separation between Imminent alerts and both Staged alerts and Cautionary alerts. This separation does not appear strong within multimodal alerts. Within the Imminent unimodal alerts, the mean time to regain control was 2.9 s (S.E. = 0.4 s), compared to a 5.3-second mean for Staged alerts (S.E. = 0.8 s) and a 6.3-second mean for Cautionary alerts (S.E. = 1.0 s).

The difference between Imminent alerts and the other alerts in terms of their mean time to regain control within multimodal alerts was not as sharp. In fact, the mean within Cautionary alerts was the lowest (1.1 s, S.E. = 0.5 s), compared to a 1.3-second mean within Imminent alerts (S.E. = 0.8 s) and a 1.4-second mean within Staged alerts (S.E. = 0.2 s). Of the 129 instances of a Staged alert in which the participant regained control of the vehicle, there were 113 instances (87.5 percent) in which participants regained control before the alert transitioned to the Imminent phase (i.e., prior to the 10-second mark following the alert). There were two instances in which a participant regained control more than 30 s after the alert: 31.0 s (Staged) and 30.5 s (Cautionary). Within Staged, multimodal alerts, there was only one instance in which a participant reached the Imminent phase before regaining control. During this instance, the participant regained control 11.0 s after the alert, despite having reacted to the alert in 1 second. However, the next highest time to regain control within Staged multimodal alerts was 5.3 s, indicating that almost all instances of a Staged, multimodal alert resulted in the participant regaining control with at least 4.6 s before the alert became imminent.

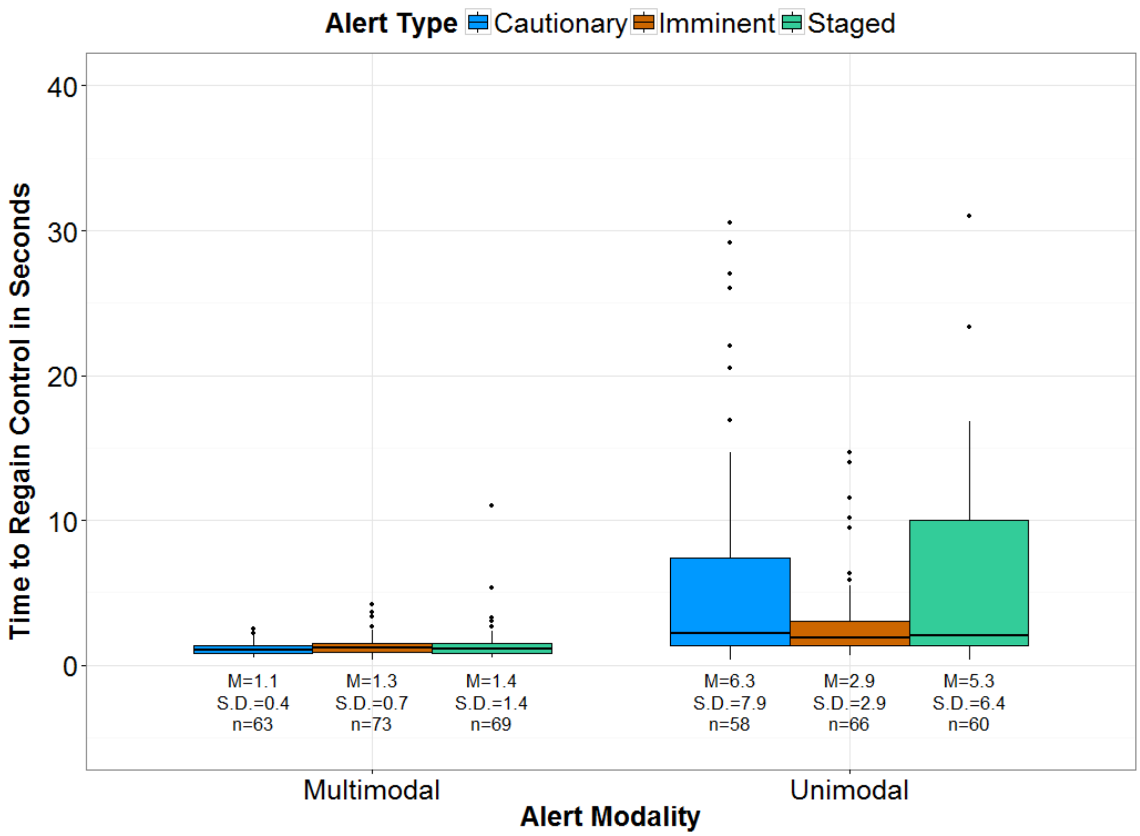



**Figure B-11. Boxplots of Time to Regain Control by Modality, Alert for Experiment 1**

Statistical Analysis

Time to regain control was log-transformed, and the analysis was performed on the log-transformed time to regain control. For this model, the different combinations of alert modality and alert type were allowed to differ in their variances, and a random intercept was used for each participant. The log of time to react was also included as a covariate, since time to regain control contains the time to react. The analysis results are displayed in Table B-3.

**Table B-3. Time to Regain Control, ANOVA Table for Experiment 1**

| Effect | Num DF | Den DF | F Value | Pr > F |
|---|---|---|---|---|
| Alert Type | 2 | 134 | 4.56 | 0.0122 |
| Alert Modality | 1 | 174 | 158.62 | <0.0001 |
| Alert Type * Alert Modality | 2 | 132 | 4.33 | 0.0151 |

Alert type ($p$ = 0.0122), alert modality ($p$ < 0.0001) and the alert modality*alert type interaction ($p$ = 0.0151) are all significant, indicating that there is evidence of a difference in the mean log of time to regain control between different alert types and different alert modalities, and evidence that the difference in alert modalities varies across alert types. A plot of least squares means of the log of time to regain control for each of the six combinations is displayed in Figure B-12.

Significance levels for alert types within alert modalities, and alert modalities within alert types, are displayed in Table B-4.

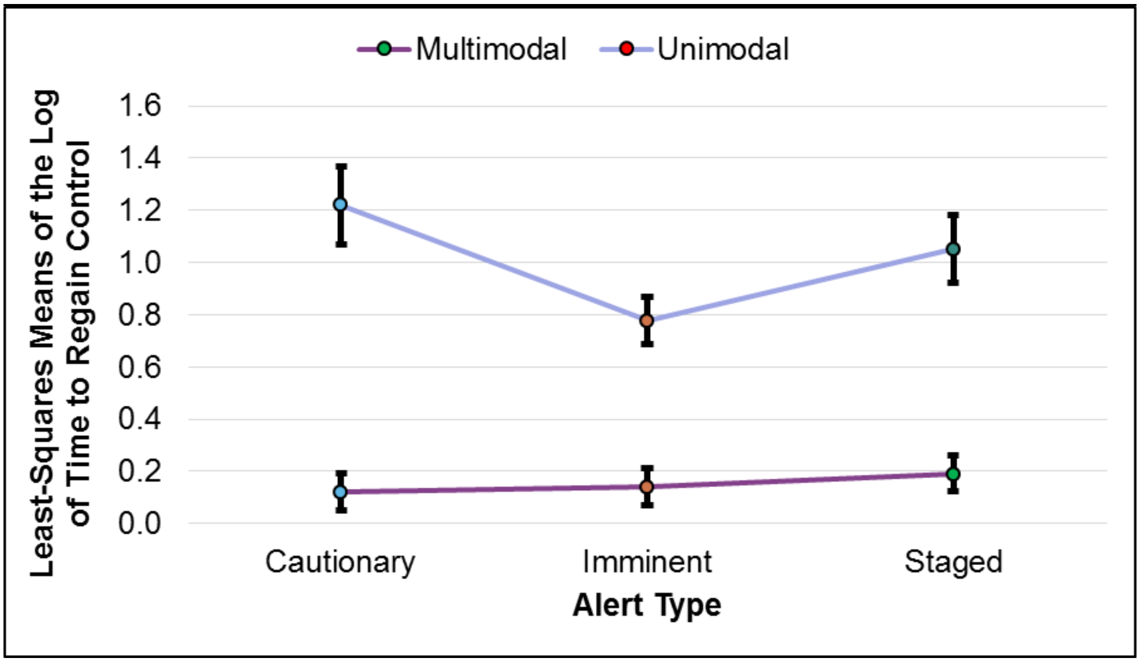



**Figure B-12. Plot of Least Squares Means for Time to Regain Control with Standard Error Bars, Alert Type by Alert Modality for Experiment 1**

**Table B-4. Significant Differences of the Log of Time to Regain Control Within Alert Modality and Alert Type for Experiment 1**

| Alert Modality | Alert Type | Num DF | Den DF | F Value | Pr > F |
|---|---|---|---|---|---|
| | **Cautionary** | **1** | **64.4** | **57.88** | **<0.0001** |
| | **Imminent** | **1** | **91.2** | **60.48** | **<0.0001** |
| | **Staged** | **1** | **75.3** | **50.05** | **<0.0001** |
| **Multimodal** | | 2 | 111 | 0.57 | 0.5657 |
| **Unimodal** | | **2** | **103** | **5.06** | **0.0080** |

For all three alert types, alert modality registered significant levels ($p < 0.0001$), indicating that the difference in the log of time to regain control between multimodal alerts and unimodal alerts is statistically significant regardless of the alert type. Similarly, for alert modality, unimodal alerts registered a significant result ($p = 0.0080$), indicating that within unimodal alerts, there is a significant effect of alert type on the log of time to regain control. However, within multimodal alerts, there was no statistically significant effect.

To investigate the differences further, the difference in log time to regain control between multimodal alerts and unimodal alerts is examined in Table B-5, and this difference is computed for Cautionary and Staged alerts. Meanwhile, in Table B-6, the significant differences in least squares means between different alert types are included, comparing only within unimodal alerts.

**Table B-5. Difference in Least Squares Means of the Log of Time to Regain Control Between Multimodal and Unimodal Alerts, Performed Within Cautionary and Staged Alerts, Respectively**

| Alert Type | Estimate | S.E. | *t* Value | Adj *p* Value | Lower Confidence Limit | Upper Confidence Limit |
|---|---|---|---|---|---|---|
| **Cautionary** | -1.10 | 0.15 | -7.61 | <0.0001 | -1.39 | -0.81 |
| **Imminent** | -0.64 | 0.08 | -7.78 | <0.0001 | -0.80 | -0.47 |
| **Staged** | -0.87 | 0.12 | -7.07 | <0.0001 | -1.11 | -0.62 |

**Table B-6. Significant Differences in Least Squares Means of the Log of Time to Regain Control Between Alert Types Within Unimodal Alerts**

| Alert Type | Comparison | Estimate | S.E. | *t* Value | Adj *p* Value | Lower Confidence Limit | Upper Confidence Limit |
|---|---|---|---|---|---|---|---|
| **Cautionary** | **Imminent** | 0.45 | 0.16 | 2.91 | 0.0162 | 0.14 | 0.77 |

The estimated difference in the log of time to regain control of -1.10 (95 percent confidence interval of -1.39 to -0.81) between multimodal alerts and unimodal alerts (within Cautionary alerts) can be interpreted as follows: for instances following a Cautionary multimodal alert, the log of the time to regain control is estimated to be 1.10 log seconds less than for Cautionary unimodal alerts. A similar interpretation can be made for the Imminent alert estimate of -0.64 (95 percent confidence interval -0.80 to -0.47) and the Staged alert estimate of -0.86 (95 percent confidence interval -1.11 to -0.62). These estimates of -1.10, -0.64 and -0.87 can be back-transformed to 0.33, 0.53, and 0.42, respectively. Hence, if an alert is Cautionary, then vehicle operators may take about 67 percent less time to regain control of the vehicle if the alert is multimodal compared to if the alert is unimodal. Similarly, if an alert is Imminent, then vehicle operators may take 47 percent less time to regain control if the alert is multimodal, compared to when the alert is unimodal. Finally, if an alert is Staged, then a vehicle operator may take 58 percent less time to regain control of the vehicle if the alert is multimodal compared to if the alert is unimodal.

Meanwhile, when comparing the different alert types within unimodal alerts, the estimated difference in the log of time to regain control of 0.45 (95 percent confidence interval of 0.14 to 0.77) between Cautionary alerts and Imminent alerts means that, for instances following a Cautionary unimodal alert, the log of the time to regain control (controlling for the log of the time to react) is estimated to be 0.45 log seconds more than for Imminent unimodal alerts. This estimate can be back-transformed to 1.57. Hence, if a unimodal alert is Cautionary, then vehicle operators may take about 57 percent more time to regain control of the vehicle compared to if the alert is Imminent.

In summary, the results provide evidence that, no matter what alert type is used, vehicle operators may regain control of the vehicle more quickly after the onset of the alert if the alert is multimodal compared to if the alert is unimodal. Moreover, if the alert is multimodal, there may not be significant gains made between alert types. However, if the alert is unimodal, an Imminent alert may result in faster times to regain control than Cautionary alerts.

### *Analysis for Time to Activate the Automation*

Only Imminent alerts required participants to activate the automation in all situations because the automation turned off, whereas in Staged alerts a participant would need to activate the automation only if the alert progressed to an Imminent alert before the participant regained control of the vehicle. Cautionary alerts did not require the participant to activate the automation because automation never deactivated. Therefore, only three combinations of alert type and alert modality had more than one valid observation: Imminent/multimodal, Imminent/unimodal, and Staged/unimodal (there were valid observations of time to activate the automation in Staged/multimodal because a Staged alert with more than 10 seconds of time to regain control resulted in an Imminent alert, and time to regain control tended to be lower in multimodal alerts). These combinations were compared directly as one factor. Figure B-13 shows the mean and standard error bar plots of time to activate the automation, stratified by these three groups.

The highest mean time to activate the automation within the three groups comes from the unimodal/Imminent group (mean = 4.0 s, S.E. = 0.3 s), while the means for the multimodal/Imminent group (3.4 s, S.E. = 0.3 s) and the unimodal/Staged group (3.4 s, S.E. = 0.9) are fairly similar. However, the mean for the unimodal/Staged alert is inflated by one value of time to activate the automation that is about 15.0 s, as shown in the boxplots in Figure B-14. During this instance the participant took approximately 9.0 s to locate the on button located on the center stack, and took an additional 6.0 s to attempt to activate the automation by pressing the on button.

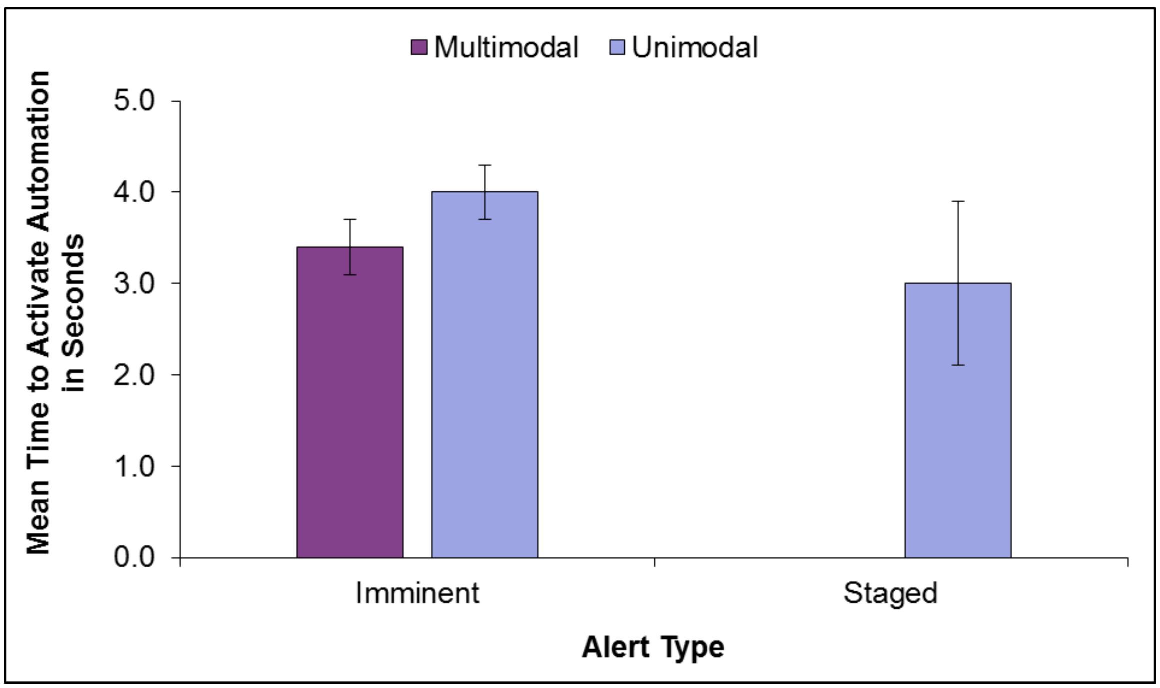



**Figure B-13. Mean and Standard Error Bar Plots of Time to Activate the Automation, Alert/Modality for Experiment 1**

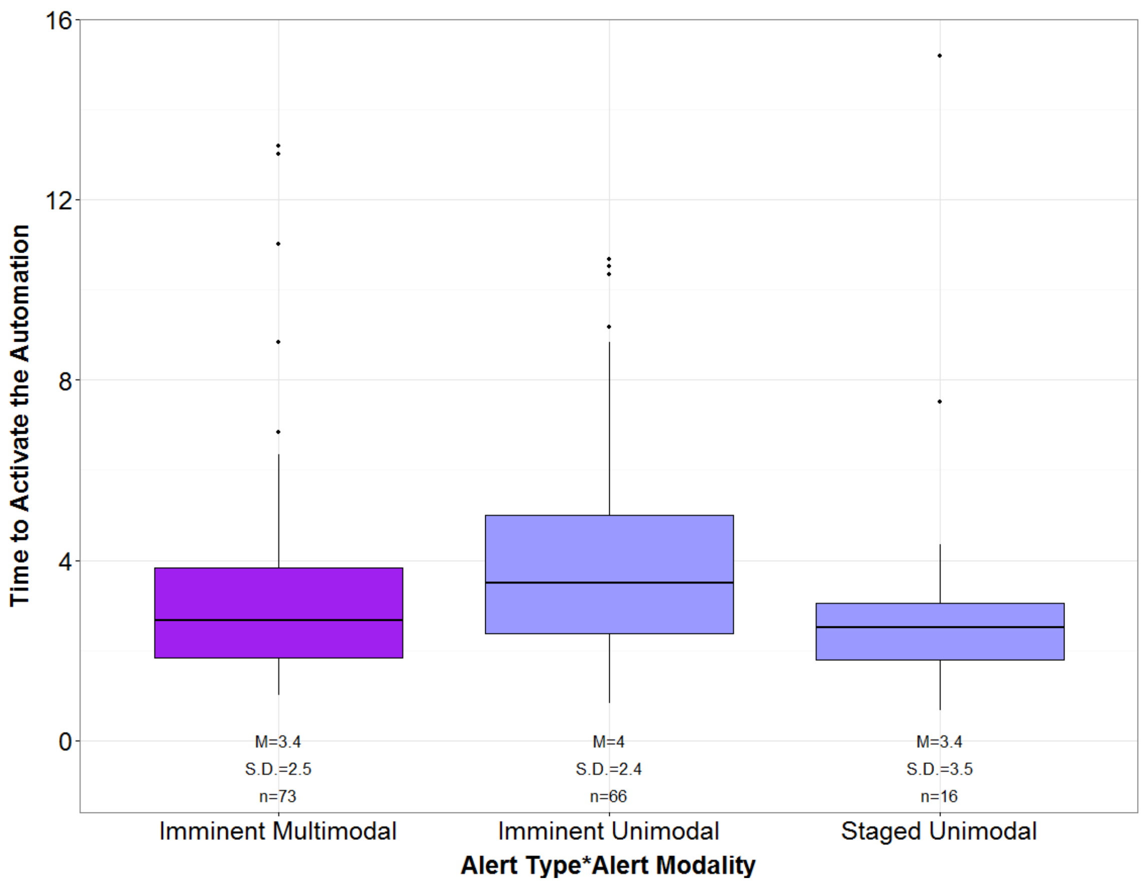



**Figure B-14. Boxplots of Time to Activate the Automation by Modality, Alert for Experiment 1**

Statistical Analysis

For this analysis, each alert modality*alert type group with more than one instance of an activation of the automation (i.e., multimodal Imminent, unimodal Imminent, and unimodal Staged) was considered one level of a single factor, called groups. The analysis was conducted with the single independent variable group, comparing the three alert modality*alert type combinations named above. The time to activate the automation was log-transformed, and the analysis was conducted on the log-transformed time. Further, a random intercept and time slope were added for each participant, but the variances of each alert type by alert modality combination was left the same. The analysis results are displayed in Table B-7.

**Table B-7. Time to Activate the Automation, ANOVA Table for Experiment 1**

| Effect | Num DF | Den DF | F Value | Pr > F |
|---|---|---|---|---|
| Group | 2 | 116 | 5.76 | 0.0041 |

The alert type by alert modality interaction is significant ($p$ = 0.0041), indicating that at least two of the three groups differ in their mean log time to activate the automation. To investigate further, least squares means are plotted in Figure B-15, and significant differences in least squares means are displayed in Table B-8.

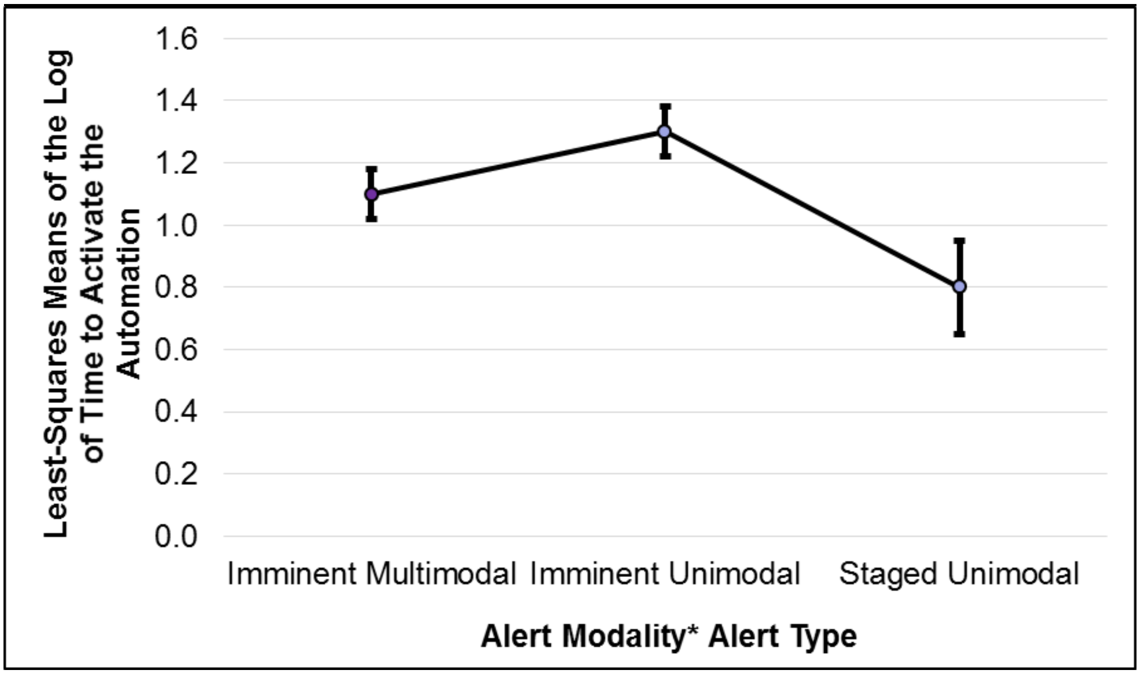



**Figure B-15. Plot of Least Squares Means for Time to Activate the Automation with Standard Error Bars, Alert/Modality for Experiment 1**

**Table B-8. Time to Activate the Automation Differences in Least Squares Means, Alert Type by Alert Modality for Experiment 1**

| Alert* Modality | Comparison | Estimate | S.E. | *t* Value | Adj *p* Value | Lower Confidence Limit | Upper Confidence Limit |
|---|---|---|---|---|---|---|---|
| **Unimodal-Imminent** | **Unimodal-Staged** | 0.42 | 0.15 | 2.83 | 0.0166 | 0.13 | 0.71 |
| **Unimodal-Imminent** | **Multimodal-Imminent** | 0.21 | 0.08 | 2.60 | 0.0321 | 0.05 | 0.37 |

The grouping of unimodal/Imminent alerts has a significantly higher mean than the other two groups. To interpret the above estimates in terms of actual time to activate the automation, the estimates are back-transformed to 1.52 (for unimodal Imminent versus unimodal Staged) and 1.24 (for unimodal Imminent versus multimodal Imminent). Hence, on average, operators may take 52 percent longer to activate the automation if they experience a unimodal Imminent alert as they would when experiencing a unimodal Staged alert, and they may take 24 percent longer to activate the automation after experiencing a unimodal Imminent alert as they would after experiencing a multimodal Imminent alert.

In summary, after having experienced an Imminent alert, operators may take longer to activate the automation if they have experienced only a visual alert compared to experiencing a visual and a haptic alert. Similarly, for those who experienced only a visual alert, it may still take longer for them to activate the automation if the alert is Imminent than if the alert is Staged.

***Analysis for Time to Release Control of Steering***

Differences between modalities in time to release control of steering do not seem strong, as seen in Figure B-16. The mean for multimodal alerts was slightly higher (3.6 s, S.E. = 0.28 s) than that for unimodal alerts (mean = 3.4 s, S.D. = 0.34 s). Hence, the amount of time participants took to release control of the steering was not influenced too greatly by different modalities.

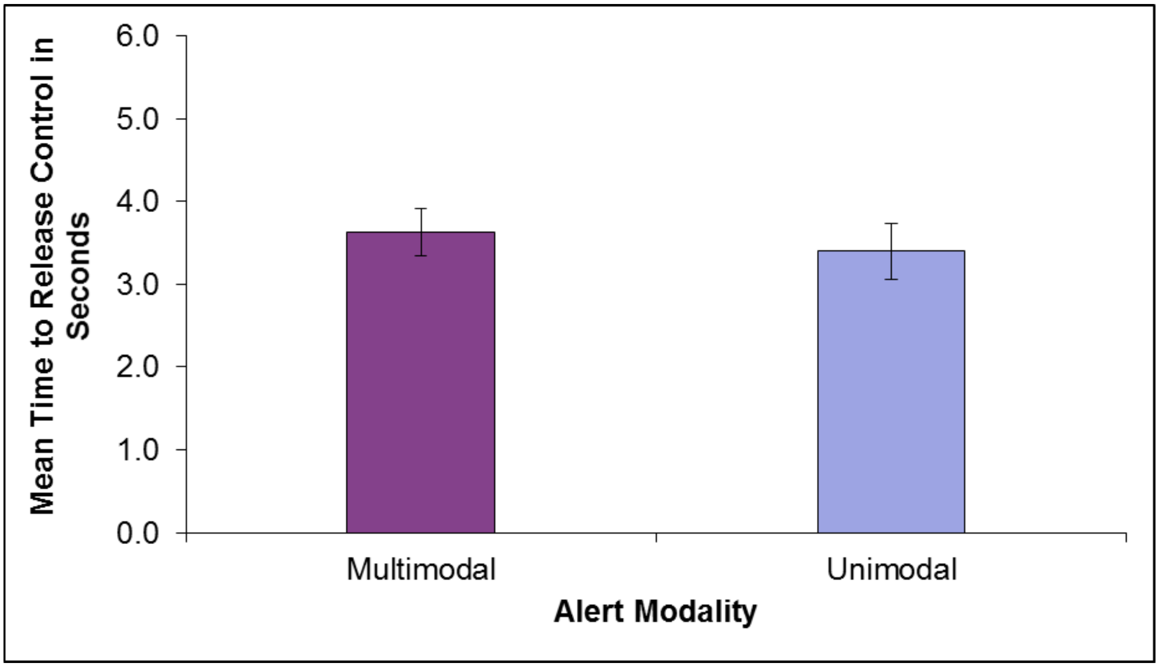



**Figure B-16. Mean and Standard Error Bar Plots of Time to Release Control of the Steering, by Alert Modality for Experiment 1**

It appears from Figure B-17 that the time to release control of steering was lower, on average, for participants after Imminent alerts than after Cautionary alerts. The mean time to release control of steering was 2.6 s after an Imminent alert (S.D. = 0.16 s) compared to means of 3.8 s and 4.3 s for Cautionary alerts and Staged alerts, respectively (S.E. = 0.3 s and 0.5 s).

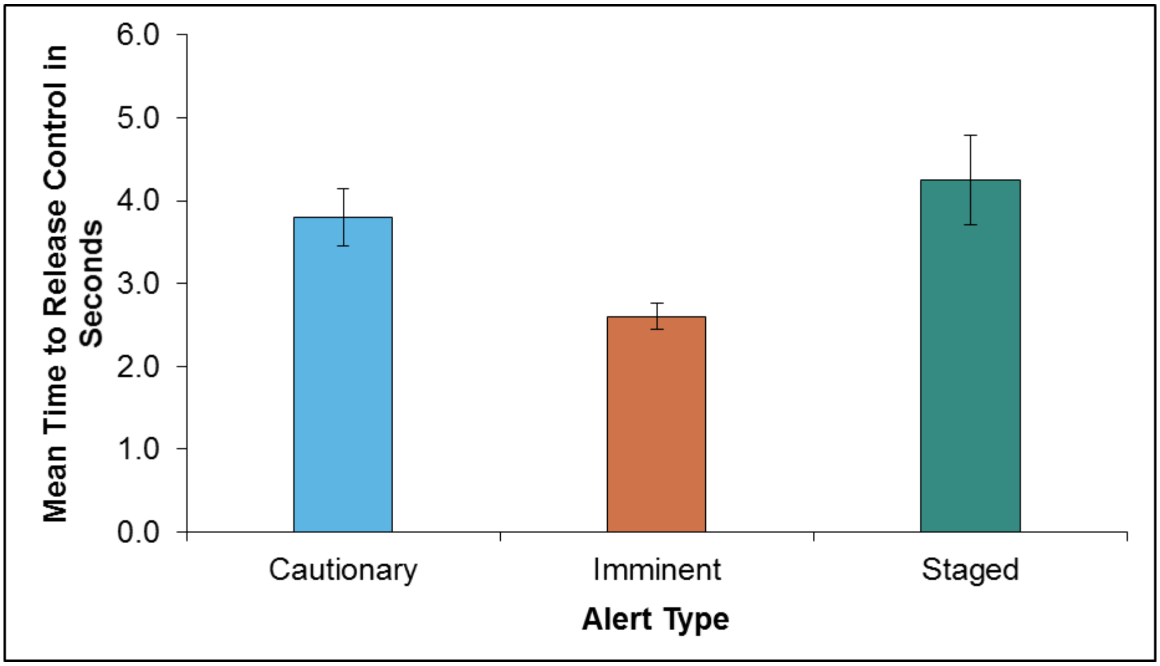



**Figure B-17. Mean and Standard Error Bar Plots of Time to Release Control of the Steering, by Alert Type for Experiment 1**

The boxplots in Figure B-18 indicate that the difference in the participants' mean time to release control among the alert types is slightly more pronounced in the multimodal alerts than in the unimodal alerts. In the multimodal alerts, the mean time to release control for Imminent alerts of 2.4 s (S.E. = 0.2 s) is 40 percent lower than the next highest mean (4.0 s for Cautionary alerts, S.D. = 0.4 s), and about 48 percent lower than the mean for Staged alerts (mean = 4.6 s, S.E. = 0.7 s). Meanwhile, the mean time to release control for unimodal Imminent alerts (mean = 2.8 s, S.E. = 0.27 s) is only about 10 percent lower than the mean for Staged alerts (mean = 3.1 s, S.D. = 0.3 s), and about 22 percent lower than the mean for Cautionary alerts (mean = 3.6 s, S.E. = 0.5 s).

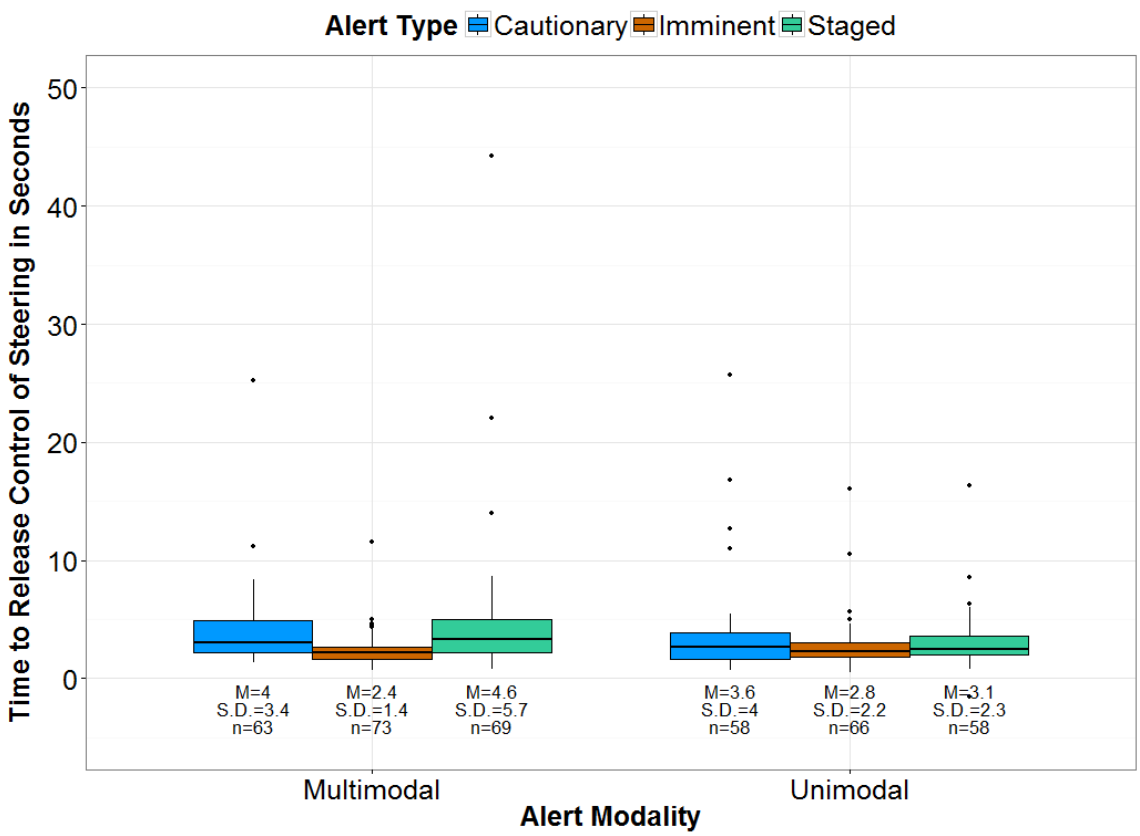



**Figure B-18. Boxplots of Time to Release Control of Steering by Modality, Alert for Experiment 1**

<u>Statistical Analysis</u>

The time to release control was log-transformed, and the analysis was performed on the log-transformed data. For this model, the variances of the alert type/alert modality combinations were kept the same, and a random intercept and time slope were added for each participant. The results for the model are displayed in Table B-9.

**Table B-9. Time to Release Control of Steering, ANOVA Table for Experiment 1**

| Effect | Num DF | Den DF | F Value | Pr > F |
|---|---|---|---|---|
| **Alert Type** | **2** | **356** | **13.33** | **<0.0001** |
| **Alert Modality** | **1** | **341** | **5.45** | **0.0202** |
| **Alert Type * Alert Modality** | **2** | **346** | **4.44** | **0.0124** |

The significant interaction between alert type and alert modality ($p$ = 0.0124) provides evidence that the effect of alert on the log time to release control varies across modalities. To examine this further, the effects of alert modality were tested for significance within each alert type, and vice versa. The $p$ value used was a Bonferroni adjusted $p$ value of less than 0.01. Mean and standard error bar plots of the least squares means are displayed in Figure B-19. Significance levels of each of these tests are displayed in Table B-10.

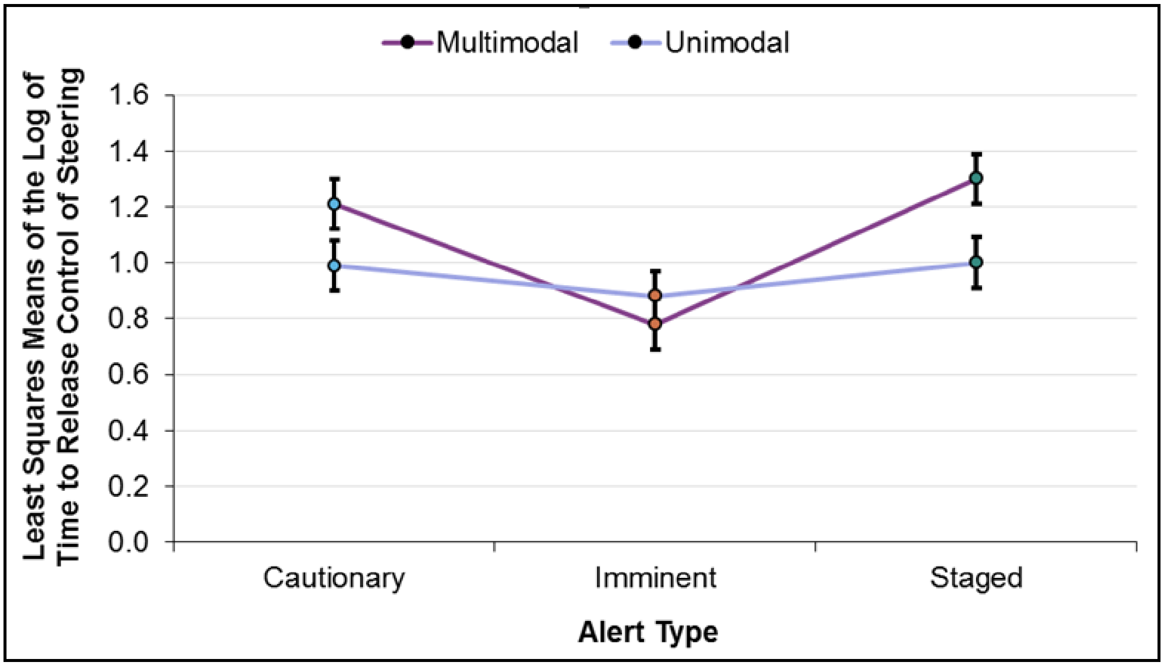



**Figure B-19. Plot of Least Squares Means for Time to Release Control of Steering with Standard Error Bars, Alert/Modality for Experiment 1**

**Table B-10. Significant Differences of the Log of Time to Release Control Within Alert Modalities and Alert Types for Experiment 1**

| Alert Modality | Alert Type | Num DF | Den DF | F Value | Pr > F |
|---|---|---|---|---|---|
| | **Cautionary** | 1 | 348 | 5.42 | 0.0205 |
| | **Imminent** | 1 | 342 | 1.15 | 0.2835 |
| | **Staged** | **1** | **344** | **6.95** | **0.0088** |
| **Multimodal** | | **2** | **351** | **17.72** | **<0.0001** |
| **Unimodal** | | 2 | 355 | 1.16 | 0.3142 |

Table B-10 indicates that there may be a significant effect of alert modality within Staged alerts, but not within Cautionary or Imminent alerts. Further, there may be a significant effect of alert type within multimodal alerts, but not within unimodal alerts. To investigate these differences, the comparison of alert modalities within Staged alerts is displayed in Table B-11, while the significantly different comparisons between alert types within multimodal alerts are displayed in Table B-12.

**Table B-11. Difference in Least Squares Means Between Multimodal and Unimodal Alerts, Within Staged Alerts, For Experiment 1**

| Alert Type | Estimate | S.E. | *t* Value | Adj *p* Value | Lower Confidence Limit | Upper Confidence Limit |
|---|---|---|---|---|---|---|
| **Staged** | 0.25 | 0.09 | 2.64 | 0.0088 | 0.07 | 0.43 |

**Table B-12. Significant Differences in Least Squares Means Between Alert Types, Within Multimodal Alerts, For Experiment 1**

| Alert Type | Comparison | Estimate | S.E. | *t* Value | Adj *p* Value | Lower Confidence Limit | Upper Confidence Limit |
|---|---|---|---|---|---|---|---|
| **Cautionary** | **Imminent** | 0.43 | 0.09 | 4.72 | <0.0001 | 0.25 | 0.60 |
| **Staged** | **Imminent** | 0.47 | 0.09 | 5.42 | <0.0001 | 0.30 | 0.64 |

Within multimodal alerts, both Cautionary alerts and Staged alerts had higher least squares means of the log time to release control than did Imminent alerts. The estimated difference between Cautionary and Imminent was 0.43 (95 percent confidence interval 0.25 to 0.60), while the estimated difference between Staged and Imminent was 0.47 (95 percent confidence interval 0.30 to 0.64). Similarly, within Staged alerts, multimodal alerts had an estimated 0.25 log seconds higher least squares mean (95 percent confidence interval 0.07 to 0.43) than unimodal alerts.

The estimated differences can be interpreted in terms of the actual time to release control of the steering by back-transforming them. The estimated difference of 0.43 between multimodal Imminent and multimodal Cautionary can be back-transformed to 1.53, indicating that operators who experienced a multimodal alert may take about 53 percent longer to release control of the steering if they have experienced a Cautionary alert compared to if the alert was Imminent. Similarly, the estimated difference between Staged and Imminent alerts of 0.47 is back-transformed to 1.60, indicating that, after experiencing a multimodal alert, participants would take 60 percent longer to release control if the alert was Staged compared to if the alert was Imminent. Finally, the estimated difference of 0.25 between multimodal and unimodal alerts (within Staged alerts) was back-transformed to 1.28, indicating that within Staged alerts, operators experiencing a multimodal alert may take 28 percent longer to release control of the steering than if they had experienced a unimodal alert.

In summary, Imminent alerts that included both the visual component and the haptic component had significantly faster times to release control of the steering than did both the Cautionary and the Staged alerts. Additionally, if the alert was Staged, then multimodal alerts had significantly longer times to release control than unimodal alerts.

***Analysis for Time to Resume a Non-driving Task***

In examining time to resume a non-driving task, the participants took, on average, slightly longer to resume a non-driving task after experiencing a multimodal alert (mean = 2.9 s, S.E. = 0.3 s) than after a unimodal alert (mean = 2.6 s, S.D. = 0.3 s), as indicated in Figure B-20.

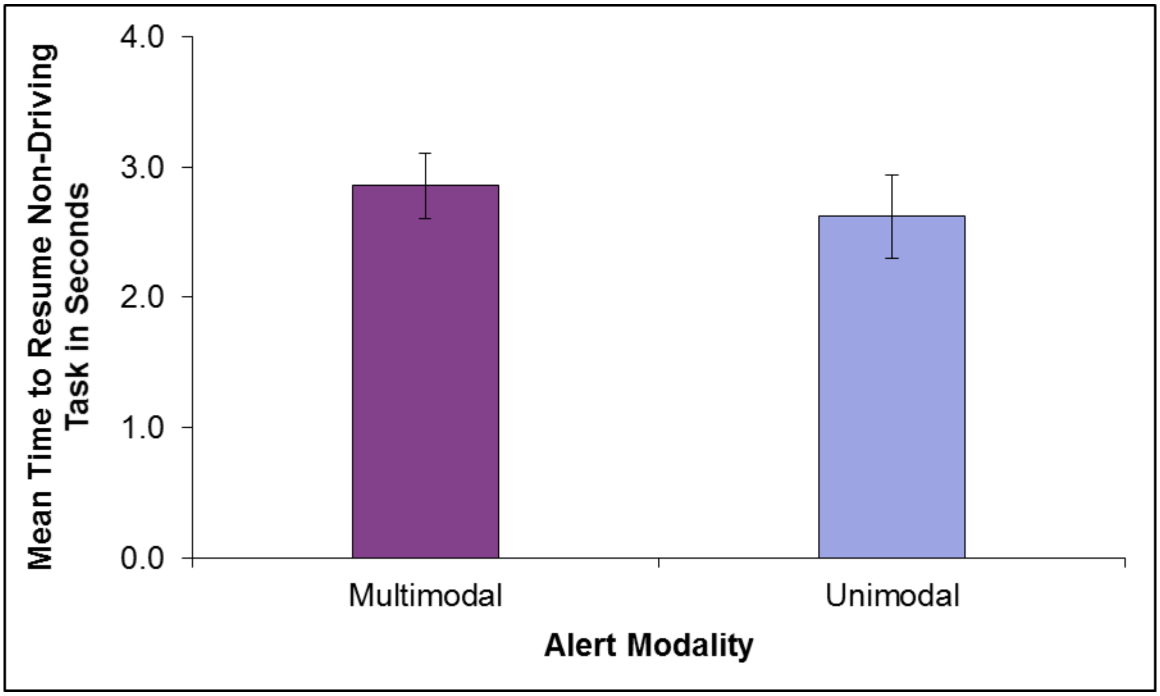



**Figure B-20. Mean and Standard Error Bar Plots of Time to Resume Non-driving Task, Modality for Experiment 1**

Additionally, after Imminent alerts, the mean time to resume a non-driving task for the participants (mean = 2.8 s, S.E. = 0.3 s) was about equal to that for Cautionary alerts (mean = 2.7 s, S.E. = 0.3 s) or Staged alerts (mean = 2.8 s, S.D. = 0.4 s), as indicated in Figure B-21.

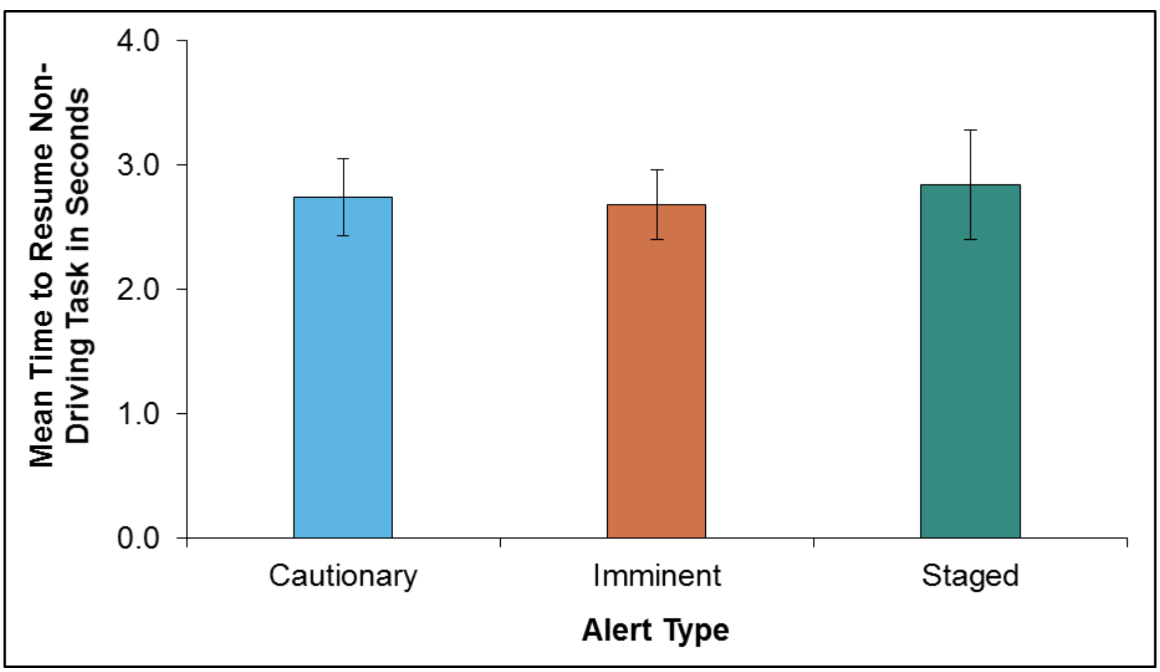



**Figure B-21. Mean and Standard Error Bar Plots of Time to Resume Non-driving Task, Alert for Experiment 1**

The boxplots in Figure B-22 indicate that both the Cautionary alert and the Staged alert have more outlying points than the Imminent alert, and this holds true in both multimodal and unimodal alerts. Although there is one instance for Imminent alerts in which the time to resume a non-driving task is above 30.0 s, only three observations are above 10.0 s, compared to six for Cautionary alerts and eight for Staged alerts. This may be pulling the means up for Cautionary alerts and Staged alerts, compared to Imminent alerts.

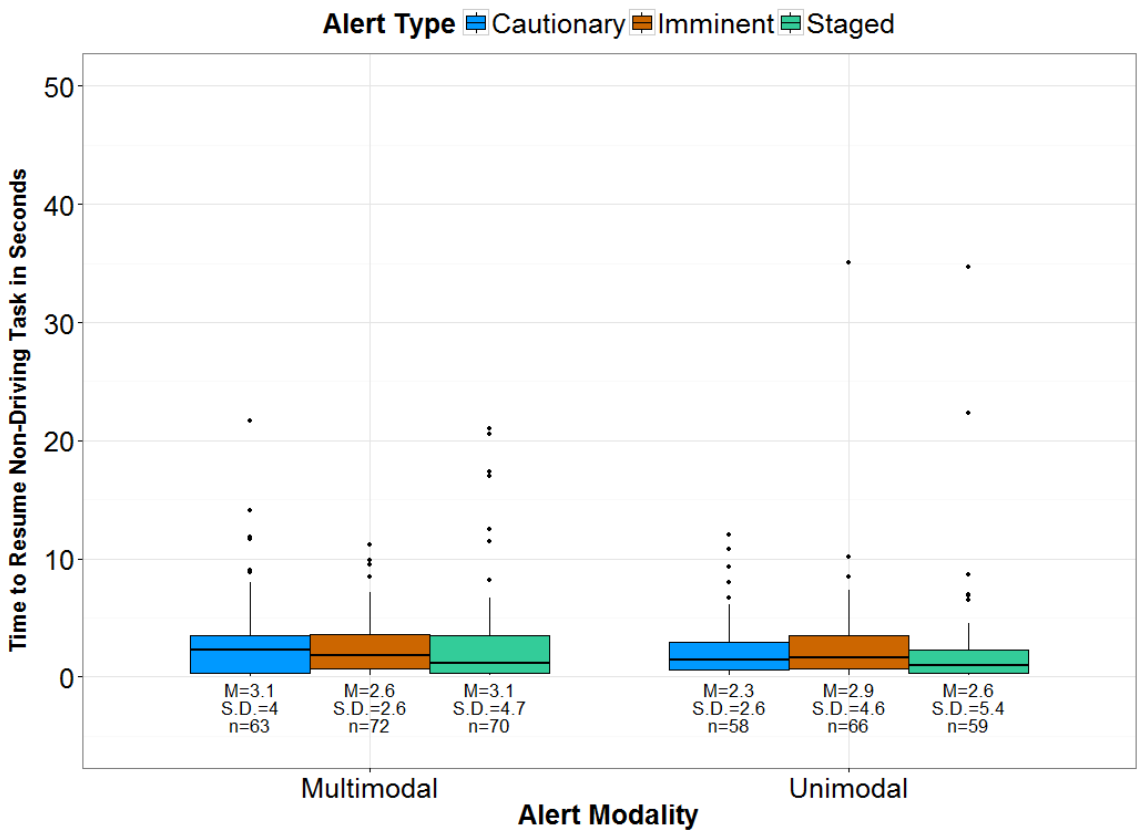



**Figure B-22. Boxplots of Time to Resume a Non-driving Task, Modality and Alert for Experiment 1**

Statistical Analysis

The time to resume a non-driving task was log-transformed, and the analysis was conducted on the log-transformed time. A random intercept and time slope were also included to account for individual correlation, but the variances for each alert type/alert modality group were given a common estimate. Results for the analysis are displayed in Table B-13.

**Table B-13. Time to Resume a Non-driving Task, ANOVA Table for Experiment 1**

| Effect | Num DF | Den DF | F Value | Pr > F |
|---|---|---|---|---|
| **Alert Type** | **2** | **353** | **4.53** | **0.0114** |
| **Alert Modality** | 1 | 338 | 0.00 | 0.9485 |

| Effect | Num DF | Den DF | F Value | Pr > F |
|---|---|---|---|---|
| **Alert Type * Alert Modality** | 2 | 342 | 0.16 | 0.8499 |

With $p = 0.0114$, alert type is significant, indicating that at least two alerts differ in their log of time to resume a non-driving task. Alert modality ($p = 0.9485$) and the alert type/alert modality interaction ($p = 0.8499$) were not significant, indicating that there is no significant difference in the log of time to resume a non-driving task between modalities, nor does the difference between alerts vary across modalities.

To investigate which pairs of alerts differ in their log of time to resume a non-driving task, error plots of least squares means are displayed in Figure B-23, while differences in least squares means between pairs of alerts are displayed in Table B-14. Only significant differences are displayed.

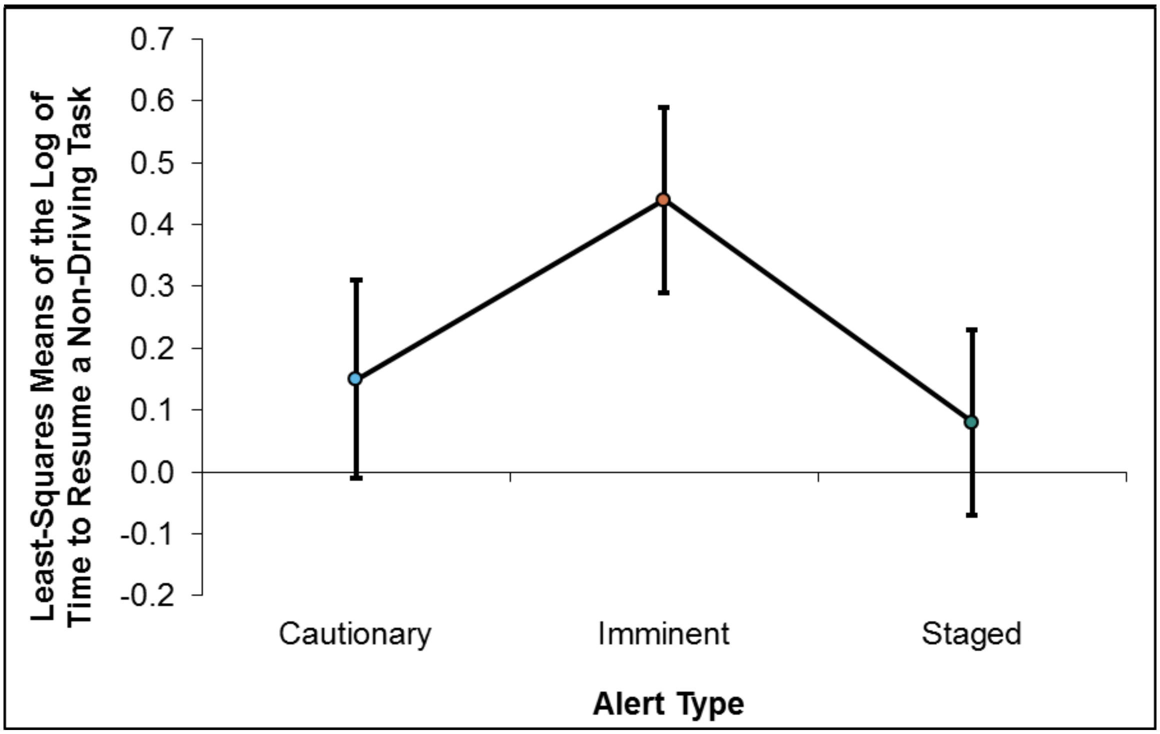



**Figure B-23. Plot of Least Squares Means for Time to Resume Non-driving Task with Standard Error Bars, Alert Type for Experiment 1**

**Table B-14. Time to Resume Non-driving Task Differences in Least Squares Means, Alert Type for Experiment 1**

| Alert | Comparison | Estimate | S.E. | *t* Value | Adj *p* Value | Lower Confidence Limit | Upper Confidence Limit |
|---|---|---|---|---|---|---|---|
| **Imminent** | **Staged** | 0.35 | 0.12 | 2.85 | 0.0141 | 0.11 | 0.60 |

The data suggest that the mean log time to resume a non-driving task may be greater after experiencing an Imminent alert than after experiencing a Staged alert. When comparing Imminent alerts to Staged alerts, the estimate of 0.35 (95 percent confidence interval 0.11 to 0.60) in the difference between Imminent and Staged alerts indicates that, on average, the log seconds of time to resume a non-driving task after experiencing an Imminent alert will be 0.35 log seconds more than for those experiencing a Staged alert. In terms of actual seconds, the estimate of 0.36 is back-transformed to 1.42, indicating that when a vehicle operator experiences an Imminent alert, the operator may take 42 percent longer to resume a non-driving task as compared to a Staged alert. There was, meanwhile, no significant difference in log time to resume a non-driving task between a Cautionary alert and either an Imminent alert or a Staged alert.

Though the finding that it took longer to resume a non-driving-related task after experiencing an Imminent alert may be surprising, given that the mean time to resume a non-driving task for Imminent alerts was the lowest of all the alerts (as referenced in Figure B-21), the median for Imminent alerts was the highest of all the alerts (1.75 s, compared to a 1.67 s median for Cautionary alerts and a 1.01 s median for Staged alerts). This implies that the means for Cautionary and Staged alerts were inflated compared to Imminent alerts by more extreme high values. The log transformation, however, is robust to high untransformed values but preserves order. Therefore, the significantly higher log-transformed time for Imminent alerts makes sense given the higher median of the Imminent untransformed times.

In summary, these data suggest that operators experiencing an Imminent alert may not resume a non-driving task as quickly after ceding control back to the automated system than they would after experiencing a Staged alert. However, an operator who experiences both a haptic alert and a visual alert may not resume a non-driving task any more or less quickly than they would if they experienced only a visual alert.

## Trust

Trust scales used in the assessment of automated systems (Lee & Moray, 1994; Luz, 2009; Jian, Bisantz, & Drudy, 2000) were reviewed in order to develop the ones used for the current study. Drawing upon the 19-question trust scale used by Luz (2009), key phrases examining different aspects of trust (overall trust, perceived reliability, perceived competence under automated control, and perceived understandability) were identified. The wording of the phrases was modified to reflect the automated system in use as opposed to the medium-term conflict detection system studied by Luz. Responses were obtained using 7-point Likert-type scales (Appendix E).

Another important aspect of the automated system and the alerts is the operators' subjective assessment of their effectiveness. While the operators' belief in the system's ability to function can help them use it effectively, a certain level of mistrust can help prevent misuse of an L2 system. The participants' subjective measure of their trust in the system was analyzed to determine if their level of trust in the system changed over time.

Various subjective measures were used to gauge participant trust and overall impressions of the system, vehicle, and experience during the experiment. A 7-point Likert-type scale was administered 10 times throughout the experimental run. Participants were read the statement "I can trust the automated system to function properly while I am doing something else" and asked to provide a rating from 1 (Strongly Disagree) to 7 (Strongly Agree) (see Appendix E).

In addition, each participant completed after-experience trust scales that listed six statements gauging overall trust, perceived reliability, perceived competence under automated control, and perceived understandability using a 7-point Likert–type scale (Appendix E). This was followed by an interview in which the experimenter asked questions regarding overall impressions, comfort level with the automation, thoughts regarding system alerts, and concerns with automated vehicle technology. The interviews were recorded and transcribed and were used to provide additional insight into the participants' trust ratings. (See Appendix F for additional details).

***Trust over Time***

This analysis helped to determine if there was a change in trust in the system over time. Table B-15 states that time was not a significant predictor of rating, indicating that as time went on, participants did not, on average, increase or decrease their trust in the system. This could potentially be due to the alerts being presented at fairly regular intervals.

**Table B-15. Summary of Trust Analysis for Experiment 1**

| Variable | Performance Over Time | Takeaways |
|---|---|---|
| **Trust** | Not Significant | Participants trusted the system, but their trust did not significantly increase or decrease over time. |

The participants often indicated that they trusted the system to some degree. Out of 249 total ratings across the experiment from the 25 participants, 227 (91.16 percent) were at least a "5," which corresponds to "slightly agree." Therefore, in about 91.16 percent of the instances, participants agree to some extent that they trusted the system. Only in 17 instances (6.8 percent) did participants express some level of disagreement, with 15 of those instances being "slightly disagree." Only in two instances did a participant "moderately disagree," and at no point did anyone "strongly disagree." Hence, in general, the level of trust the participants expressed in the system was favorable. A bar plot of rating frequency is displayed in Figure B-24.

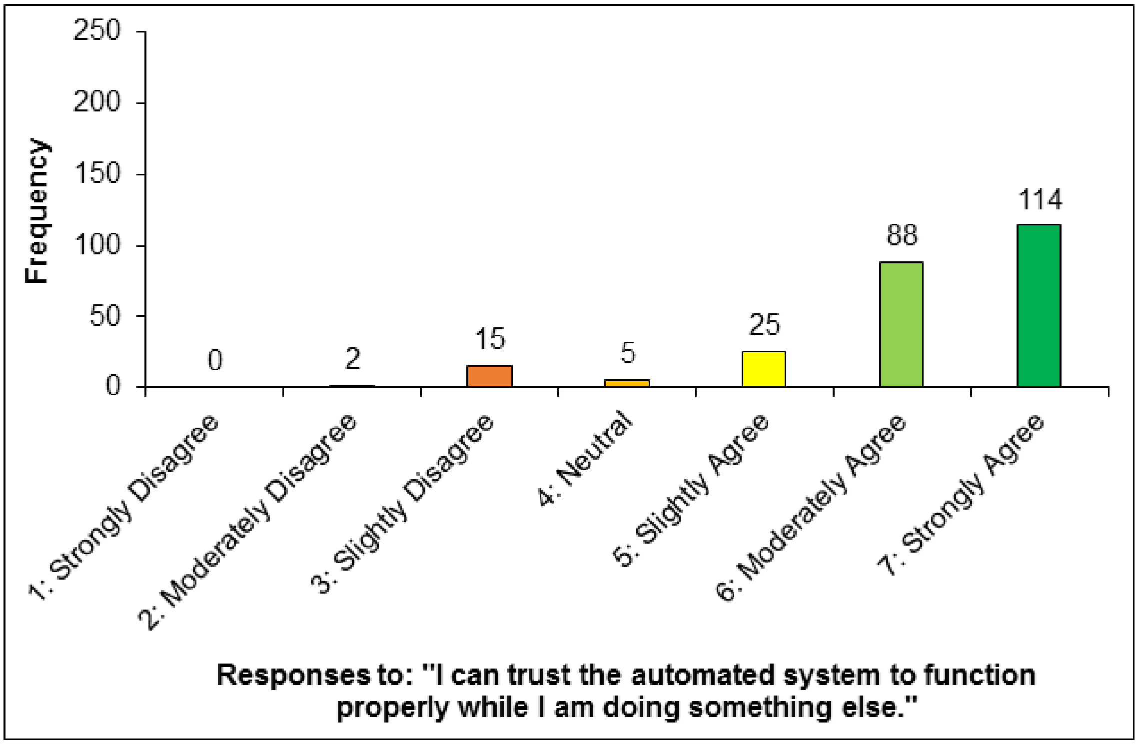



**Figure B-24. Bar Plot of Experiment 1 Trust Ratings Across Full Experiment**

Turning to the question of whether trust changes over time for individual participants, Figure B-25 displays each subject's time trend of rating. The figure indicates somewhat inconsistent trends across the participants, but a majority of the trends do not reflect much change over time. Six participants, in fact, did not alter their rating at all, while an additional six made only one change throughout the experiment, and the change was only one level (5 went up one level, 1 went down). The two "moderately disagree" ratings came from the same participant (ID=114, age=18), who also had seven "slightly disagree" ratings and one "neutral" (at the beginning of the experiment). Another participant (55+ age group), had three "neutral" ratings at the beginning of the experiment, while the rest were "moderately disagree" ratings. All but one of the ratings that were "disagree" ratings came from these two participants, and four of the five "neutral" ratings came from them as well.

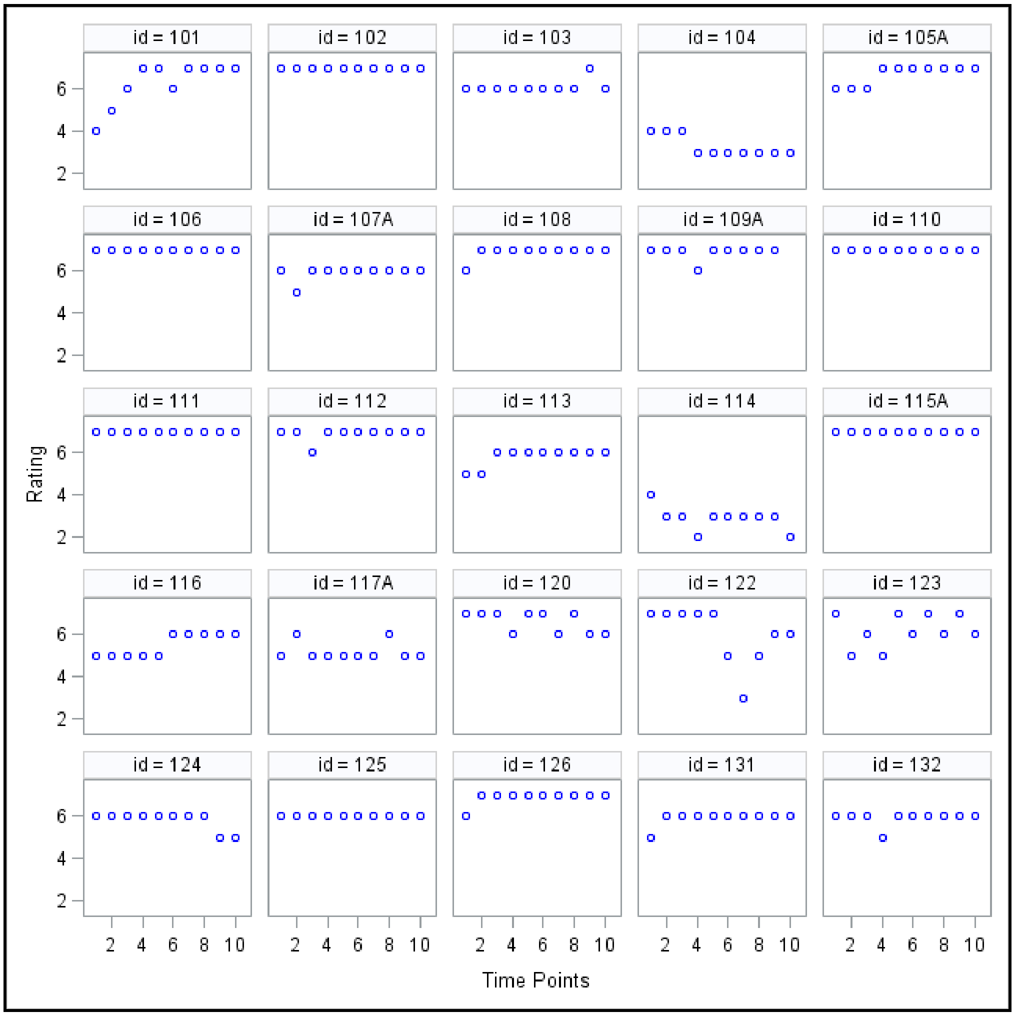



**Figure B-25. Subject Time Trends in Trust for Experiment 1**

Statistical Analysis

A mixed model was run with rating as the response and actual time in minutes as the predictor of interest, with age and events between the previous time point and the current time point as covariates. A visual residual analysis of the model revealed no major departure from the assumptions of the model, so the results for time were kept and are reported. These results are displayed below in Table B-16.

**Table B-16. Trust Model, Experiment 1**

| Effect | Num DF | Den DF | F Value | Pr > F |
|---|---|---|---|---|
| **Time** | 1 | 24 | 0.69 | 0.4137 |

With *p* = of 0.4137, time is not statistically significant, indicating that there is no statistically significant evidence of a statistically significant relationship between time and rating. The conclusion is that, although trust scale ratings are relatively high, they do not increase (or decrease) significantly over time.

***After-experience Trust Scales***

Means, standard deviations, modes, minimums, and maximums for the after-experience trust scales are provided in Table B-17.

**Table B-17. Descriptive Statistics for Responses to the After-experience Trust Scales for Experiment 1**

| Variable | Mean | S.D. | Mode (Frequency) | Minimum | Maximum |
|---|---|---|---|---|---|
| **TS1** | 6.16 | 1.11 | 6 (11) | 3 | 7 |
| **TS2** | 6.04 | 1.49 | 7 (12) | 1 | 7 |
| **TS3** | **4.00** | **2.35** | 7 and 1 (5) | 1 | 7 |
| **TS4** | 5.80 | 1.44 | 7 and 6 (8) | 2 | 7 |
| **TS5** | 6.00 | 1.47 | 7 (11) | 2 | 7 |
| **TS6** | 5.88 | 1.33 | 7 and 6 (8) | 2 | 7 |

Responses to Statement 3, “The automated system gave false alerts,” have a lower mean (4.00) than any other statement (the next lowest mean was 5.8 for Statement 4). Additionally, responses to Statement 3 had a higher amount of variability than did the responses to the other statements, with an S.D. of 2.35. The next highest was Statement 2, with an S.D. of 1.49. An equal amount of participants (*n* = 5; 25 percent) answered “strongly agree” and “strongly disagree,” and these were the two most common answers. In all other cases, the mode was either “strongly agree” or “moderately” agree or both. Hence, while participants were largely in agreement over their level of trust in the system after the experiment, they did not always agree regarding whether or not the system gave false alerts. Additional subjective data for Experiment 1 can be found in Appendix F.

# Appendix C. Experiment 2: Expanded Methods, Results, and Analysis

## Experimental Design

An a priori power analysis was conducted to determine the total number of participants required for this study to reach a power of .8 with α = .05. Assuming a large effect size (.8), a total of 45 participants were required for the design. In order to ensure all age groups were represented, participants were recruited based on the four NHTSA age groups (18–24, 25–39, 40–54, and 55 and older; Visual-Manual NHTSA Driver Distraction Guidelines for In-vehicle Devices, 2013).

## Procedure

### Screening

The experimental protocols were reviewed and approved by the Virginia Tech Institutional Review Board (IRB). Participants were recruited from the greater Detroit, Michigan, area using newspaper and Craigslist advertisements, and through word of mouth. Participants were eligible for inclusion in the study if they met all of the following criteria:

- Valid driver's license, and were at least 18 years old;
- No disqualifying moving violations in the past 3 years;
- Had not previously participated in any GM study at the Milford Proving Ground;
- Owned a smartphone;
- Normal or corrected-to-normal vision and hearing;
- Able to drive without sunglasses or photochromatic lenses;
- Able to drive an automatic transmission vehicle without accommodation;
- If pregnant, discussed the risks of participating with a physician;
- Reported not taking medications or substances that interfere with driving ability;
- No lingering health effects or any recent heath events, including high propensity towards motion sickness;
- Eligible for employment in the United States; and
- Able to wear closed-toe shoes while driving.

## Greeting

Upon arriving at the test site, participants were greeted by the in-vehicle experimenter. The experimenter escorted the participant to the preparation area and provided the informed consent documentation. The participant's driver's license was examined to ensure validity. Following informed consent, the participant was asked to undergo a basic vision screening and complete the demographics questionnaire. Hearing was not formally assessed; however, the participant was asked to respond to basic questions in the greeting process that determined if normal hearing seemed to be present.

## Participants

Data were collected from 56 participants. The age groupings used in Figure C-1 reflect those proposed in the Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Devices (2013).

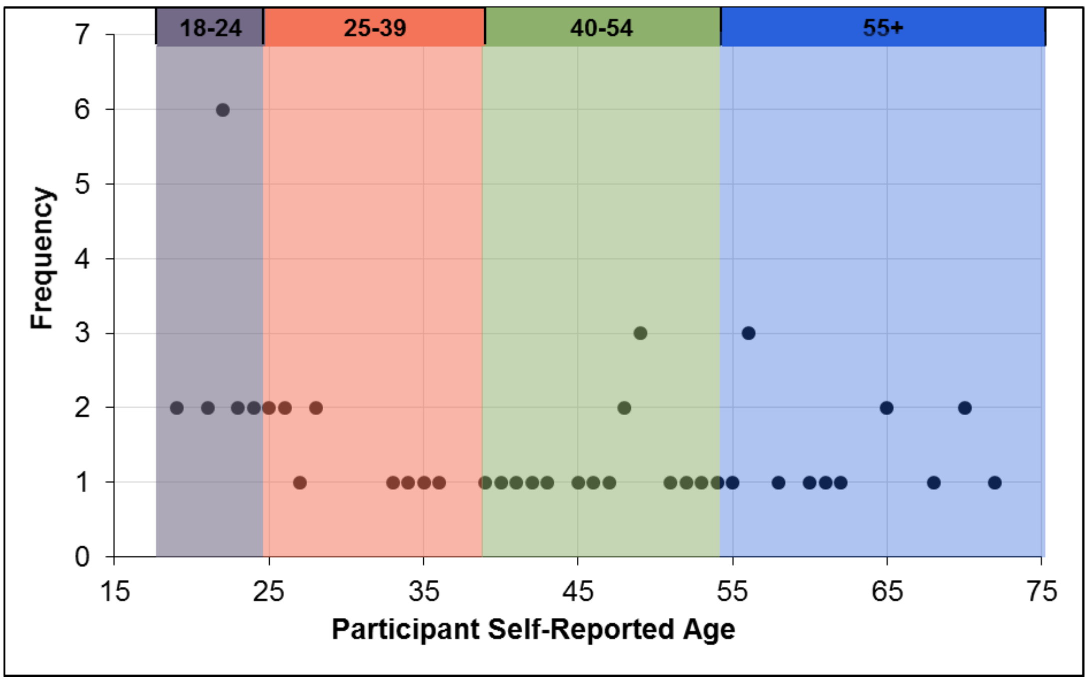



**Figure C-1. Distribution of Experiment 2 Participants Across the NHTSA Age Categories**

In addition to age, participants were asked to provide basic demographic information (e.g., gender, education level), driving experience (type of vehicles used and annualized miles driven), smartphone use and experience, and automotive technology experience. Participants were also asked about their awareness of automated vehicle technologies.

Owning a smartphone was required in order to participate in this study. This was to ensure that participants would be comfortable using the technology that would be utilized during the experiment. Participants were asked to report how long they have used a smartphone: 32.1 percent (*n* =18) indicated using a smartphone for less than 2 years; 21.4 percent (*n* = 12) indicated using a smartphone for 2 to 4 years; and 46.4 percent (*n* = 26) indicated using a smartphone for more than 4 years. These data are presented in Figure C-2 below.

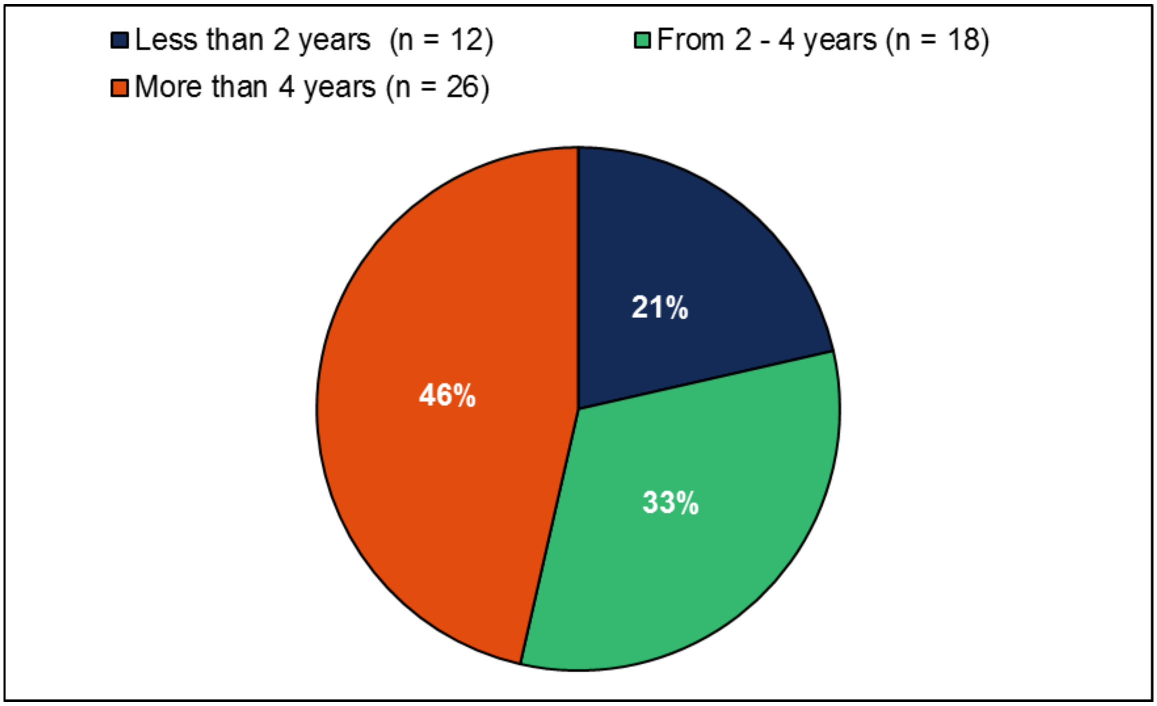



**Figure C-2. Experiment 2 Participants' Experience with Smartphones**

In addition, the participants' automotive technology experience and their familiarity with automated vehicle technologies were captured (Figure C-3). Participants were asked to report any prior experience with CWS, ACC, and LKA systems. Eleven percent of participants (*n* = 6) did report having previously driven a vehicle equipped with CWS, 3.6 percent of participants (*n* = 2) reported having previously driven a vehicle equipped with ACC, and 3.6 percent of participants (*n* = 2) reported having previously driven a vehicle equipped with an LKA system. Thirty-two percent of participants (*n* = 18) indicated being aware of self-driving/automated vehicles.

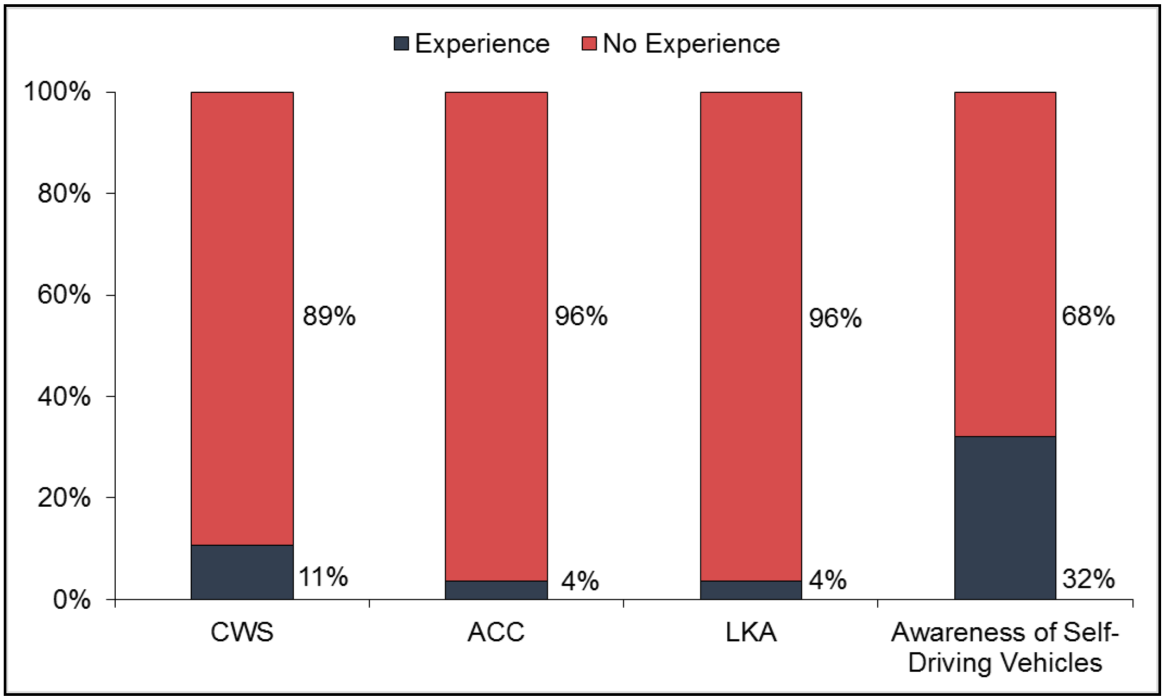



**Figure C-3. Experiment 2 Participants' Experience with and Awareness of Automated Technology for Vehicles**

# Results for Lane Drifts

The dependent variables of interest were analyzed to determine if these variables differed significantly across the different events (alert lane drift versus no alert lane drift), prompt conditions (receiving prompts 2 seconds or 7 seconds after non-attention to the roadway, or not receiving prompts at all), or between driving sessions. This section opens with an examination of operator behavior variables related to induced lane drifts, which are as follows: react to the lane drift, regain control of the vehicle, activate the automation, release control of the steering, and resume a non-driving task. Following the operator behavior variables, eye-glance behavior is analyzed via monitoring rate and the rate of non-driving-related glances per second. Following these variables related to the lane drifts, a similar set of variables are analyzed around the prompts. For operator behavior, these include time to react and time to regain control; for eye-glance, these include monitoring rate and amount of non-driving-related glances. Following these variables, trust scales are analyzed, followed by the after-experience trust scales.

For the after-experience trust scales, statistical tests were not performed because this survey came after the experiment, after all participants experienced all levels of the dependent variables. Instead, descriptive statistics were examined.

When analyzing the operator behavior variables in reaction to the lane drifts, the main independent variables of interest were event type, session, and prompt condition—and their interactions. Thus, significance was checked for these independent variables, while participant age was used as a covariate. In modeling the operator behavior variables, the log transformation was applied for each variable to account for skewed data, and analysis was performed on the log-transformed times. For test statistics, denominator degrees of freedom were adjusted using the Kenward-Roger method (Kenward & Roger, 1997). Estimates for significant differences in the least squares means (adjusted for all variables) of event type, prompt condition, or their interaction were then back-transformed in order to make an interpretation in terms of the percentage change in the operator behavior variables from one group to another. To account for correlations in the data arising from multiple measurements of participants, a linear mixed-effect model was used with random intercepts for each participant, or a covariance pattern model with event type as the repeated measures variable, was used.

Analysis was conducted in a similar manner on operator behavior in response to prompts. The difference was that prompt condition and session were considered the primary variables of interest. Prompts were randomly sampled between trust scales in each session (if there were any prompts between particular trust scales), and analysis was conducted on each of these prompts. Event type was also used in the analysis, but because these events (lane drift) happened only briefly during sessions, this was considered a secondary variable of interest, and was therefore removed if not significant. Analysis was conducted on the time to react and time to regain control. However, since few participants were required to activate the automation, this variable was analyzed descriptively.

## Operator Behavior

### *Analysis for Time to React*

The time to react was explored for each event type and each prompt condition. On average, the amount of time it took for a participant to react to the alert was 1.0 s (S.E. = 0.1 s), while the mean time it took for a participant to react to the lane drift event with no alert was 3.7 s (S.E. = 0.3 s). Thus, participants reacted almost four times as fast when they received an alert compared to when they did not receive an alert.

Meanwhile, the difference between the prompt conditions was not as strong; however, note that participants who received prompts more frequently tended to react faster (though the differences may not be statistically significant). The mean time to react for participants in the 2-second prompt condition was 2.0 s (S.E. = 0.4 s). For participants in the 7-second prompt condition, the mean time to react was 2.1 s (S.E. = 0.3 s) and for participants who did not receive any prompts, the mean time to react was 2.8 s (S.E. = 0.4 s). A mean and standard error bar plot for time to react by prompt condition is displayed in Figure C-4.

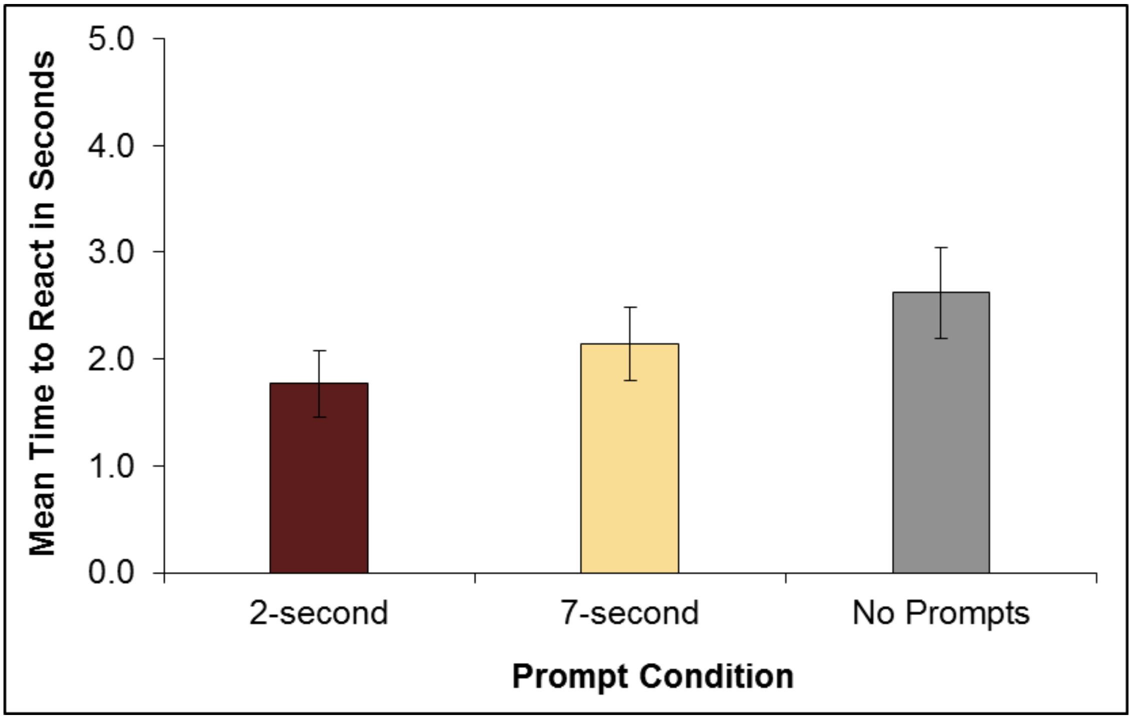



**Figure C-4. Mean and Standard Error Bar Plots, Time to React, by Prompt Condition for Experiment 2**

Participants' time to react appears to differ greatly between the two event types, but not necessarily between different prompt conditions. The highest mean within the alerted lane drifts, 1.1 s (S.E. = 0.1 s) for the No Prompts group is more than twice as fast as the fastest mean after a lane drift with no alerts (mean = 3.1 s, S.E. = 0.6 s) for participants in the 2-second condition. Within the group of instances without alerts, the reaction times increased slightly as the frequency of prompts decreased, from a mean of 3.1 s (S.E. = 0.6 s) in the 2-second group to a mean of 3.5 s (S.E. = 0.5 s) in the 7-second group to a mean of 4.5 s (S.D. = 0.7 s) in the group without prompts. This relationship is not apparent in the group of instances where an alert occurs with the lane drift. Boxplots stratifying time to react by event type and condition are displayed in Figure C-5.

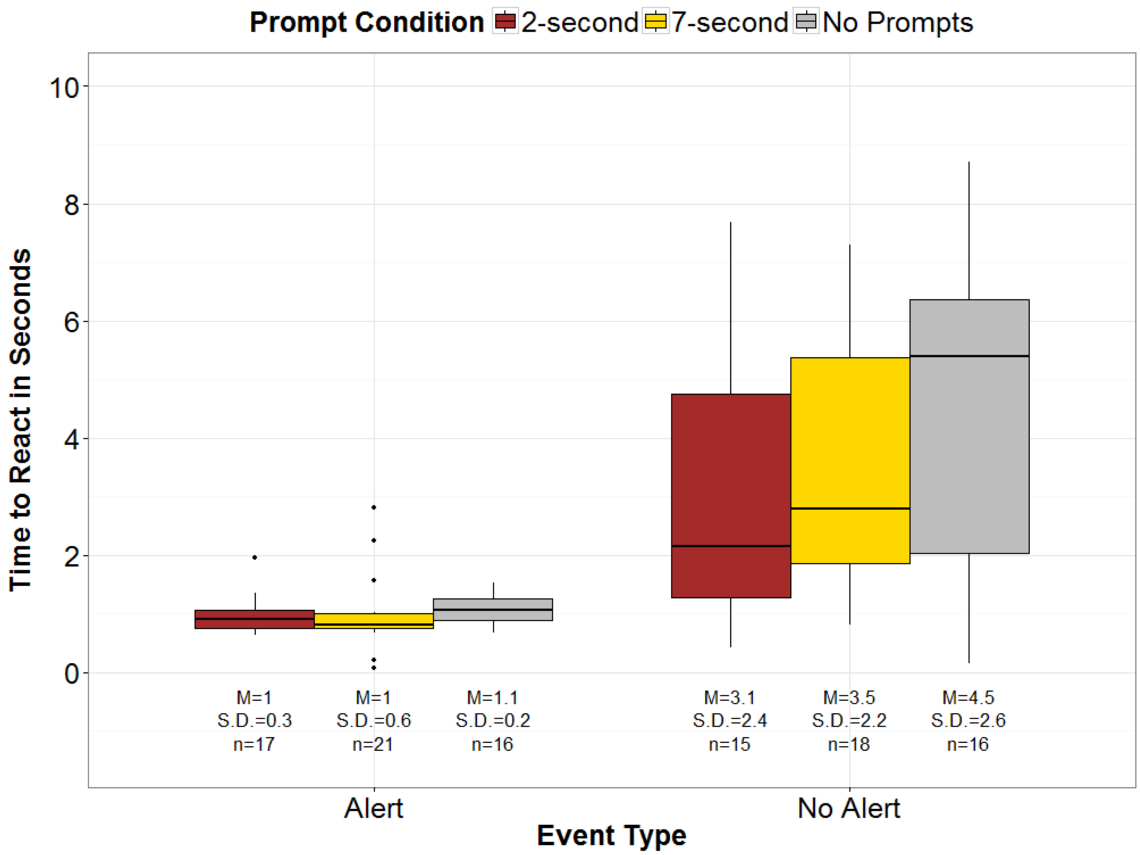



**Figure C-5. Boxplots of Time to React, Stratified by Event Type and Prompt Condition for Experiment 2**

One factor of interest that may influence reaction times is whether or not the participants are looking forward at the time of the surprise lane drift. If the participants are already looking forward, they may have adequate situational awareness of the driving situation. Hence, differing event types and prompt conditions may not make as big a difference as they would if the participants were not looking forward. It must be noted that whether or not participants were looking forward at the time of the event was determined solely by examining their eyes at the exact time stamp of the event. However, they may or may not have been consistently monitoring the roadway before the event. There were 12 instances (out of 104) in which the participants were looking forward at the time of the surprise lane drift. A bar plot of the frequencies in which participants were looking at the onset of the lane drift by event type is displayed in Figure C-6.

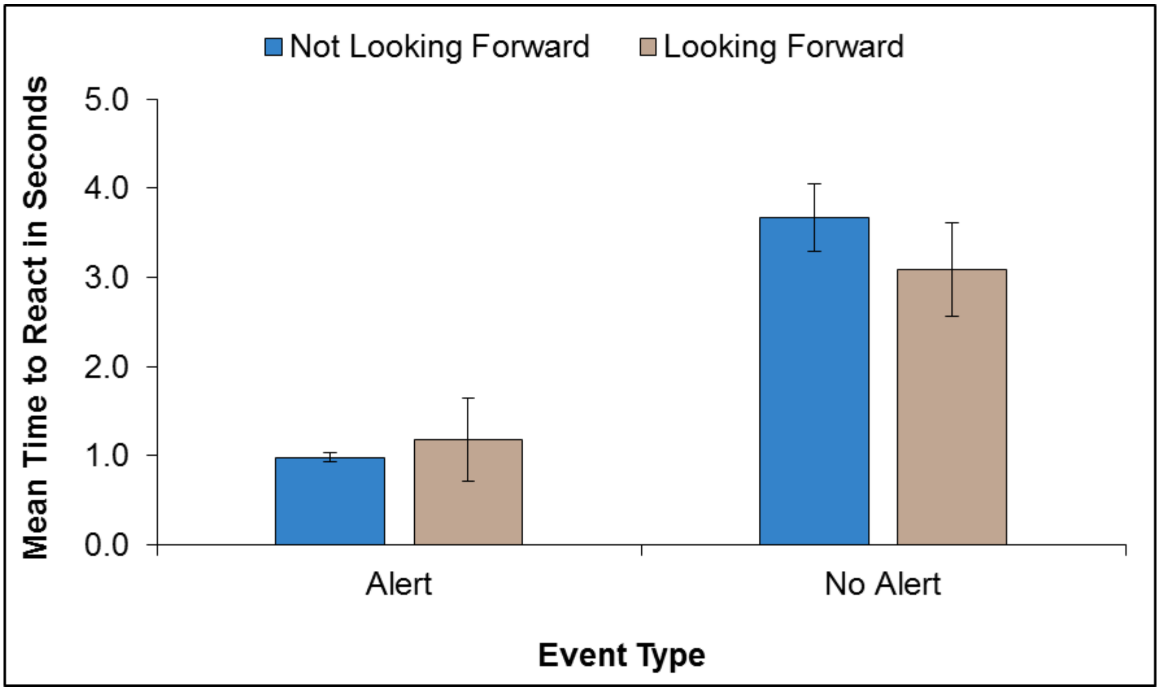



**Figure C-6. Mean and Standard Error Bar Plots, Time to React, Not Looking Forward, by Alert Type**

Statistical Analysis

The participants' reaction time to the lane drift was analyzed by applying the log transformation to reaction time. The analysis was conducted on the log of the time to react. For this model, the correlation within subjects turned out not to be significant. The best-fitting model allowed the variances within each combination of prompt condition and session to differ. Although this requires the estimation of many parameters, and therefore may reduce power, models with smaller numbers of parameters did not change the overall conclusions.

At first, a variable indicating whether or not participants were looking forward was included, as well as a variable for whether or not the participants had received a prompt within 5 s of the lane drift. However, these variables did not turn out to be significant, so they were removed. The analysis results are displayed in Table C-1.

**Table C-1. Time to React, ANOVA Table for Experiment 2**

| Effect | Num DF | Den DF | F Value | Pr > F |
|---|---|---|---|---|
| **Event Type** | **1** | **48.5** | **56.78** | **<.0001** |
| **Session** | 2 | 40.7 | 0.97 | 0.3877 |
| **Prompt Condition** | 2 | 35.2 | 0.7 | 0.5018 |
| **Event Type*Session** | 2 | 40.4 | 0.04 | 0.9575 |
| **Event Type*Prompt Condition** | 2 | 34.9 | 0.41 | 0.6658 |
| **Prompt Condition*Session** | 4 | 32.4 | 2.34 | 0.0763 |
| **Event Type*Prompt Condition*Session** | **4** | **32.5** | **4.03** | **0.0092** |

There is a significant three-way interaction between event type, prompt condition, and session (*p = .0092*), indicating that there is evidence of an interaction between event type and prompt condition that varies across sessions.

Boxplots of reaction time stratified by event type, prompt condition, and session, displayed in Figure C-7, indicate that there may be an effect of prompt condition within lane drifts without an alert within Session 1, but this does not appear as strong in the other sessions.

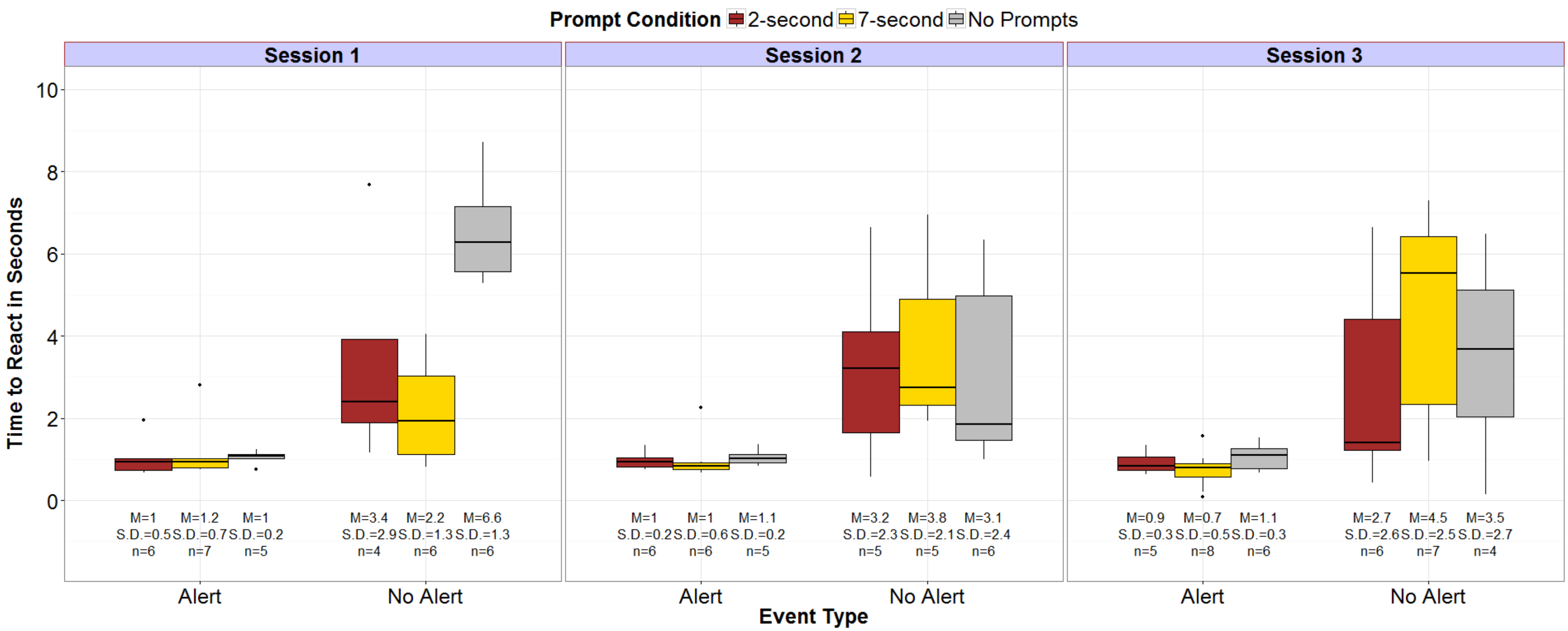



**Figure C-7. Boxplots of Time to React, Stratified by Event Type, Prompt Condition, and Session for Experiment 2**

Because the main effect of event type was highly significant ($p < 0.0001$), and because comparisons within each level of prompt condition and session will have low power because of small sample sizes and therefore may lead to spurious findings of non-significance for event type, the effects of each event type, averaged over the levels of prompt condition and session, were compared. The results are displayed in Table C-2.

**Table C-2. Comparison of Least Squares Means of the Log of Time to React Between Event Types, Averaged Across Levels of Prompt Condition and Session**

| Event Type | Comparison | Estimate | S.E. | *t* Value | *p* Value | Lower Confidence Limit | Upper Confidence Limit |
|---|---|---|---|---|---|---|---|
| Alert | No Alert | -1.05 | 0.14 | -7.54 | <.0001 | -1.33 | -0.77 |

The differences in least squares means of the log time to react between no alerted lane drifts and non-alerted lane drifts of -1.05 (95 percent confidence interval of -1.33 to -0.77) indicates that on average, the log time to react was 1.05 less after lane drifts with an alert compared to lane drifts without an alert. The estimate of -1.05 can be back-transformed to 34.9, indicating that, on average, participants reacted 65.1 percent faster after a lane drift with an alert compared to a lane drift without an alert.

To investigate the effect of prompt condition and how it may change in different sessions and event types, first, tests for a two-way interaction between prompt condition and event type were performed in all three sessions. A significant result within a particular session would imply that there is evidence that the effect of event type depends on the prompt condition. Since three comparisons were made, the Bonferroni adjusted alpha level of .017 was used. If the resulting *p* value is higher, it was deemed not significant. The results are displayed in Table C-3.

**Table C-3. Difference in Least Squares Means of the Log of Time to React Between Alerts and No Alerts**

| Session | Num DF | Den DF | F Value | Pr > F |
|---|---|---|---|---|
| **1** | **5** | **14.4** | **51.63** | **<.0001** |
| **2** | **5** | **16** | **5.42** | **0.0042** |
| **3** | 5 | 17 | 3.53 | 0.0227 |

Table C-3 implies that there is a significant interaction between prompt condition and event type in Sessions 1 and 2, but not in Session 3.

To investigate the interactions between prompt condition and event type within Session 1 and Session 2, first each relevant combination of event type and session (alert and Sessions 1 and 2, as well as no alert and Sessions 1 and 2) were inspected for an effect of prompt condition. Since four comparisons are made, the Bonferroni alpha level of .0125 was used. The results are displayed in Table C-4.

**Table C-4. Tests for Significance of Prompt Condition of the Log of Time to React Within Combinations of Sessions 1 and 2 and Event Types**

| Event Type | Session | Num DF | Den DF | F Value | Pr > F |
|---|---|---|---|---|---|
| **Alert** | **1** | 2 | 13.9 | 0.03 | 0.9662 |
| **Alert** | **2** | 2 | 16.8 | 0.04 | 0.9626 |
| **No Alert** | **1** | **2** | **12.8** | **19.02** | **0.0001** |
| **No Alert** | **2** | 2 | 16.9 | 0.65 | 0.5332 |

This table indicates that there was a significant effect of prompt condition only within lane drifts without an alert in the first session. The significant differences (using Bonferroni adjustments for multiple comparisons) between prompt conditions within Session 1 after lane drifts with no alerts are displayed in Table C-5.

**Table C-5. Comparison of Least Squares Means of the Log of Time to React Between Prompt Conditions Within Session 1, No Alert Lane Drifts**

| Prompt Condition | Comparison | Estimate | S.E. | *t* Value | Adj *p* Value | Lower Confidence Limit | Upper Confidence Limit |
|---|---|---|---|---|---|---|---|
| **2-second** | **No Prompts** | -0.88 | 0.3 | -2.95 | 0.048 | -1.55 | -0.21 |
| **7-second** | **No Prompts** | -1.41 | 0.24 | -5.83 | <0.0001 | -1.95 | -0.89 |

This table indicates that in Session 1, after lane drifts in which no alert occurred, participants that were experiencing prompts (2-second or 7-second) reacted significantly faster to the lane drift than those who were not receiving prompts. However, there was no statistically significant difference between the two prompt conditions. When comparing the 2-second prompt condition to the No Prompts condition, for example, the estimate of -0.88 (95 percent confidence interval -1.55 to -0.21) indicates that, on average, the least squares mean of the log of the time to react was 0.88 lower in the 2-second group compared to the No Prompts group. Back transforming the estimate of -0.88 to 0.41 indicates that, on average, within Session 1 and after lane drifts without an alert, those in the 2-second prompt condition reacted, on average, 59 percent faster than did those who received no prompts. A similar interpretation can be made of the estimated difference between the 7-second condition and no prompts, which was -1.41 log seconds (95 percent confidence interval -1.95 to -0.89). This can be back-transformed to 0.24, indicating that the 7-second group reacted, on average, 76 percent faster than the No Prompts group.

In summary, there is evidence that reaction time to a lane drift will be faster when vehicle operators receive a message alerting them to the lane drift, compared to when they do not receive an alert. (Note that this is an average effect of alert across the different sessions and prompt conditions.) However, statistically significant evidence of an effect of prompt condition is confined to participants experiencing a lane drift without an alert early on in the study (Session 1), in which case participants who received prompts reacted significantly faster than participants who did not receive prompts.

***Analysis for Time to Regain Control***

When examining time to regain control, it must first be noted that in a large number of instances (almost exclusively after lane drifts with no alerts), participants failed to regain control at all. In these instances, the in-vehicle experimenter instructed the participant to regain manual control of the vehicle. Out of 49 instances of lane drifts with no alerts, participants failed to regain control during 23 of these instances (46.9 percent). Meanwhile, after 55 instances of lane drifts with alerts, the participant failed to regain control after only 1 instance (1.8 percent). This one instance occurred for a participant that did not receive prompts, and the instance occurred during Session 1. Based on these data, the relative risk of failing to regain control without instruction from the experimenter was 26.1, which means that the participants were 26.1 times as likely to fail to regain control without instruction from the experimenter during a lane drift with no alert compared to during a lane drift with an alert. A bar plot of the frequencies of failing to regain control, stratified by event type is displayed in Figure 3-7.

Although there were more incorrect performances from participants without prompts in Session 1, this did not hold in Session 2 or Session 3. In Session 1, five of six participants who did not experience prompts failed to regain control without instruction from the experimenter, compared to two of four from the 2-second prompt condition group and five of six from the 7-second prompt condition group. However, 50 percent of the remaining participants in the No Prompts group ($n$ = 8) correctly performed, as compared with 50 percent ($n$ = 12) in the 7-second prompt condition group, and 64 percent ($n$ = 11) in the 2-second prompt condition group. Stacked bar charts are displayed in Figure C-8 and Figure C-9.

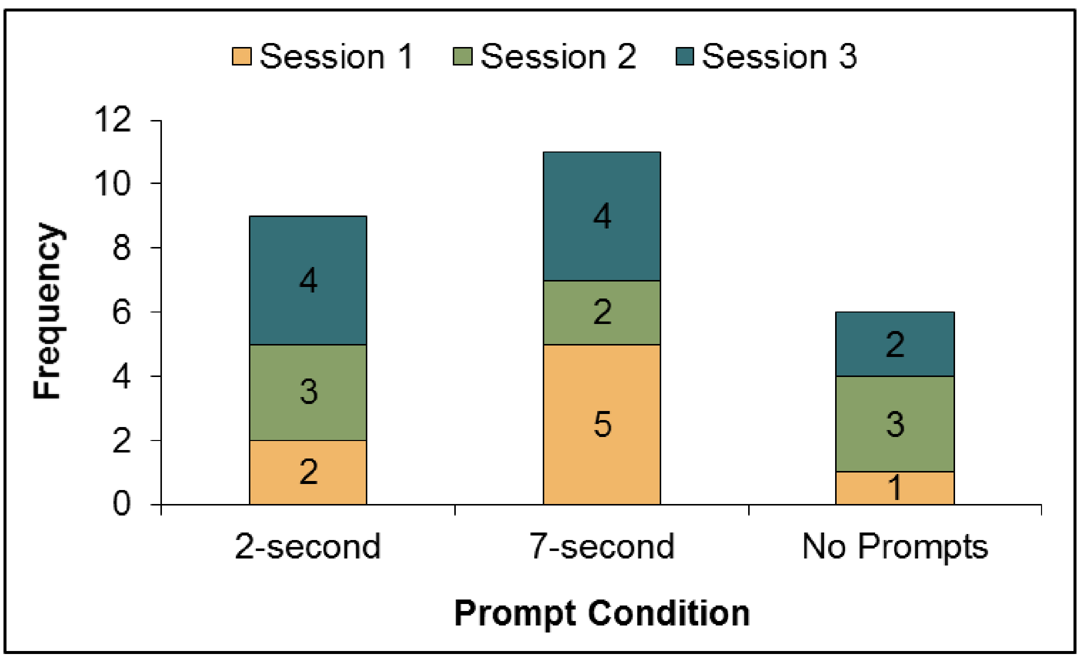



**Figure C-8. Stacked Bar Plot for Correct Responses (Regained Control Without Instruction) Stratified by Session and Prompt Condition**

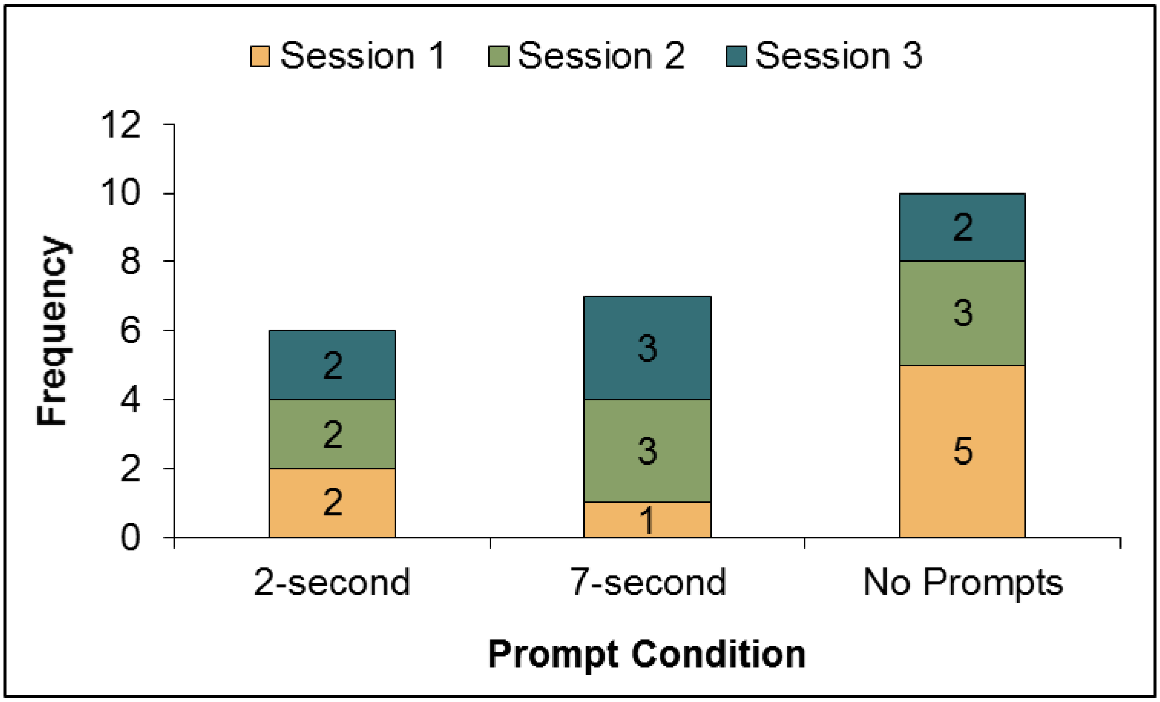



**Figure C-9. Stacked Bar Plot for Incorrect Responses (Did Not Regain Control Without Instruction from the Experimenter) Stratified by Session and Prompt Condition**

Since almost everyone who experienced a lane drift with an alert regained control on their own, the time to regain control can be examined within the lane drifts with alerts (excluding the instance in which the participant did not regain control without instruction from the experimenter) to determine if periodically receiving prompts additionally reduced the time in which participants regained control. However, the low sample size in this group makes detecting significant differences between these groups difficult. Figure C-10, which displays boxplots of the time to regain control within this group, reveals that the group with the 2-second prompt condition regained control more quickly, on average, than the other two groups (though this difference may not be statistically significant). The 2-second condition had the two lowest means across sessions, with a mean of 1.7 s in Session 1 (S.D. = 0.4), 2.5 s in Session 2 (S.D. = 0.6), and 1.8 s in Session 3 (S.D. = 0.4) (Figure C-11). There were some outlying points in the 7-second condition, including a 9.5-second regain control time in Session 1 and a 7.7-second regain control time in Session 2, but these were determined to be accurate points. The outlier in Session 1 drove this group of regain control times to the highest mean of any of the groups within the alert event type, with a mean of 3.1 s (S.D. = 3.0). In summary, although the 2-second group had more consistently lower times to regain control within the alert event type group, a low sample size may make detecting significant differences difficult.

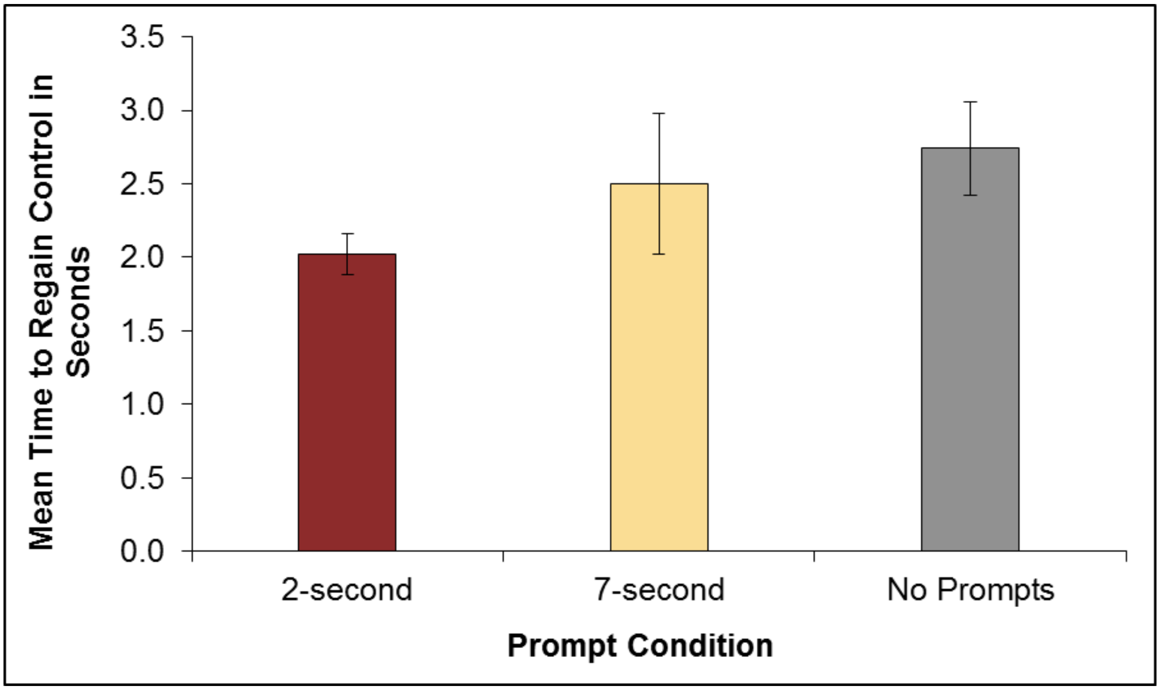



**Figure C-10. Mean and Standard Error Bar Plots, Time to Regain Control During Alerts, by Prompt Condition for Experiment 2**

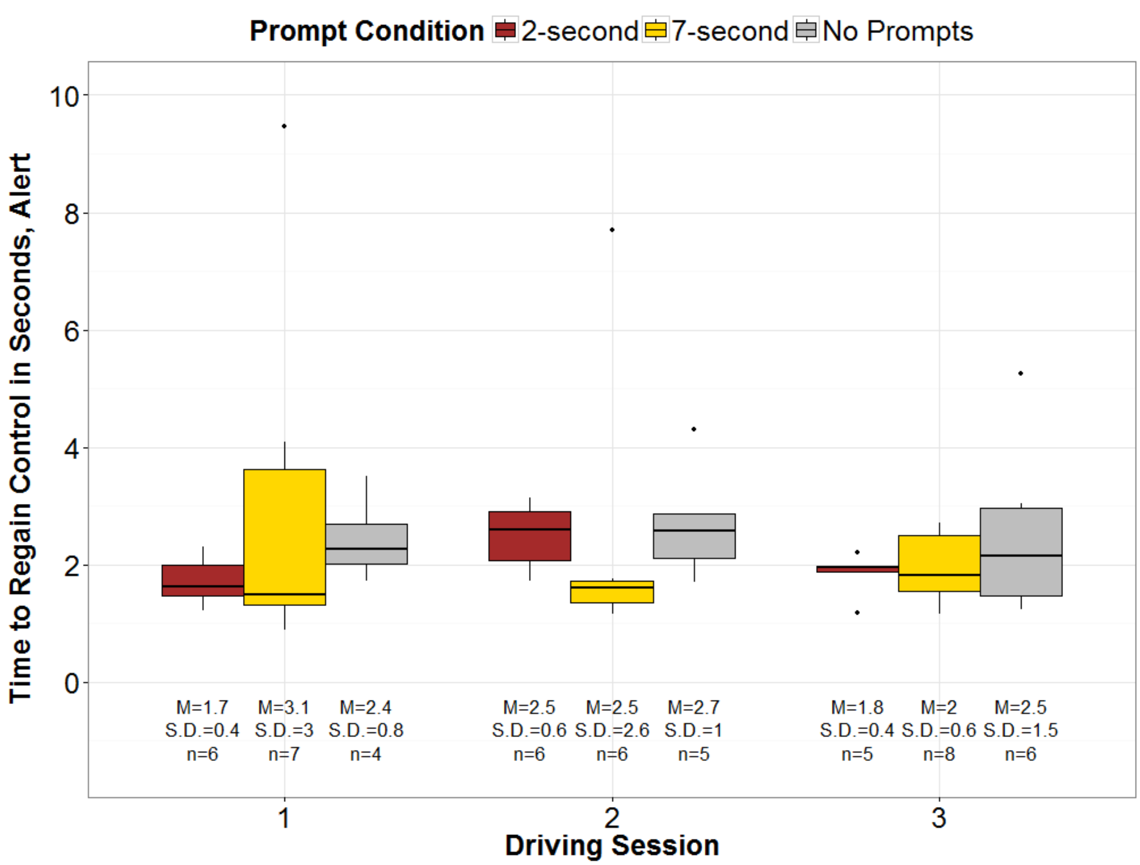


**Figure C-11. Boxplots of Time to Regain Control During the Alert Event Type, Stratified by Driving Session and Prompt Condition for Experiment 2**

Statistical Analysis

A logistic regression model with Firth's correction for quasi-separation was performed to determine if the probability of regaining control differed significantly between the variables (Firth, 1993). If there is an effect of condition and/or session within the lane drifts with no alert, there would be a significant interaction between event type and these variables, since there will not be a significant effect of these variables within lane drifts with alerts (because only one participant failed to regain control during an alert). The analysis is displayed below in Table C-6.

**Table C-6. Probability of Regaining Control Without Instruction for Experiment 2**

| Effect | DF | Chi-Square | Pr > ChiSq |
|---|---|---|---|
| **Event Type** | **1** | **15.2224** | **<0.0001** |
| **Session** | 2 | 1.4288 | 0.4895 |
| **Prompt Condition** | 2 | 1.6991 | 0.4276 |
| **Event Type*Prompt Condition** | 2 | 0.2387 | 0.8875 |
| **Session*Prompt Condition** | 4 | 2.1513 | 0.7079 |
| **Event Type*Session** | 2 | 0.1387 | 0.933 |
| **Event Type*Session* Prompt Condition** | 4 | 0.3854 | 0.9837 |

The difference between the event types is statistically significant, indicating that a participant who experienced a lane drift with an alert was significantly more likely to regain control without instruction from the experimenter compared to those who experienced a lane drift without an alert. However, neither condition nor session were significant, indicating that there is no statistically significant evidence of an effect of either time or prompt frequency on how likely the participant was to regain control without instruction from the experimenter.

A separate analysis was conducted on time to regain control for lane drifts with alerts to determine if, given that a participant has received an alert with the lane drift, there is a further difference made by prompt condition, or if there is a change over time in how quickly participants regain control. The analysis was conducted on the log of the regain control time. The analysis results are displayed in Table C-7.

**Table C-7. Analysis for the Log of Time to Regain Control, ANOVA Table**

| Effect | Num DF | Den DF | F Value | Pr > F |
|---|---|---|---|---|
| **Prompt Condition** | 2 | 25 | 1.13 | 0.3376 |
| **Session** | 2 | 35.3 | 0.61 | 0.5485 |
| **Prompt Condition*Session** | 4 | 27.2 | 0.61 | 0.6606 |

Neither condition nor session is significant at the 0.05 level. Therefore, the conclusion is that, given that the lane drift has been preceded by an alert, the speed with which operators regain control after the lane drift may not significantly increase with the addition of prompts.

In summary, there is evidence that vehicle operators may regain control of the vehicle after an unanticipated event more quickly, and more consistently, if the event is preceded by an alert as compared to no alert, and this is consistent over time. However, receiving prompts does not appear to offer statistically significant additional benefit in terms of reducing how long operators take to regain control after an unexpected lane drift.

***Analysis of Time to Activate the Automation***

Participants took longer, on average, to activate the automation following an alert than they did following no alert. The mean time to activate the automation was 7.3 s (S.E. = 0.5 s) after a lane drift with an alert, compared to a mean of 5.0 s (S.E. = 0.6 s) when there was no alert. Similarly, the time to activate the automation was lower if the participant was receiving prompts in the 2-second condition (mean = 5.2 s, S.E. = 0.7 s) than it was for both the 7-second group (mean = 7.1 s, S.E. = 0.7 s) and the No Prompts group (mean = 6.1 s, S.E. = 0.6 s). Mean and error bar plots are displayed for time to activate the automation stratified by event type in Figure C-12 and by prompt condition in Figure C-13.

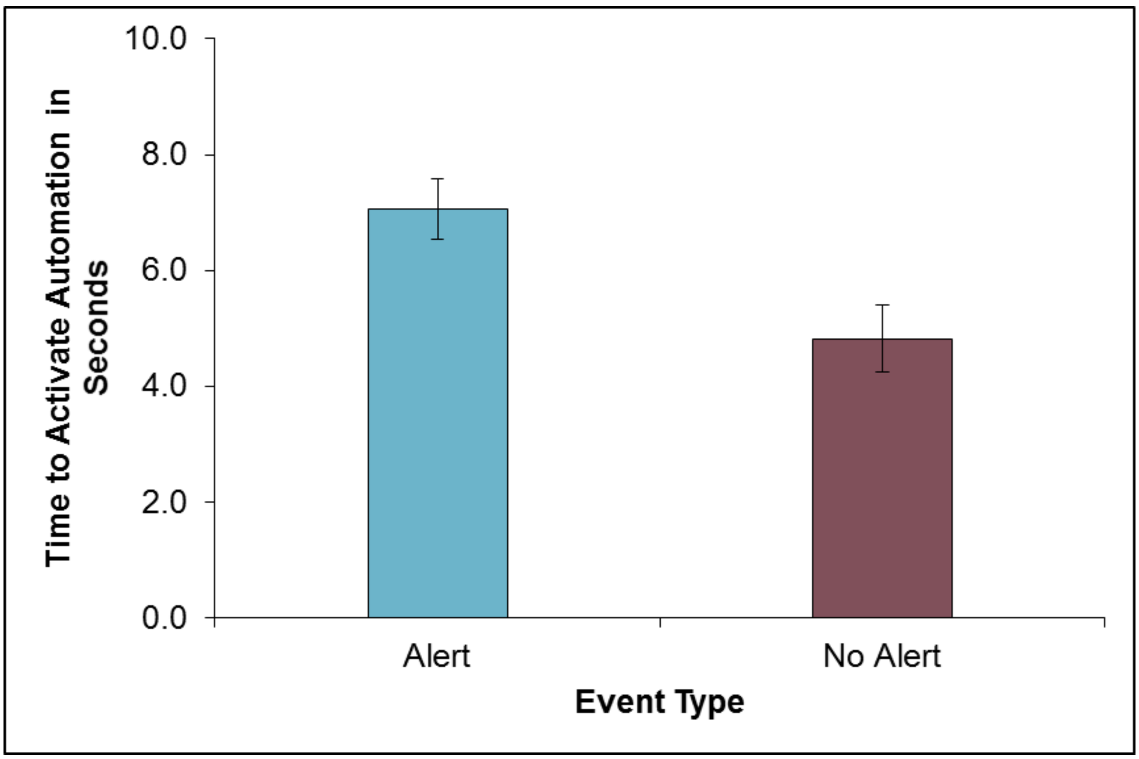



**Figure C-12. Mean and Standard Error Bar Plots, Time to Activate Automation, by Alert Type for Experiment 2**

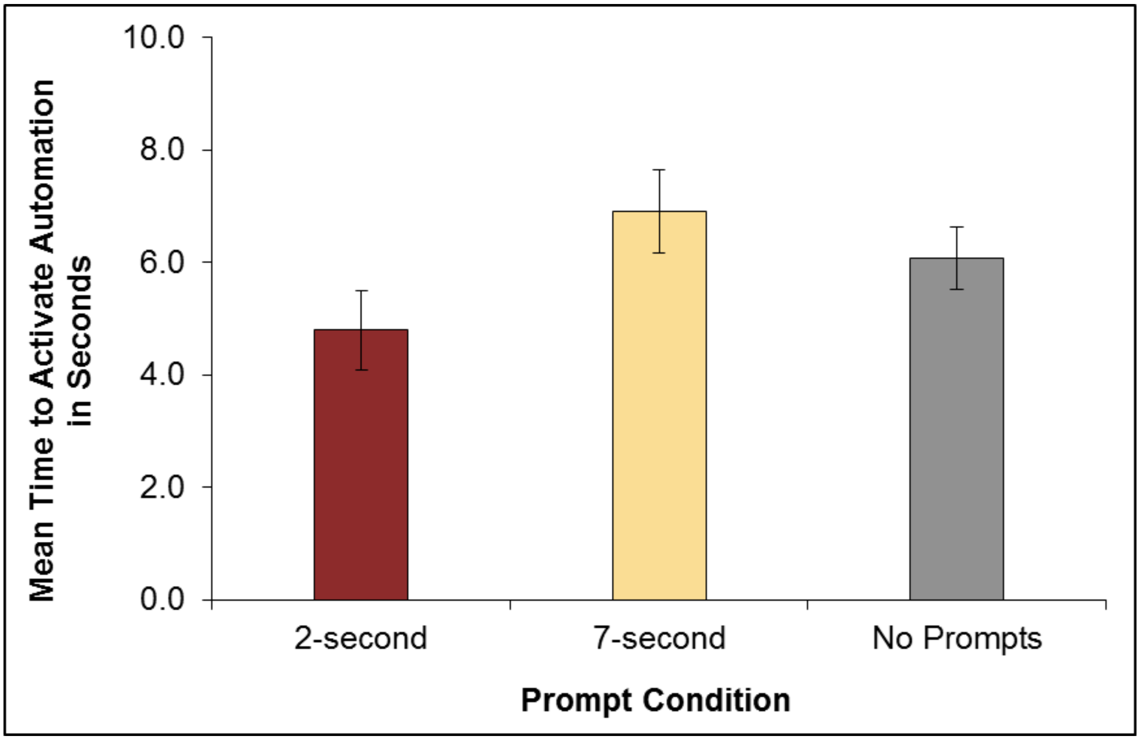



**Figure C-13. Mean and Standard Error Bar Plots, Time to Activate Automation, by Prompt Condition for Experiment 2**

The boxplots for time to activate the automation stratified by event type and prompt condition, as displayed in Figure C-14, indicate that the times to activate the automation were higher for participants when they experienced an alert with the lane drift compared to when they did not, regardless of the prompt condition. Within all three prompt conditions, the mean time to activate the automation was higher after lane drifts with alerts than after lane drifts without alerts; for the 2-second prompt condition, the mean was 5.5 s (S.D. = 4.2 s) after lane drifts with alerts compared to a mean of 3.8 s (S.D. = 3.6 s) for lane drifts without alerts (Figure C-14). The change is similar within 7-second prompt conditions (mean = 8.3 s, S.D. = 3.6 s for alerts versus mean = 5.6 s, S.D. = 5.1 s for no alerts) and No Prompts (mean = 7.2 s, S.D. = 3.0 s for alerts versus mean = 4.9 s, S.D. = 3.0 s for no alerts). The highest time to activate the automation was 22.0 s, after a lane drift without an alert with a 7-second prompt condition in Session 3. In a total of 18 instances, participants took at least 10.0 s to activate the automation, 11 of which came after lane drifts with alerts.

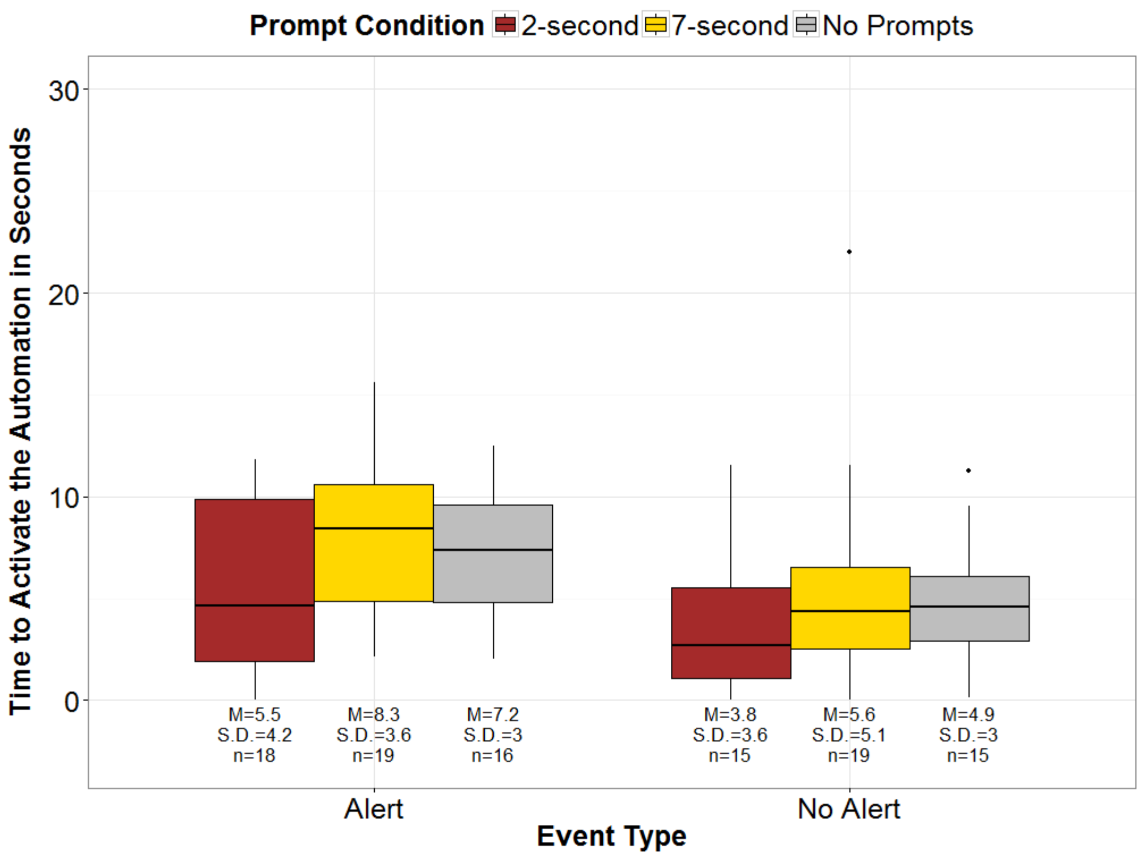



**Figure C-14. Boxplots of Time to Activate the Automation, Stratified by Event Type and Prompt Condition for Experiment 2**

<u>Statistical Analysis</u>

The analysis of the time to activate the automation was performed on the log-transformed time. The analysis results are displayed in Table C-8.

**Table C-8. Analysis for the Log of Time to Activate the Automation, ANOVA Table for Experiment 2**

| Effect | Num DF | Den DF | F Value | Pr > F |
|---|---|---|---|---|
| **Event Type** | **1** | **39.5** | **11.64** | **0.0015** |
| **Prompt Condition** | 2 | 38.3 | 2.57 | 0.0893 |
| **Event Type* Prompt Condition** | 2 | 39.6 | 0.27 | 0.7613 |
| **Session** | 2 | 57.7 | 1.50 | 0.2321 |
| **Prompt Condition*Session** | 4 | 57.5 | 0.44 | 0.7792 |
| **Event Type*Session** | **2** | **56.2** | **4.78** | **0.0121** |
| **Event Type*Prompt Condition*Session** | 4 | 56.4 | 1.04 | 0.3948 |

The interaction between the event type and session is significant (*p* = 0.0121), indicating that the effect of the event type on the time to activate the automation changes over time. Prompt condition is not significant, indicating that there is no statistically significant evidence that receiving or not receiving prompts affects the time to activate the automation.

To investigate the interaction between event type and session, the effect of event type within each session, and the effect of session within each event type, were investigated for significance. Least squares means for the log of the time to activate the automation by event type and session are displayed in Figure C-15. The significance level for these comparisons is the Bonferroni adjusted level of 0.01. The results are displayed in Table C-9.

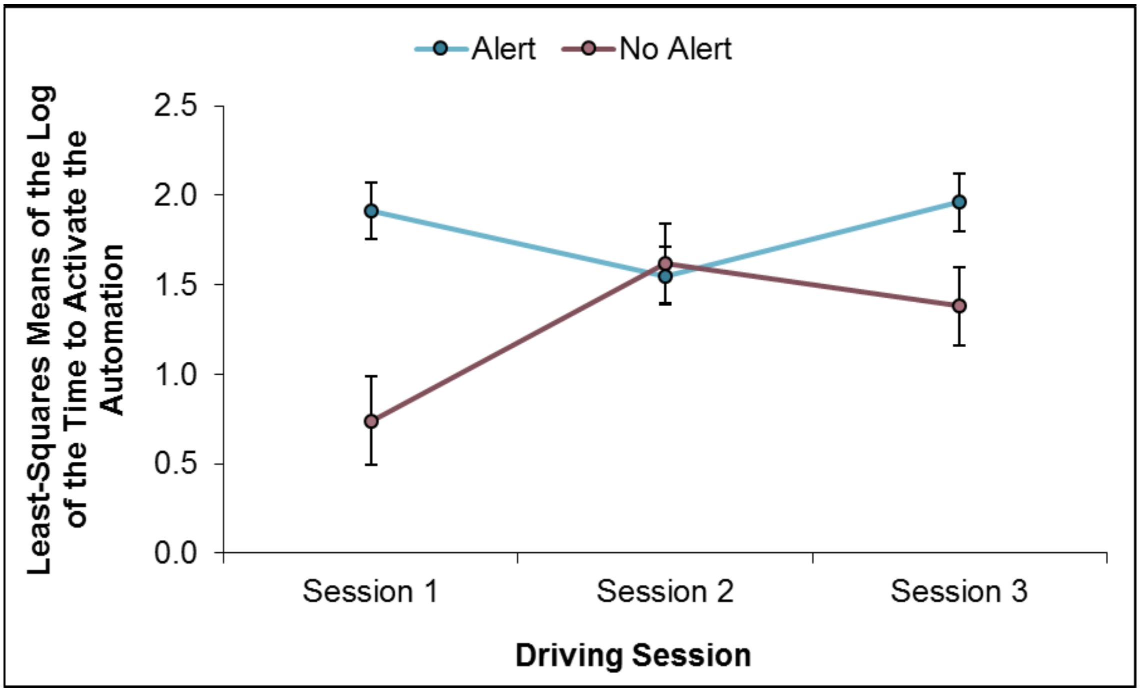



**Figure C-15. Plot of Least Squares Means for Time to Activate the Automation for Experiment 2**

Table C-9 indicates that within Session 1, there is a significant difference in the time to activate the automation depending upon whether the alert is present versus absent. However, after this point, the difference is no longer statistically significant. The difference in the least squares mean of the log of the time to activate the automation between the lane drift with an alert and the lane drift without an alert is displayed in Table C-10.

**Table C-9. Significant Differences of Time to Activate the Automation Within Session and Event Type for Experiment 2**

| Event Type | Session | Num DF | Den DF | F Value | Pr > F |
|---|---|---|---|---|---|
| | **1** | **1** | **64.5** | **15.72** | **0.0002** |
| | **2** | 1 | 70 | 0.1 | 0.7501 |
| | **3** | 1 | 69.4 | 4.17 | 0.0449 |
| **Alert** | | 2 | 41.2 | 1.72 | 0.1910 |
| **No Alert** | | 2 | 37.3 | 3.72 | 0.0336 |

**Table C-10. Difference in the Least Squares Means of the Log of Time to Activate the Automation Between Alert and No Alert, Within Session 1**

| Estimate | Standard Error | *t* Value | *p* Value | Lower Confidence Limit | Upper Confidence Limit |
|---|---|---|---|---|---|
| 1.17 | 0.30 | 3.96 | 0.0002 | 0.58 | 1.77 |

The estimate of the difference in the log of the time to activate the automation is 1.17 (95 percent confidence interval of 0.58 to 1.77). This estimate is back-transformed to 3.22, indicating that within Session 1, participants that experienced an alert at the time of the lane drift would activate the automation an estimated 222 percent more slowly than participants that experienced no alert at the time of the lane drift.

In summary, the experiment provides evidence that, if vehicle operators experience an unanticipated event early on in a driving session and thus need to regain control of the steering, they may activate the automation more slowly if the event was accompanied by an alert compared to if the event was not accompanied by an alert. However, as time passes, the difference may not be sustained.

***Analysis of Time to Release Control of Steering***

In contrast to the finding for time to activate the automation, Figure C-16, which displays the means of time to release control in seconds stratified by event type, indicates that the participants took longer to release control without the alert than they did with the alert. The mean time to release control after a lane drift without an alert was 4.2 s (S.E. = 0.7 s), compared to the mean of 3.3 s (S.E. = 0.5 s) with an alert. Regarding prompt condition, as the frequency of the prompts decreased, the time to release control decreased in the participants. The mean decreases from 4.7 s (S.E. = 1.1 s) in the 2-second group to 3.4 s (S.E. = 0.5 s) in the 7-second group, and finally down to 3.3 s (S.E. = 0.6 s) in the No Prompts group. Mean and error bar plots are displayed for time to activate the automation stratified by prompt condition in Figure C-17.

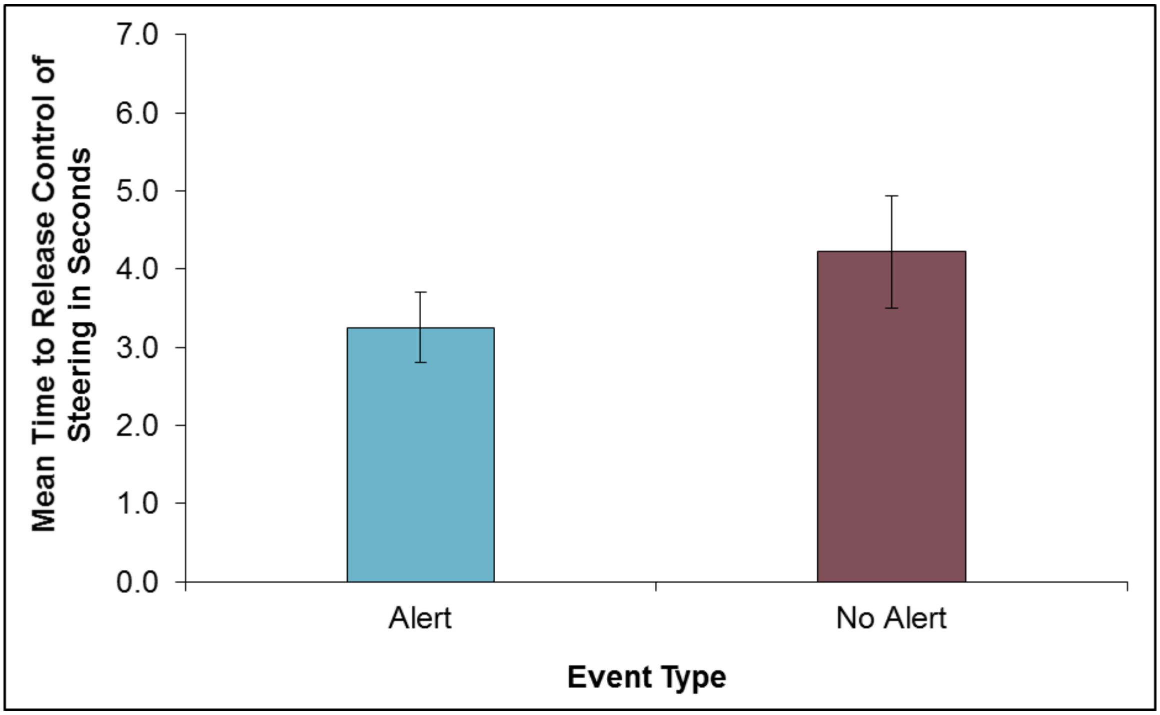



**Figure C-16. Mean and Standard Error Bar Plots, Time to Release Control of Steering by Event Type for Experiment 2**

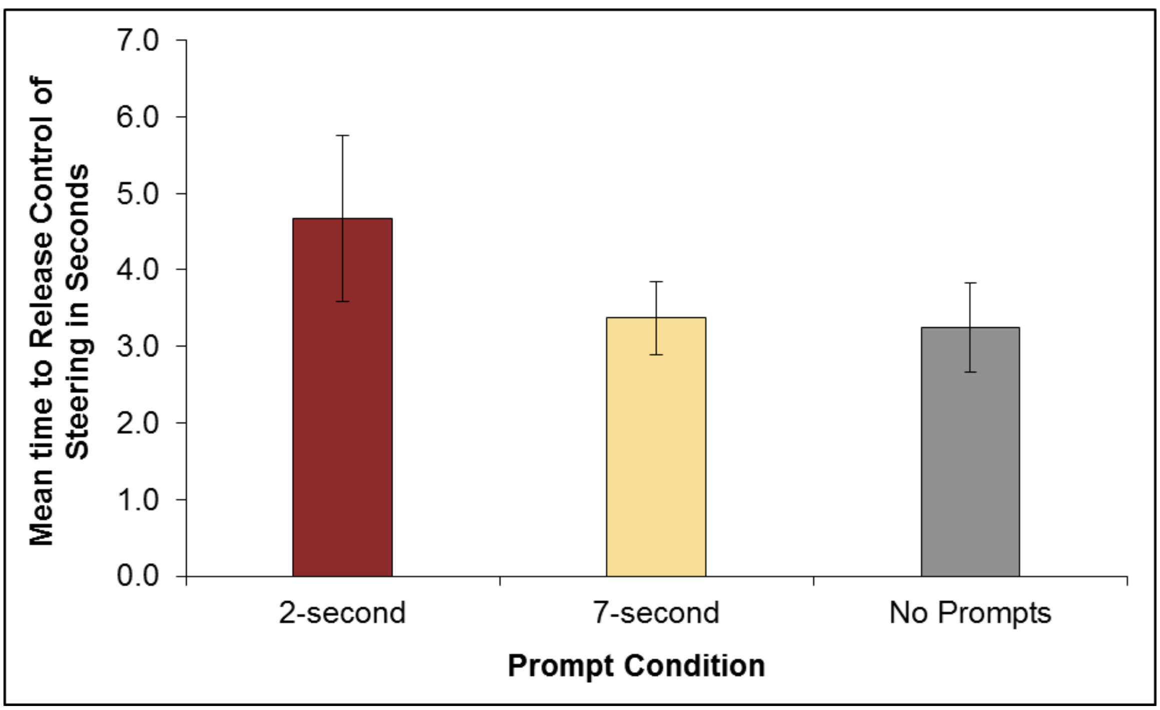



**Figure C-17. Mean and Standard Error Bar Plots, Time to Release Control of Steering by Prompt Condition for Experiment 2**

The boxplots of time to release control, stratified by event type and prompt condition, displayed in Figure C-18, indicate stronger variability in times to release control within the 2-second condition after lane drifts with no alerts than in any other group. The standard deviation in this group was 7.5 s, with a mean of 6.1 s. The next highest standard deviation was 3.7 s for the 2-second condition group after experiencing lane drifts with alerts. The mean of the 2-second lane drift with no alert group was due to this group containing the two longest times to release control. These two longer times were 25.0 s during Session 1 and 18.7 s during Session 2. In a total of seven instances, the time to release control was greater than 10.0 s, with five of these instances following lane drifts with no alerts.

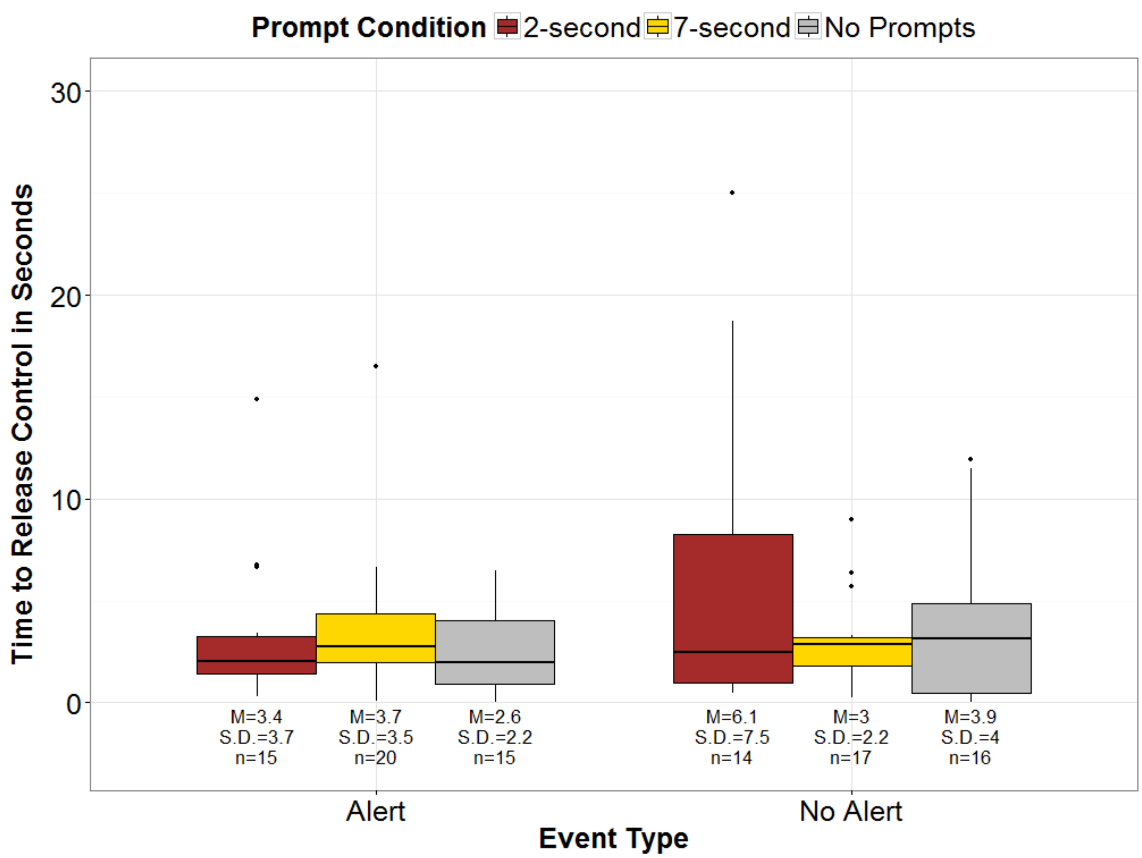



**Figure C-18. Boxplots of Time to Release Control of Steering, Stratified by Event Type and Prompt Condition for Experiment 2**

<u>Statistical Analysis</u>

The analysis for the time to release control of the steering was performed on the log-transformed data. The analysis results are displayed in Table C-11.

**Table C-11. Time to Release Control of Steering, Tests for Significance for Experiment 2**

| Effect | Num DF | Den DF | F Value | Pr > F |
|---|---|---|---|---|
| **Event Type** | 1 | 44.1 | 0.76 | 0.3880 |
| **Prompt Condition** | 2 | 47.2 | 1.45 | 0.2440 |
| **Event Type* Prompt Condition** | 2 | 44.2 | 1.00 | 0.3773 |
| **Session** | 2 | 57.4 | 0.12 | 0.8853 |
| **Prompt Condition*Session** | 4 | 57.1 | 2.46 | 0.0552 |
| **Event Type*Session** | 2 | 74.7 | 1.43 | 0.2448 |
| **Event Type*Prompt Condition*Session** | 4 | 74.4 | 0.49 | 0.7415 |

None of the variables are significant at the 0.05 level. Therefore, there is no statistically significant evidence of an effect of either event type, prompt condition, or session on the time to release control after a lane drift.

***Analysis of Time to Resume a Non-driving Task***

Figure C-19, Figure C-20, and Figure C-21 display mean and error bar plots, and then boxplots, of time to resume a non-driving task across event types and prompt conditions. The boxplots reveal large variability in the times to resume a non-driving task. Over half of instances resulted in participants resuming a non-driving task within 6.2 s (the lowest amount of time it took for a participant to resume a non-driving task was 0.005 s, which was essentially an instantaneous resuming of a non-driving task upon releasing control of the steering). On the other hand, nine participants took at least 20.0 s before resuming a non-driving task, with the highest time being 56.6 s, which occurred for a participant within the 2-second condition after experiencing no alert in Session 3. The second highest time to resume a non-driving task was 54.7 s, which occurred in Session 1 for a participant in the 7-second prompt condition who experienced no alert, while the third highest time, 44.6 s, occurred after an alert within Session 3 for a participant in the 7-second prompt condition.

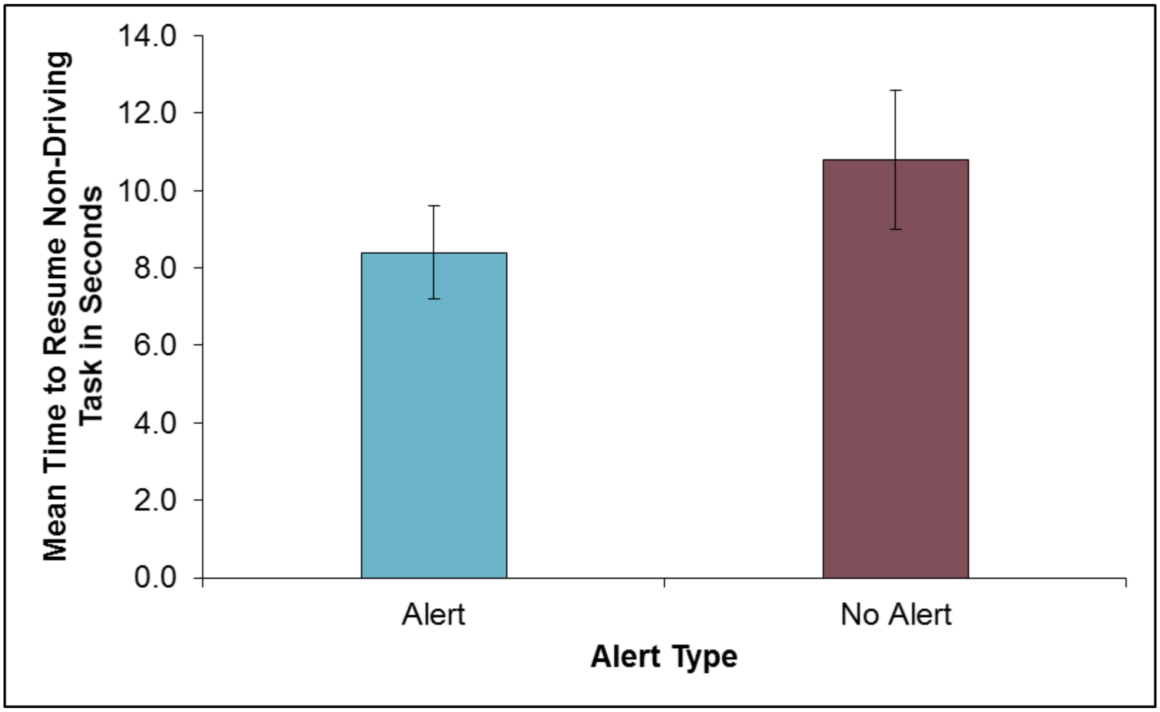



**Figure C-19. Mean and Standard Error Bar Plots, Time to Resume Non-driving Task by Event Type for Experiment 2**

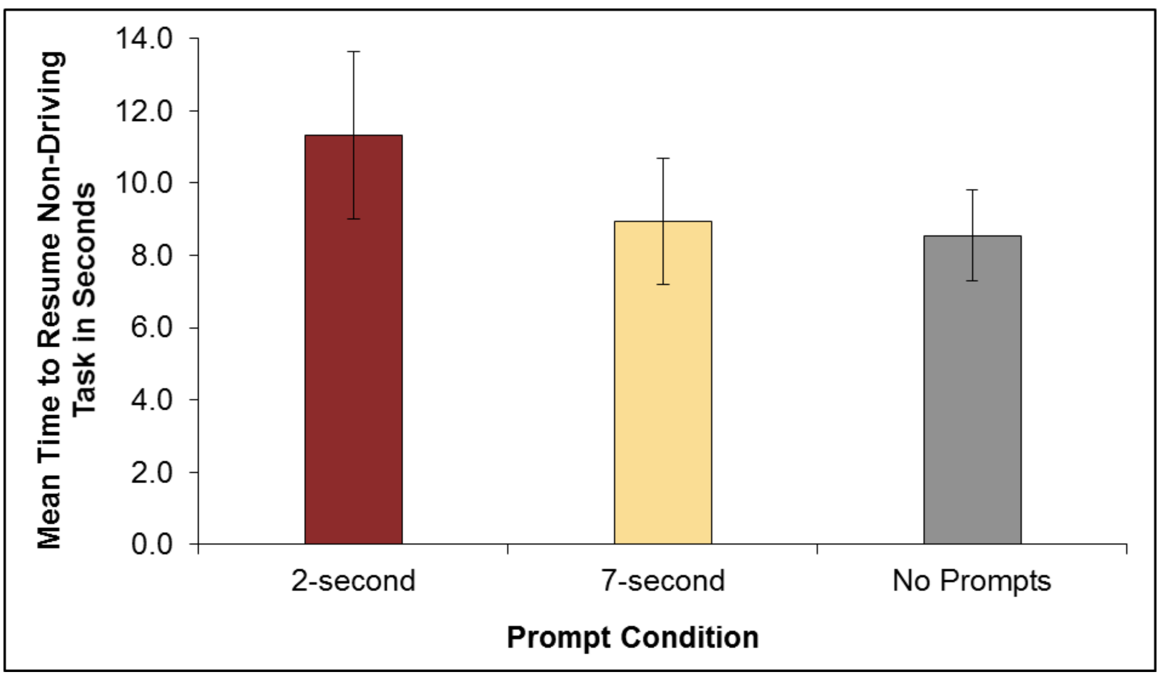



**Figure C-20. Mean and Standard Error Bar Plots, Time to Resume Non-driving Task by Prompt Condition for Experiment 2**

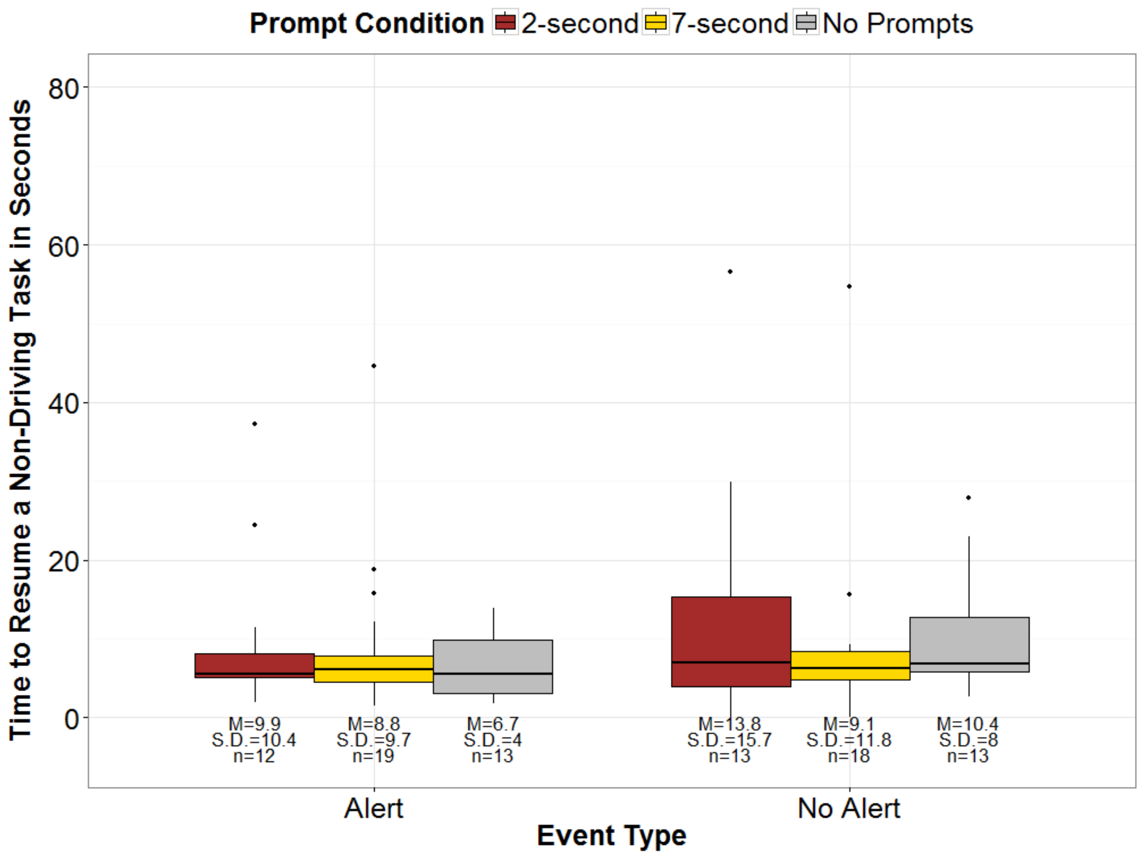



**Figure C-21. Boxplots of Time to Resume a Non-driving Task Stratified by Event Type and Prompt Condition for Experiment 2**

Statistical Analysis

The analysis for the time to resume a non-driving task was performed on the log-transformed data. The analysis results are displayed below in Table C-12.

**Table C-12. Analysis of Time to Resume a Non-driving Task, ANOVA Table, For Experiment 2**

| Effect | Num DF | Den DF | F Value | Pr > F |
|---|---|---|---|---|
| **Event Type** | 1 | 35.5 | 0.01 | 0.9291 |
| **Prompt Condition** | 2 | 35.3 | 0.40 | 0.6720 |
| **Event Type*Prompt Condition** | 2 | 35.5 | 0.39 | 0.6768 |
| **Session** | 2 | 41.4 | 0.71 | 0.4976 |
| **Prompt Condition*Session** | 4 | 41.4 | 0.60 | 0.6613 |
| **Event Type*Session** | 2 | 44.8 | 0.07 | 0.9348 |
| **Event Type*Prompt Condition*Session** | 4 | 45.0 | 1.12 | 0.3607 |

None of the variables are significant at the 0.05 level. Therefore, there is no statistically significant evidence of an effect of either event type, prompt condition, or session on the time to resume a non-driving task after an unanticipated event.

## Eye-glance Behavior

Eye-glance behavior comprises two different variables: monitoring rate and non-driving-related glances. Monitoring rate refers to the percentage of driving-related glance time in a given time interval. Non-driving-related glances refers to the number of glances operators make that do not relate to the driving task.

Eye-glance data were analyzed from 10 s before the lane drift to whenever the participant regained control or was instructed to regain control. For the prompts, eye-glance behavior was analyzed from 15 s before to 15 s after (if the participant received prompts). If the participant did not receive prompts, time points were randomly selected between those trust scales, and eye-glance behavior was analyzed from 15 s before to 15 s after the randomly selected time point. Monitoring rate was analyzed (for both lane drifts and prompts) using a beta regression model (to account for the fact that monitoring rate is bounded below by 0 and above by 1) with working covariance (estimated from the data) used to account for correlations from repeated measures from the same participants. A Generalized Estimating Equation Poisson model with working covariance (estimated from the data) was used to model the amount of non-driving-related glances in these time intervals. Figure C-22 illustrates the data calculation intervals for all event types and prompt conditions.

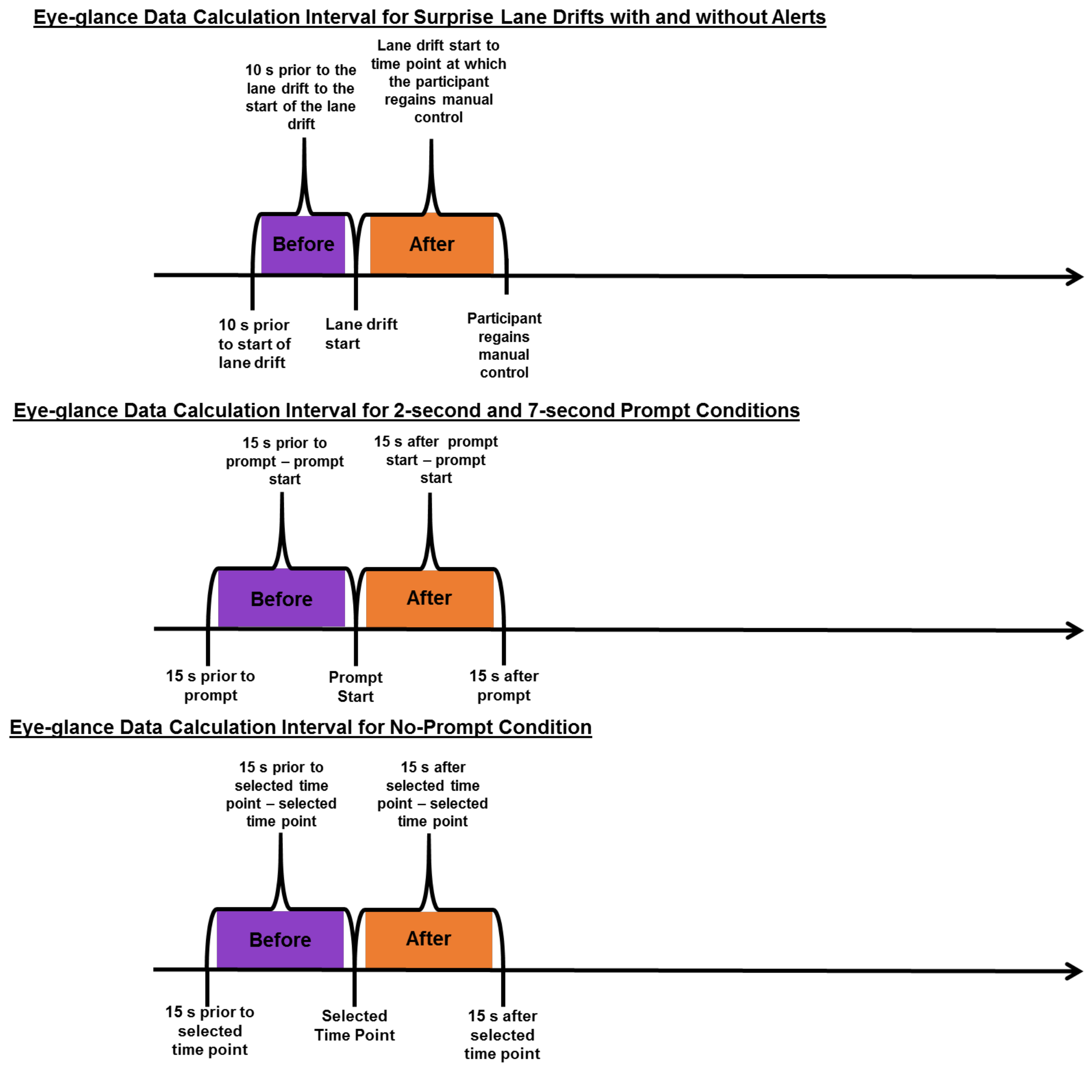



**Figure C-22. Eye-glance Data Calculation Periods for Experiment 2**

***Analysis of Monitoring Rate***

Figure C-23 indicates that the participants had a higher percentage of driving-related glance time immediately after an alert as compared to immediately after no alert. This is expected given that the alert provides participants with advance notice of the impending surprise lane drift. Further, the increase in the percentage of driving-related glance time from before the lane drift to after the lane drift was greater when the surprise lane drift was preceded by an alert compared to when there was no alert provided. In the 10-second interval before the lane drift, the proportion of driving-related glance time was, as expected, similarly low for alerts (mean = 16 percent, S.D. = 20 percent) and no alerts (mean = 19 percent, S.D. = 21 percent). However, after the lane drift until the participant regains control or is instructed to regain control, the mean proportion of driving-related glance time increases to more than twice as much after the alerts (mean = 55 percent, S.D. = 21 percent) than it does after no alerts (mean = 27 percent, S.D. = 29 percent).

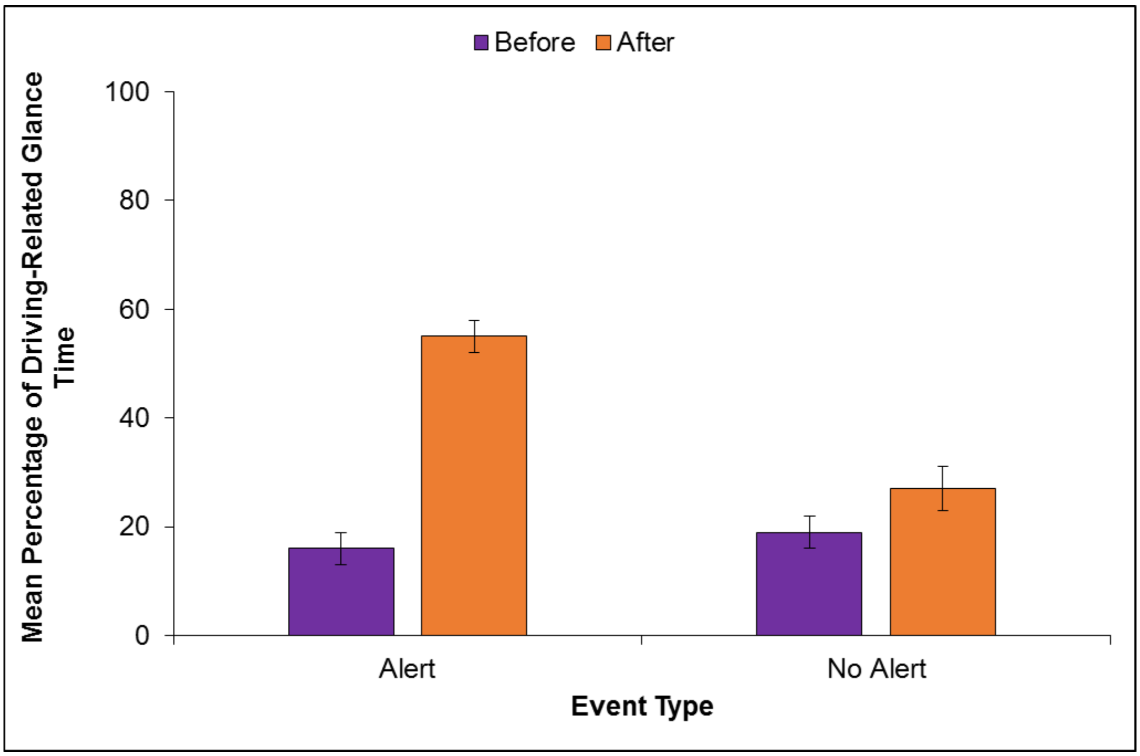



**Figure C-23. Mean and Standard Error Bar Plot of Percentage of Driving-related Glance Time, Stratified by Event Type and Time Interval Before/After the Lane Drift for Experiment 2**

Figure C-24 indicates that, overall, participants had a small percentage of driving-related glance time if they received no prompts at all, compared to receiving prompts every 2 or 7 s to increase their monitoring rate. If a participant did not receive prompts, the mean percentage of time spent on driving-related glances during the 10 s prior to the alert was 9 percent (S.D. = 15 percent), and after the alert the mean was 38 percent (S.D. = 31 percent). The percentage of time used for driving-related glances was always smaller before the alert for all prompt conditions (i.e., 2-second, 7-second, No Prompts). However, the trend shows that the prompts were able to increase the mean percentage of driving-related glance time compared to when the participants were not prompted at all. For example, under the 2-second prompts, the percentage of time participants spent monitoring the driving environment before the alert was, on average, 23 percent (S.D. = 24 percent), and it was 41 percent (S.D. = 23 percent) after the alert. That is also the case for the 7-second prompts before the alert (mean = 19 percent, S.D. = 20 percent) and after the alert (mean = 45 percent, S.D. = 31 percent). Further, this relationship holds for the participants under the No Prompts condition as well.

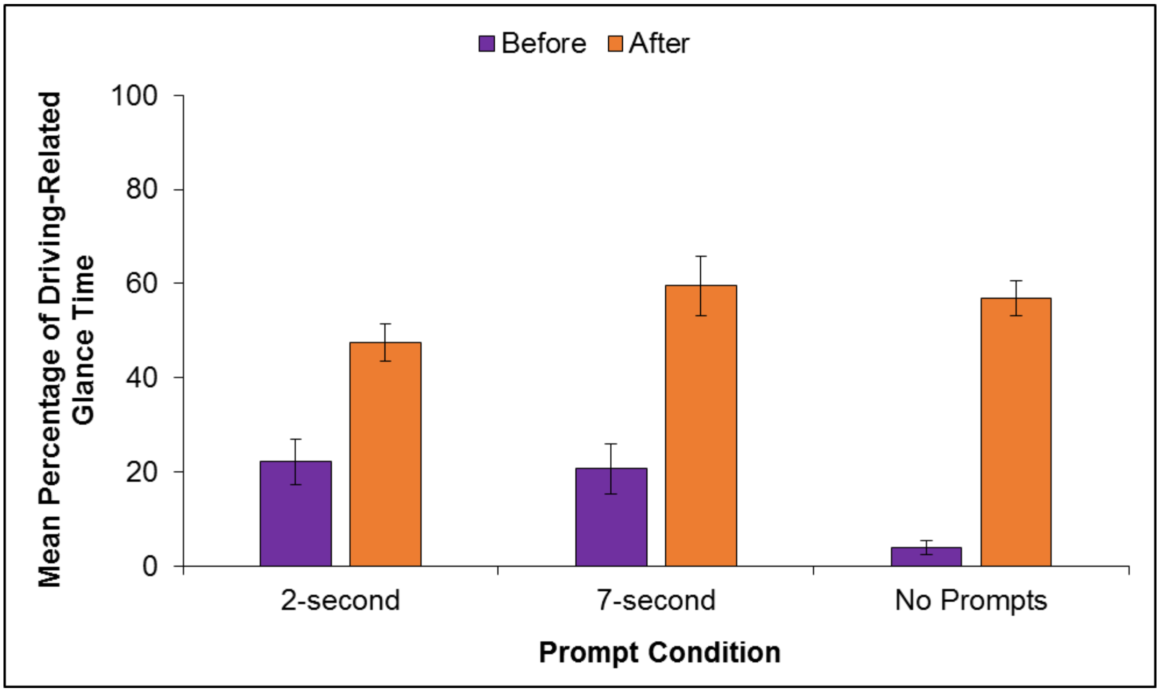



**Figure C-24. Mean and Standard Error Bar Plot Percentage of Driving-related Glance Time, Stratified by Prompt Condition and Time Interval Before/After the Alert for Experiment 2**

Figure C-25, which depicts boxplots of the percentage of driving-related glance time before and after the surprise lane drift by event type and prompt condition, indicates that the participants increased their rate of driving-related glancing after the lane drifts with alerts regardless of the prompt condition. Similarly, the increase after no alerts was not as strong, regardless of prompt condition. The largest increase is within the No Prompts condition when the alert was presented, where the increase in the mean percentage of driving-related glance time is 53 percent from before the lane drift to after the lane drift. For 2-second prompts, this increase within the alerts was 23 percent, and within 7-second prompts, the increase was 28 percent. This suggests that an alert may have had a greater effect on the change in the proportion of driving-related glance time for participants who did not have any prompts than for participants who did have prompts.

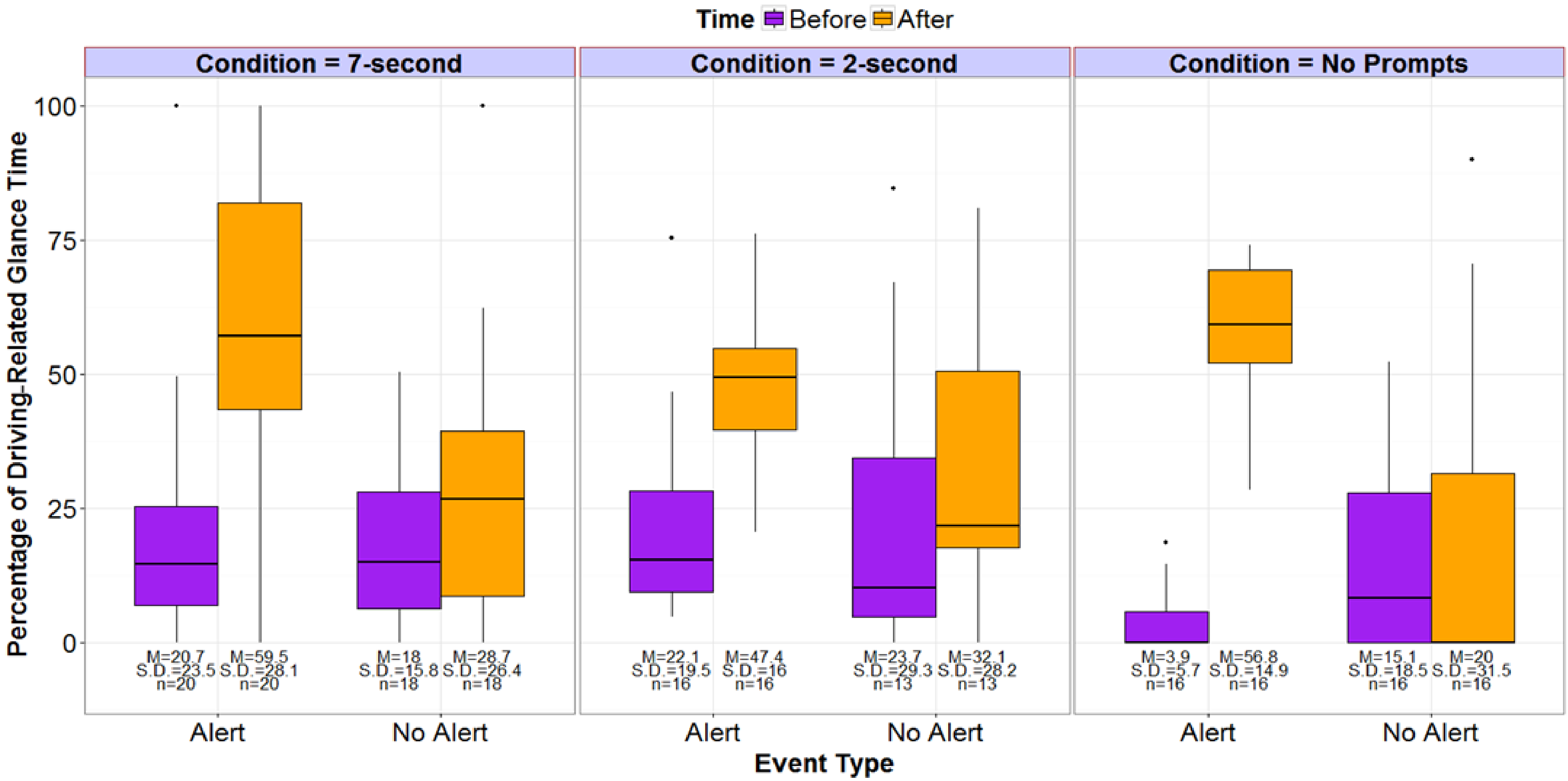



**Figure C-25. Boxplots of Monitoring, Stratified by Event Type, Condition, and Time (Before/After the Lane Drift)**

Statistical Analysis

Because these data are percentages and, therefore, bounded above and below (by 0 and 100), a beta regression model was fit (the beta distribution handles positive data bounded at both ends). This technique specifically models the log odds of monitoring the driving environment. Interpretation was then made on the odds ratios of monitoring the driving environment at a point in time. Because the model does not directly handle zeros and ones, these values were transformed via the method described in Smithson and Verkuilen, 2006, to add a small amount to the zeros and subtract a small amount from the ones. To take into account a possible subject effect in the data resulting from participants having been measured twice, a generalized estimating equation approach was used with the covariance estimated empirically from the data. The key factors of interest are event type, prompt condition, and the change from before the event to after the event. Driving session was also included. For these four factors, their main effects and two-way interactions were included in the model. In addition, the three-way interaction between event type, before/after event, and prompt condition was included to test if the effect of the event type on the change in rate before/after the event varied across prompt conditions. The participant's age was also included as a covariate, though only the effects of event type, prompt condition, before/after event, and session were reported. The analysis results are displayed in Table C-13.

**Table C-13. Analysis of Monitoring Rate, ANOVA Table, for Experiment 2**

| Effect | Num DF | Den DF | F Value | Pr > F |
|---|---|---|---|---|
| **Event Type** | **1** | **43** | **4.14** | **0.0481** |
| **Prompt Condition** | **2** | **49** | **4.95** | **0.0110** |
| **Time** | **1** | **50** | **60.20** | **<0.0001** |
| **Session** | 2 | 38 | 1.89 | 0.1646 |
| **Event Type*Prompt Condition** | 2 | 43 | 0.92 | 0.4054 |
| **Event Type*Time** | **1** | **43** | **23.29** | **<0.0001** |
| **Event Type*Session** | 2 | 38 | 0.72 | 0.4942 |
| **Prompt Condition*Time** | 2 | 50 | 3.11 | 0.0533 |
| **Prompt Condition*Session** | 4 | 38 | 0.70 | 0.5944 |
| **Time*Session** | 2 | 44 | 0.16 | 0.8564 |
| **Event Type*Prompt Condition*Time** | **2** | **43** | **3.39** | **0.0430** |

The results reveal that the three-way interaction between event type, prompt condition, and before/after event is significant ($p = 0.0430$), indicating that there is evidence that the effect of event type in the change in the percentage of driving-related glance time from before to after the event varies across prompt conditions. A check for a significant effect from before to after the lane drift revealed that the change was significant with an alert, but not without an alert (regardless of prompt condition; see Table C-14).

**Table C-14. Tests for Significance of the Change in Log Odds of Monitoring the Roadway from Before to After the Lane Drift, For Alerts and No Alerts**

| Event Type | Num DF | Den DF | F Value | Pr > F |
|---|---|---|---|---|
| **Alert** | 1 | 43 | 126.13 | <.0001 |
| **No Alert** | 1 | 43 | 2.8 | 0.1017 |

To study this further, the following procedure was performed: comparisons between before and after events with alerts were made within each prompt condition, as were comparisons before and after events with no alerts. The *p* values for these comparisons were adjusted for multiple comparisons using the Bonferroni correction. First, least squares means of the log odds of monitoring the driving environment at a point in time were plotted from before to after the event in different prompt conditions, first for alerts (Figure C-26), and then for no alerts (Figure C-27).

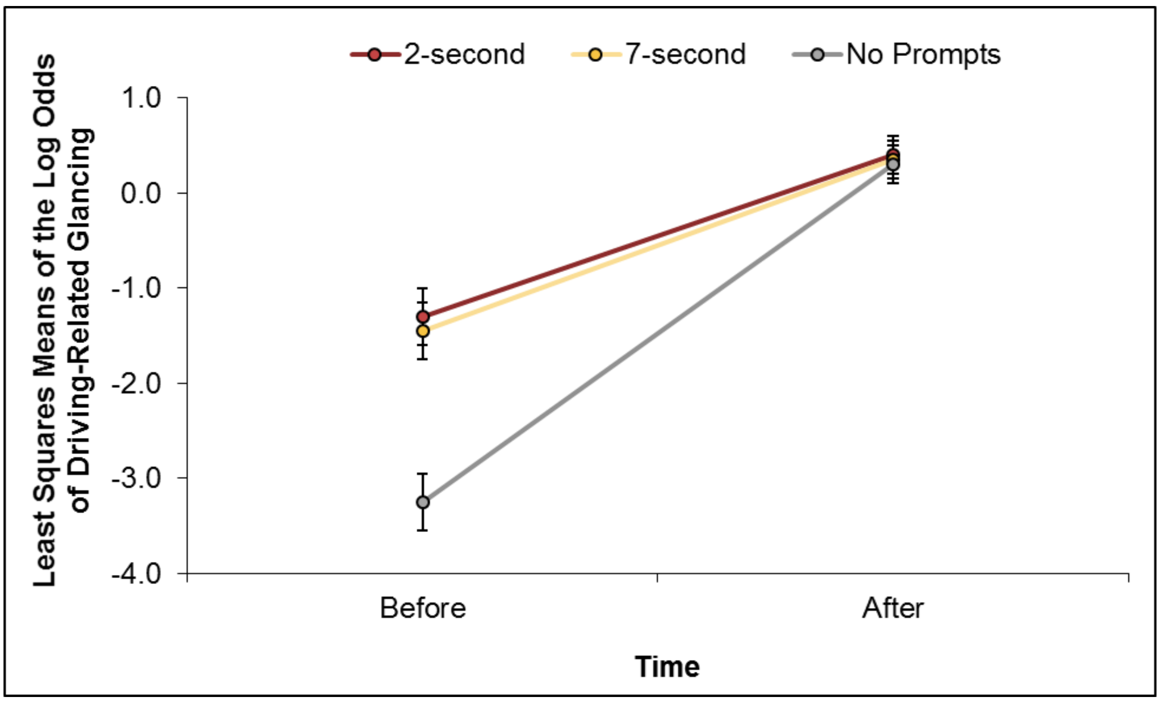



**Figure C-26. Least Squares Means of the Log Odds of Driving-related Glancing, Instances Following an Alert, Stratified by Time and Prompt Condition**

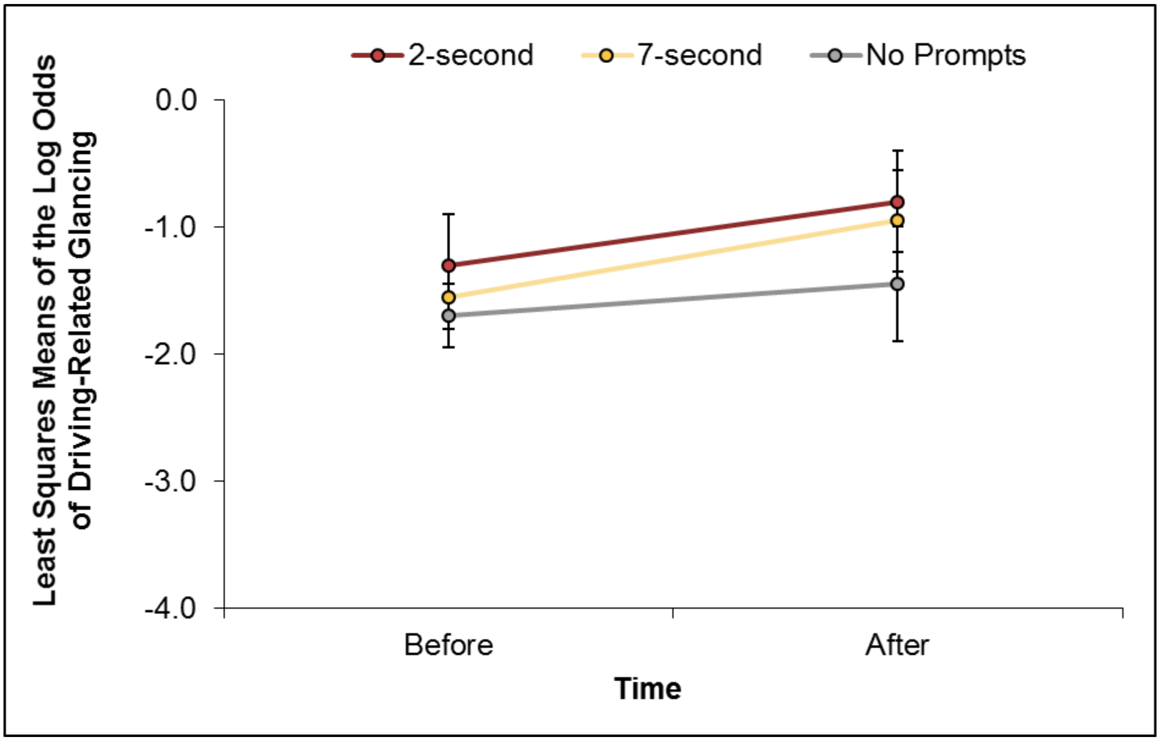



**Figure C-27. Least Squares Means of the Log Odds of Driving-related Glancing, Instances Without an Alert, Stratified by Time and Prompt Condition**

In Table C-15, the differences in the log rates from before to after the lane drift are given for all three prompt conditions separately. All three of these comparisons are significant at the adjusted level, while none of the similar comparisons for the no alert cases are significant (these are not shown).

**Table C-15. Differences in Least Squares Means of the Log Odds of Driving-related Glancing Before an Alert to After an Alert**

| Event Type | Prompt Condition | Estimate | S.E. | *t* Value | Adj *p* Value | Lower Confidence Limit | Upper Confidence Limit |
|---|---|---|---|---|---|---|---|
| **Alert** | **2-s** | 1.16 | 0.34 | 3.42 | 0.0084 | 0.48 | 1.85 |
| **Alert** | **7-s** | 1.77 | 0.32 | 5.45 | <0.0001 | 1.11 | 2.42 |
| **Alert** | **No Prompts** | 3.51 | 0.36 | 9.87 | <0.0001 | 2.79 | 4.23 |

The estimates of the change from before to after the lane drift of 1.16 for 2-second prompts (95 percent confidence interval 0.48 to 1.85), 1.77 for 7-second prompts (95 percent confidence interval 1.11 to 2.42), and 3.51 for No Prompts (95 percent confidence interval 2.79 to 4.23) can be back-transformed into odds ratio estimates of 3.2, 5.9, and 33.5, respectively. The interpretation of the odds ratio of 3.2 for the 2-second prompt is that the odds are 3.2 times as high that a vehicle operator will be making a driving-related glance at a given point in time immediately following a lane drift compared to immediately before a lane drift. Similar interpretations can be made for the 7-second prompt and No Prompts estimates.

The evidence is that an alert may successfully assist vehicle operators in monitoring the driving environment immediately following a lane drift, regardless of how often they receive prompts. However, the change still may not be the same across all three prompt conditions. Specifically, alerts may be even more helpful for a vehicle operator when he/she is receiving prompts than when he/she is not. Thus, it is of interest to test whether or not the estimates above differ significantly from each other in order to see if the change is greater for different prompt conditions. The differences are displayed in Table C-16 with adjusted *p* values.

**Table C-16. Differences Between Prompt Conditions in the Change in the Log Odds of Driving-related Glancing from Before to After the Alert**

| Label | Estimate | Standard Error | *t* Value | Adj *p* Value |
|---|---|---|---|---|
| **2-second vs. 7-second** | -0.6 | 0.48 | -1.26 | 0.6387 |
| **2-second vs. No Prompts** | **-2.35** | **0.49** | **-4.75** | **<0.0001** |
| **7-second vs. No Prompts** | **-1.75** | **0.49** | **-3.56** | **0.0024** |

The estimated difference of -2.35 between the 2-second prompts and No Prompts is significant ($p < 0.0001$), as is the estimated difference of -1.75 between the 7-second condition and the No Prompts condition ($p = 0.0024$). However, the difference between the 2-second and 7-second prompt conditions is not significant ($p = 0.6387$). Hence, the change after alerts in the rate of driving-related glance time from immediately before the alert was significantly greater for participants not receiving prompts compared to those receiving prompts. However, for those participants receiving prompts, the change in the rate of driving-related glance time was relatively similar between the 2-second and 7-second prompt conditions.

In summary, the rate at which vehicle operators monitor the driving environment may increase immediately after a lane drift if the lane drift is preceded by an alert, but this may not be the case if the lane drift does not provide an alert. The operators that did not receive an alert did not seem to clearly modify their behavior to be more vigilant. Further, the increase in the proportion of driving-related glance time from before to after the lane drifts with alerts may vary across prompt conditions. Within all prompt conditions, vehicle operators may increase their rate of driving-related glance time immediately after an alert compared to immediately before the alert, but this change may be even greater if the participant is not receiving prompts due to the low proportion of time spent monitoring the driving environment during this prompt condition.

### *Analysis for Non-driving-related Glances*

Figure C-28 shows boxplots of the rates of non-driving-related glances per second before and after the lane drift, stratified by event type and prompt condition. Although the mean rate increased consistently for participants from before the event to after the event, the increase was more pronounced after an alert compared to after no alert, and this appears consistent across all prompt conditions. The increase in the mean rate of non-driving-related glances after alerts was 0.29 for the 2-second condition, 0.21 for the 7-second condition, and 0.31 for the No Prompts condition. Meanwhile, the highest increase after no alerts was 0.08 for the 7-second prompt group. Hence, the participants responded more strongly to the alerts than to no alerts, and the different prompt condition groups were relatively consistent in this respect.

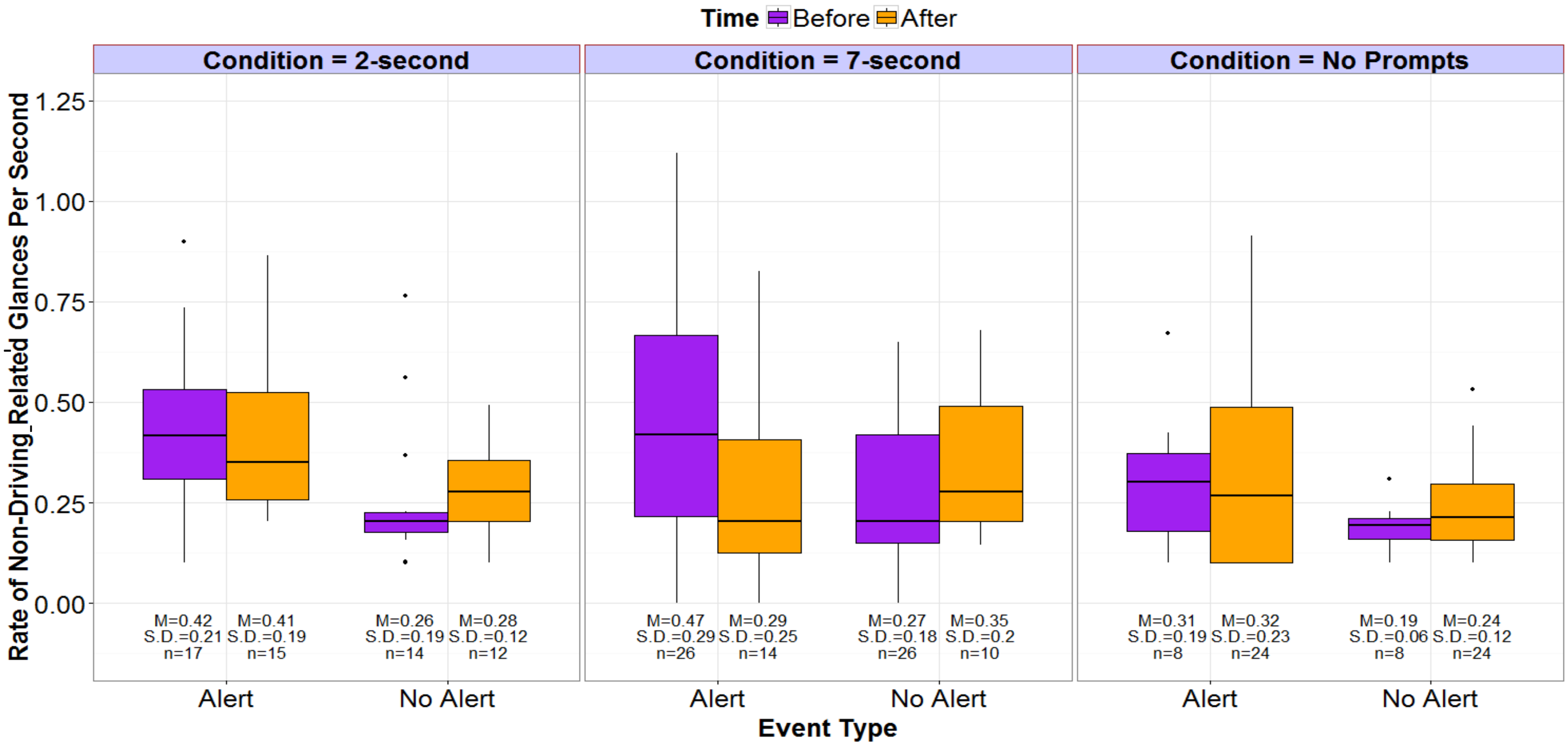



**Figure C-28. Boxplots of the Rate of Non-driving-related Glances Before and After the Lane Drift, Stratified by Alert Type and Prompt Condition**

Statistical Analysis

A generalized estimating equation Poisson model was fit with the covariance structure estimated empirically. This analysis is modeling the log of the rate of non-driving-related glances. The model results are displayed in Table C-17.

**Table C-17. Non-driving-related Glance Analysis, Lane Drifts, ANOVA Table, for Experiment 2**

| Effect | DF | Chi-Square | Pr > ChiSq |
|---|---|---|---|
| **Event Type** | **1** | **8.09** | **0.0044** |
| **Prompt Condition** | **2** | **7.98** | **0.0185** |
| **Event Type*Prompt Condition** | 2 | 2.17 | 0.3378 |
| **Time** | **1** | **20.59** | **<0.0001** |
| **Event Type*Time** | **1** | **10.73** | **0.0011** |
| **Prompt Condition*Time** | 2 | 1.54 | 0.4639 |
| **Prompt Condition*Event Type*Time** | 2 | 4.25 | 0.1193 |
| **Session** | 2 | 1.17 | 0.5558 |
| **Time*Session** | 2 | 1.11 | 0.5744 |
| **Event Type*Session** | 2 | 2.91 | 0.2338 |
| **Prompt Condition*Session** | **4** | **11.9** | **0.0181** |

The interaction between the alert type and before/after the event is significant ($p$ = 0.0011), indicating that the change in the log rate of non-driving-related glances from before to after the event is dependent on whether or not an alert is presented. Additionally, the interaction of prompt condition and session is significant, indicating that the effect of prompt condition on the log rate of non-driving-related glances varies across driving sessions and, therefore, may change over time.

To investigate the event type by before/after interaction, least squares means of the log rate are displayed in Figure C-29, while tests for the effects of event type within different levels of time, and of time within different levels of event type, are displayed in Table C-18.

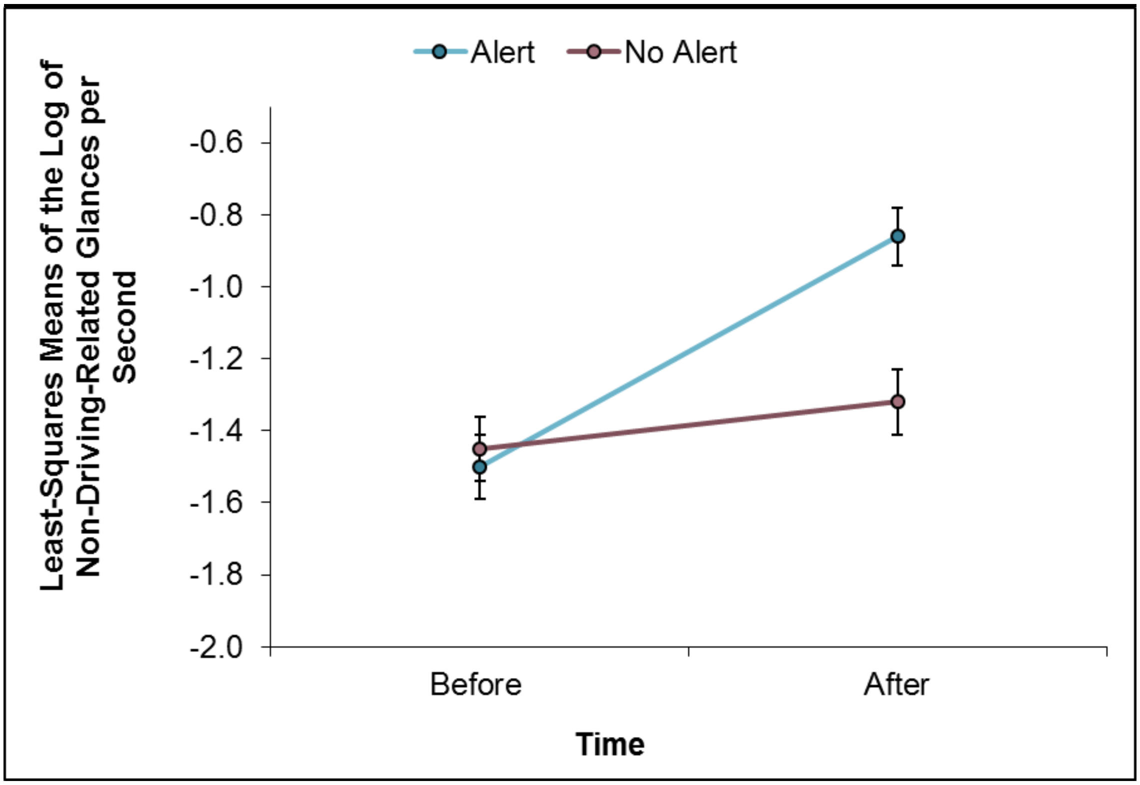



**Figure C-29. Least Squares Means of the Log of the Rate of Non-driving-related Glances Before and After the Lane Drift, Stratified by Event Type**

**Table C-18. Tests for Significant Effects of Time Within Event Type Levels, and Event Type Within Time Levels, for Experiment 2**

| Event Type | Time | Num DF | Chi-Square | Pr > ChiSq |
|---|---|---|---|---|
| **Alert** | | **1** | **53.69** | **<0.0001** |
| **No Alert** | | 1 | 1.55 | 0.2131 |
| | **Before** | 1 | 0.04 | 0.8936 |
| | **After** | **1** | **20.45** | **<0.0001** |

Table C-19 indicates that, within lane drifts with alerts, there is a significant change from before the lane drift to after, but this is not the case for lane drifts without alerts. Similarly, after the lane drift, there is a significant difference between rates if the lane drift had no alert (silent) versus if the lane drift had an alert. These differences are examined in Table C-19.

**Table C-19. Differences in Least Squares Means of the Log Rate of Non-driving-related Glances, Event Type*Time**

| Event Type* Time | Comparison | Estimate | S.E. | Z-Value | Adj *p* Value | Alpha | Lower | Upper |
|---|---|---|---|---|---|---|---|---|
| **Alert*After** | **Alert*Before** | 0.66 | 0.09 | 7.33 | <0.0001 | 0.05 | 0.48 | 0.83 |
| **Alert*After** | **No Alert*After** | 0.50 | 0.11 | 4.52 | <0.0001 | 0.05 | 0.28 | 0.71 |

The change in the log rate of non-driving-related glances between before and after a lane drift with alert was estimated to be 0.66 (95 percent confidence interval 0.48 to 0.83). This estimate is back-transformed to 1.93, indicating that the rate of non-driving related glances is about 93 percent higher after a lane drift with an alert as compared to before. A similar interpretation can be made of the estimated difference in the log rate after a lane drift with an alert as compared to after a lane drift without an alert, estimated to be 0.50 (95 percent confidence interval 0.28 to 0.71). This estimate, back-transformed to 1.65, indicates that the rate of non-driving-related glances is 65 percent higher after a lane drift with an alert as compared to a lane drift with no alert.

Part of this difference in rate between alerts and no alerts after the event may be due to increased durations of non-driving-related glances after lane drifts without alerts as compared to after lane drifts with alerts, and not due to more frequent glances after lane drifts with alerts. Although the mean amount of time monitoring the driving environment after the lane drift was similar for alerts (1.21 s) and no alerts (1.15 s), non-driving-related glances were more than three times as long after lane drifts without alerts (mean of 3.06 s) compared to lane drifts with alerts (mean of 0.94 s). In contrast, the counts of glances were relatively similar. To understand this, note that in 39 of 52 of the instances of lane drifts with alerts (75 percent), each participant made two glances (adding the driving-related and non-driving-related glances together) after the lane drift, and during all of these instances, only the last glance was driving-related.

This means that, in the majority of cases of lane drifts with an alert, participants were not monitoring the driving environment at the time of the lane drift but did take control after their first driving-related glance after the lane drift. Meanwhile, in 17 of 48 instances of lane drifts without alerts (35.4 percent), participants made only one glance (in 14 of these instances, this glance was non-driving related), while in another 17 instances, participants made two glances. Thus, after lane drifts both with alerts and without, most participants are making only one total non-driving-related glance. Therefore, the higher rate of non-driving-related glances immediately after lane drifts with alerts, compared to immediately after lane drifts without alerts, is due primarily to participants regaining control more quickly, as opposed to a change in eye-glance duration.

Regarding the significant interaction between prompt condition and session, least squares means of the log rate are plotted in Figure C-30, while tests for the effect of prompt condition within session, and session within prompt condition, are displayed in Table C-20.

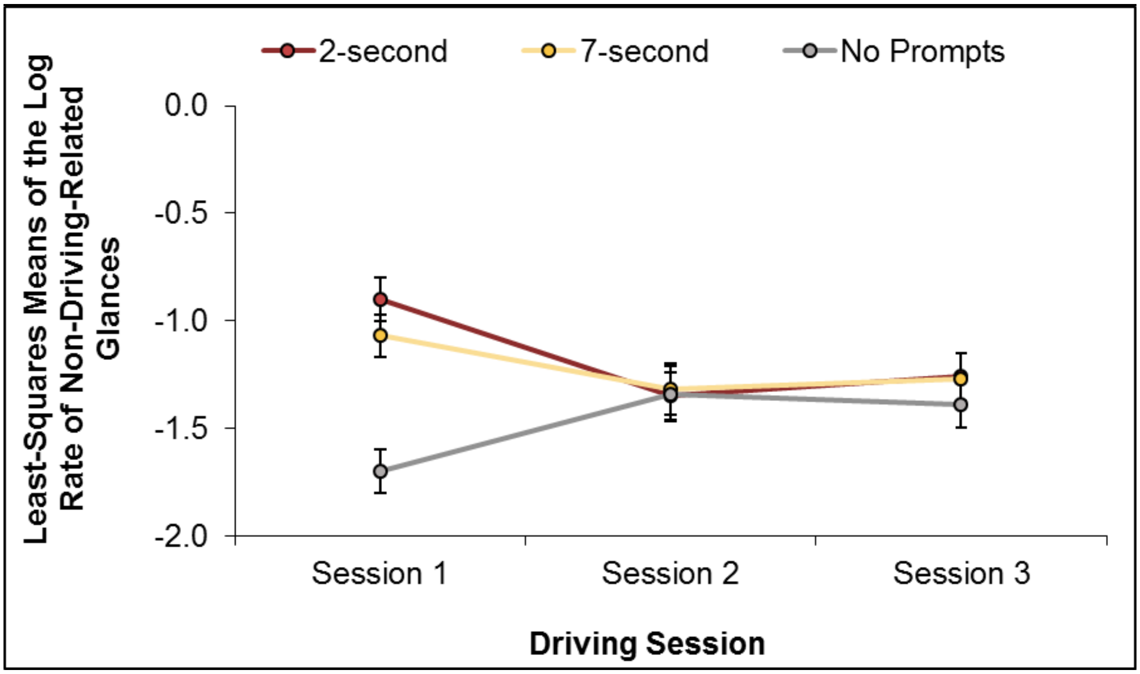



**Figure C-30. Least Squares Means of the Log of Non-driving-related Glances Across Sessions, Stratified by Prompt Condition**

**Table C-20. Tests for Significance Within Different Levels of Prompt Condition/Session**

| Prompt Condition | Session | Num DF | Chi-Square | Pr > ChiSq |
|---|---|---|---|---|
| **2-second** | | **2** | **12.68** | **0.0018** |
| **7-second** | | 2 | 1.85 | 0.3967 |
| **No Prompts** | | 2 | 8.24 | 0.0163 |
| | **1** | **2** | **34.11** | **<0.0001** |
| | **2** | 2 | 0.01 | 0.9944 |
| | **3** | 2 | 0.56 | 0.7554 |

Within the 2-second condition, there is a significant difference in the log rate of non-driving-related glances between different sessions. Similarly, within Session 1, there is a significant difference between the conditions. These differences are investigated in Table C-21, where significant differences in the least squares means are listed.

**Table C-21. Differences in Least Squares Means of the Log Rate of Non-driving-related Glances, Prompt Condition*Session**

| Prompt Condition* Session | Comparison | Estimate | S.E. | Z-Value | Adj *p* Value | Lower Confidence Limit | Upper Confidence Limit |
|---|---|---|---|---|---|---|---|
| **2-second* Session 1** | **2-second* Session 2** | 0.43 | 0.12 | 3.48 | 0.003 | 0.19 | 0.68 |
| **2-second* Session1** | **2-second* Session 3** | 0.39 | 0.14 | 2.73 | 0.0378 | 0.11 | 0.68 |
| **2-second* Session 1** | **No Prompts* Session 1** | 0.83 | 0.15 | 5.63 | <0.0001 | 0.54 | 1.12 |
| **7-second* Session 1** | **No Prompts *Session 1** | 0.67 | 0.16 | 4.09 | <0.0001 | 0.35 | 0.99 |

The significant differences within the 2-second prompt condition lie between Session 1 and the other two sessions. The estimated difference in the log rate between Session 1 and Session 2 is 0.43 (95 percent confidence interval 0.19 to 0.68), and the estimated difference between Session 1 and Session 3 is 0.39 (95 percent confidence interval 0.11 to 0.68). Hence, the rate was significantly higher in Session 1 than the other two sessions for the 2-second condition. Meanwhile, within Session 1, the No Prompts condition had a significantly lower log rate of non-driving-related glances compared to the 2-second condition (estimated difference 0.83, 95 percent confidence interval 0.54 to 1.12) and the 7-second condition (estimated difference 0.67, 95 percent confidence interval 0.35 to 0.99).

In summary, the eye-glance behavior of participants differs after a lane drift with an alert versus after a lane drift without an alert in that the participants are more likely to make driving-related glances of shorter duration after a lane drift with an alert compared to after lane drift without an alert. Additionally, based on the differences in Session 1 between the No Prompts condition and the 2-second or 7-second conditions, operators may have lower rates of non-driving-related glances if they have not experienced prompts early on in driving. However, later on, the rate of non-driving-related glances may increase, falling more in line with the rates of participants experiencing prompts.

# Results for Prompts

## Operator Behavior

### *Analysis for Time to React*

The reaction times to the prompt, as depicted in the mean and error bar plots stratified by prompt condition (Figure C-31) and driving session (Figure C-32), as well as the boxplots stratified by prompt condition and event type (Figure C-33), were higher in Session 3. The means were 2.6 s (S.D. = 2.8 s) within the 2-second group, and 2.8 s (S.D. = 32 s) in the 7-second group. Within sessions, the mean reaction time rises from about 2.4 s (S.D. = 2.4 s) in Session 1 to 3.0 s in Session 2 (S.D. = 3.7 s), with the same approximate mean in Session 3 (S.D. = 2.5 s). Overall, in 352 instances out of 466 (71.2 percent), the participant reacts before the alert reaches Stage 2 (within 5 s).

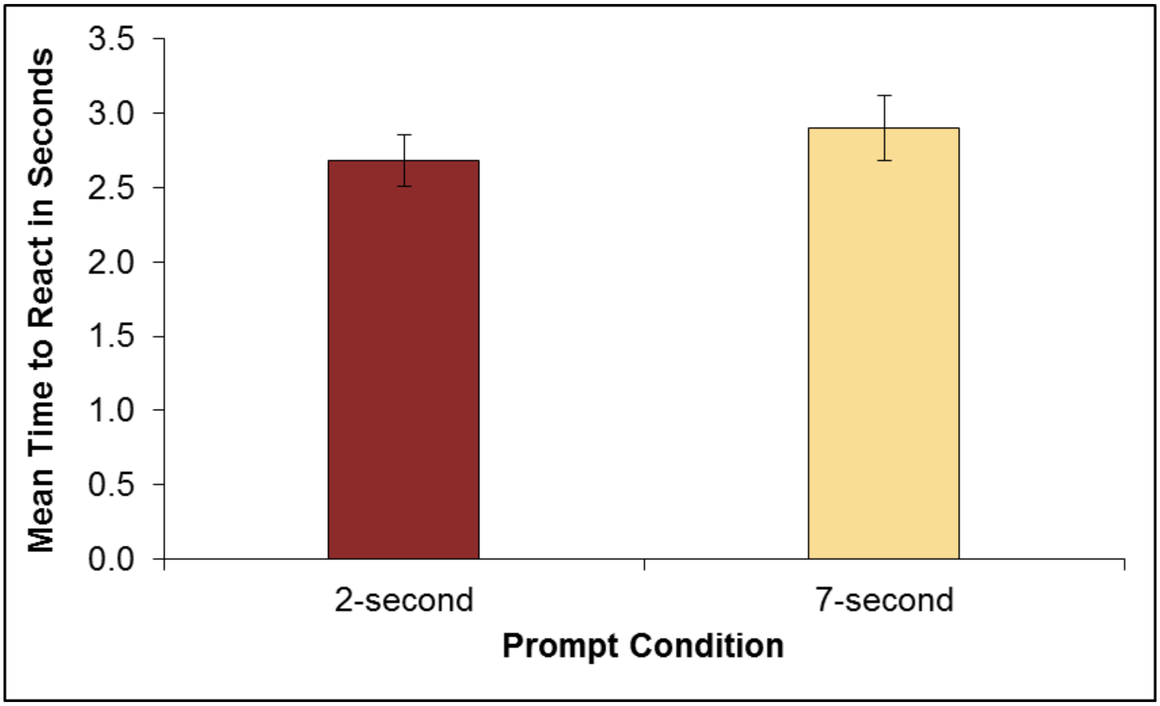



**Figure C-31. Mean and Error Bar Plots of Time to React to Prompt, by Prompt Condition**

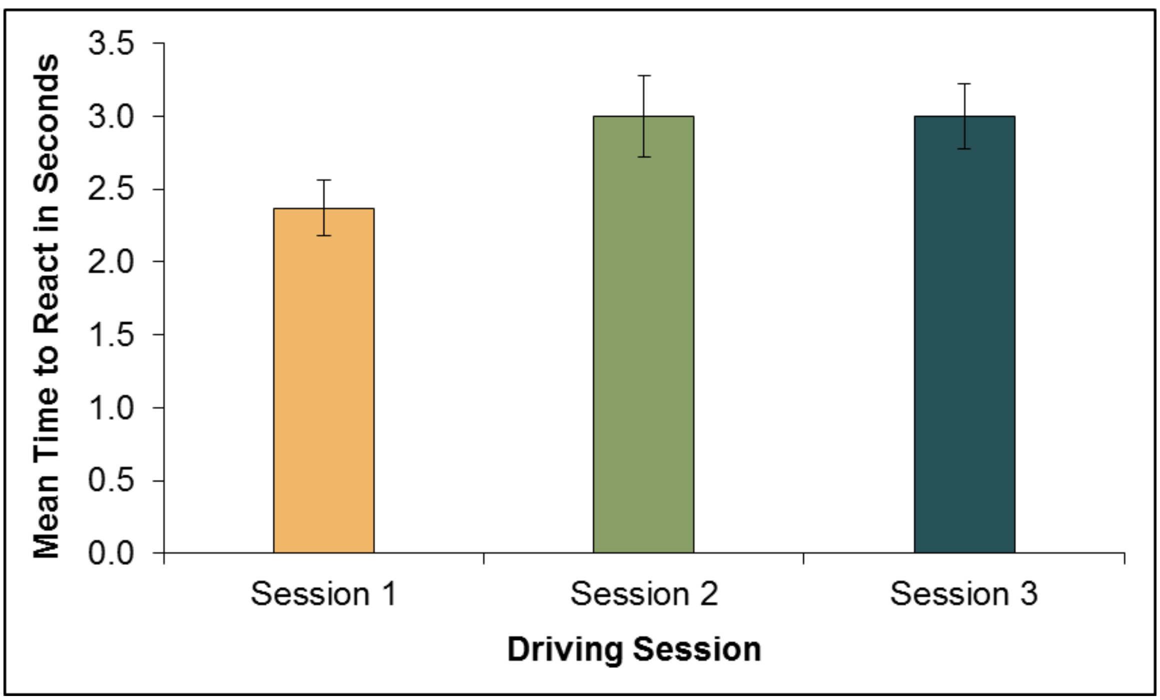


**Figure C-32. Mean and Error Bar Plots of Time to React to Prompts, by Session**

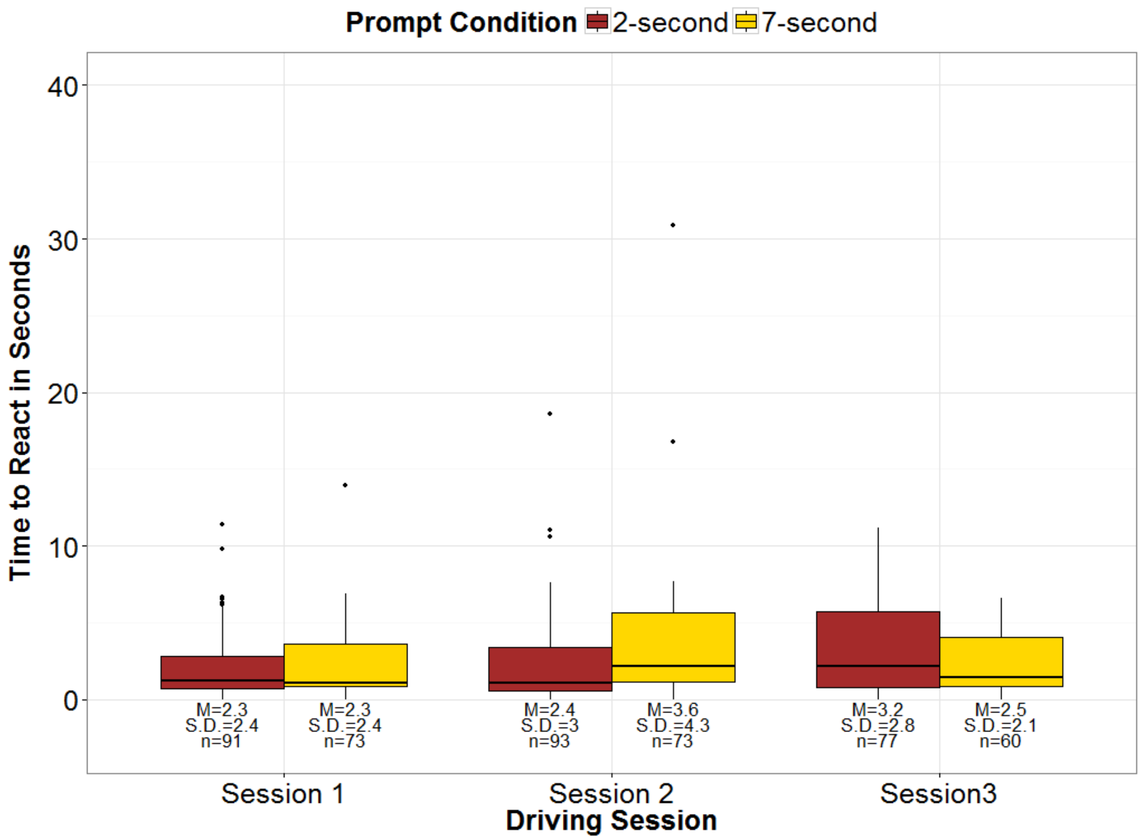



**Figure C-33. Boxplots of Time to React to the Prompts, Stratified by Session and Prompt Condition**

Statistical Analysis

For the analysis of the time to react to the prompts, the time to react was log-transformed, and the analysis performed on the log-transformed data. Although this analysis is not specifically on the lane drifts, to account for the placement of lane drifts within the experimental run, the event type (alert with lane drift, no alert with lane drift, or no alert with no lane drift) was included at first. This variable was not significant in this case, so it was removed, and the model was run with prompt condition, session, and their interaction. The results are displayed in Table C-22.

**Table C-22. Analysis of the Time to React to Prompts, ANOVA Table, for Experiment 2**

| Effect | Num DF | Den DF | F Value | Pr > F |
|---|---|---|---|---|
| **Prompt Condition** | 1 | 27.9 | 0.91 | 0.3471 |
| **Session** | **2** | **41.9** | **3.34** | **0.0449** |
| **Prompt Condition*Session** | **2** | **41.9** | **3.80** | **0.0305** |

The prompt condition*session interaction is significant ($p$ = 0.0305), indicating that the effect of the prompt conditions on the log of the reaction time to the prompts may change over time.

To investigate this further, a plot of the least squares means by prompt condition*session is displayed in Figure C-34, while significance tests at each level of prompt condition and session are displayed in Table C-23. These are considered significant if they achieve the Bonferroni adjusted significance level of 0.01.

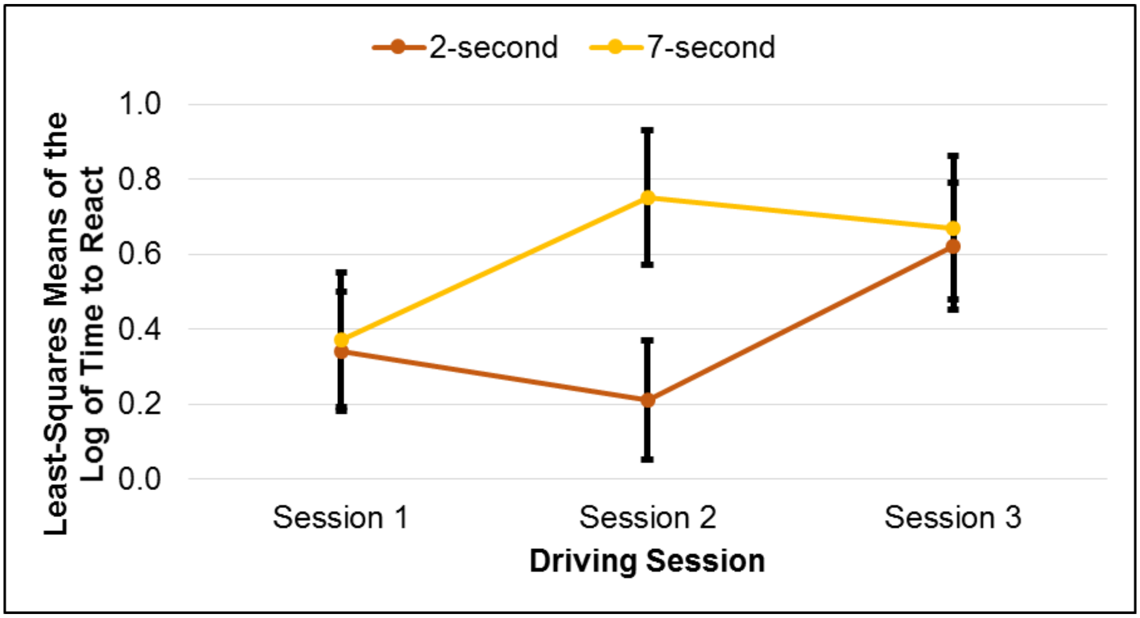



**Figure C-34. Least Squares Means of the Log of Time to React to Prompts Across Sessions**

**Table C-23. Significant Differences in Least Squares Means of the Log of Time to React to Prompts Across Sessions**

| Prompt Condition | Session | Num DF | Den DF | F Value | Pr > F |
|---|---|---|---|---|---|
| **2-second** | | 2 | 42.1 | 3.58 | 0.0367 |
| **7-second** | | 2 | 32.4 | 3.63 | 0.0378 |
| | **1** | 1 | 44.4 | 0.01 | 0.9242 |
| | **2** | 1 | 49.5 | 4.95 | 0.0307 |
| | **3** | 1 | 45.3 | 0.01 | 0.9079 |

The participants seemed to react to the prompt more quickly after Session 1 if they were in the 2-second condition, and then their reaction time slowed down in Session 3. Conversely, those in the 7-second group reacted more slowly in Session 2 compared to Session 1, and reaction time then remained relatively stable in Session 3. However, none of the comparisons within each level of these variables was significant at the Bonferroni adjusted 0.01 level.

Because the interaction effect did reveal significant multiple comparisons, the main effects of session and prompt condition were reassessed. Table C-22 indicates that session was significant ($p = 0.0449$), but prompt condition was not ($p = 0.3471$). Thus, the main effect of session across both prompt conditions was assessed. This analysis revealed a significant increase from Session 1 to Session 3 (after Bonferroni adjustment), detailed in Table C-24.

**Table C-24. Significant Differences in Least Squares Means of the Log of Time to React to Prompt Between Sessions**

| Session | Comparison | Estimate | S.E. | *t* Value | Adj *p* Value | Lower Confidence Limit | Upper Confidence Limit |
|---|---|---|---|---|---|---|---|
| **3** | **1** | 0.29 | 0.11 | -2.59 | 0.0405 | 0.06 | 0.51 |

The estimate of 0.29 (95 percent confidence interval of 0.06 to 0.51) indicates that there is evidence that the log of the reaction time increases from Session 1 to Session 3. This estimate can be back-transformed to 1.34, indicating that in Session 3, participants were 34 percent slower, on average, to react to the prompt than they were in Session 1.

***Analysis for Time to Regain Control***

Statistical Analysis

A generalized estimating equation model was used to model the probability of regaining control after a prompt. Since prompts progressing to Stage 3 needed to regain control, these were excluded, and the model was run on prompts that reached Stage 2 or Stage 1. Variables involving event type were included if they were significant. Analysis results are displayed in Table C-25.

**Table C-25. Analysis of Time to Regain Control, Prompts, ANOVA Table, for Experiment 2**

| Source | DF | Chi-Square | Pr > ChiSq |
|---|---|---|---|
| **Prompt Condition** | 1 | 0.82 | 0.3656 |
| **Session** | 2 | 2.67 | 0.2629 |
| **Prompt Condition*Session** | 2 | 0.87 | 0.6476 |

Neither prompt condition, session, nor event type were significant, indicating that there is no statistically significant evidence of an effect of these variables on the probability of regaining control after prompts that progress either to Stage 1 or Stage 2.

## Eye-glance Behavior

***Analysis for Monitoring Rate***

The monitoring rate—the percentage of driving-related glance time in a given time interval—was examined to determine if it changed for the 2- and 7-second prompt conditions from before to after the prompt, or changed between sessions. This was compared to the change from before to after the baseline time point when the participants did not receive prompts. Monitoring rate was analyzed in instances in which the participant was engaged in a non-driving task. Participants were engaged in a non-driving task during 452 out of 459 sampled prompts (98.4 percent), while participants that did not receive prompts were engaged in a non-driving task during 185 out of 280 randomly sampled time points (66.1 percent). Boxplots of the monitoring rate are displayed in Figure C-35.

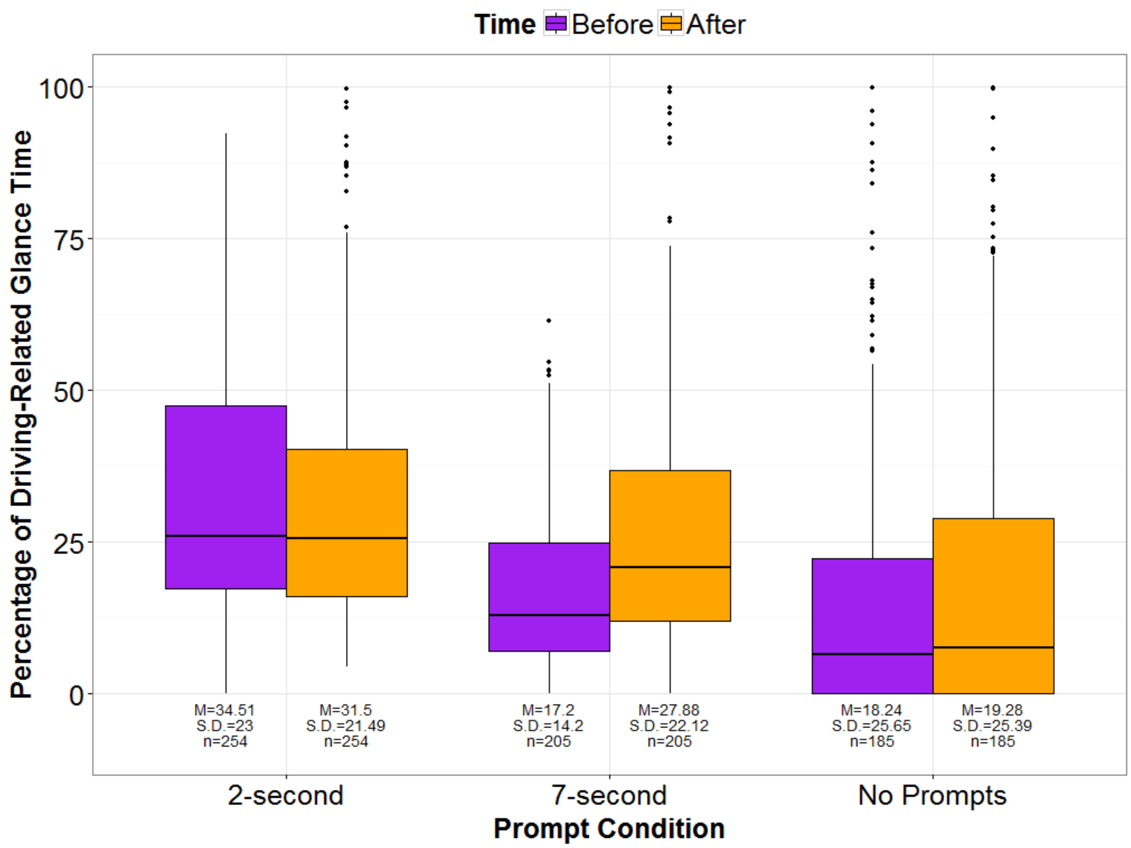



**Figure C-35. Boxplots of Monitoring Rate Around Prompts, Stratified by Prompt Condition, and Time**

Statistical Analysis

A beta regression model was used to model the monitoring rate, as the monitoring rate is a proportion. This method models the log odds of monitoring the driving environment (i.e., driving-related glance time) at any given time. Prompt condition, session, and time, as well as the two- and three-way interactions, were included. The analysis results are displayed in Table C-26.

**Table C-26. Analysis of Monitoring Rate, ANOVA Table, for Experiment 2**

| Effect | Num DF | Den DF | F Value | Pr > F |
|---|---|---|---|---|
| **Prompt Condition** | **2** | **44** | **22.68** | **<.0001** |
| **Session** | 2 | 78 | 1.26 | 0.2895 |
| **Time** | **1** | **45** | **6.55** | **0.0139** |
| **Prompt Condition*Session** | 4 | 78 | 2.15 | 0.0824 |
| **Prompt Condition*Time** | **2** | **45** | **10.66** | **0.0002** |
| **Time*Session** | 2 | 82 | 0.94 | 0.3954 |

The prompt condition*time interaction is significant ($p$ = 0.0002), indicating that the immediate change in the log odds of monitoring the roadway differs between prompt conditions. To examine this interaction, least squares means of the prompt condition*time interaction are displayed in Figure C-36. The effects within each level of these variables are displayed in Table C-27.

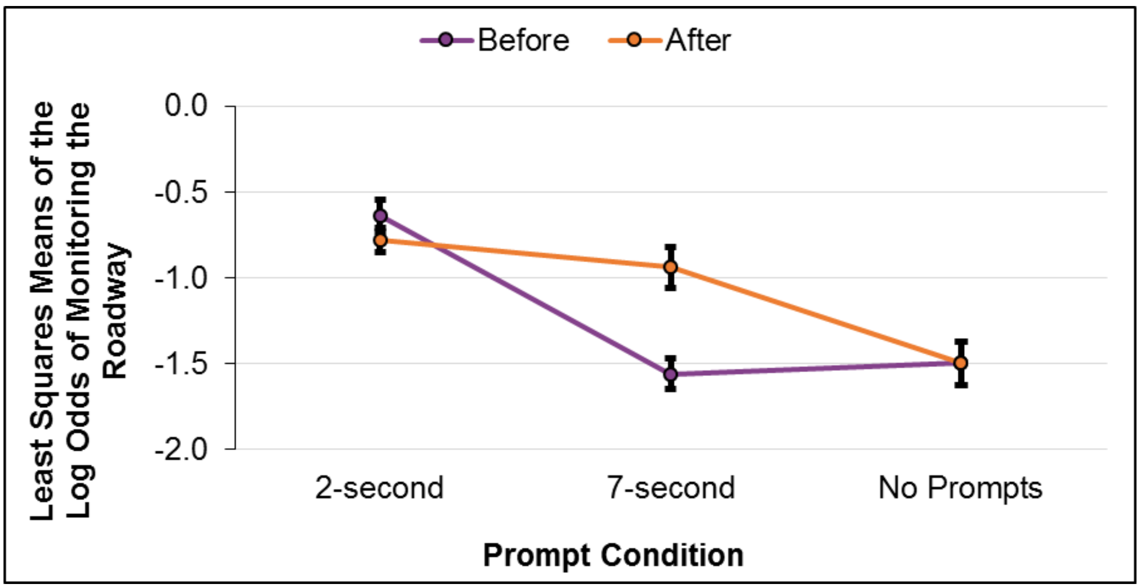



**Figure C-36. Least Squares Means of the Log Odds of Monitoring the Roadway, Stratified by Prompt Condition and Before/After the Prompt**

**Table C-27. Tests for Significance Within Different Levels of Prompt Condition for Experiment 2**

| Prompt Condition | Time | Num DF | Den DF | F Value | Pr > F |
|---|---|---|---|---|---|
| **2-second** | | 1 | 45 | 0.34 | 0.5581 |
| **7-second** | | 1 | 45 | 22.16 | <0.0001 |
| **No Prompts** | | **1** | **45** | **0.18** | **0.6745** |

| | | | | | |
|---|---|---|---|---|---|
| | **Before** | **2** | **45** | **10.43** | **0.0002** |
| | **After** | **2** | **45** | **28.73** | **<0.0001** |

Table C-27 indicates that both before and after the prompt, there is a significant difference between prompt conditions. Additionally, the only significant change from before to after the prompt occurs within the 7-second prompt condition. To compare the general monitoring rate of participants in the No Prompts condition to the worst-case scenarios in the 2-second and 7-second conditions (i.e., their monitoring rate had decreased enough to receive prompts), the participants' monitoring rates before the prompts in the 2-second and 7-second conditions, as well as in the first time interval sampled for those in the No Prompts condition, were compared. The significant differences in least squares of the log odds of monitoring the roadway are displayed in Table C-28.

**Table C-28. Significant Differences in Least Squares Means for the Log Odds of Monitoring the Roadway, Between Prompt Conditions, Significant at Bonferroni Adjusted *P* Values for Experiment 2**

| Prompt Condition | Comparison | Estimate | S.E. | *t* Value | Adj *p* Value | Lower Confidence Limit | Upper Confidence Limit |
|---|---|---|---|---|---|---|---|
| **2-second** | **No Prompts** | 0.88 | 0.17 | 5.20 | <0.0001 | 0.53 | 1.21 |
| **2-second** | **7-second** | 0.93 | 0.13 | 7.16 | <0.0001 | 0.67 | 1.19 |

Table C-28 indicates that just prior to a 2-second prompt, the monitoring rate was significantly higher than was the case for both the No Prompts condition and the 7-second condition. Note that the 7-second prompts came only in circumstances of extreme inattention, so it is not surprising that the 2-second monitoring rate is significantly higher than the 7-second monitoring rate just before the prompts. However, the fact that the 7-second monitoring rate is not significantly different from the No Prompts rate indicates that when operators experience the 7-second condition, their monitoring rate has the potential to regress back to where it would be if they did not have any prompts at all.

The estimated difference in the log odds of monitoring the roadway between 2-second prompts and no prompts of 0.88 (95 percent confidence interval of 0.53 to 1.21) can be back-transformed to an odds ratio of 2.41, indicating that just before a 2-second prompt (which occurs in cases of relatively extreme inattention), participants in the 2-second condition nevertheless had odds of monitoring the roadway 2.41 times higher than participants not receiving prompts. A similar interpretation can be made for the estimate of 0.93 (95 percent confidence interval of 0.67 to 1.19) of the difference in log odds between 2-second and 7-second prompts, which can be back-transformed to an odds ratio of 2.53.

The change in the monitoring rate from before to after the prompt in the 7-second condition is displayed in Table C-29. Since the other two conditions did not result in statistically significant changes, they are not included.

**Table C-29. Significant Differences in Least Squares Means, for Prompt Condition and Time, Significant at Bonferroni Adjusted *P* Values for Experiment 2**

| Prompt Condition* Time | Comparison | Estimate | S.E. | *t* Value | *p* Value | Lower Confidence Limit | Upper Confidence Limit |
|---|---|---|---|---|---|---|---|
| **7-second* After** | **7-second* Before** | 0.62 | 0.13 | 4.71 | <0.0001 | 0.36 | 0.89 |

There is a significant gain in the log odds of monitoring the driving environment (i.e., driving-related glance time) after the prompt compared to before the prompts within the 7-second prompt group. The estimate of 0.62 (95 percent confidence interval 0.36 to 0.89) can be back-transformed to 1.86, indicating that the odds are 1.86 times higher that a participant in the 7-second prompt group will monitor the roadway at some point in time immediately after the prompt compared to some time immediately before the prompt.

In summary, participants who received prompts in the 2-second condition kept their monitoring rate significantly higher just prior to receiving the prompt than did those in the No Prompts condition. Within the 7-second condition, however, participants' monitoring rate just prior to receiving the prompt was not significantly higher than it was for those in the No Prompts condition, indicating that there is the potential within the 7-second condition for operators' monitoring rate to regress back to baseline (note that the 7-second prompt went off only in extreme situations of time spent looking away). However, the 7-second prompt condition was effective in restoring participants' monitoring rate immediately after the prompt.

***Analysis for Non-driving-related Glances***

The amount of non-driving-related glances was investigated for participants who were engaged in a non-driving task at the randomly selected prompt (for the 2-second and 7-second conditions) or timepoint (for the No Prompts condition). The boxplots of non-driving-related glances stratified by condition and time (before and after the prompt), displayed in Figure C-37, show the strongest change from before to after the prompt occuring within the 7-second group. (Figure C-37). The mean amount of glances rose from 2.51 (S.D. = 1.1) before the prompt to 3.07 (S.D. = 1.2) after the prompt.
Additionally, those receiving prompts seemed to have a higher amount of non-driving-related glances, on average, than those who did not receive prompts, whether or not the participant was monitoring the roadway and whether or not the prompt had occurred. The highest mean for the No Prompts group was 2.45 (S.D. = 1.5). Meanwhile, the lowest mean for the 2-second or 7-second group was 2.51 s (S.D. = 1.1) which was seen in the 7-second group before the prompt. Additionally, the amount of non-driving-related glances was higher, on average, for participants in the 2-second group compared to the 7-second group. The lowest mean in the 2-second group was 3.26 (S.D. = 1.5) before the prompt, compared to a high mean of 3.07 (S.D. = 1.2) in the 7-second after the prompt.

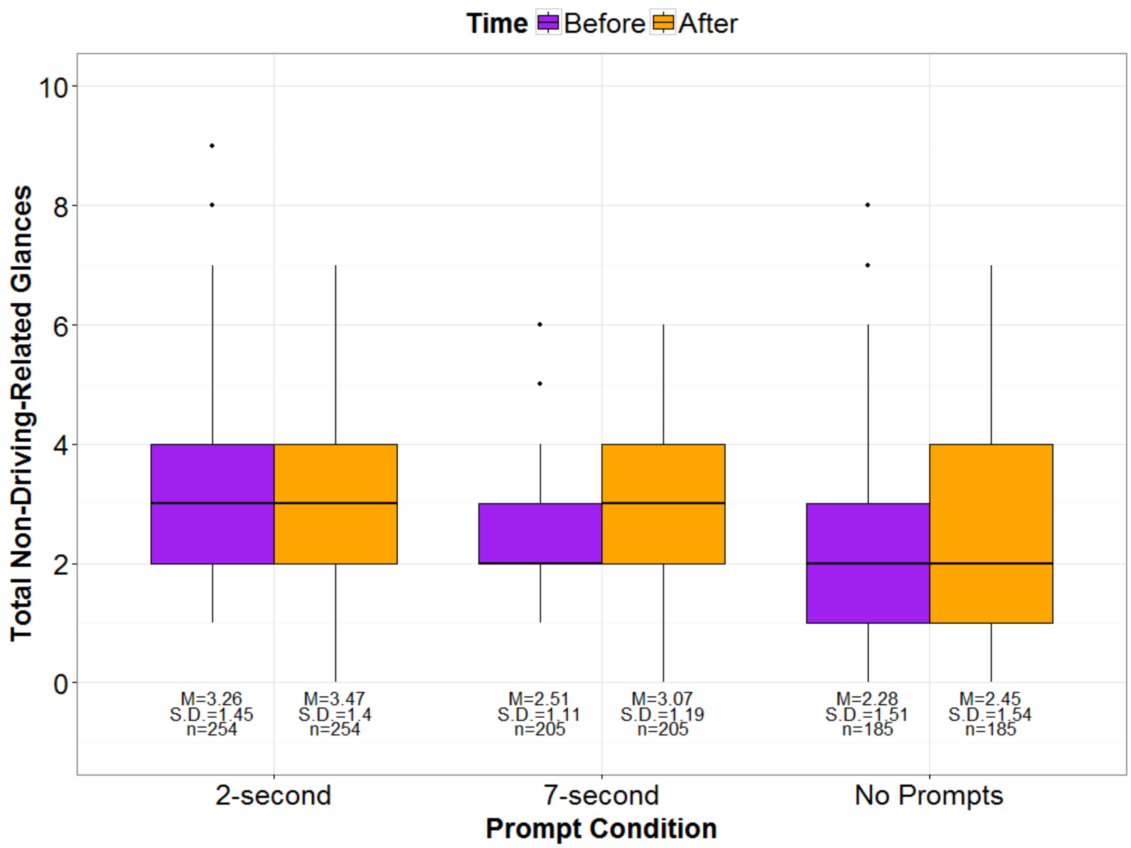



**Figure C-37. Boxplots of Non-driving-related Glances, Stratified by Prompt Condition and Time (Before/After the Prompt or Randomly Selected Time Point)**

<u>Statistical Analysis</u>

A generalized estimating equations Poisson model was used with standard errors estimated from the data. This method models the log of the amount of non-driving-related glances. In addition to the key variables of interest (prompt condition, session, and before/after the prompt), whether or not the participant was monitoring the roadway at the time of the prompt/time point sampled was used, as well as two-way interactions between roadway monitoring and the rest of the variables. Findings involving roadway monitoring that were not significant were removed. The analysis results are displayed in Table C-30.

**Table C-30. Analysis of Non-driving-related Glances Around Prompts, ANOVA Table, For Experiment 2**

| Effect | Num DF | Den DF | F Value | Pr > F |
|---|---|---|---|---|
| **Prompt Condition** | **2** | **44** | **6.37** | **0.0037** |
| **Time** | **1** | **45** | **28.66** | **<.0001** |
| **Prompt Condition*Time** | **2** | **45** | **3.27** | **0.0471** |
| **Session** | **2** | **78** | **3.81** | **0.0264** |
| **Session*Time** | 2 | 78 | 2.80 | 0.0670 |
| **Session*Prompt Condition** | 4 | 78 | 1.49 | 0.2122 |
| **Session*Prompt Condition*Time** | 4 | 78 | 1.68 | 0.2190 |

Table C-30 indicates that there is a significant interaction between prompt condition and time ($p$ = .0471), indicating that the change in the amount of non-driving-related glances from before to after the prompt may vary across prompt conditions. Additionally, session ($p$ = 0.0264) is significant, indicating that the amount of non-driving-related glances may vary across time.

To investigate these differences, least squares means plots for session and prompt condition*time, are displayed in Figure C-38 and Figure C-39, respectively.

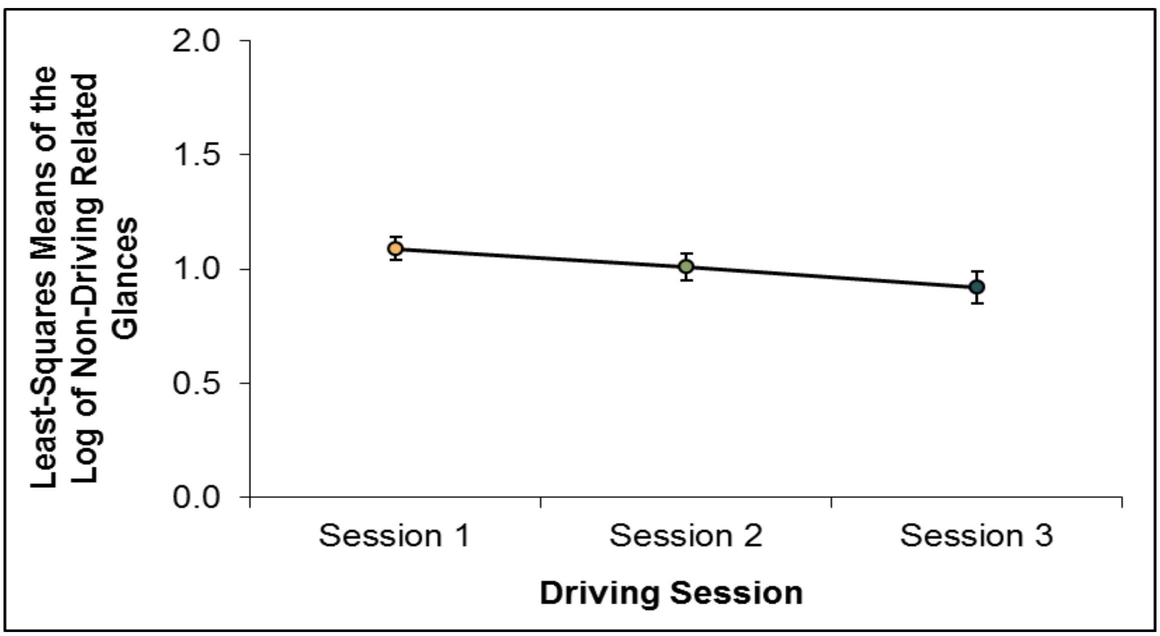



**Figure C-38. Least Squares Means of the Log of Non-driving-related Glances Across Sessions**

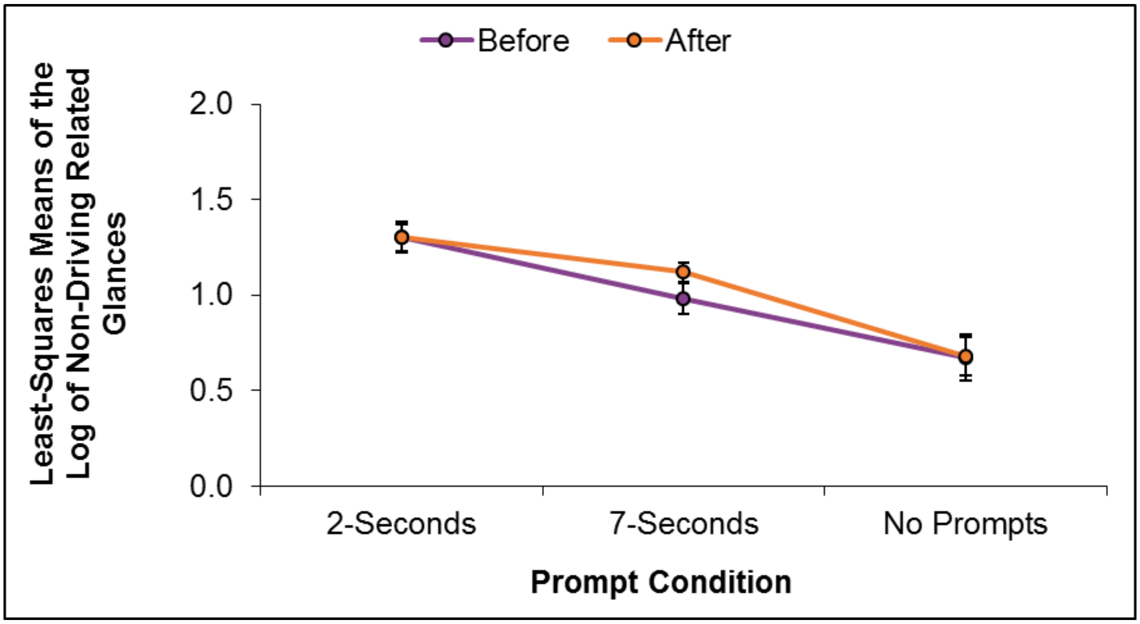



**Figure C-39. Least Squares Means of the Log of Non-driving-related Glances Before and After the Prompt, Across Prompt Conditions**

Regarding session, there appears to be a decreasing trend from earlier to later sessions. However, there were no significant comparisons with Bonferroni adjusted alpha levels between sessions.

Regarding the interaction between prompt condition and time, the significance levels (with the alpha level set at the Bonferroni adjusted 0.01) are given in Table C-31, first for time within each level of prompt condition, and then for prompt condition within each level of time. The 7-second and 2-second prompt conditions show a significant difference before and after the prompt, and there are significant differences between the conditions before the prompt and after the prompt. Table C-32 shows the estimated difference before and after the prompt within the 2-second and 7-second group, and Table C-33 shows the significant differences between the prompt conditions before the prompt (which are of primary interest).

**Table C-31. Tests for Significance Within Different Levels of Prompt Condition and Time for Experiment 2**

| Prompt Condition | Time | Num DF | Den DF | F Value | Pr > F |
|---|---|---|---|---|---|
| **2-second** | | 1 | 45 | 5.77 | 0.0207 |
| **7-second** | | **1** | **45** | **15.39** | **0.0002** |
| **No Prompts** | | 1 | 45 | 5.90 | 0.0175 |
| | **After** | **2** | **45** | **5.19** | **0.0094** |
| | **Before** | **2** | **45** | **7.87** | **0.0012** |

**Table C-32. Significant Differences in Least Squares Means, Before and After the 7-second Condition Prompt, Significant at Bonferroni Adjusted *P* Values for Experiment 2**

| Prompt Condition* Time | Comparison | Estimate | S.E. | *t* Value | Adj *p* Value | Lower Confidence Limit | Upper Confidence Limit |
|---|---|---|---|---|---|---|---|
| **2-second* After** | **2-second* Before** | 0.06 | 0.02 | 2.77 | 0.0400 | 0.02 | 0.10 |
| **7-second* After** | **7-second* Before** | 0.20 | 0.05 | 3.98 | 0.0010 | 0.10 | 0.30 |

**Table C-33. Significant Differences in Least Squares Means Between Prompt Conditions, Both Before and After the Prompt, Significant at Bonferroni Adjusted *P* Values for Experiment 2**

| Prompt Condition* Time | Comparison | Estimate | S.E. | *t* Value | Adj *p* Value | Lower Confidence Limit | Upper Confidence Limit |
|---|---|---|---|---|---|---|---|
| **2-second* Before** | **No Prompts* Before** | 0.38 | 0.11 | 3.49 | 0.0055 | 0.16 | 0.60 |
| **2-second* Before** | **7-second* Before** | 0.27 | 0.08 | 3.28 | 0.010 | 0.10 | 0.43 |

The estimated difference before and after the prompt within the 7-second condition of 0.2 (95 percent confidence interval 0.1 to 0.3) can be back-transformed to 1.22, indicating that participants made about 22 percent more non-driving-related glances immediately after the prompt compared to immediately before the prompt. A similar interpretation can be made of the estimated difference between before and after prompts in the 2-second condition, estimated to be 0.1 (95 percent confidence interval of 0.02 to 0.1), which can be back-transformed to 1.10.

Similar interpretations can be made in the differences between the prompt conditions before the prompt. For example, the estimated difference of 0.38 (95 percent confidence limit 0.38 to 0.88) before the prompt between the 2-second condition and the No Prompts condition (using the randomly sampled time point within this group as the point of comparison) can be back-transformed to 1.46, indicating that immediately after the prompt, those in the 2-second group had 46 percent more non-driving-related glances, on average, than those in the No Prompts group in the same interval before the randomly selected time point. The estimated difference between the 2-second and 7-second conditions of 0.27 (95 percent confidence interval of 0.10 to 0.43) can be back-transformed to 1.31.

In summary, when operators are engaged in a non-driving task, the change in the amount of non-driving-related glances from before to after the prompt may increase if the frequency of the prompts is lowered. Also, when operators receive prompts in the 7-second condition, their frequency of glances may have the potential to regress back to the same level as without prompts, as evidenced by the lack of significant difference between 7-second and No Prompts conditions in the amount of non-driving-related glances when a general time in the No Prompts condition is compared to the 7-second condition just before the prompt sounds. However, this is not the case for the 2-second condition, which maintains the number of glances at a significantly higher level than both other conditions just before the prompt sounds.

## Trust

Trust scale ratings were analyzed with the time point variable ("1" for the first trust scale within a session to "7" for the last trust scale) as the independent variable of interest checked for significance.

Overall, the participants often expressed that they trusted the system. Out of 1,024 total ratings (seven trust scales were given per session), 935 (91 percent) specified at least "slightly agree" (123), with 395 ratings for "moderately agree" and 417 ratings for "strongly agree." There were 42 "neutral" ratings (4 percent), which leaves 47 ratings (4.5 percent) for "disagree," with 5 for "strongly disagree," 13 for "moderately disagree," and 29 for "slightly disagree." A bar plot of total ratings is displayed in Figure C-40.

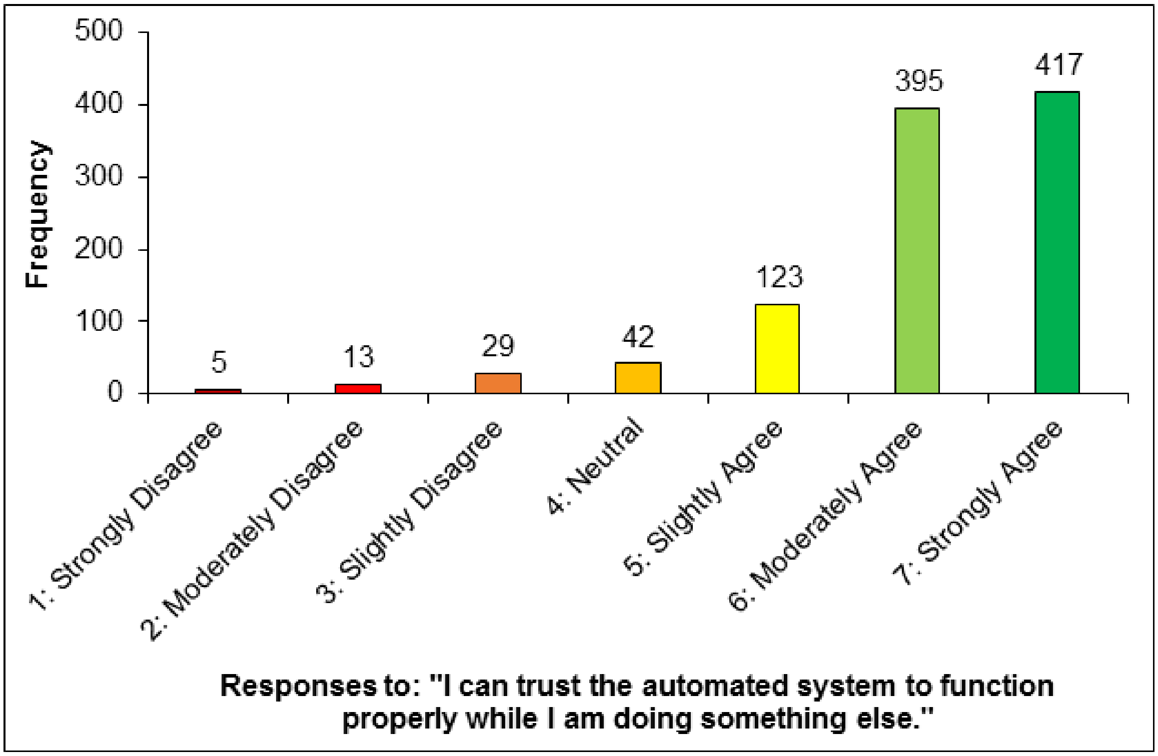



**Figure C-40. Frequency of Rating Levels for Experiment 2**

### *Analysis for Trust Over Time*

<u>Statistical Analysis</u>

A mixed model was run with time, session, condition, and the order in which participants were assigned lane drifts as predictors of trust scale rating, while random intercepts and time slopes were given to account for correlation within participants. Order was used to assess whether the flow of ratings over time was affected by the placement of lane drifts at different points in the experimental run. Analysis results are displayed in Table C-34.

**Table C-34. Analysis of Trust over Time, ANOVA Table, for Experiment 2**

| Effect | Num DF | Den DF | F Value | Pr > F |
|---|---|---|---|---|
| **Time** | 1 | 154 | 1.8 | 0.1822 |
| **Condition** | **2** | **174** | **9.54** | **0.0001** |
| **Order** | 5 | 174 | 1.00 | 0.4183 |
| **Session** | **2** | **174** | **7.91** | **0.0005** |
| **Time*Condition** | 2 | 154 | 0.62 | 0.5382 |
| **Time*Order** | 5 | 154 | 1.8 | 0.1160 |
| **Time*Session** | **2** | **154** | **5.47** | **0.0051** |

Since neither order ($p$ = 0.4183) nor the time*order interaction presented a statistically significant difference ($p$ = 0.1160), the conclusion is that the lane drifts do not appear to affect, positively or negatively, the change of ratings over time. Meanwhile, the time*session interaction ($p$ = 0.0051) is significant, indicating that the change in ratings over time depends on the driving session. The significant differences between the time slope are displayed in Table C-35.

**Table C-35. Significant Differences in Least Squares Means of Trust Change Across Sessions**

| Time* Session | Comparison | Estimate | S.E. | *t* Value | Adj *p* Value | Lower Confidence Limit | Upper Confidence Limit |
|---|---|---|---|---|---|---|---|
| **1** | **3** | 0.1 | 0.03 | 2.95 | 0.0045 | 0.03 | 0.16 |
| **1** | **2** | 0.1 | 0.03 | 2.82 | 0.0162 | 0.03 | 0.16 |

The estimates in Table C-35 indicate that the time slope of ratings in Session 1 are greater than in the other two sessions, implying that ratings start to increase, but the increase is primarily confined to the first session.

Meanwhile, condition is also significant ($p$ = 0.0001), indicating that trust differs between the different conditions. The differences in least squares means are displayed in Table C-36.

**Table C-36. Significant Differences in Trust Across Conditions**

| Condition | Comparison | Estimate | S.E. | *t* Value | Adj *P* Value | Lower Confidence Limit | Upper Confidence Limit |
|---|---|---|---|---|---|---|---|
| **2-second** | **No Prompts** | -0.67 | 0.19 | -3.5 | 0.0018 | -1.14 | -0.21 |
| **7-second** | **No Prompts** | -0.76 | 0.19 | -3.99 | 0.0003 | -1.23 | -0.30 |

The participants experiencing no prompts expressed, on average, higher trust than those who experienced prompts. Specifically, the No Prompts group expressed a mean trust rating 0.67 higher than the 2-second prompt group, and 0.76 higher than the 7-second prompt group.

***Analysis for After-experience Trust Scales***

Means, standard deviations, modes, minimums, and maximums for the trust scales are provided in Table C-37.

**Table C-37. Descriptive Statistics for Responses to the After-experience Trust Scales for Experiment 2**

| Variable | Mean | S.D. | Mode (Frequency) | Minimum | Maximum |
|---|---|---|---|---|---|
| **TS1** | 5.83 | 1.24 | 6 (61) | 1 | 7 |
| **TS2** | 5.55 | 1.52 | 6 (48) | 1 | 7 |
| **TS3** | **4.03** | **2.16** | 7 (35) | 1 | 7 |
| **TS4** | 5.68 | 1.18 | 6 (61) | 2 | 7 |
| **TS5** | 5.82 | 1.13 | 6 (65) | 1 | 7 |
| **TS6** | 5.71 | 1.28 | 6 (53) | 1 | 7 |

Responses to Statement 3, “The automated system gave false alerts,” have a lower mean than was the case for any of the other statements, with a mean of 4.03, compared to the next lowest mean (Statement 2) of 5.55. Additionally, the responses to this statement have a higher amount of variability than was the case for the other statements, with a standard deviation of 2.16 compared to the next highest (Statement 2) with a standard deviation of 1.52. Thus, while participants were largely in agreement over their level of trust in the system after the experiment, they did not always agree regarding whether or not the system gave false alerts.

***Analysis for After-session Satisfaction Scales***

Means, standard deviations, modes, minimums and maximums are provided in Table C-38.

**Table C-38. Descriptive Statistics for Responses to the Satisfaction Scale Survey for Experiment 2**

| Variable | Mean | S.D. | Mode (Frequency) | Minimum | Maximum |
|---|---|---|---|---|---|
| **1** | 6.11 | 1.06 | 7(62) | 2 | 7 |
| **2** | 5.43 | 1.57 | 6(56) | 1 | 7 |
| **3** | 5.74 | 1.17 | 6(67) | 2 | 7 |
| **4** | 6.04 | 0.99 | 6(61) | 3 | 7 |
| **5** | 5.71 | 1.28 | 6(52) | 1 | 7 |
| **6** | 5.6 | 1.58 | 7(55) | 1 | 7 |
| **7** | 5.59 | 1.70 | 7(63) | 1 | 7 |
| **8** | 6.17 | 1.34 | 7(89) | 1 | 7 |

Experiment 2 participants were asked to describe how comfortable they were with the system. As participants responses were not limited to Likert-type responses, a number of self-identified comfort-related terms emerged. Those who did not express a specific comfort level provided general descriptions of their comfort level, for example:

- "Once, after the first couple of laps, it's really simple to use. I was surprised, you know it's a push of a button." *(25–39 age group)*
- "It was pretty easy to use. It wasn't a very hard system to use, but I don't know all of the features that are available. From what you guys let me see, I thought it was easy to use." *(40–54 age group)*
- "I wish I could have done 70 so I could see how it worked." *(55+ age group)*

In terms of the alerts presented, 30 participants specifically noted that the alerts prompted them to regain situational awareness either by figuring out the appropriate response to the alert or situation or by preparing to take manual control. Fifteen participants mentioned the haptic seat alerts in response to this question. The majority noted the benefits of, and their preferences for, the haptic seat alerts (13 responses). While the benefits associated with alerts were voiced, participants (32 responses) also indicated concerns with the alerts, including concerns associated with the visual alerts' ability to adequately attract the user's attention if the user is engaged in a secondary task (6 responses). Note that alerts were fundamentally different than prompts in this study, with regard to their cause, but their differences were not explained to participants. Perhaps owing to the fact that they did not know whether they were receiving an alert or an attention-related prompt, participants voiced concerns in terms of a general confusion regarding the reason for and number of alerts (23 responses), as is illustrated by the following comment:

- "I thought sometimes it was an alert for no reason, you know, that I could see [gave a visual alert] all of a sudden, you know, why did that happen, there was nothing around. So it was curious. I'm curious to see how it would work on something that isn't a continuous roadway, if we were making, you know, turns and curves going straight, you know. I'm curious to see the kind of alerts you got in those kinds of instances." *(55+ age group)*

The lack of understanding regarding the purpose of the alerts led several participants (5 responses) to voice their frustrations with the system, resulting in two participants indicating that they ignored the alerts (i.e., participants responded that when presented with a perceived nuisance alert, they ignored it). As the following response illustrates, the comments associated with this question indicate a need for not only a better understanding of the alerts, but also for a balance in the number of alerts presented.

- "My immediate answer to that question would be 'I felt like it wouldn't shut up.' However; I suppose that's on the better biased end of the spectrum toward safety, rather than it being too quiet. The alerts were deliberate, often, and noticeable." *(18−24 age group)*

These findings suggest that designers of automated systems will need to balance the number of alerts presented. By balancing the number of alerts presented, users may be encouraged to remain vigilant in their monitoring of the system while still having trust in the system to operate effectively. Additional subjective data for Experiment 2 can be found in Appendix F.

# Appendix D. Experiment 3: Expanded Methods, Results, and Analysis

## Experimental Design

An a priori power analysis was conducted to determine the total number of participants required for this study to reach a power of .8 with $\alpha = .05$. The minimum sample size needed to detect a large difference in the dependent variables was calculated. Assuming a large effect size (.8), a total of 16 participants were required for the design, while the experiment included 25 participants. In order to ensure all age groups were represented, participants were recruited based on the four NHTSA age groups (18−24, 25−39, 40−54, and 55 and older; Visual-Manual NHTSA Driver Distraction Guidelines for In-vehicle Devices, 2013).

## Procedure

### Screening

The experimental protocols were reviewed and approved by the Virginia Tech IRB. Participants were recruited from the Blacksburg/Roanoke, Virginia region using VTTI's participant database and through word of mouth. Participants were eligible for inclusion in the study if they met all of the following criteria:

- Valid driver's license, and were at least 18 years old;
- No disqualifying moving violations in the past 3 years;
- Had not previously participated in a study at VTTI involving a "surprise" event;
- Owned a smartphone;
- Normal or corrected-to-normal vision and hearing;
- Able to drive without sunglasses or photochromatic lenses;
- Able to drive an automatic transmission vehicle without accommodation;
- If pregnant, discussed the risks of participating with a physician;
- Reported not taking medications or substances that interfere with driving ability;
- No lingering health effects or any recent heath events, including high propensity towards motion sickness;
- Eligible for employment in the United States; and
- Able to wear closed-toe shoes while driving.

## Greeting

Upon arriving at the test site, each participant was greeted by one of the in-vehicle experimenters. The experimenter escorted the participant to the preparation area and provided the informed consent documentation. The participant's driver's license was examined to ensure validity. Following informed consent, the participant was asked to undergo a basic vision screening and complete the demographics questionnaire. Hearing was not formally assessed; however, each participant was asked to respond to basic questions in the greeting process that determined if normal hearing seemed to be present.

## Participants

Data were collected from 25 participants. The age groupings reflect those proposed in the Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Devices (2013).

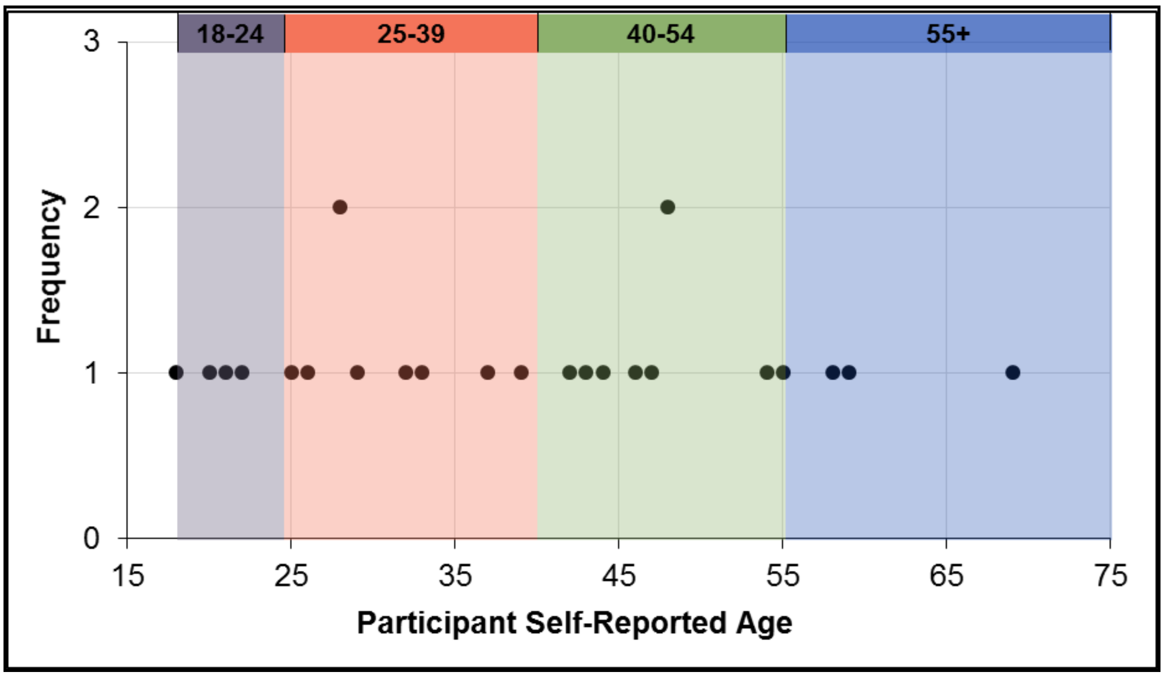



**Figure D-1. Distribution of Experiment 3 Participants Across the NHTSA Age Categories**

In addition, participants were asked to provide basic demographic information (e.g., gender, education level), driving experience (type of vehicles used and annualized miles driven), smartphone use and experience, and automotive technology experience. Participants were also asked about their awareness of automated vehicle technologies.

Smartphone ownership was a requirement to participate in this study. This was to ensure that participants would be comfortable using the technology that would be utilized during the experiment. All of the participants indicated using their smartphone on a daily basis. Figure D-2 depicts the self-reported years of experience with smartphone usage. Participants were asked to report how long they have used a smartphone: 36 percent ($n$ = 9) indicated using a smartphone for less than 2 years; 24 percent ($n$ = 6) indicated using a smartphone for 2 to 4 years; and 40 percent ($n$ = 10) indicated using a smartphone for more than 4 years.

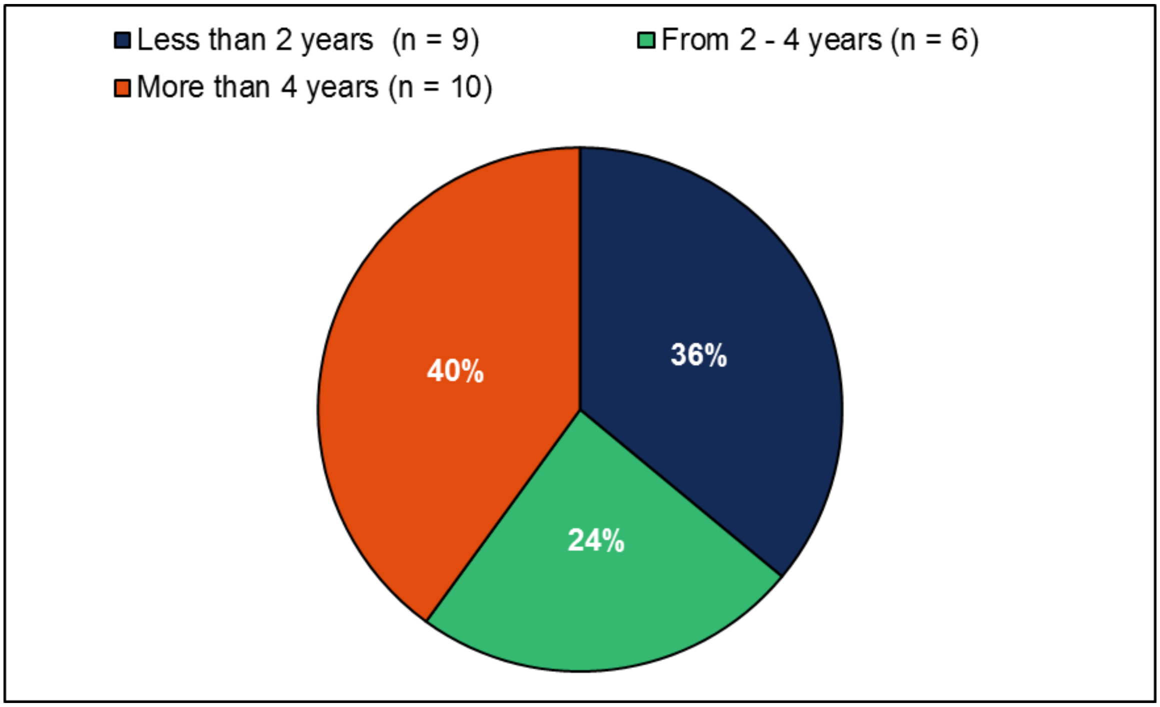



**Figure D-2. Experiment 3 Participants' Experience with Smartphones**

Participants were asked to report any prior experience with CWS, ACC, and LKA systems (Figure D-3). Sixteen percent of participants (*n* = 4) reported having previously driven a vehicle with a CWS, 8 percent of participants (*n* = 2) reported having previously driven a vehicle equipped with ACC, and 16 percent of participants (*n* = 4) reported having previously driven a vehicle with an LKA system. Forty-four percent of participants (n = 11) indicated being aware of self-driving/automated vehicles. Of the 11 participants who were previously aware of self-driving vehicles, 10 specifically mentioned the Google self-driving car.

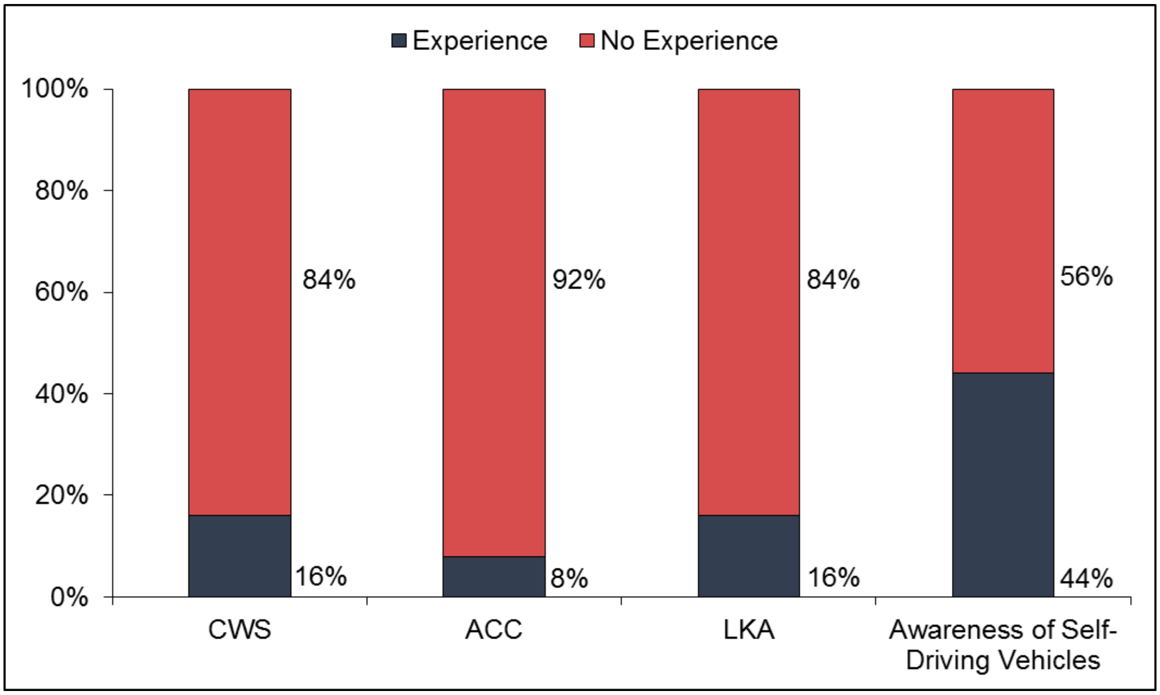



**Figure D-3. Experiment 3 Participants' Experience with and Awareness of Automated Technology for Vehicles**

# Results

Statistical analysis was performed to determine if there were any significant differences in the independent variables. This analysis aims to determine if there was evidence of differences in operator behavior across time (represented by session), between different alert types, or if the effect of the alert type depended on the session.

Since each participant was measured three times, this experiment used an analysis technique that accounted for possible correlations between instances that come from the same participants. A covariance pattern regression model was used with an unstructured error covariance with alert type as the repeated measures effect. This model accounts for varying correlations between the different alerts for each participant, as well as different variances within the different alert types. Parameters were estimated using restricted maximum likelihood (R.E.M.L.). Residual diagnostics were checked via graphical inspection to determine the adequacy of the assumptions of normality and random pattern of the error terms when accounting for the covariance structure used. If there were no drastic visual departures from normality or randomness, the response variable was used as is. Otherwise, the response was transformed using the logarithmic transformation, which is a data transformation commonly used to stabilize variances (Kutner et al., 2005).

Models for time to regain control, time to activate automation, time to release control of steering, and time to resume non-driving task were fit using alert type and session, and their two-way interaction. Age was also included as a continuous covariate, as well as the two-way interactions and the three-way interaction between age and the categorical predictors. To maximize the ability to detect significant differences in alert type and session (the primary variables of interest), insignificant terms involving age were removed. However, alert type and session, and their interaction, remained and were reported. Models for the time to react were fit with alert type, session, age, and an additional independent variable indicating whether or not the participant was looking forward at the time of the alert ("yes" or "no"). All interactions were included at first, but any insignificant terms related to age or looking forward were removed. If a significant interaction involving age or looking forward occurred with alert type or session, those were reported, as it would be of interest to note if the time effect or alert effect varied across these variables. However, the main effects of age and looking forward were not reported. The looking forward variable was not used with the other analyses because it only applies directly to time to react.

## Operator Behavior

In general, a key question is whether or not there are significant differences between Imminent alerts with and without an external threat in these measures. However, it may also be of interest to determine whether Staged alerts differed from Imminent alerts in some situations. For example, it would be expected that participants take more time to regain control after the onset of a Staged alert, due to the alert being less urgent in nature. However, it would be a noteworthy result if participants, in fact, did not take significantly longer to regain control after a Staged alert.

### *Analysis for Time to React to Alert*

The time it took for participants to react was explored for each type of alert. **Figure D-4** suggests that participants may take longer to react to the Staged alert compared to both Imminent alerts. Consistent with other research on reaction time, this may be reflective of the lower level of urgency portrayed by the Staged alert (Green, 2000; Summala, 2000). The mean time to react to the Staged alert across all three driving sessions was 1.2 s (S.E. = 0.2 s), 0.7 s for the Imminent–No External Threat alert (S.E. = 0.04 s), and 0.7 s for the Imminent–External Threat alert (S.E. = 0.06 s).

Additionally, the time it took participants to react to the alerts was examined over time. Figure D-4 depicts the mean time to react to all three alert types in each session. During Session 1, the mean time to react to the alert was 0.8 s (S.D. = 0.1 s); during Session 2, the mean time to react to the alert was 0.8 s (S.E. = 0.1 s); and during Session 3, the mean time to react to the alert was 1.0 s (S.E. = 0.2 s). While participants took longer to react to the alert after being exposed to the automated system for over an hour and having experienced two alerts prior to this point, this difference was not statistically significant.

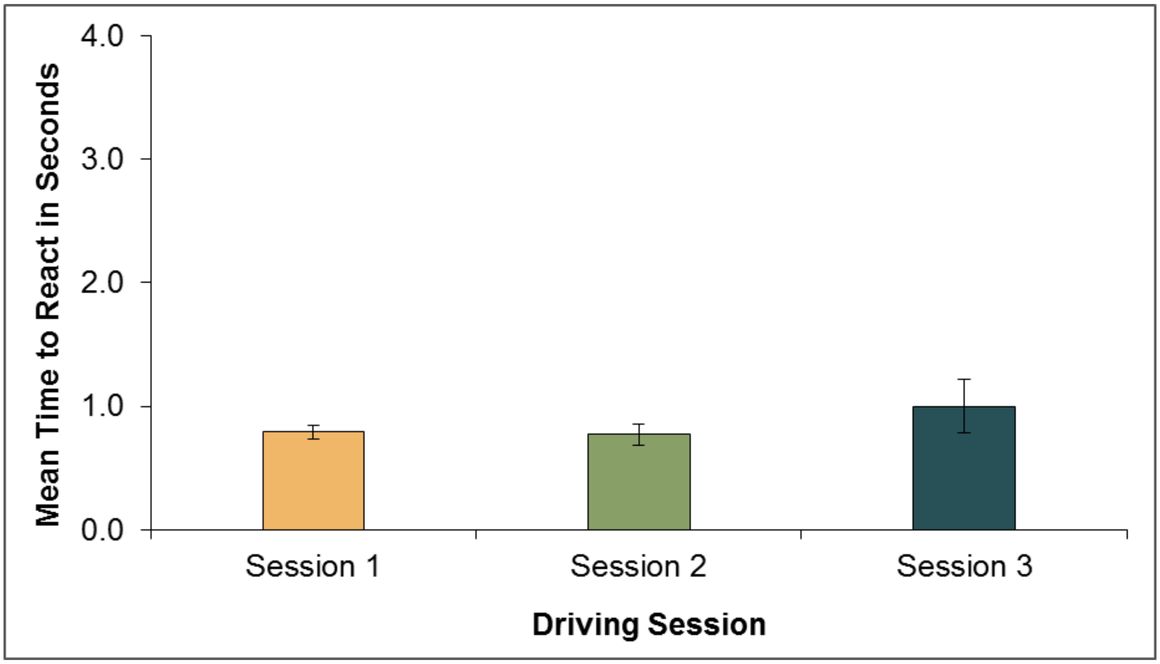



**Figure D-4. Mean and Standard Error Bar Plots, Time to React, by Session for Experiment 3**

Figure D-5 displayed below depicts the boxplots of time to react for each alert type stratified by driving session. The lowest mean time to react during the Staged alert (1.0 second during Session 1) was higher than the highest mean time to react during both the Imminent alert types across all three driving sessions (0.8 s for the Imminent–No External Threat alert during Session 1). During the second and third driving sessions, the time it took for participants to react to the Staged alert was more variable than during both Imminent alerts. However, when participants experienced the Staged alert during the first driving session, the variability in the time to react was similar to when participants experienced the Imminent alerts. The S.D. for the Staged alert was 0.6 s in Session 2 and 1.6 s in Session 3 (influenced by two outliers), while the highest S.D. for both Imminent alert types in these two sessions was 0.2 s (this was the case for the Imminent–External Threat in Session 2 and for both of the Imminent alerts in Session 3). Additionally, note that there does not appear to be a consistent trend separating the Imminent–No External Threat alert from the Imminent–External Threat alert in terms of the time it took participants to react. The presence or absence of an external threat did not appear to impact the time it took for participants to react during the higher-urgency Imminent alerts. This may suggest that operators of automated vehicles may take longer to react to a lower-urgency Staged alert compared to a higher-urgency Imminent alert attributed to either the presence of an external threat or an internal system failure.

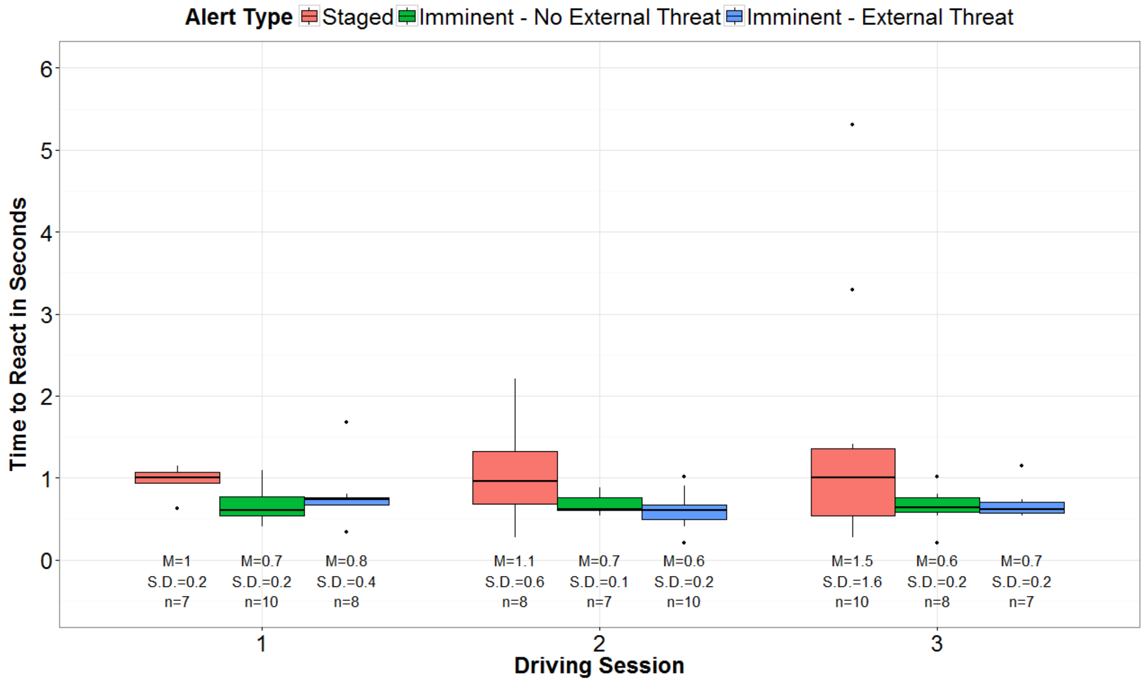



**Figure D-5. Boxplots of Time to React Stratified by Session, and Alert Type for Experiment 3**

Also of interest was the impact of non-driving task engagement on the time it took for participants to react to the alert. Non-driving tasks included the following: talking or interacting with the in-vehicle experimenters, interacting with the tablet computer, and interacting with their own smartphone. At the time at which the alert was first received, in 55 percent of the instances ($n_i$ = 41), participants were observed interacting with the tablet computer; in 27 percent of the instances ($n_i$ = 20), participants were observed interacting with their smartphone; in 1 percent of the instances ($n_i$ = 1), participants were talking with the in-vehicle experimenter; and in 17 percent of the instances ($n_i$ = 13), participants were not engaged in a non-driving task. Figure D-6 depicts the frequency with which participants were observed engaging in the aforementioned non-driving tasks.

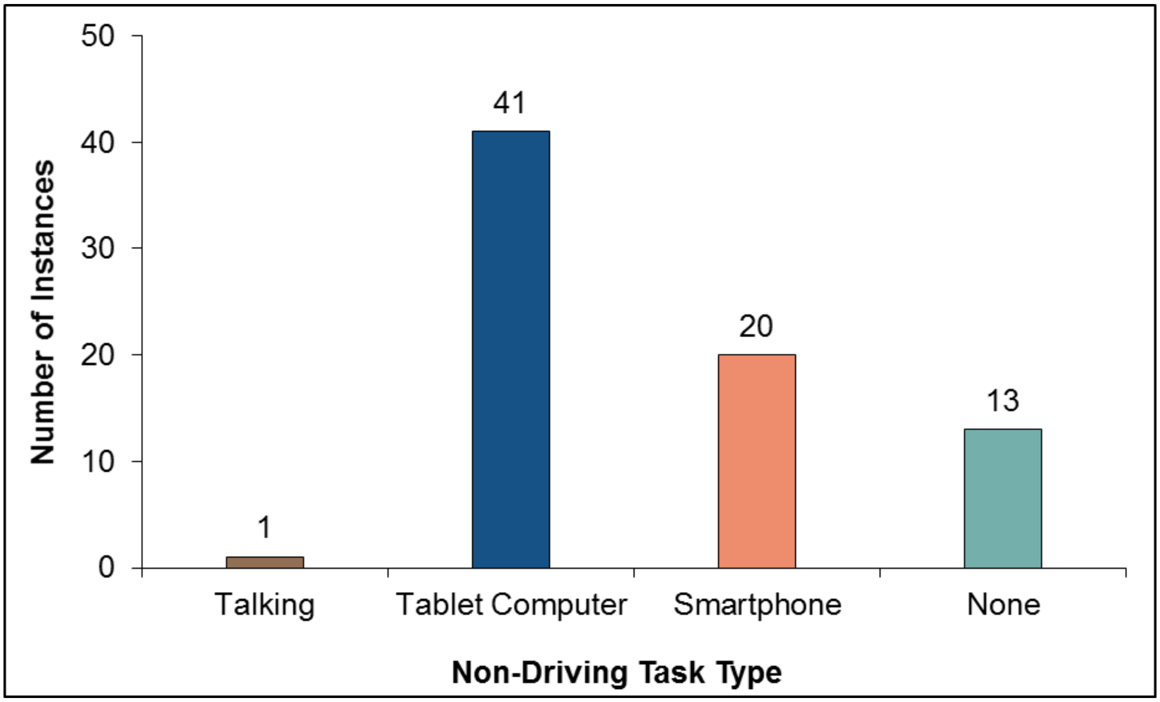



**Figure D-6. Non-driving Task Type Observed at the Alert Onset for Experiment 3**

The researchers hypothesized that participants who were engaged in a non-driving task at the time at which the alert was first received, would take longer to react to the alert compared to participants who were not engaged in a non-driving task at this point in time. There were 13 instances in which participants were not engaged in a non-driving task at the time at which the alert was first received; during the remaining 62 instances, the participants were engaged in a non-driving task at this point in time. Figure D-7 compares the mean time to react to the alert during instances in which participants were engaged in a non-driving task with the mean time to react to the alert during instances in which participants were not engaged in a non-driving task. Those engaged in a non-driving task took longer to react to the alert compared to those who were not engaged in a non-driving task. Participants who were engaged in a non-driving task took a mean of 0.9 s to react to the alert (S.D. = 0.7 s), while participants who were not engaged in a non-driving task took a mean of 0.7 s to react to the alert (S.D. = 0.3 s). However, this difference was not statistically significant.

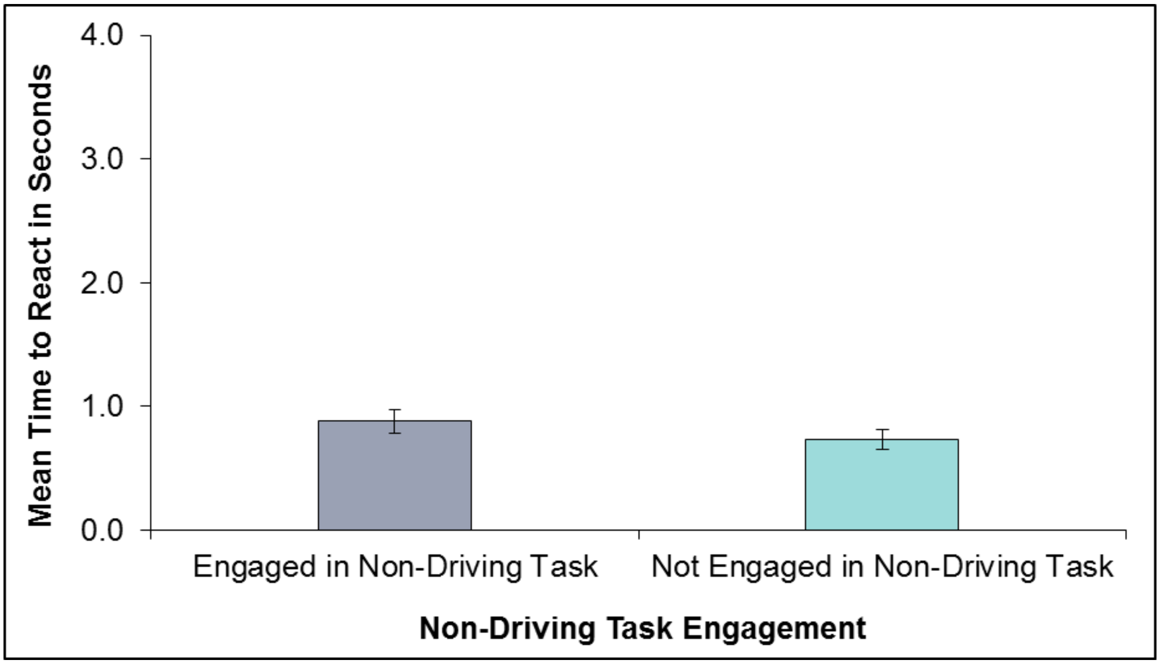



**Figure D-7. Mean and Standard Error Bar Plots, Time to React, by Non-driving Task Engagement for Experiment 3**

Additionally, the time it took for participants to react to the alert may be influenced by their glance location at the time at which the alert was first received. Participants were divided into two groups: those looking forward and those not looking forward at the time of the alert. Out of 75 instances, there were 19 instances of participants who were observed looking forward and 56 instances of participants who were observed looking elsewhere. Mean and error bar plots divided between these two groups are displayed in Figure D-8.

During the Staged alert, participants were only looking forward ($n$ = 5) in Session 3. For these five instances in which the participant was looking forward at the onset of the Staged alert, the lowest time to react was 1.0 s, which was higher than all but two of the instances in which participants were looking forward at the alert onset. All of the other instances in which participants were looking forward at the onset of the Staged alert were higher than all of the other 14 instances in which participants were looking forward at the alert onset. This was not observed with participants who were not looking forward at the alert onset. This may suggest that operators who are looking forward at the onset of a lower-urgency Staged alert may take longer to react to the alert compared to both the instances in which the operator is not looking forward at the onset of a Staged alert and the instances in which the operator receives a higher-urgency Imminent alert, regardless of whether or not he or she is looking forward at the alert onset. However, whether or not this trend would hold consistently across the sessions—or if this is merely confined to Session 3—is unclear, because none of the participants who experienced the Staged alert during Session 1 or Session 2 were looking forward at the onset of the alert.

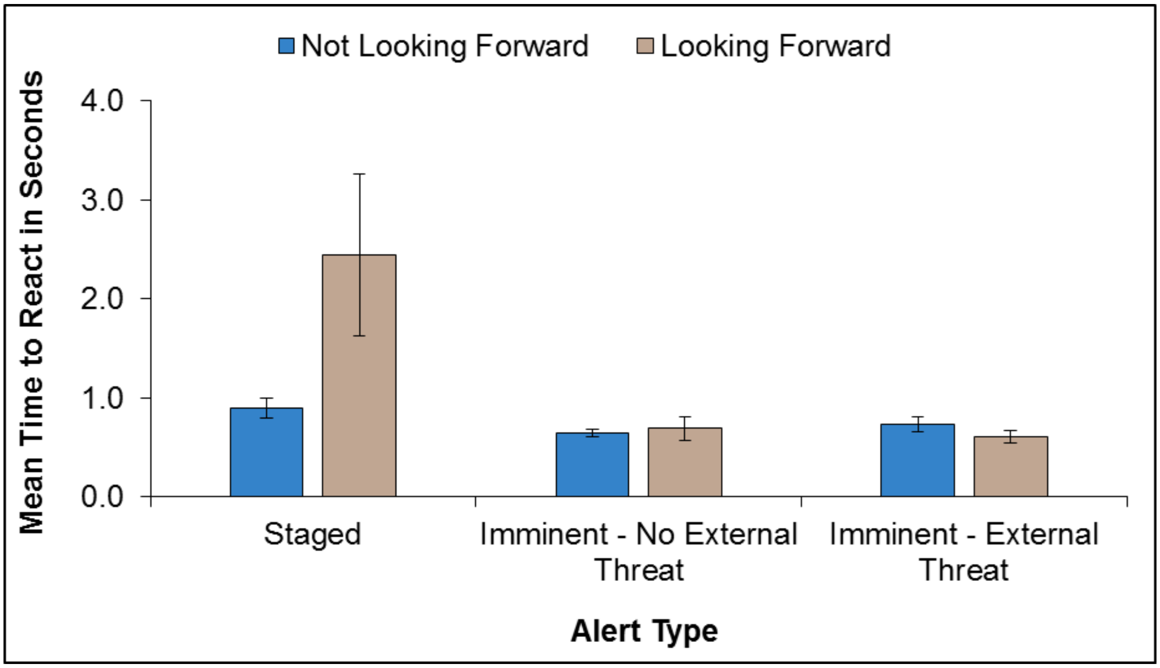



**Figure D-8. Mean and Standard Error Bar Plots, Time to React, Not Looking Forward, by Alert Type for Experiment 3**

Figure D-9 and Figure D-10 illustrate the sequence of actions taken by participants who were looking forward and not looking forward at the onset of the alert, respectively. For those looking forward at the time of the alert, the majority (73.7%) prepared to take the steering wheel as their first action, and 63.1% prepared to use a pedal as their second action (Figure D-9). For those not looking forward at the time of the alert, the majority (87.5%) looked forward as their first action, and 51.8% prepared to take the steering wheel as their second action (Figure D-10).

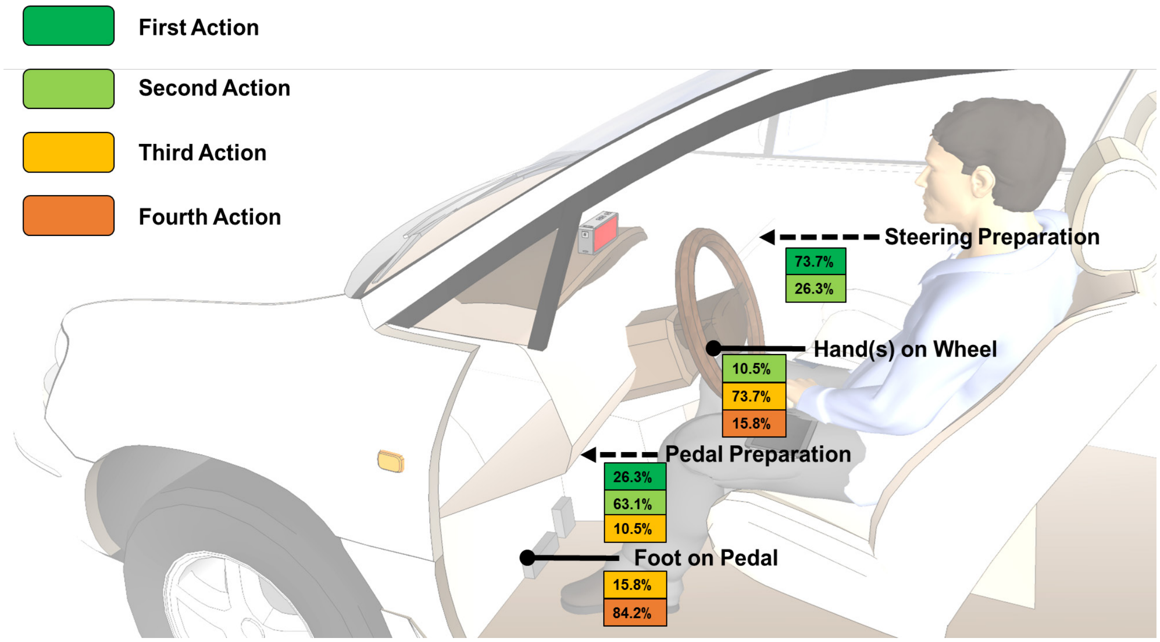



**Figure D-9. Action Sequence for Participants Who Were Looking Forward at the Time of the Alert for Experiment 3**

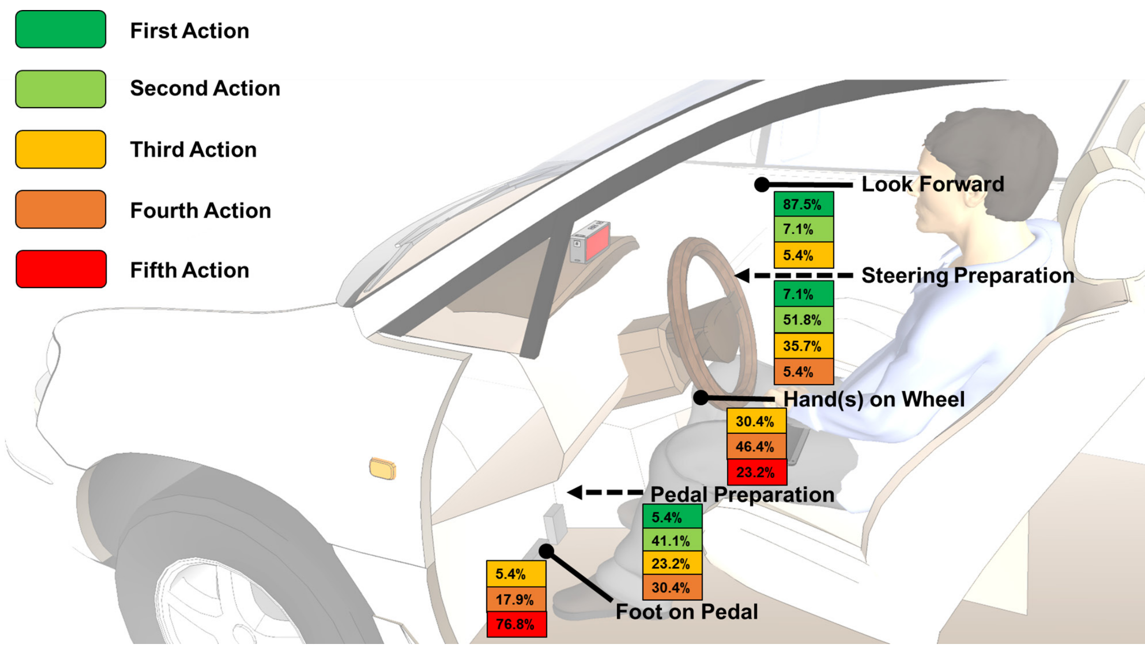



**Figure D-10. Action Sequence for Participants Who Were Not Looking Forward at the Time of the Alert for Experiment 3**

Statistical Analysis

Alert type, driving session, and glance location at the onset of the alert were used as independent variables for this analysis. Non-driving task engagement had also been included as an independent variable; however, this variable was not significant and was subsequently removed from the model. The final model is displayed in Table D-1; significant results are shown in bold.

**Table D-1. Time to React, ANOVA Table for Experiment 3**

| Effect | Num DF | Den DF | F Value | Pr > F |
|---|---|---|---|---|
| **Alert Type** | **2** | **35.8** | **8.89** | **0.0007** |
| **Session** | 2 | 27.2 | 0.52 | 0.6005 |
| **Forward** | **1** | **25.4** | **10.07** | **0.0039** |
| **Alert Type*Session** | 4 | 39 | 0.45 | 0.7712 |
| **Alert Type*Forward** | **2** | **35.4** | **5.39** | **0.0091** |

The alert type/forward interaction (*p* = 0.0091) indicates that the effect of looking forward versus not looking forward at the time of the alert type varies across alerts. A plot of least squares means with standard error bars for time to react is displayed in Figure D-11 for the six different combinations of alert type and looking forward status.

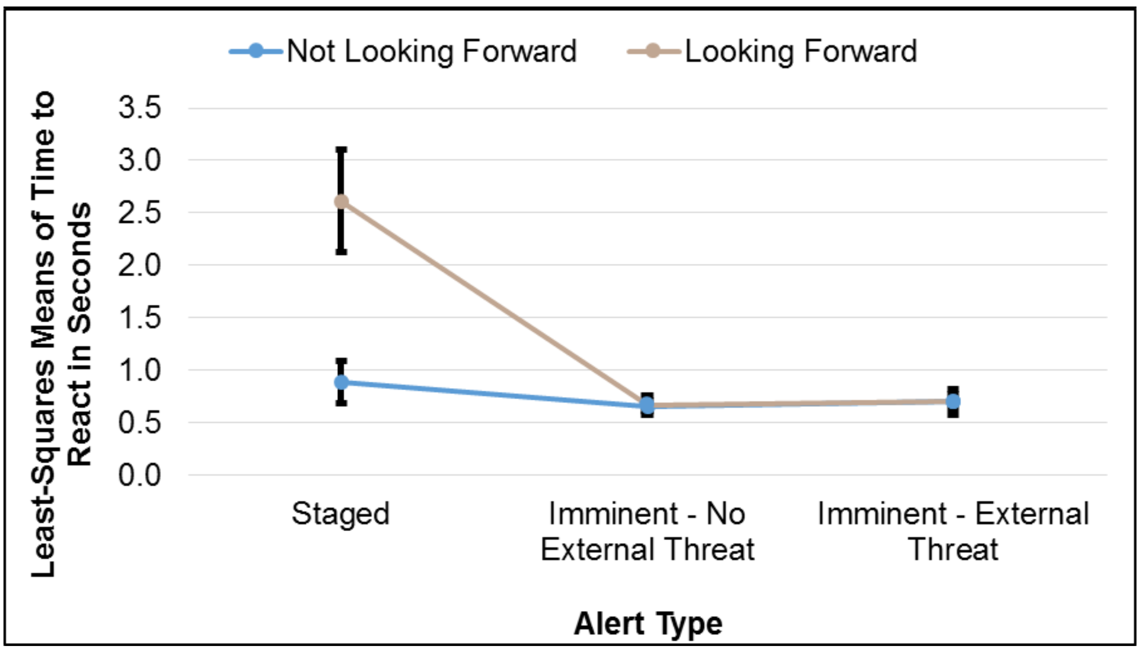



**Figure D-11. Plots of Least Squares Means for Reaction Time with Standard Error Bars, Combinations of Alert Type and Looking Forward Status for Experiment 3**

The bottom of the confidence band for the Staged/looking forward group is higher than the top of the confidence band for all other groups. This suggests that the group of looking forward participants experiencing the Staged alert had higher times to react than all other groups.

To test which levels of alert type differed in reaction times between looking forward and not looking forward, and which levels of looking forward differed in reaction times between alert types, effect slicing was performed to check for significance within each level of each of these two variables. The Bonferroni adjusted *p* value of less than 0.01 was used as the significance level. Results are displayed in Table D-2.

**Table D-2. Significant Differences of Time to React Within Alert Type and Looking Forward for Experiment 3**

| Alert Type | Looking Forward | Num DF | Den DF | F Value | Pr > F |
|---|---|---|---|---|---|
| **Staged** | | **1** | **21.2** | **11.11** | **0.0031** |
| **Imminent–No External Threat** | | 1 | 20.5 | 0.22 | 0.6470 |
| **Imminent–External Threat** | | 1 | 20.8 | 0.12 | 0.7366 |
| | **Not Looking Forward** | 2 | 35.6 | 0.69 | 0.5069 |
| | **Looking Forward** | **2** | **35.7** | **8.01** | **0.0013** |

The results in Table D-2 suggest that with Staged alerts, there is a significant difference in time to react when the participant looks forward compared to when the participant does not look forward ($p = 0.0031$). Similarly, the results suggest that within the group of instances in which participants are looking forward at the time of the alert, there is a difference in reaction times between alert types ($p = 0.0013$). The difference between not looking forward and looking forward within Staged alerts is analyzed in Table D-3, while differences between alert types within forward-looking instances are displayed in Table D-4.

**Table D-3. Difference in Time to React Between Forward-Looking and Not-Forward-Looking Group, Within Staged Alerts, For Experiment 3**

| Alert Type | Estimate | S.E. | *t* Value | Adj *P* Value | Lower Confidence Limit | Upper Confidence Limit |
|---|---|---|---|---|---|---|
| **Staged** | 1.89 | 0.57 | 3.33 | 0.0031 | 0.71 | 3.07 |

**Table D-4. Significant Differences in Reaction Times Between Alert Types, Within Forward-Looking Instances, for Experiment 3**

| Alert Type | Comparison | Estimate | S.E. | *t* Value | Adj *p* Value | Lower Confidence Limit | Upper Confidence Limit |
|---|---|---|---|---|---|---|---|
| **Staged** | **Imminent–No External Threat** | 2.05 | 0.51 | 4.00 | 0.0264 | 0.99 | 3.11 |
| **Staged** | **Imminent–External Threat** | 2.08 | 0.52 | 4.00 | 0.0243 | 1.01 | 3.16 |

The prominent finding is that when participants received a Staged alert and were looking forward at the time of the alert, they reacted significantly more slowly than for other alert types, or if they were not looking forward. For Staged alerts, participants reacted 1.89 s more slowly, on average, if they were looking forward than if they were not looking forward (95 percent confidence interval 0.71 to 3.07). Additionally, if the participant was looking forward at the time of the alert, they reacted significantly more slowly if the alert was Staged versus Imminent without an external threat (estimated difference 2.05 s, 95 percent confidence interval 0.99 to 3.11) and versus Imminent with an external threat (estimated difference 2.08 s, 95 percent confidence interval 1.01 to 3.16) The difference between Imminent alerts (with versus without an external threat) was not statistically significant.

In summary, there is some evidence of a relationship between the type of alert and the time to react to the alert, but this relationship is affected by whether or not an operator is looking forward at the time of the alert. Specifically, the data suggest that operators who are looking forward may take longer to react when they are experiencing a Staged alert. Whether or not this finding is consistent across time is a question that needs further research to answer. Meanwhile, there is no statistically significant evidence that operators who experience an Imminent alert will significantly change their reaction times if there is an external threat present at the time of the alert, nor is there evidence of a significant shift in the time to react to alerts across sessions.

***Analysis for Time to Regain Control***

The researchers examined the relationship between the alert type and the time it took the participant to regain control of the vehicle. Figure D-12 depicts the mean time to regain control, following the onset of the alert, across the three alert types. Notice that during the lower-urgency Staged alerts, participants took longer to regain control of the vehicle after the onset of the alert compared to when they experienced the higher-urgency Imminent alerts. However, in the case of Staged alerts, participants took control well before the alert progressed to the imminent phase. During the Staged alerts, participants took a mean of 17.0 s to regain control (S.D. = 6.2 s), which was equivalent to 23 s before the Staged alert progressed to the imminent phase. In fact, the 17 s was still in the informational phase of the alert, where participants were advised that they should be prepared to take control (after 20 s, the Staged alert would have progressed to the first Cautionary phase, where the participants were instructed to take control). Meanwhile, during the Imminent–No External Threat alert, participants took, on average, 2.3 s (S.D = 0.9 s), and during the Imminent–External Threat alert, participants took, on average, 2.1 s (S.D. = 0.7 s).

Additionally, the time it took for participants to regain control of the vehicle after the onset of the alert was examined over time. Figure D-12 illustrates the mean time to regain control for each driving session. The mean times to regain control across the three driving sessions were more evenly balanced in comparison to the mean times to regain control across the three alert types. During Session 1, the mean time to regain control was 6.8 s (S.E. = 1.6 s), 6.1 s during Session 2 (S.E. = 1.5 s), and 8.4 s during Session 3 (S.E. = 1.7 s). This may suggest that, upon receiving an alert, the time it takes an operator to regain control of the vehicle is not influenced by the amount of time they have been exposed to automated technology. However, whether or not driving session is statistically significantly has yet to be determined, and is discussed at a later point in this section.

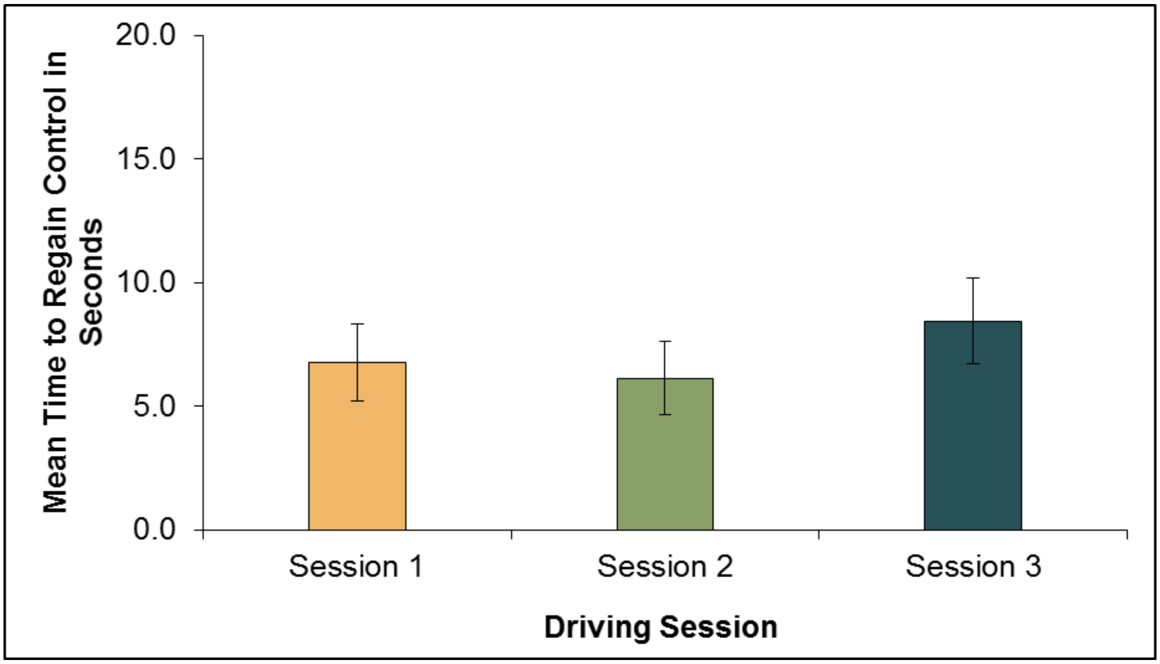



**Figure D-12. Mean and Standard Error Bar Plots, Time to Regain Control, by Session for Experiment 3**

Figure D-13 provides boxplots of time to regain control after the onset of the alert for the three alert types, stratified by the three driving sessions. Within each driving session, the mean times to regain control during the Staged alert (18.4 s in Session 1, 14.5 s in Session 2, and 17.9 s in Session 3) were at least 11.7 s higher than the mean times to regain control during both Imminent alerts. (The highest mean for an Imminent alert is 2.8 s for the Imminent–External Threat alert in Session 3). All of these means for the Staged alerts across sessions are prior to the point at which the alert would have progressed to the second phase (i.e., the first cautionary phase). These means were more than 20 s before the alert would have progressed to the imminent phase (Phase 4 of the Staged alert). Additionally, the time to regain control after the onset of the alert was more variable during the Staged alert than it was during both Imminent alerts. The lowest S.D. for the Staged alerts was 4.6 s in Session 1, while the highest S.D. for the Imminent alerts was 1.4 s for the Imminent–No External Threat alert in Session 2. There did not appear to be any difference in the time it took participants to regain control during the Imminent–No External Threat alert and the Imminent–External Threat alert. (Note that there was one participant who attempted to regain control, but failed to successfully deactivate the automation during the Imminent–

External Threat alert. This participant attempted to turn off the automation by pressing the brake pedal; however, insufficient pressure was applied and the vehicle ultimately came to a stop prior to reaching the box.)

Overall, the data suggest that operators may take longer to regain control of the vehicle during a lower-urgency Staged alert than during a higher-urgency Imminent alert, although they may still regain control with more than 20 s to spare, on average, before the Staged alert progresses to the imminent phase. Additionally, the data suggest that the absence or presence of an external threat did not affect the time it takes an operator to regain control during a high-urgency Imminent alert. (Note that there was one participant who attempted to regain control, but failed to successfully deactivate the automation during the Imminent–External Threat alert. This participant attempted to turn off the automation by pressing the brake pedal; however, insufficient pressure was applied and the vehicle ultimately came to a stop prior to reaching the box.)

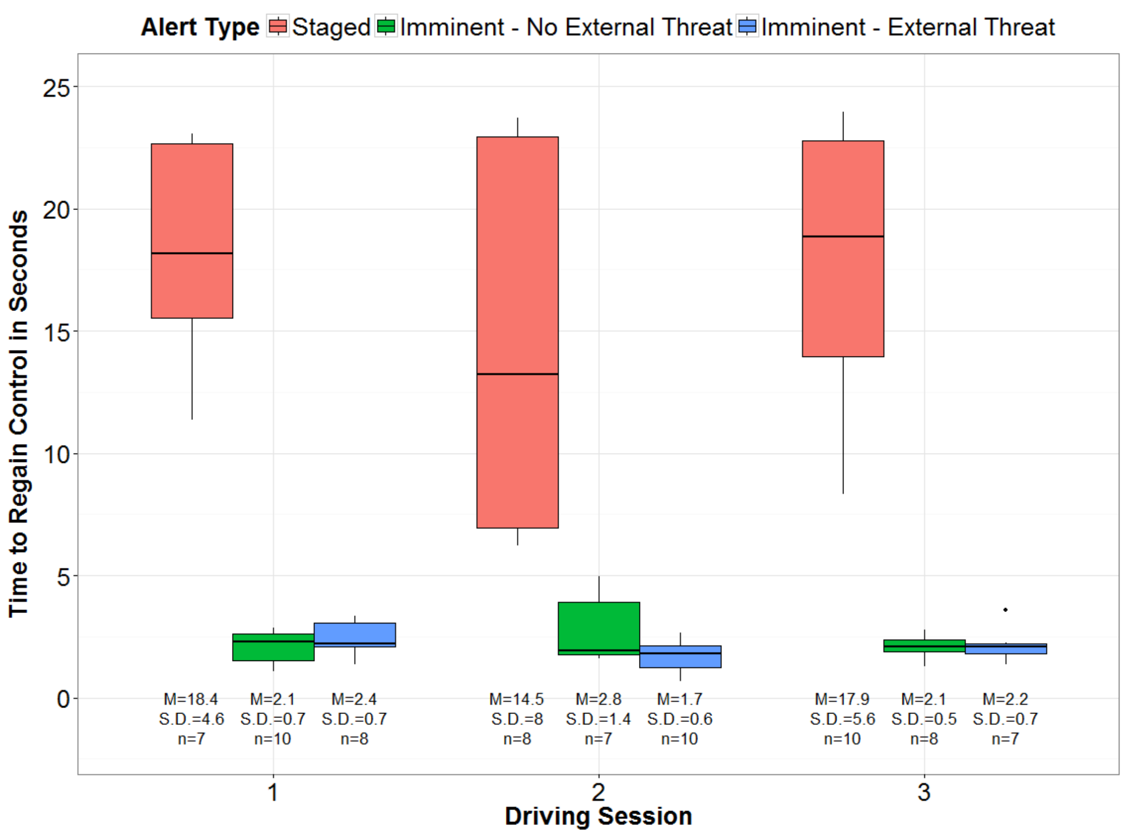



**Figure D-13. Boxplots of Time to Regain Control, Stratified by Session and Alert Type for Experiment 3**

<u>Statistical Analysis</u>

This analysis compared the time to regain control after the onset of the alert between the three alert types. The analysis for time to regain control is displayed in Table D-5.

**Table D-5. Time to Regain Control, ANOVA Table for Experiment 3**

| Effect | Num DF | Den DF | F Value | Pr > F |
|---|---|---|---|---|
| **Alert Type** | **2** | **21** | **68.18** | **<0.0001** |
| **Session** | 2 | 22.5 | 0.87 | 0.4312 |
| **Session*Alert Type** | 4 | 25.8 | 1.93 | 0.1368 |

The results show that alert type achieved significance with $p < 0.0001$. Thus, the data provide evidence that, on average, at least one level of alert type differs from at least one other level of alert type in time to regain control after the onset of the alert. Both session ($p = 0.4312$) and the alert type*session interaction ($p = 0.1368$) were not significant, indicating that there is no statistically significant evidence that, on average, time to regain control varied across sessions, nor that the variation across sessions depended on the alert type. To visualize the differences in the alert types, a plot of least squares means for time to regain control in the three alert types is displayed in Figure D-14.

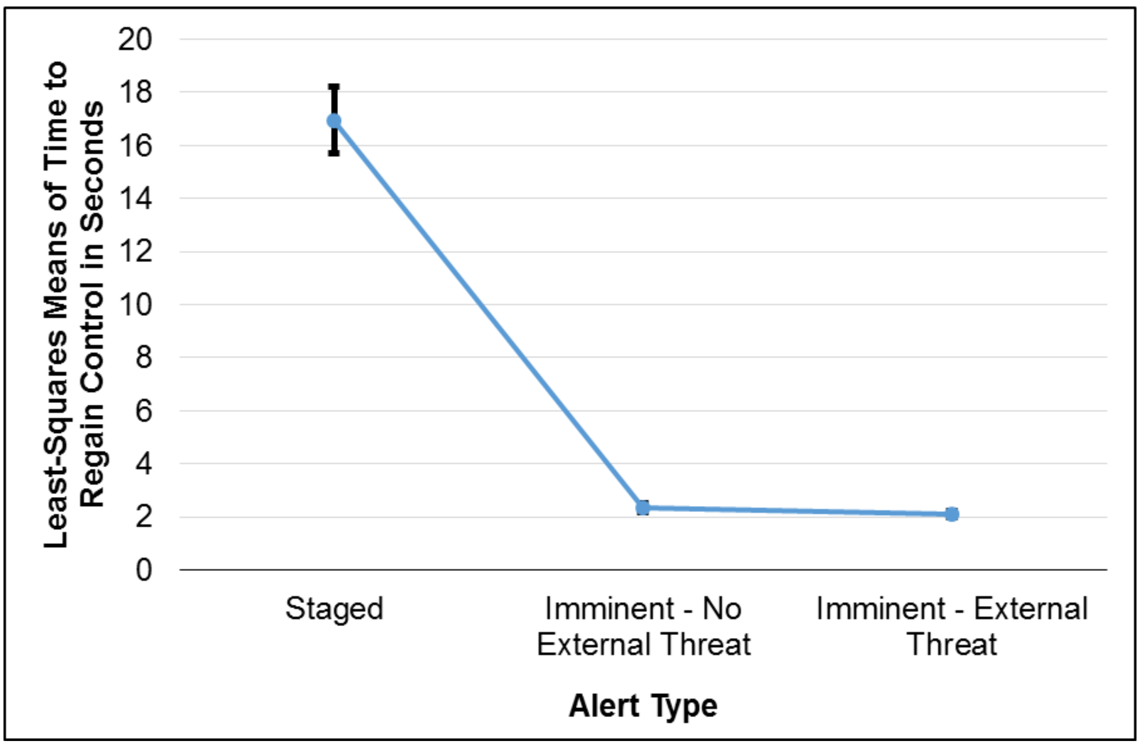



**Figure D-14. Plot of Least Squares Means for Time to Regain Control with 95 Percent Confidence Bands, Alert Type for Experiment 3**

Figure D-14 reveals that the participants generally took longer to regain control of the vehicle after receiving the Staged alert than they did after receiving either of the Imminent alerts. The figure also indicates that participants, on average, took similar amounts of time to regain control after the Imminent–No External Threat alert compared to the Imminent–External Threat alert. This suggests that there may be a significant difference in the time to regain control after the onset of the alert between the Staged alert and the Imminent alerts, but not between the two Imminent alerts. To determine if the Staged alerts indeed have significantly higher times to regain control after the onset of the alert as compared to the Imminent alerts, a post hoc analysis was conducted testing the differences in least squares means of time to regain control between the different alerts using the Bonferroni adjustment. The differences in least squares means that were statistically significant are displayed in Table D-6.

**Table D-6. Time to Regain Control Differences in Least Squares Means, Alert Type, Significant at Bonferroni Adjusted *P* Values for Experiment 3**

| Comparison | Estimate | S.E. | *t* Value | Adjusted *p* Value | Lower Confidence Limit | Upper Confidence Limit |
|---|---|---|---|---|---|---|
| **Staged versus Imminent–No External Threat** | 14.59 | 1.24 | 11.75 | <0.0001 | 12.02 | 17.17 |
| **Staged versus Imminent–External Threat** | 14.84 | 1.24 | 11.93 | <0.0001 | 11.26 | 17.41 |

The Staged alerts did indeed differ significantly from each of the two Imminent alerts in time to regain control, while the two Imminent alerts did not differ significantly from each other. The estimate of "Staged versus Imminent–No External Threat" in Table D-6 is 14.59 s (95 percent confidence interval of 12.02 to 17.17). The conclusion is that there is evidence that time to regain control is higher, on average, for the Staged alert than it is for the Imminent–No External Threat alert. Similarly, the estimate of the difference between the Staged alert and the Imminent–External Threat alert is 14.84 s (95 percent confidence interval of 11.26 to 17.41). Hence, there is evidence that the Staged alert had a higher mean time to regain control than did the Imminent–External Threat alert. Both of these results suggest that operators may regain control of the vehicle from the automated system after the onset of the alert more slowly when prompted by a Staged alert than they will when prompted by an Imminent alert. This relationship also does not appear to depend on session.

It should be noted that, although participants regained control significantly more slowly after the onset of the Staged alert than after the onset of an Imminent alert, the participants still regained control, on average, with over 20 s to spare before the Staged alert progressed to the imminent phase. Thus, whether the slower time to regain control after the onset of the Staged alert is desirable will depend on the context of the on-road situation. Meanwhile, for the comparison between the two Imminent alerts, the results suggest that an external threat will not impact the amount of time an operator needs to regain control of the vehicle after experiencing an Imminent alert.

Another topic of interest related to the Staged alert was how many participants regained control before the Staged alert progressed to the imminent phase. If the participant failed to successfully regain manual control of the vehicle within 40 s, the Staged alert would progress from the cautionary phase to the imminent phase. Thus, the participants' Staged alert times were examined to determine how many participants regained control before 40 s; that is, before the Staged alert reached the imminent alert phase. A plot of each participant's time to regain control after a Staged alert (relative to the 40-second limit before the Imminent alert goes off) is displayed in Figure D-15 (ordered by time to regain control). The four phases of the Staged Alert, with their respective lengths of time, are shown in this figure.

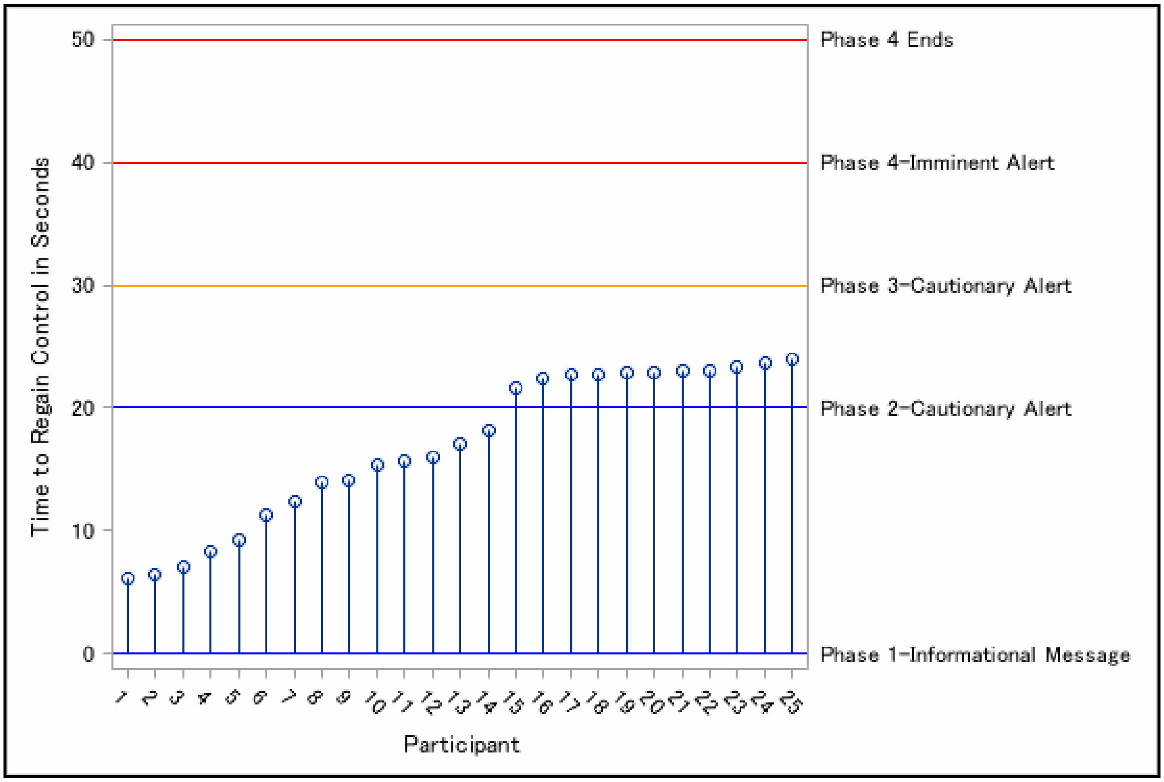



**Figure D-15. Plot of Times to Regain Control Through the Phases of a Staged Alert for Experiment 3**

The highest time to regain control during the Staged alert was 23.9 s, 16.1 s short of the 40-second time limit before the imminent phase of the alert would have commenced. This provides evidence that operators may potentially regain control of the vehicle before the Staged alert progresses from the cautionary phase to the imminent phase of the alert. Note that during the Staged alert, all participants regained control prior to 30 s and, thus, none of the participants experienced the second cautionary alert phase, which would have been a message instructing the participant to take control of the vehicle "now." Fourteen of the 25 participants regained control during the informational phase (Phase 1) of the Staged alert.

***Description for Method Used to Regain Control/Cancel Automation***

In addition to analyzing the time it took the participants to react and to regain control, and categorizing the participants' performance, the particular method each participant initially used to attempt to regain manual control was of interest. The participants could regain control of the vehicle from automation by using one of the following methods: pressing the off button on the steering wheel, moving the steering wheel, pressing the brake, or pressing the throttle.

When presented with a Staged alert, most participants used the off button on the steering wheel to attempt to regain manual control (i.e., following the animation presented as part of the Staged alert). Eleven participants waited until after the informational phase of the Staged alert was over before they attempted to regain manual control of the vehicle. After 20 s, the alert progressed to Phase 2 of the Staged alert, the first cautionary phase. This second phase of the Staged alert consisted of an auditory and visual component. Participants received verbal instructions to turn off the autodrive system. This message was also relayed on the HMI. Additionally, the HMI changed to an animated display depicting an operator deactivating the system by using the off button. Interestingly, 10 of the 11 participants who experienced the animated display used the off button to regain manual control. Figure D-16 depicts both the percentage and the number of participants who attempted to regain manual control during the Staged alert by pressing the off button (44 percent; *n* = 11), moving the steering wheel (0 percent; *n* = 0), pressing the brake (36 percent; *n* = 9), and pressing the throttle (20 percent; *n* = 5).

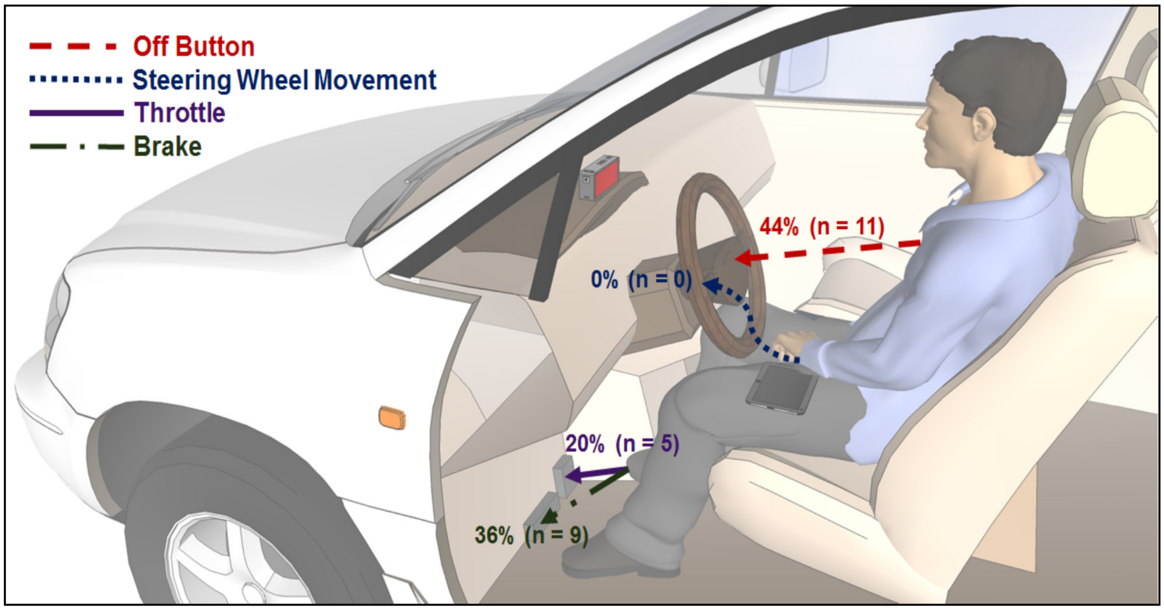



**Figure D-16. Method Used to Regain Control During the Staged Alert in Experiment 3**

For the Imminent alerts, the brake pedal was the preferred method followed by pressing the off button. In the absence of an external threat, 80 percent (*n* = 20) of participants pressed the brake pedal and 12 percent (*n* = 3) of participants used the off button in order to attempt to regain manual control. Similarly, in the presence of an external threat, 56 percent (*n* = 14) of participants pressed the brake pedal and 28 percent (*n* = 7) of participants used the off button in order to attempt to regain manual control. Figure D-17 and Figure D-18 depict both the percentage and the number of participants who attempted to regain manual control using the four possible methods when presented with an Imminent alert without an external threat and an Imminent alert with an external threat, respectively.

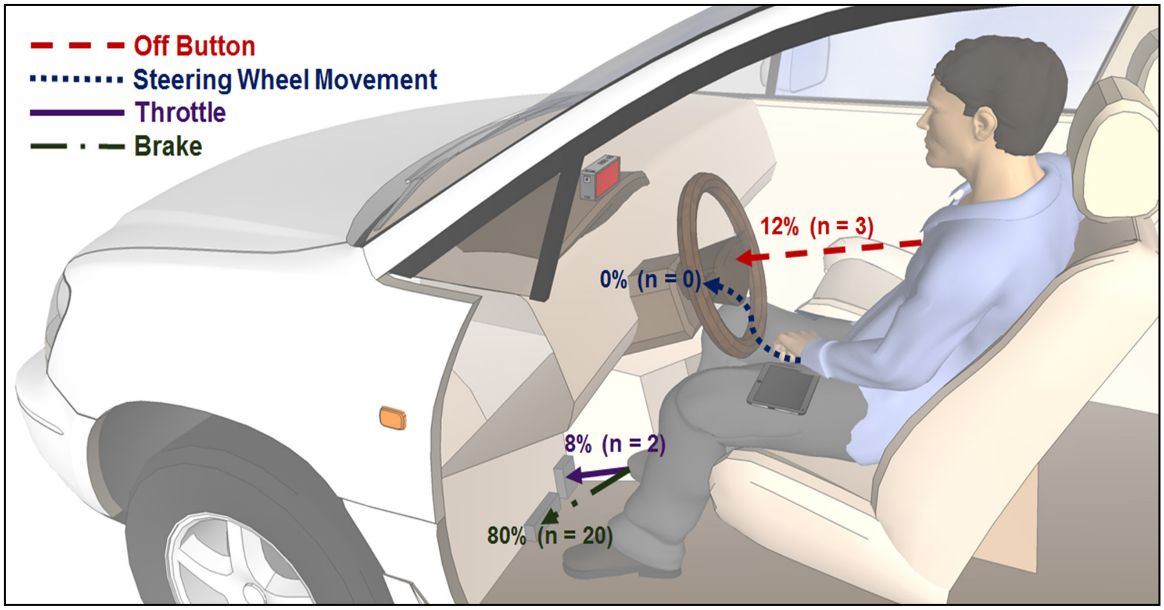



**Figure D-17. Method Used to Regain Control During the Imminent Warning due to a Simulated Internal System Failure (i.e., Imminent–No External Threat; n = 25) for Experiment 3**

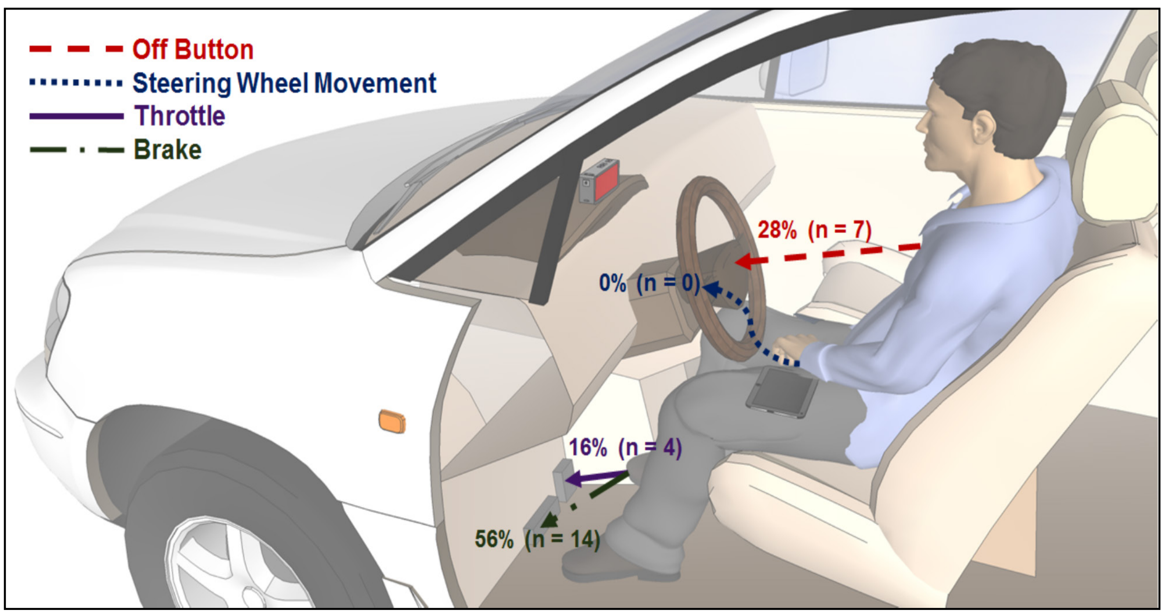



**Figure D-18. Method Used to Regain Control During the Imminent Warning due to an Object Present in the Roadway (i.e., Imminent–External Threat) for Experiment 3**

***Analysis for Time to Activate the Automation***

The time it took for participants to activate the automation after experiencing the alert was explored for each alert type. Figure D-19 (mean and error bar plots of time to activate the automation, stratified by alert type) suggests that vehicle operators may take longer to attempt to activate the automation during the higher-urgency Imminent alerts compared to the lower-urgency

Staged alert. This was particularly noticeable during the Imminent–External Threat alert. Upon experiencing the Staged alert, participants took 4.1 s, on average, to attempt to activate the automation (S.E. = 0.6 s), while upon receiving the Imminent–No External Threat this mean increased to 6.3 s (S.E. = 1.0 s) and upon receiving the Imminent–External Threat alert, the mean further increased to 8.0 s (S.E. = 1.6 s).

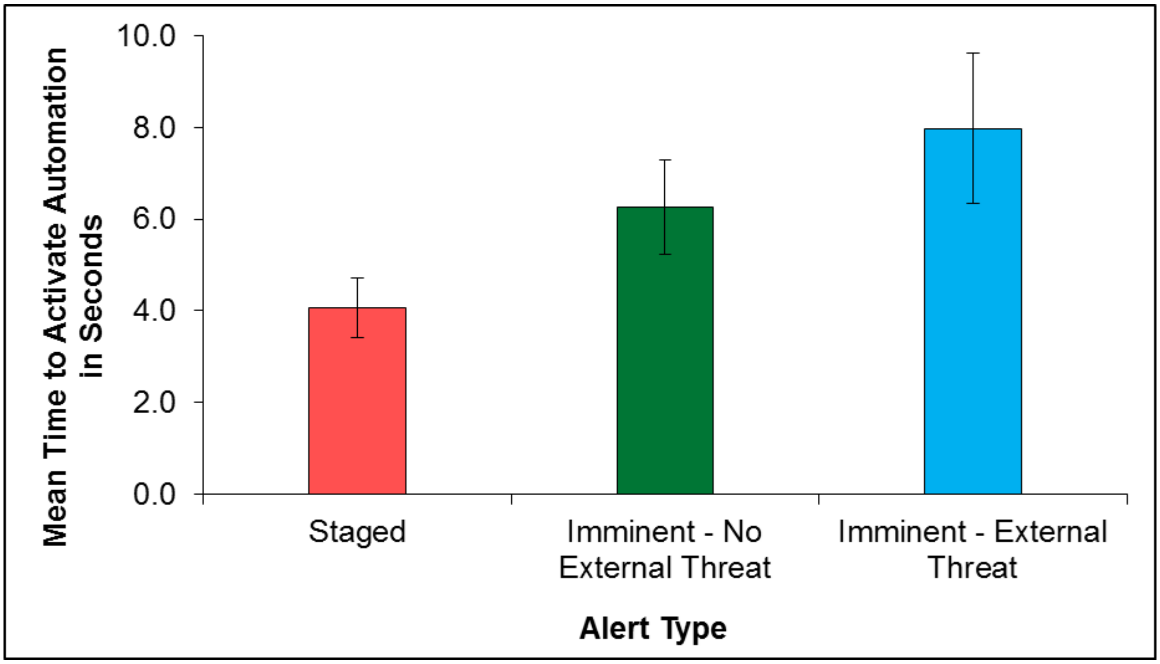



**Figure D-19. Mean and Standard Error Bar Plots, Time to Activate the Automation, by Alert Type for Experiment 3**

Additionally, the researchers analyzed how the time it took participants to attempt to activate the automation varied across the three driving sessions. Figure D-20 illustrates the mean time to attempt to activate the automation for each driving session. The figure suggests that the time it takes for the participant to attempt to activate the automation may decrease over time. The mean time to activate the automation during Session 1, across alerts, was 7.6 s (S.E. = 1.4 s), 5.4 s during Session 2 (S.E. = 1.1 s), and 5.2 s during Session 3 (S.E. = 1.0 s). This suggests that an operator's willingness to activate the automation after experiencing a system failure may increase as the time that they are exposed to the automated system increases.

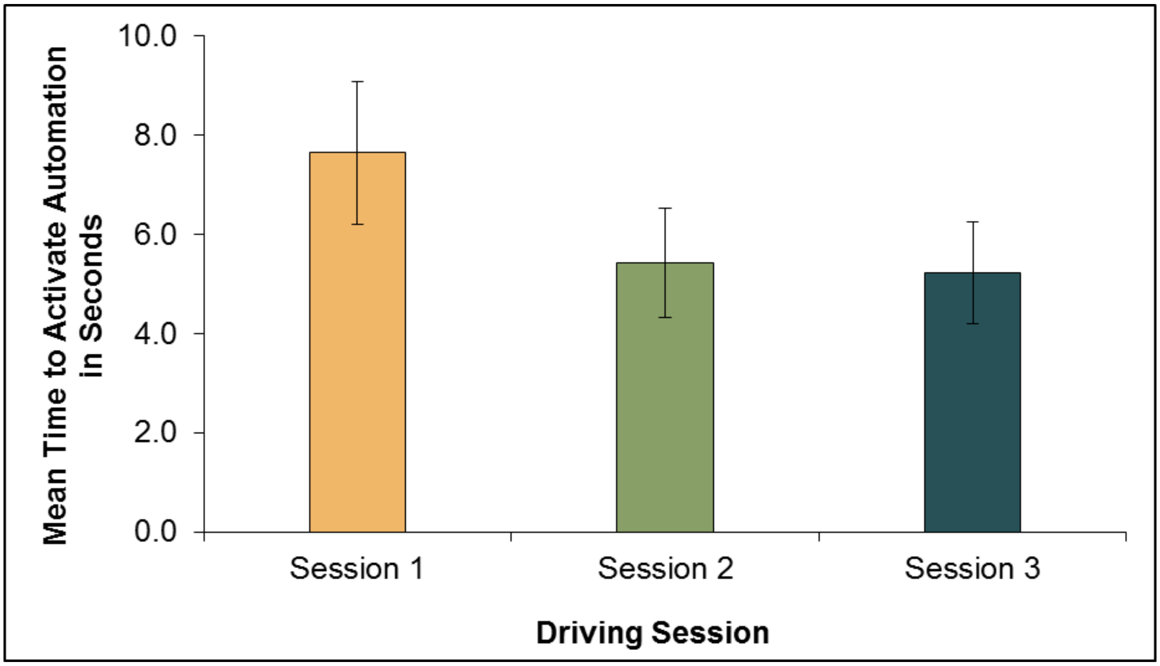



**Figure D-20. Mean and Standard Error Bar Plots, Time to Activate the Automation, by Session for Experiment 3**

Figure D-21 provides boxplots of the time to attempt to activate the automation for the three alert types, stratified by the three driving sessions. During the Imminent–External Threat alert, the amount of time it takes for participants to attempt to activate the automation may decrease over time. The mean time to activate the automation for participants experiencing the Imminent–External Threat alert in Session 1 was 10.8 s; for Session 2, it was 7.1 s; and for Session 3, it was 5.9 s. In addition, the variability of time to activate the automation decreases for participants experiencing the Imminent–External Threat alert. There is less variation in time to activate the automation as participants are exposed to the system alerts. The S.D. for time to activate the automation for the Imminent–External Threat alert was 10.1 s in Session 1, 7.9 s in Session 2, and 6.3 s in Session 3. The same trend was not observed during the Staged alert or the Imminent–No External Threat alert. Hence, the plots suggest that the amount of time operators take to activate the automation may decrease after experiencing an Imminent alert attributed to the presence of an external threat and this may become more precise over time as experience with the automated system is acquired. However, note that this trend is not shown to be the case for the other two alert types.

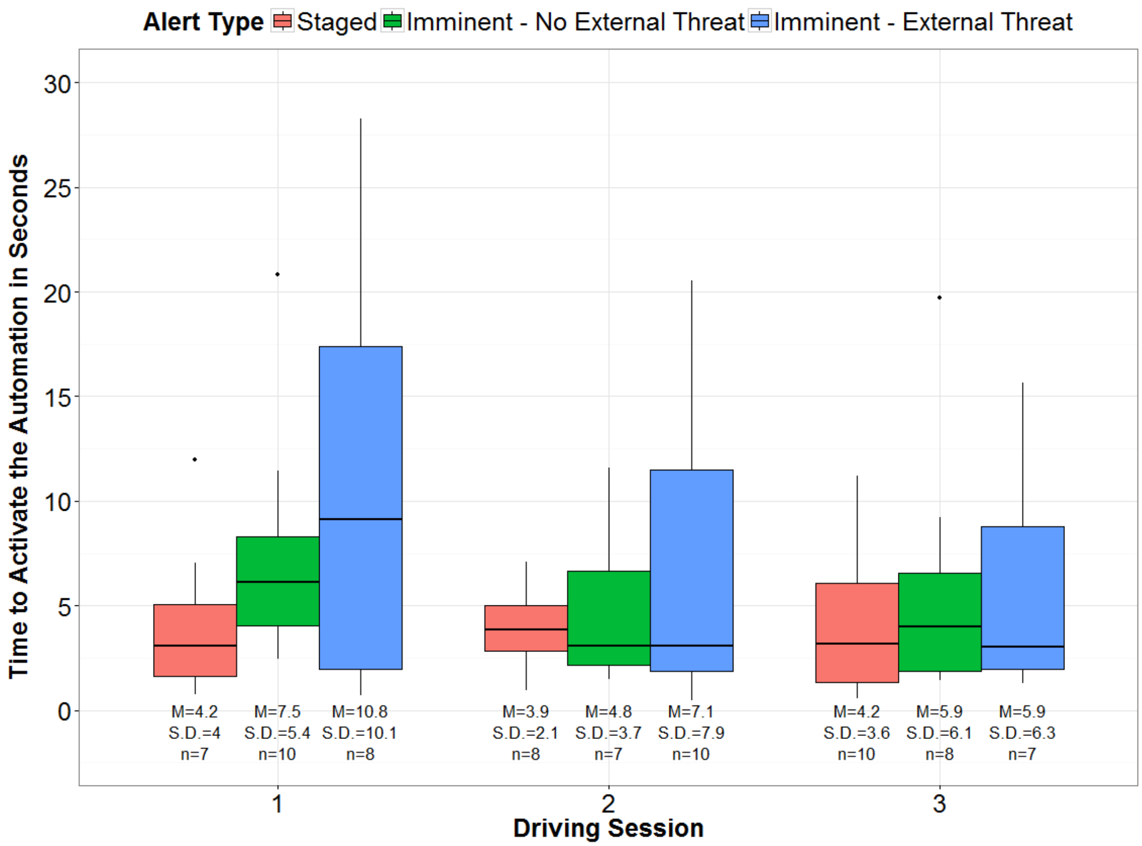



**Figure D-21. Boxplots of Time to Activate the Automation, Stratified by Session and Alert Type for Experiment 3**

<u>Statistical Analysis</u>

The time to activate the automation was log-transformed, and the analysis performed on the log-transformed times. The ANOVA table is displayed in Table D-7.

**Table D-7. Analysis for the Log of Time to Activate the Automation, ANOVA Table for Experiment 3**

| Effect | Num DF | Den DF | F Value | Pr > F |
|---|---|---|---|---|
| **Alert Type** | 2 | 21 | 1.46 | 0.2545 |
| **Session** | 2 | 36.6 | 0.65 | 0.5269 |
| **Session*Alert Type** | 4 | 28.8 | 0.30 | 0.8768 |

The analysis does not reveal any significant effects. The *p* values for alert type, session, and the session-alert type interaction were 0.2545, 0.5269, and 0.8768, respectively, which do not achieve significance. Therefore, there is no statistically significant evidence of an effect of alert type or session on time to activate the automation. Despite the potential for a trend, as shown in the previous plots, the results from the inferential tests indicate that the time to activate the automation did not change across sessions.

***Analysis for Time to Release Control of Steering***

The time it took participants to release control of the steering wheel after activating the automation was examined for each alert type. As shown in Figure D-22, it appears that participants took longer to release control of the steering wheel after experiencing the Imminent–External Threat alert compared to when they experienced the Staged alert and the Imminent–No External Threat alert. The mean time to release control of the steering wheel during the Imminent–External Threat alert across sessions was 2.7 s (S.E. = 0.5 s), compared to 1.8 s during the Imminent–No External Threat alert (S.E. = 0.3 s), and 1.4 s during the Staged alert (S.E. = 0.2 s).

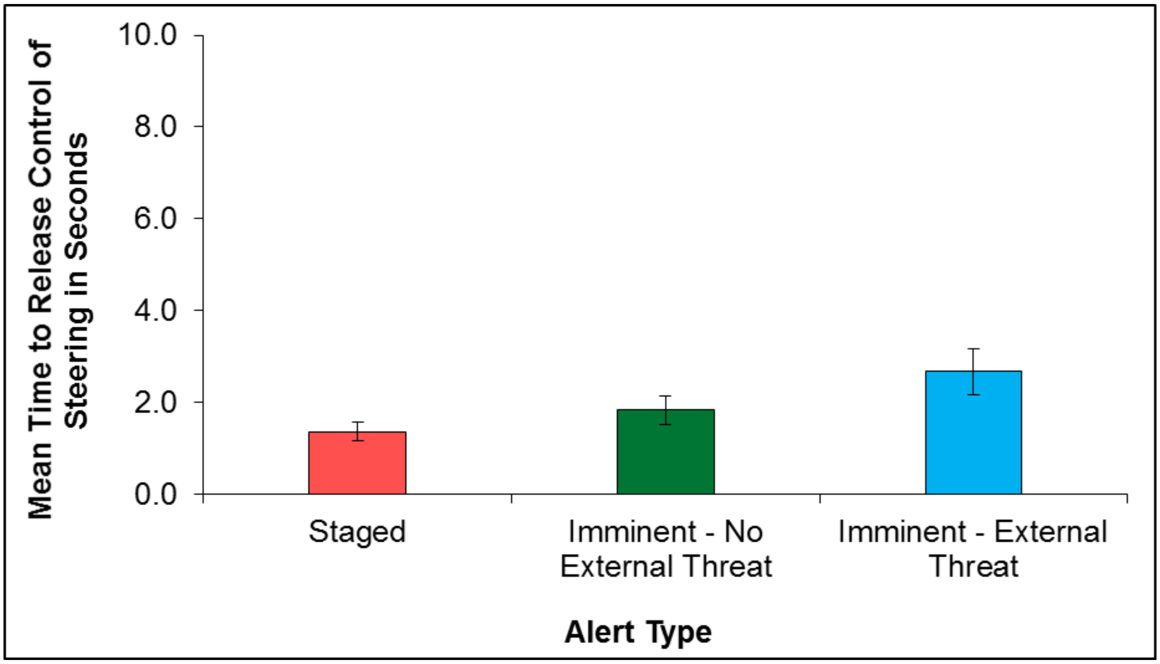



**Figure D-22. Mean and Standard Error Bar Plot, Time to Release Control of Steering, by Alert Type for Experiment 3**

Additionally, the time it took participants to release control of the steering wheel after activating the automation was examined over time. Figure D-23 depicts the mean time to release control of the steering wheel across driving sessions. The mean time to release control of the steering during Session 1, across alerts, was 1.9 s (S.E. = 0.3 s), 2.4 s for Session 2 (S.E. = 0.4 s), and 1.6 s for Session 3 (S.E. = 1.6 s). This suggests that the time it takes for participants to release control of the steering wheel does not change with increased exposure to the automation.

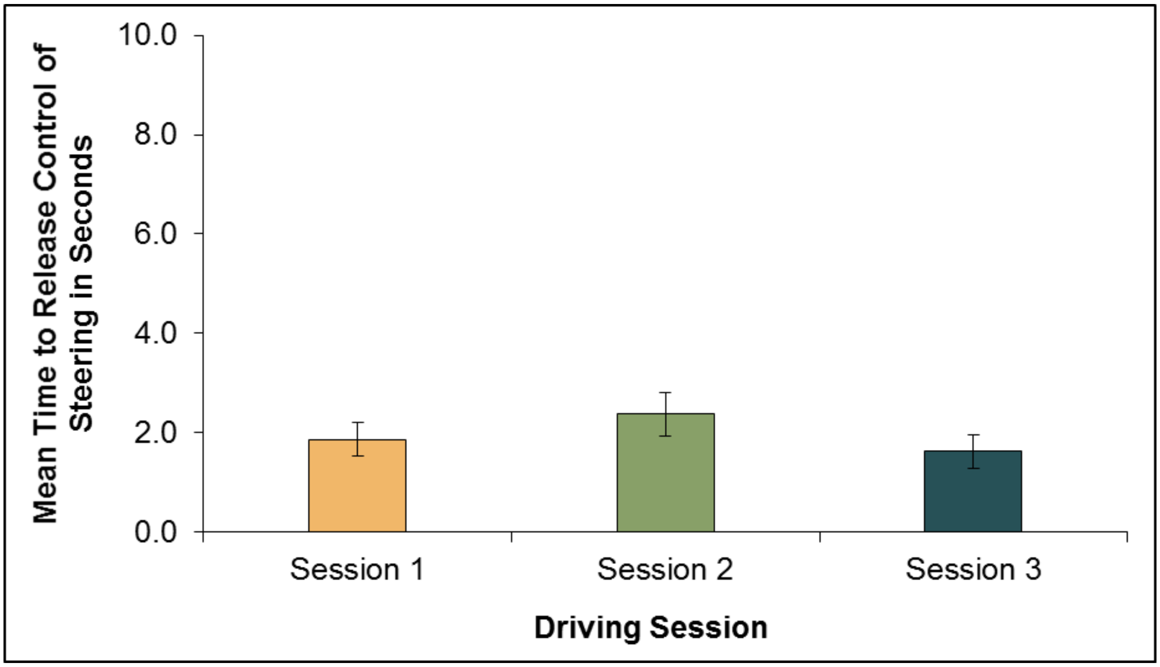



**Figure D-23. Mean and Standard Error Bar Plots, Time to Release Control of Steering, by Session for Experiment 3**

Boxplots of time to release control in alert types stratified by session are displayed in Figure D-24. The means for Imminent–External Threat alerts within sessions (2.4 s in Session 1, 3.1 s in Session 2, and 2.3 s in Session 3) are all higher than the highest mean for the other two alert types (2.1 s for the Imminent–No External Threat alert in Session 1). The variability of time to release control of steering was also higher for the Imminent–External Threat in Session 2 (S.D. = 2.9 s) and Session 3 (S.D. = 2.8 s) than it was for the other two alert types in Session 2 and Session 3. In Session 1, the variance for Imminent–No External Threat (S.D. = 1.9 s) was higher than for the Imminent–External Threat (S.D. = 1.7 s). An outlier of 8.8 s in Session 3 for Imminent–External Threat inflated the mean and S.D.

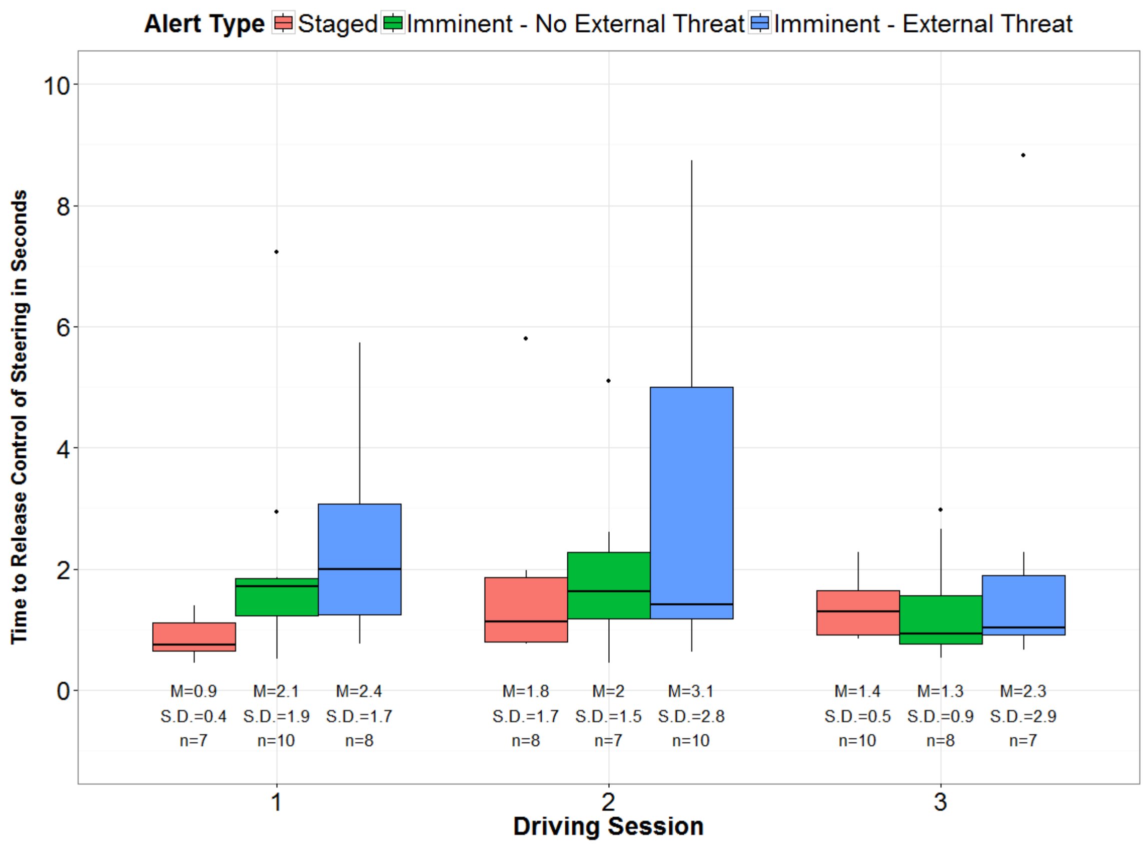



**Figure D-24. Boxplots of Time to Release Control of Steering, Stratified by Session and Alert Type for Experiment 3**

Statistical Analysis

The time to release control was log-transformed, and the analysis performed on the log-transformed times. For this analysis, the ANOVA table for time to release control of steering is displayed in Table D-8.

**Table D-8. Analysis for the Log of Time to Release Control of Steering, ANOVA Table**

| Effect | Num DF | Den DF | F Value | Pr > F |
|---|---|---|---|---|
| **Alert Type** | **2** | **21** | **6.16** | **0.0079** |
| **Session** | 2 | 42.7 | 1.99 | 0.1496 |
| **Alert Type*Session** | 4 | 29.8 | 1.92 | 0.1327 |

Since there were no departures from normality or randomness apparent in the scaled residual plots, analysis was conducted on the log of the time to release control of steering. In this analysis, alert type ($p$ = 0.0079) was significant, indicating that the log of time to release control of steering may be different in the different alerts. Session ($p$ = 0.1496) and the alert type by session interaction ($p$ = 0.1234) were not significant, indicating that there is no statistically significant evidence of an effect across time, nor that the effect of alert type depends on time. To investigate the alert type effect visually, a plot of least squares means of the log of time to release control of steering in the different alert types is displayed in Figure D-25.

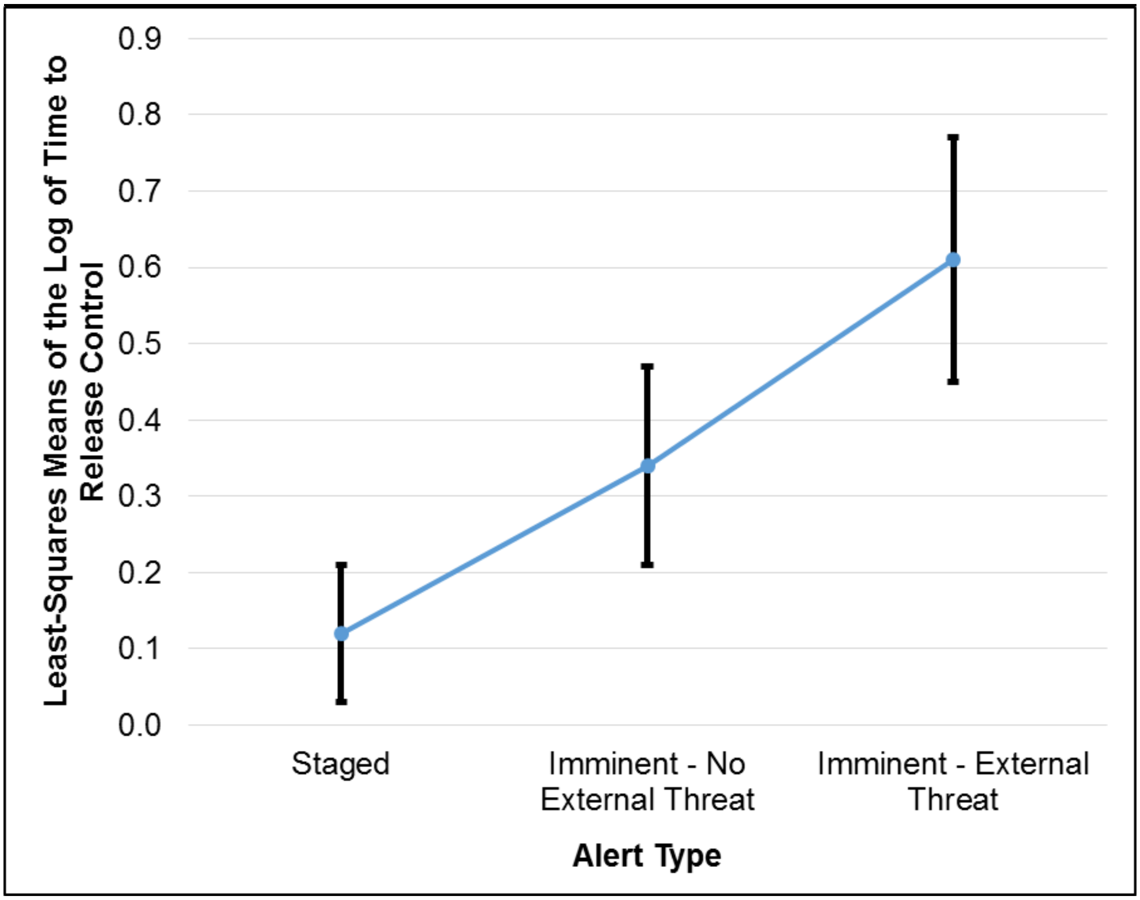



**Figure D-25. Least Squares Means of the Log of Time to Release Control of Steering with 95 Percent Confidence Bands, Alert Type for Experiment 3**

In general, the log of time to release control takes on an increasing path (on average) as the urgency increases. Specifically, the Staged alert, which has the lowest urgency, resulted in the lowest log time to release control, while the most urgent alert, the Imminent–External Threat alert, resulted in the highest. The biggest difference appears to be between the Imminent–External Threat alert and the Staged alert. Specifically, it appears that the Imminent–External Threat alert has a higher least squares mean of the time to release control of steering than does the Staged alert, which may mean that operators experiencing an Imminent–External Threat alert take longer to release control of the steering than do those experiencing a Staged alert.

To determine if this difference is statistically significant, comparisons of the least squares means of the log of time to release control of steering were conducted between each pair of alert types, with a Bonferroni adjustment. The significant differences are displayed in Table D-9.

**Table D-9. Log of Time to Release Control of Steering Differences in Least Squares Means, Alert Type, Significant at Bonferroni Adjusted *P* Values, for Experiment 3**

| Alert Type | Comparison | Estimate | S.E. | *t* Value | Adj *p* Value | Lower Confidence Limit | Upper Confidence Limit |
|---|---|---|---|---|---|---|---|
| **Staged** | **Imminent − External Threat** | -0.49 | 0.13 | -3.81 | 0.0024 | -0.75 | -0.22 |

The estimate of -0.49 indicates that in the Imminent–External Threat alert, the log of the time to release control of steering was estimated at 0.49 log seconds greater than the Staged alert. To interpret this in terms of the actual time to release control of steering, the estimate was back-transformed to 1.63, indicating that the amount of time to release control of steering for Imminent–External Threat alerts would be expected to be about 63 percent higher than for Staged alerts. The estimated differences between the log of the time to release control for Imminent–No External Threat alerts and the other two alert types were not statistically significant.

From these results, it seems that as the level of urgency of the alert increases, the amount of time it takes participants to release control back to the system may also increase. There is evidence that, when an operator experiences an Imminent–External Threat alert, it will take the operator longer, on average, to release control of the steering than it will after the operator experiences a Staged alert. However, the operator may not take significantly more or less time to release control of the steering after experiencing an Imminent–No External Threat alert compared to the other two alerts. Hence, an Imminent alert combined with an External Threat alert may make operators more reluctant to release control of steering than they would be if they did not have either an external threat or an Imminent alert.

***Analysis for Time to Resume a Non-driving Task***

Upon activating the automation and releasing control of the steering wheel, participants were free to engage in a non-driving task when they felt comfortable doing so. The majority of participants elected to engage in a non-driving task after experiencing the alert. Out of 75 instances, most of the time (61 percent, $n_i$ = 46) participants were observed interacting with the tablet computer; in 28 percent of the instances participants ($n_i$ = 21) were observed interacting with their smartphone; in 3 percent of the instances participants ($n_i$ = 2) were talking with the in-vehicle experimenter(s); and in 8 percent of the instances ($n_i$ = 6) participants did not resume a non-driving task. Figure D-26 depicts the number of instances in which participants were observed engaging in the aforementioned non-driving tasks. Notice that these non-driving-task type frequencies are similar to those observed at the time at which the alert was first presented. Thus, it appears that most participants resumed the same non-driving task that they were engaged in directly before the alert.

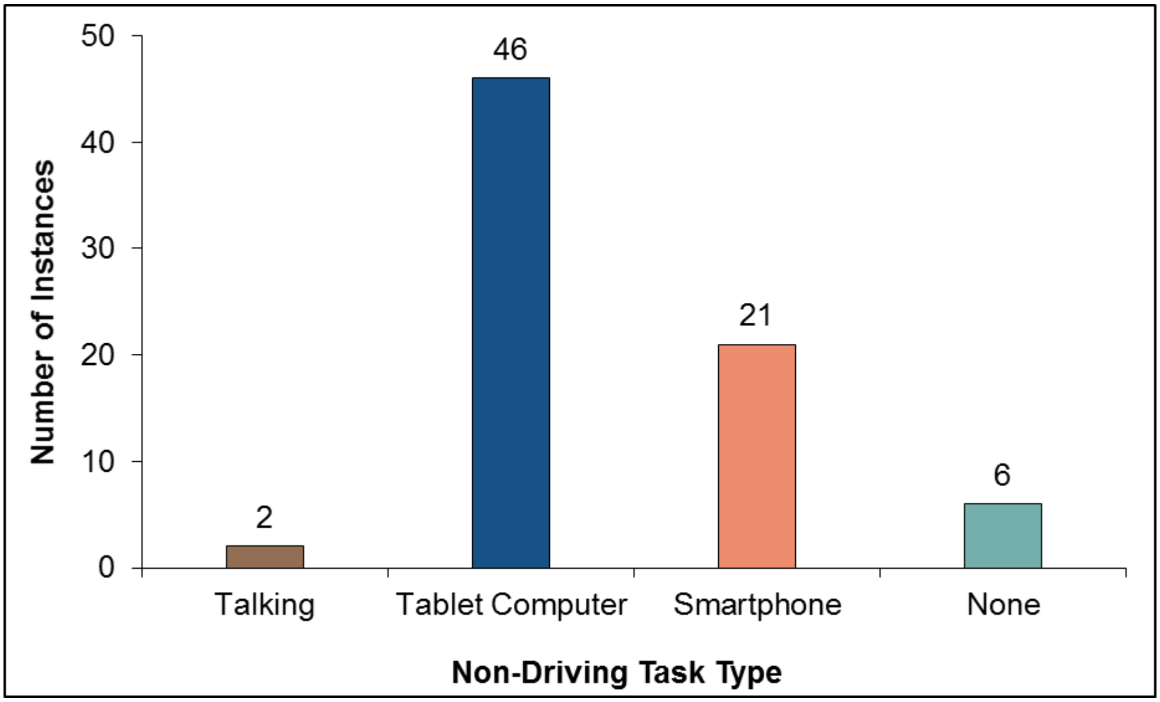



**Figure D-26. Non-driving-task Type Observed After Experiencing the Alert for Experiment 3**

The time it took participants to resume a non-driving task was examined for each alert time. Figure D-27 depicts the mean time it took participants to resume engaging in a non-driving task after experiencing each of the three alert types. Participants took longer to resume a non-driving task after experiencing the lower-urgency Staged alert compared to when they experienced the Imminent alerts. The mean time to resume a non-driving task after experiencing a Staged alert was 39.1 s (S.E. = 14.6 s), compared to 13.9 s after experiencing an Imminent–No External Threat alert (S.E. = 4.2 s), and 26.3 s after experiencing the Imminent–External Threat alert (S.E. = 9.5 s).

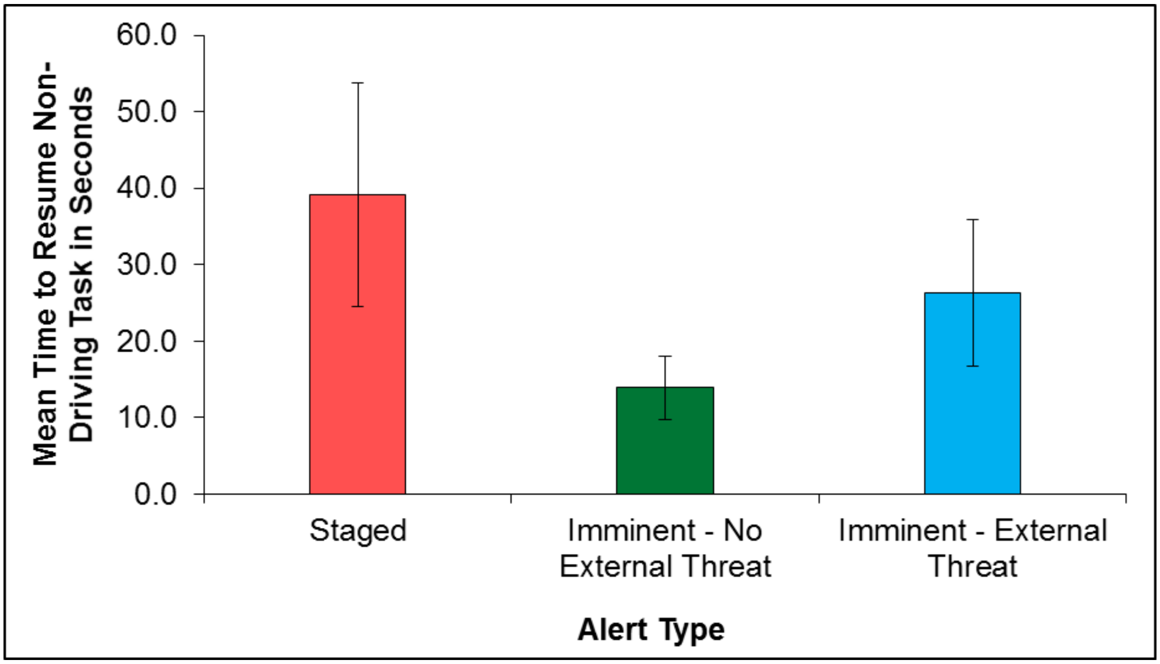



**Figure D-27. Mean and Standard Error Bar Plots, Time to Resume Non-driving Task, by Alert Type for Experiment 3**

Additionally, the time it took participants to resume a non-driving task was examined across the three driving sessions. Figure D-28 depicts the mean time to resume a non-driving task after experiencing an alert across the three driving sessions. The mean time to resume a non-driving task during Session 1 across alerts was 38.0 s (S.E. = 12.4 s), 19.3 s during Session 2 (S.E. = 7.0 s), and 21.9 s during Session 3 (S.E. = 11.5 s).

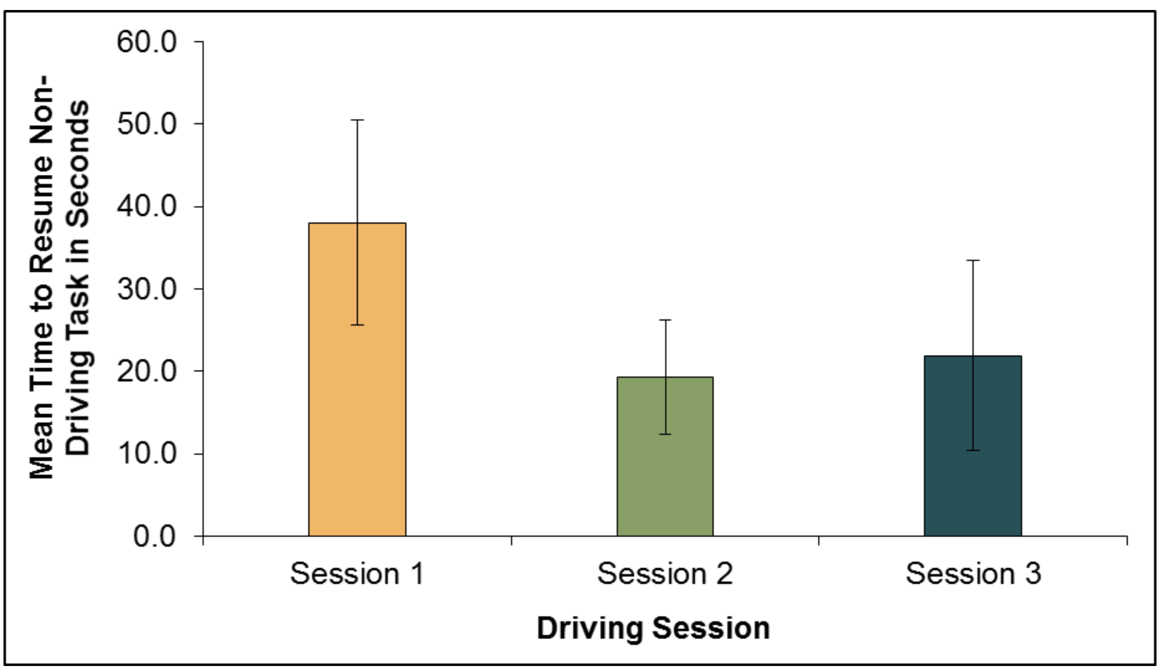


**Figure D-28. Mean and Standard Error Bar Plots for Time to Resume Non-driving Task, Session for Experiment 3**

Boxplots for time to resume non-driving task by alert type, stratified by session, are displayed in Figure D-29. Although in a majority of instances (49 out of 69), participants resumed a non-driving task within 20 s, in some instances a participant took as long as 258.3 s to resume a non-driving task after giving control of the vehicle to the automated system. In six instances, participants did not resume a non-driving task at all (one during Session 1, two during Session 2, and three during Session 3). The three highest times all followed Staged alerts. The high of 258.3 s occurred in Session 3; the next highest times, 202.2 s and 189.9 s, occurred in Session 1. Out of eight instances in which time to resume a non-driving task was greater than 50 s, only one—the highest—occurred in Session 3. Four of these instances occurred during Session 1, and two occurred during Session 2. These data suggest that operators may differ greatly in how long they will take to resume a non-driving task after activating the automation. Mean and error bar plots for time to resume non-driving task in and within alert types and sessions are provided in Figure D-29.

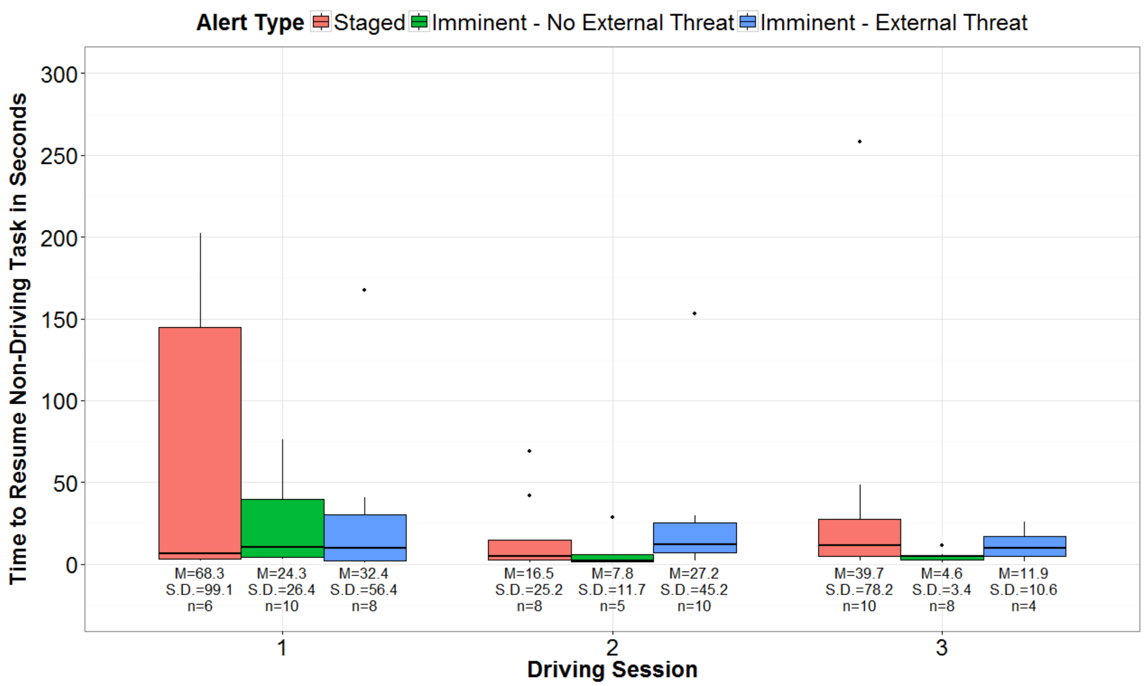



**Figure D-29. Boxplots of Time to Resume Non-driving Task Stratified by Session and Alert Type for Experiment 3**

<u>Statistical Analysis</u>

Out of the 75 instances of operating the vehicle, there were 6 instances in which a participant did not perform a non-driving task after releasing control. In all six instances, the participant had not performed a non-driving task prior to the alert. These six cases are detailed in Table D-10.

**Table D-10. Missing Times to Resume Non-driving Task for Experiment 3**

| Age Group | Session | Alert |
|---|---|---|
| **40−54** | 1 | Staged |
| **18−24** | 2 | Imminent–No External Threat |
| **40−54** | 2 | Imminent–No External Threat |
| **18−24** | 3 | Imminent–External Threat |
| **40−54** | 3 | Imminent–External Threat |
| **55+** | 3 | Imminent–External Threat |

The analysis proceeded without these instances, but included all other instances, for a total of 69 instances. The ANOVA table is displayed in Table D-11.

**Table D-11. Analysis of Time to Resume Non-driving Task, ANOVA Table for Experiment 3**

| Effect | Num DF | Den DF | F Value | Pr > F |
|---|---|---|---|---|
| **Session** | 2 | 35.7 | 1.26 | 0.2970 |
| **Alert Type** | 2 | 19.4 | 2.43 | 0.1141 |
| **Session*Alert Type** | 4 | 26.3 | 0.84 | 0.5139 |

With this analysis, neither the alert type ($p$ = 0.0972), session ($p$ = 0.3010) or the alert type by session interaction ($p$ = 0.4853) were significant, indicating no significant effect of either alert type or session. This means that the data provided no statistically significant evidence that, when an operator moves to resume a non-driving task, the time to do so will be influenced either by what type of alert they experienced or how long they have been on the road.

## Operator Behavior Analysis Summary

The analysis indicates that the influence of the alert on the operator's ability to interact with the automated system in a timely manner existed in several areas of the process: when they first needed to take control of the vehicle, when they gave control back to the automated system, and when they returned to what they were doing before the alert went off. The influence is felt particularly strongly in time to regain control, when participants in the experiments regained control significantly more slowly after experiencing a Staged alert than they did after the Imminent alerts. As mentioned previously, this is an expected result due to the urgency component of the Imminent alerts (i.e., red visual icon with voice command instructing the participant to turn off automation). Additionally, those experiencing a Staged alert may take more time to react than they would otherwise, but they may take less time to release control of the steering (after activating automation) than those who experienced an Imminent alert with an external threat. However, the key differences mainly involved the Staged alerts, meaning that nowhere in the process of taking control and returning it did the presence of an external threat increase or decrease the effect of the Imminent alert.

## Eye-glance Behavior

### *Analysis for Monitoring Rate*

This analysis helped to determine if the alert type affected eye-glance behavior and if this effect changed over time, which will aid in determining whether the effect over time should be considered in future research investigations. Table D-12 provides a summary of the results and general takeaways for monitoring rate and off-road glances. Alert type and session were marked as “**Significant”** if they achieved statistical significance, and “Not Significant” if they did not.

**Table D-12. Summary Table for Eye-glance Behavior Analysis for Experiment 3**

| Variable | Alert Type | Performance Over Time (Represented by Intervals Within Sessions) | Takeaways |
|---|---|---|---|
| **Monitoring Rate** | Not Significant | **Significant** | Participants were more likely to spend a higher percentage of time on driving-related glancing after the alert than before. Further, this relationship was consistent across alerts. |
| **Off-Road Glances** | **Significant** | **Significant** | The rate of off-road glances was higher after the Imminent alerts than after the Staged alert, likely due to the longer intervals in the latter than in the former. |

This analysis explored the eye-glance tendencies of participants over time and how these relate to different alerts. Specifically, the analysis of this experiment explored the proportion of eye-glance time that was driving-related, and the proportion that was not driving-related (i.e., while the participant was using a tablet, smartphone, talking to in-vehicle experimenter, or some other non-driving task).

During each driving session, eye-glance data were calculated for three segments: 65 to 55 s before the alert occurs, 10 s before the alert occurs to the start of the alert, and the start of the alert to when the participant regains control of the vehicle (Figure D-30). The first 10-second time interval was intended to capture the participant’s glance tendencies well before the alert was received. Similarly, the second 10-second time interval was designed to capture the participant’s glance tendencies directly before the alert was received. The third time interval varied in length as the endpoint was determined by the time at which the participant successfully regained manual control. The purpose of the third interval was to capture the participant’s eye-glance tendencies after he or she received the system alert.

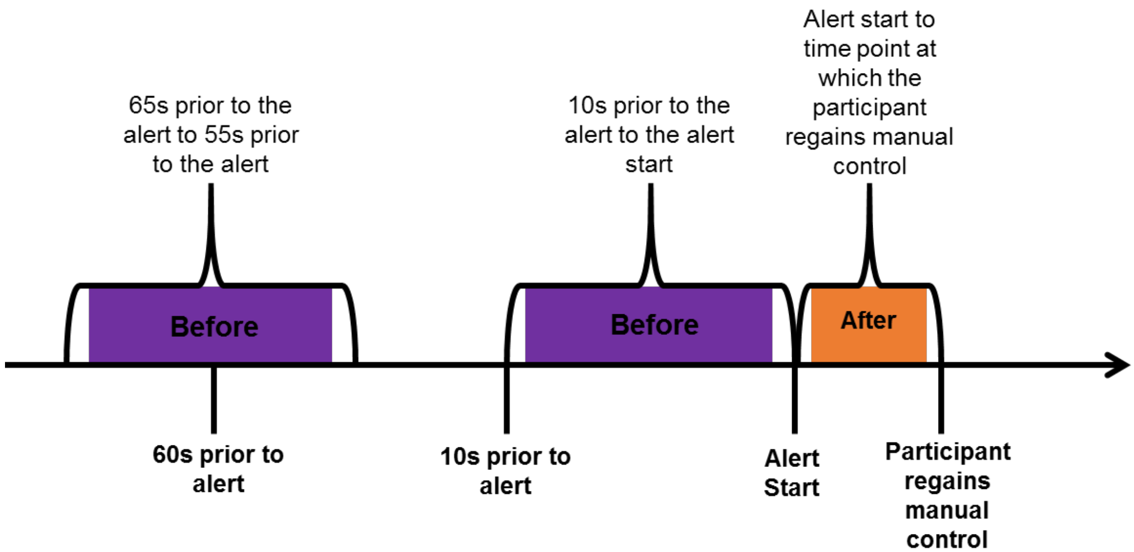



**Figure D-30. Eye-glance Data Calculation Periods for Experiment 3**

Since the first and second time intervals both occurred before the alert, these were combined and considered “Before the Alert,” while the third time interval was considered “After the Alert.” The participant’s total time performing non-driving-related glances and driving-related glances was calculated. (These do not overlap and add to the total time for that interval.) In Figure D-31, boxplots of the percentage of driving-related glance time are displayed, stratified by before/after and alert. The same general pattern holds for all three alerts: participants varied greatly in their driving-related glance time before the alert, but in the time interval after the alert, the percentage of driving-related glance time is mostly concentrated above 75 percent, indicating that the participants glanced at the forward roadway relatively more often after the alert than before, regardless of the alert.

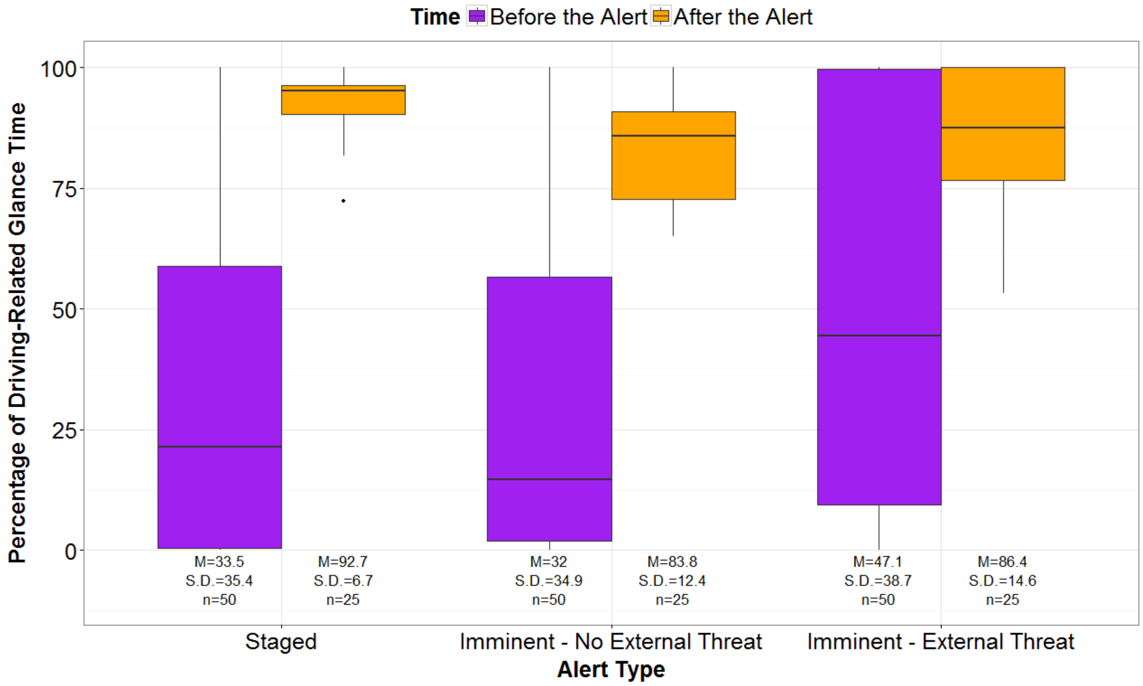



**Figure D-31. Boxplots of Percentage of Driving-related Glance Time Stratified by Alert and Time Interval for Experiment 3**

Statistical Analysis

To test for significant differences between the time intervals for percentage of driving-related glance time, Friedman's Test, a nonparametric alternative for one-way repeated measures ANOVA, was used because of the high variability of the first two time points compared to the last time point, and the non-normal distributions of the data. The test was performed separately for each alert type, then collapsed across sessions because the time intervals within sessions are the primary time components of interest in this analysis. The results, with Cochran-Mantel-Haenszel (CMH) statistics, are displayed in Table D-13.

**Table D-13. Friedman Test Results, Differences in Time Point Mean Percentage Driving-related Glance Time, by Alert Type for Experiment 3**

| Alert Type | CMH Statistic | *P* Value |
|---|---|---|
| **Staged** | **30.64** | **<0.0001** |
| **Imminent–No External Threat** | **26.40** | **<0.0001** |
| **Imminent–External Threat** | **22.05** | **<0.0001** |

Within each of the alert types (collapsed across sessions), the *p* value for the comparison of monitoring rate across the time intervals is less than 0.0001, indicating that there is evidence that the monitoring rate differs between at least two of the time intervals in all three alert types.

***Analysis for Non-driving-related Glances***

Non-driving-related glances were defined as glances not pertaining to the driving scenario, and include the following glances: smartphone, tablet computer, in-vehicle experimenter, and hygiene-related glances (Table 4-1). The goal of the analysis of the non-driving-related glances was to determine if the rate of the glances per second differed before the alert versus after the alert, and if any difference was affected by the alert type, session, or the level of engagement a participant had with a non-driving task at the time of the alert. In a majority of instances, participants made either 0 (n = 40, 17.8 percent) non-driving-related glances or 1 (n = 97, 43.1 percent) non-driving-related glance during time intervals. The maximum number of non-driving-related glances was 5, which was made in three instances (1.3 percent). None of these instances occurred after the alert took place.

Although each of the two intervals before the alert lasted 10 s, the length of the third time interval (from the start of the alert to the time at which the participant regains control of the vehicle) was determined by the point in time at which the participant regained control. The mean time to regain control for those experiencing the Staged alert was 17.1 s (S.D. = 6.2), for Imminent–No External Threat it was 2.3 s (S.D. = 0.9), and for Imminent–External Threat it was 2.1 s (S.D. = 0.7). Overall, the mean length of the third time interval is 7.2 s (S.D. = 8.0). To account for the different lengths of the after-alert interval, the analysis focused on the rate of non-driving-related glances per second. Boxplots of these rates, stratified by alert type and time interval, are displayed in Figure D-32.

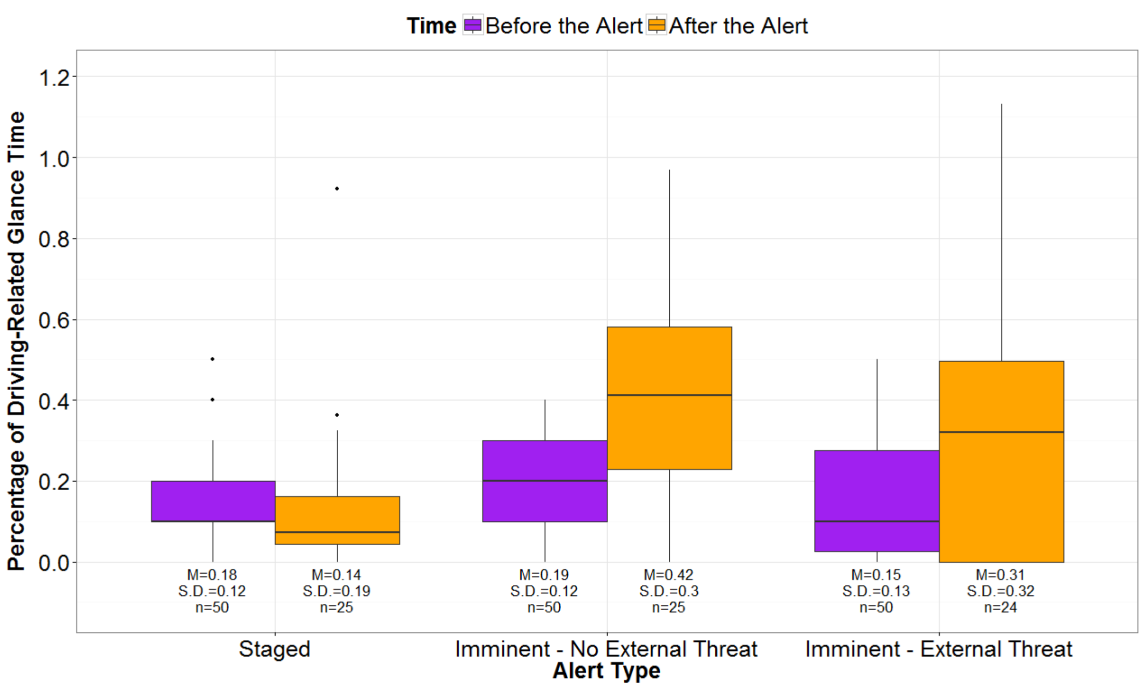



**Figure D-32. Boxplots of Rates of Non-driving-related Glances per Second, Alert Type and Time Interval for Experiment 3**

The mean rate increased after the alert for Imminent–No External Threat (0.42, S.D. = 0.3) and Imminent–External Threat (0.31, S.D. = 0.32) from before the alert (mean = 0.19, S.D. = 0.12 for

Imminent–No External Threat, mean = 0.15, S.D. = 0.13 for Imminent–External Threat). Meanwhile, in the Staged alert, the mean rate of non-driving-related glances after the alert (0.14, S.D. = 0.19) was lower than before the alert (mean= 0.18, S.D. = 0.12).

Statistical Analysis

A generalized linear mixed Poisson model was fit to account for correlations between instances in the same subject. To model the rate of non-driving-related glances per second, the log of seconds was used as an offset variable. Although there were four independent variables used (alert type, session, non-driving task, and time), only main effects and interactions involving alert type, session, and time were used because of low cell counts for the non-driving task. Note that significant differences are interpreted in this case as changes within participants, not between participants. Model tests for significance are displayed in Table D-14.

**Table D-14. Non-driving-related Glance Model, Tests for Significance for Experiment 3**

| Effect | Num DF | Den DF | *F* Value | *P* Value |
|---|---|---|---|---|
| **Non-Driving Task** | **2** | **182** | **11.08** | **<0.0001** |
| **Time** | 1 | 182 | 3.57 | 0.0605 |
| **Session** | 2 | 182 | 2.56 | 0.0802 |
| **Alert Type** | **2** | **182** | **7.61** | **0.0007** |
| **Time*Alert Type** | **2** | **182** | **10.05** | **<0.0001** |
| **Session*Alert Type** | 4 | 182 | 1.18 | 0.3222 |
| **Time*Session** | 2 | 182 | 1.92 | 0.1498 |
| **Time*Session*Alert Type** | 4 | 182 | 0.91 | 0.4617 |

The Alert Type by Time interaction ($p$ < 0.0001) was significant, indicating that the difference between time intervals in the log of the rate of non-driving-related glances per second varied across alert types. Non-driving task ($p$ <0.0001) was also significant, indicating that the level of engagement an operator has with a non-driving task is associated with their amount of non-driving-related glances.

To investigate this, the least squares means differences were calculated, with a Bonferroni adjustment to control the type 1 error rate, for the combinations of alert type and times, and for non-driving tasks. Specifically, the comparisons of interest within the alert type*time interaction are the differences in the log rate from before to after the alert within each different alert type. The analysis also revealed which level of non-driving task engagement has the greatest rate of non-driving-related glances. The significant differences are displayed in Table D-15 and Table D-16.

**Table D-15. Differences in Least Squares Means, Bonferroni Adjustment, Alert Type*Time for Experiment 3**

| Alert Type*Time | Comparison | Estimate | S.E. | *t* Value | Adj *p* Value | Lower Confidence Limit | Upper Confidence Limit |
|---|---|---|---|---|---|---|---|
| **Imminent–External Threat*After** | **Imminent–External Threat*Before** | 0.87 | 0.28 | 3.10 | 0.0069 | 0.31 | 1.42 |
| **Staged*After** | **Staged*Before** | -0.55 | 0.21 | -2.66 | 0.0258 | -0.96 | -0.14 |

**Table D-16. Differences in Least Squares Means, Bonferroni Adjustment, Non-driving Task for Experiment 3**

| Non-Driving Task Status | Comparison | Estimate | S.E. | *t* Value | Adj *p* Value | Lower Confidence Limit | Upper Confidence Limit |
|---|---|---|---|---|---|---|---|
| **Visually Engaged** | None | 1.13 | 0.26 | 4.3 | <0.0001 | 0.5 | 1.77 |

Although the log rate of non-driving-related glances increased significantly from before to after the alert during Imminent–External Threat alerts, the same did not hold true for Imminent alerts without an external threat. The estimate of 0.87 (95 percent confidence interval 0.31 to 1.42) indicates that for individual operators, their log rate of non-driving-related glances may increase by 0.87. Back-transforming this to 2.39 indicates that individual operators may make 2.39 times as many non-driving-related glances per second after an Imminent–External Threat alert as before. The same significant change did not hold for Imminent alerts without an external threat. Meanwhile, for Staged alerts, the log rate of non-driving-related glances decrease by an estimated -0.55 (95 percent confidence interval -0.96 to -0.14). Back-transforming this to 0.58 indicates that after a Staged alert, operators may make about 58 percent as many non-driving glances per second as they would before the Staged alert.

The data also suggest that when an individual operator is visually engaged in a non-driving task, they will have a significantly higher rate of non-driving-related glances than they would if they were not engaged in a non-driving task at all. The estimate of 1.13 (95 percent confidence interval 0.5 to 1.77) indicates that individual participants had a log rate of non-driving-related glances that was 1.13 times higher when they were visually engaged than when they were not engaged at all. This estimate can be back-transformed to 3.10, indicating that participants will make 3.1 times as many non-driving glances per second if they are visually engaged in a non-driving task compared to when not engaged at all. Holding was not significantly different from either visually engaged or not engaged.

Eye-Glance Analysis Summary

Participants' eye-glance behavior following the alert changed significantly from before the alert, as the data demonstrate. Participants spent significantly more time focusing their attention on operating the vehicle immediately after the alert until they regained control than they did prior to the alert. Further, this trend did not change depending on the type of alert. Participants also shifted the rate at which their eye-glances were non-driving-related after the alert versus before, but this effect may be influenced by the alert type. After experiencing an Imminent alert with an external threat, participants had significantly higher rates of non-driving-related glances than before the alert, while for Staged alerts, the rate decreased significantly for individual participants. There was no significant effect for Imminent–No External Threat alerts.

# Trust

Trust scales used in the assessment of automated systems (Lee & Moray, 1994; Luz, 2009; Jian, Bisantz, & Drudy, 2000) were reviewed in order to develop the ones used for the current study. Drawing upon the 19-question trust scale used by Luz (2009), key phrases examining different aspects of trust (overall trust, perceived reliability, perceived competence under automated control, and perceived understandability) were identified. The wording of the phrases was modified to reflect the automated system in use as opposed to the medium-term conflict detection system studied by Luz. Responses were obtained using 7-point Likert-type scales (Appendix E).

Another important aspect to the automated system and the alerts is the operators' subjective assessment of their effectiveness. The operators' belief in the system's ability to function can help them use it effectively. In this section of the analysis, the participants' subjective measure of their trust in the system was analyzed to determine if their level of trust in the system changed over time or after experiencing an alert. In addition, the research team sought to understand if any change in trust varied across the different alert types.

Various subjective measures were used to gauge participant trust and overall impressions of the system, vehicle, and experience during the experiment. A 7-point Likert-type scale was administered 12 times throughout the three driving sessions. Participants were read the statement "I can trust the automated system to function properly while I am doing something else" and asked to provide a rating from "1" (Strongly Disagree) to "7" (Strongly Agree) (Appendix E).

In addition, each participant completed the after-experience trust scales that listed six statements gauging overall trust, perceived reliability, perceived competence under automated control, and perceived understandability using a 7-point Likert-type scale (Appendix E). This was followed by an interview in which the experimenter asked questions regarding overall impressions, comfort level with the automation, thoughts regarding system alerts, and concerns with automated vehicle technology (Appendix F). The interviews were recorded and transcribed and were used to provide additional insight into the participants' trust ratings.

### *Trust over Time*

This analysis helped to determine if participants' trust in the automated vehicle system changed over time, and if that change was affected by different alerts. The results and general takeaways for trust are shown In Table D-17. Performance over time is marked as "**Significant"** if it achieved statistical significance, and alert type is marked "Not Significant" if it did not. More detail on the analysis is available following the table.

**Table D-17. Summary Table for Trust Analysis for Experiment 3**

| Variable | Alert Type | Performance Over Time | Takeaways |
|---|---|---|---|
| **Trust** | Not Significant | **Significant** | Participants trusted the system, and their trust increased over time. However, the change in trust over time was not dependent on experiencing certain types of alerts. |

Each participant orally provided a trust rating from "1" (Strongly Disagree) to "7" (Strongly Agree) a total of 12 times throughout the experiment: at the beginning and end of each session, and before and after the alert in each session. One participant (male, age 26) did not provide a rating at the end of the last session, but a rating was provided for all other instances, for a total of 299 ratings. The frequencies of the trust ratings are displayed in Figure D-33.

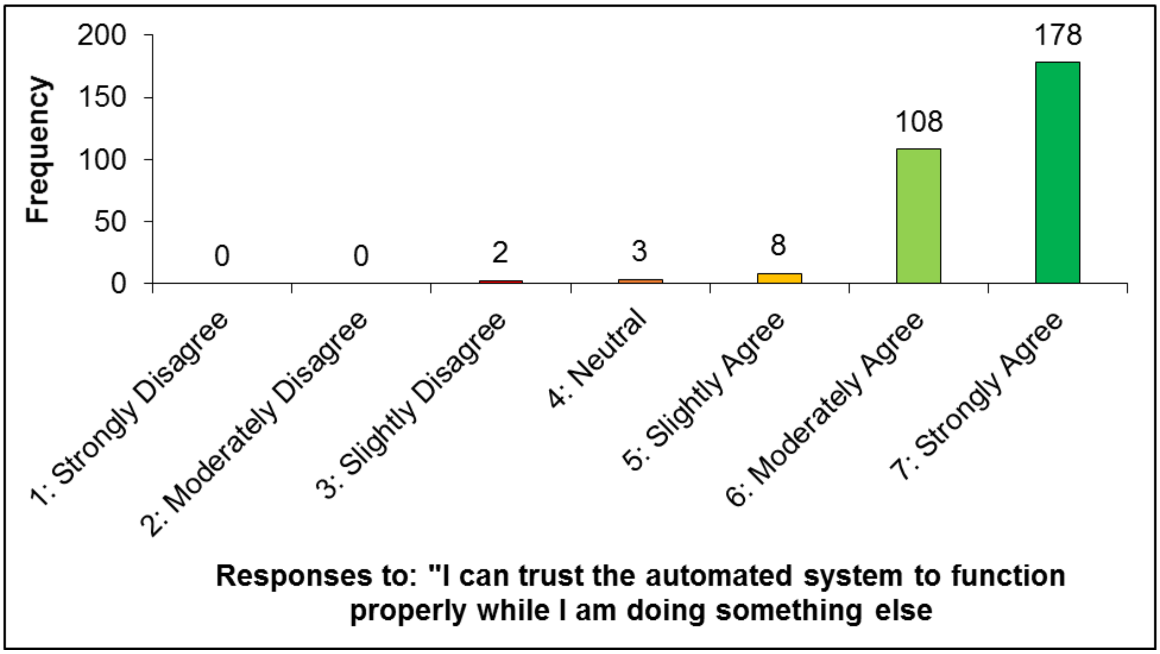



**Figure D-33. Frequency of Rating Levels for Experiment 3**

None of the ratings was below "3" and only 13 ratings (4.4 percent) were below "6." A majority of the ratings were 7 (59.5 percent, *n* = 178), where "7" corresponded to "strongly agree." There were 108 ratings of "6" (36.1 percent), where "6" corresponded to "moderately agree." The two ratings of "3" and three ratings of "4" all occurred at the beginning of the first session, along with three of the "5" ratings. After the first time point, all but 5 of the remaining 274 ratings were at least a "6." The time trends of trust rating for all individuals are displayed in Figure D-34.

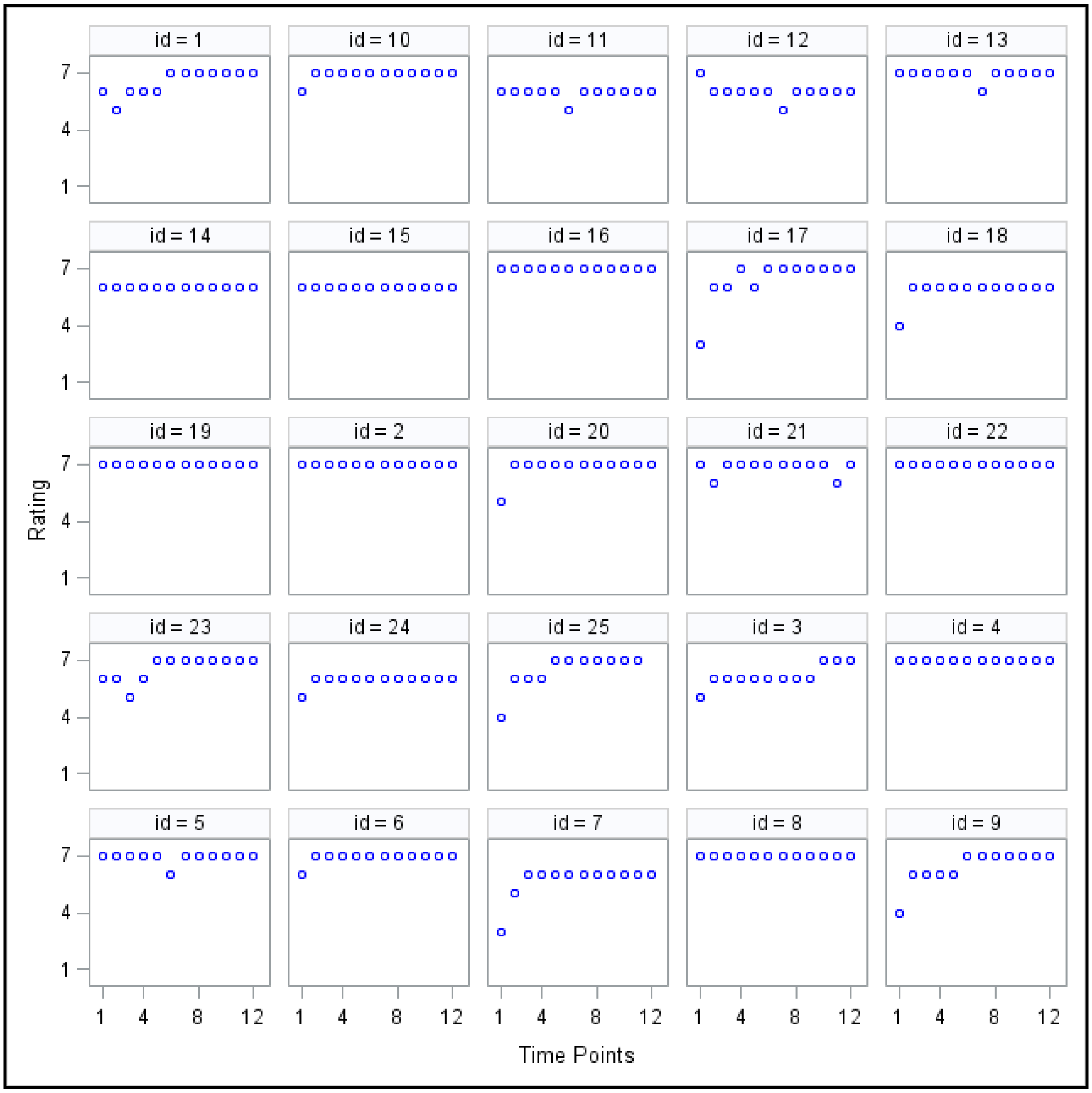



**Figure D-34. Time Trends in Trust Rating for Individuals for Experiment 3**

Because so few of the ratings were below “6”, the response was treated as a binary variable indicating whether a rating of “7” was achieved (“yes” or “no”).

Statistical Analysis

A generalized estimating equation logistic regression model with robust standard errors (taking into account the correlation between time points in the same people) was fit with time considered as a continuous predictor variable and “yes” or “no” for a “7” rating was the response. The model estimates are displayed in Table D-18.

**Table D-18. Analysis of "7" Rating Change over Time, Coefficient Estimates for Experiment 3**

| Parameter | Estimate | S.E. | Z | Pr > \|Z\| | Lower Confidence Limit | Upper Confidence Limit |
|---|---|---|---|---|---|---|
| **Intercept** | -0.28 | 0.36 | -0.78 | 0.4331 | -0.99 | 0.43 |
| **Time** | **0.13** | **0.04** | **2.78** | **0.0055** | **0.04** | **0.21** |

The time estimate of 0.13 (95 percent confidence interval of 0.04 to 0.21) is significant ($p$ = 0.0055), so there is evidence that there is a relationship between time and the probability of rating a "7." Specifically, the estimate of 0.13 indicates that as time increases by one unit, the log odds of rating a "7" increase by 0.123. This can be converted to an odds ratio of 1.13. Hence, with a one-unit increase in time, the odds of rating a "7" are multiplied by 1.13. Thus, the odds would be 3.83 times higher of rating a "7" at the end of the experiment than at the beginning.

A second analysis was conducted where only the time points during the sessions (excluding before and after the sessions) were included. Thus, only the three times before the alert (one for each session) and the three times after the alert (one for each session) were included. The purpose was to determine if any change in the probability of a "7" rating from before to after an alert depended on the alert. A time before the alert was coded as "before," and after was coded as "after." This variable was used along with session and alert (categorical variables) and all of their interactions. The analysis results are displayed in Table D-19.

**Table D-19. Analysis of "7" Rating, Tests for Effect Significance for Experiment 3**

| Effect | DF | Chi-Square | Pr > ChiSq |
|---|---|---|---|
| **Alert Type** | **2** | **6.73** | **0.0346** |
| **Session** | **2** | **6.74** | **0.0344** |
| **Before/After** | 1 | 0.01 | 0.9099 |
| **Alert Type*Session** | 4 | 6.08 | 0.1936 |
| **Alert Type*Before/After** | 2 | 1.07 | 0.5861 |
| **Session*Before/After** | 2 | 1.10 | 0.5769 |
| **Alert Type*Session*Before/After** | 4 | 1.11 | 0.8932 |

The interaction between alert and before/after was not significant ($p$ = 0.5861). The conclusion is that there is no statistically significant evidence that the change in the proportion of "7" ratings from before the alert to after the alert depends on the alert.

Alert type ($p$ = 0.0346) and session ($p$ = 0.0344) registered significance. To investigate this further, the differences in least squares means are displayed in Table D-20 and Table D-21, respectively.

**Table D-20. Differences in Least Squares Means of Log Odds of "7" Rating for Alert Types for Experiment 3**

| Alert Type | Comparison | Estimate | S.E. | Z-Value | Adj *p* Value | Lower Confidence Limit | Upper Confidence Limit |
|---|---|---|---|---|---|---|---|
| **Staged** | **Imminent—No External Threat** | **0.77** | **0.27** | **2.79** | **0.0159** | **0.23** | **1.32** |

**Table D-21. Differences in Least Squares Means of Log Odds of "7" Rating for Sessions in Experiment 3**

| Session | Comparison | Estimate | S.E. | Z-Value | Adj *p* Value | Lower Confidence Limit | Upper Confidence Limit |
|---|---|---|---|---|---|---|---|
| **3** | **1** | **0.95** | **0.35** | **-2.75** | **0.0180** | **0.27** | **1.62** |

The log odds of a "7" rating are, on average, 0.77 higher for the Staged alert than for the Imminent–No External Threat alert (95 percent confidence interval 0.23 to 1.32). This corresponds to an odds ratio of 2.17, indicating that during a Staged alert, participants were 2.17 times as likely to rate "7" for the Imminent–No External Threat. This could be reflected in the fact that more than 10 participants (40 percent) received Staged alerts in the third session compared with 8 participants (32 percent) receiving Imminent–No External Threat alerts, and previous analysis has established that time is correlated with the likelihood of rating a "7."

The log odds of a "7" rating in Session 3 were, on average, 0.95 higher than in Session 1 (95 percent confidence interval 0.27 to 1.62). This corresponds to an odds ratio of 2.58, indicating that in Session 3, participants were 2.58 times more likely to rate a "7" than in Session 1. This is reflective of the previous result that the likelihood of a "7" rating increases with time. Means, standard deviations, modes, minimums, and maximums for the after-experience trust scales are provided in Table D-22.

**Table D-22. Descriptive Statistics for Responses to the After-experience Trust Scales for Experiment 3**

| Variable | Mean | S.D. | Mode (Frequency) | Minimum | Maximum |
|---|---|---|---|---|---|
| **TS1** | 6.8 | 0.4 | 7 (19) | 6 | 7 |
| **TS2** | 6.7 | 0.6 | 7 (20) | 5 | 7 |
| **TS3** | **4.9** | **1.7** | 7 (7) | 2 | 7 |
| **TS4** | 6.3 | 0.8 | 7 (12) | 5 | 7 |
| **TS5** | 6.2 | 1.0 | 7 (13) | 4 | 7 |
| **TS6** | 6.5 | 0.9 | 7 (15) | 3 | 7 |

The recoded question 3, “the automated system gave false alerts,” has both a lower mean rating (4.9) and higher standard deviation (1.7) compared to the responses to all other statements. The mode for the recoded statement 3, given by 7 people, was “7.” This means that to the original statement, “the automated system gave false alerts,” more participants answered “1,” or “strongly disagree,” than any other answer. The highest answer to the original statement was “moderately agree,” given by one participant, implying that this participant felt moderately in agreement that the system gave false alerts. The next lowest mean belonged to statement 5, “I am familiar with the automated system,” with a 6.2 mean rating (S.D. = 1.0). All other means were at least 6.3. Hence, while participants overall agreed that they trusted the system, there were varying responses over whether or not the system gave out false alerts. Additional subjective data for Experiment 3 can be found in Appendix F.

# Appendix E. Trust and Satisfaction Scales

## In-Vehicle Trust Scale

**I can trust the automated system to function properly while I am doing something else.**

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| Strongly Disagree | Moderately Disagree | Slightly Disagree | Neither Agree Nor Disagree | Slightly Agree | Moderately Agree | Strongly Agree |

# After-experience Trust Scales

**Participant Number:**

**Circle the number that best describes your feeling or impression.**

1.) I can rely on the automated system to function properly while I am doing something else.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| Strongly Disagree | Moderately Disagree | Slightly Disagree | Neither Agree Nor Disagree | Slightly Agree | Moderately Agree | Strongly Agree |

2.) The automated system provided the alerts when needed.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| Strongly Disagree | Moderately Disagree | Slightly Disagree | Neither Agree Nor Disagree | Slightly Agree | Moderately Agree | Strongly Agree |

3.) The automated system gave false alerts.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| Strongly Disagree | Moderately Disagree | Slightly Disagree | Neither Agree Nor Disagree | Slightly Agree | Moderately Agree | Strongly Agree |

4.) The automated system is dependable.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| Strongly Disagree | Moderately Disagree | Slightly Disagree | Neither Agree Nor Disagree | Slightly Agree | Moderately Agree | Strongly Agree |

5.) I am familiar with the automated system.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| Strongly Disagree | Moderately Disagree | Slightly Disagree | Neither Agree Nor Disagree | Slightly Agree | Moderately Agree | Strongly Agree |

6.) I trust the automated system.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| Strongly Disagree | Moderately Disagree | Slightly Disagree | Neither Agree Nor Disagree | Slightly Agree | Moderately Agree | Strongly Agree |

# Satisfaction Scale

**Participant Number:**
**Mark an "x" on each line at the point which best describes your feeling or impression.**

1.) Overall, how satisfied are you with the automated system?

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| Extremely Dissatisfied | Moderately Dissatisfied | Slightly Dissatisfied | Neither Dissatisfied Nor Satisfied | Slightly Satisfied | Moderately Satisfied | Extremely Satsified |

2.) How satisfied were you with the number of alerts provided?

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| Extremely Dissatisfied | Moderately Dissatisfied | Slightly Dissatisfied | Neither Dissatisfied Nor Satisfied | Slightly Satisfied | Moderately Satisfied | Extremely Satsified |

3.) How satisfied were you with the types of alerts provided?

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| Extremely Dissatisfied | Moderately Dissatisfied | Slightly Dissatisfied | Neither Dissatisfied Nor Satisfied | Slightly Satisfied | Moderately Satisfied | Extremely Satsified |

4.) The automated system's alerts provided sufficient time to make a decision.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| Strongly Disagree | Moderately Disagree | Slightly Disagree | Neither Agree Nor Disagree | Slightly Agree | Moderately Agree | Strongly Agree |

5.) The automated system's alerts provided sufficient information to make a decision.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| Strongly Disagree | Moderately Disagree | Slightly Disagree | Neither Agree Nor Disagree | Slightly Agree | Moderately Agree | Strongly Agree |

6.) I would use this type of automated system during my normal driving.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| Strongly Disagree | Moderately Disagree | Slightly Disagree | Neither Agree Nor Disagree | Slightly Agree | Moderately Agree | Strongly Agree |

7.) I would like to have this type of automated system as part of my current vehicle.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| Strongly Disagree | Moderately Disagree | Slightly Disagree | Neither Agree Nor Disagree | Slightly Agree | Moderately Agree | Strongly Agree |

8.) I would like to have this type of automated system as part of a future vehicle.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| Strongly Disagree | Moderately Disagree | Slightly Disagree | Neither Agree Nor Disagree | Slightly Agree | Moderately Agree | Strongly Agree |

# Appendix F . Subjective Data

In addition to completing in-vehicle trust scales and after-experience trust scales (as well as satisfaction scales for Experiment 2 only), participants in all three experiments provided their impressions of the system they experienced by participating in an after-experience interview. (The satisfaction scales were used in Experiment 2 only.) The results from the after-experience interviews are summarized below.

## After-experience Interview Analysis

### Purpose

The purpose of the qualitative content analysis was to determine participants' perceptions of the experience and the automated system being tested. The results of this analysis can be used to help researchers to potentially understand participant behavior that varies from the norm (e.g., participants who elected not to engage in a non-driving task during Experiment 3).

### Data Analysis Method

The approach that was used to analyze the results of the interviews and focus groups was an adaptation of framework analysis, a methodology developed in the 1980s at the National Centre for Social Research in Britain (Ritchie, Spencer, and O'Conner, 2003).This matrix-based approach allows researchers to manage and analyze qualitative data sets in a logical and thorough manner. Through this iterative approach, data are reduced through summarization and synthesis; however, links to original data are retained. The output allows for comprehensive and transparent data analysis.

The National Centre for Social Research notes several advantages associated with the use of framework analysis (2012). First, data are ordered in descriptive groupings (columns), which aids in question-focused analysis (e.g., "why did participants fail to engage in a secondary task?"). Second, the matrix-based approach aids the search for explanation (looking across rows). However, as with any method, limitations exist. Qualitative analysis such as this is time- and labor-intensive. Additionally, researchers need to ensure that the focus remains on the process as opposed to the outcome. Further, when analyzing data, researchers should be reflexive and critical and should not force data into themes or categories (National Centre for Social Research, 2012).

The steps that were taken to carry out the framework analysis were as follows:

1. **Determining Analysis Focus:** Researchers determined that the focus of the framework analysis would be on the factors contributing to the Guidelines associated with the use of automated vehicles.

2. **Identifying and Transcribing Key Audio Clips:** Interviews were reviewed to determine the location of any discussion related to the four key areas of interest. Due to the length of the recordings, each audio file was transcribed in full. For quality control purposes, transcripts were reviewed by another researcher who was involved in the interview process. The researcher who led the interviews attempted to fill in any gaps where the transcriptionist was unable to decipher the wording of a conversation. For the two interviews without audio recordings, the researchers' written notes were consulted. The result of this thorough process is a complete set of transcripts that could be used for analysis.

3. **Familiarization**: A researcher read over each of the transcripts to become familiar with the data set.

4. **Identifying a Thematic Framework:** A researcher conducted a review of the data set and identified three primary themes: Comfort with System, Alert Reactions, and Design Suggestions. A fourth theme, Additional Thoughts, was also included to capture participant reactions that did not fit into one of the main themes. For each theme, a list of key subthemes was identified. The themes and subthemes were then arranged in a logical order and an index was created. For instance, under the theme of Alert Reactions was the subtheme Operator Actions.

5. **Indexing and Charting:** The index was systematically applied to the data set and comments were identified and marked by theme and subtheme. Multiple subthemes were assigned to the statements where appropriate. The indexed comments were arranged into one Microsoft Excel® spreadsheet. The spreadsheet was then sorted by theme. Individual spreadsheets (or thematic charts) were then created for each theme. These spreadsheets or thematic charts were further sorted by subtheme. Finally, within subthemes, categories were created. While the thematic framework and indexing stages were carried out by one primary analyst, the outcomes were reviewed by members of the research team. If there were questions about how a comment had been indexed, the research team discussed the comment and came to an agreement. Additionally, because comments are complex, they are not mutually exclusive. For example, the comment "I thought it was excellent. I had a lot of trust in it. Performed really well. I was impressed," reflects several different ideas and would thus be included in multiple categories. The result is the creation of useful "piles" of data about the same ideas. Figure F-1 illustrates the output of the framework analysis.

6. **Interpretation:** As a final step in the framework analysis process, the themes, subthemes, and categories captured and detailed in the charts were used by the research team to better understand opinions and preferences related to the automated system being tested.

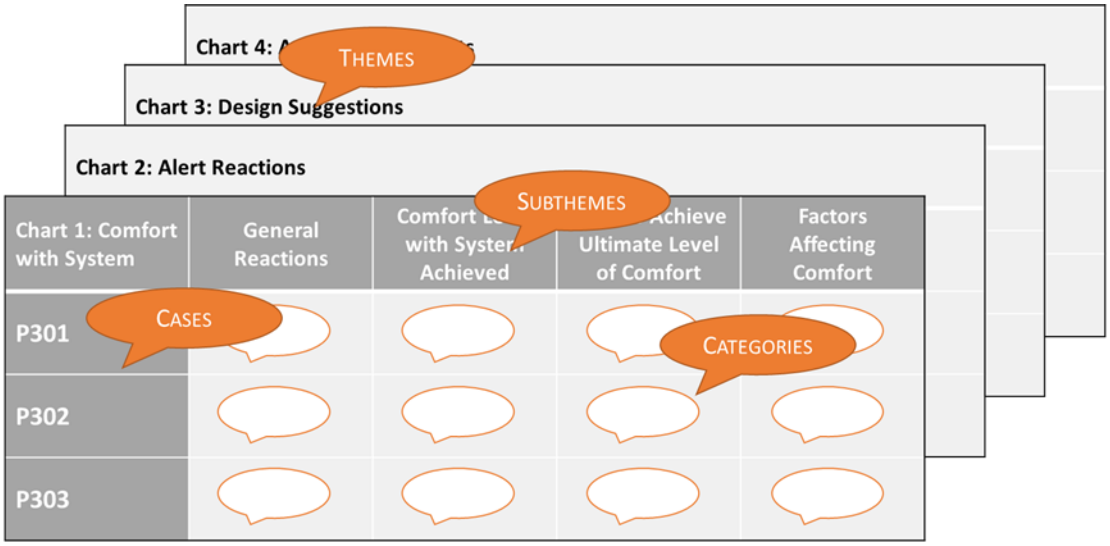



**Figure F-1. Illustration of Framework Analysis. Adapted from "The Framework Approach to Qualitative Data Analysis," by The National Centre for Social Research (2012)**

## Survey Instrument

Participants were asked the following six questions:

1. Overall what are your thoughts on the automated system? (Follow-up with questions on the key points the participant makes.)
2. Can you describe how comfortable you were using the automated system? (Follow up with:
    a. How quickly (specific or relative time) the participant reached that level of comfort?
    b. Anything that they feel affected their level of comfort?)
3. What were your thoughts when the automated system provided you with information or alerts?
4. When the automated system provided a message, what were your first thoughts and actions? (Follow-up with questions on the thoughts and actions described.)
5. If you were talking to a design team, what concerns would you have about an automated system such as you experienced?
6. Is there anything else regarding your experience today that you would like to share?

# Experiment 1 Analysis Findings

Following the in-vehicle experience, participants were asked the six questions listed above. The interviews were audio recorded as a backup to researcher notes. Digital audio recordings were made into transcripts but no individual names were attached to any comments. Responses for all 25 participants were recorded and transcribed. The transcripts and notes served as the basis for a qualitative content analysis.

# Theme 1: Comfort with System

Theme 1 addressed participant's comfort with the system. Generally, information within this theme was drawn from responses to the first two questions from the after-experience interview:

1. Overall what are your thoughts on the automated system? (Follow-up with questions on the key points the participant makes.)
2. Can you describe how comfortable you were using the automated system? (Follow up with:
    a. How quickly (specific or relative time) the participant reached that level of comfort)
    b. Anything that they feel affected their level of comfort?)

However, as framework analysis involves grouping concepts together, responses from other questions which reflected participants' comfort with the system were included here.

***Subtheme 1.1: General Reactions***

Comments were categorized into the following categories:

- Comfort with or trust in the system-related comments (13 responses)
- System performance-related reactions (15 responses)
- General reactions (11 responses)
- Societal benefits-related reactions (3 responses)
- Potential for misuse (4 responses)

Examples of representative comments follow in Table F-1. Comments suggesting factors that influenced a participant's comfort with or trust in the system were further explored in the Subtheme 1.4 Factors Affecting Comfort Levels analysis, while comments suggesting design modifications were included in the Theme 3 Design Suggestions analysis.

**Table F-1. Categories Within Subtheme 1.1 General Reactions**

| Category | Example Reaction |
|---|---|
| **1.1.1: Comfort with or trust in the system-related comments** | • "It was very reliable and dependable and I trust it completely and I've never experienced anything else like it." *(18–24 age group)*<br>• "I never gave it a full 7 because I do trust the system, but I don't trust automated driving. Just like I trust autopilot, but I would rather have a pilot land and take off, just because no machine can make decisions like we can. So, that's why I don't fully trust automated driving." *(18–24 age group)* |
| **1.1.2: Performance-related reactions** | • "It was very dependable [because] when a car came in front of me it got slower, it knew what was going on all around the car to send signals to me." *(18–24 age group)*<br>• "I liked it. I thought it performed well. [The system] kept you centered on the road. We had the one experience there where that car jumped in front of us and it slowed us right down. I thought that was a pretty good feature." *(55+ age group)* |
| **1.1.3: General reactions** | • "I think it's incredible. I like to see this kind of system being developed. It's very exciting technology." *(55+ age group)*<br>• "I thought it was nice. I never used it before, so it was my first time. I felt comfortable, overall." *(18–24 age group)*<br>• "I love it! It is so cool! Very cool. I love it." *(40–54 age group)* |
| **1.1.4: Societal benefits-related reactions** | • "Something that needs to be done for society." *(25–39 age group)*<br>• "I think the safety factor, for one thing, is great, you know, especially if other cars were doing the same thing and were aware of what was going on. I don't know that the being able to do the multitasking thing is … I don't know. I just think the safety end of it is probably the best idea." (*55+ age group)* |
| **1.1.5: Potential for misuse** | • "It's quite interesting. I can see it being beneficial. I can also see it – people utilizing it for areas where it's not designed to be utilized. [I] could see somebody getting hold of Mommy and Daddy's car that has this and driving around and enjoying time in the driver's seat that somebody normally wouldn't be able to enjoy." *(25–39 age group)*<br>• "The only thing that bothers me is that somebody may rely too much on the system and think it's a cure-all or a band aid. It's really only a Band-Aid. It won't fix everything, you know." *(55+ age group)* |

***Subtheme 1.2: Comfort Level Achieved***

Participants were asked to describe how comfortable they were in using the automated system. Levels of comfort were self-defined, i.e., not provided in response to a Likert-type scale. A summary of these self-reported comfort levels are included in Table F-2.

**Table F-2. Experiment 1 Participants' Self-reported Comfort Levels**

| Self-reported Comfort Level | Responses |
|---|---|
| **Comfortable** | 4 |
| **Decently** | 1 |
| **Moderately** | 1 |
| **Pretty** | 4 |
| **Very** | 9 |
| **Other** | 6 |

Several participants provided general descriptions of their comfort levels. Example of these general description responses are presented as follows.

- “My comfort level increased as I drove along and it got to the point where I was almost oblivious as to what was going on.” *(55+ age group)*
- “After a while I got to the point where I didn’t have to keep looking up and making sure that the system was working” *(55+ age group)*
- “I was fine with it. I had no reservations.” *(25–39 age group)*
- “Well, I at first was a little concerned, but after feeling it and sitting there I was able to play with the computer and forget about the driving.” *(55+ age group)*
- “Getting it down was fine for the most part, you know, find the center lane, kind of thing. Press the button, you know – doing it, the logistics were easy...” *(18–24 age group)*
- “At first, not as much as I am now. It definitely grew fast on me. It’s reliable and trustworthy. Drives the car, you know, sometimes I was just sitting there and just kind of laughed as drove past these cars.” *(25–39 age group)*

***Subtheme 1.3: Time to Achieve Ultimate Level of Comfort***

Participants were asked how quickly (specific or relative time) the participant reached that level of comfort). Similar to Experiments 2 and 3, participants responded in a number of ways, including:

- A specific amount of time (e.g., 15 minutes).
- A specific number of laps, which allowed for the approximation of time (e.g., 1 lap equates to 7 minutes).
- A specific point in the study.
- A temporally vague response.
    - “Very quickly” (2 responses)
    - “Pretty quickly” (3 responses)
    - “Just because it was totally new to me, when you’re coming up behind/in front of somebody, when somebody’s in front of you and you’re closing in on them, and then the system basically slows down, you know, for another vehicle, you know, moving over, whatever, and then speeds up. After encountering it a couple of times, I felt very comfortable with it. Because it was new, it was tough at first.” *(55+ age group)*

With the exception of the temporally vague responses, times were recorded or approximated. The majority of participants reported that they reached their ultimate level of comfort with the system by the 15-minute point.

### *Subtheme 1.4: Factors Affecting Comfort Levels*

Four categories emerged from the data: Time in Vehicle, System Alerts, System Performance, and Awareness of the Testing Situation.

<u>Category 1.4.1: Time in Vehicle</u>

Nine participants noted that their confidence in the system improved over time; however, one noted that his comfort decreased over time. (Note, this participant's reaction is in response to the alerts used as part of the test condition.)

- *"It started off, it was very intriguing at first so I was kind of side-tracked by it like oh, that's cool. It's doing this. So, I felt okay. And then I guess it just got deducted each time something would happen without, you know, and give me a scare. So, I guess I started off pretty comfortable and then it got deducted when it would shut off without telling me or if it [didn't vibrate] when it should have…something of that nature." (18–24 age group)*

<u>Category 1.4.2: System Alerts</u>

Sixteen participants commented on the alerts in terms of their comfort in or trust with the system. Participants specifically noted that they liked the haptic (8 responses) and visual alerts (9 responses).

- "One of the best features, I think, is the vibration in the seat to alert me. I found myself first looking [redacted] to see if the colors had changed, but when the seat would vibrate that would automatically alert me." (*55+ age group)*
- "[Even] though I was preoccupied, I was still able to visually see when the warnings went off, even feel them when the chair vibrated." (*18–24 age group)*
- "I liked how it gave the seat a buzz, and other times it was just the color changing. The buzz was a lot better. More sensory to it." *(18–24 age group)*

Of those participants, seven expressed that they wanted and/or expected the haptic alerts to be presented consistently when other alerts were issued or system faults occurred.

- "Sometimes the seat didn't vibrate, it was kind of on and off. You couldn't rely on the vibrating seat to tell me the system was shutting off or freaking out." *(18–24 age group)*
- "[I] I was a little taken aback by the fact that it didn't hold me in that lane nor did the seat vibrate." *(55+ age group)*

Similarly, several participants (5 responses) indicated that they wanted and/or expected consistent visual alerts.

- "I think the first time that it went [REDACTED] and it really swerved out of the lane, that kind of surprised me because I was expecting it to go [Cautionary] first and then I'd have the appropriate time to respond and grab the wheel. So that kind of took me off guard, but it really didn't necessarily decrease my comfort level that much." *(40–54 age group)*
- "On the negative side, when it turned off once or twice and the only indication was a lack of the light [REDACTED]. And I know the first, I think it happened twice, the first time I wasn't, it didn't really dawn on me at first, quickly enough that I needed to take control. Because, as opposed to some of the other situations, it didn't change, it didn't go to a [REDACTED] or didn't give me any seat vibration or anything like that." *(55+ age group)*

Participants also recommended the inclusion of an auditory alert (4 responses).

- "[And] I know nobody likes a woman's voice to say, "Hey, you do this," but that's… Something might be helpful." *(55+ age group)*
- "[Maybe] also if there was like a beep, like a sound, vibration, sound, light – just a trifecta, I guess you could say – would help comfort." *(18–24 age group)*
- "All the alerts seemed very good. I think maybe an audio or a verbal alert, 'system failure,' or something along those lines could be helpful." *(55+ age group)*
- "[I] needed more…the vibration on the seat was good and the flashing of the light [redacted] was very good. Maybe an audible would be good also." *(55+ age group)*

<u>Category 1.4.3: System Performance</u>

Participants (16 responses) had a variety of reactions to the system performance, both positive and negative. Contributing to participants' comfort levels was the system's ability to maintain lane position (5 responses) and regulate the vehicle's speed in response to the other vehicles (8 responses). A representative comment is the following:

- "It kept you centered on the road. We had the one experience there where that car jumped in front of us and it slowed us right down. I thought that was a pretty good feature." *(55+ age group)*

The system's effective performance enabled participants (3 responses) to comfortably focus their attention elsewhere, such as on the distraction tasks.

Ten participants made comments regarding what they perceived to be system failures and unplanned alerts. In this case, participants viewed not only the planned system failure (e.g., the lane drift test condition) but also the variations in alert presentations as a system failure. For example, as noted in the aforementioned section, participants expected both the haptic and visual alerts. Illustrative comments are as follows:

- "Negative was the one time where it just randomly shut off with no warning, but I know it's still a system in the making, so it's going to have those bugs. And other times it was really cool how it would sense the lane change in order to give you the warning. " *(18–24 age group)*
- "One thing I guess I would say that really bothered me would be, if I'm doing something else and it started going off and maybe my attention was on what I was doing instead of noticing that the alarm was going off, particularly the one time when it like had drifted over to the other lane and I didn't notice it, you know. So that was a little unsettling." *(40–54 age group)*
- "When it's done, I think it will be a good thing. Right now it was just seemed too glitchy and I don't think I'd really feel all that comfortable and certainly not in heavy traffic. Now there were times when it just cut out. And the weird thing is it usually only vibrated [REDACTED]. If it, when it went [REDACTED] it wouldn't vibrate, and then later on in the test when it would just cut out without even any indication there was no vibration." *(40–54 age group)*

Category 1.4.4: Awareness of the Testing Situation

Participants noted their awareness of the test situation. In one case, this awareness led the participant to conclude that the number of alerts experienced was probably part of the test conditions.

- "Just the random system failures, but I'm assuming that was part of the test. But, yeah, if that would happen during the drive, then I would have a problem with it – because if it had that many missed, or if it threw that many false failures, or whatever, then there would be a problem with that." *(18–24 age group)*

Two others commented on the secondary tasks associated with the experiment.

- "Because here, if you're doing it here if anything happens, you can do it, your eyes are there versus here. Now, I don't know, maybe that's what you want to test to make sure people aren't checking the road, but I think in the real world, I think it would be much easier for me to use versus doing here. You need to concentrate when you're putting things in." *(55+ age group)*

Further, participants noted that the presence of in-car experimenters (3) and the closed track conditions may have affected their comfort with the system.

- "I was comfortable with it, but I don't know whether it's because other people were in the car that really knew what it was doing… [the system] seemed to be quite dependable from what we could tell in a controlled environment. You know, how that would react on the road, I don't, I imagine it would be the same thing." *(55+ age group)*

Additionally, it is possible that some of the negative reactions associated with the number or nature of alerts which were previously noted were the result of the test situation.

## Theme 2: Alert Reactions

### *Subtheme 2.1: General Thoughts in Response to Being Provided With Information or Alerts*

When asked about their thoughts in response to being provided with information or alerts, participants (13 responses) made general comments about the alerts (e.g., helpful, effective, good) and also voiced concerns (8) associated with receiving an alert or criticism about the alert (e.g., would it function, hard to keep your eye on the screen, could it react sooner). Suggestions regarding ways to improve the alert are captured in the design-focused question.

Fourteen participants specifically noted that the alerts prompted them to regain situational awareness either by figuring out the appropriate response to the alert or situation or by preparing to take manual control. For example, participants noted:

- "I went from doing the task at hand to concentrating on the road. Fully concentrating on the road rather than the task that I was assigned to do." *(25–39 age group)*

- "It was what did I do wrong? What's happening? And then it was like okay, we just have to reset it and it's no big deal. Reset it and we went on. [Resetting] it was very easy. I felt very comfortable doing that." *(40–54 age group)*

***Subtheme 2.2: First Thoughts and Actions When the System Provided a Message***

Participants were asked for their first thoughts and actions when the system provided a message. Three categories of thoughts and actions were identified:

- Category 2.2.1: Regain situational awareness,
- Category 2.2.2: Regain manual control, and
- Category 2.2.3: Follow or respond to system instructions.

The number of responses for each category is summarized in Figure F-2.

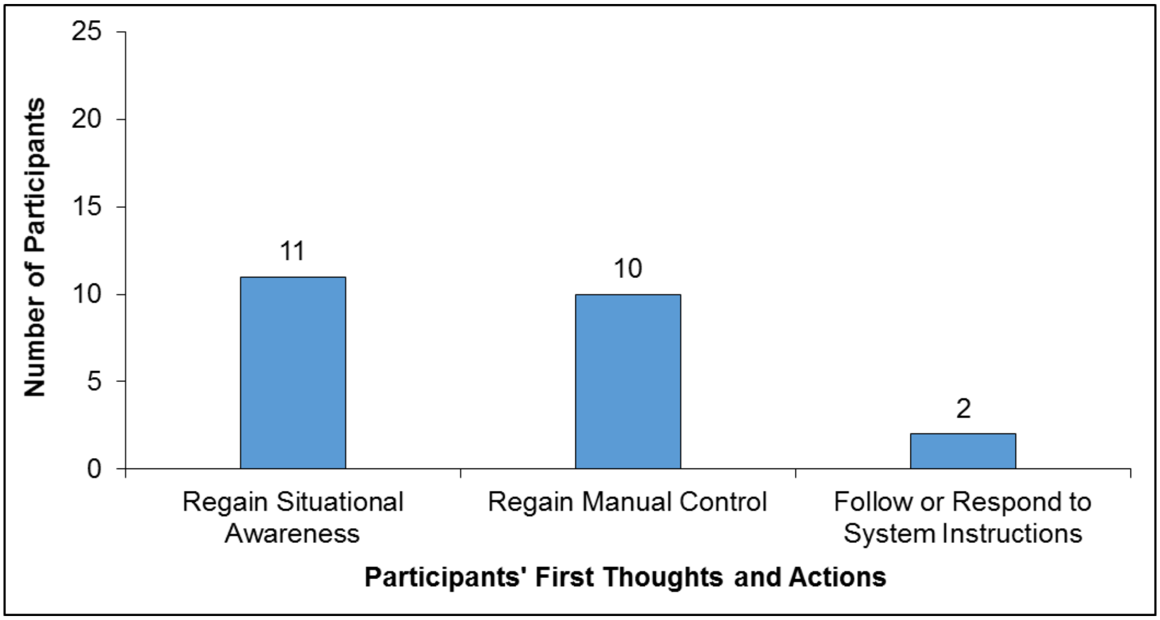



**Figure F-2. Experiment 1 Participants' First Thoughts and Actions when Provided a Message**

## Theme 3: Design Suggestions

Throughout the after-experience interview process, participants provided feedback regarding the design of the automated system. Additionally, they were specifically asked, "If you were talking to a design team, what concerns would you have about an automated system such as you experienced?" Design-related suggestions and feedback have been grouped together and analyzed as Theme 3. Four subthemes emerged from this analysis. While concerns were voiced regarding the design-aspects of the system, it should be noted that several of the concerns voiced were likely influenced by the test condition.

***Subtheme 3.1: General Design-Related Comments***

Concerns were voiced regarding the following design-related aspects of the system:

- Overall system operation, particularly in regards to making the steering smoother (3 responses) and ironing out "the glitches."

- Need for system training (1 response)
    - “I guess training would have to come with the car, that’s what I’d say.” *(25–39 age group)*
- Concerns regarding installation and maintenance (3 responses)
    - “I would want it installed in the factory, not an aftermarket guy doing it. Just because I think that if it was done in the factory it would be at the best possible place for it… it’s very important that it would be done properly.” *(40–54 age group)*
    - “Cost. If I got a new car and that broke then I wouldn’t want to pay for that. It sounds expensive.” *(18–24 age group)*
- The potential for perceived or actual system misuse (5 responses)
    - “[My] only concern would be somebody falling asleep.” *(55+ age group)*
    - “My concern would be that people would be overly dependent on it and maybe expect or maybe not realizing or not always understanding what it won’t do for them.” *(55+ age group)*

***Subtheme 3.2: Display Design***

In regards to the system display, the following display design-related suggestions were offered:

- “[I] think it could be simpler, either on or off or whatever, you know.” *(55+ age group)*
- “In terms of the traffic I had, it came up on the screen, but that wasn’t always obvious to me…I would have appreciated some other kind of activity to alert me that I was coming up on traffic even though it was adapting to the speed, if it was very gradual, you wouldn’t see that because I wasn’t keyed in on that informational screen.” *(55+ age group)*

***Subtheme 3.3: Driver-Vehicle Interface Design***

Several suggestions were offered regarding the driver-vehicle interface design. First, participants had suggestions regarding the activation button. Specifically, two participants suggested that the button be moved to the steering wheel or the control stalk *(25–39 and 55+ age groups).* Another commented that he had mistaken the button for a light *(18–24 age group).* Continuing with suggestions for the steering wheel modifications, one participant suggested that the task pad be placed in the center of the steering wheel *(55+ age group).* A second indicated potential misuse that could result from the current configuration.

- “It seemed like you don’t really have to grab it and was what the message was trying to get across – grab the steering wheel. You know, rather there’s a touch sensor or something on it, right? Because that would be easy to beat” *(25–39 age group)*

Two participants suggested that the notification display screen should be moved to a heads-up display. One had a specific recommendation regarding the size of the heads-up display:

- I would just want to make sure that [the notification display screen] stays hidden, well not hidden, but in a place where you can still see it but still see the whole scope of the windshield. *(40–54 age group)*

A final participant drew attention to the use of colors as an indicator, specifically, the similarities between two of the colors when shows in LED form.

- "Maybe that [color] light could have been a little bigger because there are some times, you know, it seems like I was just looking at it every time it went off, but like if I was looking at an email, you know, should I tell myself could I see that [color] light. Could I really see the [engaged indicator] compared to the [warning], because they are similar kind of, the [alert colors]? I suggest changing the [alert color]." *(25–39 age group)*

***Subtheme 3.4: Alert Design***

As noted in previous sections, participants indicated a desire for consistent haptic alerts and for additional alerts. Additional alert-related suggestions are as follows:

- The need for an audio alert (8 responses), specifically verbal alerts (4 responses).
    - Potential to make audio alerts optional for use during training (e.g., use the audio to train system users, but allow experienced users to disable).
- Enhanced traffic alerts to alert of oncoming traffic (2 responses) and, if possible, have the alerts be directional (i.e., to indicate traffic approaching on the left or right sides).
- A driver alertness monitor to keep drivers engaged over long periods of time (2 responses).

## Theme 4: Additional Thoughts

Participants were asked if they had any additional thoughts regarding the system or the experience that they would like to share. Where appropriate, comments were grouped with their corresponding themes. For example, comments related to the system design that were not voiced in response to the system design question were included in the design discussion. Hence, the Additional Thoughts theme encompasses only those comments not previously analyzed. The remaining comments have been categorized as follows:

- Category 4.1: Positive comments about the experience (12 responses)
- Category 4.2: Positive comments about the system (10 responses)

***Interpretation***

Generally, participants indicated that they had a positive experience and seemed impressed by the system. The majority of participants reported feeling comfortable with the system and achieving their ultimate level of comfort within the first 15 minutes of use. Participants indicated a desire for multimodal alerts that include not only visual and haptic components but also an auditory component. As system indicator alerts are tied to specific colors, the use of additional auditory and/or haptic alerts may not only address the concerns voiced by participants but also accommodate future users who are affected by color blindness.

# Experiment 2 Analysis Findings

Following the in-vehicle experience, participants were asked the six questions listed at the beginning of this appendix. The interviews were audio recorded as a backup to researcher notes. Digital audio recordings were made into transcripts but no individual names were attached to any comments. Responses for all 56 participants were recorded and transcribed. The transcripts and notes served as the basis for a qualitative content analysis.

## Theme 1: Comfort with System

Theme 1 addressed participant's comfort with the system. Generally, information within this theme was drawn from responses to the first two questions from the after-experience interview:

1. Overall what are your thoughts on the automated system? (Follow-up with questions on the key points the participant makes.)
2. Can you describe how comfortable you were using the automated system? (Follow-up with:
    a. How quickly (specific or relative time) the participant reached that level of comfort)
    b. Anything that they feel affected their level of comfort?)

However, as framework analysis involves grouping concepts together, responses from other questions which reflected participants' comfort with the system were included here.

***Subtheme 1.1: General Reactions***

Comfort with or trust in the system-related comments (20 responses)

- System performance-related reactions (43 responses)
- General reactions (23 responses)
- Benefit-related reactions (either for self or society; 10 responses)
- Potential for misuse or failure (6 responses)

Trust-related comments were often tied to an individual's positive experience with the vehicle (3 responses), overall trust (or lack of trust) in automation (12 responses), a need for greater understanding regarding the system's full range of capabilities (9 responses), and/or deviations from the participant's mental model associated with driving (7 responses). Ten respondents specifically noted that they trusted the system to take control or that they would like to have a similar system in their own vehicle.

Examples of representative comments follow in Table F-3. Comments suggesting factors that influenced a participant's comfort with or trust in the system were further explored in Subtheme 1.4 Factors Affecting Comfort Levels analysis, while comments suggesting design modifications were included in the Theme 3 Design Suggestions analysis.

**Table F-3. Categories Within Experiment 2 Subtheme 1.1 General Reactions**

| Category | Example Reaction |
|---|---|
| **1.1.1: Comfort with or trust in the system-related comments** | • "I liked it, I thought it was weird to get used to at first, just because I never used anything liked it. But I liked the freedom it gave me while I was driving." *(18–24 age group)*<br>• "I thought it was really cool, I was a little wary of it at first, but it was surprisingly very reliable and trustworthy… Just the whole idea of not being in control of the vehicle and letting go of you know, and not having control, immediate control of the wheel or speed." *(25–39 age group)*<br>• "I don't really know how useful it would be for myself. I think it's interesting, but I'm skeptical to use it and not sure it would ever be a feature I would interested in looking for in a car just because I trust myself and my judgments and my actions. There's just too much room for error, to… I don't know computers and things like that, technology are not foolproof as well, I mean, humans aren't either, but I feel more comfortable knowing that I'm in control." *(25–39 age group)* |
| **1.1.2: Performance-related reactions** | • "In general it was a lot smoother than I probably expected." *(18–24 age group)*<br>• "It's pretty nice. I can see the utility in it immediately. One of the things that I was wondering about is how it performs in different weather conditions. I know that there can be cancellations for weather conditions, so I was wondering if that would be covered as well because I feel that there will be people that will try it out during those conditions so for their safety it'd be best to test it in a secure controlled environment such as this." *(18–24 age group)*<br>• "[It] drove great, and it was great when I learned my first experience when it was slowing down when there were cars next to me, when it needed to brake and speed up. So overall I really did like that, and you really got to really learn from it – like in the beginning I didn't really know anything about it and now it's like you know what the alerts and signs and stuff are." *(25–39 age group)* |
| **1.1.3: General reactions** | • "Pretty cool. I like it." *(25–39 age group)*<br>• "Pretty decent." *(40–54 age group)*<br>• "I like it."*(40–54 age group)*<br>• "I think it's cool." *(55+ age group)*<br>• "It was interesting." *(55+ age group)* |
| **1.1.4: Benefits-related reactions** | • "It's pretty cool, it's nice, it really would work well on long distance." *(18–24 age group)*<br>• "I was surprised by how efficient it already seemed cause I had never encountered anything like that, and then it was, I think it was helpful cause you can get other things done while you're supposed to be driving" *(18–24 age group)*<br>• "I think the technology is incredible. I feel that it allows you to multi-task. I'm not sure if that's necessarily what I want from all the drivers on the road, but the technology is amazing." *(40–54 age group)*<br>• "I like the safety aspect of it, I like the convenience of it. I do a lot of traveling, and I use cruise a lot, so if I had this cruise, I could multitask which would be nice because we have computers in our vehicles and it's so tempting to look over there while I'm driving, so I actually push it because if I don't I'm going to look over there when someone's instant messaging." *(55+ age group)* |
| **1.1.5: Potential for misuse** | • "The automated system is based on a great concept, however; I have a great fear and trepidation towards the common user given that it only takes one instance, or one moment of failure in the program to cause devastating, memorable human and animal catastrophe." *(18–24 age group)*<br>• "I've been driving for a long time and I feel that with the advances technology has made, human beings haven't become more responsible using it. The tools are terrific the human beings are not being responsible." *(40–54 age group)*<br>• "It's a neat concept. I see flaws down the road. People will rely on it too much and forget to drive. And then I'll say when they've been in an accident, 'But the automated car was supposed to do that!'" *(55+ age group)* |

***Subtheme 1.2: Comfort Level Achieved***

Participants were asked to describe how comfortable they were in using the automated system. Levels of comfort were self-defined, i.e., not provided in response to a Likert-type scale (Table F-4). A summary of these self-reported comfort levels are included in Table F-5 and the subsequent discussion.

Four participants provided general descriptions of their comfort levels. These general description responses are presented as follows.

- “Once, after the first couple of laps, it’s really simple to use. I was surprised, you know it’s a push of a button.” *(25–39 age group)*
- “In the beginning, nervous, kind of uncomfortable because I wasn’t used to it – not being able to drive on your own or even use the pedal or steering wheel. But, in the end, I was extremely comfortable.” (25–39 age group)
- “It was pretty easy to use. It wasn’t a very hard system to use, but I don’t know all of the features that are available. From what you guys let me see, I thought it was easy to use.” *(40–54 age group)*
- “I wish I could have done 70 so I could see how it worked.” *(55+ age group)*

**Table F-4. Experiment 2 Participants' Self-reported Comfort Levels**

| Self-reported Comfort Level | Responses |
|---|---|
| **Moderately** | 1 |
| **Fairly** | 6 |
| **Relatively** | 1 |
| **Comfortable** | 11 |
| **Pretty** | 15 |
| **Really** | 1 |
| **Very** | 14 |
| **Extremely** | 3 |
| **Other** | 3 |

**Table F-5. Examples of Comments Associated with Experiment 2 Participant's Self-reported Levels of Comfort with the Automated System**

| Self-reported Level of Comfort | Example Response |
|---|---|
| **Moderately comfortable** | "Moderately comfortable." *(40–54 age group)* |
| **Fairly comfortable** | "Fairly comfortable. I still don't fully trust it." *(18–24 age group)* |
| **Relatively comfortable** | "Overall, it was pretty easy to operate so I'd say I felt relatively comfortable." *(55+ age group)* |
| **Comfortable** | "The beginning, I was not comfortable at all. After the third session I felt very comfortable in that you built up a trust that the car was going to do what it was supposed to." *(55+ age group)* |
| **Pretty comfortable** | "Pretty comfortable. I think the only discomfort was not fully understanding all of the errors or all of the warnings, but I could rely on the system." *(40–54 age group)* |
| **Really comfortable** | "After the first round, like the first lap, I was still unsure, then right away I got like really comfortable with it." *(18–24 age group)* |
| **Very comfortable** | "I'm very comfortable using the automated system, I don't use cruise control and so for someone who doesn't use cruise control, I found it to be easy to understand. I was a little disappointed with the number of false negatives that were evident. Meaning that the system alerted me to something and I looked up and I could not see that there truly was a problem. But I believe I would rather have it that way than the opposite way, where there is a problem and I'm not alerted." *(40–54 age group)* |
| **Extremely comfortable** | "Extremely comfortable, probably too comfortable." *(55+ age group)* |

***Subtheme 1.3: Time to Achieve Ultimate Level of Comfort***

Participants were asked how quickly (specific or relative time) the ultimate level of comfort was achieved. Similar to Experiments 2 and 3, participants responded in a number of ways, including:

- A specific amount of time (e.g., 15 minutes, in the second hour).
- A specific number of laps, which allowed for the approximation of time (e.g., one lap equates to approximately 5 minutes).
- A specific point in the study (e.g., “in the first session,” “halfway through the second session”).
- A temporally vague response.
    - “Really quickly. I think it’s pretty intuitive.” (*18–24 age group)*
    - “Very quickly.” *(40–54 age group)*

With the exception of the two temporally vague responses noted above, times were recorded or approximated. When participants provided a specific or potential range of times (e.g., “after about an hour or two of using” or “in the first session”), the greatest time was used (e.g., 120 minutes, 60 minutes).

When compared with Experiments 1 and 3, individuals noted that it took them longer to achieve their ultimate level of comfort with the system. In Experiments 1 and 3, the majority of participants achieved their ultimate level of comfort by the 15-minute point, while in Experiment 2, the majority of participants achieved their ultimate level of comfort by the 30-minute point. Additionally, nine participants noting that they did not reach their comfort level until the second half of the experiment (e.g., 90 to 180 minutes). The longest report time to achieve comfort with the system was the “first quarter of the third hour” (i.e., 135 minutes) which was reported by a member of the 55+ age group.

***Subtheme 1.4: Factors Affecting Comfort Levels***

Four categories emerged from the data: Time in Vehicle, System Alerts, System Performance, and Awareness of the Testing Situation. Again, comments found within these categories were not mutually exclusive as participants often mentioned several aspects of the experience in combination.

Category 1.4.1: Time in Vehicle

When discussing their comfort or trust in the system, 26 participants indicated that their confidence in the system changed over time. Change-related comments were, in several instances, tied to specifically noted underlying factors, including:

- Factors associated with the test conditions (8 responses)
    - “In the beginning, nervous. [I was] kind of uncomfortable because I wasn’t used to it – not being able to drive on your own or even use the pedal or steering wheel, but in the end I was extremely comfortable.” *(25–39 age group)*
    - “It’s hard to say [how long it took me to become comfortable with the system] because the car was new, the tablet I was using was new, so it was a lot of new

things. 15 minutes/half hour? If it was just, if it was a familiar car maybe it would have been quicker." *(55+ age group)*

- A greater understanding of the system's capabilities (9 responses)
    - "Once it got to be other cars on there. At first, there was no cars, so I was just okay with it. But once I saw what it could do with other cars, I felt more comfortable with it." *(25–39 age group)*
    - "It was pretty much right away, I mean I was a little uneasy trying to grab the wheel at first and try to still steer then I realized, 'Wow it's doing everything for me.' Once I realized it was doing that I was ready to trust it right away." *(18−24 age group)*
- Increased trust in the system's alerts (2 responses)
    - "At first I wasn't as comfortable because I didn't get the gist of the alerts when I was doing a task. And then, as time went on, I kind of understood. I got the concept…It will alert me if I need to take control or adjust or something like that. And I actually started to feel very comfortable, in the first round." *(40−54 age group)*

Category: 1.4.2: System Alerts

Thirty-six participants commented on the alerts in terms of their comfort in or trust with the system. Sixteen participants had positive comments about the types of alerts provided, including the visual (4 responses) and the haptic alerts (5 responses).

- "If you're looking away, looking down at the tablet or something, having that flash of light to direct your attention back to the road., having the seat buzz to let you know you're too close to the side, or something needs your attention, is good." *(40–54 age group)*
- "The alerts I like…especially the tactile alert. It's like hitting speed bumps on the side, you know. Already conditioned to feel that. Rmpt, Rmpt, Rmpt." (55+ age group)
- "I liked how it kind of vibrated the seat to let me know that it was not engaged anymore." (25–39 age group)

Participants also noted that the number of alerts, the lack of an alert associated with perceived system failures, and uncertainty associated with the meaning of the alerts affected their comfort levels. These alert-related concerns will be discussed in further detail in the alert reaction section. Regardless of the concerns individuals may have voiced regarding the alerts or the overall system, comfort-related alert comments indicated that most participants trusted the system to take control within the test environment.

- "It was pretty easy to figure out…I mean most of the time I found myself looking at the tablet and sometimes I forgot that I'm technically driving the car." (18–24 age group)
- "When I was e-mailing I didn't even think about driving. I thought I was more of a passenger than anything." (18–24 age group)

Category 1.4.3: System Performance

Participants (37 responses) had a variety of reactions to the system performance, both positive and negative. Contributing to participants' comfort levels was the system's ability to maintain lane position (8 responses) and regulate the vehicle's speed in response to the other vehicles (12 responses). A representative comment is the following:

- "As soon as [the test vehicle] locked in and I saw that it stayed in the center of the lane, I was just like that's fine, slow it down when you needed to, that's nice." *(18–24 age group)*

Several participants noted that the speed adjustments (4 responses) or the interaction of the test vehicle around other vehicles (10 responses) were not yet perfected. It should be noted that several of the participants who voiced concerns were the same respondents who reacted positively to the test system in the previous paragraph.

Category: 1.4.4: Awareness of the Testing Situation

Participants also noted their awareness of the test situation. In one case, this awareness led the participant to conclude that the number of alerts experienced was probably part of the test conditions.

- "I thought it was nice, it was really fun to drive. I'm going to assume you're like, you were doing the [visual alerts] and that type of stuff. So, normal driving circumstances, I think it's nice, I do." *(18–24 age group)*

Seventeen participants drew attention to the tasks they were asked to complete as part of the experimental design. The completion of secondary tasks often challenged participants' existing mental model regarding appropriate driving-related tasks.

- "Yeah, the thing that affected my comfort was basically like, I was trying to figure out, you know, how to do one thing and still keep my eyes on the road at the same time so I wasn't used to that." (25–39 age group)

- "As the day progressed, I felt better and better about it. I felt more comfortable. Truthfully, I think some of that was, I was becoming more comfortable using the pad, the tablet." (55+ age group)

- "Some sites are easier to navigate. It would worry me a little bit to have people doing that kind of stuff all the time, even though. It's kind of wrong, especially some of the nonsensical stuff. I could understand the navigation portion, trying to get an address. Although, just because you get the address still means you got to read the darn map. And you can't do that when you're driving very well." (55+ age group)

- "Doing the tasks on the tablet was so bizarre, I don't do those things, I listen to music, I listen to speech, I listen to radio, I listen to discussions forever, but I don't do tablets. Sometimes answer the phone but that's why I have my Bluetooth." (55+ age group)

Further, participants noted that the presence of in-car experimenters (1 response) and the closed track conditions may have affected their comfort with the system (10 responses).

- "[Experimenter's name redacted] wouldn't be interested in going in a vehicle that was not at least reasonably safe, So I put a fair, even just stepping in the door, I put a fair amount of trust in whatever I was going to drive today." *(25–39 age group)*

Final comments classified as general test-related comments include (one response each):

- That the same information could be obtained over a 30- to 40-minute time period.

- That participants were not permitted to wear their sunglasses.
- That the car was a nice size, appreciating that it was not a compact car.

## Theme 2: Alert Reactions

***Subtheme 2.1: General Thoughts in Response to Being Provided With Information or Alerts***

When asked about their thoughts in response to being provided with information or alerts, 25 responses included general comments about the alerts (e.g., helpful, effective, good).

Thirty participants specifically noted that the alerts prompted them to regain situational awareness either by figuring out the appropriate response to the alert or situation or by preparing to take manual control. Four participants further commented that they believe the alerts provided adequate time to react, while two noted that they thought the alerts were easy to learn or understand. For example, participants noted:

- "I liked the colors and everything. I thought it was a very easy system to learn, very intuitive." *(18–24 age group)*

Fifteen participants mentioned the haptic seat alerts in response to this question. The majority noted the benefits of and their preferences for the haptic seat (13 responses).

- "I thought that the vibrating seat was probably enough of an alert." *(55+ age group)*
- "I think the alerts are effective, the [visual alert]. I noticed that I became complacent as the test continued, but the vibration in the seat was terrific, it got me to respond right away. The audio cues got me to respond right away, but as I became more interested in what was happening on the tablet, the visual cues weren't as effective." *(40–54 age group)*

One participant noted confusion as to the intention of the haptic alert and another voiced concerns associated with vibrating seats.

- "It was pretty good, it was pretty good, you know, I don't like vibrating stuff you know most people like to sit in the vibrating chairs and relax, sit back and stuff, I can't stand that vibrating always made me alert yeah. But that could make people lazy, I don't know. I don't think that's something for everybody cause we already got lazy drivers that don't pay attention to what they doing now and you give them something like that oh boy you have to be sure it work for everything so I can understand the test I think it's a good test." *(55+ age group)*

Six participants noted concerns with the visual alerts, such as the following:

- "[The visual alert] didn't help that much if I was looking away." *(18–24 age group)*
- "I appreciated the vibration alert more than the lights, um sometimes I felt like maybe I kind of liked glazed it over after getting used to the [status light] all the time or the my frame kind of blocked it out but sometimes it was I had to keep checking to make sure everything was ok." *(18–24 age group)*

While participants noted the benefits of the alerts, 20 participants drew attention to the number of the alerts, specifically alerts perceived as false or unnecessary, and 14 discussed the lack of alerts associated with perceived system failures. In this case, participants viewed not only the surprise lane drift but also the variations in test conditions that required them to remain engaged with the system as false alerts or system failures.

- "There were some false alerts while driving. If you yawned too much then it didn't think you were looking at the road. I would prefer if you could turn away for more than 3 s." *(18−24 age group)*
- "Just had some discomfort when it didn't give me a signal when one was due." *(25–39 age group)*
- "When it did that swerve that one time without telling me it was going to do it; that bothered me. I hate when that happens, but other than that no." *(55+ age group)*

The comments seem to point to a desire for balance in the number of alerts, as illustrated by the following two responses:

- "I liked the fact that it gave you a lot of warnings and that's important. You want the car to give you a lot of warnings." *(40–54 age group)*
- "I was a little disappointed with the number of false negatives that were evident. Meaning that the system alerted me to something and I looked up and I could not see that there truly was a problem. But I believe I would rather have it that way than the opposite way, where there is a problem and I'm not alerted." *(40–54 age group)*

When looking at the prompt condition of the 20 participants who commented on the number of perceived false alerts, of those commenting:

- 10 were in the 2-second prompt condition.
- 8 were in the 7-second prompt condition.
- 2 were in the No Prompts condition.

When considering the prompt condition of the 14 participants who drew attention to perceived system failures:

- 7 were in the 2-second prompt condition.
- 3 were in the 7-second prompt condition.
- 4 were in the No Prompts condition.

As the following response illustrates, the comments associated with this question indicate a need for a balance in the number of alerts presented:

- "My immediate answer to that question would be "I felt like it wouldn't shut up." However; I suppose that's on the better biased end of the spectrum toward safety, rather than it being too quiet. The alerts were deliberate, often, and noticeable." *(18−24 age group)*

In addition, 35 participants indicated in some manner that they wanted or needed additional information as they were unsure why they were being alerted or asked to take control. The following comments are representative of the concerns voiced:

- “I thought it was, the way it alerted me was good, it like got my attention. The vibrating seat automatically you knew something was like not right so you had to look up. But I wasn’t sure why it was beeping, or like sometimes it [provide a visual cue] just from not using it I wasn’t sure if it was an alert or not, but that’s about it. The alerts were pretty good. Maybe like more specific on what was occurring, I mean there weren’t that many other cars out there so it wasn’t like it wasn’t coming into contact with other cars very often but I wasn’t really sure always what the alert was for. I mean I knew I had to grab the wheel or whatever but I wasn’t sure at to regarding what.” *(18–24 age group)*

- “I thought sometimes it was an alert for no reason you know that I could see [gave a visual alert] all of a sudden you know why did that happen there was nothing around. So it was curious. I’m curious to see how it would work on something that isn’t a continuous roadway, if we were making you know turns and curves going straight you know. I’m curious to see the kind of alerts you got in those kinds of instances.” *(55+ age group)*

The lack of understanding regarding the purpose of the alerts led several participants to voice their frustrations with the system.

- “My thoughts were irritation, because it happened so often. And other than that, just like why is it going off?” *(18−24 age group)*

- “I was just annoyed by the amount of alerts. It seemed like it would alert and I would look up and nothing was happening.” *(25–39 age group)*

- “That there were 99.9 percent false. No reality at all. They kept interfering with my task. Didn’t provide me any information because nothing was happening, the road was totally bare.” *(55+ age group)*

- “I couldn’t figure out why it was doing it, that’s what was bugging me. I know it’s a test, but in my mind, because I’m a technician, I’m going ‘why is it buzzing at me right now,’ and I didn’t know why. If it was in real life, the red might be saying that there’s a fault in the system, you’ve got to drive now because I screwed up, or the yellow means we’re getting too close to a vehicle, or we’re coming up blah blah blah, I don’t know. That bothered me, I’ll be honest.” *(55+ age group)*

When looking at the attention condition of 35 participants who expressed confusion with the meaning or intent of the alerts:

- 14 were in the 2-second prompt condition.

- 14 were in the 7-second prompt condition.

- 7 were in the No Prompts condition.

***Subtheme 2.2: First Thoughts and Actions When the System Provided a Message***

Participants were asked for their first thoughts and actions when the system provided a message. Three categories of thoughts and actions were identified:

- Category 2.2.1: Regain situational awareness,

- Category 2.2.2: Regain manual control,

- Category 2.2.3: Follow or respond to system instructions.

- Category 2.2.4: No action or ignored, and
- Category 2.2.5: No response provided.

The number of responses for each category is summarized in Figure F-3.

Of those noting that their first thought and action was to regain manual control (18 responses), the majority (17 responses) first indicated a steering-wheel-related maneuver (16 responses) while only one participant noted a braking maneuver first.

Eight participants noted that their first thoughts and actions were influenced by the alert presented. Somewhat concurringly, participants indicated the aforementioned confusion as to the alerts presented (e.g., why alerts were presented, uncertainty regarding the appropriate reactions to alerts) and six participants noted complacency with the alerts, even noting that they ignored the alerts.

- "[The number of alerts] bothered me, I'll be honest, and so I started to ignore the [Cautionary alerts]. I would see [the Cautionary alert] i the corner of my eye, it was just too many cautions. But again, I know they're testing it. I think in real life you don't want people to ignore a caution." *(55+ age group)*
- "To wait for autocorrect – for the system to handle its issue without my intervention" (*40 – 54 age group)*

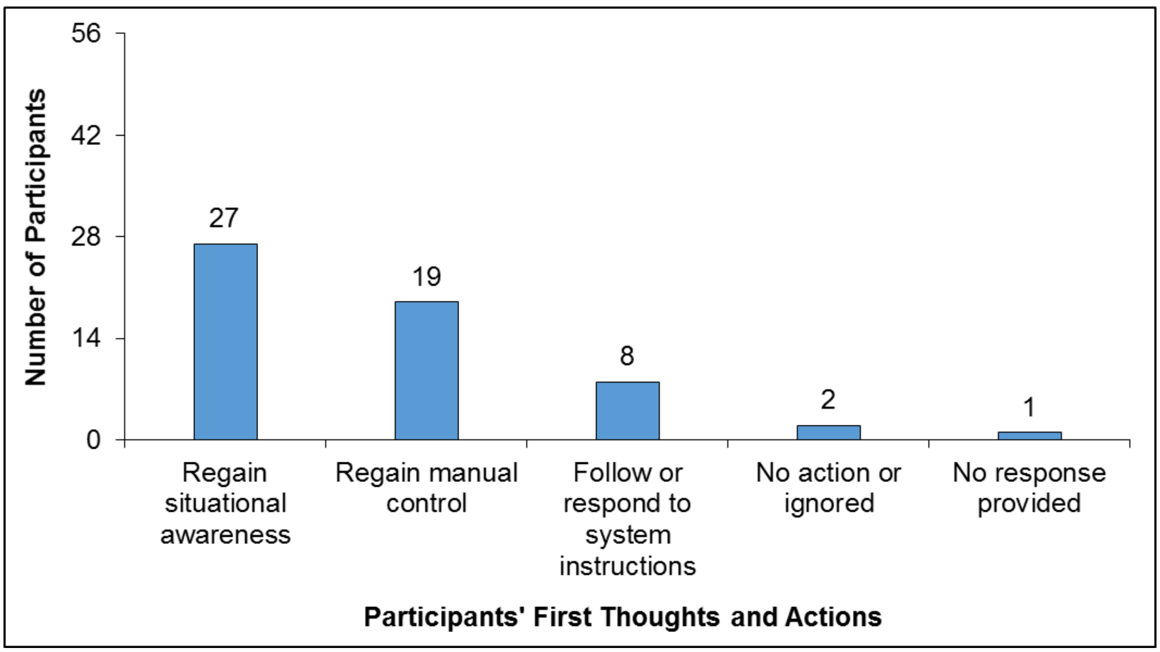



**Figure F-3. Experiment 2 Participants' First Thoughts and Actions when Provided a Message**

## Theme 3: Design Suggestions

Throughout the after-experience interview process, participants provided feedback regarding the design of the automated system. Additionally, they were specifically asked, "If you were talking to a design team, what concerns would you have about an automated system such as you experienced?" Design-related suggestions and feedback have been grouped together and analyzed as Theme 3. Four subthemes emerged from this analysis. While several concerns were voiced regarding the design-aspects of the system, it should be noted that several of the concerns voiced were likely influenced by the test condition.

***Subtheme 3.1: General Design-Related Comments***

Concerns were voiced regarding the following design-related aspects of the system:

- Overall system operation, particularly in regards to lane centering (5 responses), speed control (4 responses), and the minimization of false alerts (11 responses) and operator tasks (1 response).
- Adding features including manual override, voice commands, Bluetooth integration (1 response each).
- Need for system training, including both the system's capabilities and operating instructions (10 responses)
    - "I am sure it would come with it anyway but a nice handbook, very clear, easy to follow." *(18–24 age group)*
    - "I think the messaging, and I presume that's going to be in the manual, but most people aren't going to read the manual so understanding what those lights mean … there has to be a way to make that simple. I don't know if it's a training video, an online video, a CD that comes with it, I don't know, but to know what those alerts mean, to tell people what those alerts mean, to tell people to know what they should be looking for, doing, or potentially reacting to I think would be helpful." *(40–54 age group)*
- The potential for perceived or actual system misuse including an increase in driving distractions (13 responses).
    - "My immediate and main concern would be the overconfidence, and over-relying on the device in a common public environment, where people are too comfortable, on the device, and rely on it too heavily." *(18–24 age group)*
    - "That would be my concerns, people just getting used to it. Just like people can get used to their kid tapping them. Whatever it is that's distracting them. Being able to tune it out." *(40–54 age group)*

***Subtheme 3.2: Display Design***

In regard to the system display, five individuals suggested an additional display that alerts operators to surrounding objects (e.g., vehicles, animals, pedestrians) and one participant suggested "stronger display icons on the [activation] buttons."

***Subtheme 3.3: Driver-Vehicle Interface Design***

Several suggestions were offered regarding the driver-vehicle interface design. One participant drew upon his experience as a valet and suggested a more responsive seat similar to those with seat cushions that adjust to reflect the vehicle's direction during turns. Another participant suggested that the steering wheel include a temperature sensor which would encourage people to place their hands on the wheel. In this case the steering wheel would be warm during cold weather and cool during warm weather. Similarly, another participant suggested including a driver status monitoring sensor in the steering wheel.

Many participants voiced concerns associated with potential distractions. Thirteen participants suggested moving or adding visual displays so that they would be more conspicuous when operators were distracted or engaged in secondary tasks. One participant suggested changing the color of the indicators so that they were not as similar when shown in LED form. Nine participants suggested that moving the tablet or providing flexibility in regard to the tablet's position would allow for a better view of the visual alerts and also the forward roadway.

#### *Subtheme 3.4: Alert Design*

In addition to the previously voiced comments regarding the number of alerts, participants expressed preferences for both multimodal alerts including haptic (6 responses) and auditory alert features (7 responses) while also providing an option allowing users to select which mode of alert they prefer (1 response). Additional alert-related suggestions include:

- Adding speech alerts or enhanced alerts to inform drivers of the cause of the alert (1 response each).
- Ensuring alerts are loud enough to be heard over music, get user's attention, and get user's attention without startling (1 response each).
- Adding a lane change alert (1 response) or a driver alertness monitor to keep drivers engaged over long periods of time (5 responses).

## Theme 4: Additional Thoughts

Participants were asked if they had any additional thoughts regarding the system or the experience that they would like to share. Where appropriate, comments were grouped with their corresponding themes. For example, comments related to the system design that were not voiced in response to the system design question were included in the design discussion. Hence, the Additional Thoughts theme encompasses only those comments not previously analyzed. The remaining comments have been categorized as follows:

- Category 4.1: Positive comments about the experience (12 responses)
- Category 4.2: Positive comments about the system (10 responses)

#### *Interpretation*

Generally, participants indicated that they had a positive experience and seemed impressed by the system. The majority of participants reported feeling comfortable within the system and achieving their ultimate level of comfort within the first 30 minutes of use. Participants indicated a desire for multimodal alerts that included not only visual and haptic components but also an auditory component. Participants repeatedly questioned the number of perceived false alerts. These findings suggest that designers of automated systems will need to balance the number of alerts presented and to provide additional information as to the nature and intent of the alerts.

# Experiment 3 Analysis Findings

Following the in-vehicle experience, participants were asked the six questions listed in the Survey Instrument section in the beginning of this appendix. The interviews were audio recorded as a backup to researcher notes. Digital audio recordings were made into transcripts but no individual names were attached to any comments. Responses for 23 out of the 25 participants were recorded and transcribed. For the remaining two interviews, the researchers' written notes were consulted. The transcripts and notes served as the basis for a qualitative content analysis.

## Theme 1: Comfort with System

The first theme addressed participants' comfort with the system. Generally, information within this theme was drawn from responses to the first two questions from the after-experience interview:

1. Overall, what are your thoughts on the automated system? (Follow-up with questions on the key points the participant makes.)
2. Can you describe how comfortable you were using the automated system? (Follow-up with:
    a. How quickly (specific or relative time) the participant reached that level of comfort?
    b. Anything that they feel affected their level of comfort?)

However, as framework analysis involves grouping concepts together, responses from other questions that reflected participants' comfort with the system were included here.

#### *Subtheme 1.1: General Reactions*

Comments were categorized into trust-related reactions, performance-related reactions (e.g., ease-of-use), and general reactions. Examples of representative comments follow in Table F-6.

**Table F-6. Categories Within Subtheme 1.1 General Reactions**

| Category | Example Reaction |
|---|---|
| **1.1.1: Trust-related reactions** | • "trusting with it" *(40–54 age group)*;<br>• "had a lot of trust in it" *(55+ age group)*;<br>• "trusted more than I thought" *(25–39 age group)* |
| **1.1.2: Performance-related reactions** | • Dependable *(18–24 age group)*; maintained its speed *(25–39 age group)*<br>• "It drives better than I do." *(40–54 age group)*; "It probably drove better than I did." *(40–54 age group)*<br>• "I felt, more safety for the driver because drivers get distracted every day and I felt no matter what distractions was happening the car was still able to function as if the driver was fully operating it." *(25–39 age group)* |
| **1.1.3 General reactions** | • e.g., cool, great, interesting<br>• "Amazingly easy to use" *(40–54 age group)*<br>• "Awfully slick and pretty easy to learn" *(55+ age group)* |

***Subtheme 1.2: Comfort Level Achieved***

Participants were asked to describe how comfortable they were in using the automated system. Levels of comfort were self-defined, i.e., not provided in response to a Likert-type scale. As a result, one respondent's "fairly comfortable" may equate to another's "very comfortable." Example responses are presented in Table F-7.

**Table F-7. Example Self-reported Levels of Comfort with the Automated System for Experiment 3 Participants**

| Self-reported Level of Comfort | Example Response |
|---|---|
| **Fairly comfortable** | "I feel like I was able to trust it." *(40–54 age group)* |
| **Really comfortable** | "I was really comfortable. I looked up when I felt I needed to. The old habit of trying to be aware of what's around you while you're driving so I was aware of that, but I had trust in the system." *(18–24 age group)* |
| **Very comfortable** | "I was very comfortable. It took a few times, a few minutes to get used to it, but as you became used to the fact that it was driving and doing really well, I became extremely comfortable with it. " *(25–39 age group)* |
| **Pretty close to 100%** | "I was pretty close to 100% comfortable using it. At no time did I feel unsafe or concerned that it wasn't going to work." *(25–39 age group)* |

***Subtheme 1.3: Time to Achieve Ultimate Level of Comfort***

Participants were asked how quickly (specific or relative time) the participant reached that level of comfort. Participants responded in a number of ways, including:

- A specific amount of time (e.g., 15 minutes).
- A specific number of laps, which allowed for the approximation of time (e.g., 1 lap equates to 7 minutes).

- A specific point in the study.
    - "After the video and demonstration it was pretty much immediate." *(25–39 age group)*
    - "Maybe the second time you got me to turn it on." *(40–54 age group)*
- A temporally vague response.
    - "Fairly quickly" *(25–39 age group)*, "very quickly" *(25–39 and 40–54 age groups)*

With the exception of the three temporally vague responses, times were recorded or approximated. The majority of participants reported that they reached their ultimate level of comfort with the system by the 15-minute point (Figure 4-20). Eight participants noted that their confidence in the system improved over time. Further, several of those participants requiring the most time to achieve comfort noted that their comfort levels increased throughout the driving sessions, for example:

- **At 30 minutes:** "So at first I was a little uncomfortable, but the more I rode in it, the more it took over, I became more comfortable. So it took several times around the track." *(55+ age group – "very comfortable")*
- **At 45 minutes:** "At first, you know, with the warm-up lap and everything else, by within probably from being in the car within 45 minutes I was fairly. I was comfortable enough to the degree where I felt like I wasn't going to get much more comfortable with it. After the first probably two or three laps in the car, I would say the scales went up. But I was fairly comfortable with it, not worried. I felt like everything was going to be fine*." (18–24 age group – "really comfortable")*
- **Over 60 minutes:** "It was you know towards the end of the first one before I started becoming comfortable with it. And then the first half hour session I was kind of like, yeah, I don't know about this. But, by the third one I was pretty good." *(40–54 age group – "very comfortable")*

***Subtheme 1.4: Factors Affecting Comfort Levels***

Six categories emerged from the data: System Performance, Equating to Familiar Mental Models, Variations from Typical Driving Patterns, Uncertainty Associated with System Capabilities, and Awareness of the Test Situation.

<u>Category 1.4.1: System Performance</u>

Participants commented on the general operation of the system, specifically:

- The ability of the system to detect objects (8 responses, including 6 specific mentions of the box) and to provide timely and accurate alerts (12 responses).
- The ability of the system to maneuver the vehicle safely across the bridge (4 responses).
- The ability of the system to maintain and regulate speed in response to road conditions and design (noted in a positive way; 6 responses), including the ability to maintain lane position (5 responses).

Participants also noted that the system was user friendly, specifically, the ability to disengage the system (e.g., easy to use; 7 responses), and the system was consistent and capable (1 response).

A few participants noted concerns about the system performance. For example, three participants commented on “jerky” or “vectored” steering.

- “[After] the short, I mean, long left turn it just suddenly jerks back to normal.” *(40–54 age group)*

Several participants noted speed-related concerns. Some noted that the car was slowing too much or going too slow (4 responses).

- “[You] didn’t know if it was changing because it was a new speed limit or if it was the system.” *(25–39 age group)*
- “Just the way it slowed down, if it was a real situation someone probably would have rear-ended you…” *(25–39 age group)*

Others conveyed a desire to override the system speed and to set the speed themselves (3 responses).

- “[You] should be able to set it to the speed that you want to go instead of having the car dictate the speed.” *(55+ age group)*

While others wondered about the ability of the system to react to variations in speed limits (2 responses).

- “[O]r speed limit – like it would have to know every change in speed.” *(18–24 age group)*

<u>Category: 1.4.2: Equating to Familiar Mental Models</u>

Participants noted that the technologies being tested were similar to technologies that are currently on the market. For example, three participants responded that the system was similar to cruise control to operate:

- Used the button to engage/disengage – “muscle memory” associated with cruise control *(40–54 age group)*;
- “just turn it on, turn it off” *(18–24 age group)*;
- “automatically went to tapping the brake to disengage” *(40–54 age group)*

It was also noted that the technology would be useful in situations similar to those where cruise control is useful (e.g., highway driving [*40–54 age group*]; and non-stop-and-go conditions [*25–39 age group*]). Another participant noted “smoother braking than with ACC” *(40–54 age group).*

Two participants noted that they didn’t use cruise control in order to maintain control over the vehicle; however, they still noted comfort with the system:

- Reported that she was very comfortable with the system and obtained that level of comfort after about 15 minutes (18–24 age group).

- Reported that after a couple of times around the track (equated to 30 minutes) he was pretty comfortable with the system (55+ age group).

Category 1.4.3: Variations from Typical Driving Patterns

Four participants indicated that using the system varied in the amount of situational awareness or control compared with their current driving patterns. These comments suggest that participants were either initially hesitant with using the system or were concerned about potential effects (e.g., boredom, potential drowsiness). Regardless of the variations, all four participants noted that they were very comfortable with the system.

- "[I] don't think I've ever used cruise control and it took a little while for me to kind of stop looking up and down at the road…but eventually, it just felt really comfortable." *(18–24 age group)*

- "On the track I was very comfortable using it. I was comfortable allowing it to run its course. Again I never felt – I guess I should say I never felt completely okay to just stop looking ahead of me. I've always been taught to be aware of everything that's crossing the road. …So, again, I don't know if – I'm comfortable allowing it to run but my brain was always there and my foot was always right next to the brake, I guess. " *(24–39 age group)*

- "I think the only thing that affected is the mindset that you have when driving that there's not to be any distractions. So purposefully put your brain out of the distraction mode and into a purposed distraction mode is …you know, was probably part of that." *(40–54 age group)*

- "[I've] always been a person – I don't even really use the cruise control in my car because I like to be in control of my car. So I don't really use cruise control. So at first I was a little uncomfortable, but the more I rode in it, the more it took over, I became more comfortable." *(55+ age group)*

Category 1.4.4: Perceived System Failures

Seven participants interpreted the Imminent–No External Threat (NET) test condition as a system failure or false alarm and explicitly drew attention to the perceived failure.

- "It ran fine except for the one false alert that it gave. Yeah … Well, I only assume it's false because there was no like obvious inter-object interference with like the path that we were on. And then it kind of told me that it needed me to take over." *(18–24 age group)*

- "Sometimes it would fail out and I'd have to engage within 30 s, and other times, there was one time where it just stopped – I think it was when there was a box on the road. It just stopped." *(40–54 age group)*

- "It (the alert) did a good job, other than the false alarm." *(55+ age group)*

- "Good. There was one point where it was immediately red – I guess it was a system error – and it said I needed to take over. It was kind of unexpected, but … The other two were perfect. " *(25–39 age group)*

Category 1.4.5: Uncertainty Associated with System Capabilities

As this was a test situation, participants were purposefully not made fully aware of the system's capabilities, which resulted in uncertainty as to the system's capabilities. An example comment is as follows:

- “Well, I’m still not sure what the full capabilities of it are, but one thing would be like I said, like how small of like a thing needs to be there and how quickly can it react - like, even if, say a squirrel runs out in front of your car. And then secondly, if something like that does happen, how quickly would it stop and would it account for other stuff in the road, like a person behind, rear-end you or if like there was weather, like concerns or anything like that?” *(18–24 age group)*

For example, participants voiced concerns regarding the system’s reaction times in response to objects (i.e., how fast it can detect an object and react; 8 responses), including the ability of the system to detect people (e.g., pedestrians, small children; 2 responses), animals (e.g., wildlife; 5 responses), and the size of an object required to activate system alerts (2 responses).

- “Because there are too many variables outside the singular car that I don’t know if the car is calculating. I guess that’s… I witnessed a car accident in front of me just last week and so I don’t know how fast it can react and if you’re paying attention to something else how fast can you react to a situation that would require that avoidance, I guess.” *(25–39 age group)*

- “I just feel like it might not handle unexpected things. If like a deer ran in front of you it might not give you enough warning time or if there were things in the road that it didn’t detect or speed limit – like it would have to know every change in speed.” *(18–24 age group)*

The ability of the system to perform on different roadway types or conditions was noted. Participants questioned the ability of the system to maneuver in stop-and-go situations or where there were stop lights (3 responses) and under differing weather conditions (1 response).

- “It would have been interesting to see how it would have operated like with a vehicle passing or a vehicle or having to stop at a stop light or something like that.” *(25–39 age group)*

Additionally, participants noted potential changes in their levels of comfort with the system if the system was operating in different conditions or noted a desire to experience the system in other conditions (e.g., highways or non-test conditions; 7 responses).

- “[O]ne of the things that they should probably test for is to put a semi on the road with that because I don’t know what it is about driving right next to a semi but it makes you panic just being next to it. I don’t know how much I would trust the car right next to something that is uncontrolled and can smash you to bits fairly easily.” *(40–54 age group)*

Participants also wondered if the system would be able to detect driver state (1 response) or keep the driver engaged (2 responses).

- “The only concern that I had was that it doesn’t have any awareness of the driver, so it’s a little concerning. Someone could have a heart attack or something and the car would just stop in the middle of the highway.” *(25–39 age group)*

- “[I] think we’ll have to think about how to keep drivers engaged in some activity so that they don’t just nap. The chances of you waking up from a nap and doing the right thing are probably small.” *(18–24 age group)*

- “The only thing I can think of, and again this comes back to my boredom, is, maybe if you’re on an extended trip especially, that the system would force you to go into manual mode a few times just to break it up a little bit. Because if you don’t, your attention after you focus on whatever else you're doing and not driving. I think you need to get that driving focus back." *(40–54 age group)*

Category: 1.4.6: Awareness of the Testing Situation

Participants noted their awareness of the test situation. In three cases, this awareness led the participant to conclude that the system was set up for the experimental conditions they experienced.

- "No, I think the way it slowed down going through the curves was annoying but I think that's just for the test. It over-compensated." *(55+ age group)*
- "My thoughts, the first time I thought [the alert] was spontaneous, the second and third times I thought it was probably programmed into the test." *(40–54 age group)*
- "Programmed for the track." *(40–54 age group)*

While two others noted in some way that the test conditions, specifically having a researcher with them in the car, affected their comfort with the system.

- "I guess having someone there explaining it and having someone there who could answer questions added to the comfort level." *(25–39 age group)*
- "I mean as far as operating the system, I was very comfortable operating it, but overall usage, it was kind of like I've got two other people in the car I don't know, I'm in a car I don't know, I'm on a track I don't know. You know, it's really hard to be super comfortable doing all that." *(40–54 age group)*

## Theme 2: Alert Reactions

### *Subtheme 2.1: General Thoughts in Response to Being Provided With Information or Alerts*

Participants' comments regarding their general thoughts in response to being provided with information or alerts have been summarized in Table F-8.

**Table F-8. Categories Within Subtheme 1.1 General Reactions for Experiment 3**

| Category | Example Reaction |
|---|---|
| **2.1.1: Good, beneficial, efficient, necessary (10 responses)** | • "Good, clever. Slowed before box." *(40–54 age group)*<br>• "It was effective. It was surprising when it alerted but it was effective. I guess that's what's necessary." *(25–39 age group)* |
| **2.1.2: Timely, i.e., provided enough time for the participant to regain control (7 responses)** | • "It was always a little startling I guess, but it was supposed to be. It always gave me enough time to manually take over. But it was effective in alerting me." *(18–24 age group)* |
| **2.1.3: Got attention (10 responses)** | • "I thought it was good. Like I said it was loud enough that it got your attention but it wasn't so loud that you – that it startled you." *(25–39 age group)*<br>• "Great. Yeah. It made me jump right back into okay, I've got to do something here. So it was effective." *(40–54 age group)* |
| **2.1.4: Comforted or reassured (2 responses)** | • "It gave me a feeling of comfort, again." *(55+ age group)*<br>• "I very reassured by that and it increased my confidence." *(40–54 age group)* |
| **2.1.5: Need to regain situational awareness (i.e., figure out what to do, what was required, what was causing alert), prepare to take control (12 responses)** | • "The only time I really had a concern was when it said [REDACTED] – the tone that went off, if you're [using] digital devices you may not know that tone versus a different tone. We go some time with the car, I know that tone meant prepare to take control." *(40–54 age group)*<br>• "My thoughts, the first time I thought it was spontaneous, the second and third times I thought it was probably programmed into the test. And then, I did doubt myself because I thought I had missed the [REDACTED] level …" *(40–54 age group; note: participant did not miss the alert)*<br>• "I was prepared to put my hands on the steering wheel and take control of the vehicle." *(55+ age group)*<br>• "Just first thing to take over and then like maybe an afterthought, I guess I was just curious. I mean like you know what would actually trigger that and how sensitive is it to anything that is out on the road." *(18–24 age group)* |

***Subtheme 2.2: First Thoughts and Actions When the System Provided a Message***

Participants were asked for their first thoughts and actions when the system provided a message. In many instances, participants provided a series of thoughts and/or actions. This analysis focuses on the first thought or action identified. Three primary categories of thoughts and actions were identified:

- Category 2.2.1: Regain situational awareness,
- Category 2.2.2: Regain manual control, and
- Category 2.2.3: Follow or respond to system instructions.

The majority of participants (15) noted that upon receiving a message from the system, their first thought or action would be to regain manual control of the vehicle. Nine participants specifically noted regaining control by reaching for the steering wheel while two noted that they would perform or prepare for a braking maneuver. Four participants noted that their first thought or action was to do what the system instructed (e.g., prepare to take control, take control). The number of responses for each category is summarized in Figure F-4.

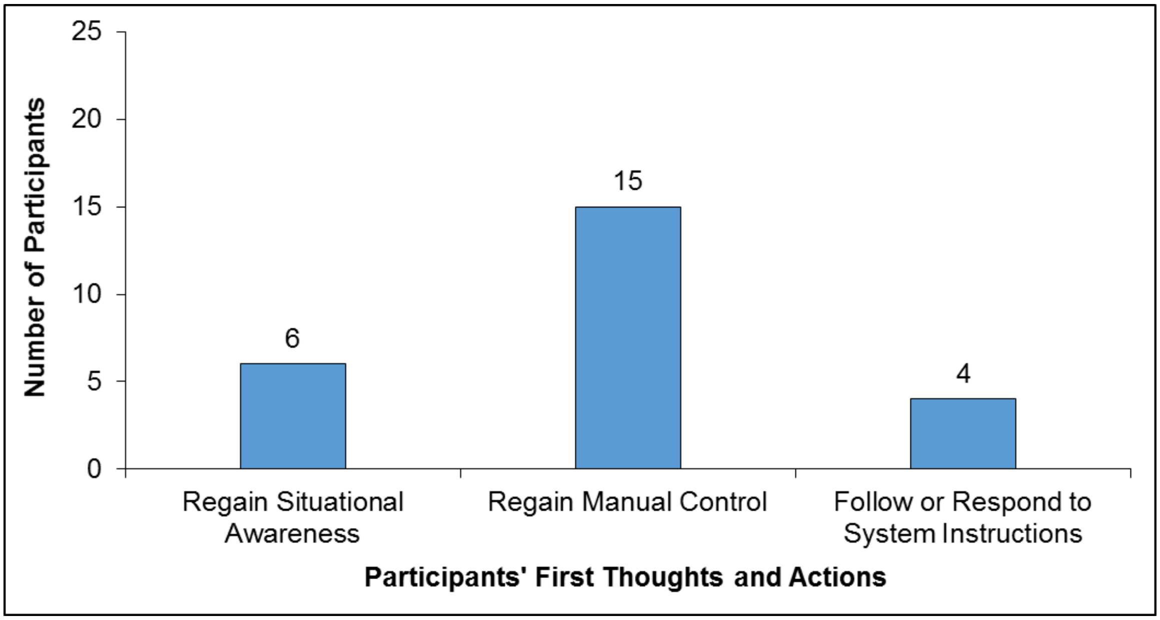



**Figure F-4. Experiment 3 Participants' First Thoughts and Actions when Provided a Message**

In looking at the full responses to this question, several (6 responses) participants noted that their thoughts and actions were influenced by the type of alert that they received, thus suggesting a conscious decision to react differently in response to a Staged or Imminent alert. For example:

- "Just depended on how severe they were. So when I had [a cautionary warning] it was a sense of urgency but it wasn't like immediately stop what I'm doing and grab the steering wheel. I could – like if I was watching a movie I could pause it and set it down safely. But when it was [Imminent] and said I immediately needed to take over it was grab the steering wheel and apply the brakes and figure out what's going on." *(25–39 age group)*
- "My instinct was to put my foot on the brake and grab the steering wheel at least with one hand. And then I would look up and read and see what it would say. The one time it – or a couple of times it was like the [Imminent warning] so I'd have to immediately take over. But that was fairly easy just with my foot on the brake. And the other times it was letting me know that I had time to take over so that was when I would use the manual button." *(18–24 age group)*

## Theme 3: Design Suggestions

Throughout the after-experience interview process, participants provided feedback regarding the design of the automated system. Additionally, they were specifically asked, "If you were talking to a design team, what concerns would you have about an automated system such as you experienced?" Design-related suggestions and feedback have been grouped together and analyzed as Theme 3. Four subthemes emerged from this analysis.

***Subtheme 3.1: General Design-Related Comments***

Category 3.1.1: Means of Engaging and Disengaging the System

Participants commented positively about the various methods available for engaging and disengaging the system (6) including the specific features of the system related to engaging and disengaging the system (e.g., the button; 7).

Category 3.1.2: General Design-Related Concerns

Concerns were voiced regarding the following design-related aspects of the system:

- Accessibility associated with system height (i.e., going through a drive-thru; *18–24 age group)*
- Car being too quiet *(40−54 age group)*
- The need for a back-up system:
    - "There's going to be flaws and as long as there's notifications. I feel like as long as it doesn't just shut off, as long as it has some kind of backup system where it can give you 5 to 10 second hey, I can't really work right now something's going down. As long as there's enough time for drivers to react I think it's fine." *(18–24 age group)*
- The need to keep it simple (*25–29 age group)* and to provide better instructions about the system capabilities (noted above as relating to comfort with the system; 8 responses).
- Ability to set speed or better regulation of speed *(40–54 and 55+ age groups).*

***Subtheme 3.2: Display Design***

In regard to the system display, the following display design-related suggestions were offered:

- Larger display screen and/or larger font *(18–24 and 55+ age groups)*
- Heads-up display *(25–39 age group)*
- A retractable display *(18–24 age group)*
    - "[I] think it would be cool to have a button or something that your auto driving screen actually goes down into the dash and then … you push a button and your screen comes up, kind of turns on the system but doesn't activate. You know, cruise control you turn it on before you activate it. Kind of the same thing… But I think something that could draw the attention of someone's eye if you're at a dealership or something you can push a button and something comes out of the dash."

### *Subtheme 3.3: Driver-Vehicle Interface Design*

Two suggestions were offered regarding the driver-vehicle interface design. One participant suggested a fixed display on the dash while a second suggested an ergonomic secondary device holder.

- "I think as far as design team in any type of accessing any type of portable device, I still don't see how that can transition into the environment. I can see how maybe using a back-up screen as a computer screen with perhaps a mouse or something, but I'm just not quite sold on the fact of you know interacting with you know other portable devices. I'd rather have the screen fixed on the front – on the dash somewhere." *(40–54 age group)*
- "I guess the comfort in using the tablet or laptop or wherever. Of course the car has its own limitations. Just not having to stare down at your lap the entire time. So kind of like a holder for your tablet or device – ergonomically correct so you're not completely looking down the whole time and at the same time if you have to take over the car you can still see what's going on and it not impeded your vision." *(25–39 age group)*

### *Subtheme 3.4: Alert Design*

With regard to the alert design, participants suggested the following additional alerts or alert types:

- An icon noting changing speeds as they happen *(25–39 age group);*
- An alert showing objects in/near the roadway so that you could decide if you wanted to avoid – e.g., "a strip of tire on the [display]" *(18–24 age group);*
- An intermittent alert to take control (to counter potential boredom; *55+ age group*);
- A haptic alert in the seat in addition to auditory and visual alerts *(25 – 39 age group).*

## Theme 4: Additional Thoughts

Participants were asked if they had any additional thoughts regarding the system or the experience that they would like to share. Where appropriate, comments were grouped with their corresponding themes. For example, comments related to the system design that were not voiced in response to the system design question were included in the design discussion. Hence, the Additional Thoughts theme encompasses only those comments not previously analyzed. The remaining comments have been categorized as follows:

- Category 4.1: Positive comments about the experience (13)
- Category 4.2: Positive comments about the system (4)
- Category 4.3: Comments noting the perceived usefulness of the technology to themselves or society (4)

### *Interpretation*

Generally, participants indicated that they had a positive experience and seemed impressed by and confident in the system. The majority of participants reported that they were very comfortable with the system and this comfort level was achieved in 15 minutes or less. Participants who indicated initial caution towards the automated system noted that by the end of the experimental test sessions they were comfortable using the system. Further, participants' level of comfort with and confidence in the system may have been improved with additional education regarding system capabilities. However, as this was a test situation, providing participants with additional education was not possible.

***Insight into Participant Behavior***

Responses obtained during this process can be used to provide insight into participant behaviors. Specifically, responses were used to provide insight as to why participants may have chosen not to perform non-driving tasks. Three participants did not participate in non-driving tasks either prior to the alert or after the alert. One participant failed to disengage the system when instructed (i.e., the system stopped the car). These participants are discussed in turn.

<u>Participant, 18 – 24 age group</u>

Participant did not participate in the following non-driving tasks:

- Non-driving task prior to the alert (2 events).
- Non-driving task after the alert (2 events).

This participant noted comfort in using the system and an interest in using and owning an automated system; however, he also noted that he was not interested in playing games or watching movies. "I was honestly, I was fairly comfortable… It gave the driver itself plenty of time to kind of notify something's going on that's in the surroundings to let them know, hey, whatever you're doing, look up, stop looking at the surroundings, stop playing the game, quit watching the movie, whatever, texting. And again, I would probably have, I'm not big into games and so if I'm not focused on the movie that's why I was…I didn't play with the tablet much but if I had my phone with me I most likely would have been texting some."

- "Again nothing off the top of my head other than as far as the tablet I wasn't big into games or anything so had I had my cell phone I would have been using it. I think it's great. I absolutely love it and I hope to see it, hope it's on the market. That's something that I, personally, would be interested in."

The participant's lack of interest in games and movies likely contributed to his choosing not to engage in non-driving tasks.

<u>Participant, 40−54 age group</u>

Participant did not participate in the following non-driving tasks:

- Non-driving task prior to the alert (2 events).
- Non-driving task after the alert (3 events).

This participant indicated that he was very comfortable with the system and that his comfort with the system was achieved quickly.

- "I was very comfortable."

- “Oh, it was about one spin around the track and I pretty much felt comfortable.”

However, this participant also indicated hesitation to engage in distraction-related tasks and to use portable devices while driving.

- “I think the only thing that affected [my comfort] is the mindset that you have when driving that there’s not to be any distractions. So purposefully put your brain out of the distraction mode and into a purposed distraction mode is … you know, was probably part of that.”

- “[The alerts were] very beneficial. I found myself even with wanting to be very on-spot alert that your mind still wanders and you’re looking at things and it’s not registering and the alert was very helpful.”

The participant’s responses suggest that he did not engage in non-driving tasks as a result of his current mental model associated with driving.

Participant, 55+ age group

Participant did not participate in the following non-driving tasks:

- Non-driving task after the alert (1 event; Session 3).

This participant noted that he trusted the system and had a desire to own an automated system.

- “I thought it was excellent. I had a lot of trust in it. Performed really well. I was impressed.”

- “I’m ready to have a car like that for myself. With as much driving as I do…”

However, he also noted that it took him a while to become comfortable with the system and that he had a desire to better understand system capabilities

- “So I don’t really use cruise control. So at first I was a little uncomfortable, but the more I rode in it, the more it took over, I became more comfortable. So it took several times around the track.”

- “Truthfully, I wouldn’t have any (design) concerns. I would just prepare the person a lot better by telling them what the vehicle can do, what it can pick up.”

Additionally, he noted that when the automated system provided him with information or alerts that he was:

- “Prepared to put my hands on the steering wheel and take control of the vehicle.”

Because this participant indicated that he was trusting of the system and wanted a system, there is no clear reason why he chose not to engage in a non-driving task after the alert. It may be that he wanted to maintain control of the vehicle, but a conclusive reason is not available.

Participant, 18–24 age group

Participant failed to disengage the system. As a result, the system stopped the car. This participant indicated trust in the system which was quickly achieved.

- “I feel like it was pretty dependable or if it was going to cut off it would tell you. I liked that it kept you right in the center of the lane. Some lane-keeping things just don’t let you go over the lines but they don’t keep you straight so you feel pretty safe.”

- “It seemed really easy to use. [Comfort was achieved after] going around the track one time.”

The participant’s trust in the system was enhanced by the alerts provided by the system.

- “I thought it was good that it knows when you should slow down and that if it’s going to shut down it warns you.”

Based on a review of the participant’s comments and the video (which clearly shows the participant tapping the brake and reaching for the wheel), the researchers concluded that the participant was unaware of having failed to disengage the system.

DOT HS 812 182
August 2015

U.S. Department
of Transportation

National Highway
Traffic Safety
Administration

NHTSA
www.nhtsa.gov

11613-081915-v6