Title
Driver Response to Variable Message Signs in a 2D Multiplayer Real-time Driving Simulator

Author
Kong, Si-Yuan

Publication Date
2018

Peer reviewed|Thesis/dissertation

University of California,
Irvine

Driver Response to Variable Message Signs in a 2D Multiplayer Real-time Driving Simulator

DISSERTATION

Submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Economics

by

Si-Yuan Kong

Dissertation Committee:
Professor Michael McBride, Chair
Professor David Brownstone
Professor John Duffy

2018

DEDICATION

To my parents and grandparents

for their boundless love and support.

And to my brother

for kindling my lifelong passion for science.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank my committee members Michael McBride, David Brownstone, and John Duffy for enabling me to achieve this significant milestone. Michael McBride has been my mentor and advisor since the summer after I completed my undergraduate at UC Irvine. He has been instrumental in my journey into, through, and beyond graduate school, and I am especially grateful for his patient and honest guidance. David Brownstone is one of the most helpful and approachable professors I have ever known. I am glad to have had the chance to work closely with him and Michael on my dissertation research project. John Duffy has greatly supported my research endeavors, and I am thankful for his advice on all matters academic and professional.

The project behind my dissertation research was conceived, proposed, and executed in close collaboration with my friend and colleague Amine Mahmassani. Together, we worked through numerous long days and nights to overcome many frustrations and successfully complete this challenging research endeavor. Among others, I would like to thank David Neumark, Jay Jayakrishnan, Jean-Daniel Saphores, Hani Mahmassani, Stephen Rassenti, and Eran Ben-Elia for their comments and feedback during various stages of this project. I acknowledge and thank our project manager Melissa Clark and other colleagues at Caltrans for their help and support. In addition, I thank our friends, colleagues, and all other individuals who helped us test and debug our experiment software.

My dissertation research was supported by the UC Irvine Center for Economics and Public Policy (Steve Borowski research award), the UC Irvine Economics department (summer and merit fellowships), the UC Irvine Experimental Social Sciences Laboratory (research merit award), the UC Transportation Center (Multiple Campus Award No. 00008817), the University of California Center on Economic Competitiveness in Transportation (Award No. 65A0529), and the US Department of Transportation Pacific Southwest Region University Transportation Center (summer and fall fellowships).

I am grateful for the help and guidance of Roger McWilliams, Gary Richardson, Guillaume Rocheteau, Donald Saari, and Daniel Bogart during my time at UCI. I also wish to thank William Branch and Cailin O'Connor for supporting me during my advancement to candidacy. I greatly appreciate the assistance of department staff and administrators, particularly John Sommerhauser, Pat Frazier, Adam Cook, Jennifer dos Santos, and Melissa Valdez.

Throughout graduate school, I have received invaluable mentorship from Michael Caldara and Alejandro Komai. I would like to thank Amine Mahmassani, Mario Silva, Debapriya Chakraborty, Karina Hermawan, Anton Leonov, and all other friends and peers from our Ph.D. cohort for helping me make it through this program. I would also like to thank Tiffany Hwu for being my chef, editor, therapist, personal trainer, research consultant, and everything else a best friend could be to help me complete this journey.

Finally, I wish to thank my mother, Wei-ping Kong, my father, Tie-jun Han, and my brother Si-ping Han for their unconditional love and support throughout all of life's highs and lows.

CURRICULUM VITAE

Si-Yuan Kong
https://sites.google.com/view/siyuank

2011	B.S. Physics, University of California-Irvine
2011	B.A. Economics, University of California-Irvine
2012 – 2017	Graduate Student Researcher, University of California-Irvine
2012 – 2018	Lab Manager, Experimental Social Science Laboratory (ESSL), University of California-Irvine
2013 – 2014, 2016 – 2017	Teaching Assistant, Economics, University of California-Irvine
2014	M.A. Economics, University of California-Irvine
2018	Ph.D. Economics, University of California-Irvine

RESEARCH INTERESTS

Behavioral economics, experimental economics, transportation economics, neural networks, machine learning, virtual worlds

PUBLICATIONS

S.Y. Kong, A. Mahmassani, D. Brownstone, and M. McBride, "An Experimental Study of Route Choice and Incident Management Using a Real-Time 2D Driving Simulator with Variable Message Signs," in *Transportation Research Board 96th Annual Meeting Compendium of Papers*, January 2017, https://trid.trb.org/View/1439693

S.Y. Kong, A. Mahmassani, D. Brownstone, and M. McBride, "Experimental Studies for Traffic Incident Management", State of California Department of Transportation, July 15, 2016, http://dot.ca.gov/research/researchreports/reports/2016/CA16-2804_FinalReport.pdf

S.Y. Kong and W. Kong, Introduction to Chinese (with companion CD), Kendall/Hunt Publishing, 1st Ed. 2007, 2nd Ed. 2008, 3rd Ed. 2013, 4th Ed. 2015.

S.Y. Kong and W. Kong, Introduction to Chinese Character Writing, Kendall/Hunt Publishing, 1st Ed. 2008, 2nd Ed. 2009, 3rd Ed. 2016.

ABSTRACT OF THE DISSERTATION

Driver Response to Variable Message Signs in a 2D Multiplayer Real-time Driving Simulator

By

Si-Yuan Kong

Doctor of Philosophy in Economics

University of California, Irvine, 2018

Professor Michael McBride, Chair

This research seeks to understand how information displayed by variable message signs (VMS) can affect driver route-choice and be better used for active traffic incident management. I study the effect of various VMS messaging strategies using a money incentivized behavioral experiment with a novel 2D real-time driving simulator that supports dozens of subjects driving on a shared virtual roadway where traffic incidents unpredictably occur. Drivers are shown a VMS display before choosing between two congestible routes. I conducted this experiment with students at the UCI Experimental Social Science Laboratory (ESSL) and with a more diverse sample of online subjects crowdsourced from the Amazon Mechanical Turk (MTurk) marketplace.

Chapter 1 will present the research motivation and methodology, the design and implementation of the experiment platform, and the results with student subjects. I find that subjects learned to efficiently operate the driving simulator, all tested VMS messaging strategies improved aggregate outcomes compared to the No VMS baseline, displaying messages didn't

cause highly volatile diversion rates, and subject gender exhibited consistent correlations with route choice.

Chapter 2 will discuss the reasons for replicating on MTurk, the methodological modifications necessary to conduct the experiment online, and how the MTurk results compare to the student results. I find that it’s viable but challenging to conduct real-time multiplayer experiments on MTurk, there are significant differences in individual characteristics between the MTurk and student subjects, and there are limited behavioral differences between the two groups.

Chapter 3 will introduce a framework using long short-term memory (LSTM) neural networks to predict driver route choice using real-time contextual data. I use varyingly limited vectors of data from my driving simulator experiments as the neural network’s input to predict driver route choice at the decision point between the two available routes. I find that the best performing model configuration can predict individual route choice with 74.0% average accuracy with in-sample cross validation and 72.2% average accuracy with out-of-sample validation.

CHAPTER 1
A Real-time 2D Driving Simulator Experiment with Variable Message Signs

1.1 Introduction

Non-recurring traffic incidents cause nearly 60% of roadway delays, prompting the need for efficient incident management (*Traffic Incident Management Handbook*, 2000). Network operators can mitigate congestion and reduce delays by diverting traffic from affected roadways onto alternate routes. One widely available tool for inducing diversions is variable message signs (VMS) – programmable electronic roadside displays that can provide travelers with timely information regarding road conditions. VMS systems in the U.S. have been used since the 1960s to direct motorists to alternate routes (Dudek, 2002), and field studies in multiple locales have confirmed their usefulness for aiding traffic incident management. (Weaver et al., 1977; Dudek et al., 1978; Dunn, Reiss, and Latoski, 1999)

However, some transportation agencies are hesitant to divert traffic for incident management because they think the risks outweigh the benefits; they fear that too many drivers will divert to alternate routes and that the ensuing congestion will undermine VMS credibility. (Dunn et al., 1999) While agencies can now determine the optimal proportion of vehicles to divert (Cragg and Demetsky, 1995), they still lack reliable methods to achieve the targeted diversion rates. Theory has shown that when many myopic travelers are presented with route-choice information, their choices may reduce road network performance in aggregate. (Mahmassani and Jayakrishnan, 1991) Case studies have demonstrated that providing traffic information using VMS does not guarantee a reduction in travel time. (Levinson and Huo, 2003)

To use VMS efficiently for managing traffic incidents, it's necessary to devise a strategy for displaying public information that will produce the desired distribution of traffic across available routes. Achieving the desired distribution is complicated by the driver's limited ability to self-coordinate. This problem has been observed in stylized route-choice games where an efficient distribution was extremely difficult to reach despite repeated trials with full information and feedback. (Iida, Akiyama, and Uchida, 1992; Selten et al., 2004) In fact, some field studies have observed unpredictable and/or non-smooth changes in diversion rates as VMS content is varied. (Chatterjee et al., 2002, Horowitz, Weisser, and Notbohm, 2003)

Although selectively provisioning information through in-vehicle systems can mitigate some of these coordination issues, such systems are not yet ubiquitous. System operators have limited control over the driver's sources of information, and third-party information providers' objectives may differ from those of the operators. Given the extant VMS infrastructure in the US and abroad, operators want to improve the effectiveness of VMS as a low-cost and readymade tool for incident management. Furthermore, studying how drivers react to public information while under the cognitive load of real-time driving will support the initiatives to build Advanced Traveler Information Systems (ATIS) that address the needs of both system operators and users who receive real-time traffic information. (Burgess, Toppen, and Harris, 2012)

I seek to understand how drivers respond to VMS information to optimize the use of VMS induced diversions for traffic incident management. To this end, I designed a money-incentivized human subject laboratory experiment to study driver response to a variety of VMS messaging schemes. I implemented my experimental platform as a networked multiplayer 2-dimensional real-time driving simulator. My platform serves as a controlled environment for observing the time-limited decision-making of drivers who possess imperfect information of the environment and influence each other’s behavior. I incentivized subjects with real monetary payments to induce a controlled value of time (VoT) preference that rewards subjects for minimizing their travel times. Within this setting, I explored how an increase or decrease in the “intensity” of VMS content, message adjustment intended to induce more or fewer drivers to divert, can produce a desired change in the diversion rate while avoiding unpredictable and/or extreme changes.

In this chapter, I will address the following key questions: 1. Do subjects learn to optimize their operation of the driving simulator to minimize their travel time for each driving scenario? 2. How effective are the messaging schemes in inducing optimal route choices and reducing travel times? 3. Does using VMS result in unpredictable or highly volatile diversion rates? 4. How do driver’s individual characteristics affect their response to and compliance with VMS? I will begin by discussing in detail the design and implementation of my experiment platform. Then, I will analyze aggregate trends such as subject demographics, learning across rounds, and average travel times. Then, I will conduct a treatment by treatment analysis of route choice optimality, predictability, and correlation with individual characteristics. Finally, I will analyze driver’s compliance for treatments with individually targeted VMS recommendations.

1.1.1 Collaborators and Funding

My research in this chapter was conducted in close collaboration with my colleague Amine Mahmassani and our advisors: Professors David Brownstone and Michael McBride. Portions of this study have been jointly published in Brownstone et al. (2016) and Kong et al. (2017). We received significant guidance and feedback from our project manager Melissa Clark and other partners at Caltrans. This study was supported by the University of California Center on Economic Competitiveness in Transportation through contract UCCONNECT-65A0529. I

received additional support from the UC Irvine Economics department, the Experimental Social Science Laboratory, and the Center for Economics and Public Policy through a generous award from Steve Borowski.

1.1.2 Related Literature

There is a substantial body of research on VMS and other real-time public traffic information systems. Previous studies have demonstrated the efficacy of information in encouraging diversions[1], identified numerous factors that influence route switching behavior[2], and confirmed the difficulty of attaining stable equilibria in route selection[3]. Among these studies, none have specifically examined the predictability of the diversion response as a function of message intensity or how the risk of over-diversion can be mitigated. The diversion rates observed and/or route choice models estimated in these studies do not reveal methods of control that operators can apply to their messages to achieve desired diversion responses over a full range of desired outcomes. These studies also estimate the effect of an alternate route's travel time savings on the probability of an individual diverting, but this knowledge is of limited use since real-world time savings are endogenous to the aggregate diversion response and cannot be known a priori.

At least two studies identify ways in which VMS content can be manipulated to produce specific aggregate changes in the diversion rate. Wardman, Bonsall, and Shires (1997) demonstrate the effects of different types of messages, while Peeta, Ramos, and Pasupathy (2000) establish a relationship between information quantity and diversion rates. However,

---

[1] Weaver et al., 1977; Dudek et al., 1978; Khattak, Schofer, and Koppelman, 1993; Horowitz et al., 2003; Levinson and Huo, 2003; Chatterjee and McDonald, 2004

[2] Allen et al., 1991; Brocken and Van der Vlist, 1991; Mahmassani and Jayakrishnan, 1991; Bonsall and Palmer, 1995; Emmerink et al., 1996; Abdel-aty, Kitamura, and Jovanis, 1997; Mahmassani and Liu, 1999; Chatterjee et al., 2002; Jou et al., 2005; Gan, 2013; Kattan et al., 2009

[3] Iida et al., 1992; Selten et al., 2004

neither study shows if and how such features can be manipulated to predictably achieve desired changes in diversion rates.

Some recent studies have used laboratory experiments in conjunction with driving simulators of varied sophistication to study the effects of VMS on drivers. Ben-Elia and Shiftan (2010) conducted a laboratory experiment to study the effect of real-time information on driver route-choice using an abstract route selection game. They show that information, experience, and risk characteristics jointly affect individual driver behavior, but their experiment does not attempt to capture group interactions or real driving dynamics. Yan and Wu (2014) use a high-fidelity 3D real driving simulator to study how subjects respond to VMS information when the layout of displayed information and physical location of the signs are varied. They demonstrate that the placement of VMS and driver's characteristics affect their response to traffic information, but they do not incorporate real incentives, traffic incidents, or group interactions in their design.

My laboratory experiment is novel in its incorporation of monetary incentives, group interactions, and realistic driving dynamics in one design. Compared to other route-choice experiments, it elicits more realistic behavior from human subjects by better simulating the context, cognitive load, and decision timing drivers face on the road. My treatments incorporate novel usage of standard and non-standard VMS verbiage not found in other studies. I also study the predictability of route choice in response to VMS – analysis that is desired by system operators yet often lacking in route choice studies.

1.2 Methodology

With my colleague Amine, I designed and implemented a 2D real-time multiplayer driving simulator that embodies several important aspects of real-world driving: 1. Vehicles move continuously and obey simplified Newtonian physics, requiring drivers to exert effort to maintain course and speed. 2. Up to 39 human participants drive together on a shared roadway to create a sense of immersive traffic and endogenous congestion. 3. The driver’s viewport into the 2D world is constrained to approximate what they can see while driving in the 3D world. 4. Drivers are incentivized to complete their trips as quickly as possible to maximize their payoff.

The simulator was written as a web app using HTML5, JavaScript, and Node.js (see Appendix B for more information). Subjects see a top-down view of the roadway where vehicles are represented as small colored squares - the driver's own vehicle is colored blue while all other vehicles are colored red. The driver's viewport constantly tracks their vehicle and allows them to see farther ahead than behind. From top to bottom, the driver's screen contains the following elements: the secondary information area that displays the current experiment round, the VMS display area, the driver's viewport, and the primary information area that displays the driver's earnings and percent completion of their itinerary in real-time.

Using the W / A / D or arrow keys, drivers control their vehicles to accelerate or change lanes left / right. All vehicles accelerate at the same rate and quickly reach the same maximum speed. If a driver stops accelerating, their vehicle will decelerate at a constant rate until it reaches the minimum speed that’s designed to prevent a driver from completely blocking their lane. While cruising, vehicles are automatically guided to stay in the center of the nearest lane. A minimum following distance is enforced between cruising vehicles to allow space for lane changes to occur. If a driver's vehicle is obstructed by another vehicle when attempting to change lanes, their vehicle will be slowed down slightly to allow them to move in behind the obstructing vehicle. Drivers are informed that there are no rewards or penalties for colliding with other objects or vehicles. In addition to human controlled vehicles, computer controlled vehicles which follow simple pre-defined control routines are used to fill in the front of the driving platoon to prevent drivers from easily knowing their starting position within the platoon.

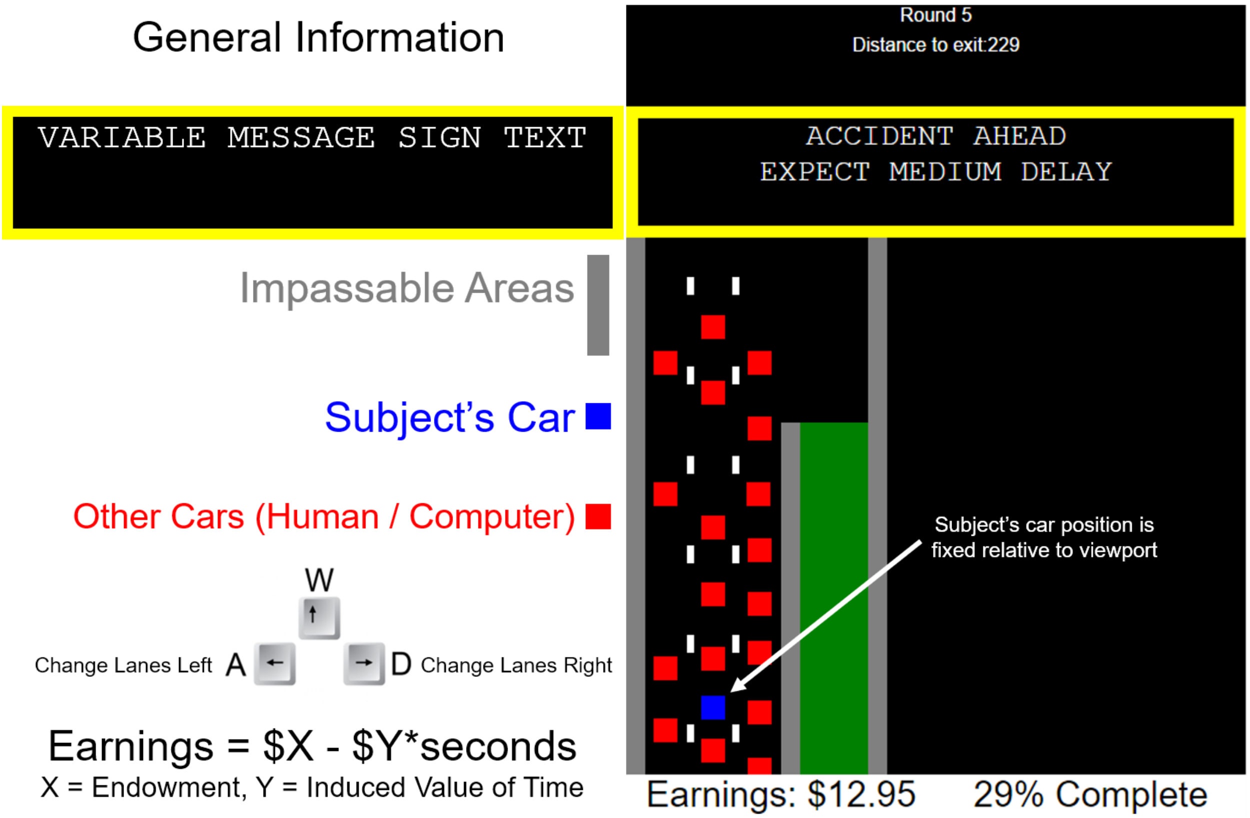


Figure 1: Driver's screen is shown on the right with feature descriptions on the left.

1.2.1 Road Network, Traffic Flow, and Incidents

The road network consists of two routes: a three-lane main freeway where traffic incidents may occur and a two-lane alternate surface street regulated by traffic signals. All vehicles start driving on the main route simultaneously at random locations on a closely spaced grid within a platoon. Their goal is to travel as quickly as possible from their starting point to a shared finish line. Soon after they leave the start, drivers encounter the VMS region where traffic information may be shown for approximately 7.5 seconds. Then, drivers encounter the exit to the alternate route where they can divert. After passing this decision-point, drivers cannot observe traffic on the route they did not choose. My choice of exit location is motivated by the finding of Dong and Mahmassani (2009) that a flow breakdown can be very difficult, if not impossible to reverse. Therefore, subjects must make their decision before evidence of a breakdown is apparent, conditioning only upon their prior experience, the actions of other visible vehicles, and any VMS information presented.

Shortly after the decision point, drivers on the main route pass through the incident area where a traffic incident may occur (see Table A2 in Appendix A for the full sequence of incidents used). The possible incident severity levels are:

0. **No incident** – roadway is clear
1. **Minor incident** – one lane is blocked, an automated "traffic cop" directs vehicles through the incident area at reduced speed
2. **Medium incident** – two lanes are blocked, vehicles queue into a three-to-one merge to pass through the incident area
3. **Major incident** – three lanes are blocked, vehicles queue into a three-to-one merge and wait for one lane to open before passing through the incident area. Cars are slowed while passing through the single open lane for a period of up to 20 seconds.
4. **Severe incident** – three lanes are blocked, vehicles must wait longer than in severity level 3 before one lane opens. Cars are slowed in the single open lane for up to 80 seconds.

When there is no incident, it is optimal for all drivers to stay on the main route, but when an incident occurs, system performance is maximized when an optimal proportion of traffic diverts to the alternate route. Drivers who divert to the alternate route will pass through two traffic signals before reaching the finish line. Each signal forms an unpassable barrier when red that prevents drivers from running the light. I conducted simulations using computer-controlled agents to determine the optimal proportion of drivers who should divert to the alternate route for each incident scenario. See Figure A1 in Appendix A for a visual overview of the road network and incident geometries.

1.3 Experiment Design

I conducted a series of experiment sessions which typically lasted for 1 hour and involved up to 39 participating subjects. My subjects were randomly recruited from UCI students registered in the Experimental Social Science Laboratory (ESSL) subject pool (IRB approval HS #2011-8378). For each session, I attempted to recruit fresh subjects with no prior experience from this experiment. However, the pool of fresh subjects was exhausted near the end of the academic year, and two sessions had a significant number of repeat subjects who had participated in previous sessions. My results analysis indicates that experience from prior sessions did not significantly affect subject performance or decision making.

Subjects receive a detailed instructional presentation at the start of each session that informs them of the simulator's controls, the layout of the road, the available routes and types of congestion that may occur on each, and how they earn money. Sessions were comprised of two parts:

Part I featured a risk elicitation task in which subjects choose between three options: receive $3.50 with certainty, receive $2.90 or $4.20 with equal chance, or receive $1.90 or $5.00 with equal chance. These options are increasing in the spread of outcomes and slightly increasing in expected value. Based on their choice, subjects are classified as risk averse, risk neutral, or risk seeking. My three-choice task is a simplified version of a well-known five choice design. (Eckel and Grossman, 2008)

Part II featured the driving task comprised of a series of 23 driving rounds – three guided practice rounds at the beginning followed by 20 normal rounds. At the beginning of each driving round, subjects started with a $14.00 endowment that decreased at $0.15 per second until they crossed the finish line. After all subjects either crossed the finish line or ran out of money, the next driving round would begin after a short delay. A single traffic incident may occur each round, and the same pre-randomized sequence of incidents was used for each experiment session to make order and learning effects comparable (see Table A2 in Appendix A for the full sequence of incidents used).

After completing Part I and Part II, subjects answer a post-experiment questionnaire regarding demographic information, route choice strategies, and other feedback. Subjects were paid the sum of their show-up payment, the realization of the lottery they selected in Part I, and the average of what they earned from three randomly chosen non-practice rounds in Part II. Averaging across three rounds helped mitigate the effect of randomized starting positions on potential earnings. Across all subjects, the average payment for Part I was $3.54 and the average for Part II was $7.65. Combined with the $7.00 show-up fee that all subjects received, the total average earnings per subject per session was $18.19.

My experiment treatments were designed to test the standard messaging content approved for use by California's system operators as well as unconventional content designed to improve the optimality and predictability of the diversion response. Standard messaging content was crafted following the *Changeable Message Sign (CMS) Guidelines* document published by California's Department of Transportation. (Wooster and Al-Khalili, 2013) Each treatment features a coherent messaging scheme that displayed information on VMS according to traffic and/or incident conditions. For static messaging schemes, a single message was displayed to all drivers within one round of driving according to the incident severity level, while dynamic messaging schemes varied the displayed message in real-time according to the diversion response. For a complete list of the individual VMS messages shown for each treatment condition and incident scenario, see Table A1 in Appendix A.

I will present detailed analysis on the messaging treatments listed below. The No VMS treatment serves as a control for the worst-case scenario where drivers receive no traffic information. Treatments with VMS are expected to have differing effects on drivers' travel times as well as the aggregate diversion response, but all are expected to improve driver outcomes relative to the No VMS baseline. On the aggregate level, the optimal aggregate diversion response is achieved when the combined travel time among all drivers is minimized for any given incident severity level. In other words, enough drivers will divert to the alternate route such that travel times will be equalized for the last drivers to finish on either route.

Treatment 1: No VMS baseline: A control treatment where no traffic information is ever displayed.

Hypothesis: Travel times will be highest in this treatment. Drivers will settle on a mixed route choice strategy that does not condition upon the incident severity level, and the aggregate diversion response will not change much from round to round.

Treatment 2: Qualitative description of incident severity – A treatment that displays a qualitative description of incident severity using Caltrans approved verbiage. (e.g. "ACCIDENT AHEAD, EXPECT MINOR DELAY" for incident severity 1). This serves as a benchmark for the efficacy of messaging strategies currently in use.

Hypothesis: Drivers will learn to condition their route choice on the variable intensity traffic messages, travel times will be shorter than the No VMS case, and the aggregate diversion response will be more optimal.

Treatment 3: Qualitative description with guidance – Same as treatment 2, but with supplemental recommendations to use the main route when there is no traffic incident (e.g. "ROAD CLEAR, ALL CARS: USE MAIN ROUTE") and to use the alternate route for severe incidents (e.g. "USE ALT RTE AHEAD"). This treatment tests whether adding a guidance recommendation can increase the diversion response.

Hypothesis: Compared to treatment 2, showing the guidance recommendation will increase the diversion rate when there's an incident and decrease the diversion rate when there isn't one. As a result, the aggregate diversion response will be more optimal.

Treatment 4: Dynamic diversion rate – In addition to a static qualitative description of incident severity, drivers are shown the current optimal rate at which they should divert to the alternate route (e.g. "1 IN 10 CARS SHOULD EXIT"). This rate is updated in real-time according to the usage of the two routes to *nudge* the diversion response towards the optimal target.

Hypothesis: The aggregate diversion response in this treatment will be more optimal than in treatment 2, but there will be more volatility.

Treatment 5: Numeric IDs – In addition to a qualitative description of incident severity, each vehicle is assigned a publicly visible numeric ID between 1 to 39. The VMS message instructs vehicles within a range of IDs to use the alternate route (e.g. "IF YOUR CAR IS #1-4, USE ALT ROUTE" for incident severity 1). The idea is to use a public message to induce individual drivers to divert based on a unique characteristic such as their vehicle's license plate number or their date of birth. Given a known distribution of characteristics, system operators would be able to target subsets of drivers in traffic.

Hypothesis: The aggregate diversion response in this treatment will be more optimal and less volatile than in non-targeted treatments.

Treatment 6: Color outlines – In addition to a qualitative description of incident severity, a subset of vehicles is outlined with a bright green border each round based on the optimal number who should divert. The VMS message instructs outlined vehicles to use the alternate route. Subjects are instructed that the best possible traffic outcome is achieved if all drivers follow the recommendations. This serves as a benchmark of driver compliance with targeted recommendations.

Hypothesis: The aggregate diversion response in this treatment will be the most optimal and least volatile among all tested treatments.

1.4 Results and Discussion

1.4.1 Subject Characteristics

Table 1: Summary of experiment sessions.

Treatment	Subjects	Repeat Subjects	Avg. Age	M	F	Licensed in USA	Avg. Weekly Hours Driven	Seen VMS
1 / No VMS baseline	30	0	20.1	14	16	88%	7.9	97%
2 / Qualitative description	39	0	19.9	9	30	77%	8.0	87%
3 / Qualitative with guidance	38	18	20.4	22	16	84%	7.6	84%
4 / Dynamic diversion rate	35	0	20.6	12	23	77%	8.0	86%
5 / Numeric IDs	37	0	19.9	14	23	73%	7.0	86%
6 / Color outlines	39	25	20.9	16	23	72%	8.0	85%

A summary of the treatments conducted is shown in Table 1. Subjects recruited from the ESSL subject pool were 20.3 years old on average. Over 80% held a valid US driving license and over 88% reported having seen VMS before. Most subjects reported that their typical driving locale is in Southern California and that they received real-time traffic information while driving. There were significantly more female than male subjects in each experiment session. In comparison, drivers in the US were nearly evenly split between males and females, and males are known to drive more miles on average within every age group. (*Highway Statistics 2016*, 2016; Sivak, 2015)

For my risk elicitation task, most subjects in each session preferred the risk averse or risk neutral options. Compared to Eckel and Grossman (2008), there was a similar risk distribution pattern with most subjects preferring the middle risk option and a greater percentage of males preferring the riskiest option than females (see Figure 2). A chi-squared test shows no statistically significant difference between male and female risk preference distributions (p-value < 0.255). While simpler to implement, my three-option task might not provide enough differentiation in riskiness to observe a statistically significant difference between male and female risk preferences.

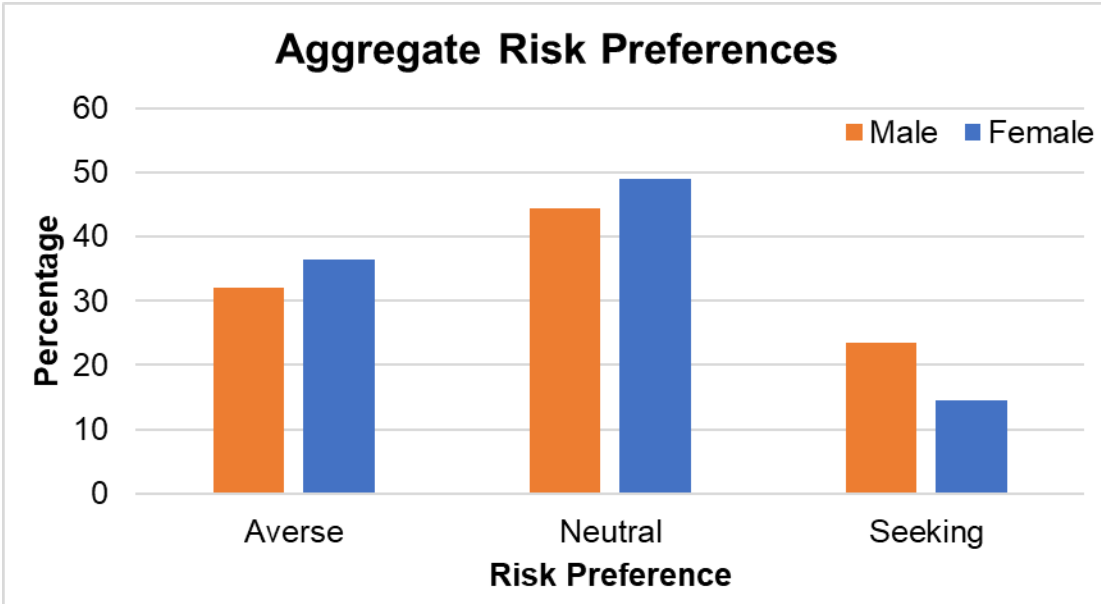


Figure 2: Elicited risk preferences among all male and female subjects.

1.4.2 Travel Times

Travel time was measured as the number of seconds it takes a driver to move from their starting point to the finish line in an experiment round. To compare average travel times between treatments, subjects' average travel times per round were regressed on VMS treatment dummies alone and on VMS treatment dummies with other treatment parameters (shown in Table 2). For the reduced regression, the intercept estimate is the average travel time in seconds for the No VMS baseline treatment, while the treatment / variable estimates are the average number of seconds saved compared to the baseline treatment. As hypothesized, all treatments with VMS improved average travel time over the No VMS treatment. Treatment 6 with color outlines resulted in the lowest average travel times. Based on the full regression, subjects who started driving in the rightmost lane had slightly longer travel times than subjects who started in the other two lanes, and I will show that this is likely due to a greater likelihood of diverting to the alternate route in all driving scenarios. The average effect of being a repeat subject with participation experience from a prior session was insignificant. Starting advantage ranged from 0 for subjects who start driving at the rear of the platoon to 1 for subjects who start at the front. The average time advantage of starting at the very front of the platoon was about 16 seconds.

Table 2: Linear regression of subject travel time per round.

Treatment / Variable	**Estimate**	**tStat**	**Estimate**	**tStat**
1 / No VMS (intercept)	**45.52**	**92.43**	**39.97**	**106.31**
4 / Dynamic diversion rate	**-2.59**	**-3.86**	**-2.68**	**-7.74**
2 / Qualitative description	**-2.70**	**-4.12**	**-2.81**	**-8.32**
5 / Numeric IDs	**-2.71**	**-4.09**	**-2.77**	**-8.11**
3 / Qualitative w/ guidance	**-3.05**	**-4.63**	**-3.15**	**-8.45**
6 / Color outlines	**-3.08**	**-4.70**	**-3.16**	**-7.97**
Started in middle lane			**0.98**	**4.22**
Started in right lane			**1.22**	**5.30**
Incident Severity 1			**7.70**	**25.83**
Incident Severity 2			**10.75**	**36.05**
Incident Severity 3			**18.66**	**62.62**
Incident Severity 4			**26.66**	**89.43**
Repeat subject			-0.04	-0.14
Starting advantage			**-15.88**	**-50.19**
F-stat vs constant model	5.90		933.00	
p-value	1.92E-05		0.00E+00	

1.4.3 Simulator Experience

As subjects gained experience operating the simulator, they should've reduced the number of extraneous inputs they make to focus on minimizing their travel time. This tendency can be confirmed by examining the trend in the average keypresses and average travel time per subject over the course of the experiment for rounds with no traffic incident (see Figure 3 and Figure 4). In these rounds, the optimal driving strategy was to continuously drive forward without changing lanes as soon as subjects realize there will be no incident based on what's displayed by the VMS. With experience, subjects learned to reduce both keypresses and travel times for no incident rounds. This indicates that subjects learned to grasp how their actions affect their travel time in the simulator and that they followed the incentive to minimize their travel time.

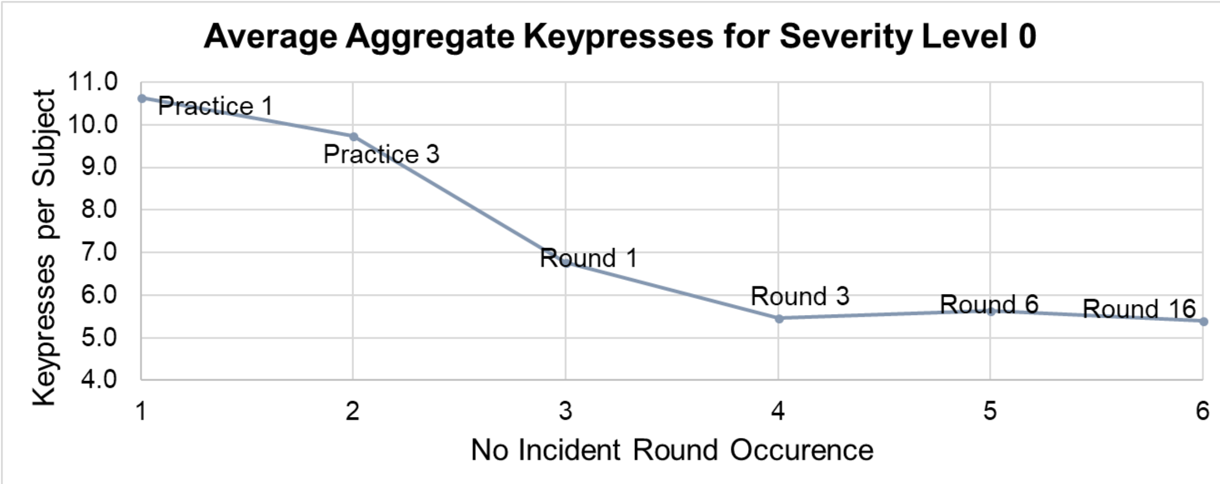


Figure 3: Average aggregate keypresses for no incident rounds.

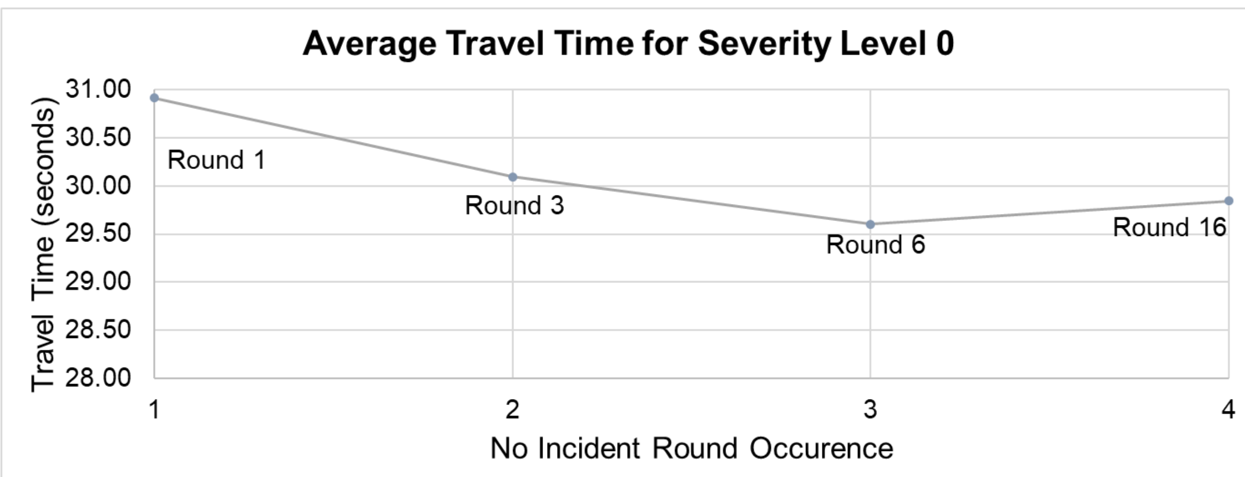


Figure 4: Average travel time for no incident rounds.

1.4.4 Treatment Analysis

The following sections present specific results from the six treatment conditions tested – one baseline No VMS treatment and five treatments with VMS. In each section, I will describe the observed behavior of and feedback received from subjects, compare the route choice of subjects to an optimal benchmark, analyze the stability of the diversion response, and use logit regressions to estimate the effect of individual characteristics on the subject's route choice.

My optimal route choice benchmark is obtained by running the experiment scenario with all computer-controlled cars to simulate different route usage proportions for each level of traffic incident severity. The optimal diversion response is determined by iterating over the number of cars that divert to the alternate route until the aggregate travel time for all cars is minimized for a given incident severity level. The amount by which subjects over or under utilized the alternate route is quantified by taking the difference between subject and optimal alternate route usage proportions to obtain a metric for "mis-diversion". Then, the overall root mean square deviation (RMSD) of route choice from optimality across all rounds is calculated from the mis-diversion per round. I calculated RMSD across all rounds and for the last three rounds (RMSD3) of each incident severity level to represent route choice performance after one round of learning for each incident type. These results are listed in Table 3. Note that RMSD3 is always lower than RMSD, indicating subjects are learning towards optimal route usage.

I examined route choice stability by calculating the change in alternate route usage between rounds of the same incident severity. Then, I quantified the overall stability by calculating the root mean square variation (RMSV) across all occurrences of an incident severity level and for the last 3 occurrences (RMSV3). These results are listed in

Table 4. Note that RMSV3 is not always lower than RMSV.

Since route choice is binary in my experiment, I used logit regressions to estimate the effects of elicited and controlled individual characteristics such as risk preference or starting lane on individual route choice. For treatments 5 and 6 with individually targeted route guidance, I used logit regressions to estimate the effects of individual characteristics on subjects' compliance with VMS recommendations. I provide tables with the estimated model coefficients and t statistics (coefficients significant at the 5% level are bolded).

Table 3: Root mean square deviation across all and last 3 rounds of each incident severity.

Root Mean Square Deviation (RMSD)		
Treatment	**All**	**Last 3**
1 / No VMS baseline	0.299	0.288
2 / Qualitative description	0.141	0.121
3 / Qualitative with guidance	0.134	0.118
4 / Dynamic diversion rate	0.172	0.146
5 / Numeric IDs	0.138	0.135
6 / Color outlines	0.099	0.098

Table 4: Root mean square variation between all and last 3 rounds of each incident severity.

Root Mean Square Variation (RMSV)		
Treatment	**All**	**Last 3**
1 / No VMS baseline	0.085	0.085
2 / Qualitative description	0.065	0.058
3 / Qualitative with guidance	0.071	0.077
4 / Dynamic diversion rate	0.104	0.096
5 / Numeric IDs	0.108	0.104
6 / Color outlines	0.058	0.057

1.4.5 Treatment 1: No VMS baseline

In this treatment, subjects received no traffic information and had no knowledge of road conditions upstream of the exit to the alternate route. As seen in Figure 5, subjects produced a flat and unreactive diversion response uncorrelated with each round's incident severity level. Based on the logit model estimates given in Table 5, subjects who preferred the riskiest lottery option were also significantly more likely to use the alternate route.

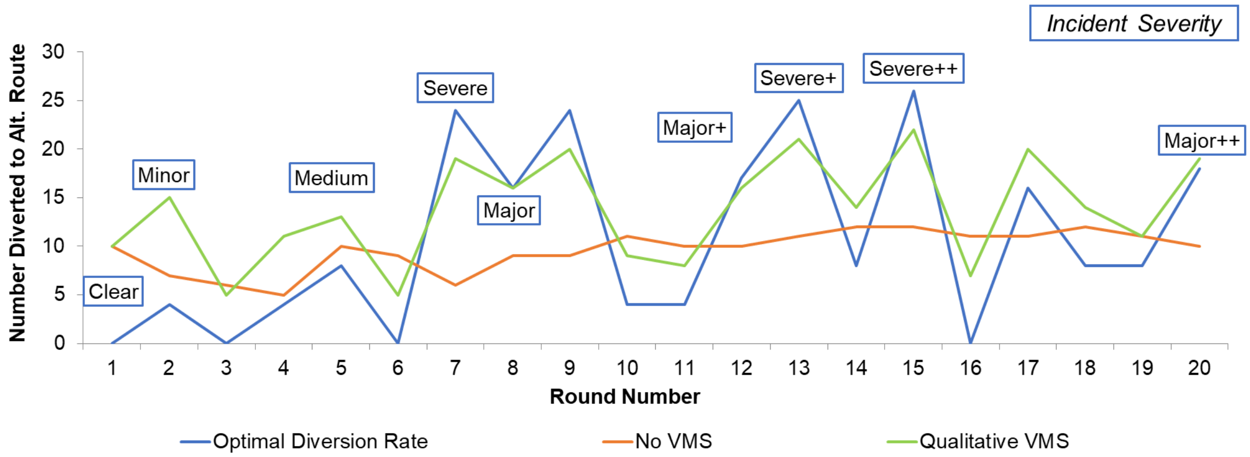


Figure 5: Diversion response graph.

Table 5: Logit regression of route choice, treatment 1.

Treatment 1: No VMS baseline			
Variable	**Meaning**	**Value**	**tStat**
(Intercept)		**-0.892**	**-3.954**
risk = 1	Risk Neutral	-0.204	-1.011
risk = 2	Risk Loving	**0.718**	**2.880**
lane = 1	Start middle lane	0.014	0.061
lane = 2	Start in right lane	0.285	1.323
gender = female		-0.026	-0.145

1.4.6 Treatment 2: Qualitative description of incident severity

In this treatment, subjects were given traffic information with a verbal description of traffic incident severity on the main route (e.g. "ACCIDENT AHEAD, EXPECT MINOR DELAY"). As shown in Figure 5, the addition of incident severity information significantly changes the diversion response. Subjects now conditioned their route choice on the traffic information displayed. They could better learn and coordinate upon the optimal diversion response for each incident severity level as they gain experience across rounds.

Mis-diversion for each incident level tended to decrease over time. Compared to the No VMS baseline, the overall diversion response was both more optimal with lower RMSD and more stable with lower RMSV. As shown in Table 6, each increasing level of VMS intensity had statistically significant and increasing effects on the probability of choosing the alternate route. Starting in the right lane increased the chance of diverting while being risk loving reduced the chance, but neither effect was statistically significant.

Table 6: Logit regression of route choice, treatment 2.

Treatment 2: Qualitative description			
Variable	**Meaning**	**Value**	**tStat**
(Intercept)		**-1.529**	**-4.753**
risk = 1	Risk Neutral	-0.026	-0.146
risk = 2	Risk Loving	-0.476	-1.932
lane = 1	Middle Lane	0.040	0.208
lane = 2	Right Lane	0.371	1.946
gender = female		-0.120	-0.616
VMS = 1	VMS Intensity 1, "Minor"	**0.604**	**2.169**
VMS = 2	VMS Intensity 2, "Medium"	**0.879**	**3.222**
VMS = 3	VMS Intensity 3, "Major"	**1.398**	**5.231**
VMS = 4	VMS Intensity 4, "Severe"	**1.685**	**6.308**

1.4.7 Treatment 3: Qualitative description with guidance

In this treatment, subjects were shown the same qualitative description of incident severity as treatment 2 with supplemental recommendations to use the main route when there is no traffic incident (e.g. “ROAD CLEAR, ALL CARS: USE MAIN ROUTE”) and to use the alternate route for some major and severe incidents (e.g. “USE ALT RTE AHEAD”). Compared to treatment 2, adding the additional guidance recommendations reduced average travel time. Diversion response optimality improved with lower RMSD, while stability was slightly worse with higher RMSV. Table 7 shows the logit regression of route choice in this treatment unpooled and pooled with data from treatment 2. The unpooled results shows a positive but statistically insignificant effect on the probability of diversion for the “Alt. rte. available" message and essentially no significant effect for the “Use alt. rte.” message. The pooled results support a large negative but statistically insignificant effect for the “Use main route” message and uphold the effects from the unpooled regression for the other two messages. Used together, the guidance recommendations in this treatment effectively improved system performance over treatment 2.

Table 7: Logit regression of route choice, treatment 3 unpooled and pooled with treatment 2.

Treatment 3: Qualitative description w/ guidance					
		Unpooled		Pooled	
Variable	**Meaning**	**Value**	**tStat**	**Value**	**tStat**
(Intercept)		**-2.411**	**-7.020**	**-1.652**	**-6.223**
risk = 1	Risk Neutral	0.148	0.751	0.061	0.474
risk = 2	Risk Loving	-0.192	-0.783	-0.326	-1.900
lane = 1	Middle Lane	-0.023	-0.110	0.007	0.047
lane = 2	Right Lane	**0.470**	**2.355**	**0.416**	**3.021**
gender = female		-0.023	-0.133	-0.056	-0.450
VMS = 1	VMS Intensity 1, "Minor"	**1.224**	**3.718**	**0.587**	**2.361**
VMS = 2	VMS Intensity 2, "Medium"	**1.452**	**4.473**	**0.839**	**3.415**
VMS = 3	VMS Intensity 3, "Major"	**2.470**	**7.264**	**1.569**	**6.370**
VMS = 4	VMS Intensity 4, "Severe"	**2.327**	**6.857**	**1.646**	**6.682**
NULL = 1	Guidance 0, "Use main route"			-0.655	-1.891
DIV = 1	Guidance 1, "Alt. rte. available'	0.334	1.153	0.464	1.812
DIV = 2	Guidance 2, "Use alt. rte."	-0.164	-0.578	-0.032	-0.126

1.4.8 Treatment 4: Dynamic diversion rate

In this treatment, subjects were shown the same qualitative description of incident severity as treatment 2 with the addition of the current optimal rate at which they should divert to the alternate route (e.g. “1 IN 10 CARS SHOULD EXIT”) updated in real-time according to route usage. This treatment provided the lowest improvement in average travel time over the No VMS baseline among all treatments tested. Route choice optimality and stability were both notably worse than treatment 2. The logit regression of route choice in Table 8 shows significant positive effects on diversion for subjects who started in the right lane and/or are risk loving.

It was a challenge to design an effective way of presenting real-time public information that could improve system outcomes. Several refinements of the messaging scheme used in this treatment were tested, and most performed worse on average compared to using the standard qualitative description. Displaying the diversion rate doesn't seem to help subjects coordinate on achieving it. Most subjects reported focusing on the qualitative description component of the VMS messages. As implemented, displaying the dynamic diversion rate as supplemental information did not improve system performance over treatment 2.

Table 8: Logit regression of route choice, treatment 4.

Treatment 4: Dynamic diversion rate			
Variable	**Meaning**	**Value**	**tStat**
(Intercept)		**-1.760**	**-5.776**
risk = 1	Risk Neutral	0.252	1.254
risk = 2	Risk Loving	**0.865**	**3.201**
lane = 1	Middle Lane	0.000	0.000
lane = 2	Right Lane	**0.463**	**2.300**
gender = female		-0.280	-1.601
VMS = 1	VMS Intensity 1, "Minor"	**0.467**	**1.594**
VMS = 2	VMS Intensity 2, "Medium"	**0.974**	**3.444**
VMS = 3	VMS Intensity 3, "Major"	**1.372**	**4.908**
VMS = 4	VMS Intensity 4, "Severe"	**1.457**	**5.220**

1.4.9 Treatment 5: Numeric IDs

In this treatment, each subject's vehicle was assigned a numeric ID, and the VMS message instructs vehicles within a range of IDs to use the alternate route (e.g. "IF YOUR CAR IS #1-4, USE ALT ROUTE" for incident severity 1). Subjects were still shown a qualitative description of incident severity as in treatment 2. Adding the targeted numeric ID based messaging did not significantly reduce average travel time or improve the optimality of the diversion response. On the other hand, route choice stability was the lowest among all treatments tested. The logit regression of route choice in Table 9 shows that targeted subjects were significantly more likely to divert to the alternate route than non-targeted subjects, and this effect dominated the effect of the qualitative description for the Minor and Medium severity levels.

Female subjects were much less likely to divert, but this effect may be confounded by their increased tendency to comply with targeted VMS that I will discuss in section 1.4.11 below.

Table 9: Logit regression of route choice, treatment 5.

Treatment 5: Numeric IDs			
Variable	**Meaning**	**Value**	**tStat**
(Intercept)		**-1.164**	**-3.741**
risk = 1	Risk Neutral	-0.211	-1.129
risk = 2	Risk Loving	-0.556	-1.572
lane = 1	Middle Lane	**0.552**	**2.633**
lane = 2	Right Lane	0.217	1.017
gender = female		**-0.898**	**-4.629**
VMS = 1	VMS Intensity 1, "Minor"	0.370	1.249
VMS = 2	VMS Intensity 2, "Medium"	0.302	0.999
VMS = 3	VMS Intensity 3, "Major"	**0.885**	**2.928**
VMS = 4	VMS Intensity 4, "Severe"	**0.942**	**2.935**
target = 1	Targeted to use alt. rte.	**1.193**	**5.795**

1.4.10 Treatment 6: Color outlines

In this treatment, a subset of vehicles is outlined with a bright green border each round based on the optimal number who should divert, and the VMS message instructs outlined vehicles to use the alternate route. Subjects were still shown a qualitative description of incident severity as in treatment 2. This treatment provided subjects with the clearest signal and easiest way to coordinate on the optimal diversion response. Overall, it performed the best with the shortest average travel time, most optimal diversion response, and highest route choice stability among all treatments tested. The logit regression of route choice in Table 10 shows that targeted subjects were significantly more likely to divert to the alternate route, and this effect was much stronger than the other effects considered. Female subjects were less likely to divert, but this effect is again confounded by their increased tendency to comply with targeted VMS.

Table 10: Logit regression of route choice, treatment 6.

Treatment 6: Color outlines			
Variable	**Meaning**	**Value**	**tStat**
(Intercept)		**-2.321**	**-6.404**
risk = 1	Risk Neutral	-0.160	-0.674
risk = 2	Risk Loving	0.339	1.412
lane = 1	Middle Lane	-0.043	-0.157
lane = 2	Right Lane	**0.818**	**2.549**
gender = female		**-0.557**	**-2.753**
VMS = 1	VMS Intensity 1, "Minor"	0.598	1.679
VMS = 2	VMS Intensity 2, "Medium"	0.427	1.126
VMS = 3	VMS Intensity 3, "Major"	**1.474**	**3.914**
VMS = 4	VMS Intensity 4, "Severe"	**0.955**	**2.278**
target = 1	Targeted to use alt. rte.	**2.063**	**6.758**

1.4.11 Compliance with Targeted Guidance

For treatments 5 (numeric IDs) and 6 (color outlines), compliance is defined as the subject choosing to divert only when instructed to do so by VMS. Overall, compliance rates were higher using color outlines than numeric IDs. This could be attributed to the ease with which outlined subjects could identify themselves as being targeted for the alternate route, whereas it was harder for numeric ID subjects to distinguish. Figure 6 depicts the difference between male and female compliance levels for both treatments combined. Female subjects tended to comply more frequently with targeted guidance, but the gap diminishes for Severe incidents. This trend is supported by the logit regression of compliance on individual characteristics in Table 11. Again, the probability of compliance decreased as incident severity increased, and female subjects were more likely to be compliant overall.

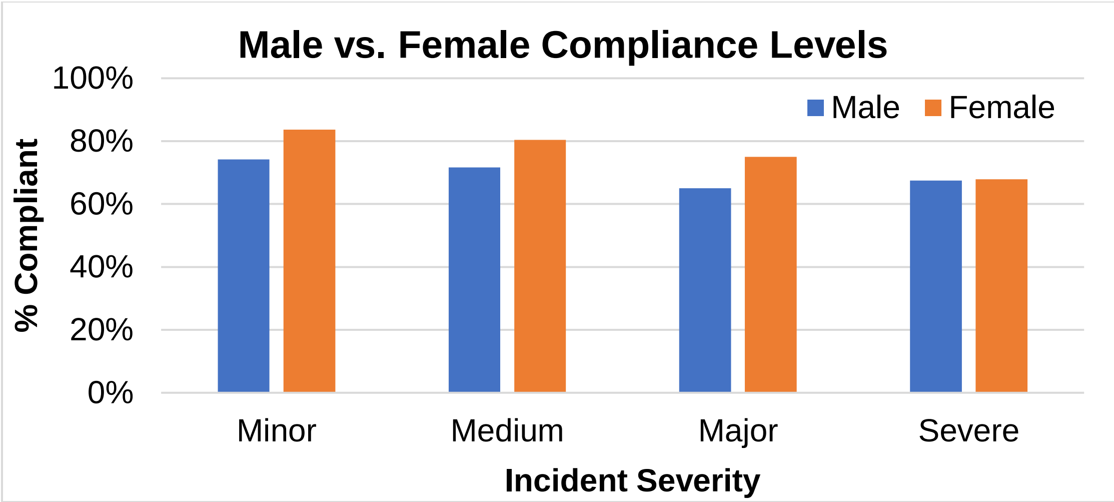


Figure 6: Male and female subject compliance with targeted VMS for treatments 5 and 6.

Table 11: Logit regression of compliance, treatments 5 and 6.

Compliance: Treatment 5 & 6			
Variable	**Meaning**	**Value**	**tStat**
(Intercept)		**1.196**	**5.859**
risk = 1	Risk Neutral	-0.125	-0.866
risk = 2	Risk Loving	0.339	1.731
lane = 1	Middle Lane	0.046	0.282
lane = 2	Right Lane	-0.162	-1.007
scenario = 2	2 Lanes Blocked	-0.177	-0.892
scenario = 3	3 Lanes Blocked, Short	**-0.491**	**-2.553**
scenario = 4	3 Lanes Blocked, Long	**-0.648**	**-3.414**
gender = female		**0.384**	**2.877**

1.4.12 Aggregate Route Choice Effects

The pooled regression across all treatments shown in Table 12 depicts the overarching effects of individual characteristics on route choice. Based on this analysis, female drivers were significantly less likely to divert than male drivers under similar driving conditions. This result is consistent with findings from many other previous studies. (Khattak et al., 1993; Emmerink et al., 1996; Abdel-Aty et al., 1997; Wardman et al., 1997; Peeta et al., 2000) Age wasn't included in this regression because there was low variation among the student subjects. My elicited risk preference didn't have a significant effect on route choice.

Notably, subjects who started in the right lane were significantly more likely to divert to the alternate route due to their close lane proximity to the exit ramp. In practice, it's typically less costly to take the exit when starting in the right lane since there's no need to slow down for an opening to change lanes right.

Table 12: Logit regression of route choice, all treatments pooled.

All Treatments			
Variable	**Meaning**	**Value**	**tStat**
(Intercept)		**-1.425**	**-13.760**
risk = 1	Risk Neutral	0.026	0.348
risk = 2	Risk Loving	0.112	1.160
lane = 1	Middle Lane	**0.204**	**2.430**
lane = 2	Right Lane	**0.642**	**7.835**
gender = female		**-0.280**	**-4.048**
VMS = 1	VMS Intensity 1, "Minor"	0.197	1.840
VMS = 2	VMS Intensity 2, "Medium"	**0.503**	**4.872**
VMS = 3	VMS Intensity 3, "Major"	**1.260**	**12.693**
VMS = 4	VMS Intensity 4, "Severe"	**1.382**	**13.913**

1.5 Conclusions

My study has shown that a real-time 2D driving simulator experiment incorporating value of time incentives and key features of the real-world environment can shed new light on how drivers decide which route to take in the face of uncertainty on the roadway. With respect to my primary research questions, I find that:

1. Subjects learned to optimize their operation of the simulated vehicle in this experiment. They reduced the amount of extraneous inputs made as well as their travel time under free flow traffic conditions.
2. All messaging schemes improved route choice and average travel times compared to the No VMS baseline, and displaying a standardized description of traffic incident severity was among the most effective strategies tested. Drivers clearly learned to differentiate

their route choice in response to message intensity. However, different information schema had significantly varied results on driver behavior. Firstly, providing additional route guidance for no incident and severe incident scenarios improved system performance. Secondly, dynamically updated information and system optima were difficult for subjects to understand – they may not improve the performance of VMS schemes without careful tuning and refinement. Finally, clear direct targeted guidance induced the most optimal route choices and resulted in the lowest average travel times.

3. Displaying VMS messages didn’t result in unpredictable or highly volatile diversion rates. Route choice variation typically stabilized after one or two periods of learning for each incident severity level.
4. Gender was the only individual characteristic with a significant effect on route choice across all treatments with females being less likely to divert than males under the same treatment conditions. This result is consistent with the findings of other studies in the literature. However, this gender effect may be confounded by an underlying effect arising from differences in risk preference between males and females that my risk elicitation task was unable to fully differentiate.

Based on these results, I conclude that system operators should utilize variable intensity message schemes for active traffic incident management. To this end, individually targeted route guidance could be the most effective method of inducing a targeted diversion response. It can be implemented using private traffic information systems, and transportation authorities should further investigate the use and development of these strategies. Otherwise, system operators should consider field testing displaying qualitative descriptions of traffic incident severity on VMS with route guidance for no incident and severe incident conditions. This strategy performed well in my experiment and is a practical candidate for expediently improving traffic incident management using extant VMS infrastructure.

CHAPTER 2
Real-time Multiplayer Driving Simulator Experiments on Amazon Mechanical Turk

2.1 Introduction

In Chapter 1, I presented a laboratory experiment to study driver response to variable message signs (VMS) using a 2D multiplayer real-time driving simulator designed by me and my colleague Amine Mahmassani. This experiment demonstrated that subjects learn to efficiently control their simulated vehicles and incorporate the information displayed by VMS to coordinate on optimal diversion responses. However, due to cost and human subject approval limitations, my experiment only recruited subjects from UCI students (mostly undergraduates) registered with the Experimental Social Science Laboratory. Although UCI is considered a diverse college campus, its subject pool is not perfectly representative of either the California or the US driving population. For example, the overwhelming majority of student subjects were between the ages of 18 and 22, whereas this age group only encompasses about 7.2% of all US drivers. (*Highway Statistics 2016*, 2016) As stated preference studies have found potential correlations between driver age and other individual characteristics on the effect of VMS on route choice (Khattak et al., 1993; Emmerink et al., 1996; Abdel-Aty et al., 1997; Wardman et al., 1997; Peeta et al., 2000) it's prudent to conduct robustness checks on the experimental findings derived from student subjects with a more representative sample even if there isn't a well-defined theoretical cause for concern. Understanding the external validity of this experiment is crucial to applying its insights to real-world traffic incident management policy.

Considering the efficacy and salience of my platform, I designed and implemented a follow-up experiment to test whether the results derived from college aged student subjects is robust and applicable to the broader US adult driver population. To this end, I undertook the substantial task of porting my driving simulator to run on the Amazon Mechanical Turk (MTurk) platform with up to 39 simultaneously connected online subjects. MTurk is an online crowdsourcing marketplace that enables requesters to recruit workers from a pool of over 500,000 registered individuals to work on Human Intelligence Tasks (HITs) – computerized tasks designed to be completed by humans. The MTurk website provides a framework for publishing HIT listings online, recruiting workers, and paying them electronically. Minor differences aside, the US MTurk worker population is comparable in demographics to the US adult working population, and over a third of MTurk requesters are academics. (Hitlin 2016) Although the MTurk worker pool won't yield a perfectly representative sample of the US driving population, it's significantly more diverse in age and other characteristics than college student subject pools (Paolacci and Chandler, 2014). I maintained maximum parity between the MTurk and laboratory experimental designs where possible and modified the experiment procedure and software where necessary to accommodate running the experiment with remote participants.

In this chapter, I will address the following key questions: 1. Is it viable to conduct real-time multiplayer experiments online using MTurk? 2. What are the characteristic differences between the MTurk and student subject groups? 3. Are there any substantial behavioral differences between the MTurk and student subject groups? I will begin by detailing the process of porting my experiment to run on MTurk and the challenges of conducting real-time multiplayer experiments online. Then, I will analyze aggregate and treatment specific topics such as subject demographics, learning, travel times, and route choices using the same methodology as in Chapter 1. Finally, I will compare the observed behavior of student versus MTurk subject groups and discuss policy implications where applicable.

2.1.1 Collaborators and Funding

My research in this chapter was conducted with help from my colleague Amine Mahmassani and our advisors: Professors David Brownstone and Michael McBride. Portions of this study have been jointly published in Brownstone et al. (2016) and Kong et al. (2017). We received significant guidance and feedback from our project manager Melissa Clark and other partners at Caltrans. This study was supported by the UC Transportation Center Multiple-Campus Award number 00008817.

2.1.2 Related Literature

Since launching in 2005, MTurk has been used to conduct countless social science research studies. Paolacci and Chandler (2014) discussed the characteristics of MTurk as a subject pool for social science research and surveyed a variety of studies conducted on the platform. In general, the authors found that MTurk workers were more diverse than college students, they truthfully and/or consistently self-reported individual characteristics, and they provided data comparable in quality to lab subjects even when paid much less. Paolacci, Chandler, and Ipeirotis (2010) and Goodman, Cryder, and Cheema (2013) found that while MTurk subjects were slightly more risk averse than college student subjects, both groups behaved similarly with respect to loss / gain framing effects. All the authors above have endorsed MTurk as a low-cost platform for conducting large scale research studies but caution that researchers must take steps to mitigate issues such as repeat participation, collusion or cheating, and worker inattentiveness during instruction or participation.

Researchers have conducted a wide variety of real-time behavioral experiments on MTurk using custom browser-based software, but few of these experiments have entailed simultaneous interaction between dozens of subjects. Crump, McDonnell, and Gureckis (2013) successfully replicated a variety of visual and real-time behavioral tasks from experimental psychology using in-browser software to trigger actions and record responses at the millisecond timescale. Studies such as theirs demonstrated the technical feasibility of implementing real-time behavioral tasks in web browsers using standard technologies such as HTML and JavaScript. Hawkins (2014) developed a platform using Node.js, HTM5 canvas, and jQuery for conducting real-time multiplayer experiments online through platforms such as MTurk. They demonstrated the ability to create two-player games with real-time physics, 2D graphics, and player input. My platform's architecture is similar in design and uses many of the same industry standard frameworks as other browser-based experiments. However, my experiment features multiplayer interaction on a much greater scale with support for 39 clients and even more entities simultaneously interacting in a single virtual environment. In addition, I'm unaware of any other real-time research-oriented driving simulator that features as many human controlled drivers sharing the same roadway as mine.

2.2 Methodology

The driving simulator in this experiment retains the core functionality and mechanics of the simulator described in Chapter 1. Subjects control 2D vehicles moving in real-time using their keyboard, up to 39 participants drive together on a shared roadway viewed from a constrained top-down perspective, and subjects are incentivized to complete their trips as quickly as possible. The road network consists of a three-lane main freeway and a two-lane alternate surface street with traffic signals. All drivers start on the main route and view the VMS display before the exit to the alternate route. Upstream of the exit, a traffic incident of random severity may occur. For a more detailed description of the simulator's core features, please see Chapter 1 section 1.2. In the following sections, I will describe the process of converting the experiment to run online with remotely connected subjects from MTurk (also see Appendix B for more information on the experiment software).

2.2.1 Amazon Mechanical Turk Integration

To conduct my experiment through the MTurk Marketplace, I adapted my procedure and software to create a Human Intelligence Task (HIT). In general, MTurk workers and requesters engage in the following process to create and complete HITs:

1. Requesters add money to their MTurk account to pay for HIT listings and/or worker bonuses.
2. Requesters create and publish their HITs using the MTurk web interface or application programming interface (API) and pay fees for doing so. Requesters can add qualification requirements to HITs to filter workers based on the presence, absence, or value of qualification attributes they possess.
3. Workers browse the HIT listings on the MTurk website and may accept HITs for which they possess the necessary qualifications attributes.
4. Workers complete and submit the HIT they accepted within the time limit specified by the requester.
5. Requesters review the results submitted by the workers and approve or reject them. Workers are automatically paid the HIT's base payment upon approval, and requesters may pay additional bonuses to any workers who have attempted a HIT. Requesters may grant workers custom qualification attributes to grant or deny access to subsequent HITs.
6. Workers review requesters and discuss HITs using 3rd party websites and forums.

The current state of the MTurk framework presents several challenges that complicate running this experiment. Firstly, the MTurk graphical web interface is ill equipped to handle the creation and management of custom web apps. They need to be hosted on separate servers and deployed/managed with the MTurk API via Amazon Web Services (AWS). Secondly, the MTurk framework was not designed for hosting tasks with concurrent multi-worker participation. Experimenters must implement their own worker pool and lobby systems to handle multiplayer experiments. Thirdly, MTurk does not provide an adequate means of communicating with workers in real-time. It's practically a necessity for the experimenter to implement a live chat system to answer subjects' questions or provide troubleshooting when things go wrong. Lastly, MTurk does not provide information or enforce requirements on workers' browser, computer, or Internet configurations. This complicates designing the experiment application for system compatibility.

To make my experiment work on MTurk, I added the following features to the driving simulator platform:

1. Integration with the MTurk API to manage HIT listings, manage worker qualifications, approve assignments, and pay worker bonuses
2. An experiment database using MongoDB to store subject and session records
3. A qualification task to determine whether workers meet system and demographic requirements to participate in the experiment
4. An experiment command and control console to monitor and manage the MTurk HIT, experiment session, and subjects in real-time
5. A live chat support system using the free service provided by tawk.to

In addition, I used pdf2htmlEX (Wang 2017) to convert and embed my instruction slides into the web app. I also used the MTurkR package for R to send mass notification emails and perform miscellaneous management on individual workers and sessions.

2.2.2 MTurk Experiment Procedure

To conduct experiments on MTurk, I first recruited workers into a subject pool through a qualification HIT, and then invited workers from the pool to participate in experiment sessions. Both the qualification and experiment HITs were restricted to US workers only using MTurk's built-in locale qualification attribute. Workers were also granted a custom qualification attribute after each HIT to prevent them from participating more than once in either the qualification or experiment HITs.

The qualification HIT consisted of three stages: a connection latency test, a browser performance test, and a pre-screening questionnaire. The latency test measured ping times between my server and the worker's computer, while the performance test measured the rendered frames per second for a test scene with a large platoon of moving vehicles. Then, the pre-screen questionnaire asked for the worker's age, possession of a valid US driver's license, ability to read English, and ability to work MTurk on weekends. Workers qualified for the experiment if their latency test averaged less than 300 ms, their performance test averaged more than 20 frames per second, and they answered yes to possessing a valid US driver's license and ability to read English. All workers were paid $0.75 for completing this HIT, their results were added to the subject pool database, and those who qualified were granted custom qualification attributes to enable them to view and accept experiment task.

Qualified workers were notified of upcoming experiment HITs one day before they were published. Once the experiment task was live, workers were asked to read through a preliminary set of instructions regarding the do's and don'ts of interacting with the experiment webpage. If they accept the task, they're redirected to the experiment site in a new browser window and placed in a waiting room until the session was launched. While they waited, workers could communicate with me using a 3rd party live chat system. This chat system proved to be critical to the success of the experiment as it enabled me to answer questions and keep workers engaged in case of delays or unexpected issues. In addition, I used sound and text alerts to notify subjects of experiment phase changes.

I typically waited up to 15 minutes for between 20 to 39 workers to connect before starting the experiment. Once started, subjects read through an instructional presentation with visual aids and screenshots. I enforced a time limit of 45 seconds per instruction slide to ensure the instruction phase lasts no longer than 15 – 17 minutes. Subjects who disconnected during instructions could reconnect and resume from the page they were on.

When all subjects finished reading the instructions, I launched the participation phase of the experiment. This phase was comprised of two parts:

Part I featured a risk elicitation task in which subjects choose between three options: receive \$3.50 with certainty, receive \$2.90 or \$4.20 with equal chance, or receive \$1.90 or \$5.00 with equal chance. These options are increasing in the spread of outcomes and slightly increasing in expected value. Based on their choice, subjects are classified as risk averse, risk neutral, or risk seeking. My three-choice task is a simplified version of a well-known five choice design. (Eckel and Grossman 2008)

Part II featured the driving task comprised of a series of 23 driving rounds – three guided practice rounds at the beginning followed by 20 normal rounds. At the beginning of each driving round, subjects started with a \$14.00 endowment that decreased at \$0.15 per second until they crossed the finish line. After all subjects either crossed the finish line or ran out of money, the next driving round would begin after a short delay. A single traffic incident may occur each round, and the same pre-randomized sequence of incidents was used for each experiment session to make order and learning effects comparable (see Table A2 in Appendix A for the full sequence of incidents used).

I monitored subjects in real-time and used alerts to remind idling subjects to pay attention to the driving task. If a subject continued idling, I situationally removed them from the experiment. Subjects who disconnected for any reason during this phase were not allowed to reconnect and were paid according to how many rounds of the experiment they completed.

After completing the participation phase, subjects filled out the post-experiment questionnaire. This questionnaire was administered using a custom Google Form that was embedded in the simulator web app but still submitted its data to Google's servers. Subjects then submitted the experiment task on MTurk, after which I approved the task submission, paid them according to their participation, and granted them a qualification attribute to prevent them from participating in future sessions.

2.2.3 MTurk Worker Compensation

Workers on MTurk are compensated through electronic payments deposited in their Amazon Pay account. Funds in the Amazon Pay account can be used to make payments directly on Amazon.com or with 3rd parties that support this payment service. Funds can also be transferred to bank accounts with no fees and a minimum transaction amount of $1.00.

When setting the payoff parameters for this experiment, I tried to strike a balance between maintaining parity with the amounts paid to the student subjects and staying within the norm of what's paid on the MTurk marketplace. In the lab, student subjects were paid a $7.00

show-up fee for arriving at the experiment on-time and a participation payment of $11.19 on average. Since sessions typically lasted around 75 minutes inclusive of the pre-experiment sign-in and post-experiment payment periods, the student subjects earned an average of $14.55 per hour. According to Hitlin 2016, 52% of the MTurk workers surveyed reported earning less than $5 per hour, 39% reported earning between $5 to $7.99 per hour, and only 8% reported earning $8 per hour or more.

Given these hourly wages, I decided to retain the original payoff structure for the participation phase of the experiment while reducing the "show-up" payment for MTurk workers to $1.50. My experiment HITs were listed with a base payment of $1.50, and my session invitations told workers to expect more than $7.00 in bonuses on average. Workers who successfully completed the entire experiment earned a participation payment of $11.18 on average. Combined with their base payment, these workers earned an average of $10.14 per hour given the typical session length of 75 minutes. Workers who accepted the HIT but did not finish the session were paid the base payment and a bonus at a rate of $7 per hour according to approximately how long they spent in the session. Workers who reported encountering a technical issue while trying to accept the HIT were situationally also paid the amount of the HIT's base payment. Since workers commonly rate and discuss MTurk tasks and requesters on several community websites, it is advisable to keep them satisfied and compensated for their time regardless of whether they complete the entire experiment.

2.2.4 Other Issues and Pitfalls

When running these experiments, I spent significant time and effort dealing with two major issues: coordination problems when starting experiment sessions and software problems due to wildly varying computer setups.

Coordination problems at the start of sessions would stem from too many workers competing for a spot in a first-come-first-serve system. Many MTurk workers use browser scripts to automatically accept new available HITs en masse but don't actually work on the tasks they accepted for a long time. This pollutes the experiment queue with workers who don't connect – an issue like students signing up but not showing up for lab experiments. I dealt with this problem by posting far more HITs than the number of workers that could be accommodated in a session and paying a "show-up fee" to excess workers who accepted the HITs. I recommend using an RSVP based system to further mitigate these problems.

Despite best efforts in testing the software, writing clear instructions, and system requirement prescreening, approximately 5% to 10% of the test subjects would disconnect from sessions due to crashes or bugs specific to their computing setup. In addition, workers would often "stress" the software through unpredictable behavior or concurrently doing other things on their computer while the experiment was running. This occasionally resulted in unhandled server-side exceptions which forced the termination of entire sessions. Over time, these issues were resolved by logging detailed information on the state of the simulator for debugging purposes. Having the live chat support system available greatly reduced the headache of appeasing frustrated workers and troubleshooting errors.

2.3 Experiment Design

Based on the best performing treatments from the laboratory experiment with student subjects, five messaging schemes were retested. Except for one treatment, the treatment conditions tested on MTurk were identical to those tested with student subjects. As before, each treatment features a coherent messaging scheme that displayed information on VMS according to traffic and/or incident conditions. For static messaging schemes, a single message was displayed to all drivers within one round of driving according to the incident severity level, while dynamic messaging schemes varied the displayed message in real-time according to the diversion response. For a complete list of the individual VMS messages shown for each treatment condition and incident scenario, see Table A1 in Appendix A.

I will present detailed analysis on the messaging treatments listed below. Only treatments with VMS were tested in the MTurk replication. When necessary, comparisons to the No VMS control treatment will be made using the data from the student subjects. I will follow the numbering and naming conventions I used in Chapter 1 Section 3. Treatments with MTurk subjects will be labeled "M#", while treatments with student subjects will be labeled "S#".

Different VMS schemes are expected to have differing effects on drivers' travel times as well as the aggregate diversion response, but all are expected to improve driver outcomes relative to the No VMS baseline. In addition, I hypothesize that there will be no significant difference in diversion response between the MTurk and student subjects under identical treatment conditions. As before, the optimal aggregate diversion response is achieved when the combined travel time among all drivers is minimized for any given incident severity level.

Treatment S1: No VMS baseline: A control treatment where no traffic information is ever displayed.

Student results: Travel times were highest in this treatment, and drivers exhibited a mixed route choice strategy. The diversion response did not change much from round to round.

Treatment S2 / M2: Qualitative description of incident severity – A treatment that displays a qualitative description of incident severity using Caltrans approved verbiage. (e.g. "ACCIDENT AHEAD, EXPECT MINOR DELAY" for incident severity 1). This serves as a benchmark for the efficacy of messaging strategies currently in use.

Student results: Drivers learned to condition their route choice on the variable intensity traffic messages, travel times were shorter than the No VMS case, and the aggregate diversion response was more optimal.

Treatment S3 / M3: Qualitative description with guidance – Same as treatment 2, but with supplemental recommendations to use the main route when there is no traffic incident (e.g. "ROAD CLEAR, ALL CARS: USE MAIN ROUTE") and to use the alternate route for severe incidents (e.g. "USE ALT RTE AHEAD"). This treatment tests whether adding a guidance recommendation can increase the diversion response.

Student results: Compared to treatment S2, showing the guidance recommendation reduced average travel time and improved the optimality of the diversion response.

Treatment S4: Dynamic diversion rate – In addition to a static qualitative description of incident severity, drivers are shown the current optimal rate at which they should divert to the alternate route (e.g. "1 IN 10 CARS SHOULD EXIT"). This rate is updated in real-time according to the usage of the two routes to *nudge* the diversion response towards the optimal target.

Student results: Displaying the desired diversion rate didn't seem to help the student subjects coordinate on achieving the optimal diversion response. Average travel time was still shorter than in the No VMS case but was the longest among treatments with VMS.

Treatment M4: Dynamic qualitative description – This treatment attempts to improve upon the dynamic messaging scheme of treatment S4 by using the same qualitative message verbiage as treatment S2 / M2 but with real-time updating of the incident severity adjective to *nudge* the diversion response towards the optimal target.

Hypothesis: Compared to treatment S4, average travel time will be shorter and the aggregate diversion response will be more optimal.

Treatment S5 / M5: Numeric IDs – In addition to a qualitative description of incident severity, each vehicle is assigned a publicly visible numeric ID between 1 to 39. The VMS message instructs vehicles within a range of IDs to use the alternate route (e.g. "IF YOUR CAR IS #1-4, USE ALT ROUTE" for incident severity 1). The idea is to use a public message to induce individual drivers to divert based on a unique characteristic such as their vehicle's license plate number or their date of birth. Given a known distribution of characteristics, system operators would be able to target subsets of drivers in traffic.

Student results: Compared to only showing the qualitative description, adding the ID targeted messages did not significantly reduce average travel time or improve the diversion response optimality. Route choice stability between rounds was the lowest among all treatments tested.

Treatment S6 / M6: Color outlines – In addition to a qualitative description of incident severity, a subset of vehicles is outlined with a bright green border each round based on the optimal number who should divert. The VMS message instructs outlined vehicles to use the alternate route. Subjects are instructed that the best possible traffic outcome is achieved if all drivers follow the recommendations. This serves as a benchmark of driver compliance with targeted recommendations.

Student results: Overall, this treatment performed the best with the shortest average travel time, most optimal diversion response, and highest route choice stability among all treatments tested with the student subjects.

2.4 Results and Discussion

In discussing the experiment results, I will focus on examining differences between the MTurk and student subject groups. For detailed analyses of the behavioral effects of VMS messaging, see Chapter 1 section 1.4.

2.4.1 Subject Characteristics

Table 13: Summary of experiment sessions.

Treatment	Subjects	Avg. Age	M	F	Licensed in USA	Avg. Weekly Hours Driven	Seen VMS
M2 / Qualitative description	36	37.2	27	9	100%	6.9	100%
M3 / Qualitative with guidance	37	37.6	20	17	100%	8.0	100%
M4 / Dynamic qualitative description	35	36.3	19	16	97%	9.6	100%
M5 / Numeric IDs	26	35.6	18	8	96%	8.3	100%
M6 / Color outlines	35	35.4	25	10	100%	8.1	100%

A summary of the treatments conducted on MTurk is shown in Table 13. MTurk subjects were 36.5 years old on average with a median age of 35, 99% held a valid US driver’s license, and all reported having seen VMS. Comparatively, student subjects recruited from the ESSL subject pool were 20.3 years old on average, 79% held a valid US driving license, and 87% reported having seen VMS before. As shown in Figure 8, the age distribution of MTurk subjects was much broader than the student subjects. Most lab subjects reported that their typical driving locale is in Southern California, while MTurk subjects hailed from across the continental US (see Figure 7). Most subjects in both groups reported receiving some form of real-time traffic information while driving. Both lab and MTurk subjects should be familiar with the physical and mental aspects of driving, but MTurk subjects have driven in a much wider variety of locales and report driving more hours per week on average (see Figure 9). With respect to race, the MTurk subjects predominantly identified as white or Caucasian, while the student subjects mainly identified as Asian (see Figure 10).

Based on results from the same risk elicitation task, the MTurk subjects were notably more risk averse than student subjects (see Figure 11). The elicited risk preference distribution is also more risk averse than that obtained by Eckel and Grossman (2008) from their university student sample, a trend that agrees with the findings of Paolacci et al. (2010) and Goodman et al. (2013). Chi-squared tests show a statistically significant difference between MTurk and student risk preference distributions (p-value < 0.001) but do not show a statistically significant difference between male and female risk preference among MTurk subjects (p-value < 0.670). Overall, the characteristic differences in age, driving habit, and risk preference imply the presence of behavioral differences between the subject groups.

Figure 7: Geographic distribution of subjects' reported driving locale.

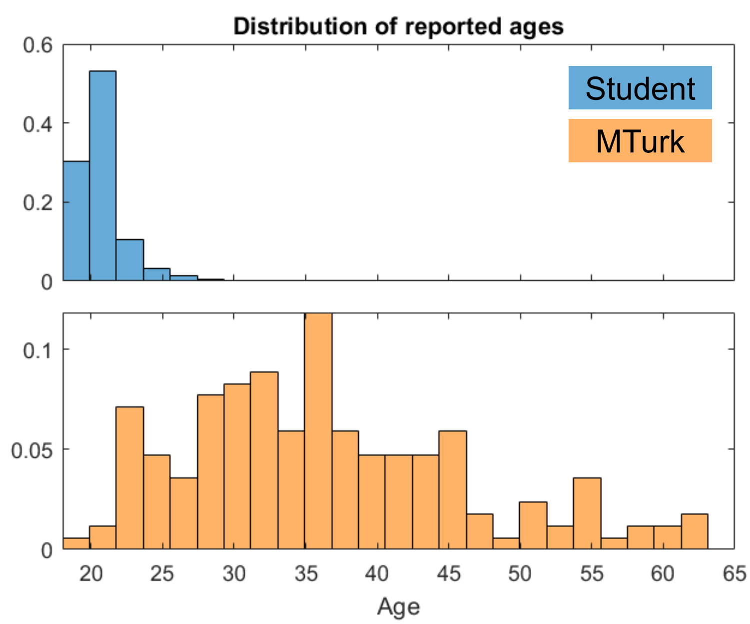


Figure 8: Distribution of subjects' reported ages.

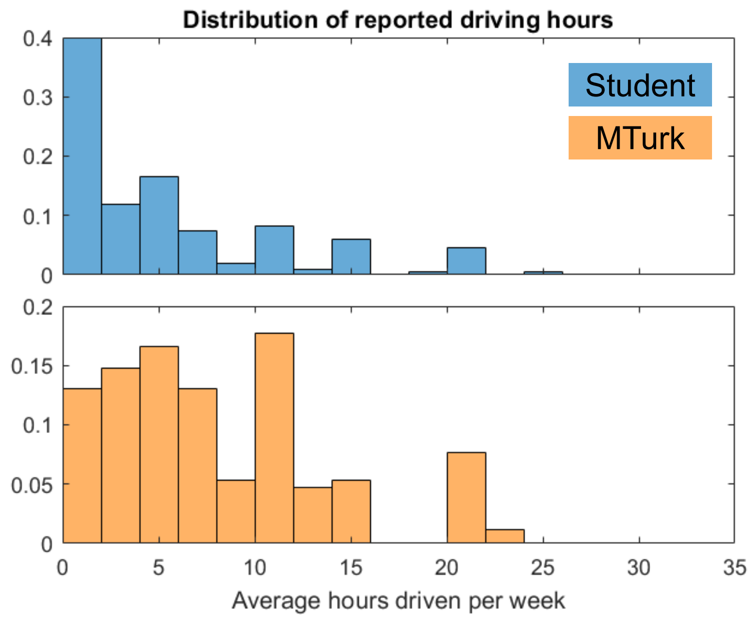


Figure 9: Distribution of subjects' reported hours driven per week.

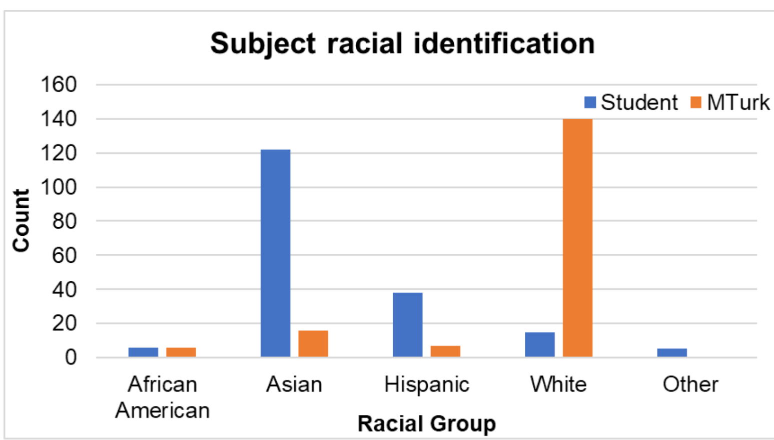


Figure 10: Reported ethnicity groups among student and MTurk subjects.

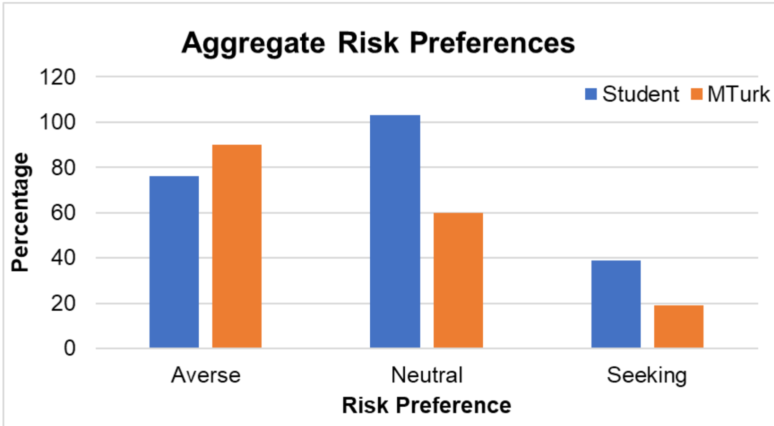


Figure 11: Elicited risk preferences among student and MTurk subjects.

2.4.2 Travel Times

Travel time is measured as the number of seconds it takes a driver to move from their starting point to the finish line in an experiment round. To compare average travel times between treatments, subjects' average travel times per round were regressed on VMS treatment dummies (shown in Table 14). In this regression, the intercept estimate is the average travel time in seconds for the No VMS baseline treatment, while the treatment / variable estimates are the average number of seconds saved compared to the baseline treatment. As expected, all MTurk VMS treatments reduce the average travel time compared to the No VMS baseline. Travel times are similar in magnitude and ranking between MTurk and student subjects for the qualitative description, qualitative description with guidance, dynamic diversion rate / description, and color outline treatments. However, average travel time for the numeric ID treatment was much shorter in the MTurk session than in the student session. As shown in Table 15, a t-test confirms that there is a statistically significant difference between the mean travel times in the MTurk and student subject sessions for that treatment. I will more detailed analyses of this treatment in the following sections to show whether there is a real behavioral anomaly.

Table 14: Linear regression of subject travel time per round.
S treatments are students, M treatments are MTurk.

Treatment / Variable	**Estimate**	**tStat**
S1 / No VMS (intercept)	**45.52**	**95.06**
M4 / Dynamic qualitative description	**-2.43**	**-3.73**
M2 / Qualitative description	**-2.51**	**-3.86**
S4 / Dynamic diversion rate	**-2.59**	**-3.97**
S2 / Qualitative description	**-2.70**	**-4.24**
S5 / Numeric IDs	**-2.71**	**-4.20**
M3 / Qualitative with guidance	**-2.92**	**-4.54**
S3 / Qualitative with guidance	**-3.05**	**-4.76**
S6 / Color outlines	**-3.08**	**-4.83**
M6 / Color outlines	**-3.14**	**-4.81**
M5 / Numeric IDs	**-4.04**	**-5.74**
F-stat vs constant model	4.18	
p-value	8.42E-06	

Table 15: t-tests for differences in mean travel time between student and MTurk sessions for the same treatment condition.

Treatment Condition Pair	**p-value**	**tStat**	**df**
S2/M2: Qualitative description	0.742	-0.330	1498
M3/S3: Qualitative with guidance	0.836	-0.208	1498
M5/S5: Numeric IDs	**0.036**	**2.099**	**1258**
M6/S6: Color outlines	0.921	0.099	1478

2.4.3 Simulator Experience

The degree to which subjects learn to operate the driving simulator efficiently can be gauged from the round-to-round trend in average keypresses and average travel time per subject for rounds with no traffic incidents. In these rounds, the optimal driving strategy is to continuously drive forward without changing lanes as soon as subjects realize there will be no incident based on what’s displayed by the VMS. As shown in Figure 12 and Figure 13, both average keypresses and travel time per subject tend to decrease over time for rounds with no incidents. This trend holds true for both MTurk and student subjects, and MTurk subjects may be slightly more efficient at operating the simulator.

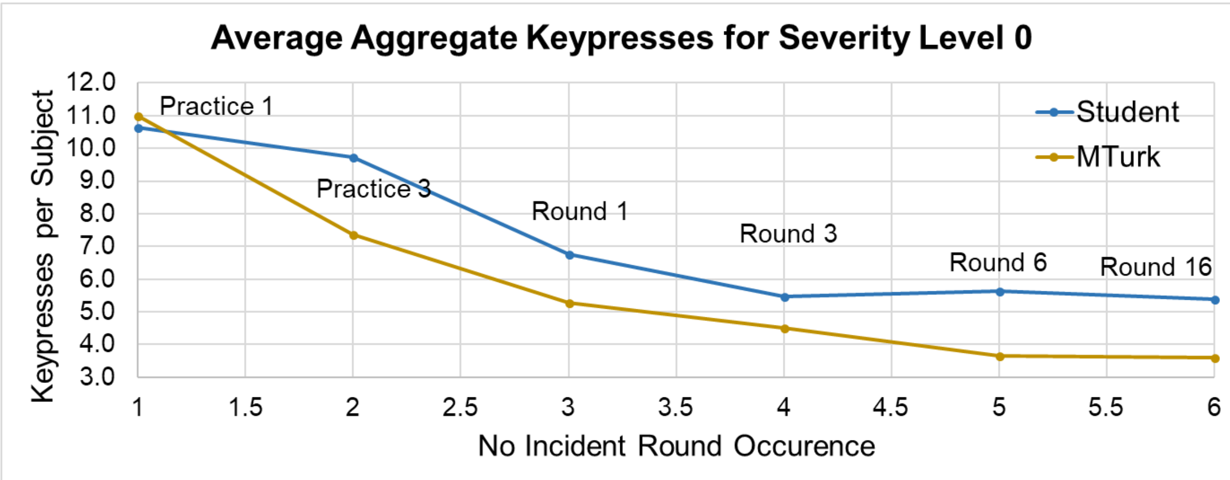


Figure 12: Average aggregate keypresses for no incident rounds.

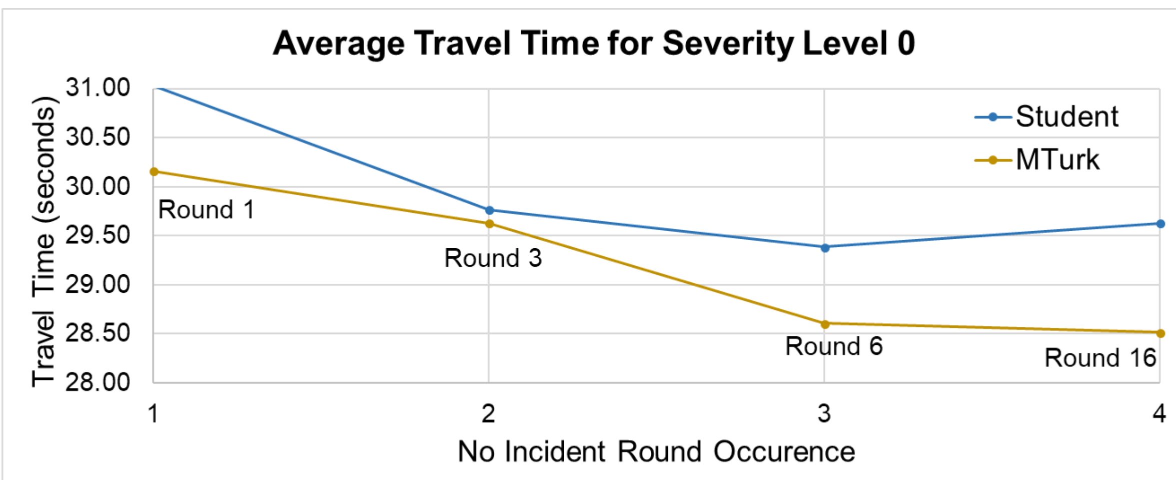


Figure 13: Average travel time for no incident rounds.

2.4.4 Treatment M4: Dynamic Qualitative Description

The dynamic qualitative description messaging treatment was the only treatment tested on MTurk that wasn't an exact replication of a treatment tested with the student subjects. The version that was tested with students displayed a dynamically updated desired diversion rate based on the remaining proportion of drivers who should divert to the alternate route. For MTurk, this treatment displayed a dynamically updated description of expected travel delay using the same standard verbiage as the static qualitative description treatments. The results indicate that this treatment didn't perform substantially better than the version tested with students.

One potential issue lies with the distinguishability of the adjectives that were displayed. The logit regression of the decision to divert (see Table 16) indicates that while there were significant increases in likelihood to divert when the incident was described as "Medium" as opposed to "Minor" or "Major" as opposed to "Medium", there was little difference in effect between incidents described as "Major" instead of "Severe" or in displaying the strong recommendation of "All cars should exit". This indicates that subjects may interpret the three strongest VMS messages very similarly without additional context. The differential impact of "Major" vs. "Severe" is also low in the static qualitative description treatments.

Table 16: Logit regression of decision to divert for treatment M4.

Treatment M4: Dynamic qualitative description			
Variable	**Meaning**	**Value**	**tStat**
(Intercept)		**-1.387**	**-5.782**
risk = 1	Risk Neutral	-0.208	-1.079
risk = 2	Risk Loving	**0.978**	**3.314**
gender = female		**0.512**	**3.012**
older = 1	Age greater than median age (35)	**-0.513**	**-2.572**
lane = 1	Middle Lane	0.112	0.548
lane = 2	Right Lane	0.019	0.093
VMS = 1	VMS Intensity 1, "Minor"	**0.603**	**2.633**
VMS = 2	VMS Intensity 2, "Medium"	**1.460**	**5.497**
VMS = 3	VMS Intensity 3, "Major"	**1.675**	**6.086**
VMS = 4	VMS Intensity 4, "Severe"	**1.625**	**6.249**
VMS = 5	VMS Intensity 5, "All cars should exit"	**1.655**	**2.733**

2.4.5 Compliance with Targeted Guidance

In the treatments with numeric IDs (M5/S5) and color outlines (M6/S6), compliance is defined as the subject choosing to divert only when instructed to do so by the individually targeted VMS messages. For the color outline treatments, student and MTurk subjects behaved similarly across all rounds (see Figure 14). For the numeric ID treatments, however, MTurk subjects were significantly more compliant with VMS guidance than their student counterparts, especially during major and severe incidents (see Figure 15). As a result, the MTurk numeric ID group also achieved a significantly shorter average travel time than the student group. For either treatment type, student subject compliance tended to decrease as incidents became more severe, whereas the opposite was true for MTurk subjects. The logit regressions of compliance shown in Table 17 confirm that MTurk subjects were significantly more likely to comply with VMS. In addition, more risk averse subjects were more likely to comply. These differences in compliance with targeted guidance could be attributed to the MTurk drivers taking the instructions more seriously or having more trust in the system operator (researcher) than the student subjects. Not complying with VMS guidance could also be indicative of exploratory behavior which may be more prevalent among student subjects. Whatever the causes may be, these results indicate that MTurk subjects can behave differently than student subjects under the same VMS treatment condition.

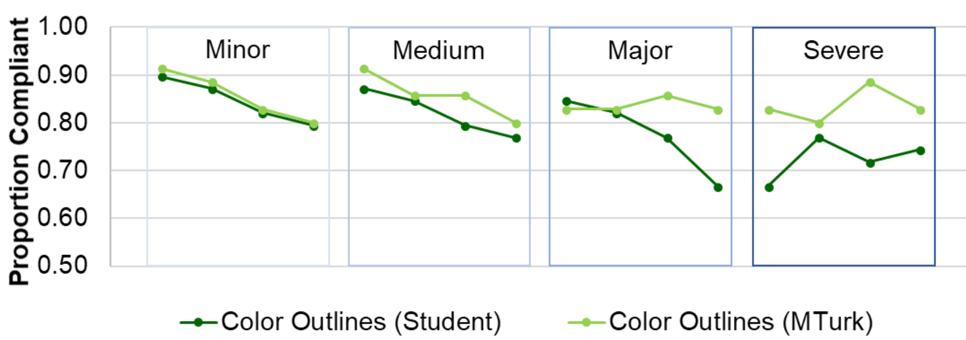


Figure 14: Compliance over time grouped by incident severity, color outline treatments.

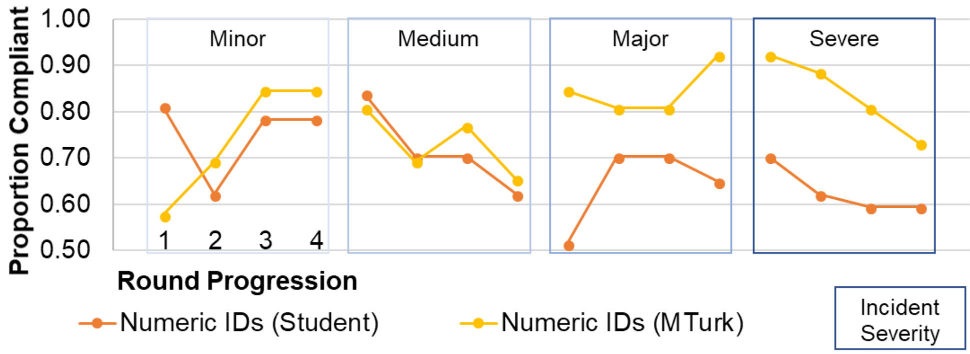


Figure 15: Compliance over time grouped by incident severity, numeric ID treatments.

Table 17: Logit regression of compliance for numeric ID and color outline treatments. (student and MTurk pooled)

		Numeric IDs		Color Outlines	
		M5 / S5		M6 / S6	
Variable	**Meaning**	**Value**	**tStat**	**Value**	**tStat**
(Intercept)		**0.928**	**3.944**	**1.410**	**5.894**
mturk	MTurk Treatment	**0.474**	**2.983**	**0.575**	**3.454**
risk = 1	Risk Neutral	**-0.353**	**-2.286**	-0.295	-1.677
risk = 2	Risk Loving	0.086	0.337	**-0.417**	**-1.970**
gender = female		0.003	0.019	**0.852**	**4.936**
lane = 1	Middle Lane	0.183	1.044	-0.029	-0.152
lane = 2	Right Lane	0.156	0.894	-0.179	-0.954
scenario = 2	2 Lanes Blocked	-0.124	-0.608	-0.104	-0.454
scenario = 3	3 Lanes Blocked, Short	-0.100	-0.491	-0.342	-1.536
scenario = 4	3 Lanes Blocked, Long	-0.168	-0.829	**-0.508**	**-2.331**

2.4.6 Aggregate Route Choice Effects

The pooled regression across all MTurk treatments shown in Table 18 depicts the overarching effects of individual characteristics on route choice. In contrast to the student results, gender did not exhibit a significant consistent overall effect on route choice across all sessions. Additionally, subjects older than the MTurk sample median age of 35 and those who were more risk seeking were more likely to divert. As with the student results, subjects who started the round further towards the right lane were more likely to divert to the alternate route due to their easier access to the exit ramp.

Table 18: Logit regression of route choice, all MTurk treatments pooled.

All MTurk Treatments			
Variable	**Meaning**	**Value**	**tStat**
(Intercept)		**-2.330**	**-18.246**
risk = 1	Risk Neutral	-0.042	-0.555
risk = 2	Risk Loving	**0.221**	**1.966**
gender = female		0.044	0.601
older = 1	Age greater than median age (35)	**0.242**	**3.390**
lane = 1	Middle Lane	**0.184**	**2.131**
lane = 2	Right Lane	**0.471**	**5.473**
VMS = 1	VMS Intensity 1, "Minor"	**0.997**	**7.845**
VMS = 2	VMS Intensity 2, "Medium"	**1.447**	**11.501**
VMS = 3	VMS Intensity 3, "Major"	**2.211**	**17.695**
VMS = 4	VMS Intensity 4, "Severe"	**2.408**	**19.242**
VMS = 5	VMS Intensity 5, "All cars should exit"	**2.408**	**4.127**

2.5 Conclusions

My MTurk results generally support the validity of the student subject results but also highlight the usefulness of replication with a broader sample of subjects. Overall, I find that:

1. It is viable to conduct real-time multiplayer experiments online using MTurk, although there are a wide variety of challenges and pitfalls that need to be mitigated.
2. There are significant differences in demographics, geographic locale, and risk preferences between MTurk and student subjects.
3. There are limited behavioral differences between MTurk and student subjects that are likely caused by hidden and revealed characteristic differences between the two groups. Most notably, MTurk subjects are much more willing to comply with targeted route guidance recommendations shown by VMS. Otherwise, the MTurk and student subject results were very similar. The change in setting between the lab and online experiments is unlikely to have significantly affected route choice or simulator operation.

Although online driving simulator experiments with crowdsourced subjects cannot replace real-world field studies, they are a useful way of checking the robustness of laboratory experiments before transitioning to more extensive and expensive field tests. The results of my MTurk experiment have further validated the potential of the proposed VMS messaging strategies to improve traffic incident management and provide system operators with a path forward to better utilize VMS infrastructure.

Through this study, I've found MTurk to be a useful venue for experimental research that makes affordable the otherwise expensive process of recruiting non-student adult subjects to participate in experiments. My driving simulator demonstrates that it's feasible to conduct large-scale multiplayer real-time experiments on this platform using browser-based software implemented with industry standard web technologies. With careful planning and consideration for expected pitfalls, the MTurk marketplace can enable researchers to study complex behavioral scenarios while sampling from populations that would be otherwise be impractical or unfeasible to recruit in the laboratory setting.

CHAPTER 3
Contextual route choice prediction using LSTM neural networks

3.1 Introduction

In the previous chapters, I demonstrated the efficacy of using variable intensity VMS messages to induce predictable changes to driver's route choice in an experimental setting. If such messaging strategies are to be incorporated into traffic incident management strategies, then the ability to predict route choice in response to the information displayed by VMS would afford system operators greater power and flexibility in utilizing VMS for incident management. The sooner the operator can predict the route choice of drivers who are exposed to VMS, the sooner they would be able to adjust VMS content displayed to drivers further downstream of the incident affected area. This will give drivers more time and opportunity to commit to and execute a route choice decision before they encounter congestion.

I implement a Long Short-Term Memory (LSTM) artificial neural network model for predicting driver route choice based on real-time information such as driver inputs, vehicle trajectory, and visible VMS leading up to the decision point between routes. Artificial neural networks are computational models that use series of interconnected signal processing nodes to transform input data into outputs. The LSTM network is a variant of recurrent neural networks (RNN) – a type of artificial neural network defined by functional units that recurrently act on both input data at a current timestep as well as the network's internal "hidden" state from a previous timestep. This hidden state functions as a type of memory that enables RNNs to internalize time-dependent correlations between input and output data features without the need to explicitly specify time-correlation parameters and structures. (See LeCun, Bengio, and Hinton (2015) for an overview of these models.)

Under the paradigm of supervised learning, RNNs can be trained through the process of backpropagation of errors, in which the gradient of the network's loss function is used to iteratively adjust the model's internal parameters or "weights" until it produces the targeted outputs from the training inputs. However, the vanilla RNN architecture encounters vanishing and exploding error gradients when backpropagating over many time steps that hinder its ability to incorporate long term correlations. The LSTM architecture mitigates this issue by introducing a special memory component to its functional unit that additively accumulates information from previous timesteps. (Hochreiter and Schmidhuber, 1997; Olah, 2015; Li, Karpathy, and Johnson, 2016) This enables the LSTM network to internalize longer term correlations than are possible with the basic RNN. In recent years, LSTMs have been used to great effect to tackle a variety of classification, prediction, and data generation problems ranging from machine translation to image captioning. (LeCun et al., 2015; Karpathy, Johnson, and Li, 2015)

I use the human driving data collected from my 2D real-time driving simulator experiment to train and validate my predictive model of driver route choice along a simple two route road network. For a complete description of the experiment methodology, see Chapter 1 section 1.2. The data captured from this experiment include vehicle trajectory, driving simulator inputs, the message displayed on VMS, route choices, and driver outcomes. Using this data, the complete sequence of simulation states can be reconstructed and distilled into inputs for the neural network, which produces probability predictions of which route drivers took.

Using the Keras deep learning framework (Chollet and others, 2015), I created, trained, and validated many variants of my model on data spanning 3 to 9 seconds (20 to 60 simulation sync-frames) before the driver's vehicle reached the decision-point that commits them to either route. I tested how different combinations of LSTM hyperparameters, input features, and input feed generators affect the model's predictive accuracy. With in-sample cross-validation, I find that a network with two 256-unit LSTM layers using a fixed 20 frame random window generator had the best accuracy for data from between three to nine seconds before the decision-point. This model predicts individual route choice with 74.0% accuracy and mis-predicts the aggregate response by 8.1%, averaged across rounds. With out-of-sample validation, this model predicts individual route choice with 72.2% accuracy and mis-predicts aggregate response by 8.8%, averaged across rounds. In both cases, the model performed best when it was not trained to full convergence and required early stopping to avoid overfitting. Overall, I find that LSTM neural networks can give operators useful predictive capability after drivers encounter VMS, but well before they reach route decision-points.

3.1.1 Funding

My research in this chapter was supported by the US Department of Transportation Pacific Southwest Region University Transportation Center Fellowship.

3.1.2 Related Literature

There is a significant body of research on using machine learning techniques to predict driver route choice, vehicle trajectories, and/or aggregate network traffic levels. Simmons et. al. (2006) build a Hidden Markov Model to predict driver route choice based on vehicle GPS coordinates in conjunction with a known map layout. They achieve up to 98% aggregate predictive accuracy after estimation and cross-validation on a sample of real driving trips, but note that the number of unforced link transitions within their dataset is typically on the order of 5%. For unforced transitions, their model's predictive accuracy falls to around 73%. Recently, an increasingly body of research has focused on using a variety of neural network architectures to predict vehicle trajectories from coordinate data or traffic imagery / visualizations. Morton and Wheeler (2016) evaluate using convolutional neural networks, LSTM neural networks, and a hybrid of both types to predict vehicle trajectory via acceleration and turn rate from the NGSIM dataset. Polson and Sokolov (2017) design a deep learning model to predict traffic flows and evaluate their model on road sensor data from Interstate I-55 during special congestion events. Ma et. al. (2015) use LSTM neural networks to perform traffic speed prediction from time-series microwave sensor data, and Ma et. al. (2017) use convolutional neural networks to perform traffic speed prediction from artificial images generated from vehicle trajectory data on sub-sections of Beijing's transportation networks. Finally, Yu et. al. (2017) combine convolutional and LSTM neural networks into a hybrid architecture that attempts to separately capture and combine spatial and temporal features of traffic data.

My methodology is unique in using complete contextual data from a controlled real-time human subject experiment as the training set for a deep learning model. This approach enables the joint incorporation of drivers' individual characteristics and their precise perceived information into my model, elements which would otherwise be latent or difficult to observe in the field today. As connected vehicles become more prevalent, more individualized driver data will become available to support the use of such models in practice.

3.2 Predictive Model

I implemented my predictive model using the Keras high-level neural networks framework. Keras provides a variety of pre-programmed neural network building blocks and enables researchers to build models quickly by connecting various neural network layers sequentially using simple function calls and parameters.

My model is comprised of the following stages: an input stage, an LSTM neural network stage, a densely connected neural network stage, and an output stage. The model is fed sequences of features extracted from simulation sync-frames to produce a vector of unit interval outputs representing the probability of each driver diverting to the alternate route. The input data feed consists of features such as the driver's lane, keyboard inputs (i.e., the desired direction of movement), VMS currently seen, and remaining distance to the decision point. At the input stage, categorical features are one-hot encoded and real-valued features are normalized to the [0, 1] interval. Each sequence of driver data is truncated to and end-aligned by the decision point. Since drivers start the round at random positions within a grid, the length of their sequences from start to decision-point will vary. For efficiency, all sequences within a round are padded to be of equal length using a signaling value such as 0 or -1 to indicate timesteps without data. At this point, the input sequences form a tensor with shape <N = number of subjects, T = number of timesteps, F = number of features / feature binaries> containing all the data of interest from the start of the round to simulation frame at which the driver reaches the decision point.

Up to four 256-unit LSTM layers with tanh activation functions act on the input data and produce outputs. If multiple LSTM layers are used, the complete output sequence at every timestep from the previous layer is fed into the next layer. The final LSTM layer returns a vector of shape <N, F> that is fed into a single fully connected neuron with sigmoid activation function, which generates a unit interval "probability" of diverting to the alternate route for each driver. During training, the binary observed route choice is used to generate error gradients for backpropagation. This process of feeding inputs, generating outputs, and backpropagation is performed for the collection (batch) of N sequences within an experiment round and repeated for batches from all other rounds use for training. Then, batch training across rounds is iteratively repeated for many training steps (epochs) until the model has reached a desired level of convergence. During prediction, a threshold such as 0.5 is used to classify which probability predictions are considered diversions, generating the final binary route choice predictions. I used a single Nvidia GeForce GTX 1080 GPU with the Theano backend during training and evaluation.

3.2.1 Model Input Feed Generation

If the model were trained on full length input sequences, it would be able to perfectly *classify* whether drivers diverted to the alternate route by learning the association between route choice and lane position. My *predictive* model was trained on incomplete sequences that end before the driver has reached the decision-point so that its weights are conditioned upon events that occur before the driver's choice is revealed. I explored using three different methods to generate partial sequences of data for training the model. First, I tested a data generator that truncates the input tensor by a fixed number of sync-frames / timesteps before the decision point but includes all available data prior to the point of truncation. Second, I tested a generator that produces a "window" tensor of fixed length (number of timesteps) from a random contiguous sub-section of the full tensor. The window's location along the full tensor is randomized from batch to batch. Third, I tested a generator that produces a window tensor of variable and randomized length from a sub-section of the full tensor. The latter two generators attempt to prevent the model from overfitting on identical sequences for each training round, which may prevent it from achieving good predictive power for sequences it has not been trained on.

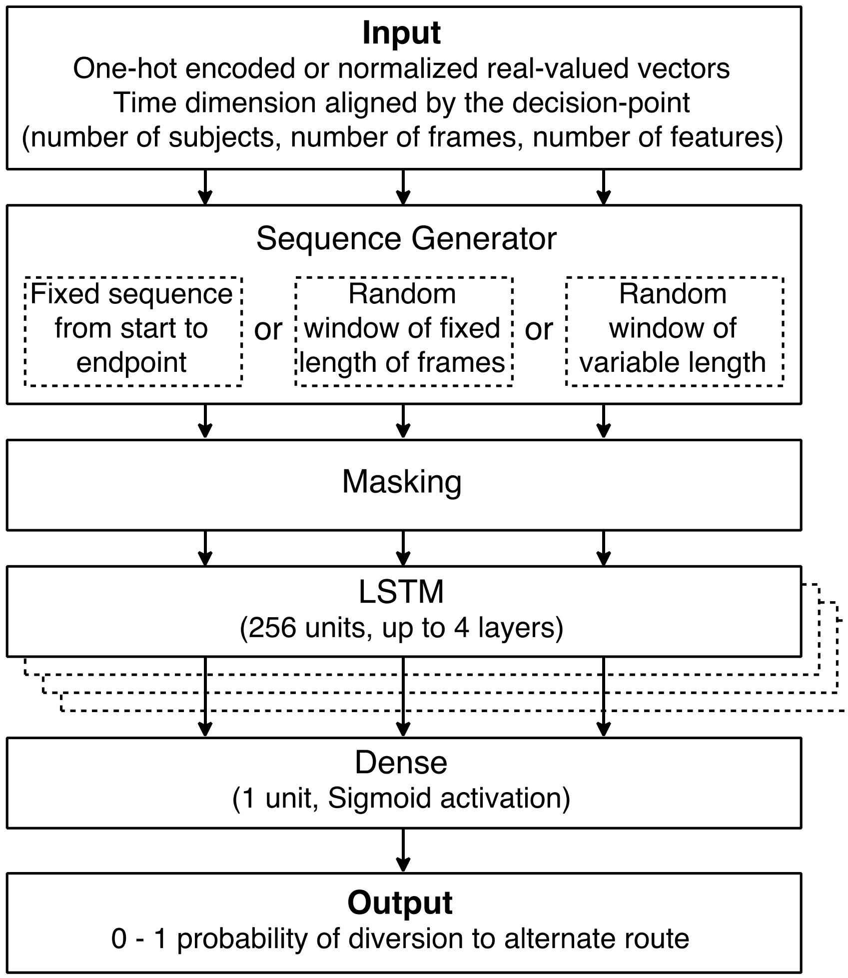


Figure 16: Block diagram of the LSTM based model.

3.3 Evaluation and Results

I used two approaches to train and evaluate my model: 1. I trained the model on a subset of rounds within an experiment session and cross-validate on the untrained subset, or 2. I trained the model on all or a subset of rounds from one experiment session and validate on out-of-sample data from a different experiment session / subject group. Between these two approaches, I expected that predictive power would be worse on average when the model is evaluated out-of-sample.

My input data consisted of the following features:

- **motion** - a vector representing the driver's input sequence to the driving simulator
- **lane** - a vector encoding the driver's current lane
- **vms** - a vector encoding the current VMS as seen by the driver
- **speed**, **direc** - vectors encoding the current traffic flow visible to the driver
- **dist** - a vector encoding the distance remaining until the driver reaches the decision-point (exit)

In the experiment, subjects should and typically always move forward at the maximum allowed speed to maximize their payoff. To divert to the alternate route, subjects must change into the exit lane on their right and remain there until they pass the decision point that commits them to either the main or alternate route. For my purposes, predictive power of my model was evaluated according to the model's ability to correctly predict route choices as early or as far downstream from the decision point as possible. Since the VMS display begins at approximately 9 seconds (20 sync-frames) before the decision-point and ends at approximately 3 seconds (60 sync-frames) before the decision-point, I performed validation for each model configuration on input sequences that terminate from between 3 to 9 seconds (20 to 60 sync-frames) before the decision point.

I present the validation results using two metrics: individual accuracy and aggregate accuracy. Individual accuracy is the percentage of route choices correctly predicted for individual drivers in a round, averaged across validation rounds. Aggregate accuracy is one minus the percentage by which the total number of predicted diverting drivers in a round deviates from the observed total number of diverting drivers, averaged across validation rounds.

3.3.1 In-sample Training and Cross-Validation

Using data from the standard description of incident severity treatment that was conducted in the laboratory, I trained the network on the first three occurrences of each incident severity type (3 rounds x 5 types = 15 rounds) and validated it on the final occurrence (1 round x 5 types = 5 rounds) within the session. A variety of networks were considered with varying hyperparameters such as the number of LSTM layers, the input data features, the input sequence generator, and the number of training epochs.

The cross-validation results are shown in Table 19. When ranked from best to worst according to overall average individual accuracy, I found that using two LSTM layers increases predictive power over one LSTM layer when all input data features are used, while using four LSTM layers seemed to worsen performance on average. The fixed and variable window sequence generators didn't dominate the fixed endpoint sequence generators, but improved average predictive accuracy and consistency across the shorter input sequences spanning 20 to 60 frames prior to the decision point. Models using the fixed endpoint generator suffered greatly in predictive accuracy when trained using early endpoints, indicating that they were not internalizing the relevant predictive features of the inputs.

Table 19: In-sample cross validation results.

Network Name	Layers	Features	Generator	Epochs	Individual Accuracy			Aggregate Accuracy		
					-9 sec	-6 sec	-3 sec	-9 sec	-6 sec	-3 sec
2 Layer, 20 Frame Window, Full Data	2 X LSTM	motion, lane, vms, speed, direc, dist	window, 20	1500	65.6%	72.8%	82.1%	92.3%	92.3%	91.3%
2 Layer, -20 Fixed Endpoint, Full Data	2 X LSTM	motion, lane, vms, speed, direc, dist	endpoint, 20	300	70.3%	72.8%	79.5%	87.7%	91.3%	95.9%
1 Layer, -20 Fixed Endpoint, Partial Data	1 X LSTM	motion, lane, vms	endpoint, 20	150	67.7%	70.3%	81.0%	84.1%	93.8%	90.3%
2 Layer, 40 Frame Window, Full Data	2 X LSTM	motion, lane, vms, speed, direc, dist	window, 40	1500	67.7%	72.8%	78.5%	93.3%	92.3%	87.7%
2 Layer, 20-60 Frame Window, Full Data	2 X LSTM	motion, lane, vms, speed, direc, dist	variable window, 20-60	1500	66.2%	73.3%	78.5%	94.9%	93.8%	93.8%

3.3.2 Out-of-sample Validation

To conduct out-of-sample validation, I trained the network on data from the standard description of incident severity treatment conducted in the laboratory and validated it on data from an identical treatment that was conducted online through Amazon Mechanical Turk (20 rounds of data for each treatment). I evaluated the best performing network structure from the in-sample case that uses two 256-unit LSTM layers with a fixed 20 frame random window generator. This network was trained and validated after 1500 and 10000 epochs.

The out-of-sample validation results are shown in Table 20. I found that the network performed slightly worse on average in individual accuracy and aggregate mis-prediction when validated out-of-sample. Additionally, network performance declined after prolonged training, indicating an overfitting issue. Early stopping before the training loss was fully minimized seemed to improve predictive accuracy. Further tests are required to better judge the correlation between training length and model performance.

Table 20: Out-of-sample validation results.

Network Name	Layers	Features	Generator	Epochs	Individual Accuracy			Aggregate Accuracy		
					-9 sec	-6 sec	-3 sec	-9 sec	-6 sec	-3 sec
2 Layer, 20 Frame Window, Full Data	2 X LSTM	motion, lane, vms, speed, direc, dist	window, 20	1500	70.1%	70.9%	76.3%	87.8%	93.2%	91.7%
2 Layer, 20 Frame Window, Full Data	2 X LSTM	motion, lane, vms, speed, direc, dist	window, 20	10000	62.8%	69.5%	78.6%	93.1%	89.5%	93.5%

3.4 Conclusions

My evaluation of LSTM models for real-time route choice prediction indicates that they can offer useful predictive power well before drivers reach the route decision point. For my best model configuration, I was able to achieve average predictive accuracies on the order of 70% (cross-validated) across different incident scenario types for input sequences that terminate in the middle of the VMS display region. In some respects, my experiment scenario represents a worst-case test of my model's predictive power since subjects may act freely with no regard to causing collisions, resulting in significantly increased jitter noise and last-minute decision making for some individuals. When applied to real-world traffic data, one can expect clearer correlations between driver lane changes and their intended route choice as well as earlier lane changes in anticipation of route selection. Additionally, one can expect to incorporate additional information such as variations in vehicle speed to further distinguish the intentions behind lane changes. Finally, the emerging public and private initiatives to rapidly transition to manufacturing connected and semi-autonomous vehicles will significantly expand both the scale of training datasets that are necessary for building predictive models as well as the volume of real-time data streams that are required for using them to mitigate traffic incidents through the dynamic provisioning of information on VMS and other sources. Neural network models such as LSTMs are well-suited towards this application as they can easily leverage big datasets to bolster predictive power. Although the training process is computationally intensive, the prediction process is not, so they can be easily used in real-time traffic management applications.

REFERENCES

Abdel-Aty, M. A., Kitamura, R., & Jovanis, P. P. (1997). Using stated preference data for studying the effect of advanced traffic information on drivers' route choice. *Transportation Research Part C: Emerging Technologies*, *5*(1), 39–50. https://doi.org/10.1016/S0968-090X(96)00023-X

Alahi, A., Goel, K., Ramanathan, V., Robicquet, A., Fei-Fei, L., & Savarese, S. (2016). Social LSTM: Human Trajectory Prediction in Crowded Spaces (pp. 961–971). IEEE. https://doi.org/10.1109/CVPR.2016.110

Allen, R. W., Stein, A. C., Rosenthal, T. J., Ziedman, D., Torres, J. F., & Halati, A. (1991). A human factors simulation investigation of driver route diversion and alternate route selection using in-vehicle navigation systems. In *Vehicle Navigation and Information Systems Conference, 1991* (Vol. 2, pp. 9–26). https://doi.org/10.1109/VNIS.1991.205745

Ben-Elia, E., & Shiftan, Y. (2010). Which road do I take? A learning-based model of route-choice behavior with real-time information. *Transportation Research Part A: Policy and Practice*, *44*(4), 249–264. https://doi.org/10.1016/j.tra.2010.01.007

Bonsall, P., & Palmer, I. (1995). The impact on route choice of direction advice and guidance from VMS signs (pp. 10–10). IEE. https://doi.org/10.1049/ic:19950251

Brocken, M. G. M., & Vlist, M. J. M. V. der. (1991). Traffic control with variable message signs. In *Vehicle Navigation and Information Systems Conference, 1991* (Vol. 2, pp. 27–45). https://doi.org/10.1109/VNIS.1991.205746

Brownstone, D., McBride, M., Kong, S.-Y., & Mahmassani, A. (2016, July 5). Experimental Studies for Traffic Incident Management. California Department of Transportation.

Retrieved from http://dot.ca.gov/research/researchreports/reports/2016/CA16-2804_FinalReport.pdf

Brownstone, D., McBride, M., Mahmassani, A., & Kong, S.-Y. (2017, June). *Experimental Studies of Traffic Incident Management with Pricing, Private Information, and Diverse Subjects*. Webinar presented at the Research Connection, Caltrans Division of Research. Retrieved from http://www.dot.ca.gov/researchconn/past_speakers/2017-06-19_DrBrownstone/Task_2770_Final_Presentation.pdf

Burgess, L., Toppen, A., & Harris, M. (2012, May). Vision and Operational Concept for Enabling Advanced Traveler Information Systems (EnableATIS). U.S. Department of Transportation. Retrieved from http://ntl.bts.gov/lib/45000/45900/45929/Final_Package_FHWA-JPO-12-052_508.pdf

Chatterjee, K, Hounsell, N. ., Firmin, P. ., & Bonsall, P. . (2002). Driver response to variable message sign information in London. *Transportation Research Part C: Emerging Technologies*, *10*(2), 149–169. https://doi.org/10.1016/S0968-090X(01)00008-0

Chatterjee, Kiron, & Mcdonald, M. (2004). Effectiveness of using variable message signs to disseminate dynamic traffic information: Evidence from field trials in European cities. *Transport Reviews*, *24*(5), 559–585. https://doi.org/10.1080/0144164042000196080

Chollet, F., & others. (2015). *Keras*. GitHub. Retrieved from https://github.com/fchollet/keras

Cragg, C. A., & Demetsky, M. J. (1995). *Simulation analysis of route diversion strategies for freeway incident management*. Charlottesville, VA: Virginia Transportation Research Council. Retrieved from https://trid.trb.org/view/448578

Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a Tool for Experimental Behavioral Research. *PLoS ONE*, *8*(3), e57410. https://doi.org/10.1371/journal.pone.0057410

Dong, J., & Mahmassani, H. S. (2009). Flow Breakdown, Travel Reliability and Real-time Information in Route Choice Behavior. In W. H. K. Lam, S. C. Wong, & H. K. Lo (Eds.), *Transportation and Traffic Theory 2009: Golden Jubilee: Papers selected for presentation at ISTTT18, a peer reviewed series since 1959* (pp. 675–695). Boston, MA: Springer US. https://doi.org/10.1007/978-1-4419-0820-9_33

Dudek, C. L. (2002, December). Guidelines for Changeable Message Sign Messages. U.S. Department of Transportation.

Dudek, C. L., Weaver, G. D., Hatcher, D. R., & Richards, S. R. (1978). Field Evaluation of Messages for Real-Time Diversion of Freeway Traffic for Special Events. *Transportation Research Record*, *682*, 37–45.

Dunn, W. M., Reiss, R. A., & Latoski, S. P. (1999). *Roadway Incident Diversion Practices*. Transportation Research Board. Retrieved from https://books.google.com/books?isbn=0309068592

Eckel, C. C., & Grossman, P. J. (2008). Forecasting risk attitudes: An experimental study using actual and forecast gamble choices. *Journal of Economic Behavior & Organization*, *68*(1), 1–17. https://doi.org/10.1016/j.jebo.2008.04.006

Emmerink, R. H. M., Nijkamp, P., Rietveld, P., & Van Ommeren, J. N. (1996). Variable message signs and radio traffic information: An integrated empirical analysis of drivers' route choice behaviour. *Transportation Research Part A: Policy and Practice*, *30*(2), 135–153. https://doi.org/10.1016/0965-8564(95)00016-X

Fernando, T., Denman, S., Sridharan, S., & Fookes, C. (2017). Soft + Hardwired Attention: An LSTM Framework for Human Trajectory Prediction and Abnormal Event Detection. *CoRR*, *abs/1702.05552*. Retrieved from http://arxiv.org/abs/1702.05552

Gan, H. (2013). Investigation of driver response to the enhanced urban freeway variable message sign information. *KSCE Journal of Civil Engineering*, *17*(6), 1455–1461. https://doi.org/10.1007/s12205-013-0231-2

Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data Collection in a Flat World: The Strengths and Weaknesses of Mechanical Turk Samples: Data Collection in a Flat World. *Journal of Behavioral Decision Making*, *26*(3), 213–224. https://doi.org/10.1002/bdm.1753

Highway Statistics 2016. (2016). Federal Highway Administration (FHWA). Retrieved from https://www.fhwa.dot.gov/policyinformation/statistics/2016/

Hitlin, P. (2016, July 11). Research in the Crowdsourcing Age, a Case Study. Pew Research Center. Retrieved from http://www.pewinternet.org/2016/07/11/research-in-the-crowdsourcing-age-a-case-study/

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Horowitz, A., Weisser, I., & Notbohm, T. (2003). Diversion from a Rural Work Zone with Traffic-Responsive Variable Message Signage System. *Transportation Research Record*, *1824*(1), 23–28. https://doi.org/10.3141/1824-03

Iida, Y., Akiyama, T., & Uchida, T. (1992). Experimental analysis of dynamic route choice behavior. *Transportation Research Part B: Methodological*, *26*(1), 17–32. https://doi.org/10.1016/0191-2615(92)90017-Q

Jou, R.-C., Lam, S.-H., Liu, Y.-H., & Chen, K.-H. (2005). Route switching behavior on freeways with the provision of different types of real-time traffic information. *Transportation Research Part A: Policy and Practice*, *39*(5), 445–461. https://doi.org/10.1016/j.tra.2005.02.004

Karpathy, A., Johnson, J., & Li, F.-F. (2015). Visualizing and Understanding Recurrent Networks. *CoRR*, *abs/1506.02078*. Retrieved from http://arxiv.org/abs/1506.02078

Kattan, L., Habib, K., Nadeem, S., & Islam, T. (2010). Modeling Travelers' Responses to Incident Information Provided by Variable Message Signs in Calgary, Canada. *Transportation Research Record: Journal of the Transportation Research Board*, *2185*, 71–80. https://doi.org/10.3141/2185-10

Khattak, A. J., Schofer, J. L., & Koppelman, F. S. (1993). Commuters' enroute diversion and return decisions: Analysis and implications for advanced traveler information systems. *Transportation Research Part A: Policy and Practice*, *27*(2), 101–111. https://doi.org/10.1016/0965-8564(93)90066-T

Kong, S.-Y., Mahmassani, A., Brownstone, D., & McBride, M. (2017). An Experimental Study of Route Choice and Incident Management Using a Real-Time 2D Driving Simulator with Variable Message Signs. In *TRB 96th Annual Meeting Compendium of Papers*. Washington DC, United States: Transportation Research Board. Retrieved from https://trid.trb.org/View/1439693

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Levinson, D., & Huo, H. (2003). Effectiveness of Variable Message Signs. Transportation Research Board. Retrieved from https://trid.trb.org/view.aspx?id=644673

Li, F.-F., Karpathy, A., & Johnson, J. (2016, February 8). CS231n Winter 2016: Lecture 10: Recurrent Neural Networks, Image Captioning, LSTM. Retrieved June 4, 2018, from https://www.youtube.com/watch?v=yCC09vCHzF8

Ma, X., Dai, Z., He, Z., Ma, J., Wang, Y., & Wang, Y. (2017). Learning Traffic as Images: A Deep Convolutional Neural Network for Large-Scale Transportation Network Speed Prediction. *Sensors*, *17*(4), 818. https://doi.org/10.3390/s17040818

Ma, X., Tao, Z., Wang, Y., Yu, H., & Wang, Y. (2015). Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. *Transportation Research Part C: Emerging Technologies*, *54*, 187–197. https://doi.org/10.1016/j.trc.2015.03.014

Mahmassani, H. S., & Jayakrishnan, R. (1991). System performance and user response under real-time information in a congested traffic corridor. *Transportation Research Part A: General*, *25*(5), 293–307. https://doi.org/10.1016/0191-2607(91)90145-G

Mahmassani, H. S., & Liu, Y.-H. (1999). Dynamics of commuting decision behaviour under advanced traveller information systems. *Transportation Research Part C: Emerging Technologies*, *7*(2–3), 91–107. https://doi.org/10.1016/S0968-090X(99)00014-5

Morton, J., & Wheeler, T. (2016). CS231n Project Report: Deep Learning of Spatial and Temporal Features for Automotive Prediction. Stanford University.

Olah, C. (2015, August 27). Understanding LSTM Networks. Retrieved June 4, 2018, from http://colah.github.io/posts/2015-08-Understanding-LSTMs/

Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a Participant Pool. *Current Directions in Psychological Science*, *23*(3), 184–188. https://doi.org/10.1177/0963721414531598

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, *5*(5), 411–419.

Peeta, S., Ramos, J. L., & Pasupathy, R. (2000). Content of Variable Message Signs and On-Line Driver Behavior. *Transportation Research Record: Journal of the Transportation Research Board*, *1725*(1), 102–108. https://doi.org/10.3141/1725-14

Polson, N. G., & Sokolov, V. O. (2017). Deep learning for short-term traffic flow prediction. *Transportation Research Part C: Emerging Technologies*, *79*, 1–17. https://doi.org/10.1016/j.trc.2017.02.024

Raju, K. (2017). *nn-trajectory-prediction*. Python. Retrieved from https://github.com/karthik4444/nn-trajectory-prediction (Original work published 2016)

Selten, R., Schreckenberg, M., Chmura, T., Pitz, T., Kube, S., Hafstein, S. F., … Wahle, J. (2004). Experimental Investigation of Day-to-Day Route-Choice Behaviour and Network Simulations of Autobahn Traffic in North Rhine-Westphalia. In M. Schreckenberg & R. Selten (Eds.), *Human Behaviour and Traffic Networks* (pp. 1–21). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-07809-9_1

Sivak, M. (2015, May). Female drivers in the United States, 1963-2013: from a minority to a majority? The University of Michigan. Retrieved from https://deepblue.lib.umich.edu/bitstream/handle/2027.42/111899/103195.pdf

Traffic Incident Management Handbook. (2000, November). Federal Highway Administration (FHWA). Retrieved from http://floridaapts.lctr.org/pdf/incident%20mgmt_handbook%20Nov00.pdf

Vemula, A. (2017). *social-lstm-tf: Tensorflow implementation of the Social LSTM model*. Python. Retrieved from https://github.com/vvanirudh/social-lstm-tf (Original work published 2016)

Wang, L. (2017). *pdf2htmlEX*. Retrieved from http://coolwanglu.github.io/pdf2htmlEX/

Wardman, M., Bonsall, P. ., & Shires, J. . (1997). Driver response to variable message signs: a stated preference investigation. *Transportation Research Part C: Emerging Technologies*, *5*(6), 389–405. https://doi.org/10.1016/S0968-090X(98)00004-7

Weaver, G. D., Dudek, C. L., Hatcher, D. R., & Stockton, W. R. (1977). Approach to Real-Time Diversion of Freeway Traffic for Special Events. *Transportation Research Record*, *644*, 57–61.

Wooster, L., & Al-Khalili, R. (2013, December 1). Changeable Message Sign (CMS) Guidelines. California Department of Transportation. Retrieved from http://www.dot.ca.gov/trafficops/tm/docs/CMS_Guidelines.pdf

Yan, X., & Wu, J. (2014). Effectiveness of Variable Message Signs on Driving Behavior Based on a Driving Simulation Experiment. *Discrete Dynamics in Nature and Society*, *2014*, 1–9. https://doi.org/10.1155/2014/206805

Yanjie Duan, Yisheng Lv, & Fei-Yue Wang. (2016). Travel time prediction with LSTM neural network (pp. 1053–1058). IEEE. https://doi.org/10.1109/ITSC.2016.7795686

Yu, H., Wu, Z., Wang, S., Wang, Y., & Ma, X. (2017). Spatiotemporal Recurrent Convolutional Networks for Traffic Prediction in Transportation Networks. *Sensors*, *17*(7), 1501. https://doi.org/10.3390/s17071501

APPENDIX A
Experiment VMS and Traffic Incidents

Table A1: VMS messages displayed for treatment and incident severity. Treatment 1, the No VMS baseline, displays nothing for any incident severity. S_ denotes a treatment with student subjects, while M denotes a treatment with MTurk subjects.

Incident Severity	VMS:	**M2/S2: Qualitative description**	**M3/S3: Qualitative with guidance**
0: No incident	LINE 1		ROAD CLEAR
	LINE 2		ALL CARS: USE MAIN ROUTE
	LINE 3		
1: One lane blocked	LINE 1	ACCIDENT AHEAD	ACCIDENT AHEAD
	LINE 2	EXPECT MINOR DELAY	EXPECT MINOR DELAY
	LINE 3		
2: Two lanes blocked	LINE 1	ACCIDENT AHEAD	ACCIDENT AHEAD
	LINE 2	EXPECT MEDIUM DELAY	EXPECT MEDIUM DELAY
	LINE 3		
3: Three lanes blocked	LINE 1	ACCIDENT AHEAD	ACCIDENT AHEAD
	LINE 2	EXPECT MAJOR DELAY	EXPECT MAJOR DELAY
10 sec delay	LINE 3		
15 sec delay	LINE 3		ALT RTE AVAILABE AHEAD
20 sec delay	LINE 3		USE ALT RTE AHEAD
4: Three lanes blocked,	LINE 1	ACCIDENT AHEAD	ACCIDENT AHEAD
extended delay	LINE 2	EXPECT SEVERE DELAY	EXPECT SEVERE DELAY
50 sec delay	LINE 3		
54 sec delay	LINE 3		ALT RTE AVAILABE AHEAD
80 sec delay	LINE 3		USE ALT RTE AHEAD

Incident Severity	VMS:	**S4: Dynamic diversion rate**	**M4: Dynamic diversion rate**
0: No incident	LINE 1		ROAD CLEAR
	LINE 2		ALL CARS: USE MAIN ROUTE
	LINE 3		
1: One lane blocked	LINE 1	MINOR ACCIDENT AHEAD	ACCIDENT AHEAD
	LINE 2	X CARS SHOULD EXIT*	**
	LINE 3		
2: Two lanes blocked	LINE 1	MEDIUM ACCIDENT AHEAD	ACCIDENT AHEAD
	LINE 2	X CARS SHOULD EXIT*	**
	LINE 3		
3: Three lanes blocked	LINE 1	MAJOR ACCIDENT AHEAD	ACCIDENT AHEAD
	LINE 2	X CARS SHOULD EXIT*	**
	LINE 3		
4: Three lanes blocked,	LINE 1	SEVERE ACCIDENT AHEAD	ACCIDENT AHEAD
extended delay	LINE 2	X CARS SHOULD EXIT*	**
	LINE 3		
		*X changes depending on the proportion P of remaining drivers who should divert: (P, X) s.t. (0.0, "NO CARS"), (0.1, "1 IN 10"), (0.2, "1 IN 5"), (0.33, "1 IN 4"), (0.5, "1 IN 3"), (0.66, "1 IN 2"), (0.88, "2 IN 3"), (>0.88, "3 IN 4")	**The message displayed changes depending on the proportion P of remaining drivers who should divert: (P, MESSAGE) s.t. (0.0, "STAY ON MAIN ROUTE"), (0.16, "EXPECT MINOR DELAY"), (0.33, "EXPECT MEDIUM DELAY"), (0.52, "EXPECT MAJOR DELAY"), (0.75, "EXPECT SEVERE DELAY"), (>0.75, "ALL CARS SHOULD EXIT")

Table A1 (continued)

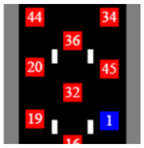
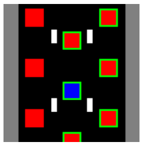
Incident Severity	VMS:	M5/S5: Numeric IDs	M6/S6: Color outlines
0: No incident	LINE 1		ROAD CLEAR
	LINE 2		ALL CARS: USE MAIN ROUTE
	LINE 3		
1: One lane blocked	LINE 1	MINOR ACCIDENT AHEAD	MINOR ACCIDENT AHEAD
	LINE 2	IF YOUR CAR IS #1-4*	GREEN OUTLINE CARS: TAKE EXIT.**
	LINE 3	USE ALT ROUTE	ALL OTHER CARS USE MAIN ROUTE
2: Two lanes blocked	LINE 1	MEDIUM ACCIDENT AHEAD	MEDIUM ACCIDENT AHEAD
	LINE 2	IF YOUR CAR IS #1-11*	GREEN OUTLINE CARS: TAKE EXIT.**
	LINE 3	USE ALT ROUTE	ALL OTHER CARS USE MAIN ROUTE
3: Three lanes blocked	LINE 1	MAJOR ACCIDENT AHEAD	MAJOR ACCIDENT AHEAD
	LINE 2	IF YOUR CAR IS #1-18*	GREEN OUTLINE CARS: TAKE EXIT.**
	LINE 3	USE ALT ROUTE	ALL OTHER CARS USE MAIN ROUTE
4: Three lanes blocked, extended delay	LINE 1	SEVERE ACCIDENT AHEAD	SEVERE ACCIDENT AHEAD
	LINE 2	IF YOUR CAR IS #1-27*	GREEN OUTLINE CARS: TAKE EXIT.**
	LINE 3	USE ALT ROUTE	ALL OTHER CARS USE MAIN ROUTE
		*Numeric IDs are displayed as white text overlaying the center of each car: 	**Green outlines are displayed as borders surrounding each car: 

Table A2: List of traffic incident severities (pre-randomized) for each experiment round. Severity levels are 0: No incident, 1: One lane blocked, 2: Two lanes blocked, 3: Three lanes blocked, and 4: Three lanes blocked, extended delay. Delay indicates the number of seconds for which traffic is slowed through the 3rd lane after the blockage clears.

Round	Severity	Delay
Practice 1	Main only	
Practice 2	Alt. only	
Practice 3	0	
1	0	
2	1	
3	0	
4	1	
5	2	
6	0	
7	4	50 s
8	3	10 s
9	4	50 s
10	1	
11	1	
12	3	15 s
13	4	54 s
14	2	
15	4	80 s
16	0	
17	3	10 s
18	2	
19	2	
20	3	20 s

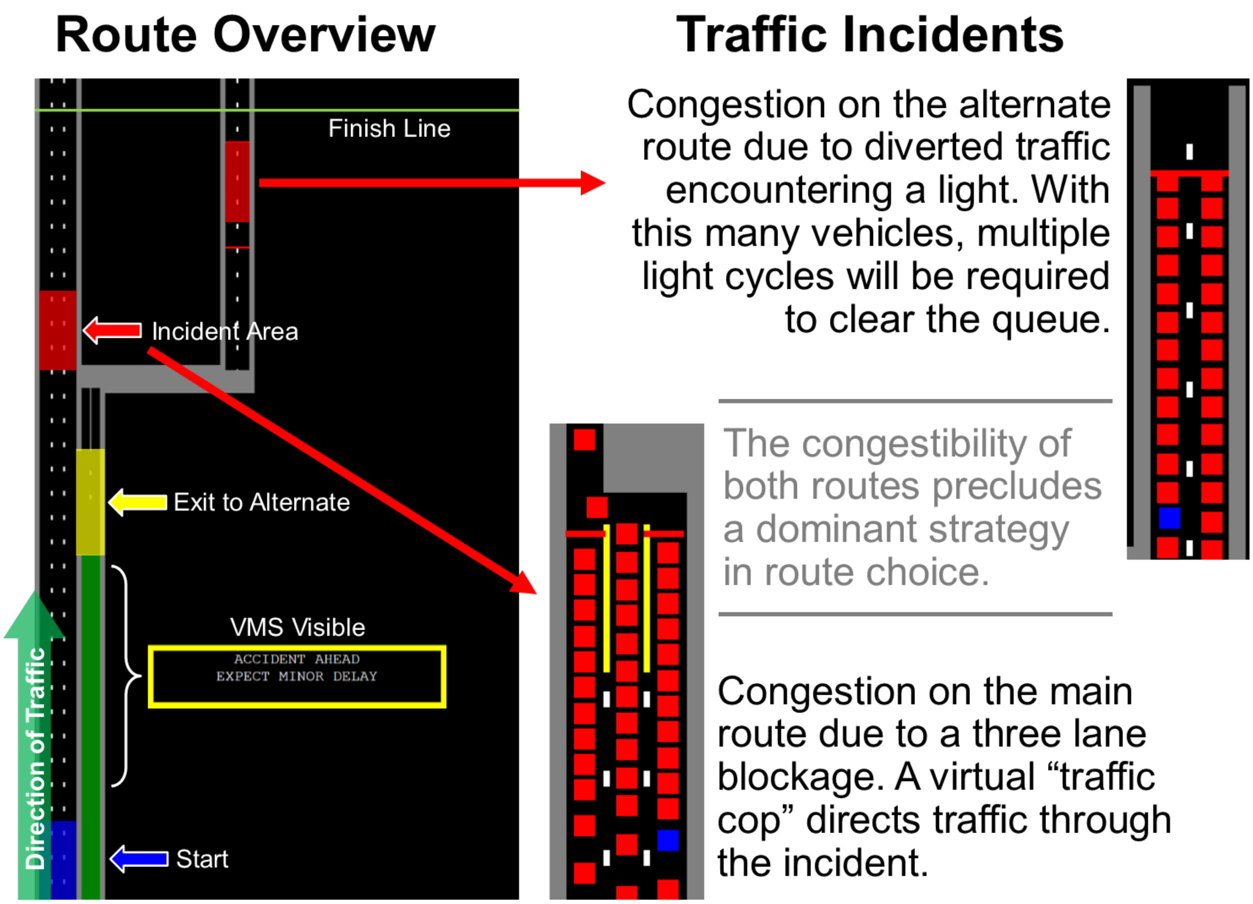


Figure A1: Route overview with traffic incident examples.

APPENDIX B
Experiment Software Implementation

B.1 Game Engine

My experiment software was written as a web application that runs in a standard HTML5 / ECMAScript 6 compliant web browser on the client side (Google Chrome preferred) and is hosted using the Node.js framework on the server side. Node.js is a server-side JavaScript run-time environment that uses the Google Chrome V8 JavaScript engine. Both the client and server-side applications are built using a self-modified version of the npm_crafty variant of the Crafty.js game engine. (See Figure B1 for an architectural overview.)

Crafty.js is a JavaScript library that features a high-level API for implementing 2D real-time videogames. It provides input handling, entity management, entity physics, graphics and animations using HTML5 Canvas or DOM, and game state management using an event / message-based system. Npm_crafty implements multiplayer functionality for Crafty by using the socket.io WebSocket library to exchange event messages between client and server instances of Crafty. Event messages received on either side can contain arbitrary message data and will trigger a user-defined callback function. The programmer can use labels to define what code runs on which client or server instances. For more information, see https://github.com/mucaho/npm_crafty.

B.2 Game Synchronization

The simplest way to implement a multiplayer game using npm_crafty is to run an identical game loop on all clients and use the server to synchronize inputs between all players. For example, a pong game can be implemented by having two clients run the same game simulation with two paddle entities and a ball with object physics. Then, each client's keyboard inputs are captured and sent as event messages containing the key inputs to the server. Finally, the server rebroadcasts these input event messages to both clients, allowing inputs commanding the movement of the paddles to be synchronized. With minimal delay, clients can now move their paddle in both their own simulation as well as their opponent's. In this case, the server only functions to receive and rebroadcast client inputs and doesn't run its own game simulation.

While this approach can work for relatively simple games with few simultaneous players, synchronization issues quickly arise as player count or simulation complexity increases. Network latency and computational deviations between clients can quickly lead to desynchronization between different client simulations, resulting in each client seeing and playing an increasingly different game!

To mitigate this issue for my driving simulator, I implemented a "master" game simulation that runs on the server instance. Player inputs are still captured and rebroadcast to all clients via the server. However, the server's master simulation also runs according to the received player inputs, and the positions of all vehicle entities on this master are collected and broadcast to all clients at regular time intervals as a sync update. When a client receives a sync update, it immediately sets all vehicle entity positions to those contained in the update. Thus, all clients and the server simulations are synchronized at regular intervals. To reduce CPU usage and conserve bandwidth, these sync updates are sent every 150 ms, or about 6.67 times a second. By letting each client's local simulation run at a target framerate of 50 frames per second in parallel to the server's master simulation, players can experience smooth framerates while remaining synchronized to each other. This technique worked flawlessly up to the 39 client limit for my experiment.

B.3 App Hosting

Npm_crafty uses the Express framework to serve files and handle HTTP requests and responses. In the laboratory, subjects are connected to the driving simulator page directly. On MTurk, subjects must click through the experiment launch page (see Figure B2) and be directed to a new tab / window containing the driving simulator page. In this case, HTTPS is required to embed the experiment's launch page in an MTurk HIT using an iframe.

When clients initially connect to the simulator, they provide an identifier via a GET query string. This string is typically either their lab computer number or their MTurk WorkerID and is used to label their vehicle entity and data within the simulator. The server will also log other connection information such as IP address, user agent, and originating HIT / assignment IDs for MTurk clients. For MTurk experiments, session and client data are stored in a MongoDB database.

B.4 Command and Control

The experiment software is launched from a command prompt / terminal and displays a variety of contextual information in the terminal window. Special localhost client instances can be used to access the command and spectator interfaces. The command interface enables the experimenter to view and manage the experiment state and connected clients. Available controls include starting or stopping the experiment, notifying or kicking clients, and publishing or expiring MTurk HITs (see Figure B3). The spectator interface uses an enlarged Crafty viewport to draw the entire road network in real-time for viewing from an all-seeing perspective.

B.5 Data Storage and Analysis

Experiment data is logged exclusively on the server as text files. Round-to-round subject outcomes are logged in CSV format, while real-time data is stored by saving client input and sync update messages to file in JSON format. Survey data is sent to Google Forms and stored as Sheets that can be downloaded. I wrote a set of MATLAB scripts to parse all types of data and perform statistical analysis. In addition, I designed a visual analysis toolkit that provides a MATLAB GUI for plotting and replaying the real-time data (see Figure B4).

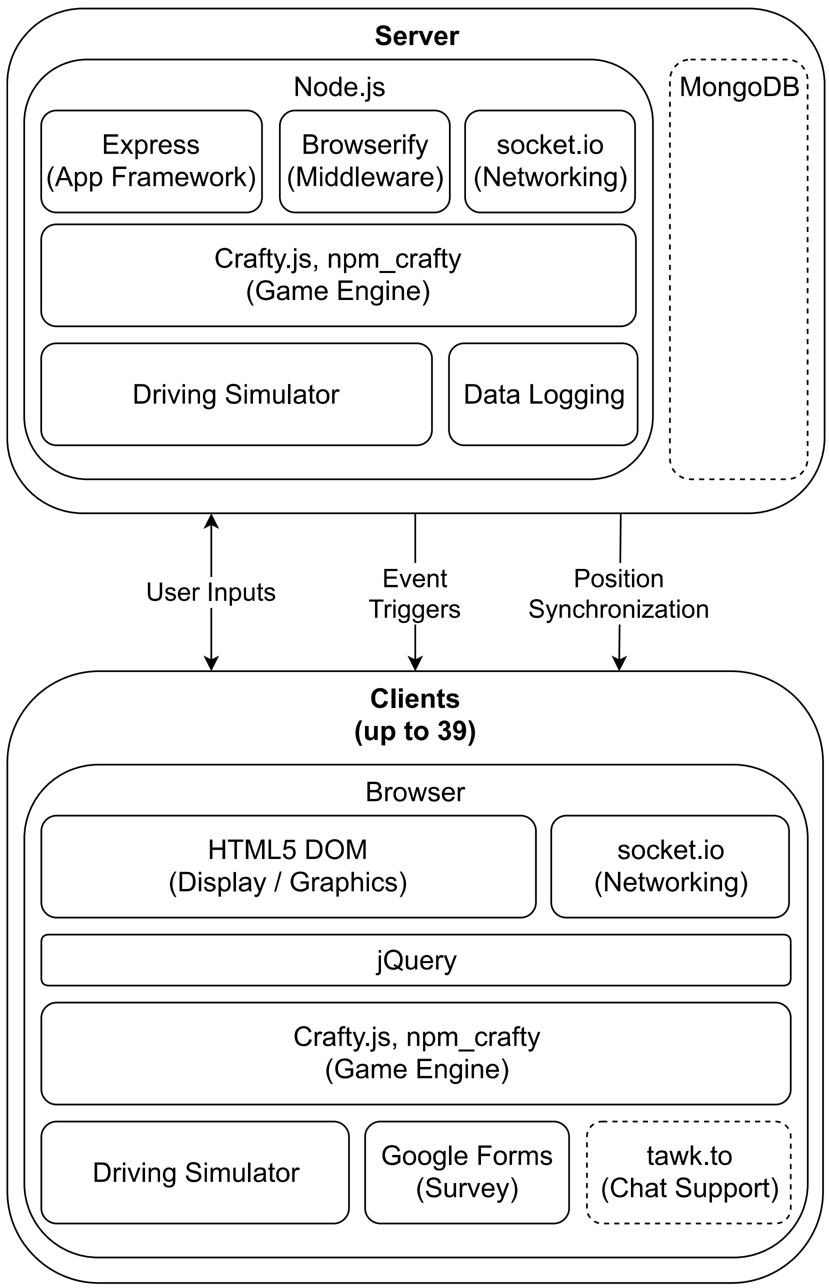


Figure B1: Experiment software architecture.
(Dashed components are only used for MTurk experiments)

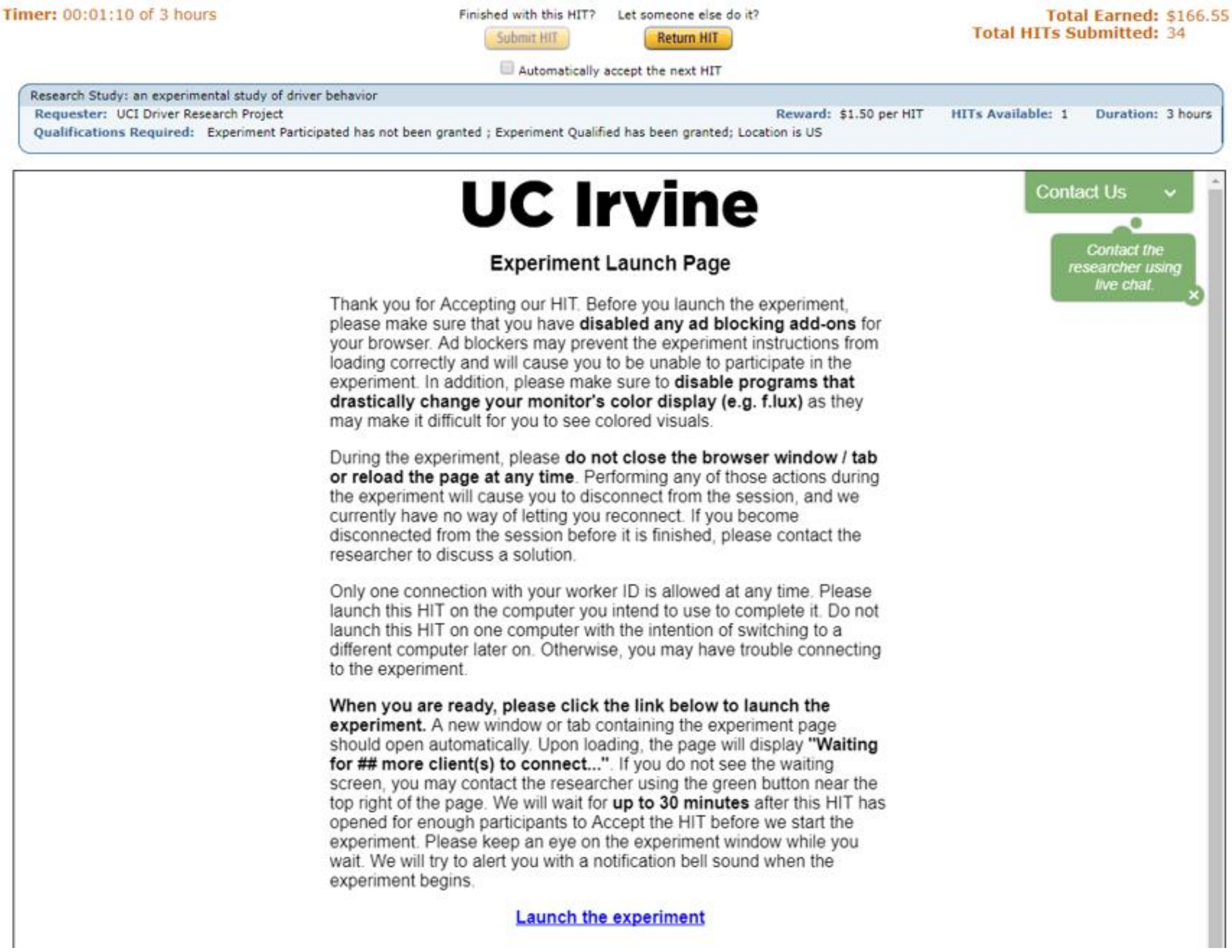


Figure B2: MTurk experiment launch page.

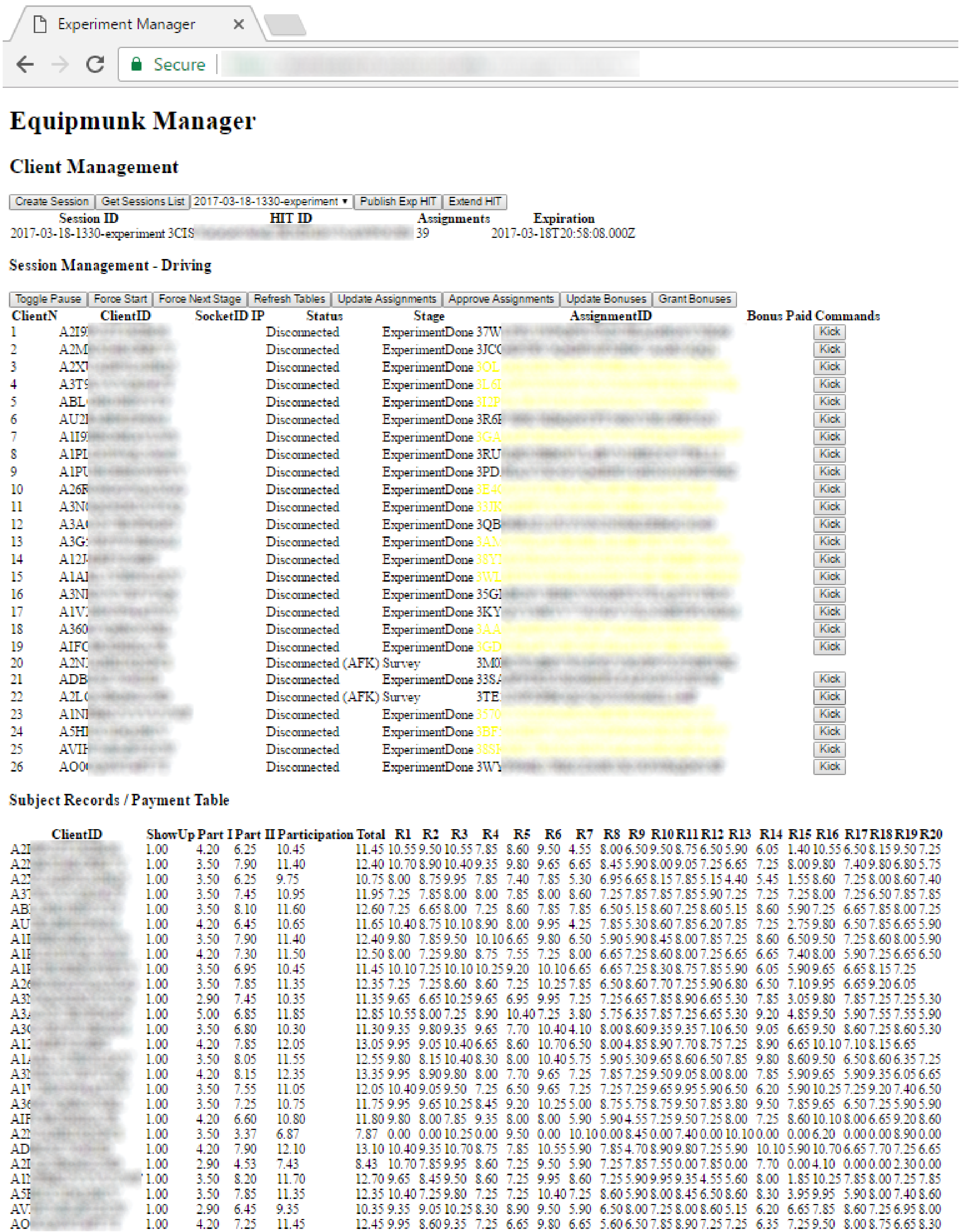


Figure B3: Experiment control interface.

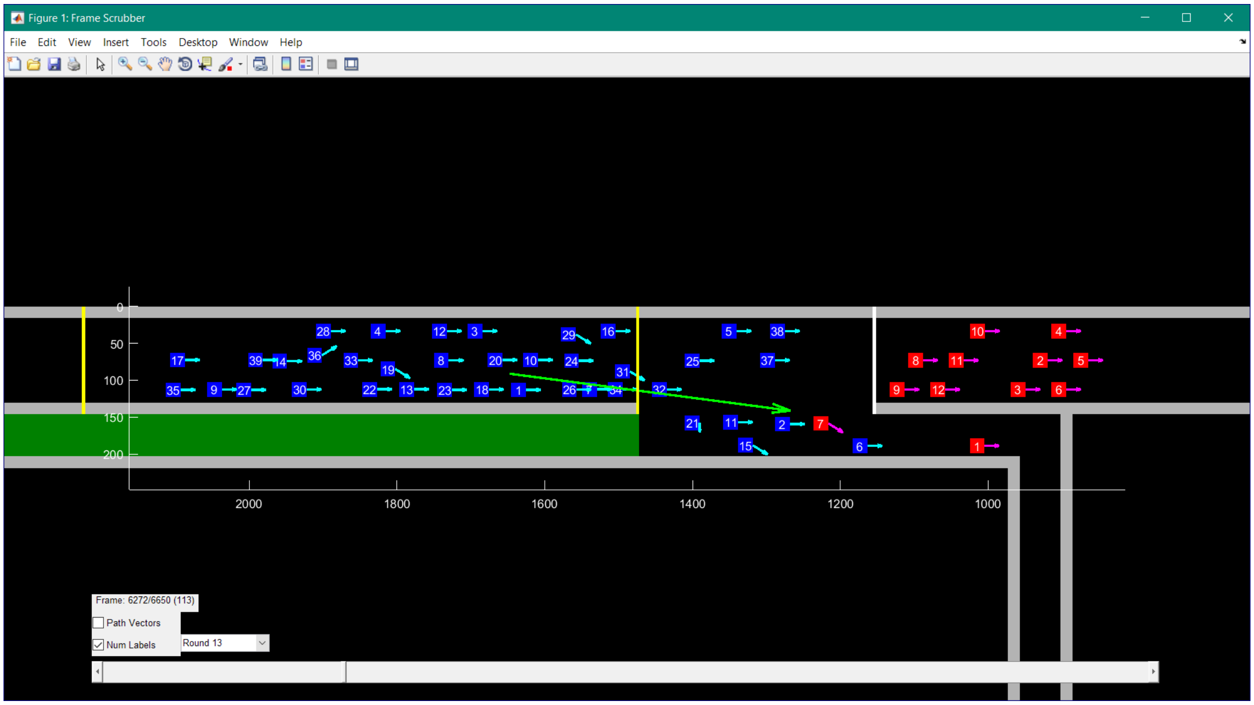


Figure B4: MATLAB experiment replay analysis tool.