

1. REPORT NUMBER CA18-2770	2. GOVERNMENT ASSOCIATION NUMBER	3. RECIPIENT'S CATALOG NUMBER
4. TITLE AND SUBTITLE Experimental Studies of Traffic Incident Management with Pricing, Private Information, and Diverse Subjects - Second Year	5. REPORT DATE 10/10/2017	
	6. PERFORMING ORGANIZATION CODE	
7. AUTHOR David Brownstone, Michael McBride, Si-Yuan Kong, Amine Mahmassani	8. PERFORMING ORGANIZATION REPORT NO.	
9. PERFORMING ORGANIZATION NAME AND ADDRESS UCCONNECT, 2614 Dwight Way, Berkeley, CA 94720-1782 University of California, Irvine (UCI), Department of Economics, 3151 Social Science Plaza, Irvine, California 92697	10. WORK UNIT NUMBER	
	11. CONTRACT OR GRANT NUMBER 65A0529 TASK ORDER 050	
12. SPONSORING AGENCY AND ADDRESS California Department of Transportation Division of Research, Innovation and System Information P.O. Box 942873 Sacramento, CA 94273-0001	13. TYPE OF REPORT AND PERIOD COVERED Final Report May 2016-June 2017	
	14. SPONSORING AGENCY CODE	
15. SUPPLEMENTARY NOTES Existing Variable Messaging Systems (VMS) give unpredictable diversion responses when used to manage traffic incidents. Is there a way to have drivers react uniformly for everyone's benefit?		
16. ABSTRACT This report documents the second year of a project using economics experimental techniques to investigate novel approaches for mitigating congestion caused by non-recurring traffic incidents. There were three goals: 1) show the results from UCI undergraduates are representative of behavior in the larger driving population, 2) investigate theoretically superior auction-based road pricing schemes, and 3) make the driving simulator more realistic. It was difficult getting a representative sample of drivers from the UCI experimental laboratory, it was decided to implement the real-time experiments using the Amazon Mechanical Turk (MTurk) to do Human Intelligence Task (HIT). This approach made it affordable to run experiments using a much larger and representative pool of experimental subjects. Experiments were completed to show the first year results were not substantially altered by using a more diverse and representative experimental subject pool. Dynamic road pricing is another possible tool for managing road congestion. However, optimal pricing requires that system operators know the distribution of the Value of Time (VOT) for road users. It is difficult to measure the VOT distribution using standard transportation survey techniques, and there is evidence that VOT varies across trip purposes and time. The second goal of this project was to investigate the possibility of using preference elicitation methods to elicit the VOT for each road user. This procedure was implemented in the experimental platform and carried out a series of experiments using UCI experimental subjects. Although the method used gives incentives for subjects to truthfully reveal their VOT, the results show that due to the cognitive complexity of the process many subjects reported erroneous VOT values. The efficiency loss due to these errors was small, demonstrating this was a promising new method of managing congestion.		
17. KEY WORDS Changeable Message Signs, Dynamic Prices, Pricing Treatments, Traffic Incident Management, Road Pricing, Pricing Schemes, Amazon Mechanical Turk, Dynamic Road Pricing, Value of Time, Variable Message Signs, Driver Behavior, non-recurring traffic incidents, Human Intelligence Task	18. DISTRIBUTION STATEMENT No restrictions. This document is available to the public through the National Technical Information Service, Springfield, VA, 22161	
19. SECURITY CLASSIFICATION (of this report) Unclassified	20. NUMBER OF PAGES 80	21. COST OF REPORT CHARGED

DISCLAIMER STATEMENT

This document is disseminated in the interest of information exchange. The contents of this report reflect the views of the authors who are responsible for the facts and accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the State of California or the Federal Highway Administration. This publication does not constitute a standard, specification or regulation. This report does not constitute an endorsement by the Department of any product described herein.

For individuals with sensory disabilities, this document is available in alternate formats. For information, call (916) 654-8899, TTY 711, or write to California Department of Transportation, Division of Research, Innovation and System Information, MS-83, P.O. Box 942873, Sacramento, CA 94273-0001.

Final Report for: Experimental Studies for Traffic Incident Management

Contractor Principal Investigator: David Brownstone and Michael McBride, PIs
Si-Yuan Kong, Amine Mahmassani
Department of Economics
University of California, Irvine

Caltrans Task Order Manager: Melissa Clark

The contents of this report reflect the views of the authors who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views of the State of California or the Federal Highway Administration. This report does not constitute a standard, specification, or regulation.

Contents

1. Introduction.....	4
2. VMS Literature Review	6
3. Experiment Platform.....	7
4. Amazon Mechanical Turk Sessions.....	10
4.1 Experiment and Treatment Design	12
4.1.1 MTurk Experiment Scenario.....	12
4.1.2 MTurk Recruitment and Session Structure.....	13
4.2 Results and Discussion	17
4.2.1 Sample Demographics	17
4.2.2 Treatment Type: Standard Qualitative Messaging.....	20
4.2.3 Treatment Type: Dynamic Qualitative Messaging	22
4.2.4 Treatment Type: Individually Targeted Messaging.....	24
4.2.5 Treatment Type: Dynamic Tolling with Qualitative Messaging	27
4.3 Concluding Remarks.....	28
5. Lane Pricing.....	30
5.1 Factors that may impact benefits in practice.....	35
5.2 Literature Review.....	36
5.3 Design/methodology	39
5.4 Results.....	44
5.4.1 Results for treatments with an assigned VOT.....	44
5.4.2 Results for Treatments with Real Subjects VOT	64
5.5 Concluding Remarks.....	73
6. References.....	75

Definitions

Acronym/Abbreviation	Meaning
AI	Computer controlled vehicles
Div	Diversion
ESSL	UCI Experimental Social Science Laboratory
HIT	Human Intelligence Task
HOT	High-Occupancy Toll
HTML	Hypertext Markup Language
ID	Identifier
IFF	If-and-only-if
MTurk	Amazon Mechanical Turk
Qual	Qualitative
Recs	Recommendations
Sc. 0 / Scen0 / s0	Incident severity level 0 (no traffic incident)
SR	State Route
VCG	Vickrey-Clarke-Groves
VMS	Variable message signs
VOT	Value of time

1. Introduction

This report documents the second year of a project using economics experimental techniques to investigate novel approaches for mitigating congestion caused by non-recurring traffic incidents. The first year demonstrated the feasibility of this approach and carried out a number of experiments using University of California, Irvine (UCI) undergraduates as experimental subjects. The experimental platform is described in Section 3 of this report. Most of the experiments conducted during the first year examined different variable message sign (VMS) wording, and later experiments examined standard road pricing schemes. It was discovered that providing any incident-related information via VMS improves system

performance relative to the no-information baseline, but also found that more complicated dynamic messaging with feedback did not always improve system performance relative to standard VMS messaging. The first year results were documented in the final report for the first year project. (Brownstone et al. 2016)

The second year of this project had three goals: 1) show the results from UCI undergraduates are representative of behavior in the larger driving population, 2) investigate theoretically superior auction-based road pricing schemes, and 3) make the driving simulator more realistic. Due to unforeseen problems with achieving the first goal, the third goal of making the simulator more realistic was abandoned. Follow-up interviews with experimental subjects did not indicate that they had troubles relating the existing simulator to real-world driving conditions.

Given the difficulty of getting a representative sample of drivers to come to the UCI experimental laboratory, it was decided to implement the real-time experiments using the Amazon Mechanical Turk (MTurk) platform. This approach made it affordable to run experiments using a much larger and representative pool of experimental subjects. Unfortunately, the limitations of the MTurk platform, coupled with the challenges of remotely administered sessions, made converting and running experiments much more difficult than anticipated. Nevertheless, enough experiments were completed to show that the first year results were not substantially altered by using a more diverse and representative experimental subject pool. The work with the MTurk platform is described in Section 4 of this report.

Dynamic road pricing is another possible tool for managing road congestion. However, optimal pricing requires that system operators know the distribution of the Value of Time (VOT) for road users. It is difficult to measure the VOT distribution using standard transportation

survey techniques, and there is evidence that VOT varies across trip purposes and time. The second goal of this project was to investigate the possibility of using preference elicitation methods to elicit the VOT for each road user. This procedure was implemented in the experimental platform and carried out a series of experiments using UCI experimental subjects. Although the method used gives incentives for subjects to truthfully reveal their VOT, the results show that due to the cognitive complexity of the process many subjects reported erroneous VOT values. Nevertheless, the efficiency loss due to these errors was small, demonstrating this is a promising new method of managing congestion. This work is described in Section 5 of this report.

2. VMS Literature Review

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 diversions¹, identified numerous factors that influence route-switching behavior², and confirmed the difficulty of attaining stable equilibria in route selection³. None of these studies has specifically examined the predictability of the diversion response as a function of message intensity, or considered how to mitigate the risk of over-diversion. Furthermore, 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. One commonly recovered parameter in studies employing econometric analysis is the effect of an alternate route's travel timesavings on the

¹ Horowitz et al. 2003, Levinson and Huo 2003, Chatterjee and McDonald 2004, Weaver et al. 1977, Dudek et al. 1978, Khattak et al. 1993

² Bonsall and Palmer 1995, Brocken and Van der Vliet 1991, Mahmassani and Liu 1999, Allen et al. 1991, Jou et al. 2005, Abdel-aty et al. 1997, Gan 2013, Mahmassani and Jayakrishnan 1991, Emmerink et al. 1996, Chatterjee et al. 2002, Kattan et al. 2010

³ Iida et al. 1992, Selten et al. 2004

probability of an individual diverting. However, this type of analysis also does not offer operators a mechanism of control since time savings are endogenous to the aggregate diversion response on a dynamic real-world network and cannot be known a priori.

At least two studies identify ways to manipulate VMS content to produce specific aggregate changes in the diversion rate. Wardman et al. (1997) demonstrate the effects of different types of messages, while Peeta et al. (2000) establish a relationship between information quantity and diversion rates. However, neither study shows if and how the manipulation of such content features to predictably achieve desired changes in diversion rates.

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. Using a learning-based model, they demonstrate that information, experience, and risk characteristics jointly affect individual driver behavior. Though informative, their study does not seek to analyze the collective effect of information on groups of drivers sharing the road. In addition, their experiment incorporates neither real incentives nor real driving.

3. Experiment Platform

Seeking to incorporate many aspects of real world driving relevant to understanding driver decision-making in a manner that is feasible to implement on ordinary lab computers. The experiment and driving simulator were designed to support the following features: 1) vehicles move in real-time and obey simplified Newtonian kinematics, requiring drivers to exert effort to maintain course and speed. 2) A large number of human drivers share the same virtual roadway to create a sense of immersive traffic. 3) Traffic and congestion are generated endogenously from a combination of the large number of vehicles and the effects of reductions in route

capacity. 4) Drivers are exposed to limited and dynamic forms of information based on that which their senses would perceive while driving on the freeway in real life, and 5) the road geometry is simplified yet retains a structure relevant to studying the questions of interest.

The experiment platform was a 2-Dimensional, real-time, top-down perspective-driving simulator implemented as a browser application using Node.js, JavaScript, and HTML5.

Subjects saw a top-down view of the roadway where vehicles were represented as small colored squares - the driver's own vehicle was colored blue while all other vehicles were colored red. The driver's viewport constantly tracked his/her vehicle and presented a fixed window of visibility around it - the driver could see farther ahead than behind to simulate the forward-focused vision of real-world drivers. From top to bottom, the driver's screen contained the following elements: the secondary information area that displayed the current experiment round, the VMS display area, the driver's viewport, and the primary information area that displayed the driver's earnings and percent completion of their itinerary in real-time.

General Information

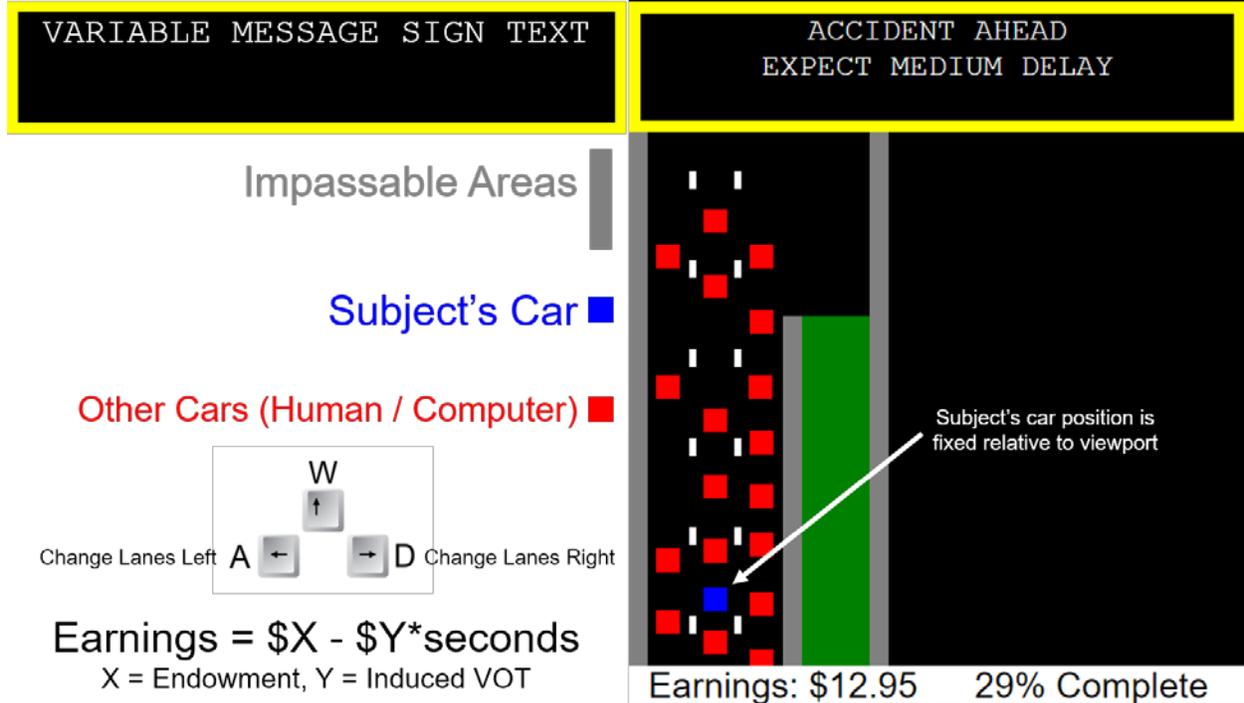


Figure 1: Driver's screen

Using the keyboard Up Arrow or letter W, Left Arrow or letter A, and Right Arrow or letter D keys, drivers' controlled their vehicles to accelerate or change lanes left and/or right. All vehicles accelerated at the same rate and quickly reached the same maximum speed. If a driver stopped accelerating, their vehicle will decelerate at a constant rate until it reached the minimum speed. The minimum speed was designed to prevent a driver from completely blocking their lane and it slowed enough, such that a driver who always traveled at the minimum speed would never complete their itinerary before their entire endowment was expended. While cruising, vehicles were automatically guided to stay in the center of the nearest lane. A minimum following distance was enforced between cruising vehicles to allow space for lane changes to occur. If another vehicle, when attempting to change lanes, obstructed a driver's vehicle, their vehicle would have slowed down slightly to allow them to move in behind the obstructing vehicle.

Drivers were informed that there were no rewards or penalties for colliding with other objects or vehicles. In addition to human controlled vehicles, the computer-controlled vehicles, which follow simple pre-defined control routines, were used to fill in the front of the driving platoon to create a sense of immersive traffic.

4. Amazon Mechanical Turk Sessions

MTurk is an online marketplace that allows requesters to crowdsource workers from a pool of over 500,000 registered individuals to work on a Human Intelligence Task (HIT). These tasks require human intelligence to complete. The MTurk website provides a framework for publishing a HIT online, recruiting workers, and paying them electronically. MTurk was used to recruit paid human subjects from across the United States (US) to participate in the VMS experiment online to replicate the student subject results with a more representative group of adult US drivers. 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) The main challenge faced in the replication was getting the simulator to run on the variety of system setups for each of the workers and ensuring they stayed connected once the experiment began.

Mechanical Turk is a marketplace for work.
 We give businesses and developers access to an on-demand, scalable workforce.
 Workers select from thousands of tasks and work whenever it's convenient.
287,132 HITs available. [View them now.](#)

Make Money by working on HITs

HITs - *Human Intelligence Tasks* - are individual tasks that you work on. [Find HITs now.](#)

As a Mechanical Turk Worker you:

- Can work from home
- Choose your own work hours
- Get paid for doing good work



or [learn more about being a Worker](#)

Get Results from Mechanical Turk Workers

Ask workers to complete HITs - *Human Intelligence Tasks* - and get results using Mechanical Turk. [Get Started.](#)

As a Mechanical Turk Requester you:

- Have access to a global, on-demand, 24 x 7 workforce
- Get thousands of HITs completed in minutes
- Pay only when you're satisfied with the results



Figure 2: The MTurk homepage

amazonmechanical turk Artificial Intelligence Your Account HITs Qualifications 245,416 HITs available now UCI Driver Research Project | Account Settings | Sign Out | Help

All HITs | HITs Available To You | HITs Assigned To You

Find HITs containing that pay at least \$ 0.00 for which you are qualified require Master Qualification GO

All HITs
 1-10 of 917 Results
 Sort by: HIT Creation Date (newest first) 60 Show all details | Hide all details 1 2 3 4 5 > Next >> Last Items per Page: 10

Purchase Date - 20 seconds Request Qualification (Why?) View a HIT in this group			
Requester: Itemizecorp	HIT Expiration Date: Jun 19, 2017 (9 hours 59 minutes)	Reward: \$0.04	HITs Available: 376
	Time Allotted: 5 minutes		
Extract summary information from 6 shopping receipts (Business, Payment Method, Total, Date) View a HIT in this group			
Requester: ScoutIt	HIT Expiration Date: Jun 26, 2017 (6 days 23 hours)	Reward: \$0.05	HITs Available: 2
	Time Allotted: 60 minutes		
Proofread English Text Take Qualification test (Why?) View a HIT in this group			
Requester: Foxtrout	HIT Expiration Date: Jun 20, 2017 (23 hours 59 minutes)	Reward: \$0.06	HITs Available: 1
	Time Allotted: 6 minutes		
Extract summary information from 6 shopping receipts (# of Items, Payment Method, Phone Number, Total) View a HIT in this group			
Requester: ScoutIt	HIT Expiration Date: Jun 26, 2017 (6 days 23 hours)	Reward: \$0.05	HITs Available: 1
	Time Allotted: 60 minutes		
Extract summary information from 8 shopping receipts (Business, Payment Method, Phone Number) View a HIT in this group			
Requester: ScoutIt	HIT Expiration Date: Jun 26, 2017 (6 days 23 hours)	Reward: \$0.05	HITs Available: 1
	Time Allotted: 60 minutes		
Extract summary information from 8 shopping receipts (Payment Method, Phone Number, Total) View a HIT in this group			
Requester: ScoutIt	HIT Expiration Date: Jun 26, 2017 (6 days 23 hours)	Reward: \$0.05	HITs Available: 1
	Time Allotted: 60 minutes		

Figure 3: The HIT listing as viewed by workers

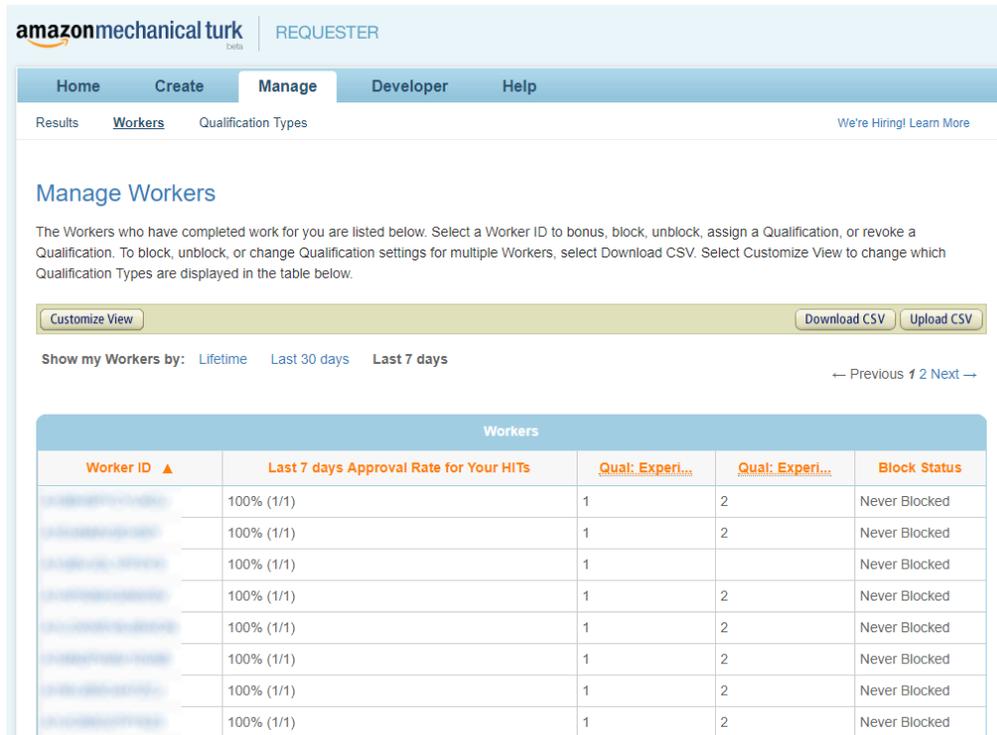


Figure 4: The requester's management interface

4.1 Experiment and Treatment Design

4.1.1 MTurk Experiment Scenario

The goal was to maintain maximum parity between the MTurk and laboratory experimental designs to ensure a comparable replication. To this end, the road network, driving dynamics, messaging schemes, and incentive structure were kept identical between settings. All vehicles started simultaneously from randomized platoon positions on three-lane highway and drove towards a shared destination. Drivers were incentivized to complete their journeys as quickly as possible and could choose between staying on the main highway or they could chose to switch to a two lane alternate surface street regulated by traffic signals. In the absence of traffic incidents, it was optimal for all drivers to use the main route. When flow on the main route is impeded, system performance was maximized when traffic was optimally split between

the two routes. If enabled, drivers were shown the VMS message on their VMS display for approximately 7.5 seconds immediately before they reached the exit to the alternate route.

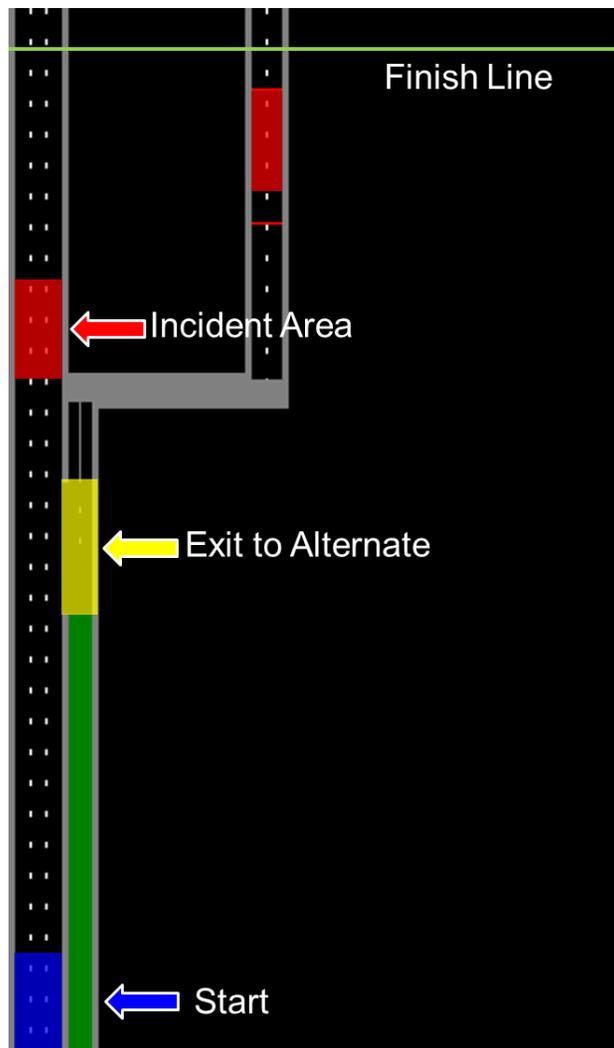


Figure 5: Road overview

4.1.2 MTurk Recruitment and Session Structure

MTurk requesters published task listings on the marketplace for workers to undertake at will. Once a worker accepted a task, they were required to complete and submit the task within the requester-defined time limit. The requester then chose to approve or reject the task submission. Workers were paid a fixed amount as a base payment upon approval of the

submission, and requesters could then pay them an optional bonus for any reason. The requester paid fees for each base and bonus payment made out to workers.

To conduct experiments on MTurk, workers were first recruited into a pool of potential subjects through a qualification task, and then the workers from the pool were invited to participate in the experiment task. Both the qualification and experiment tasks were restricted to US workers only. The qualification task consisted of four stages: an instructions page, a connection latency test, a browser performance test, and a pre-screening questionnaire. The workers were first informed of the purpose of the qualification task and the details it entailed. Then, a latency test was conducted to determine ping times between the UCI server and the workers computer. Next, a performance test was conducted to measure the rendered frames per second for a test scene with a large platoon of moving vehicles. Finally, workers answered a brief questionnaire regarding their age, possession of a valid US driver's license, ability to read English, and ability to work on MTurk over the weekends. Workers qualified for the experiment if their latency test averaged less than 300 millisecond, their performance test averaged more than 20 frames per second, and they answered yes to the questions of possessing a valid US driver's license and ability to read English. They were granted custom qualifications on MTurk to enable them to view and accept an experiment task once it was published.

Qualified workers were notified of upcoming experiment tasks one day before they were published. Once the experiment task was live online, workers were asked to read a preliminary set of instructions regarding the "do's" and "don'ts" of interacting with the experiment webpage. If they chose to continue, they would be redirected to the experiment site in a new browser window and placed in a waiting room until the session was launched. While they waited, the workers could communicate using a third party live chat system. This chat system proved to be

critical to the success of the experiments as it enabled questions to be answered regarding the experiment scenario and kept workers engaged in case of long waiting times or unexpected issues. In addition, sound and text alerts were used to notify subjects of experiment phase changes. Typical wait time was up to 15 minutes for 20 to 39 workers to connect before starting the experiment. Once started, subjects begin reading through an instructional presentation with visual aids and screenshots. A time limit of 45 seconds per instruction slide was enforced to ensure the instruction phase lasted no longer than 15 – 17 minutes. Subjects who disconnected during instructions were allowed to reconnect and resume from the page they were on and continue with the experiment.

When all subjects had finished reading the instructions, the participation phase of the experiment was launched. This phase was divided into two parts: Part I featured a risk elicitation task in which subjects chose between three lotteries with differing risk characteristics and Part II featured the driving task. Part II was comprised of a series of 23 driving rounds, during which subjects completed the same driving itinerary while being exposed to a different traffic incident scenario each round. The first three rounds were guided practice rounds that had no impact on a subject's final earnings. Each subject's final earnings were the sum of the base payment, the realization of the lottery selected in Part I, and the average of what was earned from three randomly chosen non-practice rounds in Part II. Subjects were monitored in real-time and used audible alerts to remind idling subjects to pay attention to the driving task. If a subject continuously idled while driving and provided no inputs, they may have been removed from the experiment. Subjects who disconnected for any reason during participation were not allowed to reconnect and were paid a bonus according to how many rounds of the experiment they completed.

After completing Parts I and II, subjects were asked to answer a post-experiment questionnaire to gather demographic information and optional feedback. Subjects then submitted the experiment task on MTurk and the task submission was approved. The bonus was then paid according to the earnings in Part I and Part II, and the subject was granted a custom qualification to prevent the subject from participating in future sessions of the experiment. Workers were allowed to rate and discuss MTurk tasks and requesters on several community forums and websites. It was important to keep workers satisfied and compensated them reasonably for their time regardless of the result of their participation.

The screenshot shows the MTurk interface for an experiment launch. At the top, there is a timer showing '00:01:10 of 3 hours' and buttons for 'Submit HIT' and 'Return HIT'. A status bar indicates 'Total Earned: \$166.55' and 'Total HITs Submitted: 34'. Below this, a blue box contains the following information: 'Research Study: an experimental study of driver behavior', 'Requester: UCI Driver Research Project', 'Reward: \$1.50 per HIT', 'HITs Available: 1', and 'Duration: 3 hours'. The main content area is titled 'UC Irvine Experiment Launch Page' and contains several paragraphs of instructions. A green 'Contact Us' button is visible in the top right corner, along with a live chat window. At the bottom of the page, there is a blue link that says 'Launch the experiment'.

Timer: 00:01:10 of 3 hours

Finished with this HIT? Let someone else do it?

Submit HIT Return HIT

Automatically accept the next HIT

Total Earned: \$166.55
Total HITs Submitted: 34

Research Study: an experimental study of driver behavior
Requester: UCI Driver Research Project
Qualifications Required: Experiment Participated has not been granted ; Experiment Qualified has been granted; Location is US

Reward: \$1.50 per HIT HITs Available: 1 Duration: 3 hours

UC Irvine

Experiment Launch Page

Thank you for Accepting our HIT. Before you launch the experiment, please make sure that you have **disabled any ad blocking add-ons** for your browser. Ad blockers may prevent the experiment instructions from loading correctly and will cause you to be unable to participate in the experiment. In addition, please make sure to **disable programs that drastically change your monitor's color display (e.g. f.lux)** as they may make it difficult for you to see colored visuals.

During the experiment, please **do not close the browser window / tab or reload the page at any time**. Performing any of those actions during the experiment will cause you to disconnect from the session, and we currently have no way of letting you reconnect. If you become disconnected from the session before it is finished, please contact the researcher to discuss a solution.

Only one connection with your worker ID is allowed at any time. Please launch this HIT on the computer you intend to use to complete it. Do not launch this HIT on one computer with the intention of switching to a different computer later on. Otherwise, you may have trouble connecting to the experiment.

When you are ready, please click the link below to launch the experiment. A new window or tab containing the experiment page should open automatically. Upon loading, the page will display "Waiting for ## more client(s) to connect...". If you do not see the waiting screen, you may contact the researcher using the green button near the top right of the page. We will wait for **up to 30 minutes** after this HIT has opened for enough participants to Accept the HIT before we start the experiment. Please keep an eye on the experiment window while you wait. We will try to alert you with a notification bell sound when the experiment begins.

[Launch the experiment](#)

Contact Us

Contact the researcher using live chat.

Figure 6: Experiment launch page, shown after a worker accepts the HIT

Client Management

Create Session		Get Sessions List		2017-03-18-1330-experiment		Publish Exp HIT		Extend HIT	
Session ID	HIT ID	Assignments	Expiration						
2017-03-18-1330-experiment3CIS7GG65J4MLJKXD1D37NAWFOUE939			2017-03-18T20:58:08.000Z						

Session Management - Driving									
Toggle Pause Force Start Force Next Stage Refresh Tables Update Assignments Approve Assignments Update Bonuses Grant Bonuses									
ClientN	ClientID	SocketID	IP	Status	Stage	AssignmentID	Bonus	Paid	Commands
1	A2I9PUZT19ZB08			Disconnected	ExperimentDone	37WLF8U1WPQSPUMNUEAGHG85YXK66			Kick
2	A2MKX14KCRRI7Y			Disconnected	ExperimentDone	3JCG6DTRV3Q4DF01FXH9C7AGR51QQ1			Kick
3	A2XUADP5L6HQ5			Disconnected	ExperimentDone	3OLQLKKNNSPYUP6BGMPLP9GU7LEJM			Kick
4	A3T9LYCYQIS4UT			Disconnected	ExperimentDone	3L6L49WYXWXSVOUS38ITBTEKZHW4K			Kick
5	ABL0HOIH95YT9			Disconnected	ExperimentDone	3I2PTA7R3TUL2C63SXCGLY78Z9QK9			Kick
6	AU2RMH9HZP60M			Disconnected	ExperimentDone	3R6P78PK7KBQ65YTF56SNYH15PRTGN			Kick
7	A1I9HROSHAVLWS			Disconnected	ExperimentDone	3GA6AFUKOOOZTLVWYUJQLP4QQRH3T			Kick
8	A1PLD2IWQLX4M5			Disconnected	ExperimentDone	3RUIQRXJBB0IVL4BVN1HEZCS7TRLLI			Kick
9	A1PUHCEBSOWETV			Disconnected	ExperimentDone	3PDJHANYK5GYQ4IEHYGHXO100HTH6Z			Kick
10	A26RO8GGTQANXGG			Disconnected	ExperimentDone	3E4GGUZ1T8RIAT0L0H7BK304CTVK2P			Kick
11	A3NOQZ9ZKUCCTCQ			Disconnected	ExperimentDone	33KGGHPFYCUSP3HT7CIBR4V6CTHM0N			Kick
12	A3AO2N7BNFDGF3			Disconnected	ExperimentDone	3QBD8R3Z21JU3Y0N2NSKZEBBG31049			Kick
13	A3G55RUTW3B8GM			Disconnected	ExperimentDone	3AMVWKA6YBMBL41LBB7HCYPC3760C			Kick
14	A12H4RFZ06BT			Disconnected	ExperimentDone	3EYMOXR4M2ZGNUD3CX2HYRBBF566W0			Kick
15	A1AKL5YH9NLD2V			Disconnected	ExperimentDone	3WLEIWSYHOA83ZDUW4E7BKCKCRH2I			Kick
16	A3NK57CUEJVT06			Disconnected	ExperimentDone	35GMH2SV3EHCV0XGRU5TLQ2T1VE05			Kick
17	A1V2H0UF94ATWY			Disconnected	ExperimentDone	3KYQYYS8YV773UJGVYLN38HTPUODM			Kick
18	A360G7Q6R05XRL			Disconnected	ExperimentDone	3AAJ044FG8W3KNF73GD6ZATIDUJZX			Kick
19	A1FC8R3DKKA7R			Disconnected	ExperimentDone	3GD7DDAPVUEB7AWDD1AT5975BNVIM8Z			Kick
20	A2N1AHKJQXPC1			Disconnected (AFK)	Survey	3MBCWMB8VWMT6UY80JP97X3T6HWBZ			
21	ADBZSZ70JZID			Disconnected	ExperimentDone	33SA9F9TRXUK0961FLFAF30W3UEWH			Kick
22	A2LQ38BQ8G259D			Disconnected (AFK)	Survey	3TE22NXPBCQZ5J3336186ZL144F			Win
23	A1NFBUKUYVUUYWF			Disconnected	ExperimentDone	3570Y55ZPDB8S3Q3BTRV7F9GKR9GYZ			Kick
24	A5HCCIDQ2HCV			Disconnected	ExperimentDone	3BF31CHDTVA3CTWZFP900NRQNE7H0Y			Kick
25	AVIH74S1MSTZCW			Disconnected	ExperimentDone	38SKSKUTRING8KWSAK44MJKOQPXLI4			Kick
26	A00CQ0WT4FT7T			Disconnected	ExperimentDone	3WYP994K17RKZC83RYK510WRQD6Y6F			Kick

Subject Records / Payment Table																									
ClientID	ShowUp	Part I	Part II	Participation	Total	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R16	R17	R18	R19	R20
A2I9PUZT19ZB08	1.00	4.20	6.25	10.45	11.45	10.55	9.50	10.55	7.85	8.60	9.50	4.55	8.00	6.50	9.50	8.75	6.50	5.90	6.05	1.40	10.55	6.50	8.15	9.50	7.25
A2MKX14KCRRI7Y	1.00	3.50	7.90	11.40	12.40	10.70	8.90	10.40	9.35	9.80	9.65	6.65	8.45	5.90	8.00	9.05	7.25	6.65	7.25	8.00	9.80	7.40	9.80	6.80	5.75
A2XUADP5L6HQ5	1.00	3.50	6.25	9.75	10.75	8.00	8.75	9.95	7.85	7.40	7.85	3.30	6.95	6.65	8.15	7.85	5.15	4.40	5.45	1.55	8.60	7.25	8.00	8.60	7.40
A3T9LYCYQIS4UT	1.00	3.50	7.45	10.95	11.95	7.25	7.85	8.00	8.00	7.85	8.00	8.60	7.25	7.85	7.85	7.85	5.90	7.25	7.25	7.25	8.00	7.25	6.50	7.85	7.85
ABL0HOIH95YT9	1.00	3.50	8.10	11.60	12.60	7.25	6.65	8.00	7.25	8.60	7.85	7.85	6.50	5.15	8.60	7.25	8.60	5.15	8.60	5.90	7.25	6.65	7.85	8.00	7.25
AU2RMH9HZP60M	1.00	4.20	6.45	10.65	11.65	10.40	8.75	10.10	8.90	8.00	9.95	4.25	7.85	5.30	8.60	7.85	6.20	7.85	7.25	2.75	9.80	6.50	7.85	6.65	5.90
A1I9HROSHAVLWS	1.00	3.50	7.90	11.40	12.40	9.80	7.85	9.50	10.10	6.65	9.80	6.50	5.90	5.90	8.45	8.00	7.85	7.25	8.60	5.90	5.90	7.25	8.60	8.00	5.90
A1PLD2IWQLX4M5	1.00	4.20	7.30	11.50	12.50	8.00	7.25	9.80	8.75	7.55	7.25	8.00	6.65	7.25	8.60	8.00	7.25	6.65	6.65	7.40	8.00	5.90	7.25	6.65	6.50
A1PUHCEBSOWETV	1.00	3.50	6.95	10.45	11.45	10.10	7.25	10.10	10.25	9.20	10.10	6.65	6.65	7.25	8.30	8.75	7.85	5.90	6.05	5.90	9.65	6.65	8.15	7.25	
A26RO8GGTQANXGG	1.00	3.50	7.85	11.35	12.35	7.25	7.25	8.60	8.60	7.25	10.25	7.85	6.50	8.60	7.70	7.25	5.90	6.80	6.50	7.10	9.95	6.65	9.20	6.05	
A3NOQZ9ZKUCCTCQ	1.00	2.90	7.45	10.35	11.35	9.65	6.65	10.25	9.65	6.95	9.95	7.25	7.25	6.65	7.85	8.90	6.65	5.30	7.85	3.05	9.80	7.85	7.25	7.25	5.30
A3AO2N7BNFDGF3	1.00	5.00	6.85	11.85	12.85	10.55	8.00	7.25	8.90	10.40	7.25	3.80	5.75	6.35	7.85	7.25	6.65	5.30	9.20	4.85	9.80	5.90	7.55	7.55	5.90
A3G55RUTW3B8GM	1.00	3.50	6.80	10.30	11.30	9.35	9.80	9.35	9.65	7.70	10.40	4.10	8.00	8.60	9.35	9.35	7.10	6.50	9.05	6.65	9.50	8.60	7.25	8.60	5.30

Figure 7: Server side experiment management interface

4.2 Results and Discussion

Based on the best performing treatments from the laboratory VMS experiment, six of the messaging schemes were retested. Although many other treatments in our laboratory experiments were tested, the most refined iterations of treatments with positive results were selectively chosen to replicate.

4.2.1 Sample Demographics

Discarding the results of subjects who did not successfully finish the experiment, the sampled amount was 205 MTurk workers across six experiment sessions. Compared to the lab subjects from the UCI student body, the MTurk subjects were considerably older (on average) and were skewed male as opposed to skewed female. On both a per session and aggregate basis,

more MTurk subjects played video games, possessed a US driver’s license, had driven in the past year, and reported having seen VMS while driving. The two groups were somewhat similar with respect to receiving real-time traffic information. Results from the risk elicitation lottery indicated that MTurk subjects were on average significantly more risk averse, with far fewer participants selecting the risk loving option compared to lab subjects. This might be attributed to the student subjects in the lab viewing their potential earnings in a more casual light than MTurk workers, and were more willing to pick the riskier lottery options. Student subjects were typically younger and likely under less financial pressure than MTurk workers, some of whom depended on their MTurk earnings as a major source of income. Geographically speaking, MTurk subjects lived and drove in a plethora of locations from across the continental US. In this regard, they were considerably more representative of US adult driver population as a whole than the sample from UCI.

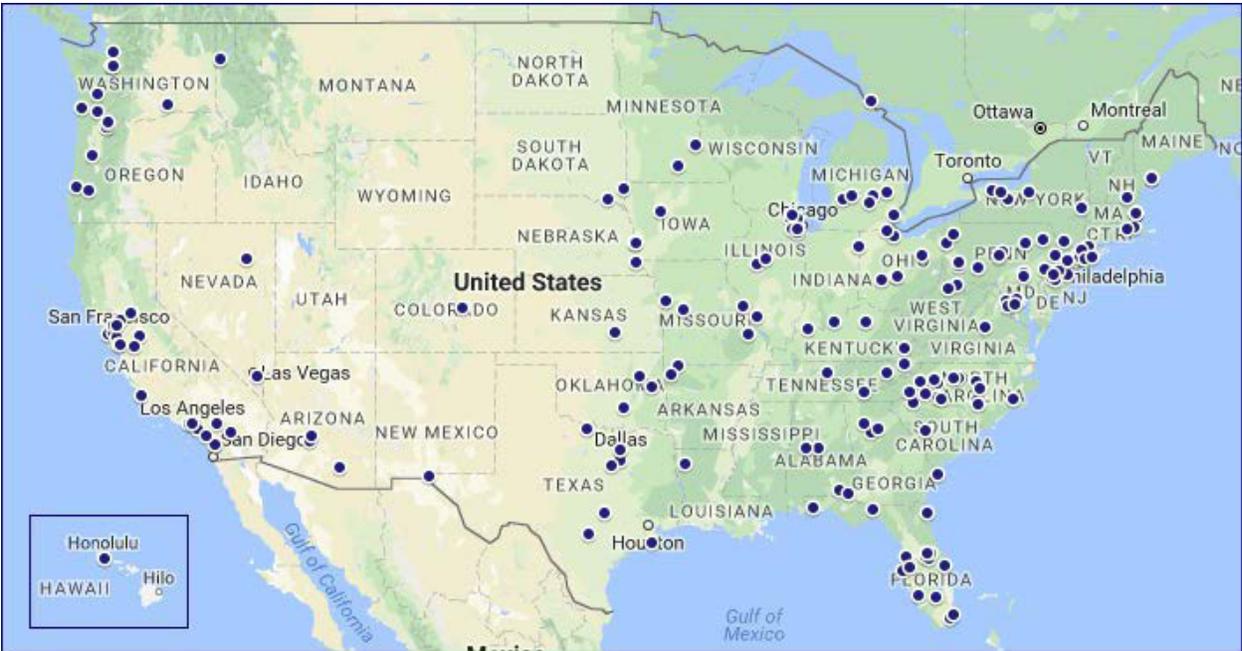


Figure 8: Map of the stated locations of MTurk subjects

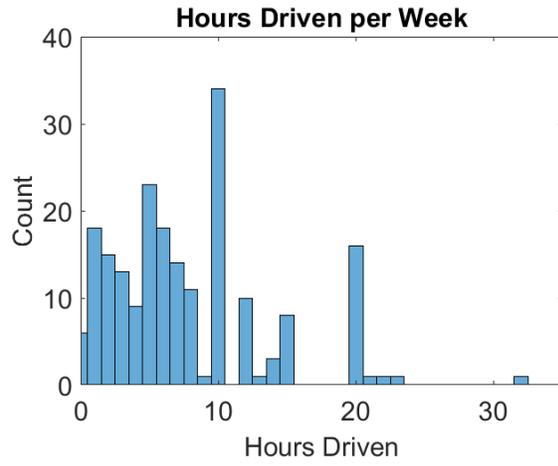
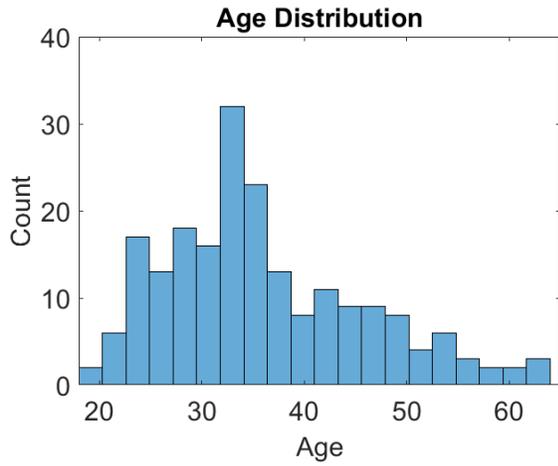


Figure 9: Age distribution of MTurk subjects (19 to 63 years)

Figure 10: Hours driven distribution of MTurk subjects (0 to 32 hours, one outlier excluded)

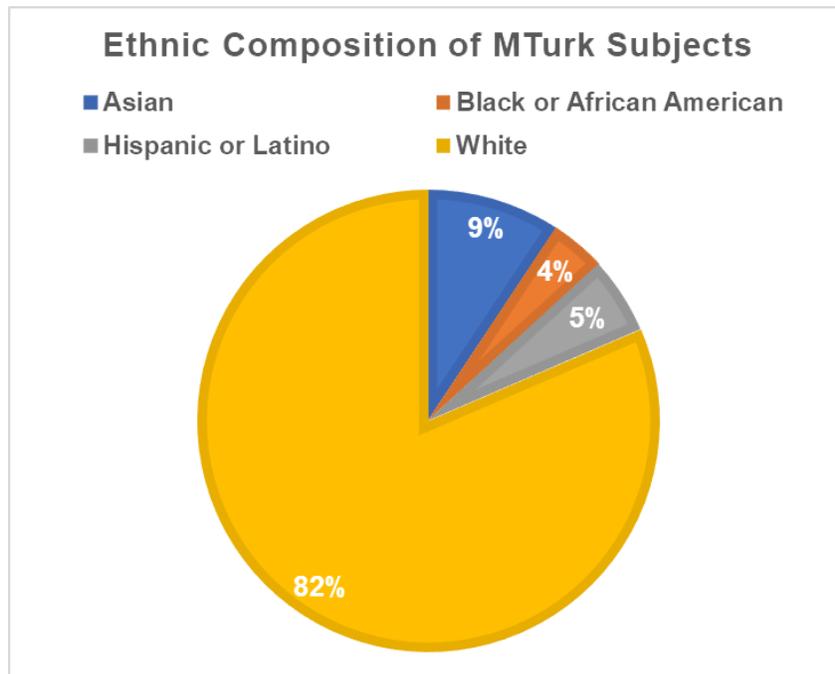


Figure 11: Ethnic composition of MTurk subjects

Table 1: Demographic and risk characteristics of Lab and MTurk subjects

Date	Setting	Treatment	Avg. Age	M:F	Play Games	USA License	Driven Past Year	Avg. Weekly Driving Hours (>0)	Seen VMS	Use Traffic Info	Risk Averse	Risk Neutral	Risk Loving
11/3/2015	Lab	Qualitative	19.9	9:30	56%	77%	82%	8	87%	77%	33%	49%	18%
3/17/2016	Lab	Qualitative, Recommend.	20.4	22:16	55%	84%	89%	7.6	84%	68%	26%	53%	21%
10/8/2015	Lab	Dynamic Description	20.1	14:24	47%	84%	89%	7.1	100%	71%	42%	45%	13%
3/3/2016	Lab	Qualitative, Outline	20.9	16:23	62%	72%	79%	8	85%	56%	41%	31%	28%
3/4/2016	Lab	Qualitative, Dynamic Toll	20.1	13:26	59%	82%	82%	6.4	82%	72%	41%	38%	21%
10/22/2015	Lab	Qualitative, Numeric ID	19.9	14:23	57%	73%	86%	7	86%	65%	43%	49%	8%
6/6/2017	MTurk	Qualitative	37.2	27:9	75%	100%	97%	6.9	100%	53%	56%	36%	8%
6/7/2017	MTurk	Qualitative, Recommend.	37.6	20:17	84%	100%	100%	8.0	100%	81%	46%	43%	11%
6/13/2017	MTurk	Dynamic Description	36.3	19:16	77%	97%	97%	9.6	100%	57%	60%	29%	11%
6/30/2017	MTurk	Qualitative, Outline	35.4	25:10	91%	100%	100%	8.1	100%	57%	54%	37%	9%
6/30/2017	MTurk	Qualitative, Dynamic Toll	33.4	21:15	86%	100%	100%	9.3	100%	67%	47%	39%	14%
6/30/2017	MTurk	Qualitative, Numeric ID	35.6	18:8	81%	96%	96%	8.3	100%	50%	50%	31%	19%

4.2.2 Treatment Type: Standard Qualitative Messaging

In the Standard Qualitative Messaging treatments, a static standard description of incident severity was displayed for each of the four types of traffic incidents that may occur on the typical roadway. Two variations were tested: one with route usage recommendations based on the incident severity and one without. Overall, MTurk and lab subjects made remarkably similar route choices. As the experiment progressed and more severe delays were experienced, MTurk subjects were more likely to divert to the alternate route than their lab counterparts were during incidents of “medium” or greater severity. This is supported by the statistical significance of the interactive terms between the VMS intensity and MTurk treatment dummies in the route choice regression model. It is conceivable that MTurk subjects, with their greater degree of driving experience, hold more pessimistic regarding the delay imposed by incidents than their lab subject

counterparts do.

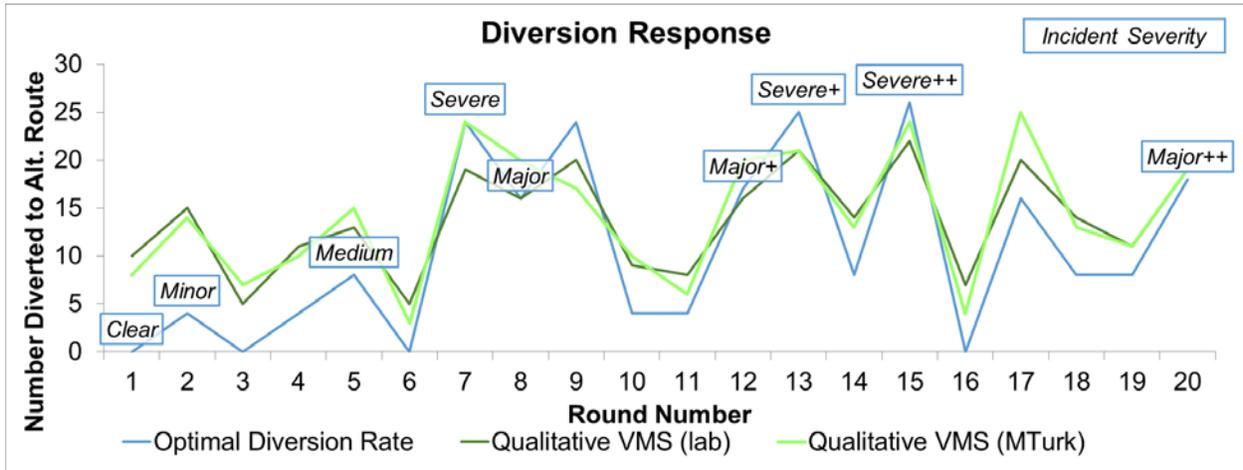


Figure 12: Diversion response of standard qualitative VMS treatments

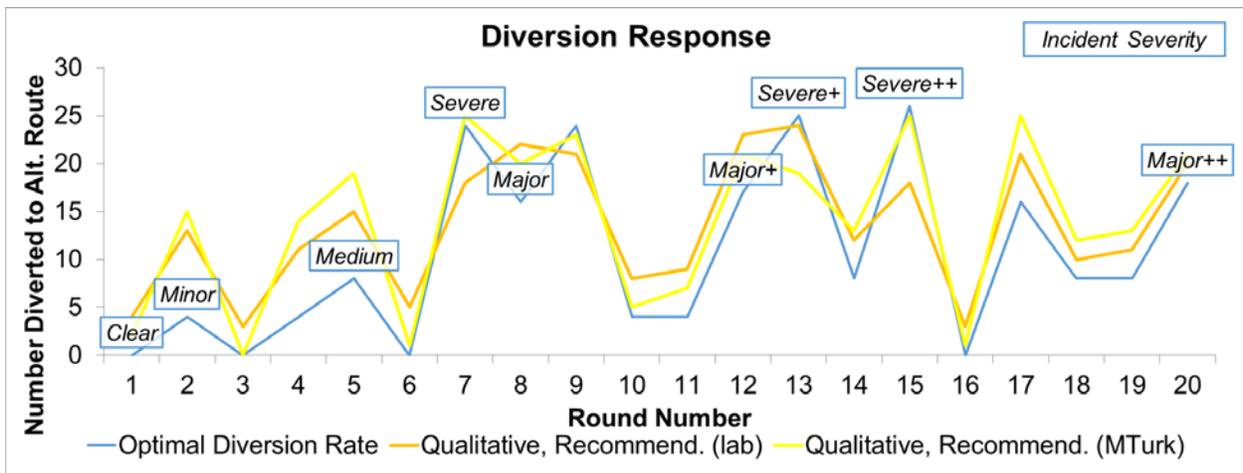


Figure 13: Diversion response of qualitative VMS with diversion recommendation treatments

Qualitative VMS, Qualitative w/ Recommendations			
Lab and MTurk Pooled, 4 Treatments			
Variable	Meaning	Value	tStat
(Intercept)		-1.981	-8.711
risk = 1	Risk Neutral	-0.068	-0.726
risk = 2	Risk Loving	0.085	0.762
lane = 1	Middle Lane	0.077	0.764
lane = 2	Right Lane	0.316	3.170
mturk	MTurk Treatment	-0.403	-1.502
VMS = 1	VMS Intensity 1, "Minor"	0.870	4.139
VMS = 2	VMS Intensity 2, "Medium"	1.121	5.427
VMS = 3	VMS Intensity 3, "Major"	1.846	8.898
VMS = 4	VMS Intensity 4, "Severe"	1.924	9.269
VMS = 1 * mturk	VMS Intensity 1 * MTurk	0.503	1.563
VMS = 2 * mturk	VMS Intensity 2 * Mturk	0.691	2.189
VMS = 3 * mturk	VMS Intensity 3 * Mturk	0.858	2.687
VMS = 4 * mturk	VMS Intensity 4 * Mturk	0.880	2.750
DIV_1	Diversion Recommend. 1, "Alt. rte. available"	0.469	1.837
DIV_2	Diversion Recommend. 2, "Use alt. rte."	-0.022	-0.086
DIV_1 * mturk	Diversion Recommend. 1 * Mturk	-0.703	-1.949
DIV_2 * mturk	Diversion Recommend. 2 * Mturk	0.134	0.372
seenvms = 'Yes'	Seen VMS on Hwy	0.173	1.338
drivehours	Hours driven per week	-0.006	-0.884
gender = 'Female'	Female	-0.017	-0.189
playgames = 'Yes'	Plays Video Games	-0.219	-2.393

Table 2: Logit regression model of *decision to divert* for qualitative VMS treatments

4.2.3 Treatment Type: Dynamic Qualitative Messaging

In the Dynamic Qualitative Messaging treatments, information was displayed and dynamically updated in relation to the anticipated route utilization. In the lab, a version was tested that displayed and dynamically updated the optimal diversion rate. On MTurk, a refined version was tested that displayed a dynamically updated standard description of expected travel delay. In this comparison, MTurk subjects performed better than lab subjects and particularly rounds with severe incidents. The dynamic description of travel delay scheme was considered superior in performance and more practical in implementation than the dynamic diversion rate scheme.

However, this treatment also illustrates a potential issue with the distinguishability of adjectives when using verbal descriptions to inform subjects. In the combined statistical analysis of MTurk sessions using standard descriptions of incident severity (Table 3), it was discovered that while there are significant increases in likelihood to divert when the incident is described as “Medium” as opposed to “Minor” or “Major” as opposed to “Medium”, there is little difference in effect between incidents described as “Major” instead of “Severe”. Using the strongest verbiage of “All cars should exit” yields a more noticeable gain in diversion response. It is possible that the terms “Major” and “Severe” are simply too synonymous for subjects to distinguish without further context. Another possibility is that subjects may have associated an adjective with the number of lanes blocked during an incident, and were thus confused when they observe that “Major” and “Severe” incidents in which both had 3 lanes blocked. However, the observations more strongly support the former theory, as the effect was still statistically significant in the dynamic description treatment where multiple adjectives may have been shown during the same type of incident.

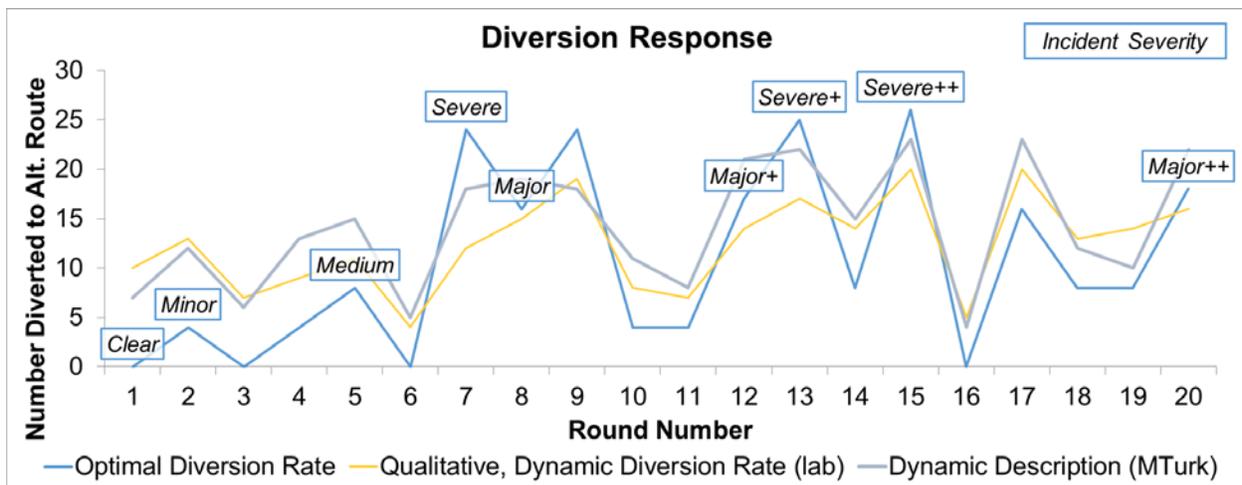


Figure 14: Diversion response of dynamic VMS treatments

Qualitative Description, Qualitative w/ Recommend., Dynamic Description			
3 MTurk Treatments			
Variable	Meaning	Value	tStat
(Intercept)		-2.470	-10.742
risk = 1	Risk Neutral	-0.163	-1.548
risk = 2	Risk Loving	0.238	1.450
lane = 1	Middle Lane	0.112	0.957
lane = 2	Right Lane	0.133	1.135
VMS = 1	VMS Intensity 1, "Minor"	0.895	5.606
VMS = 2	VMS Intensity 2, "Medium"	1.405	8.716
VMS = 3	VMS Intensity 3, "Major"	2.132	12.815
VMS = 4	VMS Intensity 4, "Severe"	2.184	13.244
VMS = 5	VMS Intensity 5, "All cars should exit"	2.308	3.917
DIV = 1	Diversion Recommend. 1, "Alt. rte. available"	-0.257	-1.027
DIV = 2	Diversion Recommend. 2, "Use alt. rte."	0.085	0.333
seenvms = 'Yes'	Seen VMS on Hwy	0.288	1.847
drivehours	Hours Driven per Week	0.022	3.486
gender = 'Female'	Female	0.226	2.269
playgames = 'Yes'	Plays Video Games	0.201	1.654

Table 3: Logit regression model of *decision to divert* for treatments with qualitative descriptions

4.2.4 Treatment Type: Individually Targeted Messaging

In the Individually Targeted Messaging treatments, VMS was utilized to display messages that targeted individual driver characteristics to influence the route choice of subsets of drivers. Two variations were tested: one in which driver vehicles were assigned publicly visible numeric identifiers and one in which driver vehicles were selectively outlined with a bright green color. VMS messages then instructed a range of driver IDs or the green outlined vehicles to divert to the alternate route based on the optimal diversion rate for a given round. The numeric ID scheme was analogous to messaging addressed to drivers based on license plate number, while the color outline scheme was designed to be a test of driver compliance under idealized conditions. In both cases, the targeted diversion instructions were displayed simultaneously with a standard description of incident severity. For the numeric ID treatments, MTurk subjects outperformed their lab counterparts after a few initial rounds of learning. For the color outline

treatments, lab and MTurk subjects performed very similarly across all rounds. In either case, some consistent trends emerge with respect to driver compliance (Table 4, Figure 17, Figure 18). MTurk subjects were significantly more likely to comply with instructions to divert than their lab counterparts were. This could be attributed to the older MTurk drivers taking the instructions more seriously or having more trust in the system operator than the younger student subjects. It is also noted that the less risk averse the subject was, the less likely they would be to comply with instructions. The subjects who reportedly played video games regularly were also less likely to comply with instructions. This could be attributed to a greater willingness to engage in exploratory behavior among these subjects.

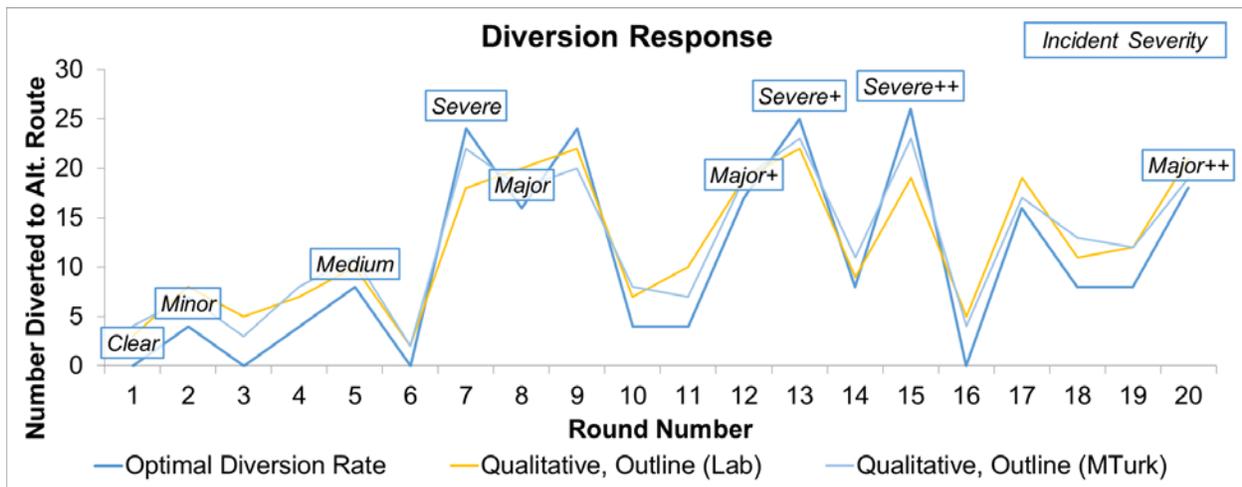


Figure 15: Diversion response of targeted messaging using outline treatments

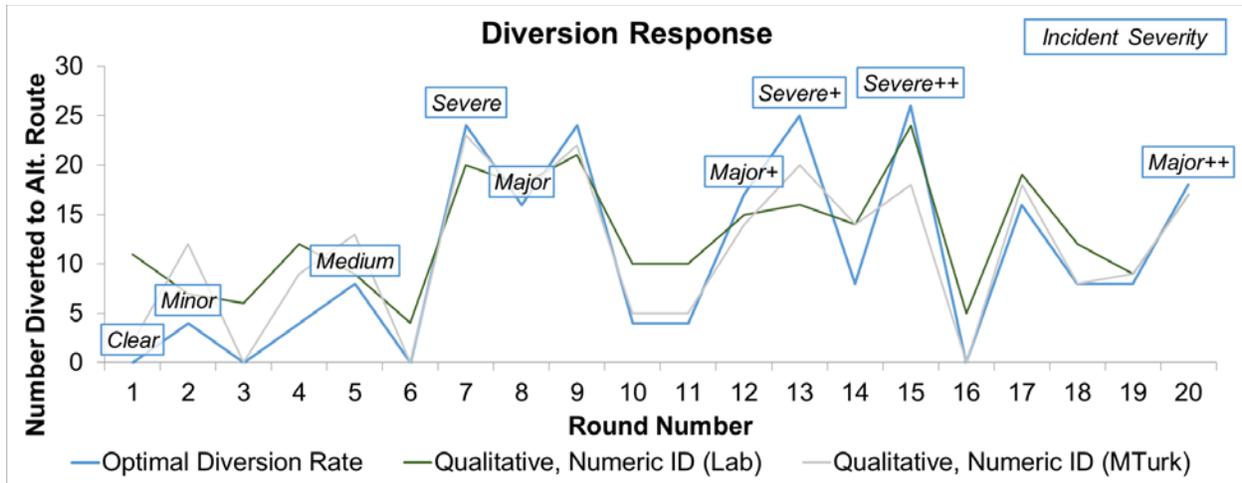


Figure 16: Diversion response of targeted messaging using numeric ID treatments

Variable	Meaning	Numeric ID Lab & MTurk		Outline Lab & MTurk	
		Value	tStat	Value	tStat
(Intercept)		1.259	3.612	2.034	4.624
mturk	MTurk Treatment	0.567	3.459	0.753	3.985
risk = 1	Risk Neutral	-0.222	-1.224	-0.398	-2.071
risk = 2	Risk Loving	-0.481	-2.376	-0.445	-2.002
lane = 1	Middle Lane	0.206	1.169	-0.059	-0.309
lane = 2	Right Lane	0.149	0.855	-0.214	-1.126
scenario = 2	2 Lanes Blocked	-0.125	-0.611	-0.104	-0.449
scenario = 3	3 Lanes Blocked, Short	-0.099	-0.482	-0.344	-1.536
scenario = 4	3 Lanes Blocked, Long	-0.168	-0.826	-0.515	-2.346
gender = 'Female'	Female	-0.135	-0.851	0.807	4.218
playgames = 'Yes'	Plays Video Games	-0.401	-2.286	-0.387	-1.727
seenvms = 'Yes'	Seen VMS on Hwy	-0.062	-0.243	-0.048	-0.142
drivehours	Hours Driven per Week	0.010	0.710	-0.039	-3.446

Table 4: Logit regression model of decision to comply with VMS for ID and outline treatments

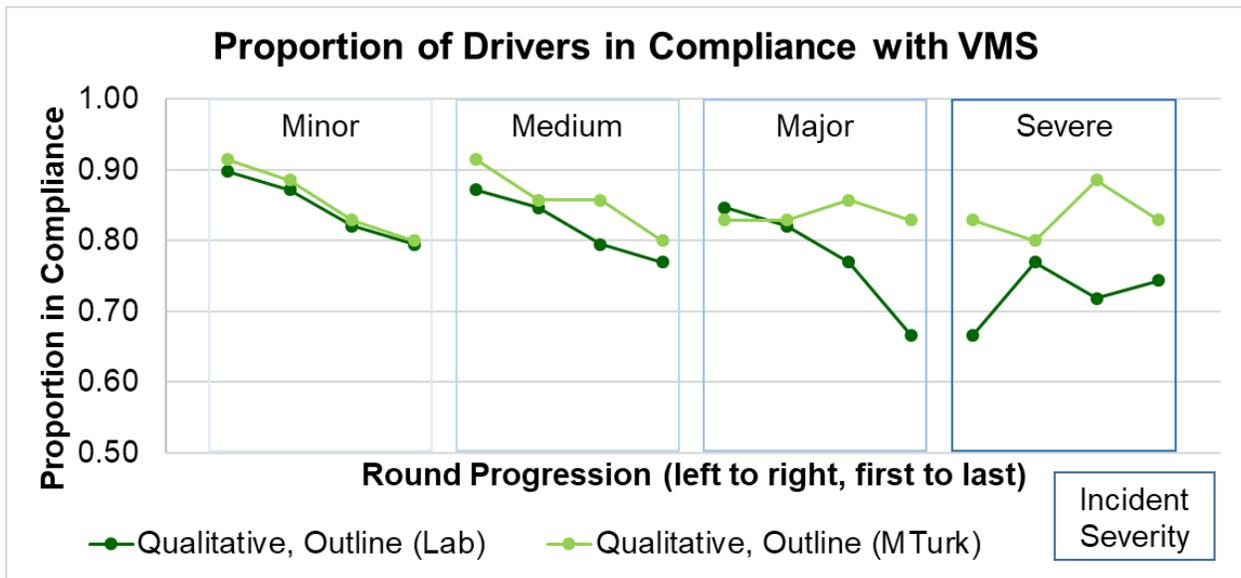


Figure 17: Compliance per incident scenario over time for outline treatments

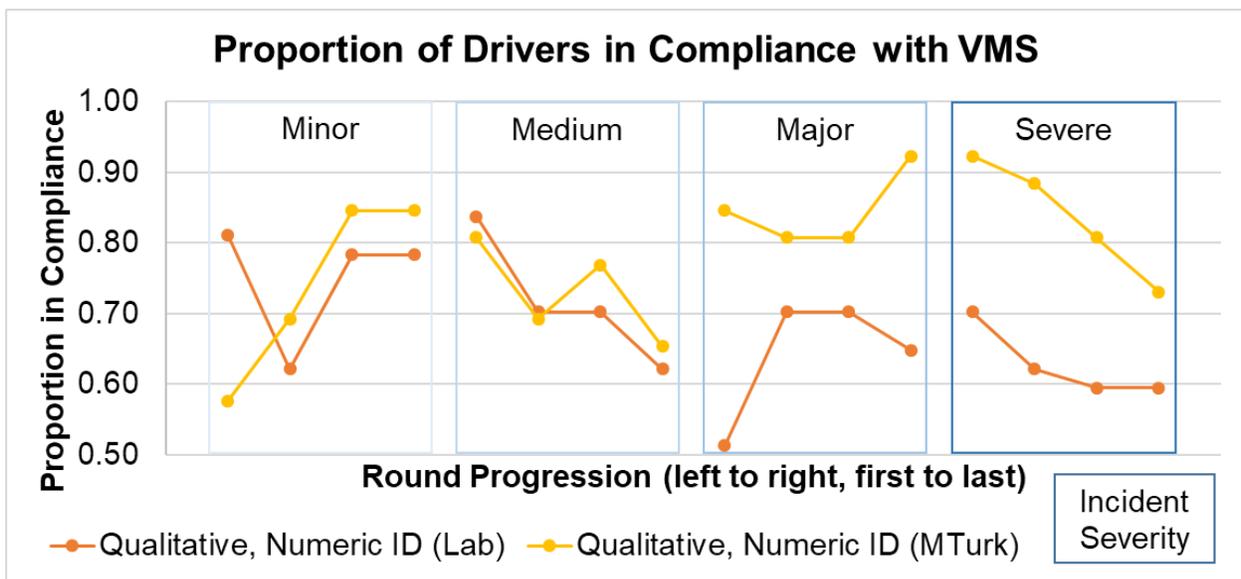


Figure 18: Compliance per incident scenario over time for numeric ID treatments

4.2.5 Treatment Type: Dynamic Tolling with Qualitative Messaging

In the Dynamic Tolling with Qualitative Messaging treatments, a standard description of incident severity was displayed and the main route was dynamically priced (the alternate route is

free to use) by charging a toll or paying a subsidy to drivers depending on the current route utilization. These tolls and subsidies were designed to be used heuristically to nudge drivers towards optimal route utilization. In the lab, suboptimal results were observed, where driver decision-making did not respond significantly to the price on the main route. For the MTurk treatment, the thresholds were adjusted to which prices changed to try to react more quickly to the current state of route usage as drivers passed the decision point. Unfortunately, there were no significant performance gains from this modification and as suspected that the amplitude of the price level needed to be increased to drive a greater response.

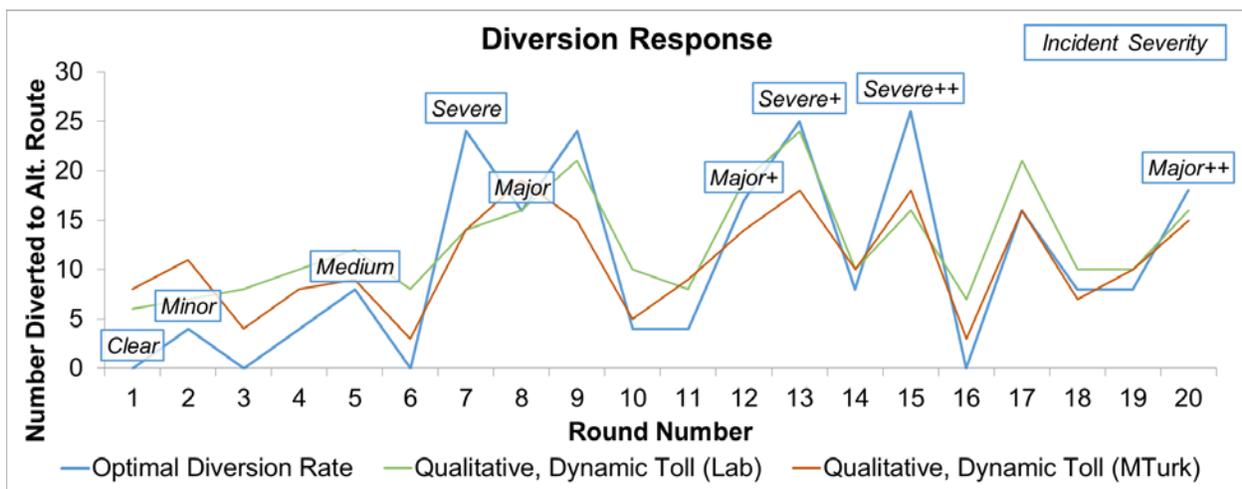


Figure 19: Diversion response of dynamic road pricing treatments

4.3 Concluding Remarks

The online replication of the laboratory experiment via the Amazon Mechanical Turk marketplace was a success that both reinforces the validity of the results in the lab as well as provides new insights with respect to driver route choice as a function of driver experience. In each tested treatment, there was a striking overall resemblance in route choice between subject groups across experiment rounds. Based on the survey questions, the MTurk subjects represented

older, more experienced drivers from across the US. They tended to respond more strongly to the descriptions of incident severity and were more likely to comply with the targeted instructions. Apart from these differences, their driving behavior in the simulator closely resembled that of the student subject group in the lab. Although MTurk subjects were more likely to lose focus on the driving task, and start to idle during driving rounds, these subjects remained in the minority, and only needed to remove one or two subjects from the experiment due to inactivity. Increased subject distraction was inevitable when the experiment was moved from a controlled laboratory setting to remotely administered sessions where there is no direct supervision or control over the subject's surroundings. Overall, this was manageable and did not detract significantly from the benefits of crowdsourced research.

After completing this project, running experiments online with large groups of subjects who interacted with each other in real-time learned to be a daunting task. During this process, significant time was spent (and effort) dealing with two major issues: coordination issues when starting an experiment sessions and software problems stemming from subject variability. Coordination issues at the start of sessions arose from too many workers competing for a spot in the first-come-first-serve session. Some MTurk workers would use browser scripts to automatically accept new available HIT, but do not plan to actually work on the tasks they accepted for a long time. This causes the HIT acceptance to stall and delay the experiment from start. This issue was dealt with by posting far more HIT than the number of workers that could be accommodated in an experiment and paying a "show-up fee" to excess workers who accepted the HIT. For future online experiments, it is recommended to switch to an RSVP based system to 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.

With careful planning and consideration for expected pitfalls, the MTurk marketplace and other online subject pools can be useful tools for conducting experimental research. It will allow a sample from subjects that would be otherwise impossible, impractical, or too expensive to recruit in the laboratory setting. Key to successfully running experiments online is maintaining a good relationship with the subject pool through clear communication and providing reasonable monetary compensation for subjects’ time and effort.

5. Lane Pricing

Priced lanes, such as those in Orange County, Los Angeles, and San Diego, can improve the efficiency of freeways by sorting drivers according to their travel time preferences (Small and Yan 2001, Arnott et al. 2002). Users who more highly value travel time savings were able to pay to access a lane(s) with faster service, while users with a lower VOT pay less and travel in a slower lane(s) (Small and Yan 2001). A commonly used pricing regime, known as “second-best” tolling, imposes a single toll on one set of lanes and leaves the remaining lanes un-tolled (Small and Yan 2001, Verhoef and Small 1999, Arnott et al. 2002). The toll is set to maximize social welfare of drivers, subject to the constraint of leaving some lanes free; users with a higher VOT will pay the toll and get a faster trip, while lower VOT users choose the free lanes and get a

slower trip. By providing extra time savings to those who value them the most highly, at the expense of slower service for those who do not, second best pricing is able to improve driver welfare relative to a freeway with all lanes untolled (Small and Yan 2001, Verhoef and Small 1999, Arnott et al. 2002).

The extent to which welfare benefits can be achieved by second-best pricing depends in part, on what planners know the correct distribution of the VOT of freeway users (Verhoef and Small 1999). If the actual distribution is different from what is assumed, the price set for the toll lane will result in a less efficient outcome. For example, if the true VOT distribution skewed more towards a high VOT compared to what was believed, the toll would be set too low, because it would be more efficient to have fewer vehicles in the toll lane so that the users with a very high VOT could travel even faster. Conversely, if the true VOT distribution were flatter than what was believed, the toll would be set too high because a more highly utilized toll lane with only slightly faster speeds, than the rest of the freeway, would be more efficient.

Unfortunately, estimating the VOT distribution can be quite difficult. VOT can vary dramatically across income groups, vehicle occupancies, household composition, and trip purposes (Koppelman 2012, Wardman 1997). Even among specific individuals, VOT can change from day to day (Parkany 1999, Sullivan 1998). Only limited survey and field data are available for determining a VOT distribution for a particular corridor served by a high-occupancy toll (HOT) lane, and VOT estimates can vary significantly for a specific locale based on data collection and model specification (Koppelman 2012). Adding to the challenge is the fact that the distribution of a VOT has been known to vary by time of day (Koppelman 2012), such as on the HOT lanes on State Route (SR) 91 in Orange County (Sullivan), and day of week (Koppelman 2012). Therefore, an accurate estimate of a VOT distribution would need to vary

by day and hour in each given locale. Furthermore, VOT and corresponding parameter estimates, from which they could be forecasted, vary from region to region (Koppelman 2012). A significant amount of effort to determine a VOT distribution is required anytime HOT lanes are introduced to a new area. Lastly, VOT has been shown to vary based on congestion levels (Koppelman 2012). This means that a typical VOT distribution may be invalid for HOT lane pricing during periods when non-recurring incidents or events result in local congestion; posing an additional challenge for incident management using pricing.

An alternative pricing scheme that uses Vickrey-Clarke-Groves (VCG) mechanism to determine tolls and lane assignments can obviate the need to estimate a VOT distribution by eliciting a VOT directly from subjects. This is a well-known mechanism used for efficiently allocating resources, but has not yet been applied to a transportation setting. The mechanism requires that freeway users remotely (through mobile or in-vehicle communications) provide the toll system with their VOT before reaching the toll lanes. The VOT they provide specifies how much money (in dollars per hour, rounded to the nearest dollar) they are willing to pay for travel time savings. Based on the reported VOT, drivers are assigned to the toll or free lanes in a way that maximizes welfare by minimizing the aggregate cost of travel delay. The drivers with the highest VOT will be assigned to the toll lanes, while those with the lowest VOT will be assigned to free lanes. The lane assignments will favor faster speeds in the toll lanes, so that time savings provided to those who value them most. In the first (best case) scenario, the toll would be set as the externality that each driver imposes on other drivers due to the impact of his/her bid on the lane allocation. In this case there would fast/slow lanes rather than toll/free lanes. This type of toll incentivizes each driver to accurately provide the system with their true VOT. If a driver overstates his/her VOT, this externality – and therefore the cost of the toll - will be greater than

the value of the driver's time savings. Conversely, if a user gives a VOT that is too low, they will likely forgo travel time savings that would cost less than what he/she values them. No user is ever able to benefit by incorrectly specifying his/her VOT. In the best case, the subject would have been just as well off had they provided their true VOT.

As second-best pricing is often more politically desirable, the mechanism can also be modified so that only drivers assigned to the faster lane pay a toll. The toll is still set as the externality that drivers assigned to the toll lane impose on other drivers, which is made up of three components:

1. The congestion cost imposed on other drivers in the fast lane due to an extra user, minus the congestion benefit for drivers in the free lane due to one less user.
2. The delay cost imposed on any marginal drivers switched from the toll lanes to the free lanes due to the VOT the user provided (compared to if that user had provided a VOT of \$0/hour).
3. The congestion cost imposed on other drivers in the slow lane minus the congestion benefit for other fast lane drivers due to any marginal drivers switched from the toll lanes to the free lanes due to the VOT the user provided (compared to if that user had provided a VOT of \$0/hour).

In most cases, a driver assigned to the toll lane displaced a single driver, who is then sent to the free lane. As a result, components 1 and 3 cancel out, which sets the toll equal to component 2 – the delay cost for the marginal driver switched whom to the free lane. This is equivalent to a second-price auction, where each “winner” pays a price equal to the cost to the marginal “loser” for not being able to use the faster toll lane. In light of the similarity of the mechanism to an auction, this report will from here on out refer to a user providing his/her VOT as “*bidding*”.

The second-best modification reduces the efficiency of the mechanism because drivers with a VOT below what would normally get them in the toll lane often have an incentive to raise their VOT. This is because such a driver will benefit if they can inflate their stated VOT and raise the average free lane VOT high enough so that another user is moved from the free lane to the toll lane – while at the same time keeping their bid low enough so that they are not switched to the toll lane.

It is typically not a strong incentive because in most cases drivers would have to raise their VOT so high to induce the switch, which would result in them becoming the ones who are switched. In addition, the prospect of being switched serves as a deterrent against much of the possible bid inflating that could occur. Simulation and experimental evidence from this project show that these perverse incentives have a very small effect on theoretical network efficiency, and an even smaller effect on actual experimental network efficiency.

As drivers approach the location where the freeway is separated into free and toll lanes, they were told through in-vehicle/mobile application which lane they were assigned. Subjects assigned to the toll lanes were charged the toll no matter which lane they chose. This removes the incentive for lower VOT users to increase free lane speeds by drastically overbidding to the point where they are assigned to the toll lane, and then staying in the free lane where they will not be charged. It also alleviates the need for enforcement of free lane assignments, since there is no reason anybody would take the slower free lanes once they were charged for toll lane access. Enforcement would still be required for the toll lanes, however, to ensure that users assigned to the free lane do not take the toll lanes instead.

5.1 Factors that may impact benefits in practice

Despite the theoretical benefits offered by lane management using a VCG mechanism, many possible factors may limit these benefits in practice. These pertain to both mechanism compliance and compatibility of the mechanism with user travel-time preferences.

Mechanism compliance: Mechanism compliance is the significant cognitive demand placed on users to understand the mechanism well-enough to both initially and persistently believe that providing their true VOT give them the best possible outcome. It is possible that subjects simply will not understand the mechanism at all, and will simply bid at random or in accordance with some other flawed heuristic. Even those with a basic understanding of the mechanism must overcome their tendency to experiment with a non-truthful VOT in an attempt to “beat the system”. The greatest test of understanding of the mechanism is that other factors will affect user outcome besides their own bids. Variation in aggregate freeway demand, incidents, and bids from other users are all factors besides a user’s own bids that will affect that user’s lane assignment, the travel times on each set of lanes, and the toll that he/she pays. A user that attributes these variations to his/her own bid will have difficulty “learning” that bidding their true VOT is in fact optimal. They may conclude that over/under-bidding was able to provide a better outcome than truth telling, and as a result be encouraged to continue to or begin providing a non-truthful VOT in the future. Furthermore, as described earlier, constraining the mechanism to leave some lanes free provides low VOT users will slightly inflate their bids. It is unclear whether users will be able to quickly recognize and exploit this perverse incentive. It is also unclear whether drivers will fail to comply with lane assignments, despite strong incentives in place for them to do so.

Compatibility of mechanism with user preferences: Even if subjects “comply” with the mechanism with an earnest attempt to provide their true VOT, subjects may not be able to identify a singular time/money tradeoff that captures their preferences. One simple reason may be that VOT is an abstract concept that is difficult to quantify. In such a case, trial and error would be required to find something that the subject deems satisfactory. Another issue is that some circumstances travel time preferences take a non-linear form. This is most common when travelers are adhering to a fixed schedule. Subjects would be unwilling to spend money to arrive extra early to their destination, but might be willing to spend significant sums of money to arrive less late. Similarly, subjects may incur discrete urgency costs simply for being late – placing a huge premium on being on time, but making them relatively indifferent between arriving one or five minutes late. Lastly, subjects with flexibility over their departure time experience forgone utility from arriving early (could have spent more time on prior activity) or late (less time for new activity), with the minimum cost occurring from a perfect on-time arrival. In each of these cases, subjects would prefer having more control of their travel time, or a lane that provides consistent/reliable travel times, so that they do not pay for time savings, they do not need (or so that they could pay more time savings they need). A shortcoming of this mechanism is that users are unable to guarantee specific quantities of time savings; instead, they only control the ratio at which the system will trade their time for money and money for time. In relevant circumstances, nonlinearities in a VOT will cause users to struggle to specify a singular value that gives the best outcome.

5.2 Literature Review

Welfare benefits of second-best value pricing: The VCG mechanism studies for this project operates on the same principle as a second-best toll that leaves one of more lanes

unpriced, and charges users with a high VOT, a toll to use a faster lane(s). When there is significant VOT heterogeneity, such that travel delay costs for some users are much higher than those of others are, this pricing scheme is able to improve driver welfare over the case where no lanes are tolled (Small and Yan 2000, Verhoef and Small 1999, Arnott et al. 1991, Braid 1996). This pricing scheme also offers better performance than other schemes currently in effect such as profit maximization and keeping the fast lane in free flow (Small and Yan 2000).

Toll setting made difficult due to a need to estimate a driver's VOT: Many studies have identified challenges associated with estimating the VOT distribution needed to set efficient prices for managed lanes. The Second Strategic Highway Research Program (SHRP 2) produced the study "Improving Our Understanding of How Highway Congestion and Pricing Affect Travel Demand" (Koppelman 2012) that describes many of these issues, which are described in the introduction to the Lane Pricing section of this study.

Wardman (2004) analyzed stated preference survey results and found that a traveler's VOT varied by income and trip purpose. Parkany (1999) found that based on revealed preference data for SR-91 users, showed that commuters chose the toll lanes for a varying subset of days of the week; drivers VOT change from day to day. Sullivan (1998) made the same observation based on the same tollway.

Drawback to using VCG to set a single second-best toll: The VCG mechanism can be used to minimize travel delay cost, but does not address all costs associated with private transport. Small, Winston, and Yan (2005) found that drivers also value travel time reliability, which means having a route that provides consistent travel times. If there are costs to a subject arriving

early or late, reliability is important to travelers (Noland and Small 1994). Braid (1996) argued these costs formed a “v-shaped” schedule delay cost function with welfare losses associated with drivers arriving both early (due to opportunity cost of leaving sooner) and late (less time spent at productive activity); he also cited empirical evidence of this functional from Small (1982).

Wardman (2004) also estimated that “late-time” is worth seven times more than typical “in-vehicle” time. Given the possibility for non-linearity in delay costs, setting lane prices based on a single linear VOT parameter may not always be appropriate. Thus, there is an argument to be made that both lane management with pricing and the analysis of lane pricing schemes should account these schedule delay costs. Furthermore, the presence of costs associated with early arrival times implies that drivers not only have preferences regarding travel delay, but also departure time. Wardman (2004) explicitly estimated these preferences, and Arnott et al. (1991) analyzed these preferences in the context of efficient toll setting (which differs from second-best value pricing).

In spite of these theorized and empirically estimated non-linearity in driver time preferences, Sullivan (1998) found that the need to be on time for a commitment (part of the V-shape) was not frequently cited among interviewed toll users as a reason for taking the toll road. Therefore the linear travel delay cost assumption is likely appropriate for many contexts.

Empirical VOT found in field studies: many researchers have estimated VOT of actual freeway drivers. The VOT can serve as a point of reference for comparison to a VOT estimated for experimental subjects in this project. One survey study of California SR-91 found that the VOT of toll users was 13-14 dollars per hour (Sullivan 1998), though this does not reveal much about the distribution of the VOT of freeway users, except that the median would be below

this figure. Small, Winston, and Yan (2005) found from revealed preference surveys of SR-91 users that the VOT distribution of freeway drivers is a normal (bell-shaped) distribution with a mean of \$21/hour and a variance of \$10/hour, while stated preference surveys find a mean of \$12/hour with a variance of \$13/hour. These differ significantly, but both show sufficient driver heterogeneity to justify priced lanes. Parkany (1999) identified some of the demographic factors that contribute to toll-lane usage. Among these, the probability of taking toll-lanes increased with age and with being a female.

5.3 Design/methodology

As part of this project, numerous experimental sessions were conducted to test the extent to which the aforementioned factors affect the theoretical benefits of the VCG mechanism in practice. As with other experiments done under the scope of this project, a unique computerized laboratory experiment was implemented, featuring incentivized real-time decisions on a dynamic interactive network. The network consists of a network with three lanes; eventually these lanes were separated so that one was tolled and two not tolled. Subjects traveled the network repeatedly and over multiple rounds, and they were incentivized to do so quickly. Before travelling the network, drivers specified their VOT. A VCG mechanism was then used to determine which lane each subject was assigned to, and what tolls he or she paid (if any). Any toll that a subject was charged was deducted from their starting endowment; the rest they were free to keep. To capture the dynamics of multi-driver mechanism use, subjects performed the bidding and driving phases together. This made user tolls and travel times a function of not just their own bids, but included those of all other experiment participants.

The experimental sessions were comprised of two broad types of treatments:

- Treatment type 1: Each subject was **assigned** a VOT

- Treatment type 2: Each subject was incentivized by their **true VOT**

For treatment type 1, subjects were randomly assigned a rate of earnings decreased in dollars per second. This determines how much money each subject lost per second of travel time, and was essentially his or her VOT. Subjects assigned a higher VOT had a stronger incentive to take the toll lanes, since each second they spent driving was more costly. The toll system did not “see” an assigned VOT, however, only the VOT that the subject chose to provide. Subjects who bid high enough to take the toll lanes would have had the cost of the toll deducted from their earnings. This tradeoff would have been worth it for subjects assigned a high rate of decrease, but not for those assigned a low one. All subjects left the experiment at the same time (they waited for other subjects at the end of each round), so their only incentive was to keep as much of their initial earnings as possible.

The assigned VOT ranges from 10 cents to 3 dollars per second, which creates significant heterogeneity among subjects. In terms of comparability to real \$/hour VOT, these are extremely high. However, to be able to have repeated rounds over the course of an experimental session, the length of the driving task needed to be short (about a minute long, with about half that time comprised of travel in toll/free lanes). The differentials between the two lanes were typically less than 10 seconds, and thus a rate of earnings decreased by cents/dollars per second were needed to make the truth-telling incentive salient.

By assigning subjects a VOT, it was possible to observe whether they provide their true VOT for the mechanism. It was also possible to observe the cost that they incur for not doing so. This type of treatment serves to answer the following questions:

- **Truth telling:** Do drivers provide a truthful VOT that matches the rate at which their earnings decrease? What factors lead to more (or less) truthful bidding?

- **Welfare:** Do losses in driver welfare due to providing a wrong VOT exceed efficiency losses from a standard toll that guesses at a VOT distribution?
- **Resiliency:** Is a truthful VOT reporting robust to changes in VOT assignment, as well as road condition changes?
- **Learning:** Do drivers learn a truthful VOT reporting?
- **Second-best implementation:** Do subjects exploit the incentive that exists for low-VOT drivers to inflate their bids, since the slow lanes are kept free?
- **Lane assignment:** Do subjects take the set of lanes to which they were assigned?

For half the treatments, subjects VOT changed twice during the experiment. This tested whether subjects understood the mechanism well enough to change their bids accordingly, or whether they simply learned through trial and error. The capacity of the free lanes also changed once during the experiment, which affected tolls and lane allocations. This tested whether subjects had the discipline to keep bidding their true VOT despite shifts in their travel times, tolls, and possibly lane assignments. For the remaining treatments, subjects VOT changed only once and there was no change in capacity. This allowed for better observation of long-term learning. Lastly, the instructions were modified between treatments to test how varying the amount, and type of information about the mechanism to subjects, affected the way that they bid.

For treatment type 2, subjects' earnings did not decrease with time. Instead, subjects left the experiment as they finish their driving tasks. Subjects in these treatments did not travel the freeway together; computer-controlled vehicles took the place of other subjects. This allowed subjects to complete the experiment at their own pace instead of waiting on others each round. As with the prior types of treatment, those who bid high enough to be assigned to the priced lanes would have the toll deducted from their initial endowment. Thus, subjects balanced their

innate desire to leave the experiment sooner (spending less time driving) against their desire to not spend money on tolls. In this case, a subject VOT was however much money they were willing to spend to leave sooner. Subjects who were in more of a hurry to leave the experiment would find this tradeoff worth it, while those who were in less of a hurry would not. Although subjects did not travel the road together, in many treatments the mechanism uses the bids from all subjects (even if they are not sharing the same roadway) to determine tolls and lane assignments. In a sense, each user's VOT was "assigned" to computer-controlled vehicles sharing the road with every other user. This preserves the dynamic aspect of bidding with the VCG mechanism.

The length of the network was significantly lengthened and the lanes were made significantly more congestible to increase the length of the rounds. This created much more noticeable differences between travel times in the toll lanes and free lanes, and as a result, was suitable for subjects to provide their VOT in dollars per hour rather dollars per second. A subject who provided a VOT of \$0 / hour each round would take about 40 minutes in total to complete every round, while a subject who provided a VOT of \$40 / hour (the max allowed) would take about 10 minutes to complete all rounds.

This type of treatment serves to answer the following questions:

- Did the mechanism elicit a plausible VOT from the subjects?
- Did a subject's elicited VOT fluctuate, drift, or remain stable? (reflects ability of subjects to know their innate VOT)
- Is the elicited VOT robust to changing experimental conditions?
- Do subjects also desire travel time reliability?
- Are users satisfied with the mechanism?

Some treatment types used a first-best VCG that priced all lanes, in order to test effects on bidding and subject satisfaction with the mechanism. Other sessions changed the units of the bidding from dollars per hour to cents per minute to see if bidding was affected. Another variant provided feedback that told subjects how their toll and travel time would have changed if they had chosen a different VOT. This information was shown for a counterfactual VOT both above and below one they reported, with the goal of both reducing the need for bid experimentation and also to help disentangle the effect on a user's outcomes from changing his/her own VOT versus the effect from other subjects changing their VOT. Lastly, some treatments gave subjects the option to pay extra money before departing; this would greatly improve the odds of being assigned to the lane they chose. This accommodated time/money preferences that go beyond a simple linear relationship. Subjects could choose extra time savings, or extra money savings, without having to over-report or under-report their VOT.

For all treatments, each session began with a presentation of instructions that explain the experiment and the VCG mechanism. Subjects then participated in two practice rounds where earnings did not count. The practice rounds familiarized subjects with inputting their VOT, receiving lane assignment instructions, choosing the correct lane, and driving to the end goal. Afterwards subjects began the main rounds, where the same process was repeated either 10 or 13 times, depending on the treatment. Lastly, subjects complete a post-experiment questionnaire. This survey solicits subject demographics, explanations for how bids were chosen, feedback about the mechanism, and other lifestyle information that may have affected how subjects bid.

5.4 Results

5.4.1 Results for treatments with an assigned VOT

A straightforward finding from the experiments was that subjects had minimal issues with complying with lane assignments.

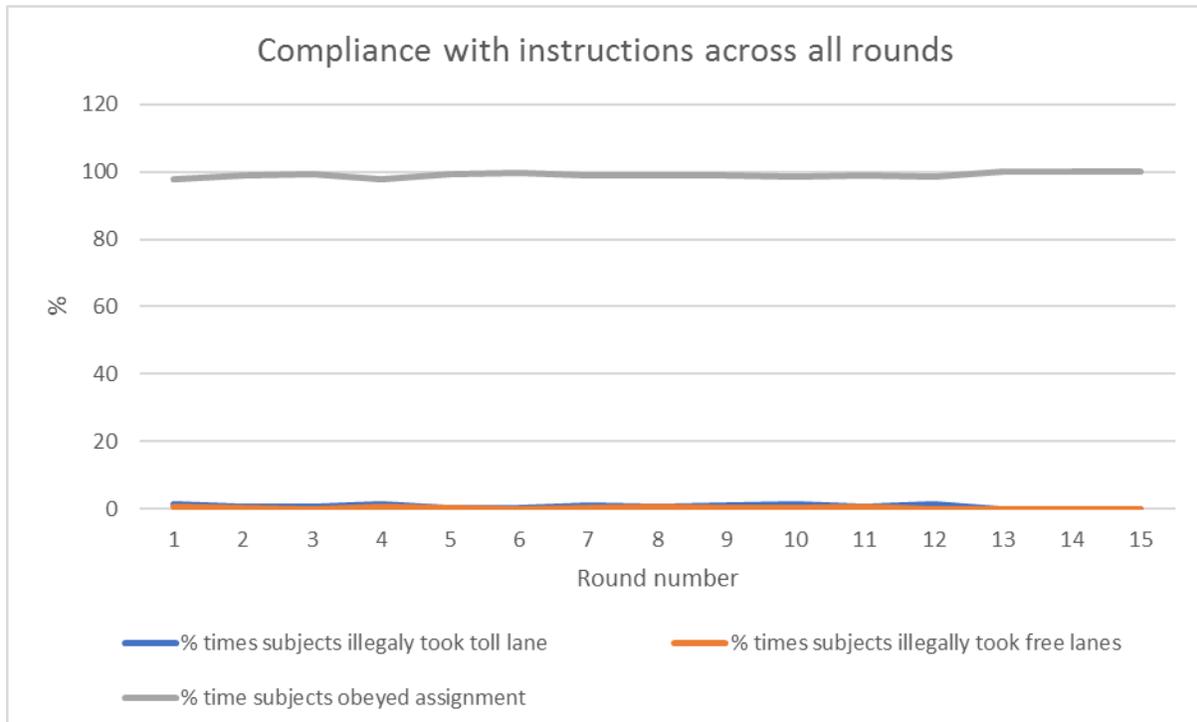


Figure 20: Subject compliance with lane assignments

Non-compliance occurred one percent of the time, decreasing as the experiment progressed. As expected, it was more common for subjects to take the toll lane when assigned to the free lanes, rather than vice versa. While most compliance was probably due to subjects' attempts to exploit the system (since it occurred more at the beginning of the sessions), some non-compliance was likely due to subjects experiencing difficulty moving into the correct lane (which many subjects stated in the post experiment questionnaire).

One of the most significant findings from the assigned-VOT treatments was that subjects frequently did not bid their true VOT. The differences between true and stated VOT varied

depending on the scenario examined. Three different scenarios were examined for each experimental session; each scenario is comprised of a specific subset of rounds intended to assess the results from a different perspective. The three different sets of rounds are summarized in the chart below, and then described in the following paragraph.

Scenario	Number of VOT re-assignments	Number of road capacity changes	Comparable between different experiment sessions?
First four rounds	0	0	Yes
Seven rounds	1	0	Yes
Entire session	1 or 2	0 or 1	No

Table 5: Summary of Differences between the Various Sets of Rounds That Were Examined

For the **first four rounds**, subjects were given a unique VOT (payment depreciation rate) which did not change. This was the most straightforward scenario for testing the mechanism. It gives pessimistic perspective on misbidding because it includes the very first rounds (giving subjects less time to learn), but an optimistic perspective because there were no changes to VOT or road conditions. The first four rounds were comparable between all sessions. The **seven rounds scenario** included the first four rounds, and the first three rounds that occurred after the first time that each subject's VOT was reassigned. This scenario incorporated a measure of how well subjects truly understood the mechanism, since subjects who learned to bid truthfully through simple reinforcement learning rather than mechanism understanding would likely fail to adjust their bids in response to a change in VOT. This scenario was also comparable between all experimental sessions. The **full session** scenario includes every round run for a given session. It included multiple VOT reassignments and (for some treatments) a

change in the freeway congestion levels towards the end of the session (for some treatments). It allowed for a comprehensive look at subject performance, but was not comparable between sessions because after the first VOT shift, the number of rounds, VOT shifts, and congestion changes were not consistent between the different sessions.

Session*	Quality of the pre-experiment explanation of the following topics:			Magnitude difference between bids and actual VOTs for the following sets of rounds (in \$/second):		
	Mechanism	Truth-telling	VOT	First four rounds, no VOT shifts, (Comparable between sessions)	Seven rounds, one VOT shift. (Comparable between sessions)	Full session, (Not comparable between sessions)
2	Sparse	Sparse	Sparse	\$0.64	\$0.73	\$0.65
3	Detailed	Sparse	Sparse	\$0.74	\$0.76	\$0.75
4	Sparse	Detailed	Sparse	\$0.57	\$0.62	\$0.63
5	Detailed	Sparse	Sparse	\$0.71	\$0.79	\$0.75
6	Detailed	Detailed	Detailed	\$0.56	\$0.63	\$0.62
7	Detailed	Detailed	Detailed	\$0.48	\$0.55	\$0.57

Table 6: Comparison of Subject Misbidding by Session and Scenario

*note: Session 1 was a control, which used a standard toll rather than a VCG mechanism. Thus, it is not included in the misbidding results table.

Misbidding ranged from \$0.48/second to \$0.74/second with no VOT shifts (no changes in subjects' payment depreciation rate), and from \$0.55/second to \$0.79/second when a VOT shift was included. The average possible misbidding was \$3.00, so a difference of \$0.60 constitutes a substantial 20% deviation. It is apparent that the amount of effort dedicated to describing both the importance of accurate VOT reporting and the concept of VOT reduced the amount of misbidding in experimental sessions. This was expected, because the toll system was complicated and additional information could help to reduce the cognitive burden on subjects. Paradoxically, explaining the mechanism in more detail without increasing the detail of the truth telling and VOT aspects of the system increased the amount of misbidding. This was perhaps due to a false confidence in subjects of understanding the mechanism well enough to try to beat the system. Not all misbidding related to understanding of the experiment was relevant to real

practice. Many subjects likely struggled to grasp the concept of an assigned VOT, and failed to internalize the payment depreciation rate as their VOT. Evidence of this comes from the fact that intra-subject variation in bids was much more substantial in assigned VOT treatments than in the real VOT treatments. In other words, subjects experimented more when their VOT was assigned. As assigned VOT has no real world bearing however, and so confusion about the concept would not be a source of mechanism inefficiency in practice.

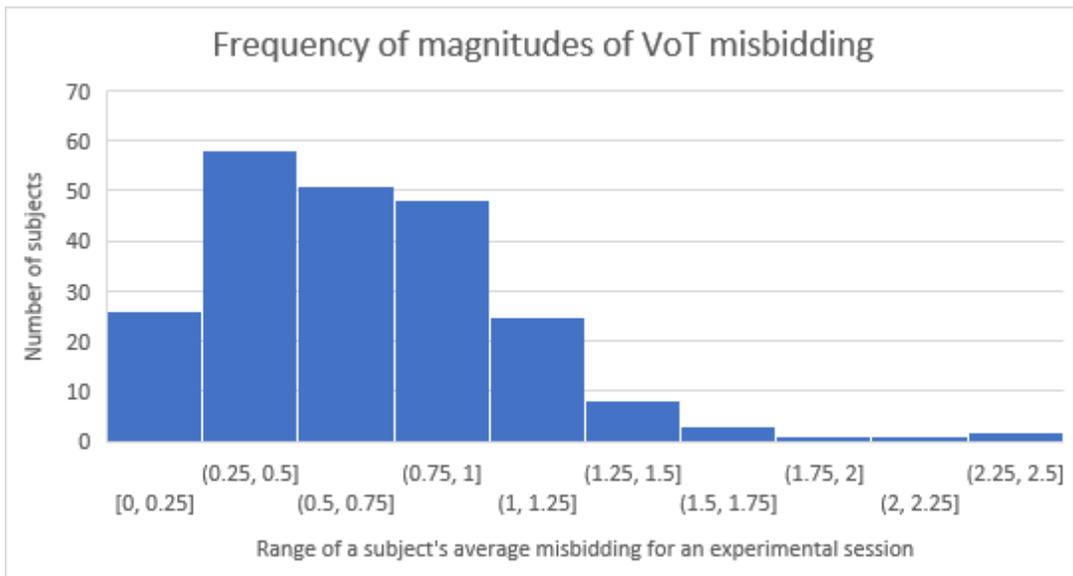


Figure 21: Frequency of different magnitudes of subject misbidding

There was significant variation between subjects for misbidding that occurred. The chart above shows that many subjects bid within 25 cents/second of their true VOT on average, while slightly more misbid by a dollar/second on average. The most common amount of average misbidding was 25 to 50 cents/second.

Many other factors also affected the degree to which subjects provided VOT that matched their payment depreciation rate.

Factor affecting VOT reporting accuracy	Effect on misbidding magnitude (\$/sec.)
Not having a U.S. driver's license	+ 0.12
Never encountered a variable message sign	+ 0.14
Difficult Major	- 0.13

Table 7: Factors affecting subject VOT reporting accuracy

Not having a US driver's license was correlated with a 12 cent/second increase in the magnitude of misbidding. One possible reason for this was that these subjects are international students. If their English proficiency was lower on average than domestic students were, then they might have had a weaker understanding of the pre-experiment explanation of the toll system.

Never having encountered a variable message sign was correlated with a 14 cent/second increase in the magnitude of misbidding. This variable seems to be a proxy for having driven on a freeway before. It was unclear why experience on a freeway would affect the propensity of a subject to provide an accurate VOT.

Regardless of the reasons, both groups would be unlikely to encounter the mechanism in real life. Therefore, omitting subjects with no license who would not drive on freeways yields a more optimistic assessment of the willingness of subjects to provide their true VOT. Doing so changes the range of the magnitude of the different between true and stated to \$0.22/second to \$0.48/second for rounds with no VOT shifts, or about 12% on average of the possible magnitude of the error.

Attitude	Effect on misbidding magnitude (\$/sec)
Toll lane was slow	+ 0.27

Table 8: Self-reported subject attitudes associated with misbidding

Strategy	Effect on misbidding magnitude (\$/sec)
Gibberish explanation of bidding	+ 0.23
Always avoided toll	+ 0.31
Trial and error bidding approach	+ 0.29
Bid a multiple of true VOT	+ 0.19
Randomly chose bids	+ 0.41

Table 9: Self-reported bidding strategies associated with misbidding (vs. ideal bidding strategy)

Many attitudes and bidding strategies were significantly correlated with misbidding. Of those, always trying to avoid toll road (regardless of VOT) and randomly choosing bids (no understanding of mechanism) were the most important. The signs of the variables were all in agreement with intuition.

Although an obvious indicator of mechanism performance, the magnitude of the VOT deviations do not tell the full story. Usually a subject misbidding does not result in any changes to his/her toll or travel time unless it causes the subject to be assigned to a different lane than if they had provided a truthful VOT. For example, going from \$2.00 to \$3.00 per second will likely not affect the subject, but going from \$1.00 to \$2.00 per second likely will because the cutoff to take the toll lane was usually a VOT between \$1.00 and \$2.00 per second). Subjects do not have strong incentives to provide their true VOT. The graph below shows the relationship between a subject's average misbidding, and average experiment earnings, after controlling a subject's average payment depreciation rate. This is done because subjects have a slightly higher earnings potential when their payment depreciation rate was higher (it was too difficult to perfectly calibrate the starting endowment and rate of decrease to make payouts equal for everyone).

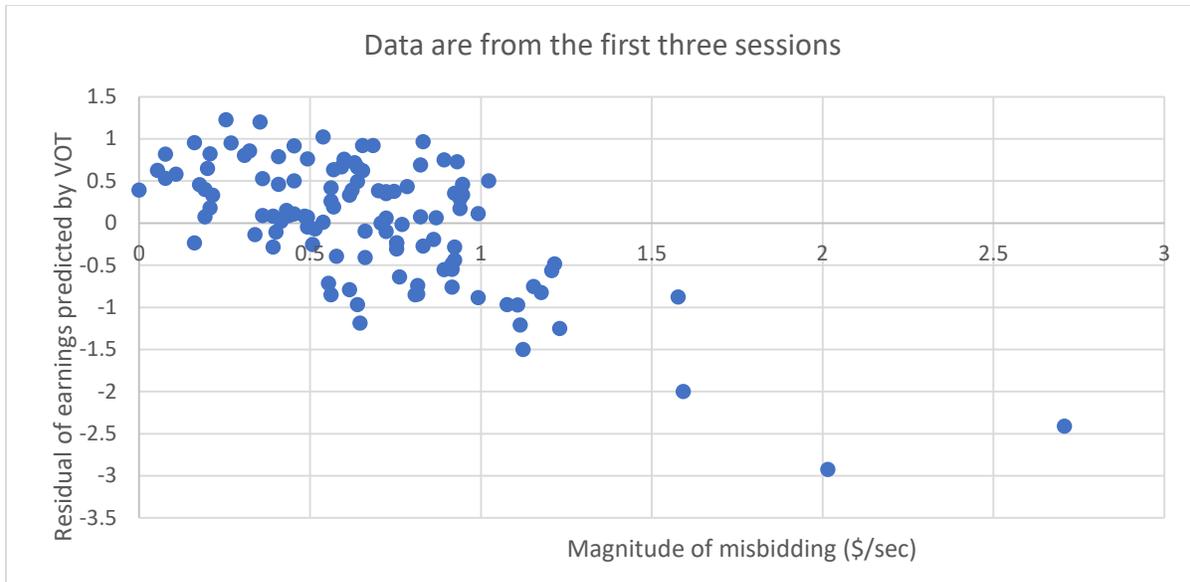


Figure 22: Relationship between misbidding and earnings, controlling for assigned VOT

It was clear that on average subjects would lose more money the more they misbid. However, it was possible for subjects to improve their payout without significantly reducing misbidding as long (as previously described), as long as they could ensure they were assigned to the correct lane. Furthermore, misbidding often cancels out in aggregate. For example, if one person with a VOT of \$0.50 bids \$0.00 /second and another bids \$1.00/second, there will likely be no net effect on the system even though the two subjects misbid by \$0.50/second on average.

Therefore, it is more informative to observe what the aggregate loss from driver welfare due to misbidding was. It was clear from the graph below that the correspondence is not one-to-one, which if true would appear as a straight line.

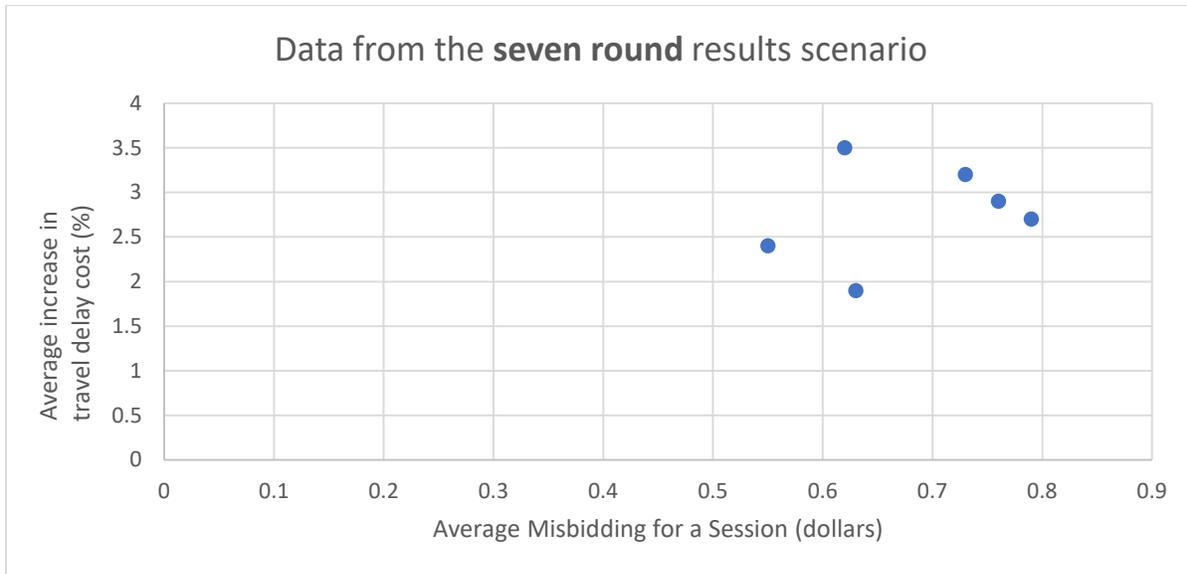


Figure 23: Relationship between misbidding and mechanism efficiency loss

The session with the most amount of misbidding was in the middle of the pack for efficiency loss; demonstrate the need to look at overall driver welfare. The simplest way of doing so was to assume that the only relevant welfare measure was travel delay cost. This was computed as the sum of each driver’s VOT multiplied by the amount of travel time spent reaching his or her destination. Another way welfare could be measured would also consider toll expenditures (for example, did subjects overpay for time savings in some cases?). Incorporating tolls into the analysis was much more complicated, however, because it depends on the way that toll revenue was allocated.

When assessing efficiency loss from aggregate travel delay cost due to misbidding, the system must be looked at as a whole. This was because misbidding can affect not only the subject doing so, but other subjects as well. A single subject misbidding could result in too many / too few drivers assigned to the toll lane (which affects all drivers), or could result in another subject being “bumped” from the toll lane to the free lanes. This could be done by comparing the minimum aggregate travel delay cost possible to the aggregate delay cost actually

achieved by subjects. The minimum travel delay cost was what would have been obtained if the system knew everyone’s exact VOT and assigned everybody to the correct lane. The actual delay cost was determined by the mechanism’s lane assignments based on VOTs reported by subjects. When subjects misbid, the mechanism makes sub-optimal lane assignments resulting in higher travel delay costs.

Session	Quality of the pre-experiment explanation of the following topics:			Efficiency loss from mechanism for the following sets of rounds (measured as % increase in travel delay cost above optimal):		
	Mechanism	Truth-telling	VOT	First four rounds, no VOT shifts, (Comparable between sessions)	Seven rounds, one VOT shift. (Comparable between sessions)	Full session, (Not comparable between sessions)
2	Sparse	Sparse	Sparse	1.8 %	3.2 %	3.2 %
3	Detailed	Sparse	Sparse	2.4 %	2.9 %	3.1 %
4	Sparse	Detailed	Sparse	2.3 %	3.5 %	3.4 %
5	Detailed	Sparse	Sparse	2.4 %	2.7 %	2.4 %
6	Detailed	Detailed	Detailed	1.5 %	1.9 %	1.6 %
7	Detailed	Detailed	Detailed	1.3 %	2.4 %	2.2 %

Table 10: Mechanism efficiency loss by scenario and treatment type

The chart above shows the welfare impacts associated with the VCG mechanism due to subjects misbidding. Depending on the scenario examined, providing incorrect VOT increased the delay costs of drivers on the road anywhere from 1.3 % to 3.5 %. When looking at only the first four rounds, the lowest range of efficiency losses were found (1.3% to 2.4%), while the highest were found for the seven-round sample (1.9% to 3.5%). As with misbidding, the pre-experiment information provided to subjects had an impact on efficiency losses. Losses were lowest when detailed information about the mechanism, importance of bidding truthfully, and concept of VOT were provided.

These efficiency losses are put in better context by comparing them to travel delay costs in the absence of the VCG mechanism. With no pricing mechanism at all, travel delay costs only increase by 3%. This means that for samples of rounds that included VOT shifts, the misbidding was extensive enough to where there was no point in having priced lanes at all. For both sessions with thorough pre-experiment instructions, however, the mechanism was still always beneficial.

The VCG mechanism's performance can be put into further perspective by comparing it to the issue it was intended to fix – efficiency losses with second-best value pricing due operators setting the toll based on incorrect assumptions about a users' VOT. Simulations can be used to estimate these losses based on various assumptions about planners' ability to know the true VOT of drivers. These simulated scenarios were compared to an ideal case, where planners know every driver's VOT perfectly, and then set the ideal second-best value-pricing toll. The simulation assumes that drivers always know to choose the toll if it is worth it to them and choose the correct lane.

Two value-pricing scenarios were considered. In one type of scenario, planners knew the correct distribution of drivers VOT for a certain set of conditions, but did not know the assortment of drivers VOT at any given time. In this scenario, the simulated “planners” know that drivers VOT will come from a uniform distribution between .1 and 2.4 dollars per second, which matches the payment depreciation rates assigned to subjects in the VCG experiment. The second-best toll was set according to the expected value of the distribution, and the “coin flip” is 39 draws from that distribution which determines the actual VOT of simulated drivers. If the sample produces a set subject VOT that perfectly matches the expected assortment, then the ex-ante toll set by planners ends up being optimal, and there is no efficiency loss. If the sample of a

VOT is non-representative of the underlying distribution, however, then the ex-ante toll will be suboptimal; a better toll could have been set had operators known the true set of subjects VOT. To determine the efficiency loss that results from this sampling error, the travel delay cost given the ex-ante toll was compared to the travel delay cost possible with the toll that would have been set had the true subjects VOT been known perfectly. These samples were taken dozens of times and the subsequent efficiency loss calculations were averaged together over all the samples. It is important to note that these efficiency losses from sampling are sensitive to the number of drivers thought to be affected by any one driver's lane assignment. The higher this number was the lower efficiency losses from sampling would have been, simply by virtue of increasing the sample size. The amount of driver interdependency was determined by complex traffic dynamics that were beyond the scope of this project; it was assumed each vehicle affects 38 other vehicles in order to match the dynamic from the experimental sessions (which were constrained to 39 subjects).

In another second-best toll scenario, planners knew the correct shape of the VOT distribution (uniform), but they were incorrect about the mean of that distribution by a fixed percentage. Efficiency losses were higher in this scenario, since it was almost guaranteed that a set of drivers VOT would not match the expected value of the incorrect distribution. Obviously the further off the sample mean was from the true mean, the higher the resulting efficiency loss would be. This scenario did not depend nearly as heavily on the amount of vehicles that interact in the freeway system.

The chart below shows the increase in travel delay loss associated with various scenarios for both the “standard” second-best toll and the VCG mechanism.

Efficiency loss measured as % increase in travel delay cost above optimal				
“Standard” second-best toll for following scenarios:			VCG Mechanism	
VOT guesses from correct distribution	Correct shape of VOT distribution, but mean is off by		Best case	Worst case
	10%	25%	No VOT shift	No VOT shift
1.2 %	1.7 %	2.5 %	1.3 %	2.4%

Table 11: Efficiency loss comparison between for “standard” and VCG tolling

The chart shows that even in the best case, the VCG mechanism is less efficient than a second-best toll when operators perfectly know the underlying distribution of subject VOT. This means that the subjects in the VCG experiments misbid to a great enough extent that the mechanism was unable to make up for random variation in a subjects VOT that reduce efficiency in second-best value pricing schemes. If the correct distribution was not known within 10% of the true mean, however, the VCG mechanism could be more efficient than second-best pricing. In this case, increase in travel delay resulting from misbidding with the VCG mechanism does not offset the efficiency loss from planners setting a single toll using second-best pricing. Even when operators set a toll based on a distribution with a mean estimate off by 10%, the VCG mechanism was not always superior. The worst case for the VCG mechanism, where the mechanism was not well explained to subjects – was only superior to the second-best toll when planners estimated the mean incorrectly by 25% of more. Although 25% seems high, however given how difficult estimating a VOT distribution in practice is, this was not an unlikely scenario.

The experimental VCG results would be even more favorable if the welfare calculation was restricted to subjects with driver's licenses. Subjects without a license misbid significantly more than other subjects. These subjects did not represent drivers on the road, and thus should not be included in analysis. These subjects and their outcomes cannot simply be dropped from the analysis, since their bidding also affected the travel times of other drivers.

Learning

These experiments reveal mixed results regarding learning effects for subjects to provide their true VOT. In the first four rounds, before any VOT shifts, the magnitude of misbidding was reduced by only 8%. Efficiency improves slightly more over that interval, but still by only 12%.

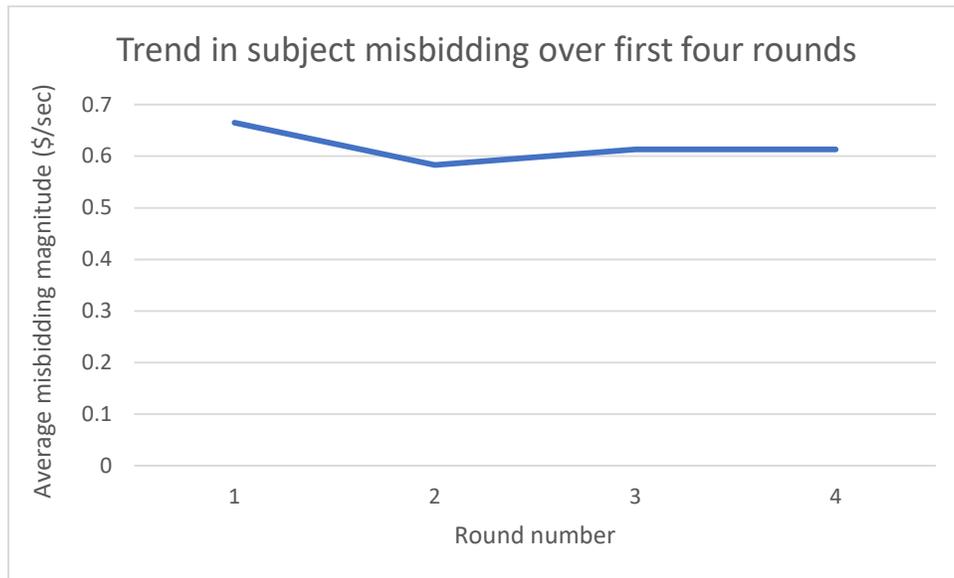


Figure 24: Trend in subject misbidding over the first four rounds (averaged across each treatment)

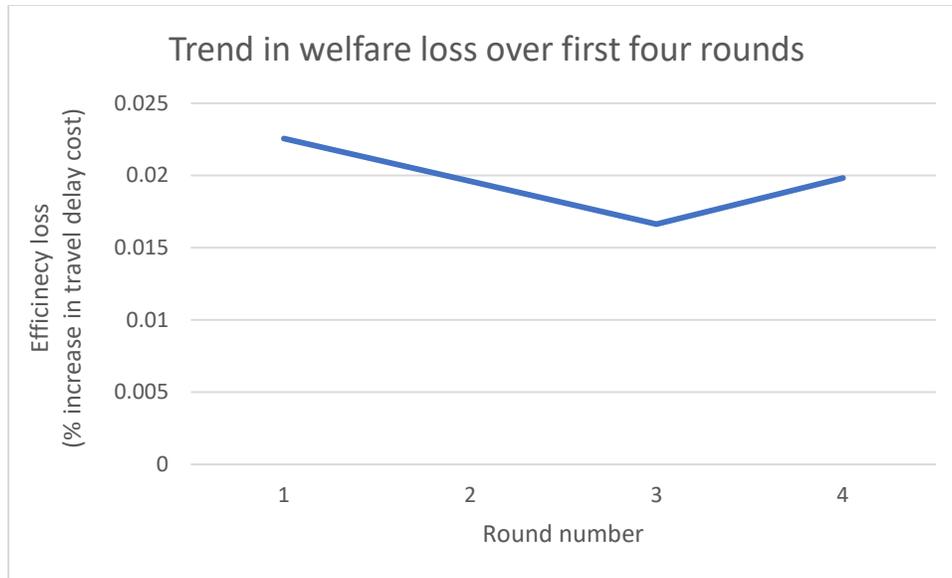


Figure 25: Trend in efficiency loss over the first four rounds (averaged across each treatment)

It was not determined why subjects failed to learn to bid their true VOT at a faster rate. One possibility is that many subjects never understood that choosing their payment depreciate rate as their VOT was the best strategy for them, and instead experimented with their bid throughout the round to try to find the best outcome. Due to the nature of the mechanism and the driving dynamics in the experiment, reinforcement learning of this kind would have been difficult for subjects. Changes in the bids of other subjects, as well as random slowdowns that occurred to subjects' switches, created stochasticity in the tolls and travel times experienced by subjects each round. The resulting fluctuations in subjects' payouts created a "noisy signal" that obscured the relationship between a subject's bid and his/her payout in a given round. In some cases it appeared as though some subjects learned to misbid after instances of misbidding coincided with favorable lane switching dynamics that gave them a better payout. This conjecture was supported by results from an experimental session when lane assignments were given to subjects sooner, in order to reduce the extent to which subjects interfered with one another in trying to change lanes. This in turn reduced the stochasticity in travel times

experienced by subjects each round. As a result, learning over the first four rounds occurred more quickly.

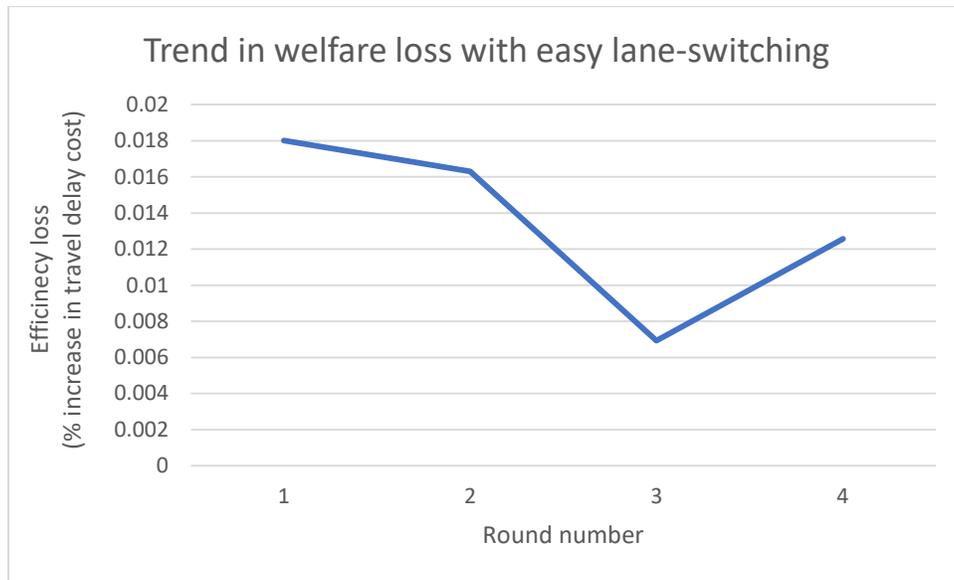


Figure 26: Trend in efficiency loss over the first four rounds of treatment with easy lane switching

Over the first four rounds, efficiency losses were reduced by 30% when travel times varied less due to lane switching issues. The discrepancy between learning with and without costly lane changes highlights a unique challenge posed by transportation applications of this mechanism. Random variability in travel times due to arbitrary non-congestion slowdowns can obscure the relationships between a subject's bid and outcome. However, this issue will be less prevalent in the real world than the experiment, because the experiment has a higher VOT than what is realistic. Therefore, each random slowdown was more costly in the experiment.

Learning plays a much more prominent role in subject bidding behavior once experimental conditions change. When subjects are assigned a new VOT and when the capacity of the free lanes change, misbidding temporarily increases significantly. Subjects quickly learn

to bid closer to their true VOT after this initial shock. The graph below illustrates learning after the first time a subjects' VOT were reassigned.

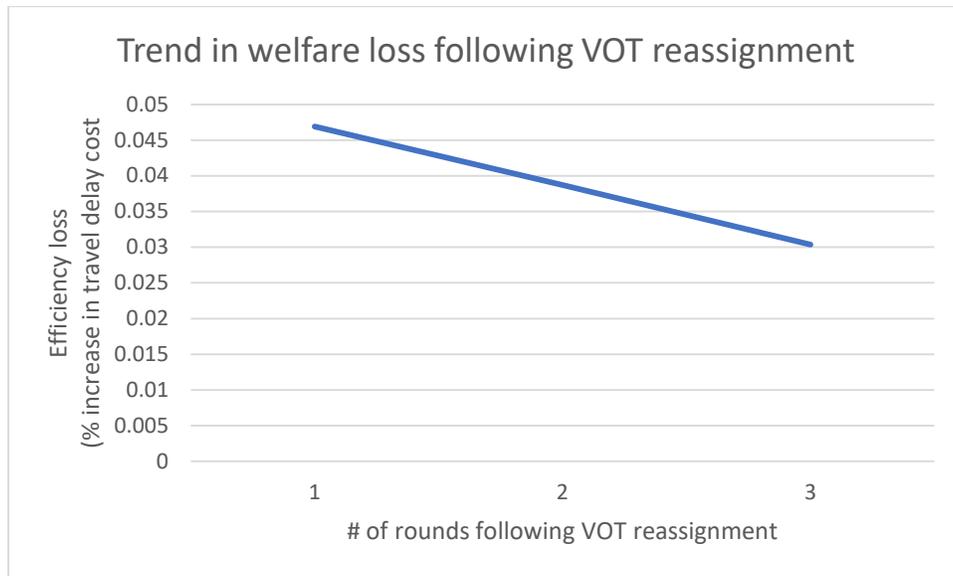


Figure 27: Trend in efficiency loss over the first three rounds following VOT re-assignment

In this circumstance, efficiency loss was reduced by 35% over three rounds as subjects adjust to the VOT reassignment.

Resiliency

Preliminary findings show mixed evidence regarding the resiliency of the mechanism to changing capacity conditions. This aspect of the mechanism's performance was indicative of its ability to be used in incident management contexts.

In two of three instances, the magnitude of subject misbidding was completely unaffected. These sessions are shown below:

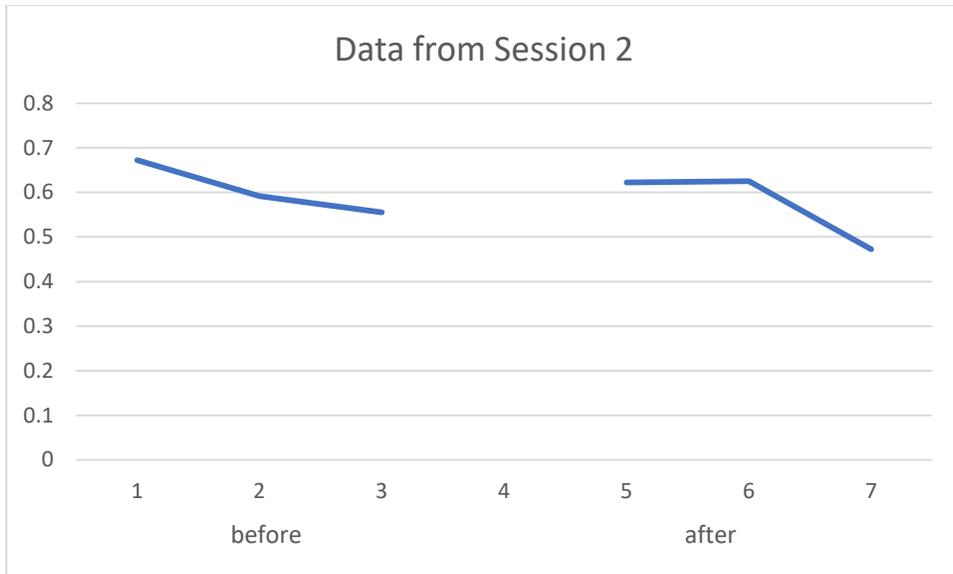


Figure 28: Average magnitude of misbidding before and after road capacity shift (session 2)

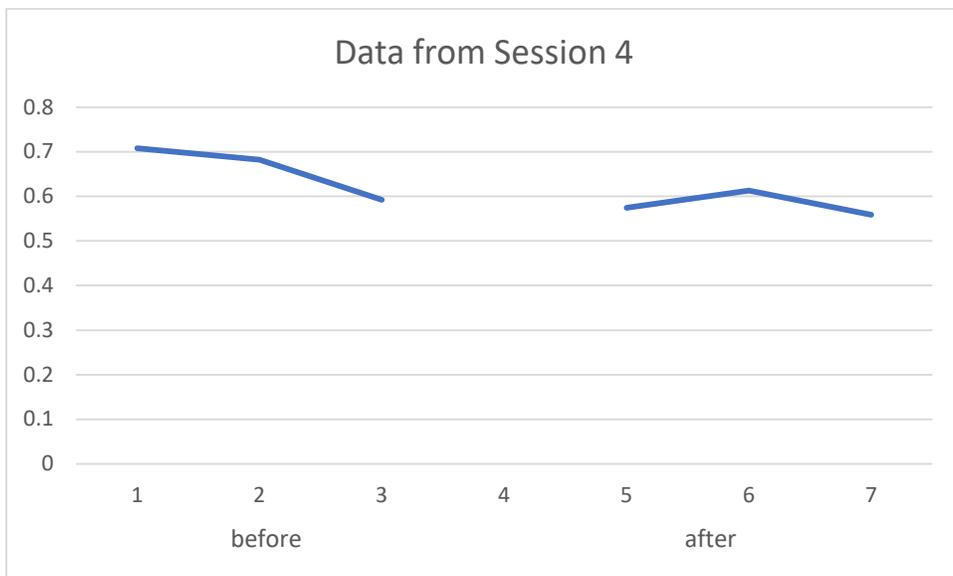


Figure 29: Average magnitude of misbidding before and after road capacity shift (session 4)

This is encouraging, because it demonstrates that even in adverse conditions the mechanism continues to work just as well. Trying to estimate a VOT for a standard second-best toll in these conditions, however, would likely be more difficult - making the VCG a superior option.

In one of three sessions, misbidding increased after the shift in road capacity.



Figure 30: Average magnitude of misbidding before and after road capacity shift (session 3)

This shift was problematic, because it reflected shaky trust in / understanding of the mechanism by the subjects. After tolls, travel times, and lane allocations changed due to the capacity decrease, subjects “panicked” and began to experiment with a new VOT.

The road capacity shifts were implemented as part of the three treatments for which the least amount of mechanism/VOT explanation was provided. Therefore, although the capacity shift affected subject misbidding in one of the three sessions, this should be considered a worst case finding that would likely improve had the subjects been provided with better pre-experiment information?

The welfare analysis, which was a true measure of the impact of capacity shifts, was forthcoming. Although efficiency losses increase after the loss of free lane capacity, this was likely because the differential between the toll and free lane increases, misbidding becomes more costly. One must therefore put these increased efficiency losses in context by comparing it to standard second-best tolling efficiency losses given the same capacity loss.

Analysis of performance affects from implementing the VCG mechanism as a second-best solution

There were no significant drawbacks to the VCG mechanism associated with implementing it as a second-best solution, where only the fast lane was tolled. The table below shows the efficiency losses incurred by restricting the slow lanes to be free. The row headings differentiate between three efficiency-loss metrics analyzed. The column headings differentiate between combinations of three distinct VOT distributions tested in the analysis and two distinct free-lane road capacities tested.

Scenario	VOT array 1 Capacity array 1	VOT array 2 Capacity array 1	VOT array 3 Capacity array 1	VOT array 3 Capacity array 2
Incentivized Efficiency Loss in equilibrium	0%	1%	0%	0%
Worst possible non-dis-incentivized efficiency loss in equilibrium	1%	1%	0.5%	0%
Experimental efficiency loss	< .2%	< .2%	< .2%	0%

Table 12: Summary of efficiency losses from leaving slow lane untolled

1) Incentivized efficiency loss in equilibrium

Although in theory leaving the slow lanes unpriced could create slight incentives for low VOT to inflate their bids, these incentives were typically not present given experimental parameters. This was tested by simulating the experiment in the case where every subject truthfully provided a VOT that was equal to his or her payment depreciation rate. Next, instances were identified where a subject could increase their earnings by stating a VOT that was

higher than their payment depreciation rate. Only one experimental scenario out of provided such an opportunity. Exploiting this opportunity increases the social cost of travel delay by 1%. Obviously, in the other cases, there was a 0% increase since there was no opportunity. This constitutes what theoretical efficiency losses due to the second-best implementation would be in equilibrium. On average, these losses were .25%. To put this loss in perspective, it was equal to approximately 1/5th of the best-case efficiency loss when using a standard toll. Therefore, in theory, the second-best implementation was not a deal-breaker for the VCG. Furthermore, for these losses to be manifested, however, subjects would need to identify and exploit these opportunities.

2) Worst possible non-disincentivized efficiency loss in equilibrium

Although in most cases subjects did not have an individual incentive to inflate their bid, there was also no disincentive for them to inflate their bid. Thus in the worst case of every low VOT increasing their bid up until the very point where they would be switched to the toll lane, it was possible for further efficiency losses to occur. This worst case, which corresponds to every free-lane user bidding the maximum possible VOT that still lets them use the free lane, was simulated. It was found that for every scenario but one, efficiency losses due to extra users on the toll lane would equal between .5% and 1%. On average, this loss was equal to .625%. To put this loss in perspective, it was equal to approximately 1/2th of the best-case efficiency loss when using a standard toll. Thus even in the worst possible case, the second-best implementation of the VCG was still beneficial, but half the benefits were eroded. This worst case was almost impossible to be achieved in practice. There was virtually no way subjects would know exactly where the toll lane cutoff would be, and attempting to do so would almost

certainly end up with them frequently being assigned to the toll lane (which would make them worse off).

3) Experimental efficiency loss

It was not possible to determine the social cost of using the second-best implementation of the VCG in practice, because no sessions were run for assigned-VOT treatments that used an implementation where all users (even those in the slow-lane) were charged. The best estimate from the available data was obtained by identifying the average bid-inflation of experimental low VOT users, and identifying what welfare losses this amount of bid-inflation would cause if all other subjects were to bid their true VOT. The average experimental bid inflation was about 15 cents per second, or 20% of the average free lane VOT. This amount of bid inflation never increased overall travel delay cost by more than .2% when simulated. Furthermore, overbidding was just as prevalent among high-VOT users who had no incentive to do so. Lastly, there were only a handful of instances where subjects who had an incentive to inflate their bid actually did. Therefore being constrained to the second-best implementation poses minor theoretical efficiency concerns for the VCG, and negligible practical ones. This shows that the mechanism was compatible with the desire of planners to leave one set of lanes unpriced.

5.4.2 Results for Treatments with Real Subjects VOT

The results that follow were for treatments where subjects completed the experiment at their own pace. The sooner they finish their driving tasks, the sooner they have to leave. Thus, their own personal VOT was relevant to their tradeoffs between paying tolls and saving travel time.

Plausibility

The average VOT of subjects ranged from 8-12 dollars per hour depending on the treatment. This could not be directly compared to empirical studies of SR-91 users, which found the average VOT of \$12-13/hour (stated preference) and \$21/hour (revealed preference). The reason was that the SR-91 study samples a representative group of toll-road users, while the sample for these treatments were limited to students. The students will likely have lower average incomes, and more importantly were participating in experimental sessions that typically pay \$15/hour. In light of these factors, the experimental VOT elicited by the mechanism seemed reasonable. The VOT was below the real-world toll road VOT as expected, and below what subjects were being paid per hour for participating in experiments.

Some individual subjects made seemingly irrational bids. These subjects bid well over \$20 dollars on average, which did not make economic sense given that subjects only make 15 dollars per hour on average (and therefore had implicitly indicated that they value their time less than that).

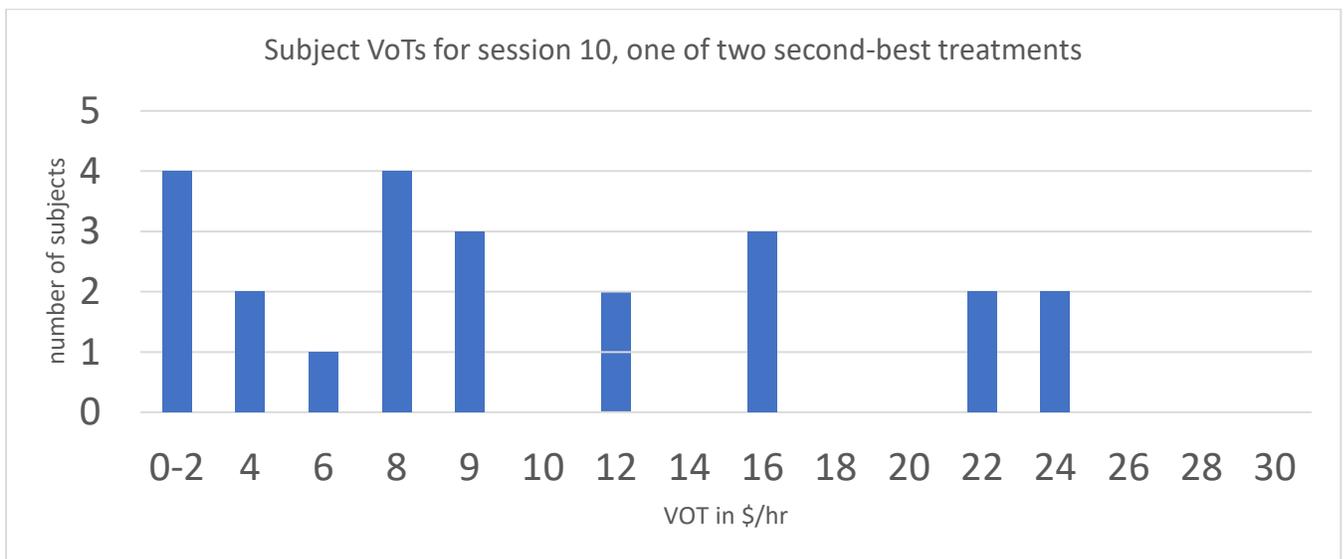


Figure 31: Distribution of Subjects' Average VOT Elicited from a Second-best VCG Mechanism

The above figure shows the distribution of elicited bids for two types of “real VOT” treatments; one type used a second-best modification of the VCG mechanism described earlier in the report, and the other used a full “first-best” VCG where drivers in any lane may be charge – even the slow lanes (fast-lane users will still be charged more).

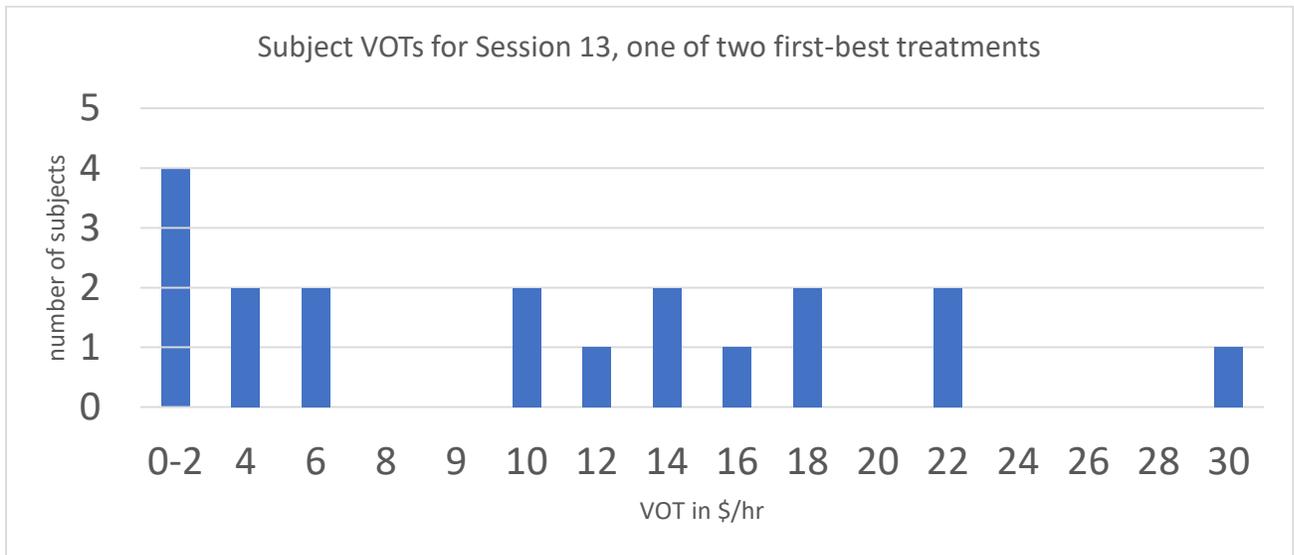


Figure 32: Distribution of Subjects’ Average VOT Elicited from a First-best VCG Mechanism

When comparing the two distributions, there are no obvious difference between the first-best and second-best implementation of the VCG mechanism. Both have roughly the same shape, and the same mean of \$10/hour.

Both treatments also have some subjects make unreasonably high bids and an unusually high frequency of near-zero bids. Possible reasons for this unusual bidding behavior were confusion about the mechanism (specifically the relationship between bids and tolls paid / time saved), boredom, or the inability of the mechanism to accommodate true time preferences as a single linear VOT. To elaborate on the last reason - if subjects wanted to ensure that they were given access to the toll lane to meet some sort of deadline, they would not be able to guarantee

this by bidding their true VOT, since their lane assignment depended on the VOT of other subjects as well. They could only come close to guaranteeing a toll assignment by bidding significantly higher than their true VOT.

Evidence supporting mechanism confusion was proven when extra feedback was provided, subject bids looked much more sensible. The figure below shows that when subjects were given feedback about impact of their bid, and hypothetical outcomes had they bid differently, abnormally high bids were somewhat less prevalent and extreme. The share of near-zero bids also became more sensible.

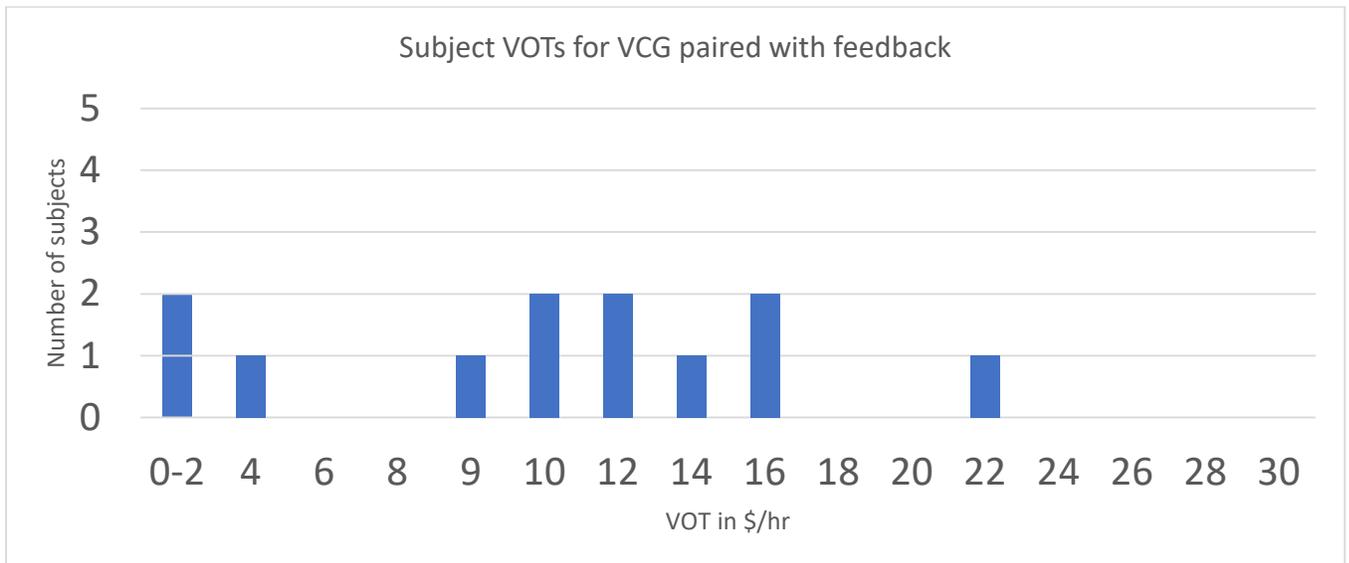


Figure 33: Distribution of Subjects' Average VOT Elicited from a VCG paired with feedback

The feedback likely accomplished this by both removing the need for subjects to experiment with different bids, and also by making the relationship between bids and toll / lane assignments much more clear to subjects.

Evidence for the inability of the mechanism to accommodate true time preferences as a single linear VOT as a cause for unrealistic elicited VOT is quite strong. The figure below shows that when subjects were given the option to pay extra money to greatly increase the odds of being assigned to their preferred lanes, bidding anomalies were essentially eliminated.

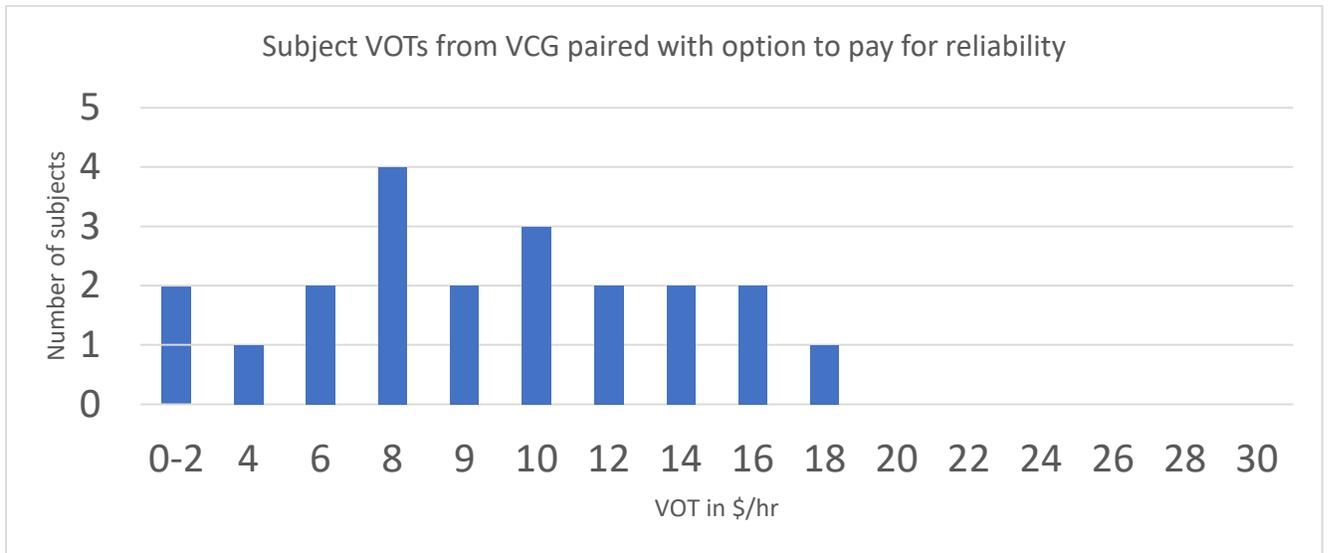


Figure 34: Distribution of Subjects' Average VOT Elicited from a VCG paired with option to pay for reliability

Instances of abnormally high bids decreased dramatically, as did the share of near-zero bids. This indicates a desire for "reliability" the simple VCG mechanism does not accommodate. Thus, the VCG mechanism may not be appropriate for all users and contexts.

VOT Stability

Ideally, subjects VOT would not fluctuate much. This would indicate that subjects quickly grasp the mechanism and can immediately identify their own VOT. It also would

indicate that a single linear VOT was suitable for subjects throughout the duration of the experimental session (Individual VOT shifts following a trend might indicate VOT non-linearity).

The average VOT tended upward slightly for every treatment except the treatment where one can pay for reliability. The figures below shows examples of this.

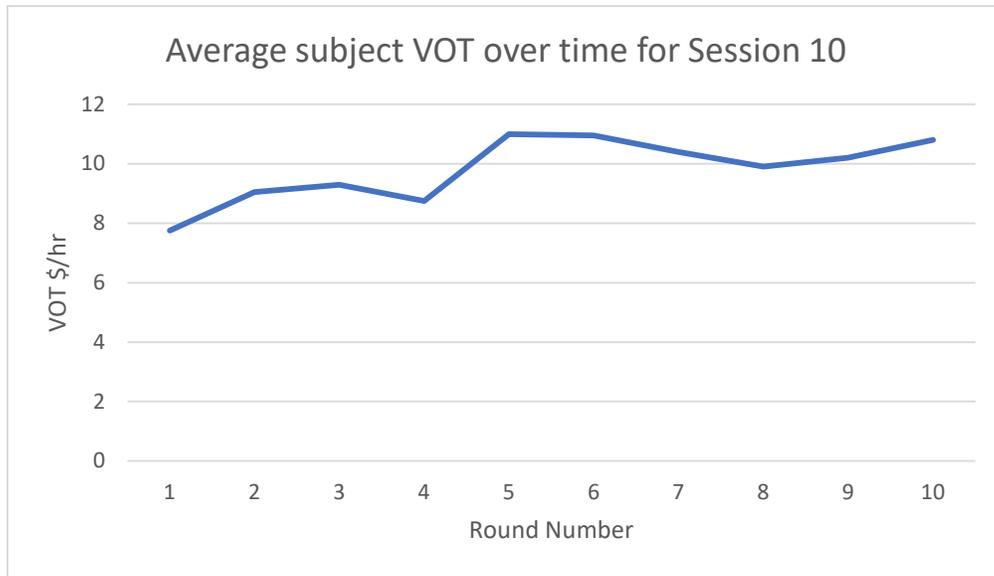


Figure 35: Trend in average subject VOT over time - typical pattern (Session 10)

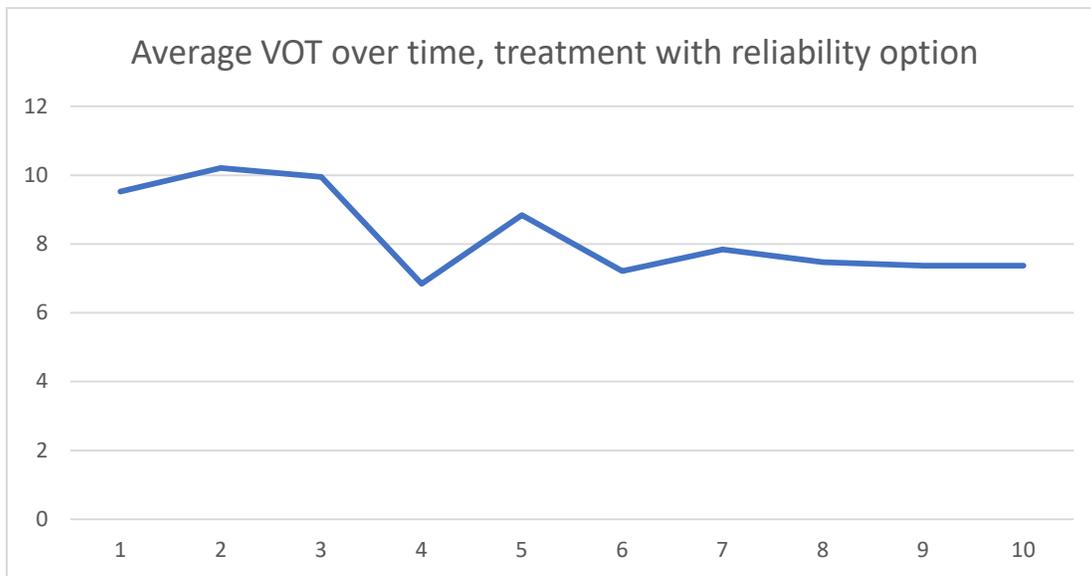


Figure 36: Trend in average subject VOT over time for VCG paired with reliability option

The upward trend in subjects VOT for most treatments might be due to boredom, regret, or subjects not understanding their true VOT right off the bat. Boredom was shown to be a significant factor determining subject bids, and would make sense that boredom would increase over time.

Another sign of VOT instability was the presence of significant volatility in the reported VOT for some subjects. The post-experiment questionnaire responses might suggest that subjects need a “break-in” period to familiarize themselves with the mechanism and understand what their own VOT should be.

Mechanism satisfaction

Subjects’ satisfaction with the toll system in the various experiment treatments ranged from 70 – 100% saying they liked it. Further study of the questionnaire is required to understand what leads to subject dissatisfaction. Factors mentioned in the post-experiment questionnaire were opposition to toll systems in general, uncertainty of travel times, and being charged to use the slow lanes (for first-best treatments). The large majority of subjects did not express regret at the VOT they chose, i.e. they did not wish they had chosen a higher or a lower VOT when the experiment was over.

The table below shows how self-reported mechanism satisfaction varied with the type of experimental session subjects participated in.

Treatment	Second-best VCG	Second-best VCG Subjects shown cents/min VOT.	First-best VCG	VCG w feedback	VCG w reliability
% Reporting Satisfaction with toll system	66	100	63	91	84

Table 13: Summary of subject satisfaction with VCG by treatment

Perhaps subjects like a cents/minute VOT, and, the option to pay for reliability. Further analysis is required to control for other factors, and to explore reasons for subjects' satisfaction or dissatisfaction.

Factors affecting subject bids

Gender: Being male was associated with a VOT that was roughly \$1.75/hour lower (statistically significant at 10% level). The sign was consistent with the finding in Parkany (1999) that female drivers were more likely to use toll lanes.

Ethnicity: Compared to subjects identifying as white or "Asian" (specifically using the word "Asian", suggesting a domestic student), subjects that identified as Hispanic/Latino/Latina chose a VOT that was \$2.50 higher on average (statistically significant at 10% level). The reason for this was unclear.

Schedule: A subject's schedule strongly determined the VOT he or she chose. The base case was a subject who said they were essentially committed to staying at the experimental for the entire schedule duration (e.g., has to wait around for a ride). This base case should be associated with a very low VOT. By comparison, subjects who either mentioned nothing or said they had class / office hours had an average VOT that was \$4/hour higher (statistically significant at the 5% level). Subjects who mentioned a deadline or something important to do later had a VOT that was \$10/hour higher on average (statistically significant at the 5% level). This was a good check of a VOT elicited by the mechanism. This strongly corresponds with intuition.

Toll History: Given a past real-world opportunity to have taken a toll lane, those who reported ever having actually chosen the toll lane reported a VOT of \$3/hour higher than those who hadn't (statistically significant at the 5% level). This was another verification that confirms the intuition. Those who have used toll lanes in real life should have a higher VOT.

License: Subjects with a driver's license had a statistically significant increase in reported VOT of over \$4/hour. It is unclear why this is the case. Although increased mobility increases the opportunity cost of participating in the experiment, many members of the younger generation are able to use ride-sharing services rather than drive themselves and thus do not need to get a license.

Age: Subjects older than the age of a typical undergrad reported a VOT of over \$3/hour higher on average. This was not statistically significant. However, this matches the Parkany (1999) finding that age increased the probability of choosing the toll lane.

Boredom: In a separate regression, subjects who reported being bored also chose a VOT that was about \$4/hour higher on average (statistically significant at the 1% level). This is another indication that an elicited VOT matched intuition.

5.5 Concluding Remarks

The experimental tests of a VCG mechanism used to manage priced lanes revealed that there were significant disparities between the theoretical and realized benefits of the mechanism. The theoretical benefit of the mechanism was that operators would not need to guess the VOT of freeway users in order to set a toll, and could avoid efficiency losses that result due to errors in the VOT being estimated. This disparity between theory and practice mainly stems from the fact that subjects often do not truthfully/accurately provide their VOT for the mechanism, despite being incentivized to do so.

The discrepancies between subjects' truthful and reported VOT reduced the performance of the VCG mechanism to a level below that of a perfectly implemented second-best toll, which corresponds to what, is currently implemented on freeways. In other words, by the most conservative measure, the VCG did more harm than the problem it was intended to solve. For the more realistic scenarios, where operators/planners do not correctly know the true parameters of the distribution of drivers VOT, the VCG mechanism performed better. Therefore, the VCG

mechanism could likely improve toll lane performance relative to the standard second-best tolls currently being implemented.

Much of the misbidding by subjects stemmed from a lack of understanding of how the mechanism works. To this end, it was encouraging those greater efforts to explain the mechanism to subjects before the experiment resulted in significantly better performance.

Therefore, significant efforts to educate potential users should be made if such a mechanism were to be implemented in the future.

Experimental results from treatments that elicited an actual VOT from subjects, show that the VCG mechanism was well-enough understood to elicit a sensible VOT. Furthermore, most subjects reported being satisfied with the mechanism and reported feeling confident in their ability to know and provide their true VOT.

Some subjects, however, provided an unrealistic VOT that was far too high. This is evidence that the mechanism does not work well for everyone. Abnormal bidding behavior was somewhat reduced by providing detailed feedback to drivers after every round, and was greatly reduced by adding a feature that allowed giving subjects more certainty in their travel times. More certainty was facilitated by giving subjects the option to pay extra money to influence which lane they were assigned to, without having to modify the VOT they provided. This feature was not actually a part of the mechanism, but rather supplemented it. Drivers gave bids that are more plausible and were more satisfied with the mechanism when given this option. The positive response to this modification was strong evidence that a single linear VOT – which was the only time preference parameter that lane-assignment and toll setting was based on for the mechanism – was insufficient to capture all the time preferences of drivers. This was especially true when drivers have scheduling constraints that are more salient than a simple trade-off

between time and money. Future work will focus on whether the VCG mechanism itself can accommodate these extra preferences for travel time certainty.

6. References

- Abdel-Aty, Mohamed A., Ryuichi Kitamura, and Paul P. Jovanis. "Using Stated Preference Data for Studying the Effect of Advanced Traffic Information on Drivers' Route Choice." *Transportation Research Part C: Emerging Technologies* 5, no. 1 (February 1997): 39–50. doi: 10.1016/S0968-090X(96)00023-X.
- 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. doi:10.1109/VNIS.1991.205745.
- Arnott, Richard, et al. "Route Choice with Heterogeneous Drivers and Group-Specific Congestion Costs." *Regional Science and Urban Economics*, North-Holland, 15 Apr. 2002, www.sciencedirect.com/science/article/pii/016604629290026W.
- Ben-Elia, Eran, and Yoram Shiftan. "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, no. 4 (May 2010): 249–64. doi:10.1016/j.tra.2010.01.007.
- Bonsall, P., and I. Palmer. "The Impact on Route Choice of Direction Advice and Guidance from VMS Signs," 10–10. IEE, 1995. doi: 10.1049/ic: 19950251.
- Braid, Ralph M. "Peak-Load Pricing of a Transportation Route with an Unpriced substitute." *Journal of Urban Economics*, Academic Press, 1996, www.sciencedirect.com/science/article/pii/S0094119096900285.
- Brocken, M. G. M., & Van der Vlist, M. J. M. (1991). Traffic control with variable message signs. In *Vehicle Navigation and Information Systems Conference*, 1991 (Vol. 2, pp. 27–45). doi:10.1109/VNIS.1991.205746.
- Brownstone, David, Michael McBride, Si-Yuan Kong, and Amine Mahmassani. "Experimental Studies for Traffic Incident Management." California Department of Transportation, July 5, 2016. http://dot.ca.gov/research/researchreports/reports/2016/CA16-2804_FinalReport.pdf.
- Chatterjee, K, N.B Hounsell, P.E Firmin, and P.W Bonsall. "Driver Response to Variable Message Sign Information in London." *Transportation Research Part C: Emerging Technologies* 10, no. 2 (April 2002): 149–69. doi: 10.1016/S0968-090X(01)00008-0.
- Chatterjee, Kiron, and Mike McDonald. "Effectiveness of Using Variable Message Signs to Disseminate Dynamic Traffic Information: Evidence from Field Trails in European

- Cities.” *Transport Reviews* 24, no. 5 (September 2004): 559–85. doi: 10.1080/0144164042000196080.
- Dudek, C.L., G.D. Weaver, D.R. Hatcher, and S.R. Richards. “Field Evaluation of Messages for Real-Time Diversion of Freeway Traffic for Special Events.” *Transportation Research Record* 682 (1978): 37–45.
- Emmerink, Richard H.M., Peter Nijkamp, Piet Rietveld, and Jos N. Van Ommeren. “Variable Message Signs and Radio Traffic Information: An Integrated Empirical Analysis of Drivers’ Route Choice Behaviour.” *Transportation Research Part A: Policy and Practice* 30, no. 2 (March 1996): 135–53. doi: 10.1016/0965-8564(95)00016-X.
- Gan, Hongcheng. “Investigation of Driver Response to the Enhanced Urban Freeway Variable Message Sign Information.” *KSCE Journal of Civil Engineering* 17, no. 6 (September 2013): 1455–61. doi: 10.1007/s12205-013-0231-2.
- Hitlin, Paul. “Research in the Crowdsourcing Age, a Case Study.” Pew Research Center, July 11, 2016. <http://www.pewinternet.org/2016/07/11/research-in-the-crowdsourcing-age-a-case-study/>.
- Horowitz, Alan, Ian Weisser, and Thomas Notbohm. “Diversion from a Rural Work Zone with Traffic-Responsive Variable Message Signage System.” *Transportation Research Record* 1824, no. 1 (January 1, 2003): 23–28. doi: 10.3141/1824-03.
- Iida, Yasunori, Takamasa Akiyama, and Takashi Uchida. “Experimental Analysis of Dynamic Route Choice Behavior.” *Transportation Research Part B: Methodological* 26, no. 1 (February 1992): 17–32. doi: 10.1016/0191-2615(92)90017-Q.
- Jou, Rong-Chang, Soi-Hoi Lam, Yu-Hsin Liu, and Ke-Hong Chen. “Route Switching Behavior on Freeways with the Provision of Different Types of Real-Time Traffic Information.” *Transportation Research Part A: Policy and Practice* 39, no. 5 (June 2005): 445–61. doi:10.1016/j.tra.2005.02.004.
- Kattan, Lina, Khandker Habib, Shahid Nadeem, and Tazul Islam. “Modeling Travelers’ Responses to Incident Information Provided by Variable Message Signs in Calgary, Canada.” *Transportation Research Record: Journal of the Transportation Research Board* 2185 (December 2010): 71–80. doi: 10.3141/2185-10.
- Khattak, Asad J., Joseph L. Schofer, and Frank S. Koppelman. “Commuters’ Enroute Diversion and Return Decisions: Analysis and Implications for Advanced Traveler Information Systems.” *Transportation Research Part A: Policy and Practice* 27, no. 2 (April 1993): 101–11. doi: 10.1016/0965-8564(93)90066-T.
- Koppelman, Frank. “Improving Our Understanding of How Highway Congestion and Price Affect Travel Demand.” *SHRP2 Report*, 30 Nov. 2012, trid.trb.org/view/1226379.
- Levinson, D. and Huo, H. (2003). Effectiveness of Variable Message Signs. TRB Conference 565.

- Mahmassani, Hani S., and R. Jayakrishnan. "System Performance and User Response under Real-Time Information in a Congested Traffic Corridor." *Transportation Research Part A: General* 25, no. 5 (September 1991): 293–307. doi: 10.1016/0191-2607(91)90145-G.
- Mahmassani, Hani S., and Yu-Hsin Liu. "Dynamics of Commuting Decision Behavior under Advanced Traveler Information Systems." *Transportation Research Part C: Emerging Technologies* 7, no. 2–3 (April 1999): 91–107. doi: 10.1016/S0968-090X(99)00014-5.
- Noland, Robert B, and Kenneth A Small. "Travel-Time Uncertainty, Departure Time Choice, and the Cost of Morning Commutes." *Transportation Research Record*, 30 Nov. 1994, trid.trb.org/view.aspx?id=452605.
- Parkany, Emily. "Traveler Responses to New Choices: Toll vs. Free Alternatives in a Congested Corridor." PhD Dissertation, 1999.
- Peeta, Srinivas, Jorge L. Ramos, and Raghubhushan Pasupathy. "Content of Variable Message Signs and On-Line Driver Behavior." *Transportation Research Record: Journal of the Transportation Research Board* 1725, no. 1 (January 1, 2000): 102–8. doi: 10.3141/1725-14.
- Selten, Reinhard, Michael Schreckenberg, Thorsten Chmura, Thomas Pitz, Sebastian Kube, Sigurður F. Hafstein, Roland Chrobok, Andreas Pottmeier, and Joachim Wahle. "Experimental Investigation of Day-to-Day Route-Choice Behavior and Network Simulations of Autobahn Traffic in North Rhine-Westphalia." In *Human Behavior and Traffic Networks*, edited by Michael Schreckenberg and Reinhard Selten, 1–21. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004. doi: 10.1007/978-3-662-07809-9_1.
- Small, Kenneth A., et al. "Uncovering the Distribution of Motorists' Preferences for Travel Time and Reliability." *Econometrica*, Blackwell Publishing Ltd, 13 June 2005, onlinelibrary.wiley.com/doi/10.1111/j.1468-0262.2005.00619.x/abstract.
- Small, Kenneth, and Jia Yan. "The Value of 'Value Pricing' of Roads: Second-Best Pricing and Product Differentiation." *EconPapers*, 1 Jan. 2001, econpapers.repec.org/RePEc:cdl:uctcwp:qt9569k1sz.
- Small, Kenneth. "The Scheduling of Consumer Activities: Work Trips." *American Economic Review*, American Economic Association, 1 Jan. 1982, econpapers.repec.org/RePEc:aea:aecrev:v:72:y:1982:i:3:p:467-79.
- Sullivan, E. "EVALUATING THE IMPACTS OF THE SR 91 VARIABLE-TOLL EXPRESS LANE FACILITY. FINAL REPORT." *Home - Transport Research International Documentation - TRID*, 30 Apr. 1998, trid.trb.org/view.aspx?id=707271.
- Verhoef, Erik, and Kenneth Small. "Product Differentiation on Roads." *Journal of Transport Economics and Policy*, University of Bath, 1 Jan. 1999, econpapers.repec.org/RePEc:tpe:jtecpo:v:38:y:2004:i:1:p:127-156.

Wardman, M, P.W Bonsall, and J.D Shires. "Driver Response to Variable Message Signs: A Stated Preference Investigation." *Transportation Research Part C: Emerging Technologies* 5, no. 6 (December 1997): 389–405. doi: 10.1016/S0968-090X(98)00004-7.

Wardman, Mark. "Public Transport Values of Time." *Transport Policy*, Pergamon, 30 July 2004, www.sciencedirect.com/science/article/pii/S0967070X04000319.

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