AN INVESTIGATION OF LANE-CHANGING RELATED ENVIRONMENTAL FACTORS AND POSSIBLE LANE-CHANGING INDICATORS ON HIGHWAY

by

Xiaojian Jin

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

School of Industrial Engineering

West Lafayette, Indiana

May 2022

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Steven J Landry, Co-Chair
School of Industrial Engineering (Penn State University)

Dr. Brandon J Pitts, Co-Chair
School of Industrial Engineering

Dr. Caitlin Surakitbanharn
Apple Inc.

Dr. Zachary J Hass
School of Industrial Engineering

Approved by:
Dr. Barrett S. Caldwell

ACKNOWLEDGMENTS

To my parents, I am truly grateful for your support both mentally and financially.

To Dr. Landry, you changed the way I think, thank you so much.

To my committees, thank you for all your help to get this work complete.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Unsafe lane changes have been identified as a common factor in motor vehicle accidents. It would be helpful, particularly for automated vehicles, to know if there are behaviors of vehicles, beyond a directional signal, or characteristics of the traffic environment that correlated with a higher probability of an unsafe lane change (lane changes without a directional signal). This work investigates what the observable cues are that drivers use to determine the relative safety when overtaking front vehicles, and if drivers make more lane changes under certain conditions on highways. This study utilizes interviews, surveys, 3D animation software, and highway driving public footage for data collection and experiments. It is found that a side-to-side motion of the front vehicle or a factor that might trigger a side-to-side motion of the front vehicle in the environment is the key marker that indicates a possible unsafe lane change, and it is also found that traffic speed, time of day, traffic flow, and a combination of traffic density & number of lanes & vehicle count all have effects on drive's decision on making lane changes on different levels.

CHAPTER 1. INTRODUCTION

Unsafe lane changes have been identified as a common factor in motor vehicle accidents, and NHTSA estimated that about 9 percent of all motor vehicle accidents occur due to some kind of lane changing or merging collision (NHTSA, 2003). It would be helpful, particularly for automated vehicles, to know if there are behaviors of vehicles, beyond a directional signal, or characteristics of the traffic environment that correlated with a higher probability of an unsafe lane change.

We do not know what markers exist that would indicate that a vehicle will (unsafely) change lanes without a directional signal; such markers, if detectable by an autonomous vehicle, would help the vehicle prevent accidents caused by such unsafe lane changes. And we do not know if drivers are more likely to make a lane-change under certain conditions; such environmental conditions, if detected by an autonomous vehicle, would help the vehicle to execute appropriate actions beforehand to mitigate unforeseeable risks accordingly. It is crucial for the development of autonomous vehicle highway safety to identify observable markers that indicate a possible unsafe lane change, or if there is any correlation between the environmental conditions and the possibility of lane-changes to fill this gap.

This work examines two research questions to address the knowledge gap:

1. *Are lane changes more frequent under certain conditions?*
2. *Are there observable cues that drivers use to determine the relative safety of passing vehicles?*

For research question one, it utilized the highway traffic surveillance system of California to collect public visual data to investigate the relationship between the likelihood of a lane change on highway and five environmental factors, traffic speed, traffic density, traffic flow, time of day, and number of lanes. The reason why the state of California was chosen for this study is mostly because it has the most developed highway traffic surveillance system across the states, in terms of camera image quality and camera quantity, and the high traffic volume of California is very desirable for the study. The reason why higher traffic volume is desirable for the study is because higher traffic density is needed to be observed in this study, there is a higher chance to observe this in higher volume traffic. For research question two, this study interviewed a group of professional drivers such as truck drivers and taxi drivers to collect what are the possible observable visual cues. Then I replicated the markers in animation and show them to another group of amateur drivers for validation.

The current self-driving technologies still struggle to understand if the manual vehicle ahead is going to make a lane-change unless the directional signal is on. The results of this study will provide guidelines to assist the development of the self-driving technologies, particularly their ability to perceive an impending lane change without the input of a directional signal.

The rest of the document introduces the background information and a comprehensive literature review that discusses the previous studies on the topic of lane changes behavior and markers, a detailed explanation of the methods that are used in the two studies, followed by the results of the two studies, a discussion on the interpretation of the results, the conclusions that present the takeaways from the work, and the future work that layout a research agenda at the end.

CHAPTER 2. BACKGROUND AND LITERATURE REVIEW

2.1 Lane change behavior and markers

In 1986, Gipps established the first lane change model, and since 1990, researchers have been modeling lane-changes for traffic microsimulation software (Yang, 2019). Lane-changes can be categorized into mandatory lane-changes and discretionary lane-changes depending on the motivation of the driver. To elaborate, when the driver needs to get into another lane for a turn, to enter or exit a highway, to prevent a potential accident, this can be categorized as mandatory lane-changes in which the driver must leave the current lane. In cases where the driver is making lane-changes to improve the current driving condition such as move to a lane to drive at a more comfortable speed, to follow a certain vehicle, or to overtake a vehicle in front, this kind of lane-changes are categorized as discretionary lane-changes (Toledo, 2003; Gipps, 1986)).

The lane changing mode that Gipps proposed in 1986 stated that lane changing is a complex decision that depends on a number of objectives, for example, if a car needs to make a right turn on the most right lane in 100 ft but it still has to change lane to the left just to avoid a stopped vehicle ahead (Gipps, 1986). A driver thinks about three questions before making a lane change:

- Is it possible to change lanes?
- Is it necessary to change lanes?
- Is it desirable to change lanes?

Previous studies categorized lane-changing tasks into mandatory lane-changing(MLC) and discretionary lane-changing(DLC) depending on the intention of the behavior(Gipps, 1986; Hidas, 2005). In general, MLC is the lane-change that just keeps the driver and the vehicle on the required route. DLC is the lane-change that can provide the driver a better situational awareness, a better position, or at a higher/lower speed, such as overtaking a slow semi-truck or moving into a slower lane to drive at a more comfortable speed. The intention and classification of lane changes can affect driver's lane change maneuvers, such as drivers may make a MLC when the gap is smaller, and the drivers may not make the lane change if it's a DLC for that same amount of gap. (Toledo, 2003)

A lane change maneuver defined by the NHTSA, is a driving maneuver that move a vehicle from one lane to another where both lanes have the same direction of travel (Fitch, 2009). The lane changing markers mentioned in this study are the indicators exhibited by the front vehicle before a lane change maneuver is executed, except for the directional signals. In this study, the combination of indicators can be a series of things, it can be the single action of the front vehicle itself such as wobbling sideways gently, or it can be the multiple actions of the front vehicle itself such as wobbling sideways with the brake lights on, or it can also be the multiple actions of the front vehicle itself and some other conditions such as wobbling sideways, brakes lights on, and approach an exit. The markers mentioned in study two occur in the realm where the rear driver cannot clearly see the danger of the front vehicle, but the driver can definitely sense there is something strange going on in his/her field of view based on his/her driving experience. If these markers are not easily identifiable by experienced drivers, it can be difficult for an autonomous vehicle to identify through cameras and sensors based on plain metrics.

Previous works in the related area targeted developing novel models capable of detecting and tracking unsafe lane departure events to provide support for at-risk medical or aging populations or building an algorithm that can efficiently process large-scale naturalistic driving videos in detecting lane-change events by monitoring the pixel changes. However, no studies have been found in the area of identifying observable cues of a front vehicle that indicates a possible unsafe lane-change.

In recent years, lane-changing detection and lane-changing behavior have been receiving attention in traffic flow modeling. Numerous studies have been conducted on designing and developing complex algorithms to simulate and analyze urban streets and highways traffic patterns from a systematic standpoint, without considering the driver's personal characteristics. A study published by researcher Sun in 2012 at Shanghai Jiao Tong University studied driver's lane-changing behavior on urban streets from an in-vehicle perspective, found that driver's personality significantly affect their driving style such as aggressive or passive. And it is validated that asking a sample of drivers or the drivers' friends and family can accurately categorize the drivers into aggressive driving and passive driving groups which helps with micro-simulations by implementing this finding to replicate driver behaviors in urban street networks (Sun, 2012).

In 2020, a study focusing on the development of a model capable of detecting and predicting unsafe lane departure events that might happen more often in at-risk driver populations such as medical or aging populations. It utilized low-resolution driving video footage through a semantic lane detection pre-processor (Mask R-CNN) to detect and track the hull centroid, the results show a robust lane departure event detection rate at 81.82% (Riera, 2020) .

Although this study is closely related to study two topic which looks at lane change markers, this study is more focused on developing an algorithm to detect an unsafe lane departure event by utilizing video footage of the front vehicle while study one focuses on providing guidelines on the overall precursors that might lead to an unsafe lane change including the front vehicle behavior and the other factors, such as the environmental factors.

Knoop from Delft University of Technology, Netherlands conducted a study focusing on lane change behaviors on freeways by utilizing an online survey using video clips is very similar to the method used in study one which is using a combination of traffic video clip plus a multiple-choice question after each clip. It aimed to find the pattern of drivers when it comes to making choices for lane changes, and provide some insights on traffic operations and assist the development of microscopic simulation models. The practical usage of these models are for designing road layouts, traffic management measurements, such applications require a good understanding of driving behaviors and drivers distributions (Knoop, 2018). Knoop found that the majority of the participants would like to drive at a comfortable speed in the beginning and stay with it which is the strategy number one. A slightly less chosen strategy 2 is that they will choose a comfortable lane in the beginning and then adapt the traffic speed of that lane. The third strategy is very similar to the first one, but except for sticking with a comfortable speed, they speed up when overtaking another vehicle in a different lane. The least chosen strategy number 4 is that they neither stick to a certain speed nor stick to a certain lane.

Knoop's study is very interesting because it shows the distribution of driver behaviors which can be used as a reference while designing the script for study 1 after the survey results are collected from the professional drivers, this is one extra material to consider on top of the survey results.

Among the four driving strategies listed in Knoop's study in 2018, some of these driving strategies can provide some guidelines when it comes to lane changing: (Knoop, 2018)

1. Strategy 1: Drivers maintain the speed of the vehicle at a comfortable level on the freeways and try to stay with that speed as much as possible. If necessary, they will change lanes to overtake forward vehicles to maintain that speed. While making the overtake maneuver, they would stay at the same speed.
2. Strategy 2: Similar to the strategy 1, drivers drive at a comfortable speed in the beginning as much as possible and make lane changes to overtake other forward vehicles if necessary. However, they speed up while overtaking other vehicles and then they usually stop staying with the original speed.
3. Strategy 3: Drivers select a comfortable lane in the beginning and then accommodate to the traffic speed of that lane to stay in there, however they usually leave that lane if the speed becomes too slow or too fast, the margins are usually plus or minus 40 kilometer per hour.
4. Strategy 4: Drivers do not have a specific speed or a specific lane they would stick to, they will just follow the traffic on the road and go as they like.

Besides the lane-changing setup, the survey design can be looked at and used as a reference as well since both study 1 and Knoop's study in 2018 utilizes a survey to collect results.(Knoop, 2018) Knoop's driving behavior study used a survey that starts with a picture of a freeway that has 3 lanes and a speed limit of 100 kilometer per hour. The participants were asked which lane they will go to and also what speed they will be at that moment, then a personalized survey is presented to the participant based on the answer. The personalized survey contains a video that has a speed that match with the participant's desired speed, here is a table of the groups of different driving speeds:

Desired speed	Category	Speed in video clip
< 103 km/h	“Slow” driver	~100 km/h
103- 110 km/h	“Average” driver	~107 km/h
> 110 km/h	“Fast” driver	~115 km/h

Figure 1. Previous study screenshot 1

According to the answer that the participant indicated in the previous question, a personalized survey containing a corresponding video that drives at a matched speed is presented to the participant. The idea of the survey is to investigate how the drivers choose a certain lane by utilizing a number of questions and videos. The traffic video clip is reworded to let the participants feel they are driving at their normal speed as much as possible. The pic here is a screenshot of the survey video that Knoop used, a front video of the vehicle to show the traffic is presented with another video feed facing the rear of the vehicle, the current speed of the vehicle is given on the top right corner of the view and it is matched with the participant’s selection of speed categorizes in the previous question.

Figure 2. Previous study screenshot 2

The participants are asked what they would act if they are in the driver seat in the vehicle, this method is surprisingly similar to the method that is going to be used in study 1, which is asking the participants how they would act if they are in the driver seat trying to overtake the front vehicle. This similarity in method building that putting the emphasis of the questions on the participants instead of putting it on the other factors such as the front vehicles indirectly shows the feasibility and validity of the method in study 1; it is more intuitive for the participants to think about the next action of themselves instead of the other vehicles', such prediction of action can be used to find out what the participants think the front vehicle is going to do next more accurately.

Throughout the survey there are 14 questions with corresponding video clips populated, they all aim to fulfill the goals that whether the participant is following strategy 1, 2, 3, or 4, and whether the participant is driving to obey the keep right rule, and does the participant use the right lane to overtake the front vehicle, and does the participant make any lane change to create space for a merging car? A screenshot of all the questions and corresponding videos is listed on the next page.

After the survey design, the validation part of Knoop's study can be a good reference. Knoop used 25 actual driving cases to validate the survey results, and with the instrumented vehicles, the validation part is actually used for filming as well. The same 25 participants are involved in the actual driving(seating in the passenger seat to observe), a short interview was carried out after the drive, and later on, using exactly the same video footage to validate the result in the survey.

Table 2: Overview of the questions: screenshot (all but the first row cropped due to space limitation), question number, theme of question (in bold), and description.

1: **Right overtaking**. You are driving in a tunnel on a 4-lane freeway on the third lane from the left. A truck is driving up ahead in the rightmost lane and a person car is driving in the second lane from the left while there is space for that driver to keep right using your lane. You are getting closer to that person car.	2: **Strategies 1, 2, and 3**. You are driving on a 3-lane freeway in the right lane. A slightly slower predecessor is driving in front of you and you are approaching that vehicle, while a black car just passed you with high speed.	3: **Strategies 1, 2, and 3**. You are overtaking the white car, which increases its speed a bit, resulting in a very slow overtaking maneuver (this scenario is only shown to the respondents who want to overtake the vehicle with a constant speed).
4: **Strategies 1, 2, and 3**. You are driving on a 2-lane freeway during congestion in the left lane. You are approaching the tail of a stop-and-go wave, while traffic in the right lane drives smoother and with a higher speed.	5: **Right overtaking**. You just entered the freeway via an on-ramp, while in the lane to the left of you a black car is driving with a low speed. You need to proceed on this carriageway to follow your route.	6: **Strategies 1, 2, and 3**. You merge onto the main road from an on-ramp, while to continue your route you need to take the second exit from that point which is 3100 meters up ahead.
7: **Keep right/Strategy 4**. You are driving on a 3-lane freeway in the center lane while having passed a truck in the right lane. There is space to change lanes to the right after you have passed the truck, while a faster driver is approaching you from behind.	8: **Keep right**. You are driving on a 3-lane freeway in the center lane, while passing two trucks in the right lane. After you have passed the trucks there is an empty road ahead of you.	9: **Strategies 1, 2, and 3**. You are driving on a 3-lane freeway in the rightmost lane. A truck is driving in front of you in the same lane, while a faster driver is approaching you from behind on the center lane.
10: **Keep right**. You are driving on a 2-lane freeway in the left lane, while you pass a truck that is driving in the right lane. Up ahead another truck is driving there, while a faster driver is approaching you from behind.	11: **Strategy 4**. You are driving on a 3-lane freeway in the rightmost lane with a speed of 100 km/h. However, all other vehicles around you drive much faster and you are overtaken by them.	12: **Courtesy lane change**. You are driving on a 2-lane freeway in the right lane, while you approach a merging lane on which a vehicle is driving. However, the driver is not using its blinker.
13: **Strategies 1, 2, and 3**. You are driving on a 3-lane freeway in the center lane. To follow your route you have to take the exit that starts at 600 meters from the moment the video stops. On that moment a truck is driving ahead in the right lane.	14: **Keep right**. (Static picture). You are driving on a 3-lane freeway in the rightmost normal lane. This road section is equipped with a rush-hour lane which is currently open for traffic. You will pass the truck that drives on the rush-hour lane.	

Figure 3. Previous study screenshot 3

2.2 Lane changing environmental factors

The lane changing environmental factors mentioned in this study specifically mean, from the driver's standpoint, all the factors and conditions exist in the environment that contribute to the occurrence of a lane change, such as the traffic speed, time of day, and traffic density.

In a previous study, Knoop states that lane-changing and the drivers' choice of speed is not well understood on freeways, most of the drivers choose a speed and stick to it, another group with a smaller number of drivers choose to stick to a lane and adapt the speed based on that lane, the third group has a desired speed in the beginning and speed up when overtaking a vehicle on another lane (Knoop, 2018). Drivers have different strategies throughout their trips, and the environmental factors are suspected to be one of the elements that influence their decision making.

Over 90% of the traffic crashes globally are caused by human error according to a study about traffic faults in accidents in 2011 (Karacasu, 2011). Most of the studies that tried to find out the relationship between risky driving and traffic crashes addressed speeding, tailgating, the usage of alcohol and secondary task hand-held communication devices and failure to follow local traffic regulations and ignorance of traffic signs/signals are the most factors that caused traffic accidents. A study conducted by Shawky in Cairo that looked at the factors affecting lane change crashes addressed that the results show that the occurrence of lane change crashes is significantly related to driver's personal factors such gender, nationality, years of experience, and the location plus the environmental factors (surrounding condition) such as junction/non-junction location, light and road surface condition and the roads features such as the road type, the number of lanes and the speed limit (Shawky, 2020). The results also show that the participants look at the side mirrors before making the lane change are 4.61 times less likely to get involved into a lane change related crash than others, and the participants look outside the window before making a lane change are 3.85 times less likely to get into a lane change related crash than others. The results of this study showed that the number of lanes is statistically significant to the occurrence of lane change crashes. One of the differences, however, between Shawky's study in 2020 and this lane changing study is that Shawky investigated the factors affecting lane change crashes, and this study investigates the factors affecting lane changes in general, not just crashes.

2.3 Driver distractions and driving performance

Driver distractions can be one of the major cause for unsafe lane changes, the previous studies on the cause for driver distractions such as hand-held devices, the effect of hand-held devices on driving performance, driver's attentions, and the methods to measure driver's performance are reviewed in this section.

The growing development of personal devices and the increasing usage of these technologies have led to a more and more complex in-vehicle information system which results in the concerns for drivers' distractions over the usage of these technologies. Driver distraction can be considered as a diversion of attention from the driving operation, it usually causes a delay in recognition and information processing that is required for performing the driving operation in a safe manner (Khushaba et al., 2013; Jin, 2015).

Many countries have taken account into how important it is to forbid the usage of secondary task related hand-held device during driving operation, according to the World Health Organization in 2011, United States, China, and Canada have all established regulations against the usage of secondary task related hand-held devices such as cell-phones, hand-held GPS service devices when driving. (WHO, 2011)

In 2005, Jin conducted an experiment that uses a novel method to evaluate in-vehicle secondary task driving safety. There are in total five distracting tasks during the driving operation: (Jin, 2015)

1. Tuning the radio on the infotainment system to a specific local station
2. Navigating through the cell phone to select a certain song
3. Talking with a lab assistant
4. Answering a phone call through Bluetooth headphones
5. Navigating through an iPad 4 to find the map app

Forty young participants finished the driving task inside a driving simulator. The measurements of fixations, saccades, and number of blinking were collected and processed. The measurement of saccade is the process of eye movements measured in terms of average saccade speed, amplitude, number of saccade. Results show that there is a significant difference between the baseline and the distracted condition in the eye movement related measurements, and distracted driving caused by secondary tasks greatly reduces driving attention on the roads.

In 2010, Madden conducted a survey on adults and cell phone distractions that investigated the population of adults that text and drive. The primary finding is almost half of the adults that text, reported that they have sent out or they have received a cell phone text message while driving. And nearly half of the adults said that they have been in a car when the driver utilized the cell phone in a way that put themselves and other in danger (Madden, 2010).

The attention of drivers sometime can be influenced by over saturation of information in the view, in 2013, Park and their team conducted a study on a system that implements augmented reality heads-up display into vehicle in order to better convey the driving information for safety purposes. The system achieved a 73% recognition ratio towards the driving safety related information and a frame rate at 15 frame per second for both on road vehicles and pedestrians. The studies stated that sometimes the distraction of drivers are caused by too much information displayed in front of the driver, therefore a concise and filtering system too process the information before showing to the human drivers is the next step in the field. (Park, 2013)

In 2002, Recarte and Nunes investigated the relationship between attention and speed control. They suspected that drivers need more attention to maintain their speed when there is a posted speed limit on the road compared to without a posted speed limit, such as more glance on the speedometer. A secondary task is performed to test the hypothesis, it is found when a secondary task is performed, it does not matter if the speedometer is visible to the driver or not, the speed increased when there is a posted speed limit, and no speed increase under no speed limit condition (Recarte, 2002). In the highway lane change observational study, the factor traffic speed is suspected to be one of the factors that can be significantly influential to the possibility of drivers changing lanes because this study in 2002 shows that a secondary task can affect speed and speed control stability.

A study published in 2013 that looked at the influence of drive while texting between speech-based texting and handheld device texting showed that, no matter it's speech-based texting or handheld device texting, driving performance is impaired by both on different level. The speech-based texting has a less effect on driving performance than the handheld device texting when both texting task required the participants to perform the same front vehicle following task. The handheld device texting caused more variation on the lane position and the speed control than the speech-based texting type. The study concluded that both types of texting while driving impaired the driving performance but the handheld device texting is more mentally demanding and require more attention, so the drivers cannot focus on maintaining the speed and lane position as well as the speech-based texting. (He, 2013)

The texting while driving task also has a psychosocial influence on teenager's intention to text since there are many teenagers doing so. A study conducted in 2010 investigating the psychosocial influences on why young cell phone users would like to texting while driving, the study measured a number of parameters that includes attitudes, norm, intentions, social influence measure of group and moral norm. The study found attitude predicted intentions of sending and reading text messages while driving, intention predicted texting behaviors and both group and moral norm improved the predicting ability of the model. (Nemme, 2010)

In 2015, Hill performed an anonymous survey towards the college students to identify driving distractions, 4964 responses were collected and 91% of the responses reported that they have done either texting or phone calls when driving, 90% reported that they have done texting when driving, and 87% reported that they have done texting when waiting for the traffic light. The phenomenon of text while driving is getting more and more popular, especially among college students who consider themselves as good at multitasking. And the majority of the students think that the policy would change their behavior towards text while driving. (Hill, 2015)

One way to minimize driver distraction and bring their attention on the road is by utilizing the augmented reality through heads-up displays on the windshield, one study conducted by scientists and researcher from Honda Research Institute in Mountain View, California and Stanford University, California and Max-Plank-Institute fur Informatik of Germany looked at the user-centered design and user centered perspective for AR(augmented reality) in car. The study argued that augmented reality in the field of automobile contains the potential of changing the driver's driving experience significantly, the development of augmented reality show follow the principle of user-centered perspective. A user-centered process should be developed for the development of augmented reality so that the AR can better serve the role in between the technologies and the driver. Three pieces of perspectives should be considered as the nature of this development process: the understanding of human perception, the understanding of driver distractions, and the understanding of human driver behaviors. It also stated that the design of augment reality should focus on display the immediate information to the driver, rather than the secondary tasks to decrease the driver's distraction and improve driver's judgement and driving performance. (Bark, 2013)

Another study conducted in 1991 on the subject of driver distractions performed experiments on the effects of telephone usage during driving, it performed the experiment in 3 different traffic environments, light traffic motorway, heavy traffic on a 4-lane road, and city traffic. Twelve participants were gathered for the experiment, none of them were familiar with the telephone. The results of the study shown that the effect of the telephone usage during driving significantly influenced the driver's performance compared to normal driving, such as not using a telephone. Furthermore, the participants who performed the telephone task with a handheld telephone resulted in worse performance than those who performed the telephone task with the handsfree telephone set. The participants who performed the driving task with the handsfree telephone had better control over the testing vehicle, the performance of the drivers were measured in terms of the steering wheel's movements. (Brookhuis, 1991)

Driver performance can also be affect by the temperature of the environment which can caused the body temperature to raise and drop. Schmidt conducted a study that looked at the short-term cooling effect on driver's performance and decision making during a simulated driving experiment. The study argued that cognitive fatigue can potentially lead to discomfort during driving and impair the safety of the driver and the passengers, as well as the on road traffic. The method to counter this fatigue is required, however, there is hardly any preexisting study that looks at the method to measure the effect of cooling when driving. A simulator driving experiment with 34 participants is performed to collect psychological and the data of the vehicle, the cooling condition and the control condition were compared. The results shown that the cooling condition provide an environment that allows the participants to have less sleepiness and better alternes, the cooling measurement has a positive short-term effect on driver's wakefulness and a three minute cooling has the best result. (Schmidt, 2017)

Another study published by Nunes in 2002 looked at the driving performance and distractions found when drivers perform secondary tasks such as a conversation during driving, there is no difference found between talking on the hands-free phone and talking in a live conversation with a person in the vehicle, but there is a difference found in driving performance when a more demanding cognitive task is performed. The driving performance is measured in terms of driving speed, visual search behavior, visual detection, and response selection capacities. The study concluded that as long as the hand-held cell phone is no longer needed to be held during driving, a secondary task like a conversation does not affect driving performance and it is the same as talking to a person. However, the only risk is the level of demand of the secondary task. (Nunes, 2002)

2.4 Autonomous driving safety and lane changes

A study conducted by Khelfa at the University of Wuppertal in 2020 proposed a dynamic safety analysis for the trajectory planning of automated vehicles on highway and also to predict any lane-changes, it mentioned that it is crucial now for the industry to develop a method to model drivers behavior on multiple-lane roads and a way to predict the lane-changing events for the traffic safety.

Advanced driver-assistance system (ADAS) is an electronic system that assists drivers in driving and parking situations though a human-machine interface to increase car and road safety, it utilizes sensors and cameras to detect nearby objects and respond to drivers errors. One of the functions inside this system called Adaptive Cruise Control (ACC) that can automatically control the vehicle speed to maintain the safe distance between the car and the front moving vehicle has been implemented widely in modern cars, it has been noticed that during lane-changing events, such as the driver moving into another lane or another vehicle cut-in into the driver's lane, drivers always have to readjust their speed and spacing in a specific way to maintain the safe distance (Carvalho, 2016). Therefore, ACC systems must work with some kind of lane changing prediction algorithms in order to fully operate safe cruising control on highways (Khelfa, 2020).

Khelfa states that such integration of ACC and lane-change prediction model can solve issues such as the overbraking that can cause uncomfortable jerking of the vehicle and also creating an unnecessary spacing between the driver and the front car (Khelfa, 2020), Khelfa also shows that the distance difference between the driver and the surrounding vehicles and the speed difference between the driver and the surrounding vehicles can be used as explanatory variables as input data in lane-changing estimations in the preliminary data analysis.

The term safe is ambiguous, it is a relative concept which depends on the observer's personal perception (Reschak, 2016). According to ISO 26262 standard, for an automated car, it means the possibility of accidents and the severity of the personal injury should be minimized and predicted under all of the circumstances in which the course of action that can lead to a reasonable and acceptable level of risk must be identified.

In 2016, Reschak wrote in the publication about safety concepts for autonomous vehicles, there is currently no robust and reliable safety concept that can meet all requirements of vehicles without a safety driver on public roads. The study reviewed both vehicles in product and in an experimental stage that presents high-level autonomous driving functions including adaptive cruise control and fully autonomous driving. One example derived from the study is the Mercedes-Benz model that has a system called "Distronic Plus with Steering Assist and Stop & Go Pilot" which provides assistance to the driver in both lateral and longitude directions. However, the drivers are still required to maintain attention on the roads and react accordingly, and the steering assistance system turns itself off if the drivers do not put their hands back on the steering wheel after a certain amount of time. Such a system is considered as a semi-automated system or as partial automation. Another example raised by the study about the safety concepts for autonomous vehicles is the self-driving car project that Google has been working on, as of the stage in 2016, Google has been testing their systems on series production vehicles that were

equipped with sensors and cameras that can safely operate on the public roads in Nevada and California, US. Although the information leaked from the development and research at Google is not much, it seems without using a safety driver at present is still not feasible. There are simply too many situations that make the processing power and robustness of the algorithm too demanding for the system, the safe state that was mentioned before still means that it is a must for the driver to take control of the vehicle. A prototype has been shown in 2014 which can be classified as fully automated car without the driving factors of drivers and other assistance, however, there is not vehicle has been released to the public roads as of right now.
One study in the field of automated driving that investigated the automated driving to manual driving take-over time found that the key factor that determines the time of take-over when human drivers take-over the vehicle from the automated system is the cognitive and not motor process that affect the performance. (Zeeb, 2015)

2.5 Traffic density

The topic of traffic density was first raised before 2000 in traffic data analysis work. In the field of traffic data analysis, there are three types of data that is related to each other, speed, flow and density. The speed is simply how much time an object consume to move within a certain amount of space, the flow is simply how many objects reach a certain point on a road, described in terms of how many objects per hour. And the traffic density is the term that is used to explain how close the vehicles/objects are, whether it is dense or spread out, it is also normally measured within a certain length of a road, and can be called traffic density or traffic occupancy. (DOT, 2000)

In the definition published by the department of transportation in the April of 2000, the traffic density is measured by using the flow of the traffic divided by the speed of the traffic. The unit of flow is the number of vehicles per hour and the speed unit is miles per hour, the unit of the traffic density is vehicles per mile. The traditional and widely used method of calculating the traffic density is by count how many vehicles within a mile which is the number of vehicle per mile.

Al-Sobky conducted a research that look at traffic density determination with cell phones, the study provided the method for traffic density calculation by using cell phone data. It stated the method to calculate the traffic density is by counting the number of cell phone within a certain length segment, use each recorded count to divide the length segment to get the density of the traffic. (Al-Sobky, 2016)

In the study published by Benjamin Coifman of Ohio State University in 2003 that looked at the traffic density and lane inflow on a freeway, Coifman utilized the arrival of vehicles at two detector stations within a designated space to measure the traffic density (Coifman, 2003). This measurement is one step over the traditional method of calculating the traffic density.

One study that looked at the relationship between driver performance and the road complexity, the traffic density for example, stated that traffic density is shown to be one of the traffic complexity that affect driver’s workload. Teh examined the effect of traffic complexity on driver’s workload and driving performance in a dynamic traffic setting. The study shown that the driver’s workload increases when a front vehicle makes a lane-change maneuver inside driver field of view, and when the lane change occurred at a closer distance, the workload increase as well. Driver’s performance is measured in terms of speed change, lateral position variation and mean time headway. (Teh, 2014)

CHAPTER 3. STUDY 1 – ENVIRONMENTAL FACTORS

3.1 Method of the first study

Study one aims to investigate if there is any correlation between the environmental factors on highways and the possibility of lane changes occurring. We do not know if drivers make more lane changes under certain conditions, such conditions, if detected by an AV, would help the vehicle to execute appropriate actions beforehand to mitigate unforeseeable risks accordingly, and improve the overall highway safety. This study collects data through the highway surveillance system of CA, analyzes the observational data, and performs a full factorial design, reporting findings on the significant correlations at the end.

Research question:
Are lane changes more frequent under certain conditions?

3.1.1 Environmental factors

The first step to investigate the correlation between the environmental factors and the possibility of lane changes occurring is to list all the possible environmental factors with their levels. There are five factors listed in this study that are suspected to be significantly correlated to the possibility of lane changes occurring, they are traffic speed, the number of lanes, traffic density, traffic flow and time of day, each one is explained in the following paragraphs.

Traffic speed. The levels of this factor are high and low, this study suspects that drivers make more lane changes at a lower speed compared to at a higher speed. The reason is that when a vehicle is moving at a higher speed, drivers receive more information through visual and audio cues comparing to moving at a lower speed in a same time period, therefore, drivers bear heavier mental load at a higher speed than at a lower speed in order to process all the information to execute appropriate actions. Making a lane-change under a high-speed situation is one extra burden on top of the existing mental load, this study suspects that under a higher speed situation the drivers would only perform a lane-change when it is necessary to do so while making a lane-changing under a lower speed situation is much less mentally stressful therefore lane changing maneuver might occur not just limited to mandatory intentions. The high and low level of this factor can be distinguished after the sampling is done and the clear divider has been determined in the data collection phase.

Traffic density. The levels of this factor are high and low which are high traffic density and low traffic density. This study suspects that drivers make more lane-changes in lower density traffic rather than high density traffic. The reason is that when on a low-density traffic, drivers have more available space and opportunities to make lane changes than driving on a high-density traffic where it's more crowded. The other reason is that on a high-density traffic, drivers will have to keep the distance away from the other vehicles and this task requires a substantial amount of mental load to maintain the safe distance. Lane changing itself is one extra burden on top of the existing mental load, so this study suspects that drivers should do less discretionary lane changes on high density traffic compared to driving on a low-density traffic. The high and low level of this factor can be distinguished after the sampling is done and the clear divider has been determined in the data collection phase.

Traffic flow. The levels of this Factor are high and low which are unimpeded traffic and impeded traffic. This study suspects that drivers make more lane changes in impeded traffic than in unimpeded traffic, the reason is that in a stop and go traffic, drivers do not patiently stay in one lane, in order to optimize their routes, drivers are constantly moving to a quicker lane as soon as a feasible gap is found. A lot of times, this kind of stop and go traffic lane changes are the most dangerous ones. However, in an unimpeded traffic, when the traffic speed is high, drivers normally don't feel the need to change lanes, unless it's mandatory lane changes. When the traffic speed is low in an unimpeded traffic, making lane changes is still not as easy as in an impeded traffic due to less gaps. The high level and the low level of this factor can be distinguished by looking at the traffic feed with naked eyes, if the traffic is in a "stop-and-go" status with the majority of the vehicles braking constantly, then mark it down as impeded. Otherwise unimpeded if the traffic is running smoothly, no matter the speed.

Time of day. The levels of this factor are high and low which are rush hours and non-rush hours in the bay area of California. This study suspects that drivers make more unsafe lane changes during rush hour than during the non-rush hour, the reason is that during rush hours drivers may tend to do discretionary lane changes more often than mandatory lane changes, the concept of discretionary/mandatory lane changes were discussed in the background and literature review chapter of this report. The nature of discretionary lane changes is that driver change lanes based on their needs to reach their destination, it can be either reach the destination as quickly as possible, or drive at a speed that the driver is most comfortable with. However, most of the drivers are in a rush to reach their destination during the rush hours, thus, more discretionary lane changes may happen and lead to overall more lane changes during the rush hour than the non-rush hours. The other reason is that, the rush hour time in the bay area of California is 7-10 am and 4-7 pm, the non-rush time is the rest of the times. This study suspect that drivers make less lane changes during the nighttime due to the fact that, the traffic density during the nighttime should be lower than the daytime, and the drivers can usually see further and clearer during the daytime than the nighttime, thus they should have less confidence to make lane changes during nighttime than daytime., less confidence leads to less overall lane changes. The high level and the low level of this factor can be distinguished from the video feed, the local time is shown on the website below the video feed as in figure 4 (just an example).

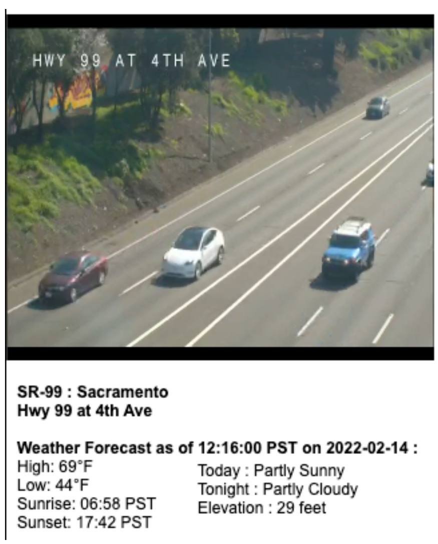


Figure 4. Traffic camera footage screenshot

The number of lanes. The levels of this factor are high and low, which are 4-lane highway and 2-lane highway, this study suspects that drivers make more lane changes on a 4-lane highway than on a 2-lane highway. The reason is compared to driving on a 2-lane highway, drivers feel more open wide on a 4-lane highway where there is more available space and opportunities to make lane-changes. Furthermore, driving on a 4-lane highway requires more lane-changing maneuvers than driving on a 2-lane highway to achieve the same goal, such as exiting the highway on the right lane, a vehicle on the fast-lane on the left must move to the right lane in advance, on a 2-lane highway it is only 1 lane-changing maneuver, on a 4-lane highway it is 3 lane-changing maneuvers. And Shawky concluded in 2020 that the number of lane changes is showing as one of the factors that affect the occurrence of lane changes crashes significantly (Shawky,2020).The higher level and the lower level of this factor can be easily distinguished by counting the number of lanes with naked eyes on the video feed.

3.1.2 Traffic speed calculation and cluster analysis

The method to estimate the traffic speed by observing highway traffic is very straightforward, the distance that the vehicle travels down the road can be measured by counting how many dashed lines the vehicle has passed. The length of dashed lines on highways are regulated by the US department of transportation, each dashed line on the highway is 10 feet and each gap between the dashed lines is 30 feet, so one dashed line plus one gap is 40 feet in total. The T(time) that the vehicle has traveled can be simply measured by using a stopwatch, when the L(distance) and the T(time) are known, the V (traffic speed) can be estimated by:

$$V=L/T$$

Furthermore, in order to categorize the traffic speed into high and low groups, a cluster analysis must be conducted before the data collection in order to determine the divider that categorize the data collected later into higher and lower groups. First, a standardized way for sampling needs to be established for the cluster analysis. Traffic lights tend to release traffic in clumps, and the pattern of traffic will affect the sampling. Thus, when sampling traffic data, measures that help to prevent the pattern of the traffic influencing the sampling must be established as well. In this study, before the start of sampling, a 2-minute video, a 5-minute video, and a 10-minute video are recorded for each location. Then for each video, a 1-minute clip is randomly chosen from each video. Next, calculate the traffic speed of each one-minute video and compare the traffic speed among the 3 videos. If there is no difference found in traffic speed among all the 3 videos, then just use the 2-minute video. If there is a difference found between the 2-minute video and the 5-minute video and there is no difference between the 5-minute video and the 10-minute video, then just use the 5-minute video. If there is a difference between the 2-minute video and the 5-minute video and there is also a difference between the 5-minute video and the 10-minute video, longer samples need to be collected until the traffic speed is no longer fluctuating which means the length of the video is independent from the flow of the traffic. The reason why the 2-minute video is sufficient when there is no difference found between the 2-minute video and the 5-minute video is because when the 5-minute video has the same traffic speed as the 2-minute video that means the 2-minute video has the same information as the 5-minute video contains. After the length of the video sample has been determined for each selected location, observe different cameras at different locations to sample data points, collect the observed speed for all locations and save them on a spreadsheet.

After the observed speed has been recorded and saved, the next step is to perform the cluster analysis and determine the divider for categorizing data collected later. Due to the fact that each road has different speed limit, the method used in this traffic speed cluster analysis process involves a simple transformation relative to the posted speed limit of that location:

$$Transformed\ speed\ =\ Observed\ speed - posted\ speed\ limit$$

The purpose of this transformation is to help standardize the roadway and make the speed more comparable to each other, since each road has its own posted speed limit. After the observed speed is transformed, take the median of the transformed speed. In this study, the median of the transformed value is -7.4. Due to most of the traffic captured in the bay area of California for the traffic speed cluster analysis purpose are running under the posted speed limit, this value is below zero. This value is used later in the data collection process of traffic speed, to categorize observed speed into higher and lower groups, speed lower than -7.4 is categorized as the lower group, and speed higher than -7.4 is categorized as the higher group. The observed speed in data collection is transformed in regards with the posted speed limit of that roadway as well, so that comparability among different roadway can be met.

3.1.3 Traffic density calculation and cluster analysis

The method of measuring the traffic density by calculating the free space percentage in a certain area introduced in this study, is a slight modification on the basis of the established method in the field to measure traffic density. In previous studies, the widely accepted method to measure traffic density is by counting the number of vehicles per lane per mile. However, this method does not fully fit the needs for this study and a modified measurement of traffic density is developed.

First of all, conducting studies on the topic of lane changing itself requires a multi-lane fashion setting, counting the number of vehicles per lane does not quite meet this requirement. Secondly, different types of vehicles have different body lengths, and the length of vehicles plays a critical role in this study, just counting the number of vehicles and ignoring the different length of them does not fit the requirements of this study. For example, the average length of a sedan/SUV/hatchback is about 15 ft and the average length of a semi-truck is about 70 ft. On a 300 ft road it can fit roughly 20 sedans or 4 semi-trucks, if the number of vehicles is the only variable counted here, 20 vehicles seem to be a lot higher than 4 vehicles, but they are both bumper-to-bumper traffic with roughly the same traffic density, plain numbers cannot show that property. This is the reason why measuring free space instead of measuring the traditional traffic density works better for this study, the free space measurement ignores different types of vehicles and the traditional traffic density measurement does not ignore it. Lastly, the field of view of the traffic cams on the highways cannot reach as far as one mile, and each one of them has different views, zoom, and resolutions. The only way to standardize this is to set a big enough area in each footage, such as 10 units which is about 400 ft, one unit is one dashed line (10 ft) plus one gap (30 ft), then calculate the free space percentage in a certain area so that it can be a standardized variable to be compared across different locations.

The method to calculate the traffic density by observing the highway camera surveillance system is to first calculate how much free space there is on a highway traffic screenshot. For example, on a four-lane highway with 5 observable units' distance (one unit is 40 feet), the total length of the space is 5 * 40 feet * 4 lanes which is 800 feet. Then count how many semi-trucks and how many other vehicles within this five units area, combine all the occupied distance, the rest of the space within the 800 feet is the free space available on this road. The average length of a semi-truck is 70 feet and the average length of other vehicles including sedans, SUVs, pick-ups, and hatchbacks is 15 ft, plug it into the equation and the free space is calculated as:

$$Free\ space = Total\ length - occupied\ length$$

$$Free\ space = n_u * 40 * n_l - 70n_S - 15n_c$$

$where, n_u = the\ number\ of\ units$

$n_l = the\ number\ of\ lanes$

$n_S = the\ number\ of\ semitrucks$

$n_c = the\ number\ of\ the\ average\ length\ vehicles$

Since different cameras have different fidelity, different field of view, and different light quality, there is no guarantee that the observable total length can be the same across all the cameras, a standardized measurement must be used in order to compare across all of them, use the free space divided by the total length times 100% is the free space rate of this road which can be a standardized measurement used across all different cameras at different locations.

$$Free\ space\ rate = \frac{Free\ space}{Total\ length} * 100\%$$

$$Traffic\ density = 1 - Free\ space\ rate$$

After the method to calculate the traffic density has been determined, a measure to do a cluster analysis for the traffic density in order to put the observations into high and low groups must be established as well. Similar to the method of performing cluster analysis for the traffic speed, sampling is also required for this part of analysis. It’s very common for traffic to have a pattern, sometimes it's caused by traffic lights which tend to release traffic in clumps. A measure to make sure that the way the screenshots are taken is independent from the flow of the traffic.

First, for each location selected, the length of the video must be determined in the previous part before working on this part of the cluster analysis. Then use the same time, day, and week that was used in determining the length of the video for that location and record a short clip at a certain time length determined in the previous part.

Take screenshots in three different ways. The first way; take two screenshots, one at the beginning of the video and the other one at the end of the video. The second way; take 3 screenshots, one at the beginning, the second one in the middle, and the third one in the end. The third way; take 5 screenshots, at 0%, at 25%, at 50%, at 75%, and at 100% of the video. Calculate the traffic density for every screenshot, then take the median to determine the traffic density of each way of taking screenshots. If there is no difference found between the three ways then just use the first one, if there is a difference found between the first one and the second one and there's no difference found between the second one and the third one, then just use the second one. If there is a difference found between the first one and the second one, and also a difference found between the second one and the third one, more frequent screenshots must be taken until the traffic density is no longer fluctuating and the way the screenshots are taken is independent from the flow of the traffic.

Similar to performing a cluster analysis in the previous part that determined the divider for traffic speed, save the observed traffic density in terms of free space rate in percentage on a spreadsheet for the next step of analysis.

This process is not exactly the same as the traffic speed cluster analysis, there is no transformation performed. The value of observed traffic density is recorded in a column, take the median of that column, and the divider is determined. The divider in this study for traffic density is at 9.38%, any traffic density collected in the data collection phase higher than this rate is categorized as the higher group, any traffic density collected in the data collection phase lower than this rate is categorized as the lower group.

3.1.4 Data collection process of the first study

After the factors, levels, and the divider for categorizing traffic speed and traffic density are determined, the next step is to determine the method of data collection. Firstly, the state of California has the most advanced highway traffic surveillance system around the United States, it is publicly accessible and the high-volume traffic in California is very desirable for the purpose of this research. Secondly, the ultimate goal of this study is to contribute to the overall highway safety of the autonomous driving industry, and currently, the most advanced autonomous driving technology in the world exists in the state of California, so utilizing the traffic data in California can provide direct results that researchers and scientists can refer to without making assumptions or further investigations. Therefore, this study utilizes the public highway traffic surveillance system of California to collect public traffic data by observing the video feed.

This study designed a 2^5 factorial experiment and the design table is attached below:

Table 1. DOE table

#	traffic speed (high+/low-)	number of lanes (4+/2-)	traffic density (high+/low-)	traffic flow (unimpeded+/ impeded-)	time of day (rush+/ nonrush-)	number of lane changes
1	0	0	1	0	1	
2	1	0	1	1	1	
3	1	1	1	0	1	
4	1	0	1	0	0	
5	0	1	1	0	1	
6	0	0	0	0	1	
7	1	1	1	1	1	
8	0	1	0	1	1	
9	1	0	1	0	1	
10	0	0	1	1	1	
11	0	1	0	1	0	
12	0	1	0	0	0	
13	1	1	0	1	0	
14	0	0	0	1	0	
15	0	0	0	0	0	
16	1	1	0	0	1	
17	1	0	0	0	1	
18	0	0	1	1	0	

Table 1 continued

19	0	0	0	1	1	
20	0	1	1	0	0	
21	0	1	1	1	0	
22	0	1	1	1	1	
23	1	1	0	0	0	
24	0	1	0	0	1	
25	1	0	0	1	0	
26	1	0	1	1	0	
27	1	1	0	1	1	
28	1	1	1	0	0	
29	0	0	1	0	0	
30	1	1	1	1	0	
31	1	0	0	1	1	
32	1	0	0	0	0	

The rows colored in grey are the combinations that are impossible to happen, and it is also noted in the last column on the right as "NULL". For example, for combination #3, high traffic speed, 4 lanes, high traffic density, impeded traffic and rush hour time, it is impossible for the high traffic speed and impeded traffic to happen together. When the traffic speed is high, the traffic cannot be impeded. Thus, 12 combinations are removed from the factorial design table, and the factorial design is not a full one. Besides the five factors on the left side of the table, the 2 columns on the right are the number of lane changes, and the situation might occur. The number of lane changes is the response variable in this observational experiment, it is measured as counted data.

In the process of actual data collection, the steps are listed as follow:

1. Use the factorial design table (Table 1) as the basis for data collection, there are a total of 32 combinations from the independent factors, with 20 possible combinations.
2. Use the CA public surveillance website: https://cwwp2.dot.ca.gov, and pick 32 locations with high quality cameras with 4 or 2 lanes. (a screenshot of that traffic surveillance website is shown in figure 5)

a. The cameras that are pointing at the highway

b. The cameras with good resolution and lighting, 6 units must be visible and the dashed line must be distinguishable from each other

3. For each highway camera location, screen-record for 2 minutes the whole width of the highway and at least 6 visible units. Each unit is defined as one dash plus one break.
4. While screen-recording the 6 units, mark the start/end of the unit with the mouse so that the same 6 units are used for later calculations.
5. While screen-recording, count the number of lane changes within 6 units, and record this number on an excel sheet for all locations, each as its own row.
6. After screen-recording for all locations, go through the recorded videos and calculate the factors from the factorial design table using excel. The steps for calculating traffic density and traffic speed is explained in the previous sub-sections.

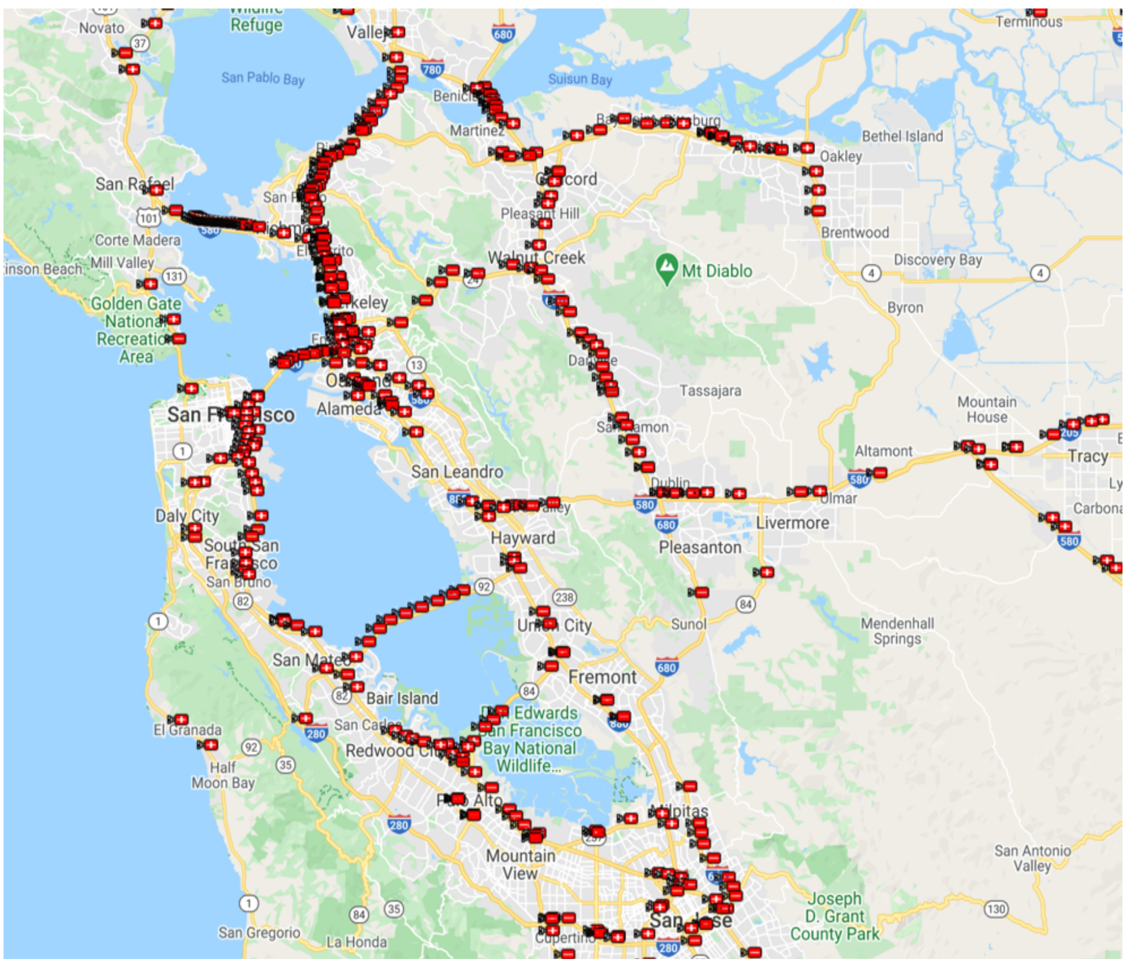


Figure 5. Caltrans map screenshot of the Bay area

7. Using the factorial design table, find matching combinations from the excel sheet and record its corresponding lane change number in the design table. Only do this for possible combinations as marked in the design table.
8. If there are combinations from the factorial design table that are not filled in, revisit the surveillance website at another time and:
    a. Find a camera that has similar conditions as the missing combination
    b. screen-record 2 minutes of at least 6 units & categorize the five factors
    c. record the number of lane changes and the value of the factors on the excel sheet separately
9. Find the matching combinations from the new data collected in step 8 with the missing combinations in the factorial design table. Repeat step 8 if necessary until all combinations are filled in the factorial design table.
10. After the factorial design table has been filled out, repeat steps 1 through 9 for two more replications, a total of three replications. One replication can be finished in one week, thus this study used a total of 3 weeks for data collection.

The number of replicates is estimated in Minitab 17, figure 5 shows the power calculation to estimate the number of replicates necessary if the target power is at 80%.

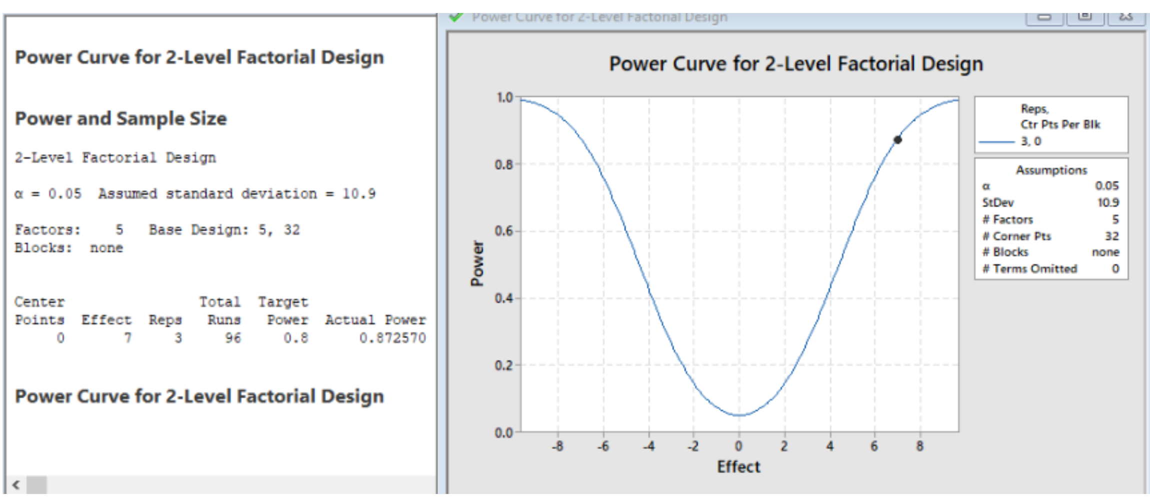


Figure 6. Power calculation

3.1.5 Data analysis of the first study

The data analysis process of the highway environmental factors study starts with data cleaning that looks at the raw data with three replicates, check if there is any cell that is blank in the data sheet and determine the reason for the missing data, whether it is an observer error or it's just because of the fact that there is no lane changes in that 2 minutes window. After all the cells are checked, pull the three replicates into one spreadsheet as Table 2.

Table 2 Number of lane changes data sheet

combination	traffic speed (high+/low-)	number of lanes (4+/2-)	traffic density (high+/low-)	traffic flow (unimpeded+/impeded-)	time of day (rush+/non rush-)	number of lane changes	vehicle counts
1	0	0	1	0	1	1	67
2	1	0	1	1	1	0	72
3	0	1	1	0	1	7	224
4	1	1	1	1	1	5	157
5	0	1	0	1	1	4	163
6	0	0	1	1	1	0	102
7	0	1	0	1	0	0	207
8	1	1	0	1	0	4	173
9	0	0	0	1	0	0	75
10	0	0	1	1	0	1	46
11	0	0	0	1	1	1	59
12	0	1	1	0	0	5	223
13	0	1	1	1	0	5	214
14	0	1	1	1	1	0	134
15	1	0	0	1	0	2	72
16	1	0	1	1	0	0	42
17	1	1	0	1	1	3	160
18	0	0	1	0	0	0	88
19	1	1	1	1	0	0	113
20	1	0	0	1	1	1	21

Table 2 continued

21	0	0	1	0	1	5	65
22	1	0	1	1	1	2	62
23	0	1	1	0	1	3	150
24	1	1	1	1	1	6	188
25	0	1	0	1	1	1	135
26	0	0	1	1	1	2	132
27	0	1	0	1	0	1	129
28	1	1	0	1	0	1	197
29	0	0	0	1	0	0	35
30	0	0	1	1	0	0	68
31	0	0	0	1	1	2	77
32	0	1	1	0	0	4	160
33	0	1	1	1	0	8	182
34	0	1	1	1	1	5	179
35	1	0	0	1	0	1	14
36	1	0	1	1	0	0	46
37	1	1	0	1	1	4	168
38	0	0	1	0	0	1	117
39	1	1	1	1	0	7	189
40	1	0	0	1	1	0	22
41	0	0	1	0	1	2	68
42	1	0	1	1	1	1	118
43	0	1	1	0	1	4	165
44	1	1	1	1	1	4	241
45	0	1	0	1	1	0	105
46	0	0	1	1	1	1	91
47	0	1	0	1	0	3	251
48	1	1	0	1	0	1	96
49	0	0	0	1	0	3	57
50	0	0	1	1	0	0	47
51	0	0	0	1	1	1	59
52	0	1	1	0	0	0	105

Table 2 continued

53	0	1	1	1	0	2	162
54	0	1	1	1	1	5	210
55	1	0	0	1	0	1	39
56	1	0	1	1	0	2	50
57	1	1	0	1	1	3	150
58	0	0	1	0	0	1	119
59	1	1	1	1	0	3	154
60	1	0	0	1	1	0	72

As mentioned in the method part of this study, this study performed a 2^5 factorial design but not all the combinations are possible to achieve in reality. This table above contains all the lane changes data collected within the 3 weeks of data collection process, row 1-20, 21-40, and 41-60 each part represents 1 replicate of 20 combinations, and after the 3 replicates are pulled together, it's 60 rows of data in total. The reason for pulling the replicates together into one analysis is to increase the statistical power of the analysis, and it is a common practice in statistical analysis.

This study used Poisson regression to model the response variable since the response variable is the number of lane changes, it's non-negative count data, and Poisson regression is the most suitable one for this type of response variable. Along with the regression analysis, no issue was found in the normal probability plot as shown in figure 8.

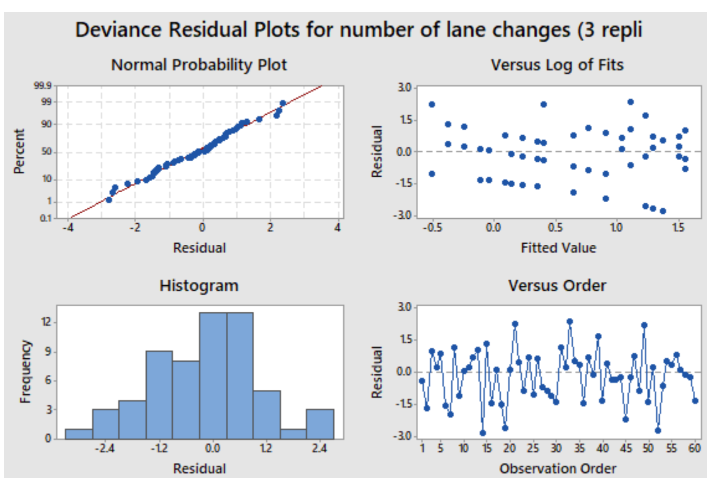


Figure 7. Four-in-one plot of Poisson regression

Besides the normal probability plot, the dispersion parameter for negative binomial is at 8.3955 which shows the model should have properly estimated errors and P-values, shown in figure 6. After the Poisson regression analysis is done, the main effects plot and the interactions plot is also created to check how does the number of lane changes variate in respect with the change between lower and higher levels of the five environmental factors.

The interactions are also checked in data analysis, both in interactions plot and in regression analysis, the results are shown in the next part of this report.

During the analysis, a strong correlation between the number of lanes and the volume of traffic within the 2 mins time period is suspected, in order to see the correlation between these two factors for better understanding of the relationship among the factors, a new variable called "vehicle counts" is measured by going back to the recorded video and manually counted, it records the number of vehicles that traveled through the 6 units space that other factors used within the 2 mins time period, and the data is stored in Table 3. Besides recording the continuous data for vehicle counts, the column was also transformed into text data with a higher and lower level in order to see the interactions between vehicle counts and other factors. The divider for vehicle counts is 115 which is the median, any observations above 115 is in the higher level, vice versa. (Only the continuous data of vehicle counts is included in this report)

3.2 Results of the first study

In total, this study observed 7086 vehicles with 129 lane changes, and the results of this study are shown below:

- Found a strong correlation between the number of lanes and the vehicle counts, but no correlation between vehicle counts with other environmental factors is found
    - For the upper 30 combinations above the median, only 4 of them are 2-lane highways, the rest 26 are all 4-lane highways.
    - For the lower 30 combinations below the median, only 4 of them are 4-lane highways, the rest 26 are all 2-lane highway

- just by looking at the data, without doing any fancy data analysis, I can already see a strong correlation between the number of lanes and the vehicle counts. Then I did the same thing with the rest of the factors, but I do not see any correlations with the other factors.
- The coefficient estimates and the P-value of the number of lanes and vehicle counts further supports this point in Figure 8 and 9.

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-0.2873	0.3718	-0.773	0.4397
TS	0.1147	0.2342	0.490	0.6242
NL	1.1474	0.2256	5.087	3.65e-07
TD	0.4361	0.2374	1.837	0.0662
TF	-0.2005	0.2837	-0.707	0.4798
TOD	0.2706	0.2044	1.324	0.1856

Figure 8 Poisson regression output 1

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-0.4093	0.3378	-1.212	0.2257
TS	0.1598	0.2098	0.762	0.4463
NL	0.5449	0.2856	1.908	0.0564
TD	0.3686	0.2167	1.701	0.0890
TF	-0.2000	0.2488	-0.804	0.4216
TOD	0.1926	0.1832	1.051	0.2930
VC	0.8724	0.3014	2.894	0.0038

Figure 9 Poisson regression output 2

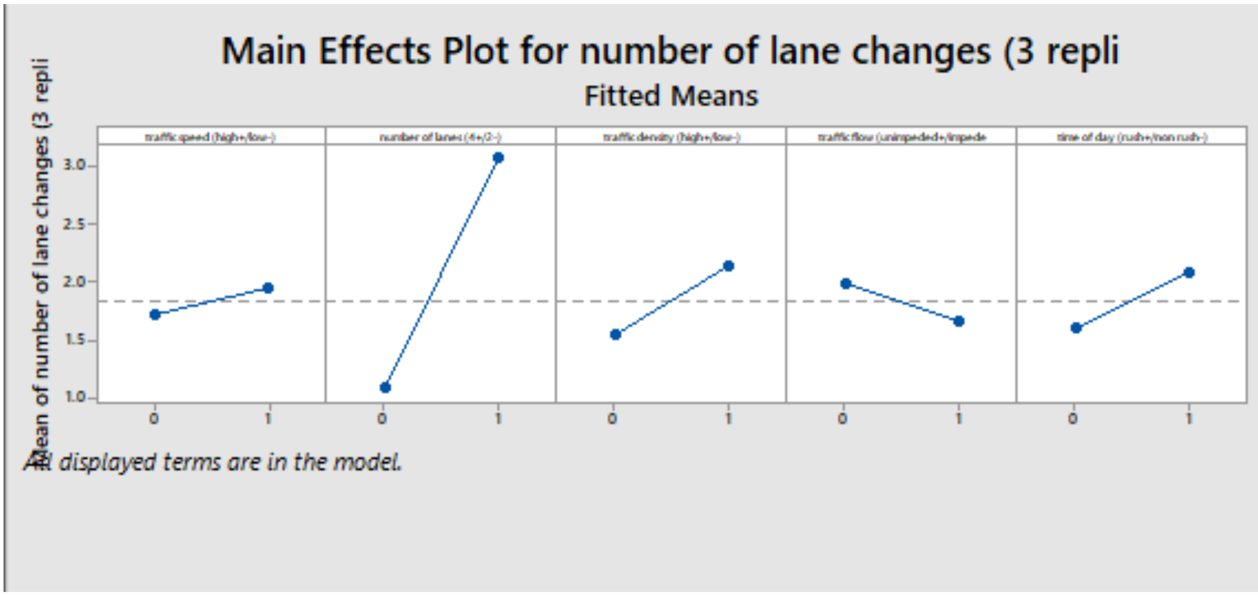


Figure 10. Main Effects Plot for number of lane changes

- In figure 10 and 11, it is clear to see that after the vehicle counts is included, the slope of the number of lanes is not as steep as before. But for the rest of the factors, not a lot of change.
- The aliased effect between all the 5 main factors + Vehicle counts (treated as a main effect) and all the 2-way interactions were checked for aliasing due to some factor combinations not occurring in the data, no aliased effect were found.

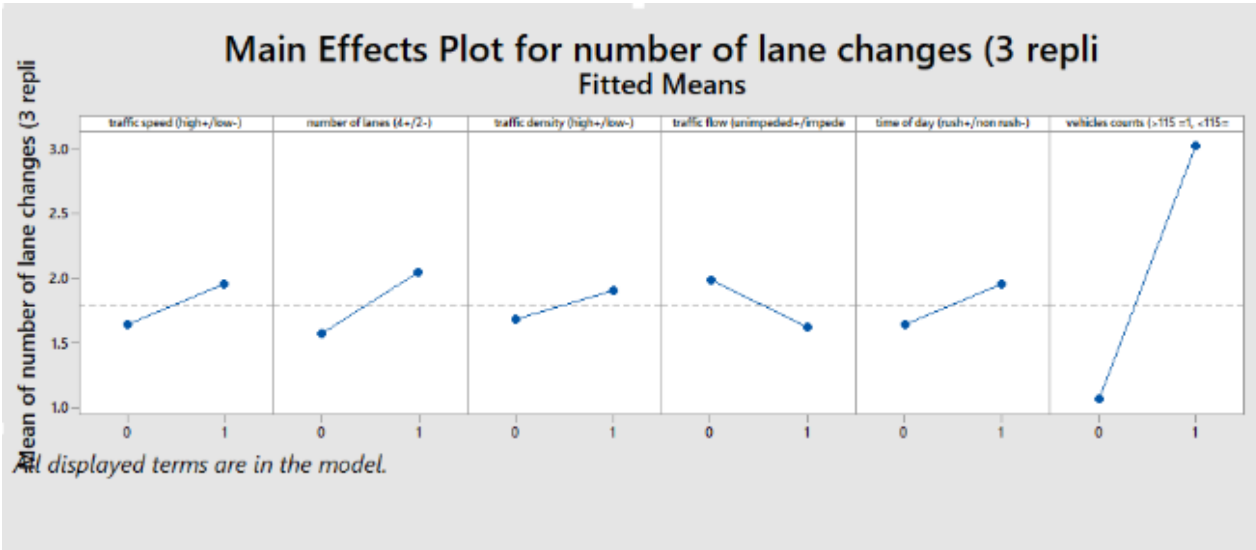


Figure 11 Main Effects Plot with Vehicle counts included

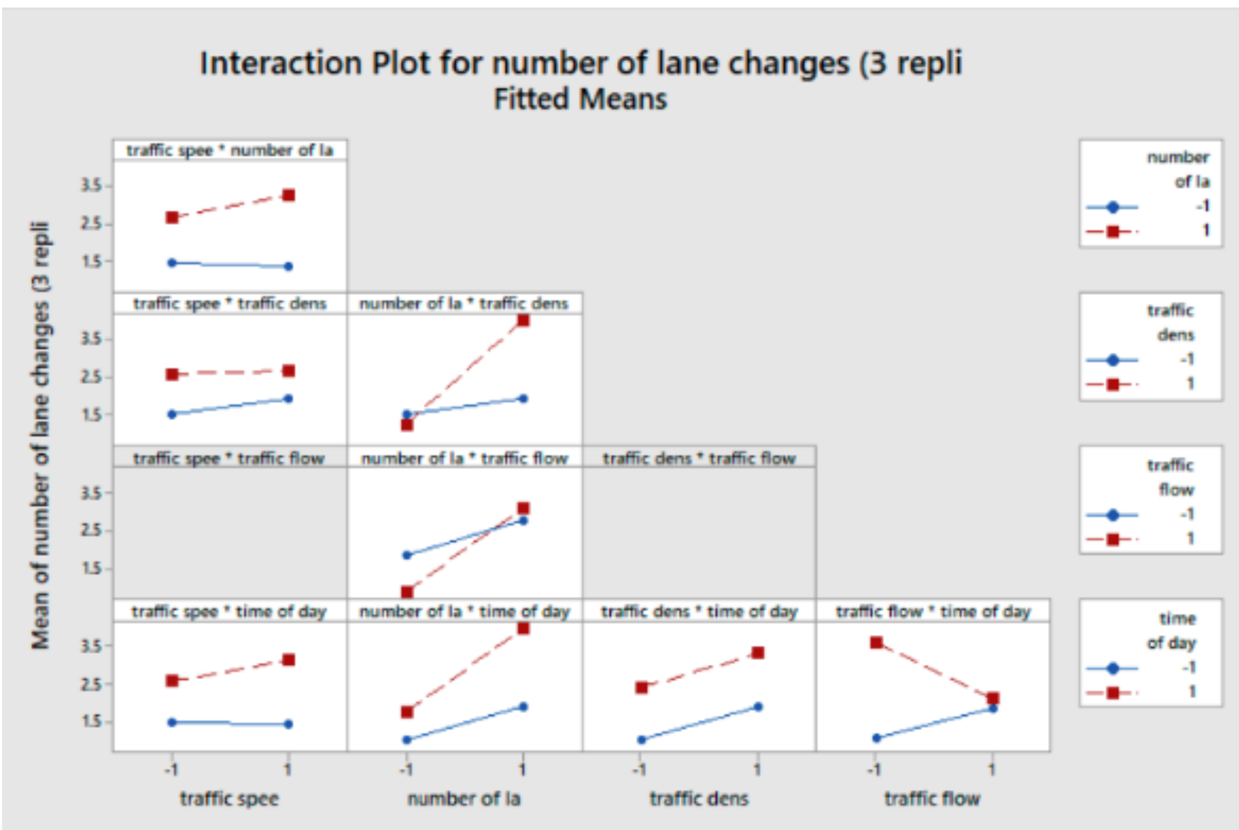


Figure 12 Interactions plot for number of lane changes

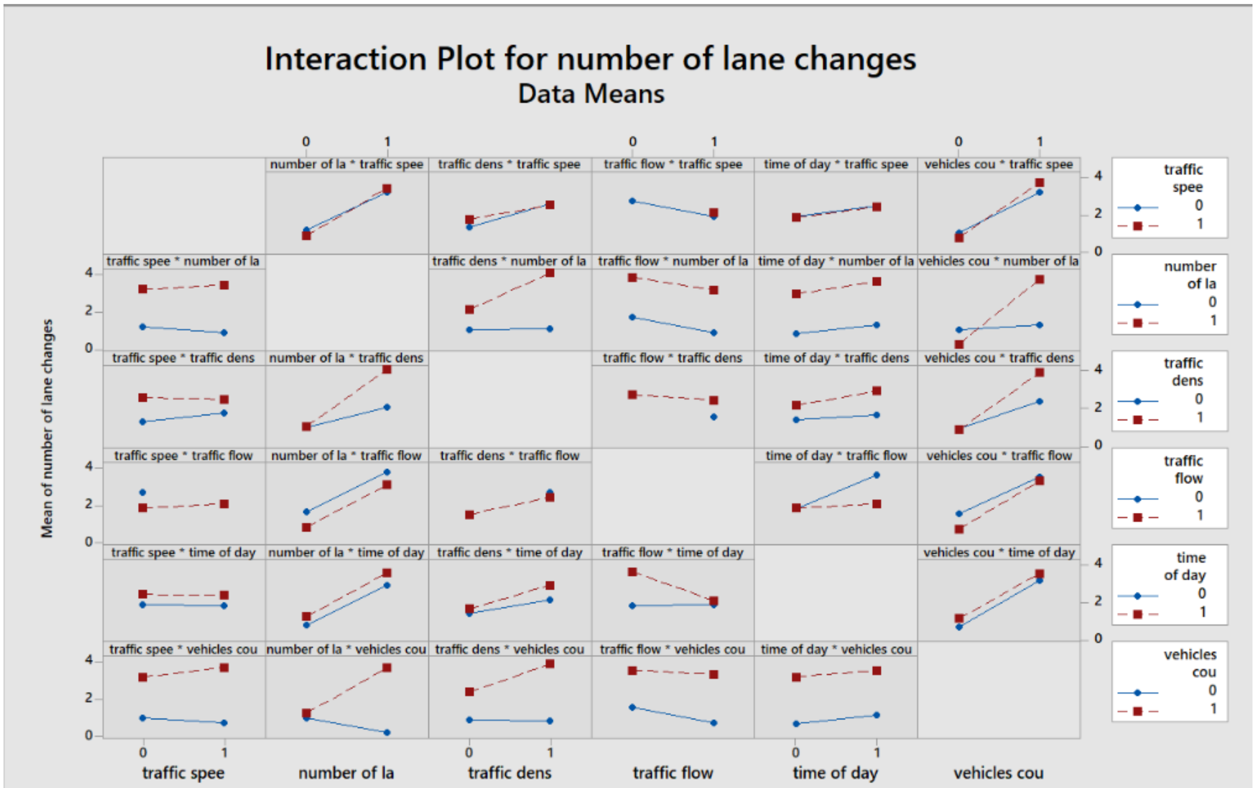


Figure 13. Interaction Plot for number of lane change (with vehicle counts)

- The interactions plot in Figure 12 shows the interactions between factors that worth further investigation, the further investigation of interactions is in the discussion and conclusion section where the suspected reasons for these interactions are discussed.
- The interaction plot in Figure 13 shows the interaction plots after the vehicle counts are included.

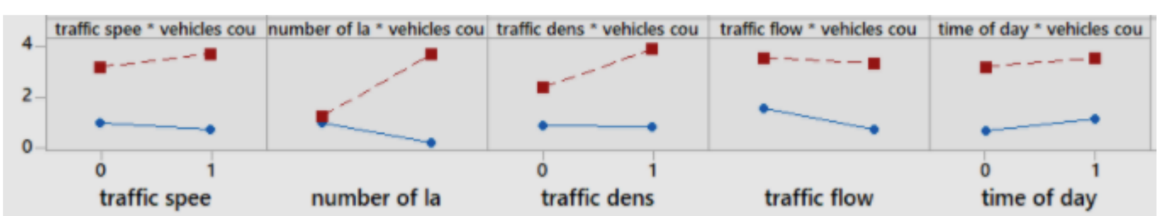


Figure 14 Interactions plot of vehicle counts

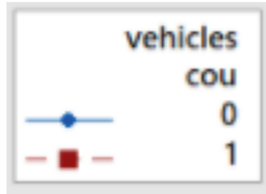


Figure 15 Interactions plot of vehicle counts attachment

- Figure 14 and 15 show the interactions only between the vehicle counts and other factors
- A strong correlation between the number of lanes and the vehicle counts has been found
    - For the upper 30 combinations that have the number of vehicles larger than the average, only 4 of them are 2-lane highways, the rest 26 are all 4-lane highways
    - For the lower 30 combinations that have the number of vehicles smaller than the average, only 4 of them are 4-lane highways, the rest 26 are all 2-lane highways
- No other factors show a strong correlation with the vehicle counts, to illustrate:
    - With respect to traffic speed:
        - For the upper 30 combinations over the average of the number of vehicles, 11 of them are high traffic speed,19 are lower traffic speed
        - For the lower 30 combinations below the average of the number of vehicles, 10 of them are high traffic speed, 20 are lower traffic speed
    - With respect to time of day:
        - For the upper 30 combinations over the average of the number of vehicles, 16 of them are in rush hours, 14 are in non-rush hours
        - For the lower 30 combinations below the average of the number of vehicles, 14 of them are in rush hours, 16 are in non-rush hours
    - The same thing with traffic density, and traffic flow, no strong correlation with the vehicle counts has been found
- The interactions plot between the vehicle counts and the rest of the factors are shown in Figure 14.
    - The interactions plot between the vehicle counts and the rest of the factors are almost the same as the interactions plot between the number of lanes and the rest of the factors

```
Call:
glm.nb(formula = NLC ~ TS + NL + TD + TF + TOD + VC + NL * TD +
    NL * VC, data = pooled_data, init.theta = 12986.54234, link = log)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.8565  -1.0231  -0.0940   0.6445   2.5994

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.10406    0.37529   0.277   0.7816
TS           0.19710    0.20550   0.959   0.3375
NL          -1.74170    1.05503  -1.651   0.0988 .
TD          -0.05141    0.39993  -0.129   0.8977
TF          -0.26834    0.24303  -1.104   0.2695
TOD          0.11827    0.17913   0.660   0.5091
VC           0.19356    0.52263   0.370   0.7111
NL:TD        0.60255    0.45676   1.319   0.1871
NL:VC        2.44923    1.13781   2.153   0.0314 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for Negative Binomial(12986.54) family taken to be 1)

    Null deviance: 133.81  on 59  degrees of freedom
Residual deviance:  66.84  on 51  degrees of freedom
AIC: 208.79

Number of Fisher Scoring iterations: 1


              Theta:  12987
          Std. Err.:  226812
Warning while fitting theta: iteration limit reached

 2 x log-likelihood:  -188.786
```

Figure 16. Poisson regression analysis output 3

```
Call:
glm.nb(formula = NLC ~ TS + NL + TD + TF + TOD + VC + NL * VC,
    data = pooled_data, init.theta = 12077.90833, link = log)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.8040  -1.0550  -0.1072   0.5229   2.3158

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.164664   0.336223  -0.490   0.6243
TS           0.193076   0.205514   0.939   0.3475
NL          -1.376439   1.021819  -1.347   0.1780
TD           0.402256   0.211132   1.905   0.0568 .
TF          -0.269461   0.243153  -1.108   0.2678
TOD          0.121975   0.179140   0.681   0.4959
VC           0.008434   0.494873   0.017   0.9864
NL:VC        2.646979   1.123894   2.355   0.0185 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for Negative Binomial(12077.91) family taken to be 1)

    Null deviance: 133.812  on 59  degrees of freedom
Residual deviance:  68.564  on 52  degrees of freedom
AIC: 208.51

Number of Fisher Scoring iterations: 1


              Theta:  12078
          Std. Err.:  220722
Warning while fitting theta: iteration limit reached

 2 x log-likelihood:  -190.511
```

Figure 17 Poisson regression analysis output 4

3.3 Discussion of the first study

This study aimed to investigate the relationship between the possibility of drivers decision on making lane changes on highways and the conditions of the road environment, if such condition that influence driver’s decision on making a lane changes can be found, it can provide some information for the autonomous vehicles to better understand human driver behaviors and take proactive maneuvers to prevent a possible collision. The previous studies have stated that drivers’ choice of speed is not well understood on freeways (Knoop, 2018), most of the drivers would choose a speed and stick with it and some might stick to a lane and adapt the speed of that lane. Knoop also stated that drivers have different strategies throughout their trips, and the environmental factors are suspected to be one of the elements that influence their decision making. This study suggest that if the drivers choose their speed on freeways based on the environmental factors, it should be reasonable for the drivers to choose their lane or make lane changes based on the environmental factors.

Although not much study was found in the relationship between the environmental factors and the driver’s decision on making lane changes, one lane change crashes study in 2020 conducted by Shawky in Cairo concluded that the occurrence of lane change crashes is significantly related to driver’s personal factors such as gender, nationality, years of experience, and the location plus the environmental factors (surrounding condition) such as junction/non-junction location, light and road surface condition and the roads features such as the road type, the number of lanes and the speed limit (Shawky, 2020). This study does not investigate the factors that affect lane changes crashes, however, the conclusion published by Shawky in 2020 reflects that the environmental factors that might affect driver’s decision on making a lane is worth investigating.

Based on the results, both the number of lanes and the time of day caused the number of lane changes to double its average mean when switched from lower to higher level. The traffic speed, the traffic density, and the traffic flow caused some level of change in the mean of the response when switching from the lower level to the higher level. This result does not fully support what was speculated in the study, both traffic speed and traffic density shows opposite effect on the probability of driver’s making lane changes. The following parts discuss about the interactions effect among the factors, and also how the response variable changed.

Traffic speed: this study suspects that drivers make more lane changes at a lower speed compared to at a higher speed, the result suggests the opposite. The main effects plot and the coefficient of the traffic speed show that the average mean of the response increases slightly when the traffic speed switches from low to high, but 3 interactions are identified in the interactions plot, one with number of lanes, one with traffic density, one with vehicle counts..

- From the interactions plot between traffic speed and number of lanes, it is clear to see that:
    a. On a 4-lane highway, a higher traffic speed has slightly more lane changes
    b. On a 2-lane highway, a lower traffic speed has slightly more lane changes
- From the interaction plot between traffic speed and traffic density, it is clear to see that:
    a. In higher density traffic, there is no difference in the number of lane changes between higher or lower traffic speed.
    b. In lower density traffic, higher traffic speed has an average about 0.3 more lane changes than lower traffic speed.
- From the interaction plot between traffic speed and vehicle counts, it is clear to see that:
    a. In higher vehicle counts, a higher traffic speed has slightly more lane changes
    b. In lower vehicle counts, a lower traffic speed has slightly more lane changes, can almost be ignored.
- Considering the coefficient estimates, the results suggest that drivers do not make more or less lane changes at a higher speed compared to at a lower speed, at least there is no practical effect.

Traffic density: this study suspects that drivers make more lane-changes in lower density traffic rather than higher density traffic, the results show the opposite. The main effects plot and the coefficient estimates show that the average mean of the response increases slightly when the traffic density switches from low to high, and 2 interactions are identified.

- From the interaction between traffic density and traffic speed, it is clear to see that
    a. No matter if the traffic speed is high or low, there are always more lane changes in higher density traffic than in a lower density traffic.
- From the interaction between traffic density and number of lanes, it is clear to see that

a. Cross interaction is identified, in a 2-lane highway, there are more lane changes in low density traffic, this provide some support to the claim of the study on traffic density but it is only by about 0.2 lane changes on average, so this can be a rare case and the support is very minor. And in a 4-lane highway, there are on average about 2 more lane changes in a higher density traffic than in a lower density traffic. This can be potentially caused by there are more vehicles in a higher density traffic than in a lower density traffic but there is no correlation found between these 2 variables, so this reason does not stand. This shows drivers do make more lane changes in a higher density traffic than in a lower density traffic.

- The results show that drivers make more lane changes in higher density traffic than in a lower density traffic in general.

Traffic flow: This study suspects that drivers make more lane changes in impeded traffic rather than in unimpeded traffic, the results align with this argument on some level. The main effects plot and the coefficient estimates of the traffic flow show that the average mean of the response decreased slightly when the traffic flow switches from impeded to unimpeded traffic, so it suggests drivers make more lane changes in impeded traffic. And there are 2 factors identified to have an interaction effect with traffic flow.

- From the interaction between traffic flow and number of lanes, it is clear to see that
    a. No matter if the traffic is impeded or not, the number of lane changes increased as the number of lanes increased.
    b. Cross interaction identified, when the number of lanes is 2, there are more lane changes when the traffic flow is impeded by 1 lane change on average. This provides support on the speculation of the study regarding traffic flow. But when the number of lanes is 4, there are more lane changes when the traffic is unimpeded, although just by around 0.3 lane changes on average.
    c. More lane changes on a 4-lane unimpeded traffic than a 4-lane impeded traffic is not expected and the reason for this unexpected outcome can be
        i. Due to observer error and not enough data points, larger data set might show different results

ii. Some factors we do not know that makes drivers want to make more lane changes in a 4-lane, unimpeded traffic than a 4-lane, impeded traffic, this finding can be further investigated in future studies

- From the interaction plot of the traffic flow and the time of day, it is clear to see that
    a. During the rush-hours, the number of lane changes decreased as the traffic goes from impeded to unimpeded by 1.5 lane changes on average. This provide supports for what is suspected in this study regarding traffic flow, more lane changes in impeded traffic than unimpeded traffic.
    b. And during the non-rush hours, the number of lane changes increased as the traffic flow goes from impeded to unimpeded. This can be caused by some particularly type of drivers drive more often during the non-rush hours and they do not drive the same way as the drivers who drive during the rush-hours, or it's simply because impeded traffic is rare during the non-rush hours, causing the overall number of lane changes in impeded traffic during non-rush hours is low.
- The results suggest that drivers make more lane changes in impeded traffic than in unimpeded traffic on highways, the difference is more notable when on a 2-lane highway or during rush hours.

Time of day: This study suspects that drivers make more unsafe lane changes during rush hours rather than during the non-rush hours, the main effects plot and the coefficient estimates highly aligns with the argument. The average mean of the response doubles when the level of the time of day switches from the non-rush hours to rush hours in the main effects plot, but 3 interactions are identified.

- From the interaction plot between the traffic speed and time of day, it is clear to see that
    a. No matter if the traffic speed is in the higher or the lower level, there are more lane changes during the rush hours than the non-rush hours. This provides support on the speculation of the study regarding time of day that drivers make more lane changes during rush hours than non-rush hours.
    b. during rush-hours, as the traffic goes faster, more lane changes occurred, and it is the opposite when during the non-rush hours, but the decrease of lane changes is very slightly, almost can be ignored.

- From the interaction plot of the number of lanes and time of day, it is clear to see that
    a. There are more lane changes during rush hours and non-rush hours no matter if it's on a 2-lane or a 4-lane highway, on a 2-lane highway there are on average 0.5 more lane changes and on a 4-lane highway there are on average 2 more lane changes. This provides support on the speculation of the study regarding time of day that drivers make more lane changes during rush hours than non-rush hours.
    b. On a 4-lane highway, the increase of the number of lane changes is steeper than on a 2-lane highway, this can be caused by the strong correlation between the number of lanes and the vehicle counts.
- From the interaction plot between traffic flow and time of day, it is clear to see that
    a. No matter if the traffic is impeded or not impeded, there are more lane changes during rush hours than during non-rush hours, this provides support on the speculation of the study regarding time of day that drivers make more lane changes during rush hours than non-rush hours.
    b. The number of lane changes is almost the same when the traffic is unimpeded for rush hours and non-rush hours, however rush hours is still slightly higher.
- The results suggest that drivers make more lane changes during rush hours than in non-rush hours on highways regardless of all the other factors.

The number of lanes: This study suspects that drivers make more lane changes on a 4-lane highway rather than on a 2-lane highway, the main effects plot and the coefficient estimates highly aligns with the argument but this factor has interactions with every other factors, and it is strongly correlated with the new variable "vehicle counts".

- From the interaction plot between the number of lanes and traffic speed, it is clear to see
    a. No matter if it's higher or lower traffic speed, there are more lane changes on a 4-lane highway than on a 2-lane highway.
- From the interaction plot between the number of lanes and traffic density, it is clear to see
    a. No matter if it's higher or lower traffic density, there are more lane changes on a 4-lane highway than on a 2-lane highway.
- From the interaction plot between the number of lanes and traffic flow, it is clear to see

a. No matter if it's the traffic is impeded or unimpeded, there are more lane changes on a 4-lane highway than on a 2-lane highway.

- From the interaction plot between the number of lanes and time of day, it is clear to see
    a. No matter if it's rush hours or non-rush hours, there are more lane changes on a 4-lane highway than on a 2-lane highway.
- From the interaction plot between the number of lanes and vehicle counts, it is clear to see
    a. When the vehicle counts is in the higher level, there are more lane changes on a 4-lane highway than on a 2-lane highway
    b. When the vehicle counts is in the lower level, there are more lane changes on a 2-lane highway than on a 4-lane highway
        i. It is reasonable for there to be more lane changes on a 4-lane highway than on a 2-lane highway when the vehicle counts is in the higher level because more vehicles passed through in that 2 mins time period can this fact directly leads to more lane changes. However the reason for there to be more lane changes on a 2-lane highway than on a 4-lane highway when the vehicle counts is in the lower level can be when the vehicle counts is low on a 4-lane highway, that means the vehicles are very spread out. As we discussed in the traffic flow section, drivers make more lane changes when the traffic is impeded than unimpeded, thus it can be rare for a driver to make a lane change on a 4-lane, low vehicle counts highway when compared to a 2-lane, low vehicle counts highway, because they just don't feel the need to change lanes.
    c. When the number of lanes is 2, the number of lane changes are the same for the lower and higher level of vehicle counts
    d. When the number of lanes is 4, there are on average 3 more lane changes when the vehicle counts is in the higher level than in the lower level
        i. This provides support on when the number of lanes is 4, the vehicle counts can cause more fluctuation on the number of lane changes than when the number of lanes is 2.

- The results show there are more lane changes on a 4-lane highway than on a 2-lane highway, however it is certain that a large part of this result is contributed by the effect of vehicle counts since it has a strong correlation with the number of lanes. Thus, no clear conclusion can be drawn from this result, but we know that now, the approach of counting the number of lane changes on highways with different number of lanes to see on which highway do drivers make lane changes more frequently does not work, because more lanes is highly correlated with more vehicles, and more vehicle naturally leads to more lane changes.

From the regression outputs, only the factor number of lanes has a P-value that is lower than 0.05 but this factors is highly correlated with vehicle counts so the coefficient estimate does not mean anything, and the traffic density being very close to 0.05 at 0.07. No interactions is showing statistical significance as well, perhaps with more replicates or a larger data set, some interactions might be statistical significant to show that the results is consistent, such as the interaction between the number of lanes and time of day.

Although the P-value of the traffic speed in the Poisson regression is not low enough to conclude that the results about the traffic speed is statistically valid, the high P-value can be suffered by the factorial design is not full because certain combinations do not exist together, and a bigger sample size likely would show statistical significance.

It seems to be reasonable to discussion traffic density and number of lanes and vehicle counts all together, because they are all related to space, after looking at the effect of these three factors with respect to the response variable in the interactions plots, coefficient estimates, main effects plot, I found drivers make more lane changes when there are more space to driver.

Before putting everything together, we suspect that there might be a slight shortcoming to categorizing traffic speed, traffic density and vehicle counts into dichotomous groups. As the study is an observational study that mimics a factorial design, we have no control over any factor or the environment. Additionally, the method of turning the continuous data of traffic speed, traffic density and vehicle counts into categorical data might obscure some pattens present in the continuous data. Thus, the Poisson regression analysis was repeated using the raw continuous data of traffic speed, traffic density, and vehicle counts in place of the dichotomous variables. It was found that, traffic density is showed no effect on the response variable, vehicle counts continued to show a significant effect on the response, and the traffic speed is showed a slightly negative effect on the response, all the other factors had the same effect as in the previous model results.

There are several possible reasons for this change, one is that, as mentioned in the previous paragraph, the categorizing step actually obscured the underlying pattern in the continuous data. A second possibility is that the dividers of these three factors are chosen based on the median of the continuous data, might not be the correct cut off point due to heterogeneity in one or both of the created factor levels.

To sum up, the results align with some of the speculations of this study regarding how the number of lane changes should fluctuate depending on the 5 environmental factors, and the findings are listed below:

- Traffic speed
    a. Drivers make slightly more lane changes at a lower speed compared to at a higher speed. (A higher traffic speed is defined as on average, the traffic is moving faster than 7 mile per hour below the posted speed limit, any traffic speed under 7 miles per hour below the posted speed limit is consider as a lower traffic speed.)
- Traffic density &Number of lanes & Vehicle counts
    a. Drivers make more lane changes when there are more space to drive
- Traffic flow
    a. Drivers make more lane changes in impeded traffic than in unimpeded traffic on highways, the difference is more notable when on a 2-lane highway or during rush hours.
- Time of day
    a. Drivers make more lane changes during rush hours than in non-rush hours on highways in general.

CHAPTER 4. STUDY 2 – LANE CHANGE MARKERS

Study two aims to investigate if there are any observable cues that drivers use to determine the relative safety of passing vehicles, we do not know what markers exist that would indicate that a vehicle will (unsafely) change lanes without a directional signal; such markers, if detectable by an autonomous vehicle, would help the vehicle prevent accidents caused by such unsafe lane changes. The research question of study two is:

- *Are there observable cues that drivers use to determine the relative safety of passing cars?*

The method part of this study consist of 2 major sections, markers gathering and markers validation, this study collects data through survey and human subject experiments, analyzes the experiment results, and reports findings at the end.

4.1 Method of the second study

As one of the most common causes of accidents on highways, unsafe lane changes are usually very difficult for human drivers to predict, especially for those who lack years of driving experience and solid driving skills. However, in real life driving, there are situations where experienced drivers are able to predict an unsafe lane change and take precautions to avoid collisions. This phenomenon usually happens when experienced drivers do not feel safe overtaking or staying close to front vehicles. This study believes that single/multiple visual cues exist in the driver's field of view that cause the driver to decide not to overtake the front vehicle or stay close to it. If experienced drivers can truly pick-up detailed visual cues to predict an unsafe lane change, there should be a method to identify these critical visual cues. The visual cues in this study are called "lane change markers", or just "markers".

One challenge to identify these markers is that there is no existing related study in this field that can provide guidelines on, to what extent, can a potential marker be categorized as a maker.

There is no reference on the threshold, or existing database to help create this threshold. One way to handle this challenge is to report what has been found, and what are the most likely markers rather than categorizing them as markers, and provide guidelines or advice on what properties the potential markers might have for future works.

The methodological approach in the first study consists of two major sections, section one is an online interview for markers collecting, section two is an online survey for validation.

4.1.1 Section one - markers gathering

The first step to investigate if there is any marker that indicates an unsafe lane-change on highways is to have a pool of all the possible markers. First, an interview sheet is created to gather information from both amateur drivers and professional drivers. The reason why interviews are used in this part of the research is because interviews, compared to surveys, can gather much more profound and much more accurate information than surveys. While this marker pool is the foundation of the study one, all the rest starts from here. It is very important that the information gathered in this section is profound and accurate.

The interview first gathers basic information such as age and gender, and then it asks the participant 10 open-ended questions about what the driver has seen on the highways. Specifically, the interview focuses on what do the participants look at before passing a front vehicle and what visual cues might cause the participants to feel unconfident to pass a front vehicle. However, not every driver has experienced an unsafe lane change, and for those who have, not every one of them can remember it and recall the details when asked straightforwardly. Thus, this interview is formatted in a way to help the participant recall the markers as detailed as possible. Only question 7 is the key question that asks about the visual cues before passing a front vehicle. All the rest of the questions are “beating around the bush” type of question, they are either for warm up purpose to assist the participant fit into the role better and quicker and get ready for question 7, or lengthening the conversation on the topic of highway lane changes and unsafe lane change, so that during the conversation, it might trigger a hidden past memory of the participants that he/she is not aware of in the beginning. There are questions such as, have you ever seen vehicles make lane changes without using their blinkers (directional signals)? Have you ever predicted any of these unsafe lane changes?

The interview is carried out through phone calls, the original plan is to interview 10 amateur drivers, 10 professional drivers, and 10 autonomous truck drivers. The amateur drivers in this study are defined as those who drive their vehicles routinely, such as driving from home to work on the weekdays, driving to supermarkets for groceries, or an occasional cross country road trip, they only drive for commute purposes. The requirements for potential amateur drivers are they must be car owners, and they must have at least 2 years of driving experience for commute purposes. The professional drivers in this study are defined as those who are hired or paid to drive, such as taxi drivers, shuttle drivers, and semi-truck drivers. Due to the fact that highways are the road environment that this study focuses on, semi-truck drivers are the main targeted participants for the interview, but 2 shuttle drivers and 2 taxi drivers are also included in the interview. The requirements for potential truck drivers to participate in this study are they must be currently hired or paid to drive as an active professional driver with a minimum of 1000 miles of travel distance weekly, and at least 10 years of continuous professional driving career up to now. According to the Department of Transportation of the US, truck drivers may not drive if they have worked 70 hours or more in the previous 8 days. Which means if a truck driver uses all 70 weekly legal working hours on highways at a typical speed limit of 65 mile per hour all the time, the truck driver can travel at a maximum distance of 4550 miles per week. However, that travel distance is not realistic since truck drivers log their driving hours not only when they are at the highest speed, but also during pre-trip and post-trip inspections, driving in cities, getting off the highway to get fuel, etc., according to multiple online professional truck driving community websites. Nearly 10% of the total work time is spent below 65 mile per hour, and a travel distance of 3000 miles is a good benchmark for a productive week. The average distance for a truck driver to travel weekly in the US is between 2000-3000 miles. The Bureau of Labor Statistics estimates that the average age of a commercial truck driver in the U.S. is 55 years old.

Due to the cost and time, this study can only reach out to a certain number of potential truck drivers for selection. Thus, to ensure the quality of the responses and also taking the difficulties of finding the ideal participants into consideration. An active truck driver that has a minimum of travel distance at 1000 miles weekly and a minimum of 10 years of continuous professional driving seems to be a reasonable middle ground. This standard is not enforced on shuttle drivers and taxi drivers as they do not drive on highways as frequently as truck drivers, but they can sometimes provide fresh perspectives and new elements which helps the comprehensiveness of the markers. However, their approximate traveled mileage each year and driving age are recorded.

The Autonomous truck drivers in this study are defined as those who previously worked as traditional truck drivers for a living, and then were hired by an autonomous truck company, went through a comprehensive safety and monitoring training provided by the company, and received the qualification to operate autonomous trucks for cargo transportations. The reason why autonomous truck drivers are included in this interview is because, although there already is a large enough population of traditional truck drivers to interview, the autonomous truck drivers have underwent unique safety and monitoring training compared to traditional truck drivers, they are professionally trained to identify highway safety hazards during the monitoring operation, so their opinion could be valuable for collecting markers in this study.

One challenge during the recruiting process is that the population of autonomous truck drivers is limited in the US currently. Although this study was able to reach out to 3 autonomous truck companies that contain around 25-30 autonomous truck drivers, only one autonomous truck driver is willing to participate in the study.

Due to the current policy of social distancing, the interviews are all phone based. Table 1 contains the standard questions that are used during this interview, and the conversations are recorded as audio data for later analysis.

Table 3. Driver interview question sheet

Driver interview question sheet	
Participants ID:	AV /Traditional/ Amateur
Age: 18-69/above 70	Gender: M/F
Question	Answer
1. If a professional driver: How many years have you been driving as a paid professional driver? Do you drive every day, every other day, or weekly? If an amateur driver: How many years have you been driving? How many miles have you driven in your life, just give an estimation. Do you drive every day, every other day, or weekly?	
2. Have you ever seen vehicles make lane-changes without using their blinkers?	Yes / No
3. If yes, have you ever predicted any of these unsafe lane-changes (make sure to clarify the "unsafe" is without using the blinkers)?	Yes / No
4. If yes, how did you predict it? (What are the actions or movements that the front vehicle does that makes you feel it's going to make an unsafe lane change?)	
5. What do you think are the most common reasons that some drivers make unsafe lane-changes (without blinkers)？	
6. What do you think are the possible reasons that some drivers make unexpected lane-changes?	
7. Imagine yourself driving on a highway like you normally do, there is a vehicle driving in front of you. At a certain point you want to pass the front vehicle due to some reason, but you do not feel safe passing it, you would rather stay back for a little longer and keep a safe distance. What do you think the reason might be or what did you see that might stop you from passing the vehicle right away? Try to describe the reason as detailed as possible.	

Table 3 continued

8. Have you seen any dangerous maneuvers on the highway except for what you already told me? What is your interpretation of dangerous here? And what are some examples that you have personally experienced, try to describe the surrounding environment and the situation as detailed as possible.	
9. (For AV truck drivers) When you are monitoring the truck driving itself, what are the aspects that you look at specifically? This includes not only your truck, but also other possible aspects, for example the environment condition and the surroundings.	
10. Outside of the question that I have already asked, is there anything that you would like to share that you think might be helpful to this research?	

The workflow of conducting this interview starts with greeting the participant, expressing gratitude for participating in this study. Then, inform the participant that they will be asked ten questions in this interview, all of them will be related to their highway driving experience. After the participants have understood what this study is majorly about, their approval on audio recording is needed. Then, collect basic information such as age, gender, and use the standard question sheet as the guide in Table 1 to collect responses, record their answers on the right column. One challenge during the interview is that each participant has different levels of driving skills and different levels of spatial imagination ability. It is crucial for the interviewer to maintain a consistent understanding of the questions among all the participants, it requires vivid description of the scenario, and also multiple attempts of back-and-forth confirmation between the interviewer and the participant. However, after 4-5 interviews, the interviewer should be very familiar with the question sheet, and will know how to navigate the interview so that the participant can understand the questions more accurately and efficiently.

After all the interviews have been conducted, the next step is to review the completed interview sheet, and summarize the keywords to create a table of markers. It is important to review every detailed information in the completed interview sheet, as well as the information in the audio data. The workflow to review and summarize the markers is by each participant. For example, when reviewing participant 1, have the completed interview sheet opened, and play the audio file in the background. Review along the interview sheet as the audio plays, just like during the interview and check if there are any details missed on the sheet. Make sure everything worth mentioning is listed in the sheet, it can be in a sentence, a paragraph of description or a combination of keywords.

The challenge of this step is that different participants can mention the same thing with different words, such as “an unexpected lane change”, can be “that car came out of nowhere, I had no idea that he would cut me off like that”. Twenty participants might have 20 different ways to describe this. Thus, using automated audio-to-text software that contains keywords searching and categorizing capability is not accurate and efficient in this situation, at least no existing audio-to-text software is capable of handling data this random and meets the requirements of this study. Therefore, listening to the audio files and reviewing the completed interview sheet to pick up potential markers can only be performed by human researchers. However, relying on human researchers to process this data has other risks. Different researchers can have different interpretations of what the participant is trying to describe due to bias. To lower this risk, 2 researchers are involved in this process, and work independently in their workflow.

Here is the audio-to-text workflow that summarizes the keywords and markers. The researchers work independently, listen to the audio file and review the interview sheet at the same time. Whenever a potential marker is identified, the researcher puts it in a new document that is named keywords-to-markers table, attached in the appendix. Then, after the researchers have reviewed all the audio files and interview sheets, each of them should have a completed table of keywords-to-markers that contains the information of:

1. All the potential markers
2. All the keywords of each marker

The next step is standardizing the table, 2 researchers would sit down and look at each other’s keywords-to-markers table, compare, discuss, and summarize the two tables into one standardized table. This is a time consuming approach but with multiple benefits:

1. The researcher would have deeper understanding of the data compared to using automated software for summarizing and categorizing, therefore, less details are missed
2. Two researchers working independently and compare afterwards can lower the error of subjective interpretation which might introduce bias
3. The keywords-to-markers table can be one of the contributions of this research, it is a standardized tool to be utilized whenever markers need to be extracted from text or audio based data

After the table of keywords-to-markers has been summarized by different researchers, compared and standardized into one table. Here is the list of all the potential markers that will be transformed into script so that a clear and detailed instruction can be followed when building the animation in Unity 3D.

The list of potential markers are listed below, this list is summarized from the interview results:

1. Swerving and leaning towards one side of the lane
2. Making consecutive lane changes
3. Brake lights turn on once
4. Brake lights turned on continuously
5. Wobble motion
6. Near an exit
7. Near an Exit + brake light
8. Overtaking a semi-truck
9. Police car on the road shoulder
10. Construction site on the road shoulder

4.1.2 Section two - markers validation

With this list of potential markers, the process moves on to the next major section, markers validation. First, we created the script for replicating each marker; the full version of the marker-to-script table is attached in the appendices. To explain the marker-to-script table, A is the prop sedan, B is the participant's sedan, the animation's viewing angle is always "what the participants should see if they are sitting in vehicle B (driver seat）". From the left, the first column is the number of markers. The second column is the information of the marker, showing which marker it is. The third column is the further information of the marker, usually, this column is used for putting some notes either for explaining a question or keeping track of a piece of thought. The last column is the key information of the table which is the script for the animation. It is very similar to a script that actors use to practice performance, it describes the environments, relative position of the objects, and actions step by step in a timely fashion. The script has to be very detailed and very specific so that the animation building process has standardized instruction to follow, it's the necessary measure to ensure an accurate and even performance across all animations, the marker to script table is attached in the appendix.

With the specific and standardized script table, the next step is to build the animations based on the information written in the script. The tool used in this study to build the animation is called Unity 3D, it is a widely used video game development tool, because of its cross-platform feature, both hobby game developers and professional game studios use this software to develop games. It is relatively easy to learn and use, there are a lot of learning materials of this tool online for self-learning, and it offers a much higher degree of flexibility than a driving simulator testing environment software.

Below is a screenshot of the animation at the beginning of the animation build, it only contains a one-direction 2-lane highway, a Mustang-like sedan that is free to use, in a very basic landscape environment with no clouds. The next figure below is a screenshot of one of the fully built markers, just to demonstrate what the actual image looks like.

Figure 18. Marker animation screenshot 1

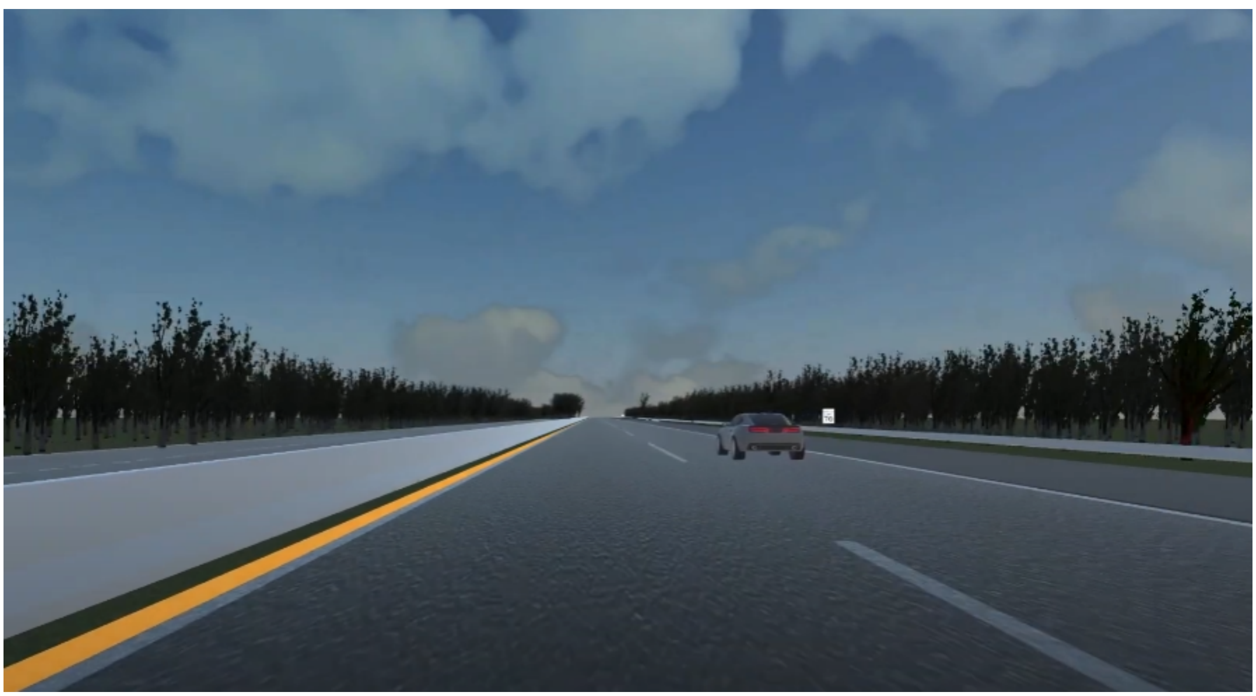


Figure 19. Marker animation screenshot 2

The actual marker animation contains traffic signs on the side of the road, it can be a speed limit sign, or other signs depending on the setting of that marker. It contains a 2 lane highway and the opposite-direction highway on the left with no traffic in that direction. A solid yellow line on the edge of the highway, trees, and clouds. The vehicle model remains the same as before.

The animations built in this study are for the purpose of validation. Typically, traffic scenarios are replicated through driving simulator software, and participants can come to the lab and drive the simulator. The reason why 3D animation is chosen in this study but not other methods, such as driving simulators and real-life traffic recording, is because it requires the least amount of extra work, such as the legal issues, but provides a great amount of flexibility and possibility for experimental environment design. To illustrate, Most driving simulator software is specifically created for the specific hardware of the manufacturer, which means without the specific software, the hardware cannot work on itself. While most of the simulator software does not offer a great amount of flexibility in designing, it does not help with the study. Regarding the real-life traffic recording method, although it offers more flexibility and control of the testing environment, it requires legal approval from the local law enforcement and legal approval from the University to perform controlled experiments and filming in real-life traffic. Thus, it is also not an ideal approach for the study.

After the markers have been summarized with minimum bias, scripts for replicating the animations have been created, and the animations have been built through animation software, the next step is to design the survey for validation.

The purpose of validation in this study is to see to what extent, does the markers that were summarized from the interview, are considered to be one of the observable cues that help the daily amateur drivers to determine the relative safety of passing cars. To achieve this purpose, a survey is designed to show the animation clips to the participants and collect their responses.

In section one, the survey starts with a filtering question at the beginning that asks if the participant drive routinely or not. If the participant chooses no, the participant is not qualified to take the survey and be directed to quit.

1. Do you drive routinely? (such as drive to work, Monday to Friday, or drive to buy groceries from time to time) (Y/N)

In section two, the survey collects the participants' basic info:

2. How many years have you been driving routinely?
    a. 0-5 years
    b. 5-10 years
    c. 10-15 years
    d. Over 15 years
3. Approximately, how many miles do you drive each year?
    a. 0-5k miles
    b. 5k-10k miles
    c. 10k-15k miles
    d. Over 15k miles
4. What is your gender?
    a. Male
    b. Female
    c. Prefer not to say

Question 2-4 collect participants' demographics which allows the researcher to know approximately how much driving experience does the population have and whether the population is skewed towards one gender or not.

In section 3, the survey lays out the definition of "a sudden lane change" that is mentioned in the survey:

5. Here is the definition of a "sudden lane change" in this study. (Term definition)
    a. A lane change that occurred with no signal beforehand (such as a directional signal, "blinker")

In section 4, the data collection starts with the first marker question:

1/10. Watch the video carefully and answer the question. You can watch the video as many times as you want. (Click on the YouTube video link below to access the video, please watch in full screen) https://youtu.be/1wfa0IBtjjg

1/10. What would you do at this moment?

1. Overtake without any hesitation
2. Overtake with caution
3. Not sure what to do (stay put)
4. Decide not to overtake (stay put)
5. Decide not to overtake, slow down, and keep a safe distance
6. Other:_______

There are 10 questions in total, each question contains the same question body but with a different video link for the participants to access the 10 different animations. The question body is designed following a Likert scale fashion that allows the participants to respond with more flexibility, and if the participants find the first five selections not usable, the participants can manually type their answer in selection number 6. Inside the animations of each marker, a sentence that sets the scene of the animation is presented to the participants at the beginning to let them know what the basic situation is. For example, in marker 1, this sentence is presented at the beginning "You are driving normally on a 2-lane highway, the car in front of you swerves and lean towards the left side of the lane". Then the participant sees an animated highway driving scene as if he/she is driving on the left lane (shown in figure 21).

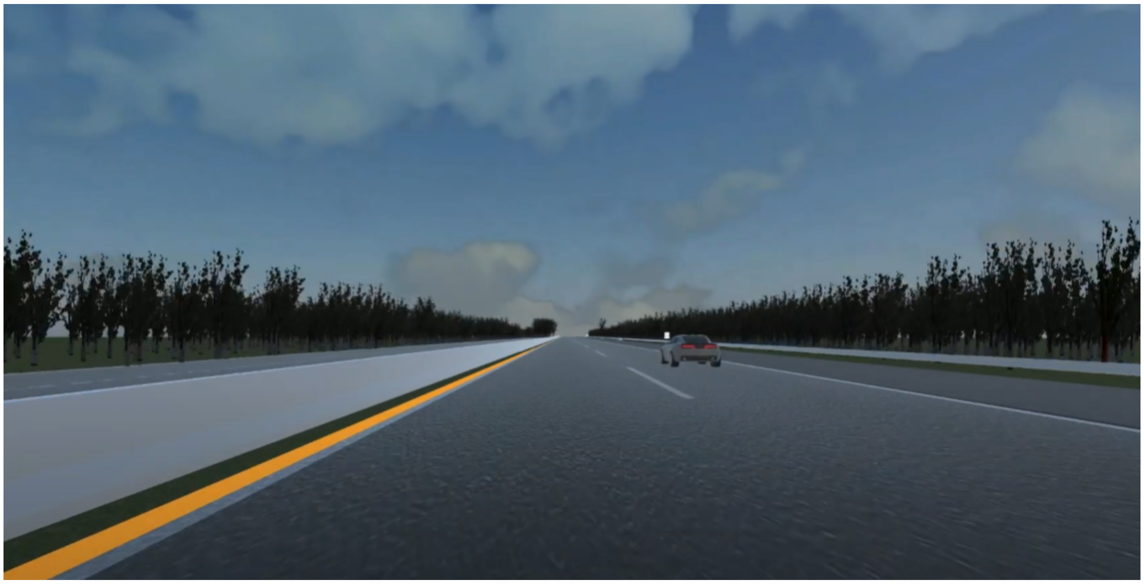


Figure 20. Marker animation screenshot 3

After driving for a few seconds, the front vehicle starts to swerve and lean towards the left side of its lane, circled in blue and also paused for one second in the animation(shown in figure 22).

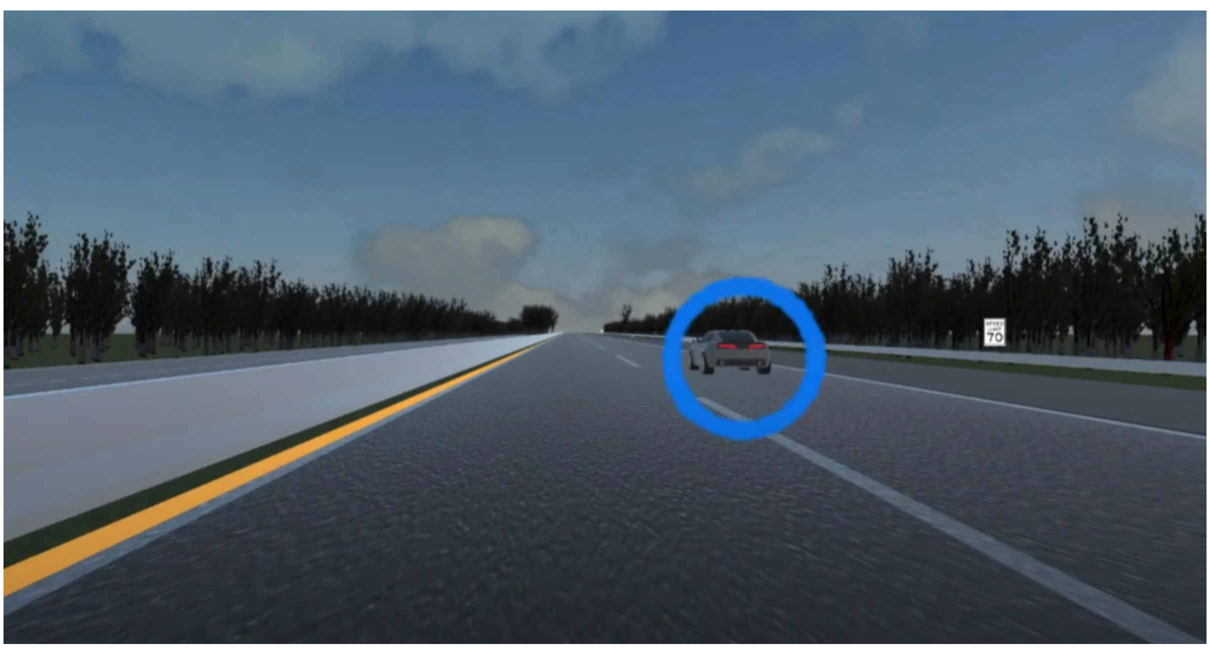


Figure 21. Marker animation screenshot 4

One important note that was later found out in the survey testing before the launch is that, if the key information about the marker is not emphasized by bright color circles and paused for the participants to look at, most of the participants cannot fully understand what the animation video is trying to convey because they simply do not know where they should look at, therefore leads to a useless survey result. Thus, key information about the marker should be either highlighted with bright color indications or paused to allow the participants a moment to digest or in this study, both measures are been used to ensure the information is conveyed as accurately as possible.

After the video is paused for 1 second and the key information has been highlighted in the animation, the front vehicle starts to swerve and lean towards the left side of its lane until the left wheel touch the dashed lines in the middle of the highway, as shown in figure 23. This frame is paused again for the participants to read the statement, depending on the length of the statement, this pause can vary from a few seconds to dozens of seconds. An arrow points at the front vehicle to ensure that the participants are looking at the correct object.

Figure 22. Marker animation screenshot 5

Then, the survey question appears in the animation, as shown in figure 24. The reason why the survey question appears in the animation first rather than only in the survey is that prompting the survey question in the animation allows the participants to generate their decision while they are looking at the traffic at that moment. Compared to turning off the video and coming back to the survey, then reading the survey question without any related traffic animation on the screen, making their decision right at the end of the animation with the traffic animation in the background of the question is simply a more intuitive and natural way. This difference between prompting the survey question right after the animation and just keep the survey question in the survey is verified through a survey pilot study before the survey launch.

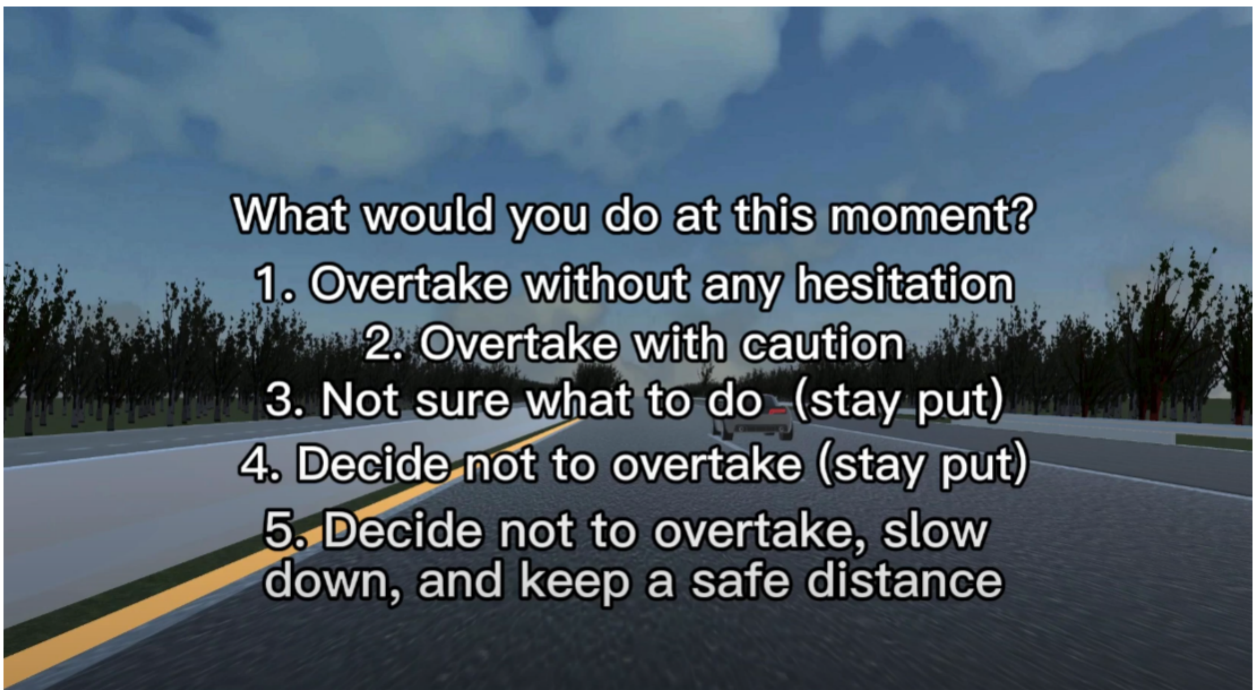


Figure 23. Marker animation screenshot 6

When the animation stops, the participants go back to the survey and respond to the same question that was prompted in the video. Then, the participants move on to the next page for the next marker, watch the video and respond to the same question until all the markers have been watched and all the questions have been answered.

This survey is entirely conducted online through Amazon Mechanical Turk which is a widely used online survey tool that can collect a large number of responses at an acceptable cost. The average length of finishing the survey is about 10-12 minutes.

4.2 Results of the second study

The interview conducted in the markers gathering section collected 21 participants in total, in which there are 10 amateur drivers, 10 professional drivers (6 traditional truck drivers, 2 shuttle drivers, 2 taxi drivers), and 1 autonomous truck driver. Among the 21 participants in this interview:

- 24% are female, 76% are male.
- 58% drive over 15k miles every year, and 4% drive under 5k miles every year
- 52% have been driving for over 15 years, 10% have been driving for less than 5 years (drive routinely, such as drive to work, and drive for basic commute)

The experiment conducted in the markers validation section gathered 100 participants in total online, among the 100 participants in this experiment:

- 58% are male, 41% are female
- 18% drive over 15k miles every year, 39% drive between 5k and 10k miles every year
- 63% have been driving for over 15 years, 2% have been driving for less than 5 years (drive routinely, such as drive to work, and drive for basic commute)

Below are the experiment responses of the participants after watching each marker animation video, shown in a visualized format.

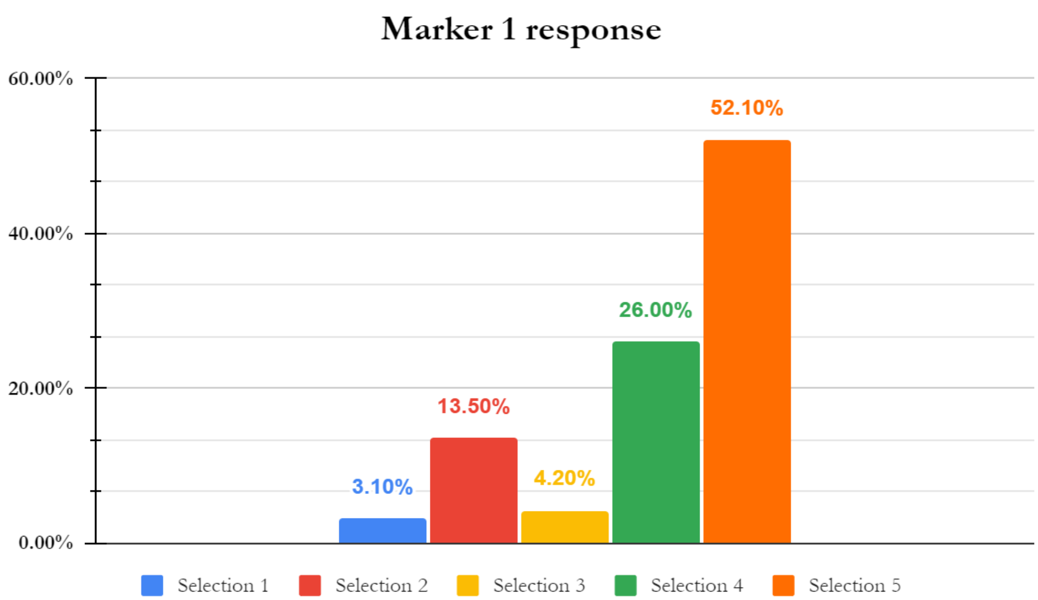


Figure 24: Marker 1 responses bar chart

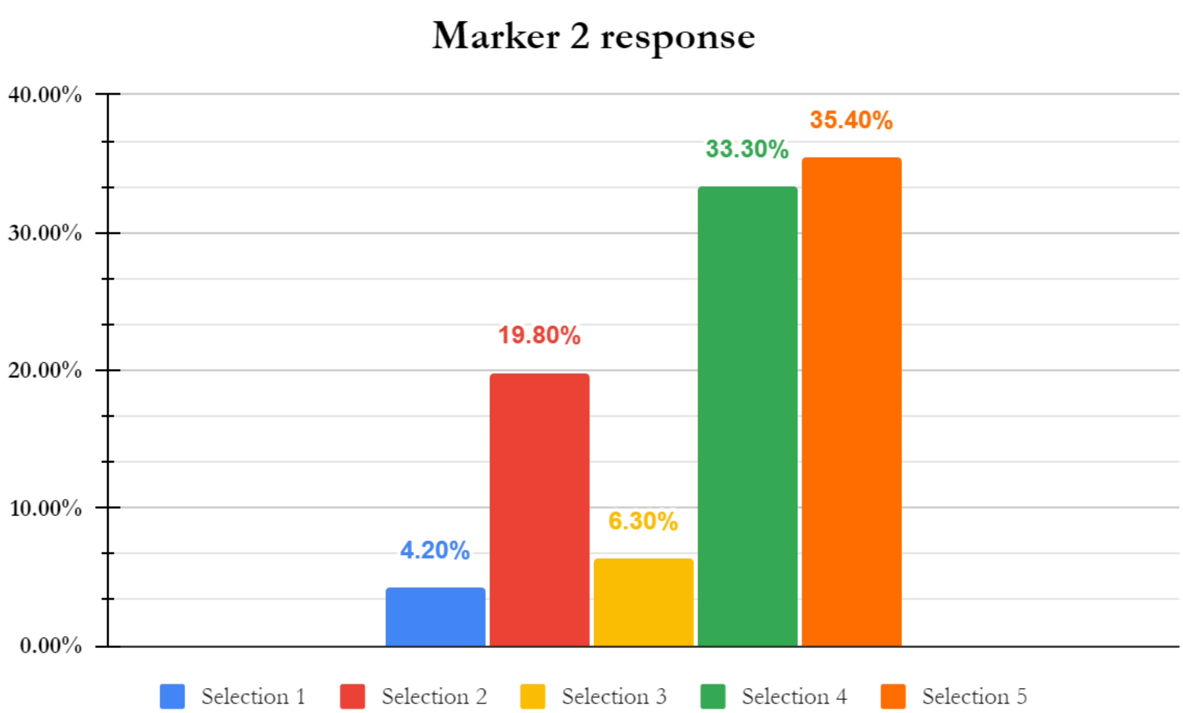


Figure 25: Marker 2 responses bar chart

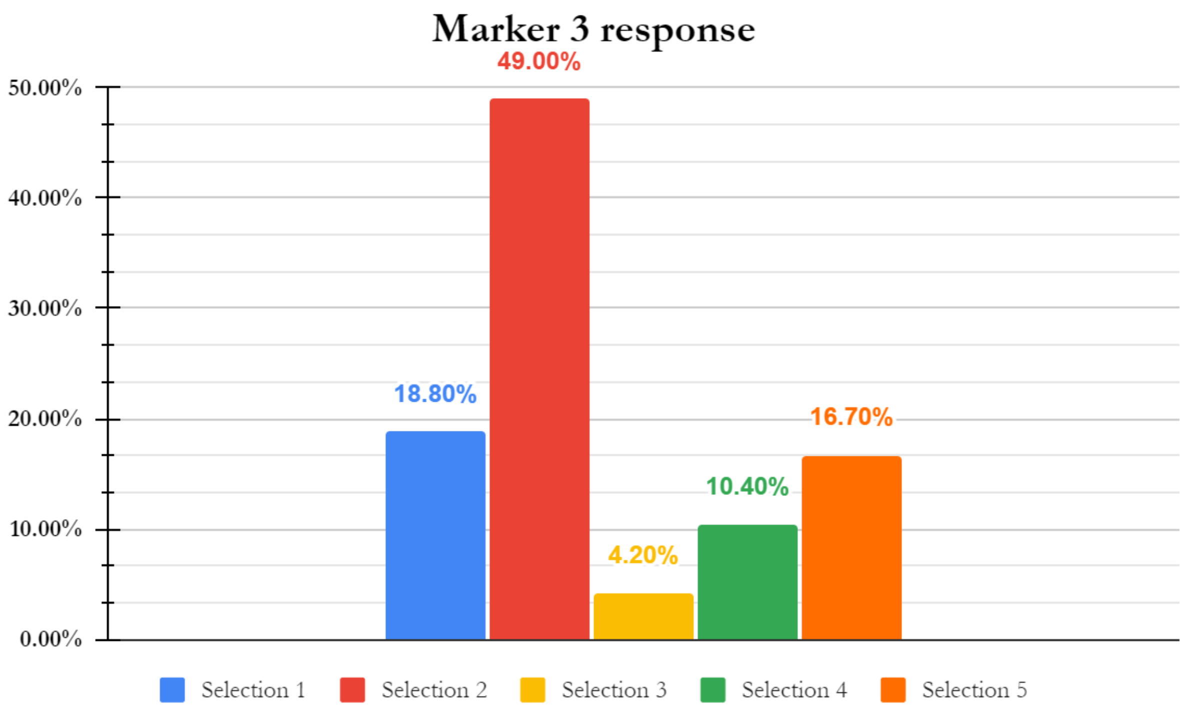


Figure 26: Marker 3 responses bar chart

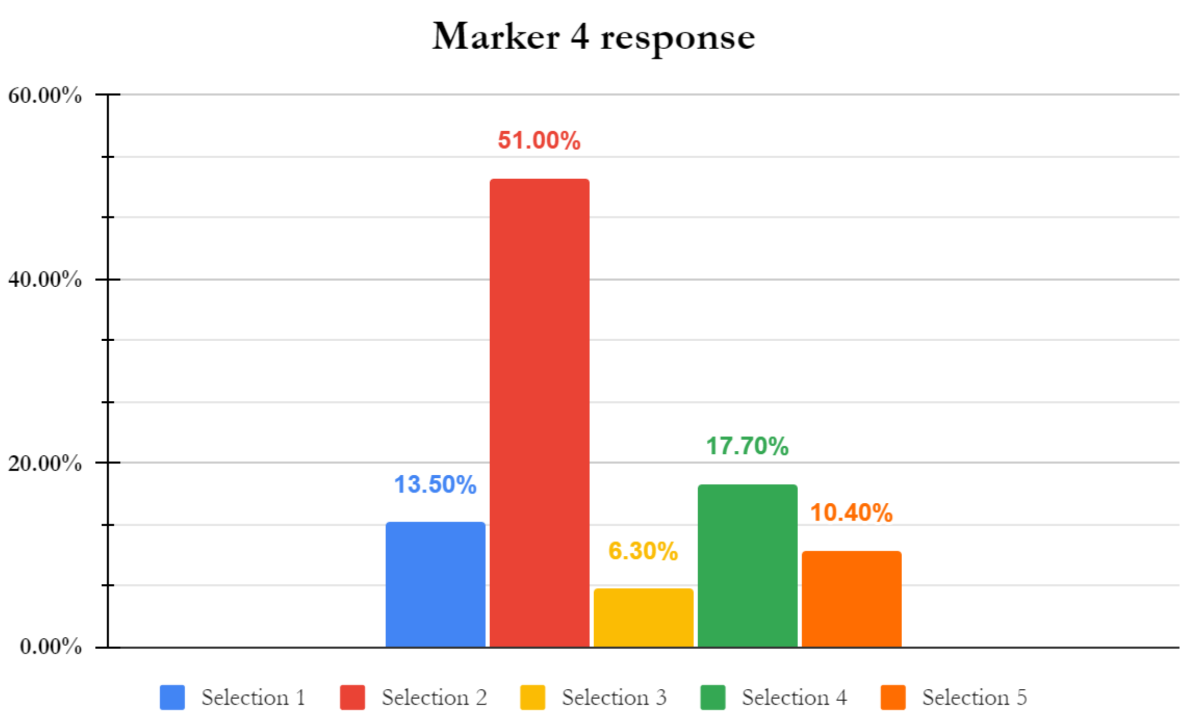


Figure 27: Marker 4 responses bar chart

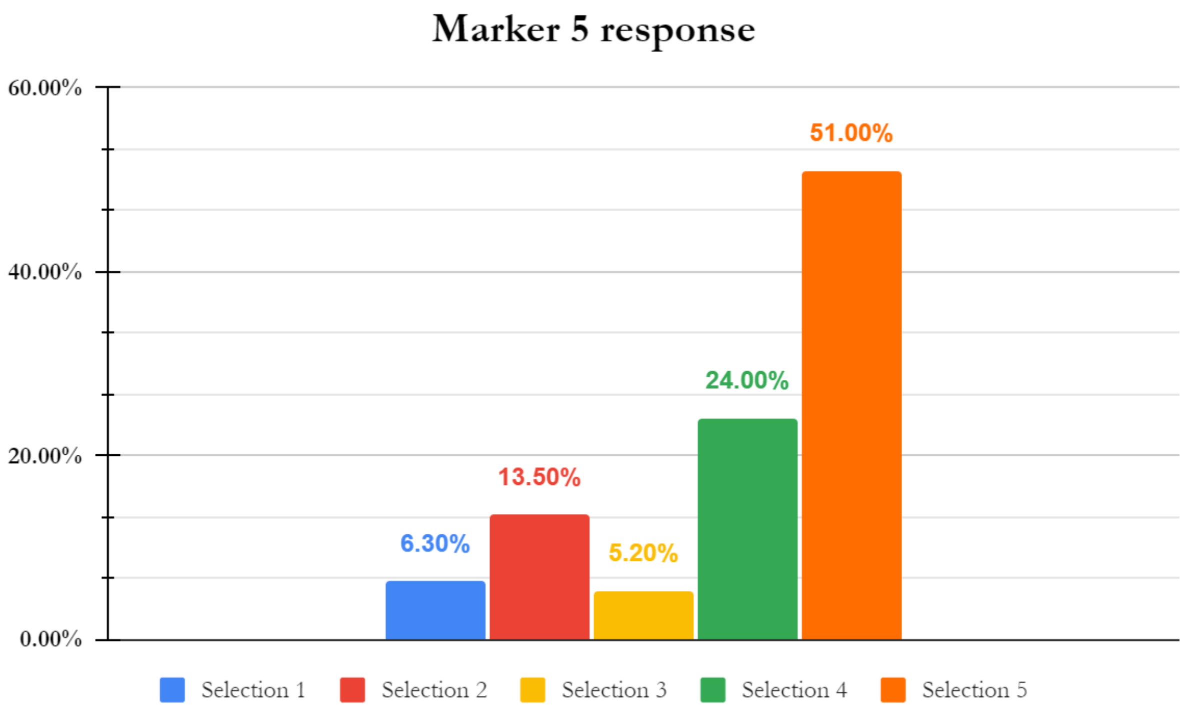


Figure 28: Marker 5 responses bar chart

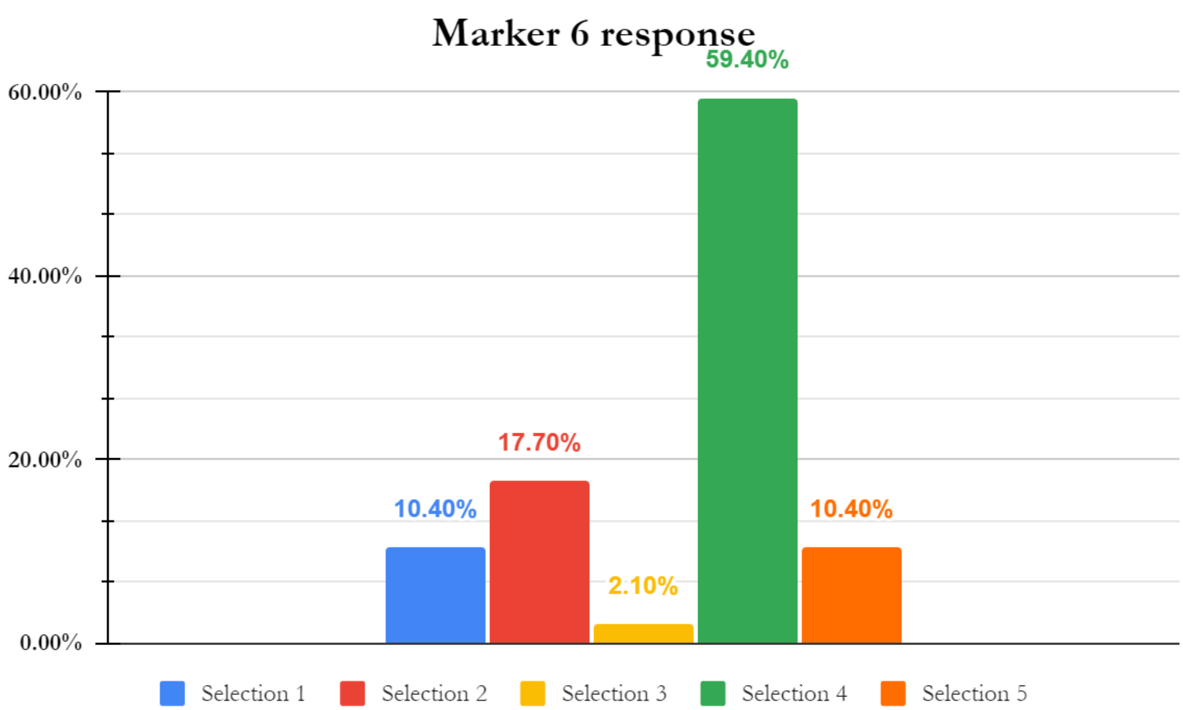


Figure 29: Marker 6 responses bar chart

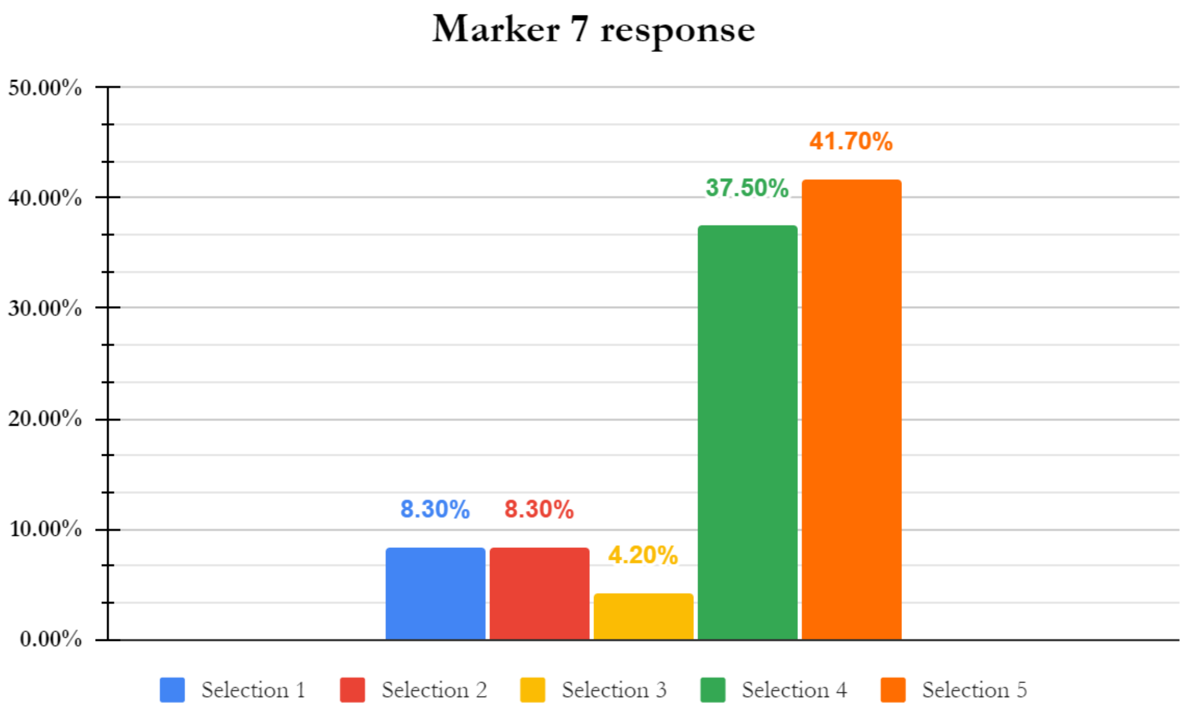


Figure 30: Marker 7 responses bar chart

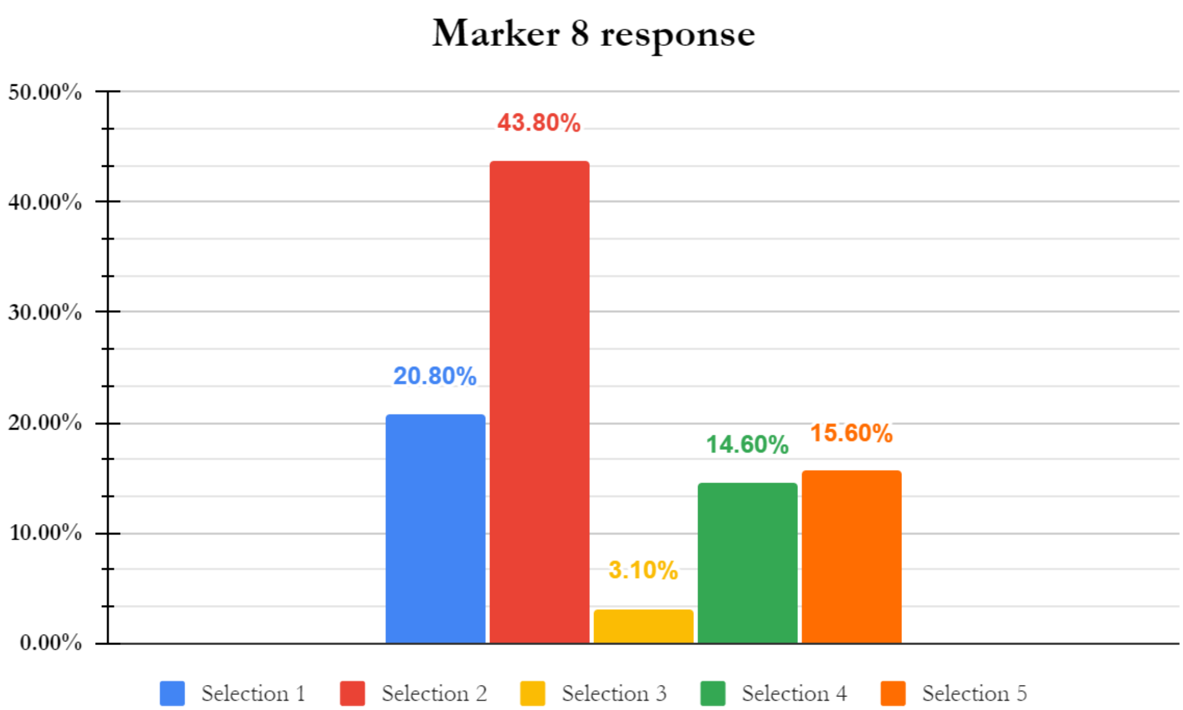


Figure 31: Marker 8 responses bar chart

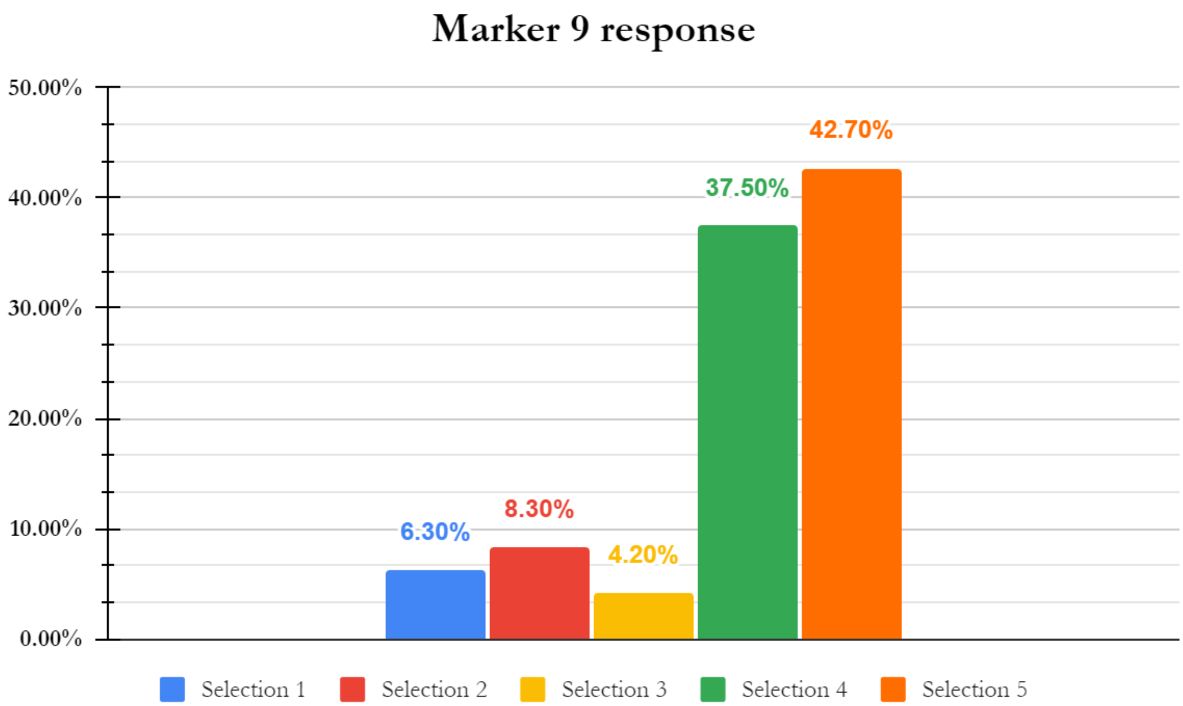


Figure 32: Marker 9 responses bar chart

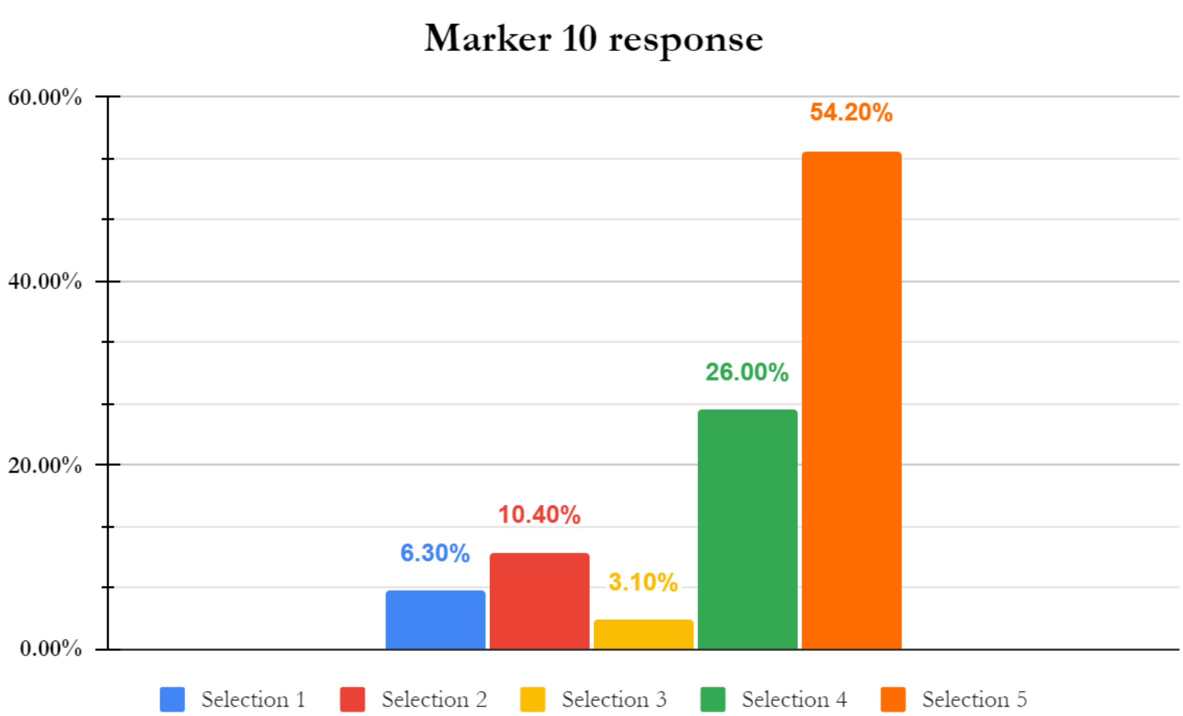


Figure 33: Marker 10 responses bar chart

The averaged selection among 100 responses of each marker is shown below in Table 5 and Figure 35, this statistics shows that:

- Marker 3, marker 4, and marker 8 have the averaged selection below 3
- The majority of the markers contain a response over 3
- Marker 1 and marker 10 have largest response at 4.12

Table 4: Averaged selection among 100 participants of each marker

Marker 1	Marker 2	Marker 3	Marker 4	Marker 5	Marker 6	Marker 7	Marker 8	Marker 9	Marker 10
4.12	3.79	2.64	2.65	3.98	3.44	3.95	2.62	4.01	4.12

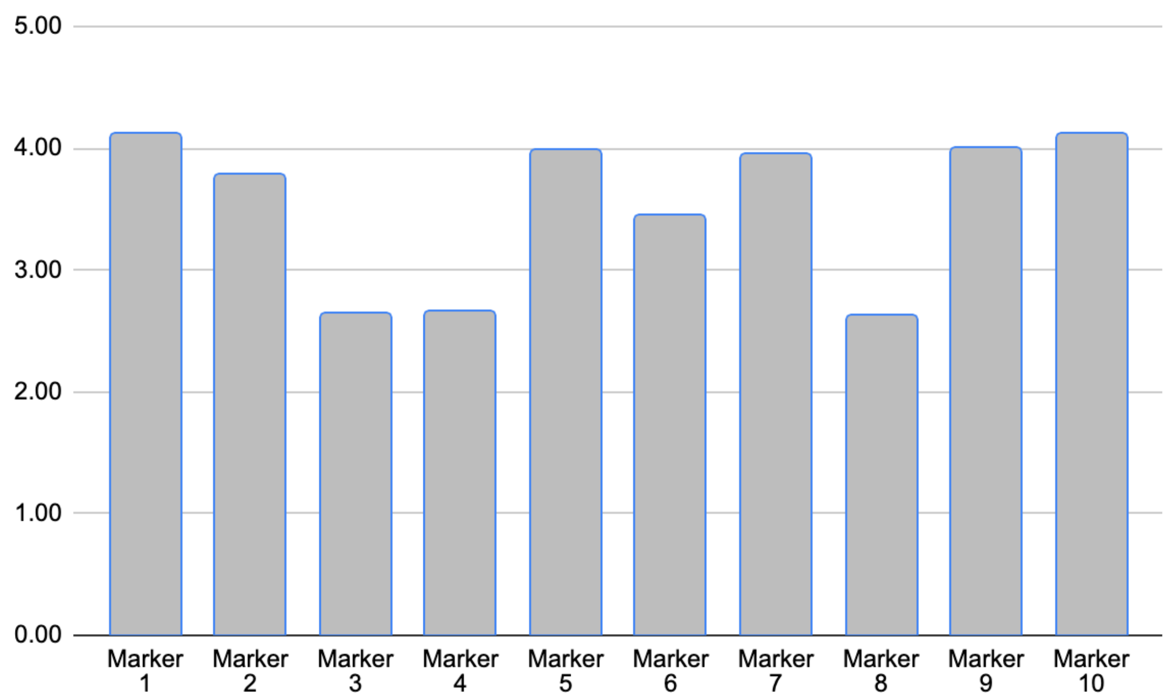


Figure 34: Averaged selection bar chart

Table 5. Summarized responses

Marker	Selections						
Marker	1. Overtake without any hesitation	2. Overtake with caution	3. Not sure what to do (stay put)	4. Decide not to overtake (stay put)	5. Decide not to overtake, slow down, and keep a safe distance	Averaged selection	Majority selection
1	3.10%	13.50%	4.20%	26.00%	52.10%	4.12	5
2	4.20%	19.80%	6.30%	33.30%	35.40%	3.79	5
3	18.80%	49.00%	4.20%	10.40%	16.70%	2.64	2
4	13.50%	51.00%	6.30%	17.70%	10.40%	2.65	2
5	6.30%	13.50%	5.20%	24.00%	51.00%	3.98	5
6	10.40%	17.70%	2.10%	59.40%	10.40%	3.44	4
7	8.30%	8.30%	4.20%	37.50%	41.70%	3.95	5
8	20.80%	43.80%	3.10%	14.60%	15.60%	2.62	2
9	6.30%	8.30%	4.20%	37.50%	42.70%	4.01	5
10	6.30%	10.40%	3.10%	26.00%	54.20%	4.12	5
Mean	9.80%	23.53%	4.29%	28.64%	33.02%		

From the bar charts of each marker in Figure 25 to Figure 34, and the summarized responses in Table 5, the results can be listed as follow:

1. Selection 3 has the lowest percentage among the other selections for every marker (based on the 4.29% at the bottom row of the table)
2. Marker 3,4,8 have the majority of the participants choose to overtake with caution (based on the selections under the Majority selection column)
3. Marker 1,2,5,7,9,10 have the majority of the participants choose not to overtake, slow down and keep a safe distance (based on the selections under the Majority selection column)
4. Marker 6 have the majority of the participants choose not to overtake and stay put (based on the selections under the Majority selection column)

The results can then be further summarized by grouping the markers with similar properties together and rank them based on the percentage of the majority selection of each marker from high to low for a cleaner observation, shown in Table 6 below:

Table 6. Markers ranking

#	Marker	Markers that present a potential safety concern (high to low)
1	10	Construction site on the road shoulder
2	1	Swerving and leaning towards one side of the lane
3	5	Wobble motion
4	9	Police car on the road shoulder
5	7	Near an Exit + brake lights
6	2	Making consecutive lane changes
7	6	Near an exit
8	8	Overtaking a semi-truck
9	3	Brake lights turn on once
10	4	Brake lights turned on continuously

The markers listed in Table 6 above are categorized into 3 groups, red, yellow, and green. The markers in red exhibit a higher level of safety concern before overtaking from the driver's perspective, the markers in green exhibit a lower level of safety concern before overtaking from the driver's perspective, and the marker in yellow is somewhere in between, it does not exhibit a higher nor a lower level of safety concern from the driver's perspective, it generally implies driver's would not do anything rather than stay put and assess.

4.3 Discussion of the second study

This study aimed to investigate if there are any observable cues of the front vehicle that can indicate an unsafe lane change with a directional signal, such observable cues or markers, if detectable by an autonomous vehicle, would help the vehicle prevent accidents caused by such unsafe lane changes. In recent years, lane-changing detection and lane-changing behavior have been receiving attention in traffic flow modeling. Numerous studies have been conducted on designing and developing complex algorithms to simulate and analyze urban streets and highways traffic patterns from a systematic standpoint. However, no studies have been found in the area of identifying observable cues of a front vehicle that indicates a possible unsafe lane-change.

This study first designed and conducted an interview in order to gather the potential markers that might indicate an unsafe lane change on highways, then 10 lane-change markers were extracted from the interview results, and then turned into scripts for the animation build to follow. Secondly, 10 animation video clips were built based on the 10 markers, each animation contains the scenario and the information of a marker. Then, a standardized question designed in a Likert scale fashion that can gather participants responses after watching each animation video clip was embedded into an online survey for validation. The online survey consists of 10 questions, each question contains 1 marker animation and 1 standardized question. Lastly, 100 participants joined the validation experiment online with different but sufficient driving experience, and the participants are not heavily skewed towards one gender.

Based on the results summarized in Table 5 that shows the summarized responses, selection 3 which is “not sure what to do (stay put)” has the lowest percentage among all the other selections for every marker. It is not very useful to compare the averaged percentage of each selection as shown in the last row of Table 5, because every marker contains different scenarios and background information, averaging this percentage does not help revealing anything relates to the characteristic of the markers. However, one useful information can be derived by comparing this averaged percentage of each selection. If on average, only 4.29 out of the 100 participants selected choice 3 across all the markers, this means that the majority of the participants are able to make up their minds in either direction of choices after watching the markers animation video, which can reflect that the question design and the animation build for the online validation survey is not difficult for the participants to understand, otherwise this number should be much higher than 4.29, this provides some support to show the validity and reliability of the results.

Regarding the results in Table 5 and Table 6 on the grouping and ranking of the markers, it is important to stated that the ranking of the markers within one group is only based on which one has a higher percentage in their majority selection, it does not exclusively mean that a marker with a higher ranking exhibit more safety concern than a marker with a lower ranking within the same group. To further explain the reason behind this statement, marker 10 and marker 1 need to be used as an example. From Table 5, the majority selection of marker 10 and the majority selection of marker 1 are both 5, which means the percentage of the participants that choose “not to overtake, slow down, and keep a safe distance” is the highest among all the other selections after watching the markers animation video. Marker 10 is ranked higher than marker 1 in Table 6 is only because of the percentage of the majority selection of marker 10 is a few percentage higher than what marker 1 has, which is 54.20% against 52.10%. If we assume that, from the driver’s perspective, marker 10 exhibits “more safety concern before overtaking” than marker 1 exclusively just due to the fact that the participants choose “not to overtake, slow down, and keep a safe distance” for marker 10 is 2% higher than marker 1. The assumption would collapse the moment when the percentage of other selections are included. For example, the percentage of selection 1 of marker 10 is 6.3%, and 3.1% for marker 1. This means 6.3% of the participants choose to “overtake without hesitation” after watching marker 10, and 3.1% choose this selection after watching marker 1. A higher percentage of participants choose to “overtake without hesitation” for marker 10 compared to marker 1 reflects that marker 10 exhibits less safety concern before overtaking from the driver’s perspective than marker 1. This tells exactly the opposite of what the percentage of the majority selection of markers is showing. Considering the fact that the percentage of the majority selection of markers should contains more weights than the others, the grouping principle is established on the percentage of the majority selection of markers rather than others, but a higher ranking within the same group does not necessarily mean an exclusively higher level of safety concern. However, red group does exhibit more safety concern than the yellow group in general, and the yellow group does exhibit more safety concern than the green group in general, from the driver’s perspective.

The most important question to ask is, why would the participants feel a higher level of safety concern before overtaking in the red groups and feel a lower level in the green groups? Or more broadly, if we assume the results from this study is believable and representable of the public, why would people feel safer to pass in certain markers? What are the common characteristics among these markers that can be extracted from the results? This study aims to investigate if there are any observable cues that drivers use to determine the relative safety when overtaking front vehicles, although the detailed and fractional information about why drivers don't feel safe to overtake the front vehicle in a certain marker is useful to the autonomous vehicle industry, the common characteristics within the same group of markers is the main reason that caused the driver to make the decision, and that is the most valuable information for the autonomous vehicle industry, and also what this study cares the most.

To find out the common characteristic among the same group of markers that can cause the drivers to make their decisions, more detailed information such as the surrounding environment of the vehicle of each marker need to be revisited.

In the red group, there are 6 markers in total. Construction site on the road shoulder, swerving and leaning towards on side of the lane, wobble motion, police car on the road shoulder, near an exit and brake lights on, making consecutive lane changes. One common property among these markers is that they either exhibit some kind of side-to-side motion such as "swerving and leaning towards one side of the lane", "wobble motion", and "making consecutive lane changes", or they contain some factors that can potentially trigger a side-to-side motion such as "construction site on the road should", "police car on the road shoulder", and "near and exit and brake lights on". The approaching construction site and the police car on the road shoulder show the driver that there is a good chance that the front vehicle might switch to the left lane to create a safe space away from the objects on the road shoulder. One comment from the online survey wrote "slow down for the police car and expect the other car to move over is the law here", this comment reflects that some drivers have the mindset to keep away from the objects on the road shoulder. Another interesting note, almost 60% of the participants choose not to overtake and stay put after watching marker 6 which is "near an exit" and made marker 6 the only marker in yellow group. However, when a pair of brake lights are added in marker 7, which is the marker that the participants watched right after watching marker 6, the majority of the participants chose "not to overtake, slow down and keep a safe distance", and a good drop appeared in the percentage of selection 1 and 2 (the sum of selection 1 and 2 of marker 6 is 28.10%, and the sum of selection 1 and 2 of marker 7 dropped to 16.60%), which means less people were willing to overtake. The reason for this drop should be related to the factor that only appeared in marker 7 compared to marker 6, a pair of brake lights. From the driver's perspective, a sudden illumination of a pair of brake lights near an exit on the left lane can be a cue for a sudden lane change to the right. Almost every professional driver mentioned that one of the most dangerous and unexpected maneuvers that the front vehicles can make on the highway, is a sudden lane change to the right, right before an exit ramp. It can be due to the driver of the front vehicle did not pay attention to the navigation, or was distracted at the moment and forgot to change to the right lane beforehand. In this case, a pair of brake lights shows a good chance that the front vehicle might make a sudden lane change to the right, which is a factor that can potentially trigger a side-to-side motion.

When comparing marker 6 from the yellow group and marker 2 from the red group, although there are 35.4% of participants choose “not to overtake, slow down and keep a safe distance” in marker 2 and 10.40% choose this option in marker 6, the sum of selection 4 and 5 of marker 6 is 1% higher than what marker 2 has. Which means out of the 100 participants that watched the markers animation video, marker 6 is 1 person more than marker 2 when it comes to deciding to overtake or not, if we consider the sum of selection of 4 and 5 means “the will of not overtaking in general”. From Table 6, marker 2 and marker 6 are very close to each other. Marker 2 is “making consecutive lane changes”, and marker 6 is “near an exit”.

In the green group, there are 3 markers in total. Overtaking a semi-truck, brake lights turn on once, and brake lights turn on continuously. One common property among these three markers is that none of them exhibit a side-to-side motion or contain any factor that can trigger a side-to-side motion in the driver’s perspective. One comment from the online survey regarding marker 3 “brake lights turn on once” wrote “get what they are braking for, then overtake with caution”, the same comment appeared again in marker 4 “brake light turn on continuously”. The results show that the most important factor that drivers care about in marker 3 and 4 is not how many times the brake lights of the front vehicle turn on, it’s the reason why they brake, and if they do not see the factor that can predict the next step, they normally won’t do anything other than stay put and assess. Another comment from the online survey regarding marker 8 “overtaking a semi-truck” wrote “honk to get the truck driver’s attention and then overtake with caution”, this reflects that the drivers care about the factor that might trigger the next step, and as long as the truck driver aware of his/her presence, which means the possibility of the semi-truck’s side-to-side motion is highly reduced, the driver is willing to overtake with caution because the driver can see the factor that predicts the next step, at least from the driver's perspective.

Thus, based on the results of this study, the observable cues or the markers that the drivers check the most to determine the relative safety of passing cars, is the side-to-side motion of the front vehicle. Either it's an existing side-to-side motion such as:

- The front vehicle is wobbling
- The front vehicle is swerving and leaning towards one side of the lane
- The front vehicle is making consecutive lane changes

Or there is a factor that can potentially trigger a side-to-side motion, and the factors that can trigger a side-to-side motion are:

- Approaching an exit, front vehicle not in the exit lane
- Approaching an exit, brake lights turn on, front vehicle not in the exit lane
- Approaching a construction site or a police car on the road shoulder, front vehicle is in the lane that is right next to the road shoulder

CHAPTER 5. CONCLUSION, LIMITATION AND CONTRIBUTION

5.1 Conclusions

Conclusions of study 1: before the completion of this study, we do not know if drivers are more likely to make a lane-change under certain conditions, or if there is any correlation between the environmental conditions and the possibility of driver's decision on making a lane change. After the completion of this study, we now discovered some trends:

- Drivers make slightly more lane changes at a lower speed compared to at a higher speed.
- Drivers make more lane changes when there are more space to drive
- Drivers make more lane changes in impeded traffic than in unimpeded traffic on highways, the difference is more notable when during rush hours.
- Drivers make more lane changes during rush hours than in non-rush hours on highways in general.

Conclusions of study 2: before the completion of this study, we do not know what markers exist that would indicate that a vehicle will (unsafely) change lanes without a directional signal, and we do not know what the autonomous vehicles can do to avoid an accident caused by an unsafe lane change, at least the current self-driving technology is not advanced enough that autonomous vehicles can predict unsafe lane changes when passing a slower manual vehicle. After the completion of this study, there are some markers that can help the autonomous vehicles to better understand human driver behaviors and potentially predict these unsafe lane changes to avoid collisions. These markers are either existing side-to-side motion of the front vehicle or a factor that might trigger a side-to-side motion of the front vehicle in the environment.

5.2 Contributions

This dissertation contributes to the body of knowledge on lane changing precursors that come from human driver behaviors and the environments, it impacts the autonomous driving industry by providing a fresh angle on top of the existing widely used method of building mathematical models to predict vehicle lane changes, it provides an angle from a human factors perspective when looking at this problem, and it may inspire others to think about the new approaches on how to better understand human drivers behaviors and also its relationship with the environments, and keep on enrich the knowledge body of this topic. Eventually, all manual vehicles will be replaced by fully autonomous vehicles in human society, but before that, understanding the meaning of certain human driver's maneuver, how do human drivers maneuver in certain environmental conditions, and an ongoing enrichment of this knowledge base is still worth looking for the development of autonomous driving. This dissertation also provides contribution to the field by developing 2 standardized tools. One is the keyword-to-marker table that helps the researchers to extract lane change markers from the interview record by identifying keywords, the other one is a novel method of calculating traffic density.

CHAPTER 6. FUTURE WORK

Section 1 – Future work on lane changes

In study 1, the environmental factors on highways that might influence a driver's decision on making a lane change were investigated, and in study 2, the markers that might indicate a possible (unsafe) lane change of the front vehicle were investigated, no previous study was found on either of the topic. This is the first study on the topic of the lane change markers and environmental factors related to lane changes and there can be numerous future work for the people in the community to work on, the topic listed below worth furth investigation:

Firstly, we don't know if all the environmental factors are included or not, and we don't know how much of the effect they have on driver's decision of making lane changes. If we do, that can first contribute to the body of knowledge on this topic, and it would be a more comprehensive reference for the researchers and engineers in the autonomous driving industry to review.

Secondly, we don't know if certain drivers drive more often during the non-rush hours vs the rush hours, and if they prioritize the factors to make lane changes differently than the rush-hour drivers. If we know this, the engineers in the autonomous driving industry can anticipate these type of drivers and teach the self-driving cars to act accordingly.

Thirdly, we don't know how to quantify the effects into probabilities, I kept on discussing the effects that these environmental factors have on driver's decision of making lane changes, but how much is the effect, is there 25% more chance for drivers to make lane changes on an impeded road than on an unimpeded road? We are not sure, the conclusions in this study are more like ideas rather than actual figures, if we can somehow quantify the effects and package the conclusions in a way that is easier for the autonomous car companies to implement, that would be more meaningful.

Fourthly, we don't know if all the lane changing markers are included or not. There can be a lot more lane changing markers that drivers use to determine the relative safety when pass a front vehicle, especially a combination of motion and environment. If we do, it would be a more comprehensive reference for the researchers and engineers to review.

Number five is that we don't have standardized tools or clear thresholds to categorize markers, for instance, the keyword-to-marker table I made in this study, that is one tool that other researchers can use when they extract markers from interview records, but we are still not clear on, such as "to what extent, can we say a lane change marker should be considered as a marker that presents a higher level of safety concern than other ones? What is the clear distinguishing line here?" If we do have more tools and standardized thresholds on this, the results in this field can be more meaningful and trustworthy.

And the last one is, we are not sure about the validity of these tools, so they still need to be verified, this can be applied to a lot of human factors tools and standards.

Section 2 – Lane changes of autonomous vehicles

At the end of this report, I would like to talk about what additional work would have to be accomplished to provide a direct useful information to those who work in that autonomous vehicle industry. The environmental factors and markers investigated in this work, they either have some levels of effect on driver's decision of making a lane change, such as drivers make less lane change when there are more space to driver; or they exhibit a higher possibility that a driver is going to make a lane change, such as an existing side-to-side motion of the front vehicle. However, these effects and possibility still need to be quantified into actual figures so that it can be directly used. One possible way is to look at the recorded data such traffic footage and see if the listed predictors/markers in this study actually have an influence on the driver's lane changes. This way the results are validated through truth, not just what a person think if a front car is going to make a lane change. This might be the next step for quantifying the possibility of driver's decision on making a lane change, if traffic footage is accessible.
Another important question that worth some further investigation and discussion is that how can the results from this study be used in making an autonomous vehicle more predictable when it comes to lane changes, so that during this period of manual vehicle and autonomous vehicle

transition, both the human driver in the AV and in the other manual vehicles on road can easily predict if an AV is going to make a lane change.

One perspective of this study is that, it looks at how human drivers behave before making an unsafe lane change, what are the possible markers before an unsafe lane change, so that the engineers and researchers can teach the AV to look for these types of behaviors when driving around manual vehicles to avoid potential collision.

We believe that the findings of this study can reflect some key points for the AVs to mimic so that when an AV makes an unsafe lane change, it can act naturally in a way that fits how human drivers predict an unsafe lane change of a front vehicle. And therefore, not leaving them an impression that AVs are unpredictable, but rather fairly stable on the roads. Theoretically speaking, AVs are more predictable than human drivers, they follow a certain set of rules written by their manufacturers. Unlike human drivers that might forget to use turn signals, they should use turn signals for all planned lane changes. However, under certain circumstances where they must execute unplanned lane changes for certain reasons, what the AV should do to naturally exhibit its lane changing intention to the drivers behind it? One of the possible way that the AV can do is to exhibit a side-to-side motion to warn the other drivers that it is going to make a lane change, because based on the result of the lane change markers study, whenever a driver sees a side-to-side motion of the front vehicle, they intuitively assume this vehicle is not in a stable condition and overtaking it is not a safe thing to do. However, the magnitude of the side-to-side motion of the AV still need to be looked at, if the magnitude of the motion is too much, it might introduce safety hazards of the AV itself such as losing control. If the magnitude is too little, the other human driver might not be able to catch that warning.

We also believe that the findings of this study can reflect some key points on how can an AV make itself more predictable towards the driver/monitor inside of it before it makes an lane change. One possible approach is to let the driver/monitor inside decide whether it should execute a lane change, such as, prompting a question at an appropriate place inside the vehicle "Would you like me to make a lane change here?", then let the driver/monitor make a decision. A lot of times, the driver inside an AV is not comfortable with the AV's movement is due to the lack of information, or not well informed with feedbacks. If the driver/monitor inside an AV is well informed, and making lane change decisions, they might feel safer and more comfortable with the AV's movements. However, humans can get bored easily after a few rounds of repetitive decision making, and considering how "safe" they feel about the vehicle, they can loss interest in takin control of the vehicle and hand it over to the AV, then they might shift their attention elsewhere rather than the traffic. So, it is not a long-term measurement to let human drivers/monitors to stay in control all the time and expect them to stay alert while the AV is performing normally. Thus, the more important question here might be, how should the AVs convey its intention of making a lane change to the driver/monitor naturally so that they feel well informed, safe, and comfortable with the movement of the AV.

In one of the online surveys, one participant wrote "watch to see what is the thing they are braking for, then overtake with caution" after watching marker 4 (marker 4 is front vehicle's brake lights turn on continuously). This shows drivers do not feel confident to make their next move unless they have a good sense of what the next step of the current situation might be. The same principle applies here as well, drivers/monitors need to be informed through some type of media otherwise they don't feel safe and comfortable with the movement of the AV. Thus, look for a natural way to convey the lane change intention of the AV to the driver/ monitor is the next problem to solve.

One possible natural way is to convey this message through tactile signal through seats to the driver/monitor inside of an AV. In the lane change markers study, we found a side-to-side motion of a front vehicle is how drivers tell if a front vehicle might make an unsafe lane change. If a side-to-side motion of a car implies "a potential lane change" inside the drivers' head, it might worth some investigation to see if we can naturally convey a piece of vectorial information in an non-obstructing way to the driver/monitor inside of an AV. The picture can look like this, the AV can let the driver/monitor decide on the lane changes at first. After the driver choose to handover all the controls and the AV is going to make a lane change, the AV can simply give a tactile signal on the different side the seats to inform which direction it will make a lane change to, a few seconds before it actually execute the lane change. However, this is nothing more than a guess and imagination, there are a lot of things need to be tested, such as:

- What type of tactile signal can intuitively convey this information to the driver accurately
- Whether human consider this “tactile signal a few seconds before executing” is intuitive, and makes the whole process smoother than without
- How do drivers/monitors understand the tactile information under secondary tasks such as checking cell phone, laptop, conversation, GPS, etc.
- Tactile signal conveying accuracy might be related to cultures, gender, educational background, field of work, a baselining process might be necessary for any testing on information conveying.

REFERENCES

Al-Sobky, A.-S. A., & Mousa, R. M. (2016). Traffic density determination and its applications using smartphone. *Alexandria Engineering Journal*, *55*(1), 513–523. https://doi.org/10.1016/j.aej.2015.12.010

Brookhuis, K. A., de Vries, G., & de Waard, D. (1991). The effects of mobile telephoning on driving performance. *Accident Analysis & Prevention*, *23*(4), 309–316. https://doi.org/10.1016/0001-4575(91)90008-s

Carvalho, A., Williams, A., Lefèvre, S., & Borrelli, F. (2016). Autonomous cruise control with cut-in target vehicle detection. *Advanced Vehicle Control AVEC'16*, 93–98. https://doi.org/10.1201/9781315265285-16

Coifman, B. (2003). Estimating density and Lane inflow on a freeway segment. *Transportation Research Part A: Policy and Practice*, *37*(8), 689–701. https://doi.org/10.1016/s0965-8564(03)00025-9

Faubel, F., Georges, M., Kumatani, K., Bruhn, A., & Klakow, D. (2011). Improving hands-free speech recognition in a car through audio-visual voice activity detection. *2011 Joint Workshop on Hands-Free Speech Communication and Microphone Arrays*. https://doi.org/10.1109/hscma.2011.5942412

Fitch, G. M., Lee, S. E., Klauer, S. G., Hankey, J. M., Sudweeks, J. D., & Dingus, T. A., Analysis of lane-change crashes and near-crashes (2009). Washington, DC; U.S. Dept. of Transportation, National Highway Traffic Safety Administration.

Gipps, P. G. (1986). A model for the structure of Lane-changing decisions. *Transportation Research Part B: Methodological*, *20*(5), 403–414. https://doi.org/10.1016/0191-2615(86)90012-3

He, J., Chaparro, A., Nguyen, B., Burge, R. J., Crandall, J., Chaparro, B., Ni, R., & Cao, S. (2014). Texting while driving: Is speech-based text entry less risky than handheld text entry? *Accident Analysis & Prevention*, *72*, 287–295. https://doi.org/10.1016/j.aap.2014.07.014

Hill, L., Rybar, J., Styer, T., Fram, E., Merchant, G., & Eastman, A. (2015). Prevalence of and attitudes about distracted driving in college students. *Traffic Injury Prevention*, *16*(4), 362–367. https://doi.org/10.1080/15389588.2014.949340

Huang, F. L., & Cornell, D. G. (2012). Pick your poisson: A tutorial on analyzing counts of student victimization data. *Journal of School Violence*, *11*(3), 187–206. https://doi.org/10.1080/15388220.2012.682010

Jin, L., Xian, H., Niu, Q., & Bie, J. (2015). Research on Safety Evaluation Model for in-vehicle secondary task driving. *Accident Analysis & Prevention*, *81*, 243–250. https://doi.org/10.1016/j.aap.2014.08.013

Karacasu, M., & Er, A. (2011). An analysis on distribution of traffic faults in accidents, based on driver's age and gender: Eskisehir Case. *Procedia - Social and Behavioral Sciences*, *20*, 776–785. https://doi.org/10.1016/j.sbspro.2011.08.086

Khelfa, B., & Tordeux, A. (2020). Dynamic Safety Analysis for automated driving. *Proceedings of the 30th European Safety and Reliability Conference and 15th Probabilistic Safety Assessment and Management Conference*. https://doi.org/10.3850/978-981-14-8593-0_4205-cd

Khushaba, R. N., Kodagoda, S., Liu, D., & Dissanayake, G. (2013). Muscle computer interfaces for driver distraction reduction. *Computer Methods and Programs in Biomedicine*, *110*(2), 137–149. https://doi.org/10.1016/j.cmpb.2012.11.002

Knoop, V. L., Keyvan-Ekbatani, M., de Baat, M., Taale, H., & Hoogendoorn, S. P. (2018). Lane change behavior on freeways: An online survey using video clips. *Journal of Advanced Transportation*, *2018*, 1–11. https://doi.org/10.1155/2018/9236028

Ma, Y., Yin, B., Jiang, X., Du, J., & Chan, C. (2020). Psychological and environmental factors affecting driver's frequent lane-changing behaviour: A national sample of drivers in China. *IET Intelligent Transport Systems*, *14*(8), 825–833. https://doi.org/10.1049/iet-its.2019.0558

Madden, M., & Rainie, L. (2010). Adults and Cell Phone Distractions. *Pew Internet & American Life Project*.

Nemme, H. E., & White, K. M. (2010). Texting while driving: Psychosocial influences on young people's texting intentions and behaviour. *Accident Analysis & Prevention*, *42*(4), 1257–1265. https://doi.org/10.1016/j.aap.2010.01.019

Ng-Thow-Hing, V., Bark, K., Beckwith, L., Tran, C., Bhandari, R., & Sridhar, S. (2013). User-centered perspectives for Automotive Augmented Reality. *2013 IEEE International Symposium on Mixed and Augmented Reality - Arts, Media, and Humanities (ISMAR-AMH)*. https://doi.org/10.1109/ismar-amh.2013.6671262

Nunes, L., & Recarte, M. A. (2002). Cognitive demands of hands-free-phone conversation while driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *5*(2), 133–144. https://doi.org/10.1016/s1369-8478(02)00012-8

Park, H. S. (2013). In-vehicle ar-HUD system to provide driving-safety information. *ETRI Journal*, *35*(6), 1038–1047. https://doi.org/10.4218/etrij.13.2013.0041

Recarte, M. A., & Nunes, L. (2002). Mental load and loss of control over speed in real driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *5*(2), 111–122. https://doi.org/10.1016/s1369-8478(02)00010-4

Riera, L., Ozcan, K., Merickel, J., Rizzo, M., Sarkar, S., & Sharma, A. (2020). Detecting and tracking unsafe lane departure events for predicting driver safety in challenging naturalistic driving data. *2020 IEEE Intelligent Vehicles Symposium (IV)*. https://doi.org/10.1109/iv47402.2020.9304536

Salvucci, D. D., & Liu, A. (2002). The time course of a lane change: Driver control and eye-movement behavior. *Transportation Research Part F: Traffic Psychology and Behaviour*, *5*(2), 123–132. https://doi.org/10.1016/s1369-8478(02)00011-6

Schmidt, E., Decke, R., Rasshofer, R., & Bullinger, A. C. (2017). Psychophysiological responses to short-term cooling during a simulated monotonous driving task. *Applied Ergonomics*, *62*, 9–18. https://doi.org/10.1016/j.apergo.2017.01.017

Sen, B., Smith, J. D., & Najm, W. G., Analysis of lane-change crashes 9–9 (2003). Cambridge, MA; U.S. Dept. of Transportation, National Highway Traffic Safety Administration.

Shawky, M. (2020). Factors affecting lane change crashes. *IATSS Research*, *44*(2), 155–161. https://doi.org/10.1016/j.iatssr.2019.12.002

Sun, D. J., & Elefteriadou, L. (2012). Lane-Changing Behavior on Urban Streets: An "in-vehicle" field experiment-based study. *Computer-Aided Civil and Infrastructure Engineering*, *27*(7), 525–542. https://doi.org/10.1111/j.1467-8667.2011.00747.x

Teh, E., Jamson, S., Carsten, O., & Jamson, H. (2014). Temporal fluctuations in driving demand: The effect of traffic complexity on subjective measures of workload and driving performance. *Transportation Research Part F: Traffic Psychology and Behaviour*, *22*, 207–217. https://doi.org/10.1016/j.trf.2013.12.005

Toledo, T., Koutsopoulos, H. N., & Ben-Akiva, M. E. (2003). Modeling integrated lane-changing behavior. *Transportation Research Record: Journal of the Transportation Research Board*, *1857*(1), 30–38. https://doi.org/10.3141/1857-04

Vechione, M., Balal, E., & Cheu, R. L. (2018). Comparisons of mandatory and discretionary lane changing behavior on freeways. *International Journal of Transportation Science and Technology*, *7*(2), 124–136. https://doi.org/10.1016/j.ijtst.2018.02.002

Velichkovsky, B. M., Dornhoefer, S. M., Kopf, M., Helmert, J., & Joos, M. (2002). Change detection and occlusion modes in road-traffic scenarios. *Transportation Research Part F: Traffic Psychology and Behaviour*, *5*(2), 99–109. https://doi.org/10.1016/s1369-8478(02)00009-8

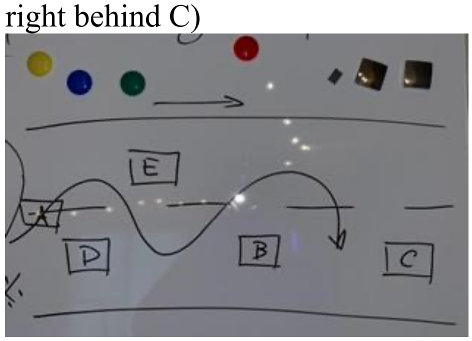
World Health Organization. (2011, February 14). *Mobile phone use: A growing problem of driver distraction*. World Health Organization. Retrieved February 12, 2022, from https://www.who.int/publications-detail-redirect/mobile-phone-use-a-growing-problem-of-driver-distraction

Yang, M., Wang, X., & Quddus, M. (2019). Examining lane change gap acceptance, duration and impact using naturalistic driving data. *Transportation Research Part C: Emerging Technologies*, *104*, 317–331. https://doi.org/10.1016/j.trc.2019.05.024

Zeeb, K., Buchner, A., & Schrauf, M. (2015). What determines the take-over time? an integrated model approach of driver take-over after automated driving. *Accident Analysis & Prevention*, *78*, 212–221. https://doi.org/10.1016/j.aap.2015.02.023

APPENDIX 1: MARKER-TO-SCRIPT TABLE

A is the prop sedan, B is the participant's sedan, the animation's viewing angle is always "what the participants should see if they are sitting in vehicle B （driver seat）"

#	Markers	Further information	Script for animation
1	Swerving and leaning towards one side of the lane	Testing one side should be sufficient already for this maker, adding testing for another side should not change the result.	1. A and B are both driving normally on a 2-lane highway 2. A is on the right lane, B is on the left lane 3. A is in front of B 4. B accelerates slightly to initiate the "overtake" action 5. A slowly swerve and lean towards the left side of the lane 6. Stop the animation when A's left wheels are about to touch the dashed lines
2	Making consecutive lane changes	Multiple other vehicles are required in this testing, C is in front of B, D and E are behind B on different lanes, as shown in this picture, and all vehicles are moving to the right direction (the pic is no longer 100% correct, C moved to the left lane, and A stops right behind C) 	1. All vehicles drive normally as shown in the picture (move C to the left lane, and A stop right behind C) 2. A appears in the rear mirror of B, and starts making consecutive lane changes and pass multiple vehicles in the rear mirror (participant sees A's maneuvers in the rear mirror) 3. A overtakes B on the left lane 4. A stops right behind C 7. B accelerates slightly to initiate the "overtake" action 5. Stop the animation

3	Brake lights turn on once		1. A and B are both driving normally on a 2-lane highway 2. A is on the right lane, B is on the left lane 3. A is in front of B 4. B accelerates slightly to initiate the “overtake” action 5. A’s brake lights turn on once and reduce speed (speed reduction is standard across all conditions) 6. Stop the animation when this action appears
4	Brake lights turned on continuously		1. A and B are both driving normally on a 2-lane highway 2. A is on the right lane, B is on the left lane 3. A is in front of B 4. B accelerates slightly to initiate the “overtake” action 5. A’s brake lights turn on continuously and reduce speed slightly (speed reduction is standard across all conditions) 6. Stop the animation when this action appears
5	Wobble motion		1. A and B are both driving normally on a 2-lane highway 2. A is on the right lane, B is on the left lane 3. A is in front of B 4. B accelerates slightly to initiate the “overtake” action 5. A wobbles from the left to the right side of the lane slightly 6. Stop the animation when this action appears

6	Near an exit		1. A and B are both driving normally on a 2-lane highway 2. A is on the left lane, B is on the right lane 3. A is in front of B 4. B accelerates slightly to initiate the “overtake” action 5. Highway exit enters the frame 6. Stop the animation when the exit is at an appropriate distance away from A (a distance that allows A to change lane and exist)
7	Overtaking a semi-truck	A is a semi-truck, C is a prop sedan in front of A	1. A and B and C are all driving normally on a 2-lane highway 2. A and C are on the right lane, B is on the left lane 3. C is in front of A, A is in front of B 4. C is slightly slower than A, A is slightly slower than B 5. When B is close to A 6. B initiate the “overtake” action 7. B gets closer to A as B is moving faster than A 8. A gets closer to C as A is moving fast than C 9. Stop the animation when B is about to pass A

8	Police car on the road shoulder		1. A and B are both driving normally on a 2-lane highway 2. A is on the right lane, B is on the left lane 3. A is in front of B 4. B accelerates slightly to initiate the “overtake” action 5. A police car with red and blue siren light appears on the right road shoulder and gets closer 6. Stop the animation when the police car is at an appropriate distance 7. Appropriate distance means A still have the room to change to the left lane before passing the police car
9	Construction site on the road shoulder	Construction site will look like a few orange cones on the ground, with a worker holding a “slow” sign on his hand facing the traffic  	1. Everything is the same as above except for replace the construction site with the police car

10	Near an Exit + brake light		1. A and B are both driving normally on a 2-lane highway 2. A is on the left lane, B is on the right lane 3. A is in front of B 4. B accelerates slightly to initiate the “overtake” action 5. Highway exit enters the frame 6. A hits brake once 7. Stop the animation when the exit is at an appropriate distance away from A (a distance that allows A to change lane and exist)

APPENDIX 2: KEYWORD-TO-MARKER TABLE

Keywords-to-marker table

Keywords	Answer	Marker
Consecutive lane changes	"if they have been constantly lane changing"	Making consecutive lane changes
Consecutive lane changes	"passing many cars while speeding, like a fish pattern"	
Consecutive lane changes	"when they drive unconventionally, merging between cars snake like"	
Sudden traffic ahead	"instead of using brakes, they just merge over"	
Speed	"when they are going really fast"	
Lane position	"can feel if the car wants to make lane change based on its position"	Swerving and leaning towards one side of the lane
Driving habit	"some drivers are wild, it is all the driving habit"	
Driving habit	"it's a problem of driving habit, a lot of people don't have the habit of using blinkers"	
Lane position	"they will lean towards the dashed line"	
Intuition	"by their car direction, it is easy to tell and predict they are going to change lanes"	
Intuition	"we start preparing when we feel that the car is about to change lane"	
Learning influence	"driving school does not teach very well and nowadays people only need to spend a month in driving school	
Lane position	"the head of the vehicle will lean towards the middle"	
Learning influence	"did not have using blinkers in mind when learning how to drive"	
Lane position	"their tire angle and vehicle speed"	
Lane position	"when they want to merge, they will naturally lean towards your lane"	
Intuition	"just by looking at them you can just tell they will merge"	
Learning influence	"some people were not taught the importance of it while learning how to drive"	

Lane position	"if they are riding on the dashed line, attempting to make the lane change"	
Speed	"if their speed is close to mine then they might cut in"	
intuition	"you can just feel when they are about to make a lane change"	
Sleep	"if they are sleeping/tired"	
Sudden traffic ahead	"if there is traffic ahead and I don't want to suddenly stop"	Brake lights turn on once
Obstacle ahead	"if there are obstacles ahead"	
Speed	"if they want to merge they will hit the brake, even though they don't use blinkers"	
Obstacle ahead	"if there are unusual objects ahead"	
Sudden traffic ahead	"if there is traffic ahead and the driver has to make a sudden stop"	
Speed	"if they suddenly speed up or suddenly slow down, that is an indication"	Brake lights turned on continuously
traffic	"when they are in traffic and there is an open spot in the adjacent lane"	
Speed	"if they slow down, or speed up to match the traffic speed to try to merge"	
Driving habit	"doesn't realize or feel the need to use blinker, it is like they don't care"	
Using phone	"if one of their hands is occupied while calling"	Wobble motion
Lane position	"they want to merge left, so they lean left, then back, and left again"	
Lane position	"if they are switching between driving towards the left and then driving to the right"	
Lane position	"the car doesn't look safe and steady, they might be asleep or drunk"	
carelessness	"If they think it doesn't matter whether or not to use blinkers"	
intuition	"based on their body language and where their attention is"	
Missed Exit	"when they're almost to an intersection and suddenly merge"	Near an exit
Carelessness	"they forgot"	Near an exit + brake light
Missed Exit	"most of the time close to an exit or to a rest area, they brake and doesn't have time to use blinker"	
Missed exit	"some drivers don't pay attention and they make a sudden stop and lane change"	
Missed exit	"missed rest area or missed exit"	
Missed exit	"people will be close to an exit and realize they missed it, so they will merge 3-4 lanes at once"	

Missed exit	“many people get distracted and realize they are about to miss an exit”	
Missed exit	“they’re unfamiliar with the road or are using GSP”	