Title
Latent Market Segments for the Adoption of Fully Automated Vehicles in California

Author
Matson, Grant

Publication Date
2021

Peer reviewed|Thesis/dissertation

Latent Market Segments for the Adoption of Fully Automated Vehicles in California

By

GRANT MATSON
THESIS

Submitted in partial satisfaction of the requirements for the degree of

MASTER OF SCIENCE

in

Transportation Technology and Policy

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

____

Giovanni Circella, Chair

____

Susan Handy

____

Yongsung Lee

____

Daniel Sperling

Committee in Charge

2021

Table of Contents

List of Tables

List of Figures

List of Acronyms

Acronym	Meaning
AIC	Akaike's Information Criterion
AV	Automated Vehicle
BIC	Bayesian Information Criterion
BEV	Battery Electric Vehicle
CACC	Connected Adaptive Cruise Control
CAV	Connected Automated Vehicle
EFA	Exploratory Factor Analysis
IC	Information Criteria
ICEV	Internal Combustion Vehicle
LCA	Latent Class Analysis
LL	Log-likelihood
MTC	Metropolitan Transportation Commission
NCST	National Center for Sustainable Transportation
PHEV0	Plug-in Hybrid Vehicle
PMT	Person Miles Traveled
SACOG	Sacramento Area Council of Governments
SANDAG	San Diego Associations of Governments
SAV	Shared Automated Vehicle
SCAG	Southern California Association of Governments
SPSS	Statistical Product and Service Solution
TNC	Transportation Network Companies
USDOT	United States Department of Transportation
VMT	Vehicle Miles Traveled
VOT	Value of Time
WTP	Willingness to Pay

Abstract

Automobile manufacturers are pushing the rapid development of automated vehicles (AVs) despite a limited understanding of consumer demand and potential impacts on travel behavior. An effective policy response depends on an improved understanding of who will be interested in (early) adoption of AVs, what the users' preferred business models (private vs. shared) will be, and how the eventual adoption of shared automated vehicle (SAV) services will likely impact personal/household vehicle ownership levels. This study addresses these topics through a market segmentation analysis using latent-class modeling of data from a custom-designed transportation survey of California residents. I use sociodemographics, general attitudes, and current AV familiarity to define the segments. The analysis uncovered three latent classes: (1) *AV Early Adopters* who are enthusiastic about the fully automated AV scenarios presented (private ownership and shared services) and are ready to adopt shared AVs instead of owning other personal/household vehicles, (2) *AV Curious* who are less enthusiastic about AVs than the prior class and more likely to maintain their current vehicle ownership in the future even when using shared AV services, and (3) *AV Hesitant* who are resistant to using either private or shared AVs and would not be interested in reducing their vehicle ownership levels if they had access to SAV services. The characteristics of the three classes provide a basis for actionable recommendations for both private companies and transportation policy makers.

Keywords: Latent Class Analysis, Automated Vehicles, Vehicle Ownership, Market Segmentation

1. Introduction

As the possibility of fully automated vehicles (AVs) comes closer to reality researchers can play an important role in the development of a policy response by examining how this new technology will be accepted by consumers and how it will impact travel behavior. Recent research into AVs has shown that AV deployment and adoption have many potential positive as well as negative effects. The positive effects range from (eventual) reduced traffic congestion and pollutant emissions, higher roadway throughput, increased mobility, to a reduction of safer roadway costs with the inverse of these as the potential negatives (Litman 2014). Beyond basic underlying acceptance of the new technology, it will be important to understand the rate of adoption and the adoption patterns among various segments of the population when this technology becomes available on the market. Anticipating the preferred ownership model and the potential impacts that AV adoption might have on household vehicle ownership is also imperative, particularly because it currently appears that shared automated vehicle (SAV) fleets that operate like transportation network companies (TNCs) will be deployed before private ownership becomes available. As the TNC deployment of AVs is driven by the development cycle of the technology and not necessarily underlying consumer preference, my thesis investigates consumer response to both shared and privately-owned AVs. Developing an understanding of the potential for the speed and magnitude of adoption during the transition period between the current transportation status quo (with no/extremely limited AVs) to a transportation future dominated by AVs is important for addressing issues such as road safety or congestion that may arise during the transition, thereby delaying the realization of the potential benefits. Two studies on this topic suggested that until the AV market reaches a saturation point of 30% the benefits of AVs will not be actualized (Ye and Yamamoto 2018, Nishimura, Fujita et al. 2019).

While it is easy to get tunnel vision on all the benefits of AVs, it is important to take a step back and consider the potential negative effects of AVs. AVs have the potential to further cement us into our current land-use and development preference of single-passenger vehicle focused infrastructure by making travel in AVs so easy people would not even consider alternative modes or designs. This would likely also negatively impact public transportation as AVs could siphon off passengers which would put further financial stress on these public services. It is important to have a robust public transportation system as it provides critical services to its users and when used at sufficient levels is more environmentally sustainable. It is important researchers, developers, and policymakers acknowledge the negative effects of AVs so they can find ways to mitigate the negative effects directly in the product design or indirectly via broader policies to ensure AVs are deployed equitably and sustainably.

Businesses are driving the development of AV-related technology at such an accelerated pace that the research community must conduct forward-looking research to try and anticipate the implications of this technological surge, and policy makers need to get ready to regulate the many aspects associated with this disruptive technology. It is estimated that an additional $80 billion in revenue for AV manufacturers captured by 2030 with the sales of automated vehicles and automated features (Jiang, Petrovic et al. 2015). With this significant future revenue stream available, there is so much interest from manufacturers and technology companies in bringing it to the market as soon as possible AVs and its supporting technology.

One of the critical requirements for AVs to produce many of their positive outcomes is achieving widespread acceptance of shared use over private ownership (Sperling 2018). One of the societal benefits of the shared ownership model of AVs is the potential for the associated car shedding, *i.e.*, reducing the number of cars a person/household owns (Lavieri, Garikapati et al. 2017). While car shedding may reduce the revenue of AV manufacturers, in recent years we have seen them enter the shared vehicle market, which would provide the means for them to recapture this revenue. This would generate a reduced need for parking as well as other implications for land use and land consumption in cities thus prompting a potentially more efficient use of the existing resources. Whatever the business model behind the deployment of AVs, a person’s willingness to reduce car ownership is an indicator of a pertinent shift in travel behavior that can lead to realizing the benefits of AVs. That is one of the topics I planned to investigate in my thesis.

Framing AVs in this manner lays the foundation for the study to seek answers to the following questions:

- Who will compose the different segments of the early AV adoption market?
- What ownership model would be preferred, *i.e.,* private vs. shared. By whom and why?

While research in this field is inherently explorative given the lack of individuals' exposure to the topic, as well as the degree of misinformation in some media (Charness, Yoon et al. 2018) and the speculative nature of predicting future behavior, it is still important to conduct this research as it can help inform other research, policy makers, and business in the development of the AV field into a thriving and beneficial component of future transportation systems.

To aid in the reading of this thesis, the following maps out the structure of the remaining sections. First, in the next section I summarize a review of the literature to present the current body of work related to the topic of this thesis, which will cover topics such as AV adoption, its benefits and costs, and related studies that highlight the gaps this research will address. Next, Section 3 presents a detailed account of the data collection efforts conducted for this research. While the data collection was part of a larger research project on which I worked during my degree at UC Davis, the relevant sections of the survey were designed with my research goals in mind, and I led the administration and management of the data collection process. Then the analytical methods utilized in the research are presented to provide a thorough accounting of the underlying thought processes for the major components of the data analysis, which is centered on the application of the latent class analysis (LCA) but also includes a factor analysis of individual attitudes. The following section delves into the conclusions that can be drawn from the analysis with an emphasis on practical policy implications that address current concerns among transportation planners and policy makers. This is then followed by a brief discussion on the limitations of this research and how future research efforts on this topic could address them while continually pushing the body of knowledge in the field forward.

2. Literature Review

The rapid growth in technological development and general interest in AVs has led to the development of a wide range of research. One topic of interest has been quantifying the potential benefits of using privately owned AVs and shared AVs. A major selling point for AVs is the potential for a cost reduction. Litman (2020) did a meta-analysis on the topic by compiling cost estimates from six previous efforts to estimate the associated costs of ownership and use (Litman 2009, Johnston and Walker 2016, Stephens, Gonder et al. 2016, Association 2017, Bösch, Becker et al. 2017, Keeney 2017). This analysis concluded that AVs provide the biggest cost savings over privately-owned, human-driven vehicles when part of a shared AV service. Private ownership of AVs has an increased cost over non-automated vehicles because of the added expense of purchasing the new technology (Litman 2020). While this study combined the work of six other studies they were all slightly incomplete as they did not capture all costs associated with mode choice which are a combination of monetary and non-monetary elements. This gap was recently addressed by the research by Compostella, Fulton et al. (2020). This study did an in-depth analysis of all costs (fix, variable, and hedonic) for a variety of vehicle powertrains (internal combustion engine (ICEV), plug-in hybrid (PHEV), and battery electric (BEV) and ownership models (privately owned, ridesourcing, and shared ridesourcing) for AVs and traditional vehicles. Figure 1 summarizes some of these findings and shows that cost savings are fully realized until automated vehicles are used in a shared ridehailing service. This is the results of the vastly greater expected yearly VMT of a vehicle in a ridehailing service (80,000 miles) compared to a privately-owned vehicle (15,000 miles). While this will increase the variable costs, such as fuel and cleaning, it allows for the fixed costs to amortized over a much larger annual VMT thus pushing the cost per mile down.

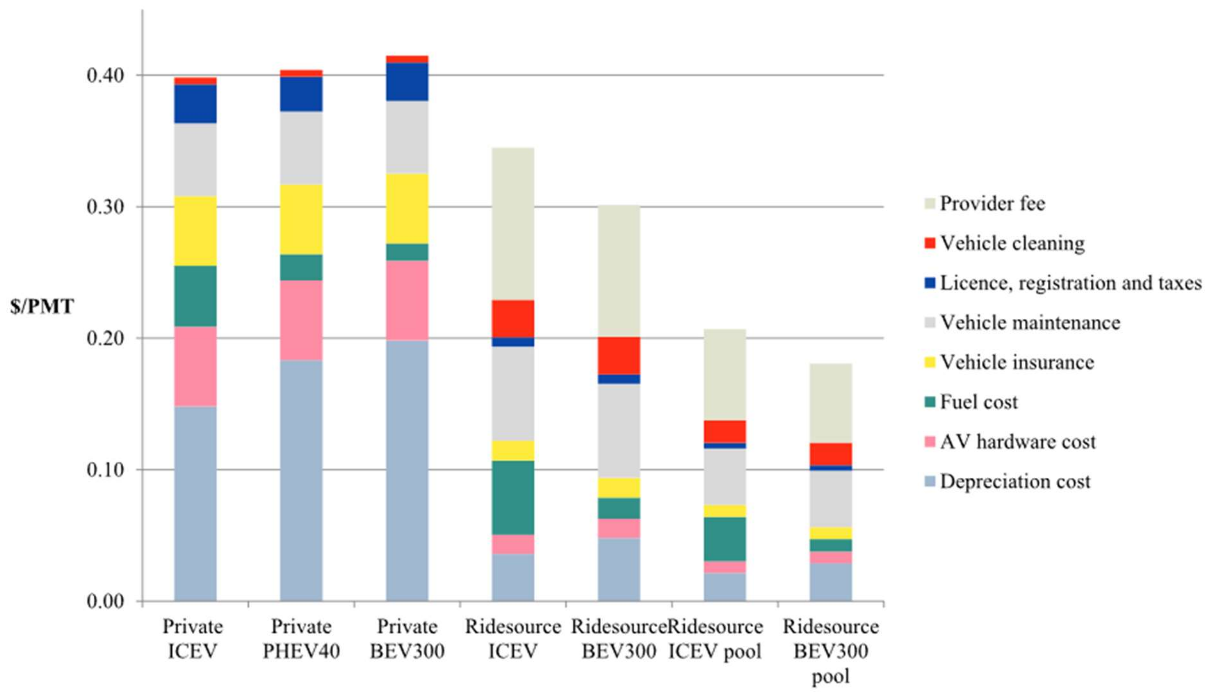


Figure 1. Automated Midsize Car Costs ($/Person Miles Traveled)
Source: Reprinted with approval from Compostella, Fulton et al. (2020)

The operating costs are only part of the picture with another key element to consider as part of the user's cost is accounting for the value of time (VOT) associated with travel in AVs because if a person can regain some valuable time during their travels the value proposition of AVs grows substantially. For example, Malokin et al. (2021) investigated the impacts of multitasking while traveling on VOT among commuters in northern California. Using revealed-preference data to estimate a mode choice model, they concluded that the ability to multitask while traveling might considerably modify the VOT, posing a potentially large advantage for AVs (Malokin et al. 2021). They also observed that while younger adults (millennials) are in general more price sensitive they are willing to pay more than non-millennials for the ability to use a laptop while commuting (Malokin, Circella et al. 2021). This study illustrates the potential for AVs to be embraced as it easily allows for this type of multitasking.

However, Singleton's 2018 analysis of the body of work related to AV VOT observed that the assumption of VOT of many simulation-based studies used the higher bounds of the benefits of highly automated AVs that are shown in Table 1 (Singleton 2018). Singleton suspected this to be influenced by the popular media and technology developers driving the conversation and highlights a growing set up studies (Cyganski, Fraedrich et al. 2015, Sivak and Schoettle 2016, Milakis, van Arem et al. 2017) that suggest an alternative impact on VOT from AVs, i.e. users will not utilize the time for productive activities so the effects should be more modest than what the simulation studies assumed. While there is no consensus on the scale of impact of AVs on VOT, at least the direction of the change does seem to always be consistently positive, pointing to a reduction in VOT associated with AV deployment.

Table 1. Autonomous Vehicle Value of Time (VOT) Assumptions in Simulation Studies

Study	Area	AV VOT Assumptions
Gucwa (2014)	San Francisco Bay Area, CA	100% of high-quality rail VOT; 50% of car driver VOT; zero
Spieser et al. (2014)	Singapore	30% of car driver VOT
Childress, Nichols, Charlton, and Coe (2015)	Seattle, WA	65% of car driver VOT (for high-income travelers only)
Davidson and Spinoulas (2015)	Brisbane, Australia	75-90% of car driver VOT (for lower-level AVs); 50-90% of car driver VOT (for higher-level AVs)
Kim, Rousseau, Freedman, and Nicholson (2015)	Atlanta, GA	50% of car driver VOT
van den Bern and Verhoef (2016)	United States and Netherlands	61-100% of car driver VOT
LaMondia, Fagnant, Qu, Barrett, and Kockelman (2016)	Michigan	75% of car driver VOT
Auld, Sokolov, and Stephens (2017)	Chicago, IL	25%, 50%, 75%, 100% of car driver VOT
Kockelman et al. (2017)	Austin, TX	100% of transit VOT; 50% of car driver VOT; zero

Source: Adapted from Singleton 2018

Another potential benefit of AVs is the ability to reduce crashes and other road safety issues. Arbib and Seba (2017) state in their report that “current safety data suggests at least a 90% reduction in the number of accidents involving A-Evs [automated-electric vehicles], relative to [manually-driven] ICE” because 94% of current accidents are caused by human error, which would not be a factor in a fully automated vehicle (Arbib and Seba 2017). In contrast, Sivak and Schoettle (2015) have a more practical view of the situation and suggest that the transition to a fully automated fleet will likely take several decades, and therefore the reduction in crashes will be minimal and may increase in the shorter term due to the mixed fleet (Sivak and Schoettle 2015).

AVs are also expected to alleviate many issues that currently affect the road network, such as congestion, inefficient signals, and parking limitations. Shaldover, Su et al. (2012) examined the effects of low levels of automation via adaptive cruise control and connected adaptive cruise control (CACC) by using a microsimulation model. They observed "conventional ACC is unlikely to produce any significant change in the capacity of highways" as users tend to set them at manual driving time gaps; however, "lane capacity increases approximately linearly from 2000 to 4000 as the percentage of CACC vehicles increases from zero to one hundred" (Shladover, Su et al. 2012). Mattas, Makridis et al. (2018) expanded on this initial research to quantify the benefits of AVs and connected AVs (CAV) for roadway capacities. Utilizing a traffic simulation model for different mixture scenarios of AVs and CAVs at different penetrations rates from 0% to 100% they observed that AVs generally have a negative effect on traffic speeds because of "more conservative driving of AVs compared to human drivers in acceleration behavior, but also in lateral movement" however when the CAV penetration was greater than 60% there were large increases in speed for the road network (Mattas, Makridis et al. 2018). Examining another element of roadway capacity, Pourmehrab, Elefteriadou et al. (2019) studied the effects of AVs on signalized intersections with an intelligent intersection control system via a traffic simulation model. When compared to a convention signal, a connected intersection signal to AVs experienced a reduction of 38-52% in average wait times depending on the AV utilization rate (Pourmehrab, Elefteriadou et al. 2019). These studies begin to illustrate the need for AVs (both connected and not connected) to achieve high penetration rates for their benefits on roadway capacity to be realized so understanding who will adopt this new technology, and when, will be crucial to achieving the desired levels.

It is anticipated that AVs will cause a dramatic reduction in parking demand as they will be able to perform other tasks such as drive home, circle a location, or pick up another passenger instead of being parked. Zhang, Guhathakurta et al. (2015) examined the effects of SAV on urban parking demand through an agent-based simulation. Their results suggested that users of SAV services would eliminate 90% of their parking demand (1.8% of total parking demand) even when the service has a low penetration rate of 2% of the total vehicle fleet (Zhang, Guhathakurta et al. 2015). It is worth noting that to achieve this there was a substantial increase in vehicle miles traveled (VMT) due to the associated cruise time waiting for their next customer. With this reduction in demand Zhang and Guhathakurta 2017 continued this line of inquiry to see how much this would in turn impact parking land use. Their simulation results showed that through higher vehicle utilization and lower ownership level, "parking land use can be reduced by approximately 4.5% once the SAVs start to serve 5% of the trips" which equates to each SAV freeing up more than 20 parking spaces (Zhang and Guhathakurta 2017). AVs can also change the physical design of parking spaces as space for vehicle access is eliminated and/or limited to two rows. Nourinejad, Bahrami et al. (2018) analyzed this topic and found that optimizing the design for AV parking lots can "decrease the need for parking space by an average of 62% and a maximum of 87%" (Nourinejad, Bahrami et al. 2018) thus creating denser parking lots within the existing footprint. Medina-Tapia and Robusté (2018) modeled the effects of AVs on urban mobility and observed that the cost associated with looking for parking was reduced by 28.8% compared to a manual vehicle given the ability of AVs to park themselves thus freeing up that time for the user and only incurring the operating costs during that time (Medina-Tapia and Robusté 2018). The research suggests that AVs likely will not demand as much land for parking while requiring smaller spaces and costing less time when needed, but they are not without concerns. For example, AVs allow for shifting parking location from the urban core to suburbs or less affluent neighborhoods to save on parking cost and time – this raises concerns about equity or the additional deadheading and cruising time, which will have negative effects on the environment due to the increase in VMT and the potential encouragement of urban sprawl.

In his book *Three Revolutions: Steering Automated, Shared, and Electric Vehicles to a Bette Future*, Sperling (2018) coalesces all of the benefits and issues of AVs into two clear scenarios of a Heaven and

Hell depending on how AVs are implemented. He argues to achieve the Heaven outcome of reduced VMT, reduction in costs, higher transit ridership, and repurposing of parking land use a broad set of forward-thinking policies need to be enacted early enough to ensure the desired outcome. In contrast, the Hell outcome of increased VMT, worsening traffic congestion and exacerbating environmental issues will happen by following the current business-as-usual case of private ownership with an aversion to shared services while letting for-profit companies set policy.

Despite the murky outlook for realizing the benefits of AVs for society and users, the technology developers are forging ahead with the development as they see a new market brimming with profits while steering policy to their needs. Jiang, Petrovic et al. (2015) describe two very different development paths by the automotive and technology companies. Automotive companies, *e.g.,* BMW, Ford, and Volvo, among others, are taking a more conservative approach with the incremental implementation of AV features while many technology companies, *e.g.,* Google's Waymo division, are implementing a more disruptive approach by jumping straight to full Level 5 automation (Jiang, Petrovic et al. 2015), which is tested typically in geofenced areas, at first, in preparation for broader market deployment. While it is still too early to tell which approach will lead to widely available AVs first, it is clear that they both have limitations as seen in the death caused by Uber's AV taxi service striking a pedestrian (Griggs and Wakabayashi 2018) and Tesla's partial automated vehicles being involved in fatal crashes (Krisher 2020).

Given the varying approaches to AV development, understanding how this new technology will be adopted in terms of the time frame and velocity are important given the opposed outcomes of the varying levels (low vs high penetration rates) and types (private vs shared) of AV adoption. Lavasani, Jin et al. (2016) examined this topic by estimating a market diffusion model based on the adoption of previous technologies, *i.e.*, conventional automobiles and hybrid electric vehicles for cost and pricing, and internet and cellphones for usage.

Figure 2 illustrates their market penetration curve for AVs at different market sizes of the current US fleet. Assuming a market size of 75%, the model estimates it would take about 10 years to sell the first 8 million units and would take 35 years for full saturation (Lavasani, Jin et al. 2016).

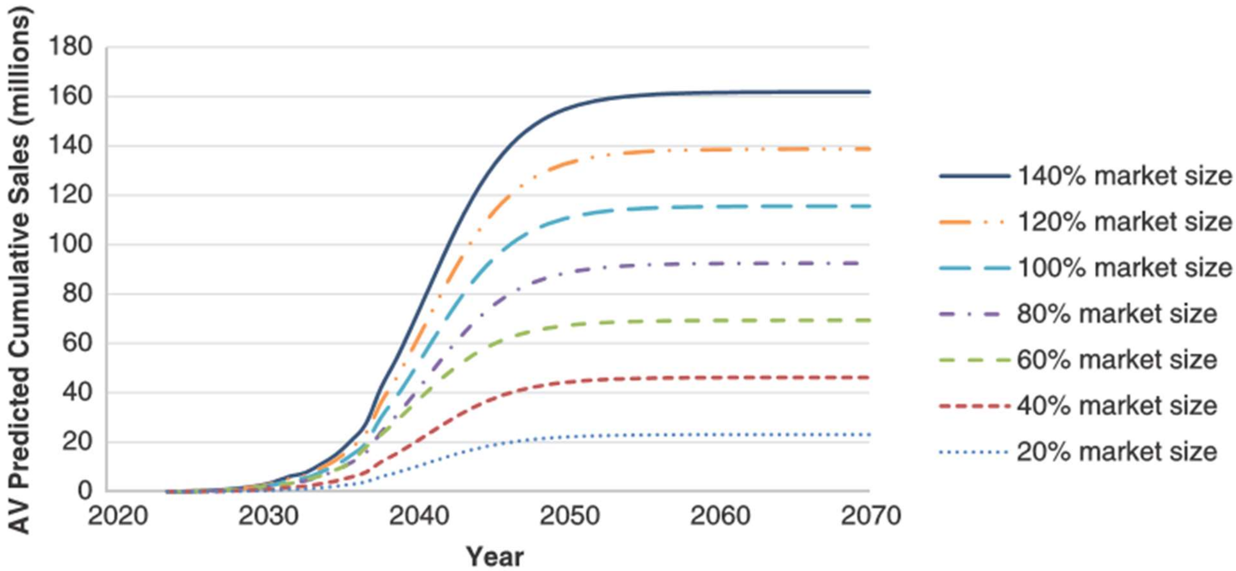


Figure 2. Sensitivity Analysis Results on Market Size
Source: Reprinted with approval from Lavasani, Jin et al. 2016

Bansal and Kockelman (2017) investigated AV adoption by using a simulated fleet modeling calibrated from survey data and their results suggest "without a rise in most people's WTP [willingness to pay], or policies that promote or require technologies, or unusually rapid reductions in technology costs" it seems unlikely that AVs will be the primary mode of transportation by 2045. Using a similar methodology, Nieuwenhuijsen, de Almeida Correia et al. (2018) further this line of research via a simulation model utilizing system dynamics to capture the uncertain, complex and dynamic nature of the innovation diffusion for AV adoption for each of the 5 SAE level of automation. They tested multiple scenarios but their "AV in bloom" closely mirrors the state of AV development in California (at least in the pre-COVID-19 pandemic society) as it is defined by having conditions in which "customer attitude is positive, economic growth is strong, the technology development is high and the policy is supportive". Figure 3 represents the progressive scenario results, which show that, even with it being the more aggressive of this study's scenarios, it would take an estimated period of 60 years for highly automated vehicles (Level 4 and 5) to reach full market saturation. While this research seems more methodologically robust than Lavasani, Jin et al. 2016 (2016)and Bansal and Kockelman 2017, I feel they make an unrealistic assumption that manual vehicles (Level 0) or the lower levels of automation (Level 1 and 2) will eventually be completely replaced, which does not seem reasonable as there will likely always be people that prefer manual driving for the enjoyment of it or out of necessity, *e.g.*, off-roading.

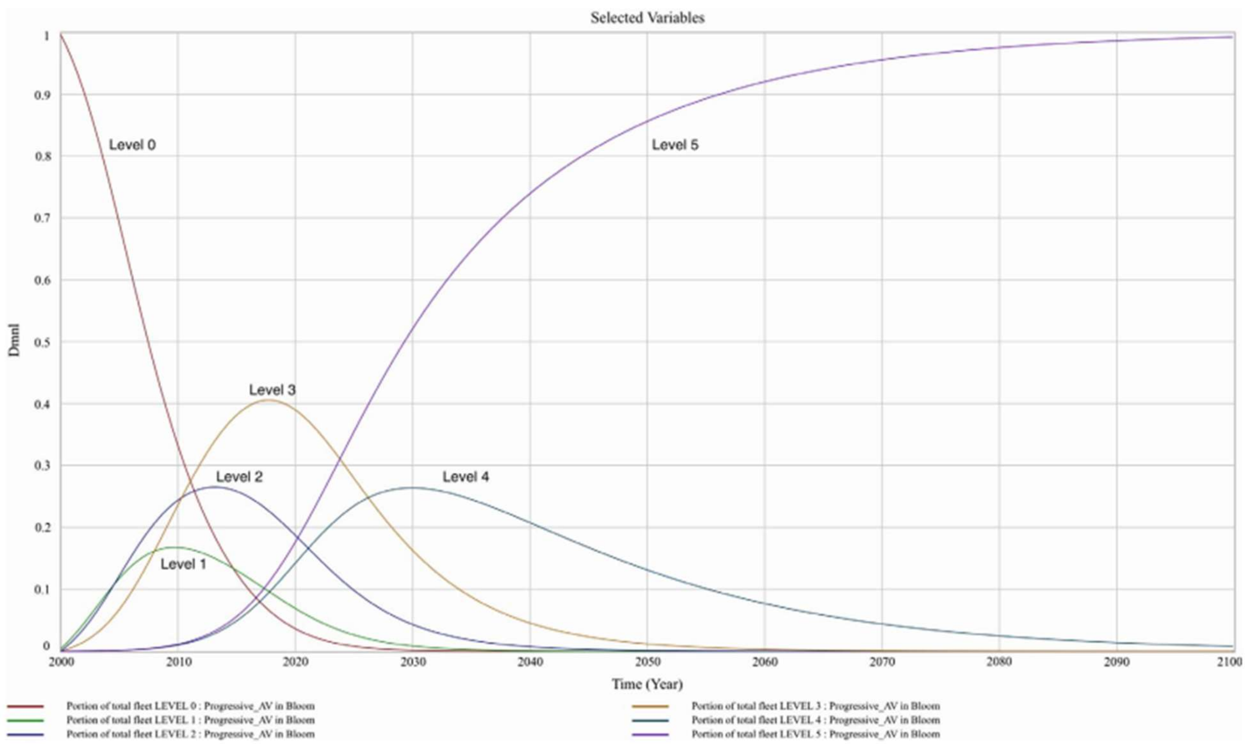


Figure 3. Market Penetration of "AV in Bloom" Progressive Scenario
Source: Reprinted with approval from Nieuwenhuijsen, de Almeida Correia et al. 2018

Webb, Wilson et al. (2019) looked at acceptance rates for new AV-based models for people living in a central business district. Their results showed only a small portion of people (16%) would not be willing to incorporate AVs into their travel planning and saw wealth, length of commute, and marital status as influencers of their willingness to adopt AV use (Webb, Wilson et al. 2019).

Knowing the perception of AVs and risk on the road is crucial in understanding the steps to broad adoption. A recent study on this topic showed that while AVs are perceived, on average, as a low-risk mode, this perception varied as passenger AVs were perceived as riskier than human-driven vehicles while pedestrians perceived AVs as less risky than their human-driven counterparts (Hulse, Xie et al. 2018). The work by Nazari, Noruzoliaee et al. (2018) studied the latent factors influencing AV adoption and found that safety had the largest marginal effect followed by a desire for green transportation and savviness for mobility on demand (Nazari, Noruzoliaee et al. 2018).

Another area that has been studied along with user acceptance is the user experience in the vehicle itself, and the interaction between the two. One study showed that user acceptance factors in user experience, such as perceived ease of use and attitude towards using the system and weakens desire to use AVs at all levels of automation based on levels of familiarity with AVs (Rödel, Stadler et al. 2014). This area of study related to AV adoption has been the focus of many other studies producing similar results (Charness, Yoon et al. 2018, Herrenkind, Nastjuk et al. 2019). In a literature review of 16 surveys conducted on this topic between 2012 and 2016, the authors observed a consensus that being young, male, and technology savvy were positive indicators of willingness to use AVs (Becker and Axhausen 2017). Another review of surveys on AV adoption analyzed 43 surveys conducted between 2012 and 2018 and again confirmed these findings (Gkartzonikas and Gkritza 2019).

Wang and Akar (2019) investigated the adoption of AVs by commuters and determined that solo car commuters are more likely to adopt AVs than commuters who use other modes while also noticing a decrease in interest in AVs among commuters (Wang and Akar 2019). In contrast, Penmetsa, Adanu et al. (2019) suggested that increased interaction and familiarity with AVs tend to generate a more positive attitude towards AVs so the interest in AV will only continue to grow as the technology advances (Penmetsa, Adanu et al. 2019). Continuing this line of research related to AV adoption has led to studies investigating the differences between countries. Kyriakidis, Happee et al. (2015) studied what different factors in country attributes influence acceptance of AVs and found that manual driving was preferred across the board. Respondents from nearly all the countries in the study were most concerned with hacking, legal concerns, and safety while only respondents from more highly developed countries expressed concern about data security (Kyriakidis, Happee et al. 2015).

A popular topic in AV research has been the willingness to accept shared AVs since it is often seen as the key feature that will reduce AV costs, achieve environmental targets, and reduce congestion (Sperling 2018). Being younger and living close to a central business district are factors influencing the acceptance of shared AVs (Webb, Wilson et al. 2019). Other studies have produced similar results, such as the analysis from Lavieri, Garikapati et al. (2017), which concurred that younger urban individuals are more likely to adopt AVs but also identified higher levels of technology savviness which positively correlate with shared AV use (Lavieri, Garikapati et al. 2017). In contrast, Clayton, Paddeu et al. (2020) conducted a survey to assess the willingness to adopt four types of fully automated transportation service (privately owned, synchronously shared, asynchronously shared, and public) and found significant uncertainty associated with the willingness to adopt either shared or private automated vehicles (Clayton, Paddeu et al. 2020). As no clear consensus has been reached on the likely acceptance of shared AV services, it is important to continue to study this topic and follow its development.

Sweet and Laidlaw (2019) dug deeper into the motivation that influences interest in AVs and its effect on vehicle ownership by conducting a survey using both revealed preference and stated preference questions to determine attitudinal motivators. They observed “significant differences between individuals’ motivations for vehicle ownership and AV adoption and use” (Sweet and Laidlaw 2019). This is an interesting finding as it illustrates that while it is easy to assume vehicle ownership and AV adoption would be related as it seems like a mode choice decision, private vehicles have other values and purposes than just transportation. For example, Sweet and Laidlaw (2019) noted that vehicle ownership was strongly associated with the attitudinal constructs of car freedom and control.

Much of the research related to AVs is application-based utilizing a variety of theoretical frameworks. Acheampong and Cugurullo (2019) conducted research to establish and test standardized conceptual models for AV adoption. They developed four conceptual frameworks to be used in AV adoption intentions which included identifying latent factors influencing shared AV use, the propensity to use AV in public transportation, and identifying factors impacting the likelihood of AV ownership (Acheampong and Cugurullo 2019). By establishing a set of suggested variables to use in researching these topics, this study allows future researchers to focus their specific analyses while only marginally optimizing the well-developed theoretical frameworks and models. Utilizing a rank-ordered probit model, Nair, Astroza et al. (2018) provide a foundation for establishing which variables to use in AV choice modeling. Their results suggest age, gender, household income, education, employment status, and household composition are significant variables in determining the preference of AV adoption and usage models (Nair, Astroza et al. 2018).

Some research has been conducted on AV adoption using latent class analysis which is the foundation for the research presented in this thesis. A 2019 survey of Australians segmented the potential AV market based on their propensity to use shared vs private AVs and determined 5 discrete classes to include non-adopters, ride-sharing, AV ambivalent, likely adopters, and first movers (Pettigrew, Dana et al. 2019). The significant indicators for being a non-adopter were being older, lower-income, higher enjoyment from driving, and a lower number of crashes in their driving history while first movers' key characteristics were higher education, driving enjoyment, and if they were more likely to have to transport elderly or disable passengers. As suggested in the paper, "it thus appears that the segments are substantively different rather than being polar opposites" (Pettigrew, Dana et al. 2019). Continuing this research on this topic, Kim, Circella et al. (2019)identified seven latent classes based on their mode-use propensity in a mature AV market, which is defined as being when the vehicle fleet is dominated by fully automated AVs. This study utilized a survey questionary similar in its content to the one used in my thesis, as there was a collaboration between the research teams at Georgia Tech and the University of California, Davis. Kim, Circella et al. (2019) provided a thorough analysis with many insights into the market segmentation of full AV users but does not explore the impacts on vehicle ownership, which is an important factor for transportation policymakers to understand as they work to create a more sustainable transportation system and will be explored in my thesis. The studies by Pettigrew, Dana et al. (2019) and Kim, Circella et al. (2019) show the interest in and multifaceted nature of latent class analysis of AV users and their future behaviors within the transportation research field as a rich vein for research. Related research into the latent attitudinal factors influencing AV adoption suggests other key factors include joy of driving, mode choice reasoning, trust, and technology savviness (Asgari and Jin 2019). Hardman, Lee et al. (2019) conducted research on current users of partially automated vehicles, which limited the study to Tesla owners/drivers and its partial automation solution given that Tesla vehicles were the only widely available partially automated vehicles at the time of the study. They identified four latent classes of users, which were labeled as very frequent users, frequent users, semi-frequent users, and infrequent users (Hardman, Lee et al. 2019). While this previous study started to identify real-world user behaviors it might not be as generalizable to the full automation AV market given the differences in the level of automation, expected use cases, pricing, and ownership models.

The studies referenced in this literature review provided the underpinnings for the research described in this thesis, from a theoretical foundation to the exploration and development of the analysis. Drawing inspiration from the previously conducted LCA-based research I seek to improve upon their work by including additional data associated with attitudes towards AVs as well as the expected impacts of the adoption of shared AVs on household vehicle ownership while refining the modeling approaches to create a model specifically designed for California. The data that underpins my research has been collected via a bespoke questionnaire designed with this research in mind, so it provides a rich set of data spanning variables that measure broad lifestyle factors, *e.g.*, attitudinal statements, to AV-specific topics, which provide a robust basis for this type of analysis. The area of study for my thesis is California, which is the epicenter of the AV deployment and will likely be able to provide unique insights into newer behavioral and attitudinal patterns that might later be observed elsewhere. Using data from California will allow statewide planners and policy makers to have data and analysis that is directly applicable to their locality without being concerned about the generalizable nature of prior studies. It is with these identified gaps in the research in mind that this thesis is conducted.

3. Data Collection

In this section, I present the methods used to collect the data for this thesis. This was one of my primary tasks as a researcher during my graduate program and was part of a larger research project that led to the 2018 data collection as part of the California Mobility Panel Study. For full details on this project please see the full report, *Panel Study of Emerging Transportation Technologies and Trends in California: Phase 2 Data Collection* (Circella, Matson et al. 2019). The remainder of this section repurposes the parts of that report that I wrote as it comprehensively summarizes the data collection process for this project.

For clarity and consistency, the nomenclature presented in Table 2 will be used throughout this section when referring to the different sampling methods used in the study.

Table 2. Sampling Method Nomenclature

Sample ID	Sampling Method	Details
A	**Mail Survey**	**30,000 paper questionnaires mailed to a random sample of California residents**
A.1	Returned via mail	Respondent opted to complete the provided paper survey and return it via regular mail
A.2	Completed via online survey system	Respondent opted to complete the survey via the online survey platform
B	**Online Opinion Panel**	**Approximately 2,000 respondents were collected via an online opinion panel provider**
B.1	Longitudinal	~1,000 respondents completed the 2018 survey and agreed to participate in future iterations of the study
B.2	Cross-sectional	~1,000 respondents completed only the 2018 survey
C	**Recontact 2015 Respondents**	**All respondents from the 2015 survey (N=1,975) were re-contacted to solicit participation in the 2018 wave of the data collection**

3.1 Survey Design

In the following subsection, I discuss the survey design process and administration. This was a collaborative effort by all members of the research team and in which I played a central role in both coordinating and actively working at the survey design and all stages of the data collection. This research was designed as a continuation of the *Panel Study of Emerging Technologies and Transportation Trends*, which already included a previous round of data collection in 2015. Therefore, the survey design utilized the content of the 2015 survey as the foundation, which was revised and expanded for the 2018 iteration of the data collection. It was important for the study's longitudinal design to ensure consistency between the two rounds of data collected, which allows for the analysis of trends over time. Therefore, the 2018 survey retains a similar structure and the core content of the 2015 survey, but some sections were simplified to reduce the response burden and new questions were included to collect information on new topics deemed important during this stage of the research (*e.g.,* use of shared ride-hailing services, and attitudes towards AVs). The research team utilized their access to colleagues at other research institutions, strategic business associates, staff at regional and state agencies, and other partner organizations to provide input on the survey design in the various areas of expertise.

Survey Section Overview

The survey collects information on personal attitudes and preferences, adoption of various communication technologies, personal lifestyles and work-styles (including telecommuting and mobile work, and adoption of e-shopping), cultural background, residential location and living arrangements, current travel behavior, use of cars vs. non-motorized transportation modes, adoption of alternative-fuel (*e.g.,* electric) vehicles, availability and use of new transportation modes and shared mobility services, aspirations for future travel and purchase of vehicles, and sociodemographic traits. As previously mentioned, I strived to maintain consistency, wherever possible, with the previous survey from 2015 to maintain the longitudinal component of the research while making additions and modifications in the 2018 surveys to address current or future transportation trends that were not well established in 2015.

Specific efforts were made to maintain question structure and neutral wording as much as possible to not add sources of bias related to the questions that were presented and phrased.

The following sub-sections of the report present the content of each section in the final version of the survey that was used for data collection in the project.

Cover Letter

Two versions of the cover letter were created for the two sampling methods, online and mail. The online versions (Samples B and C) provided a brief introduction to the research project, stated their responses would be confidential and thanked them for participating in the study. This was kept brief to not overburden the respondents with a long page to read.

The mail survey (Sample A) included all the information in the online version as well as additional information about the survey since it would be a respondents' only source. The cover letter provided the instructions on whom should take the survey, *i.e.,* the adult (18+) with the closest birthday to the date they received the survey. This additional randomization was intended to minimize a potential sampling bias of having responses from only the main person in the household that retrieves/opens the mails or the head of the household. The cover letter also provided instruction on how to complete the survey online if that was their preferred method. Providing two options to complete the survey was intended to encourage the highest response rate possible. To further this objective, an incentive was offered to participate in a random drawing for one of many gift cards from a major online shopping retailer. The research team's contact information was provided in the form of an email address and a toll-free 1-800 number to facilitate answering any questions from the respondents.

It is worth noting that on the outside of the survey, there were instructions in English and Spanish to inform the reader that if they wanted to complete the survey in Spanish, a Spanish version of the questionnaire was available via the online survey platform, or a Spanish translation of the survey would be provided upon request. While not explicitly stated as an option in the instructions, four requests for the

Spanish surveys were received through the toll-free number, which was set up with a virtual switch board to direct calls to an English- or Spanish-speaking member of the research team.

Section A: Your Opinions on Various Topics

This section began with a brief introduction to the topic and a reminder that there is no "right" or "wrong" way to answer the survey if it was their honest opinion. Section A focuses on attitudes and preferences by gauging their agreement with 29 attitudinal statements on a 5-level Likert-type scale. The core constructs these statements measured were:

- Active Lifestyle/pro-exercise
- Environment
    - Environmental concern
    - Environmental concern and travel behavior
- General Value/Beliefs
    - Trust
    - Variety-seeking
- Role of government on regulating car use
- Perceptions on land-use
- General life satisfaction
- Travel Modes Perceptions
    - Bike
    - Car
    - Transit
- Sensitivity to autonomy in driving
- Materialism – sharing vs. private ownership
- Technology trends
    - Early adopter
    - Technical savviness
- Time use/multi-tasking
- Perception on commute time

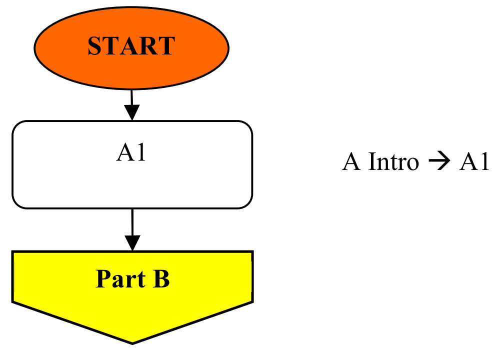


Figure 4. Logic Flow and Branching for Section A

Since this section was structured as one question with 29 statements, it had a high potential for respondents to flatline, *i.e.,* mark all of one response to the whole list. To prevent this and to test for full engagement with the survey, trap questions were included in this section. These questions asked the respondent to enter a specific response and if they failed this task, it would indicate that their survey may need to be dropped for quality assurance reasons. In the paper version (Sample A.1) of the survey, one trap question was used while the online versions (Sample A.2, B, C) used two trap questions. An additional trap question was used in the online version to prevent inattentive respondents "flatlining" the section and automated "bots" from attempting to complete surveys distributed through online survey panel vendors. This section consisted of one question so there was no logic/branching in this section (see Figure 4).

Section B: Your Use of Technology

Section B collects information regarding the respondent's familiarity with internet-connected devices and services.

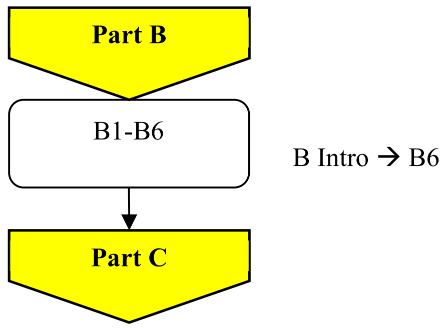


Figure 5. Logic Flow and Branching for Section B

Questions in the section asked about ownership of devices and the frequency apps/services are used. Online shopping is a prominent activity for people with internet access, which impacts their travel behavior as it correlates to the number of trips made to stores and their use of the freight network. The survey collected information on recent purchases, preferred shipping timeframe, and how this has impacted their item search/purchase patterns. This section has no logic/branching through the six questions in the section (see Figure 5).

Section C: Key Aspects of Your Lifestyle

Section C collected information regarding key aspects of their lifestyle such as current housing arrangements, major life events, residential location choice, and car ownership status.

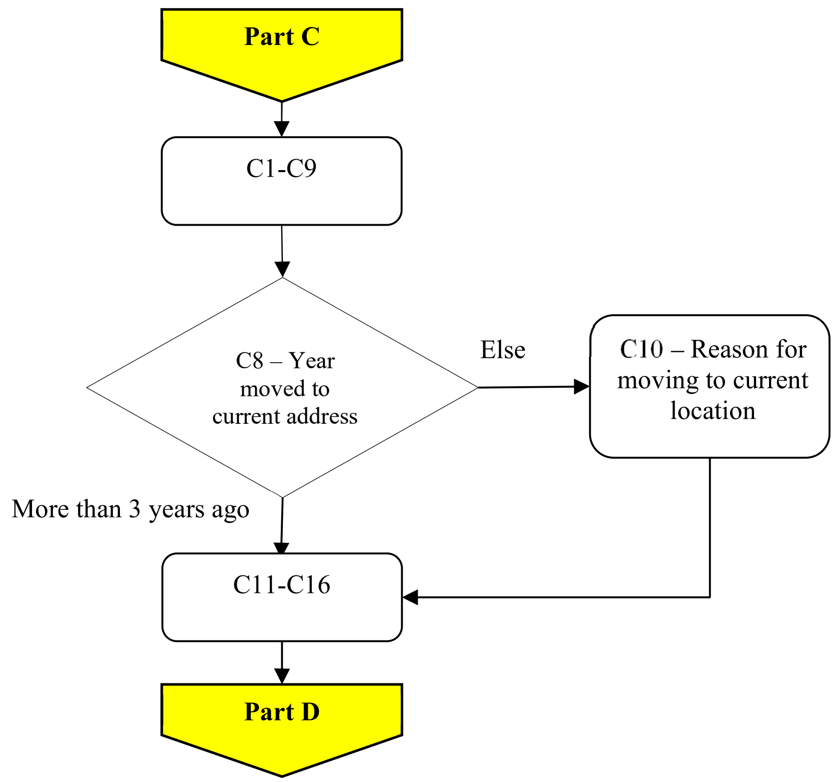


Figure 6. Logic Flow and Branching for Section C

This section has skip logic following the ninth question (see Figure 6). Using the information provided in the eighth question (the year they moved to their current address) and if a respondent had lived there for more than three years, they skipped the tenth question which asked for the reason they moved to their current location.

Section D: Employment and Work/Study Activities

Section D collected information on employment status, student status, and the corresponding schedule for those activities. After providing their work and student status, skip logic was used to not burden respondents with unrelated questions (see Figure 7). If a respondent was not employed or a student, they were instructed to skip to question 6 in Section E. Section E began with questions related to commute travel, so it would not pertain to these respondents. If a respondent was only a student, they skipped question three which asked how many hours they work in a week. Finally, the remaining people that work and either are or are not a student progressed through the section without skipping any questions.

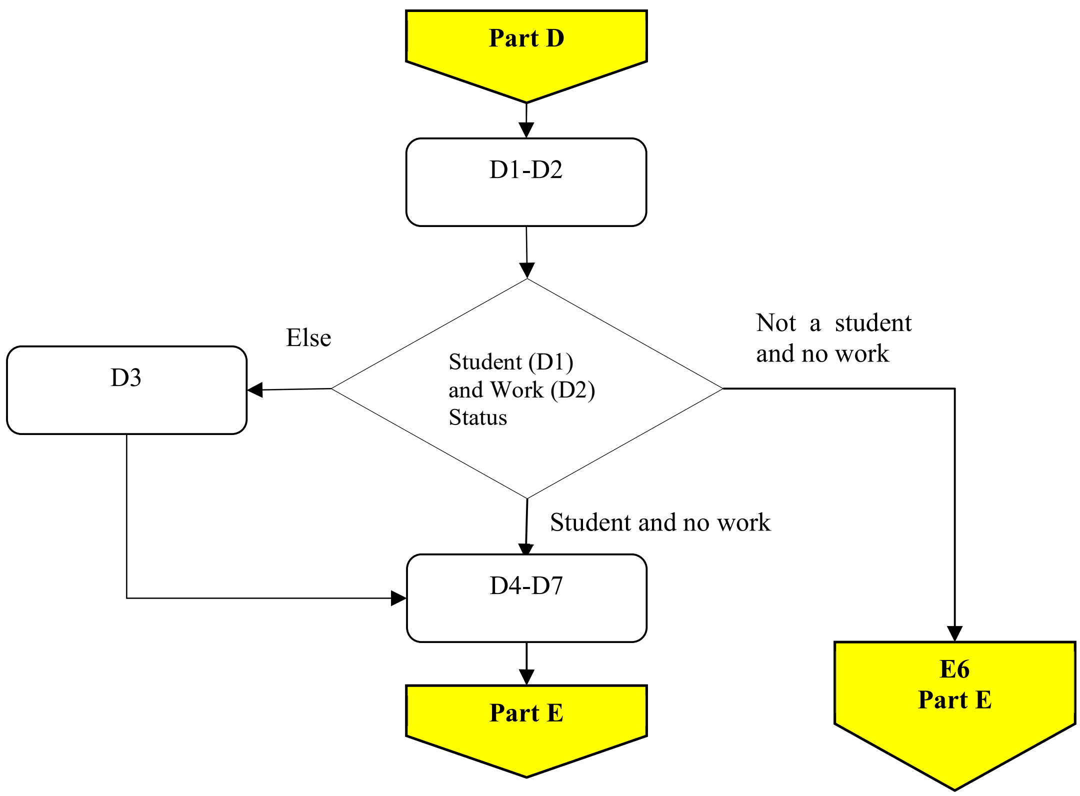


Figure 7. Logic Flow and Branching for Section D

__Section E: Your Current Travel Choices__

Section E collected information on current travel behavior. At the beginning of this section, there was a brief introduction that defined terms that were used in the remainder of the survey. This was required since terms such as “trip” and “car” can be interpreted differently and defining them it would minimize any ambiguities. The content of this section is related to current commute patterns, most recent leisure/social/shopping trip, average monthly transportation costs, how much they like current modes of travel, and long-distance travel. In addition to the overall travel choices, there were also questions relating to multi-tasking during these trips, the presence of any physical or other conditions preventing/limiting travel, and the influence of the internet on daily travel. In its totality, this section provided a robust understanding of the respondent’s current travel choices.

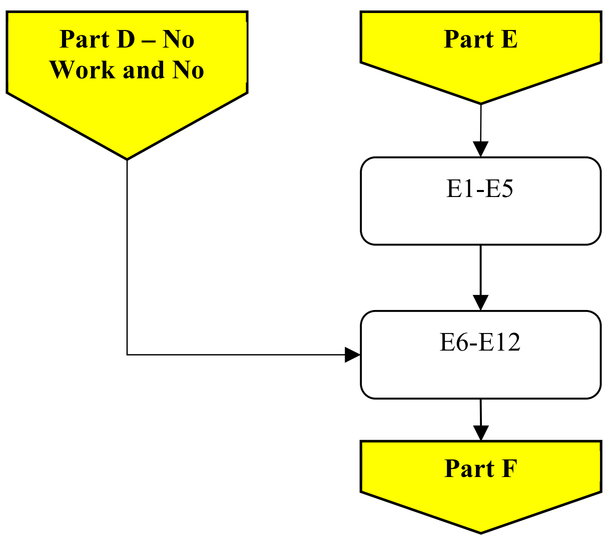


Figure 8. Logic Flow and Branching for Section E

There is no skip logic/branching starting in the section; but it does have respondents who neither worked nor were students skip to question 6, which is the first non-work/study related travel question (see Figure 8).

Section F: Emerging Transportation Services

Section F collects information on the emerging app-enabled ride-hailing services, *e.g.,* Lyft, Uber, and Via. Questions in this section include frequency of use of the services, a detailed examination of their last ride-hailing trip, impact on other modes, and an assessment of current dependency on the services. Following these questions on established mobility services, the survey probes the respondents on their willingness to use the emerging shared ride-hailing services, *e.g.,* Lyft Line or UberPOOL, and measures their interest in theoretical mobility-as-a-service implementations.

This section was expanded compared to the 2015 survey given the rise in popularity of these services and their potential to affect other components of travel behavior such as travel cost, convenience, and security (National Academies of Sciences and Medicine 2016). The 2015 survey already included a sizable section on shared mobility, collecting detailed information on the awareness, adoption, and frequency of use of the most common ride-hailing services. In the 2018 survey, we also collected information for additional types of shared mobility services that have been introduced during recent years, specifically shared ride-hailing services.

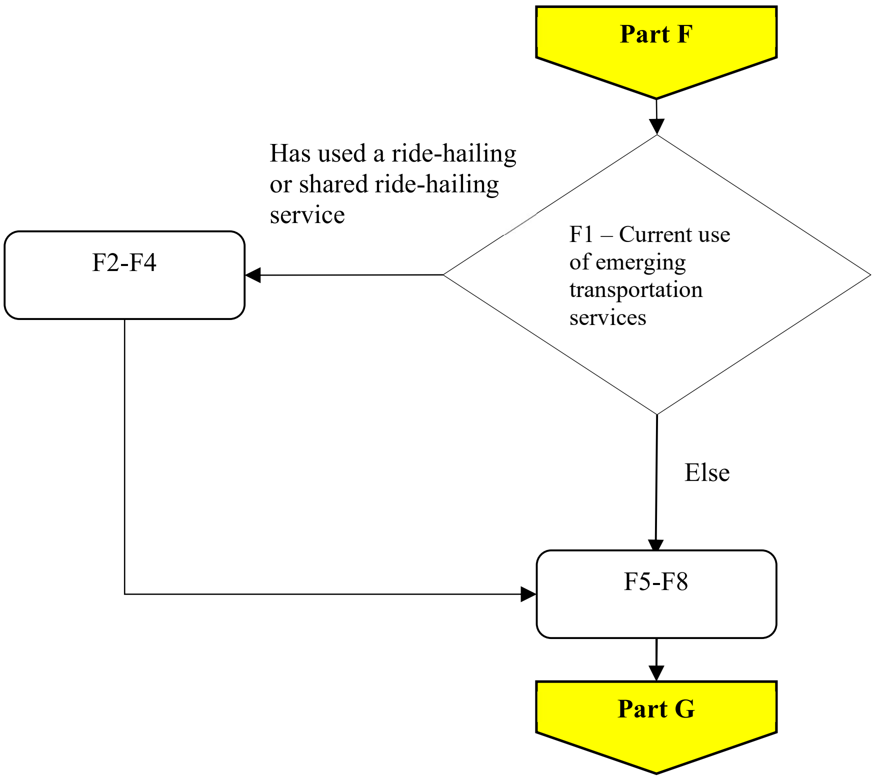


Figure 9. Logic Flow and Branching for Section F

To measure the likelihood a person would be willing to share a ride via a ride-hailing service if a discount was provided, a unique question design was used. The second question in the section consists of a table of 13 sub-questions with the last sub-question asking how long they would be willing to wait for a shared ride over a single occupant ride for a discount over the single occupant cost. To test people's price sensitivity, four versions of this question were created with discounts of 10%, 20%, 30%, and 40%. This was easily implemented in the online versions (Samples A.2, B, C) with one of the four variants being randomly presented to the respondent. The mail version (Sample A.1) of the survey required four versions of the survey to be created and then randomly assigned one of the four versions.

There is one point of logic/branching in this section, and it is based on the first question which asked about the frequency they have used a list of emerging transportation services (see Figure 9). If they have ever used a ride-hailing or shared ride-hailing service they would proceed to the second question, otherwise, they would skip to question five. The paper version had a simplified version of this by instructing the reader to check a box if they have never used a ride hailing service (shared or single) and then skip to question 5. This approach was used as it was a simpler approach than having to review the previous question for a list of specific responses thereby reducing the cognitive burden placed on the respondents.

Section G: Future Mobility

Section G collected information on how new technologies and service may impact their current vehicle ownership and, as an addition to the 2015 survey question, a new block of questions related to AVs was added. These new questions focused on collecting information about a respondent's perceptions and propensity to adopt AVs in scenarios of shared-ownership/shared-use or personal ownership. Before the new block of AV-related questions which included a brief introduction to the technology, the respondents were asked to provide their current level of awareness of AVs. This was done to measure the respondent's unprimed opinion on the subject. In both the online and paper version, a page break was added after this question so the introduction AVs would not bias the previous question. The introduction to AVs was designed to illustrate the wide range of possibility AV may provide and was presented it in two different ways (pictures and text) to stimulate as much thought and engagement about this new technology. This section of the survey collected the primary information that is the main interest for the analysis in my thesis.

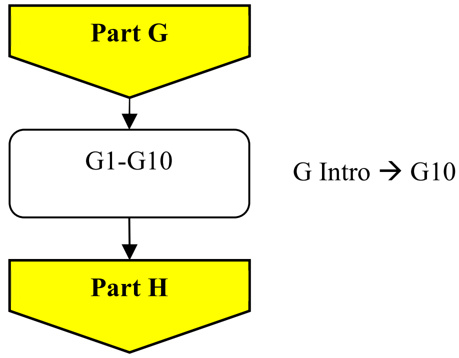


Figure 10. Logic Flow and Branching for Section G

There is no logic/branching in the section (see Figure 10). As this is the penultimate section in the survey and there is potential for fatigue from the respondents, a third trap question was included in question five. The results of this trap question will be considered in conjunction with the prior trap questions and other data validation tests to determine if a respondent should be dropped from the dataset.

Section H: Some Background about Yourself

Section H collects basic sociodemographic questions from the respondents. Collecting data on their age, ethnicity, gender, education, and income will allow the researchers to generalize the findings from the sample to the population of California.

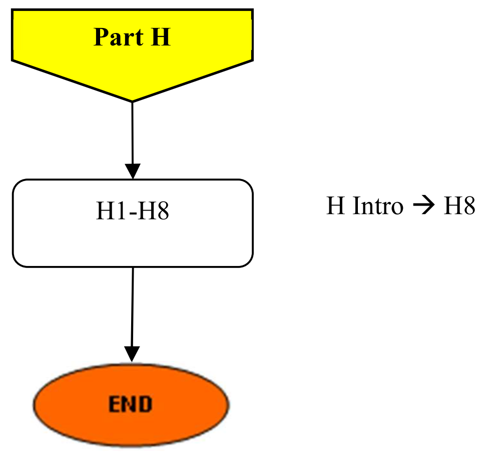


Figure 11. Logic Flow and Branching for Section H

There is no logic/branching in this section (see Figure 11). Following the end of the survey, there is a heartfelt thank you to the respondents for their time and effort and up to two additional questions, depending on the distribution method. The mail survey (Sample A.1) and its corresponding online survey (Sample A.2) ends first with a question asking if the researcher team can contact them for one of the following reasons: 1) entry in the drawing for the incentive, 2) availability for any follow-up questions based on their responses to the survey, and 3) willingness to participate in future iterations of the survey, *i.e.,* the longitudinal component of this ongoing research effort. As part of that question, the respondents are asked to provide their preferred contact information. This is followed by an open response question for any additional comments about the survey itself or topics raised in the survey. The online-only versions (Sample B, C) of the survey only have the second of those two questions. The panel vendors directly handle the distribution of the incentives for their respondents and normally do not allow personally identifiable information, such as contact information, to be provided unless in special cases like our efforts to build on our longitudinal panel (Sample B.1), which will be discussed in detail in the Data Collection Methodology section.

Survey Length

After completing the 2015 survey many respondents noted that the survey was too long. Reducing the time needed for completion was an important issue that needed to be addressed. As the length of the survey increases the potential for fatigue increases thus calling into question the validity of responses from later parts of the survey. This was addressed by consolidating the number of sections in the survey from eleven in 2015 to eight in 2018. While the number of sections is cosmetic, the reshaping of the survey flow did focus the content to be reduced to stay within the topics of each section. This led to a reduction of 31% in the number of questions from the 2015 survey, which consisted of 159 questions, and the 2018 survey which consisted of 109 questions. This resulted in a survey that was designed to be completed in 30 minutes.

Distribution Method Considerations

Given the unique natures of the two distribution methods (online opinion panel and mailed questionnaire), I made a concerted effort to design the survey in a manner to have consistency between them, thereby not introducing an unintended source of error or bias in the data collection. The paper survey drove the design as it would be the most limiting of the versions due to the inability to force responses, no response validation on open-ended questions, and the additional effort required by respondents to correctly follow skip logic/branching. Efforts were primarily directed at reducing and simplifying the skip logic/branching. It was used three times and spread out across three sections to not over-burden the survey respondents. Also, the criteria to skip certain questions were presented as simply as possible with the key criteria being in bold typeface to stand out on the page in the paper version of the survey.

Testing and Survey Design Quality Assurance

When the content and formatting of the survey were nearly finalized, I tested the survey by conducting a multistage pretest. In the first round of pretesting, team members took the survey to conduct a final review of copyediting as well as to experience taking the survey in the same manner as a respondent. This provided feedback that was useful in identifying small items that needed correction. In the next stage of the pretest, I utilized a convenience sample of colleagues and other peers to run a pretest of people that had not seen the survey before. They were able to complete the survey either via the online platform or by printing a copy of the mail survey. A total of 18 pretest surveys were completed and based on their feedback, I was able to modify questions for clarity, correct small typos, and confirm the estimated time to complete the survey at 30 minutes.

3.2 Data Collection Process

Figure 12 summarizes the sampling strategy for the second wave of data collection in this panel study. The target sample size at the end of this phase of the data collection was 3,300. However, we exceeded this with the final sample totaling 3,834 cases.

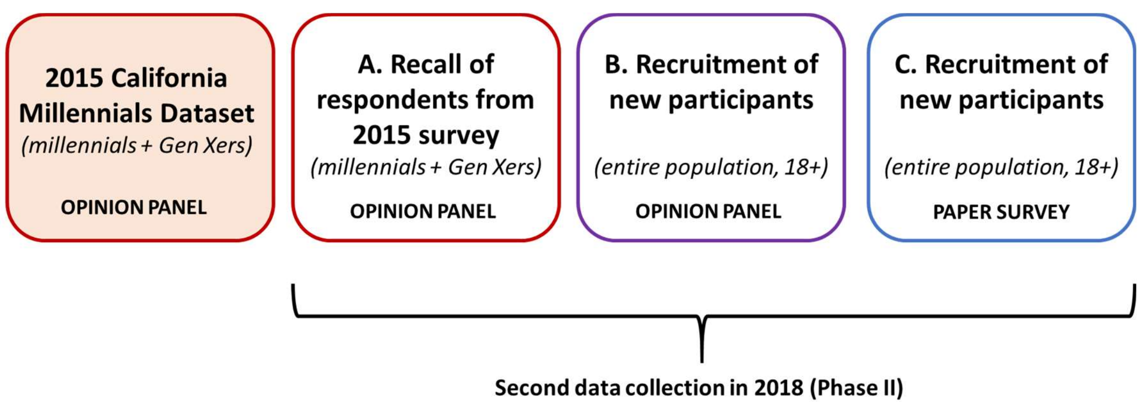


Figure 12. Sampling Strategy for Second Round of Data Collection in Phase II of the Research

2018 California Dataset

A combination of sampling strategies was used for the second wave of data collection to create a sample that would minimize the sampling issues associated with each recruiting and sampling channel.

For these studies, California was divided into six major regions (see Figure 13):

- San Francisco Bay Area, identified by the boundaries of the Metropolitan Transportation Commission (MTC)
- Los Angeles/Southern California, identified by the boundaries of the Southern California Council of Governments (SCAG)
- Sacramento region, identified by the boundaries of the Sacramento Area Council of Governments (SACOG)
- San Diego, identified by the boundaries of the San Diego Association of Governments (SANDAG)
- Central Valley, corresponding to the eight counties in the central San Joaquin Valley
- Northern California and Others, which includes the rest of State not included in previous regions.

Figure 13. Regions of California Included in This Study

Sample A – Mail Survey

A paper version of the survey was mailed to a random sample of 30,000 addresses in the state. This approach allowed us to reach most major segments of the population, including the elderly or people not familiar with technology, who are less likely to be part of online opinion panels.

The mailing addresses were provided by a vendor that maintains comprehensive lists of postal addresses in California. They created the mailing list based on the sampling rates presented in Table 3. Given the population disparities between the different regions, the researchers adjusted the sampling rates to obtain sizable numbers of respondents in all six regions. Using this stratified random sampling approach helped reduce the very large number of respondents from the Los Angeles region that would be recruited if true random sampling was used. Before mailing the surveys, all addresses were validated with the United States Post Office to make sure they were bona fide addresses and that the person was still a resident at that address.

Table 3. Sampling Rate for Each Region for Mail Out/Mail Back Survey

Population	% of California Population	# Invitations with Constant Sampling Rate		Final Sampling Rate	Final # Invitations	% Over/Under Sampling
Central Valley	4219,854	10.67%	3,202	0.091%	3,820	119.27%
MTC	7756,158	19.62%	5,885	0.090%	6,982	118.61%
NorCal and Others	2752,763	6.96%	2,089	0.140%	3,854	184.50%
SACOG	2498,563	6.32%	1,896	0.150%	3,749	197.68%
SANDAG	3337,685	8.44%	2,533	0.120%	4,006	158.15%
SCAG	18,971,630	47.98%	14,395	0.040%	7,589	52.72%
Total	39,536,653	100.00%	30,000	0.076%	30,000	100.00%

Sample B – Online Opinion Panel

The panel was also refreshed by adding a group of new participants in this wave of data collection that were collected via an online opinion panel. Similarly, for future waves of data collection in this panel study we will continue to refresh the panel at each round of data collection with a similar approach to replace the naturally accruing dropping out of respondents from the panel. Sample B collected 2,000 participants and utilized a similar methodology to what was used for the 2015 California Millennial

Dataset data collection. This method also facilitated expanding the age cohorts in the study with younger respondents between 18 and 21, *i.e.,* members of Gen Z, and baby boomers who were not included in the data collection in 2015.

The sampling conducted for this sub-sample utilized a quota methodology. This was used as online panels tend to be a skewed sampling frame (towards younger, more often female, unemployed respondents), so the quota system would allow for this issue to be corrected. The quotas were established by using the most current 5-year estimates from the American Community Survey (see

Table 4-5). This also allowed us to control for varying travel behavior associated with land use/neighborhood type in the study, *i.e.,* quotas were established for each pairing of region and neighborhood type for each key age group.

Table 4. Sampling Quotas by Region

Region	Percent of Sample	Target Sample Size
Central Valley	12%	120
MTC	27%	270
SACOG	10%	100
SANDAG	12%	120
SCAG	29%	290
NorCal & others	10%	100
Total		1,000

Table 5. Sampling Quotas by Neighborhood Type (Nested Within Region)

Neighborhood type	Percent of Sample	Target Sample Size
Rural	17%	170
Suburban	44%	440
Urban	39%	390
Total		1,000

Table 6. Sampling Quotas by Age (Nested Within Neighborhood Type)

Age	Percent of Sample	Target Sample Size
18-38	46%	460
39-53	35%	350
54+	19%	190
Total		1,000

I also controlled for other sociodemographic factors beyond region and neighborhood type to minimize the non-representativeness of the sample and mimic the characteristics of the population of California. The quotas presented in Table 7 were soft quota targets which allowed for +/- 5% deviations on the targets and, if the quota proved to be hard to attain after a diligent effort, they could be relaxed to +/- 10% of the target.

Table 7. Targets by Age Group for Eight Key Sociodemographic Factors

	Ages 18-37		Ages 38-53		Ages 54+	
	% of Sample	Sample Size	% of Sample	Sample Size	% of Sample	Sample Size
Sample Size/Age Group		460		350		190
Gender						
Male	50.00%	230	50.00%	175	50.00%	95
Female	50.00%	230	50.00%	175	50.00%	95
Children in Household						
Yes	35.00%	161	35.00%	123	35.00%	67
No	65.00%	299	65.00%	228	65.00%	124
Household Income						
Less than $24,999	18.40%	85	18.40%	64	18.40%	35
$25,000 to $49,999	19.40%	89	19.40%	68	19.40%	37
$50,000 to $74,999	16.30%	75	16.30%	57	16.30%	31
$75,000 to $99,999	12.20%	56	12.20%	43	12.20%	23
$100,000 to $149,999	15.70%	72	15.70%	55	15.70%	30
$150,000 or more	18.00%	83	18.00%	63	18.00%	34
Age						
18-27	50.20%	231				
28-37	49.80%	229				
38-46			55.90%	196		
47-53			44.10%	154		
54+					100.00%	190
Race						
White	53.40%	246	62.70%	219	68.30%	130
African American	5.80%	27	6.20%	22	5.70%	11
Asian	13.80%	63	17.00%	60	15.30%	29
Other	27.10%	125	14.10%	49	10.80%	21
Hispanic						
Yes	41.60%	191	40.70%	142	23.50%	45
No	58.40%	269	59.30%	208	76.50%	145
Work Status						
Employed	68.00%	313	76.20%	267	39.70%	75
Not Employed	32.00%	147	23.80%	83	60.30%	115
School Status						
Enrolled	30.30%		2.00%	7	1.00%	2
Not Enrolled	69.70%	321	98.00%	343	99.00%	188

Sample C – Recontact of 2015 Respondents

Finally, I recalled the respondents that completed the 2015 survey using the same online opinion panel vendor. Even if we initially expected to be able to retain close to 50% of the respondents from the 2015 data collection (*e.g.*, 38pprox.. 1,000 respondents), during the data collection process it became clear that this goal would not be achieved for three key reasons. First, when the respondents took the first survey in 2015 there was no mention of future surveys so there was no expectation to be contacted again. Second, there was no contact in the intervening 3 years to keep them engaged and invested with the success of the project which is needed to garner high levels of participation in longitudinal surveys. Third, there was no guarantee the full panel would remain active within the panel vendor's research efforts over the 3 years – endangering the only means to contact these respondents available to the researchers. These factors led to a low number of responses with 246 completed surveys among the 1,975 respondents from the 2015 survey. The lessons learned from this data collection also helped identify strategies to improve the retention rate of respondents and contribute to the "panel building" component of the study, which will be applied for more efficacy in the longitudinal data collection in future related research.

3.3 Survey Administration

The 30,000 paper surveys were printed by the UC Davis in-house printing services, while the envelope stuffing and mailing were executed by the UC Davis mailing office. To encourage a higher response rate a postcard was sent to all the addresses that had not already returned a survey. The creation and processing of the postcards were handled by the same vendors. The online surveys were created on the Qualtrics online survey platform. The distribution of the online survey was conducted by two different online panel vendors for Sample B and Sample C.

Table 8 presents the response rates that were achieved by the time of writing this report. For Sample A, we anticipated a response rate of 6-8% given the length and content of the survey and considering the continual general decline of response rates expected for unsolicited survey requests. The results for recruitment for this method were near the lower bounds of our expectations. At the time of writing this report, information on the number of email invitations sent by the opinion panel vendor for Sample B was not available. Therefore, the response rate for the online panel could not be calculated. As is typical with panel vendors, the 2015 California Millennial Dataset was created from a combination of an internal list of respondents as well as external lists from other sources. So, for Sample C it is worth noting that of the full 1975 respondents in the 2015 dataset only 315 were still active with the vendor in 2018. This means that the response rate for this sample could be as high as 77% based on respondents that were certainly still active within the panel (N=315). However, a more accurate estimate of the response rate for this sample is approximately 12%, if we consider that the opinion panel vendor was able to reach the remaining respondents from the 2015 survey through partner online panels.

It is worth noting that the research team is using this experience to shape plans for improving the panel for future iterations of this study. We are directly collecting contact information for the majority of respondents in the 2018 sample, so we will be able to recontact respondents to provide updates on the research and have already primed them with the idea of being re-contacted later for future phases of the data collection. This will have the benefit of reducing future costs of the data collection, reduce attrition, and help stimulate a higher response rate in future data collection efforts.

Table 8. Response Rate by Sub-Sample (Raw Rata Without Any Filtering)

Type	Number of Invitations	Number received	Response Rate
A. Mail Survey	30,000	1,992	6.64%
A.1 Returned via mail		*1,620*	*5.40%*
A.2 Completed via online survey system		*372*	*1.24%*
B.1 Online Opinion Panel – Longitudinal	N/A**	830	-
B.2 Online Opinion Panel – Cross-Sectional	N/A**	1,003	-
C.2015 Panel Recontact	1939*	246	12.69%
Total		4,071	

* Maximum number of invitations possible
** Number of invitations sent has not been provided by panel vendor

Table 9. Data composition (raw data without any filtering)

	Online	Paper	Total
A. Mail Survey			
Count	**372**	**1,620**	**1,992**
Percent by row	18.67%	81.33%	100.00%
Percent by total	9.14%	39.79%	48.93%
B.1 Online Opinion Panel – Longitudinal			
Count	**830**	**N/A**	**830**
Percent by total	20.39%	N/A	20.39%
B.2 Online Opinion Panel – Cross-Sectional			
Count	**1,003**	**N/A**	**1,003**
Percent by total	24.64%	N/A	24.64%
C. 2015 Panel Recontact			
Count	**246**	**N/A**	**246**
Percent by total	6.04%	N/A	6.04%
Total	**2,451**	**1,620**	**4,071**
Percent	60.21%	39.79%	100.00%

Time to Complete the Survey

As previously discussed, the survey was designed to be completed in 30 minutes to reduce response fatigue, encourage higher survey engagement and response rates thus providing higher quality data. This was proven to be an accurate estimated time to complete the survey as seen with the results from the online surveys (see Table 10). It is worth noting that Sample A.2 did exceed the desired time to complete with a median time of 37.8 minutes. The time difference can likely be attributed to different levels of experience with online surveys across samples, but this will warrant further examination during the analysis stage of the research to see if this hypothesis holds.

Table 10. Completion Times for Online Versions of Survey

Online Survey Version	Median Completion Time (Minutes)
A.2 Mail Survey completed via online survey systems (N=372)	37.8
B.1 Online Opinion Panel – Longitudinal (N=830)	29.9
B.2 Online Opinion Panel – Cross-sectional (N=1,003)	31.4
C. Recontact 2015 Respondents (N=246)	27.8

Response over Time by Distribution Method

2015 Panel Recall Sample

The recall of the 2015 panelists began on July 23, 2018 and lasted until August 6, 2018 (see Figure 14). Half of the sample was achieved by the second day, and it experienced a slowdown after the 7th day as the sampling worked to gain the last 11.78% of the sample. This was a fast and productive sampling as the sampling frame was limited in size and thus it quickly achieved its targets.

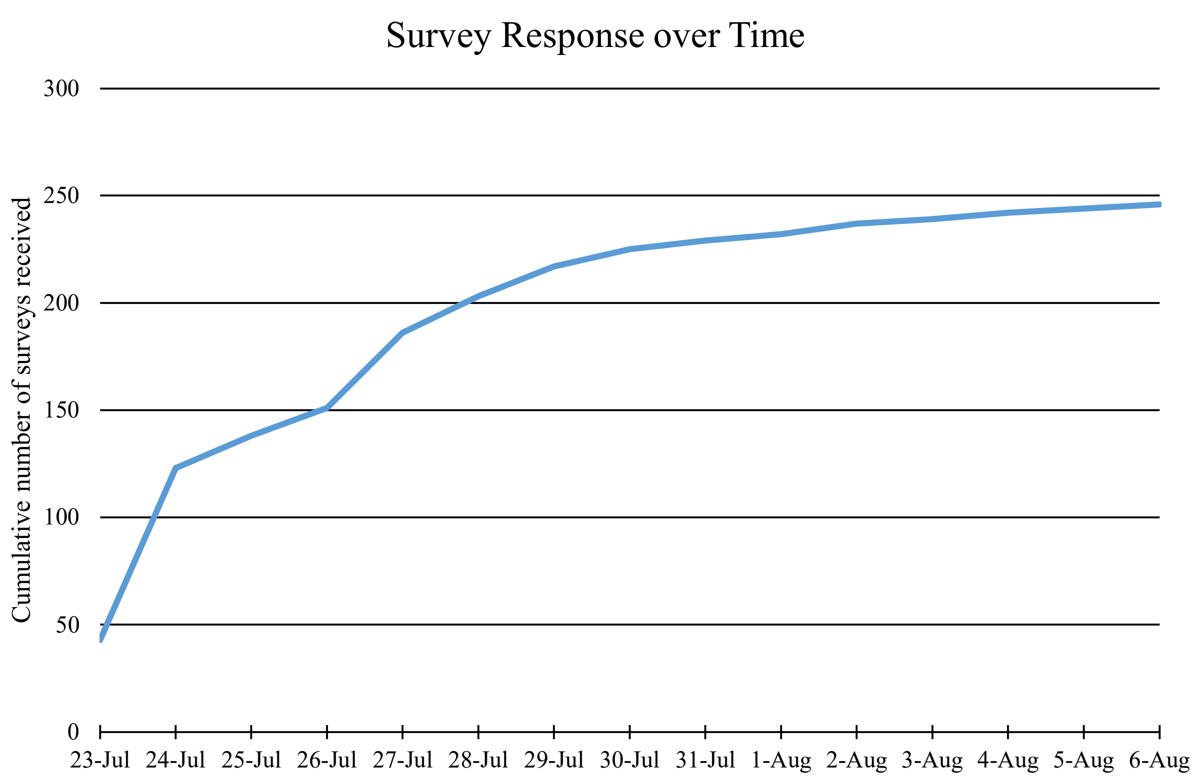


Figure 14. Survey Response Over Time for 2015 Panel Recall

Online Panel Cross-sectional Sample

The online panel cross-sectional sampling being on August 15, 2018 and was completed on September 26, 2018 (see Figure 15). 50% of the sample was achieved by the 9th day and experienced a long tail of modest gains following the second big push by the panel vendor to meet the quota targets, which started on September 4th. This sampling was rather complicated with the multiple nested quotas, so it took longer to achieve the final sample. This was driven by the challenges to reach smaller, specific targets, such as urban residents in the SANDAG region or Hispanics throughout the state.

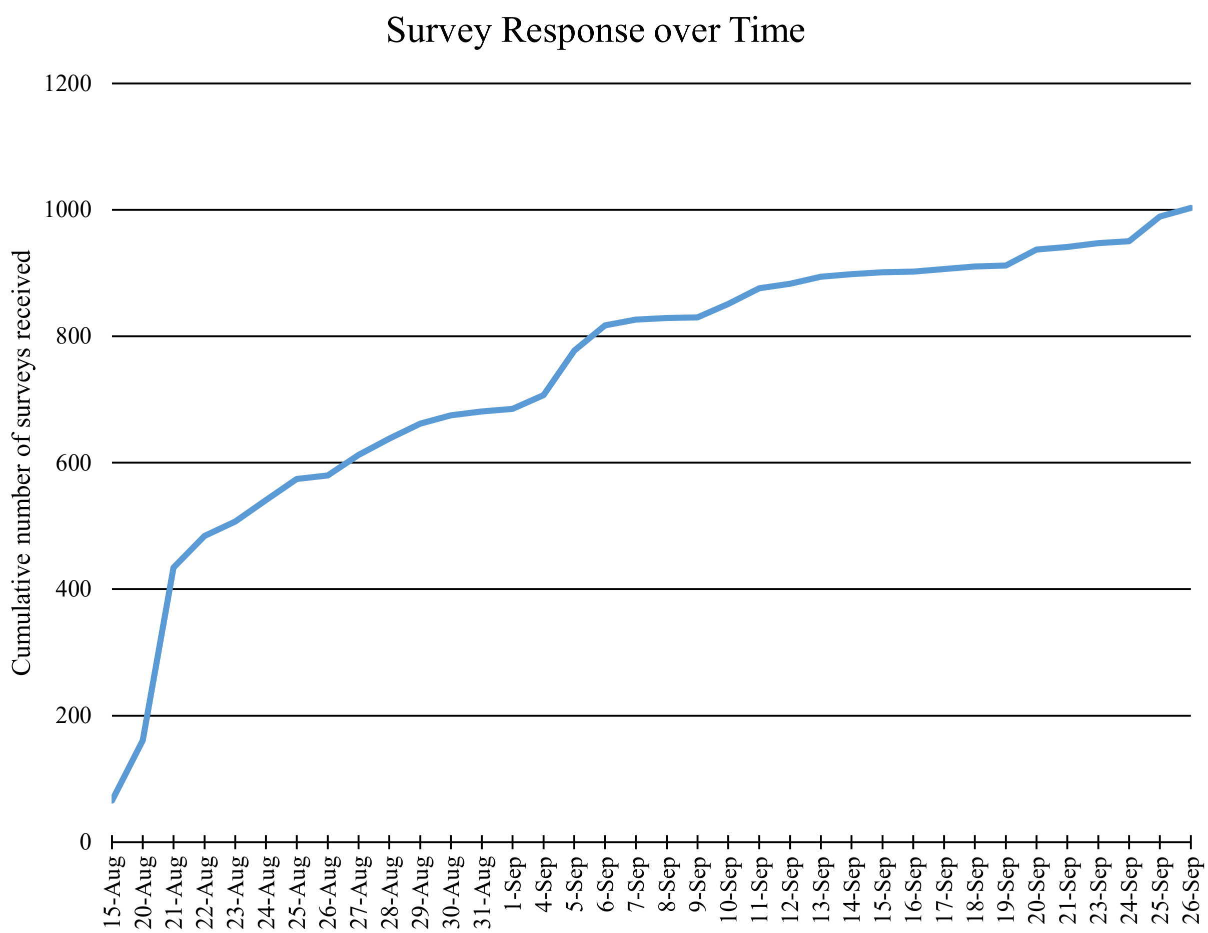


Figure 15. Survey Response Over Time for Online Panel Cross-Sectional Sample

Online Panel Longitudinal Sample

The online panel longitudinal sample began on August 6, 2018 and was still ongoing as of the writing of this report as it is short 170 responses for its 1000 target (see Figure 16). This sampling process has proved to be a level more challenging than the previous sample as it has similar quotas and the added challenge of finding participants that want to be part of a multiyear panel. The 50% mark was achieved during the 29th day which was quickly followed by a large jump in the following days as the panel vendor added a new external contact list provider. This sample's quotas have been relaxed in the following areas to aid in the timely completion of the sampling: Ages 18-27, Males, Race – Other, Income $150,000+, and Northern California and Other residents. Besides the Age 18-27 quota, these were expected trouble areas given the nature of online opinion panels. By relaxing the quotas, the sample will be able to be completed but extra attention is placed on the weighting of the final sample to bring it back in line with the actual sociodemographic composition of the state.

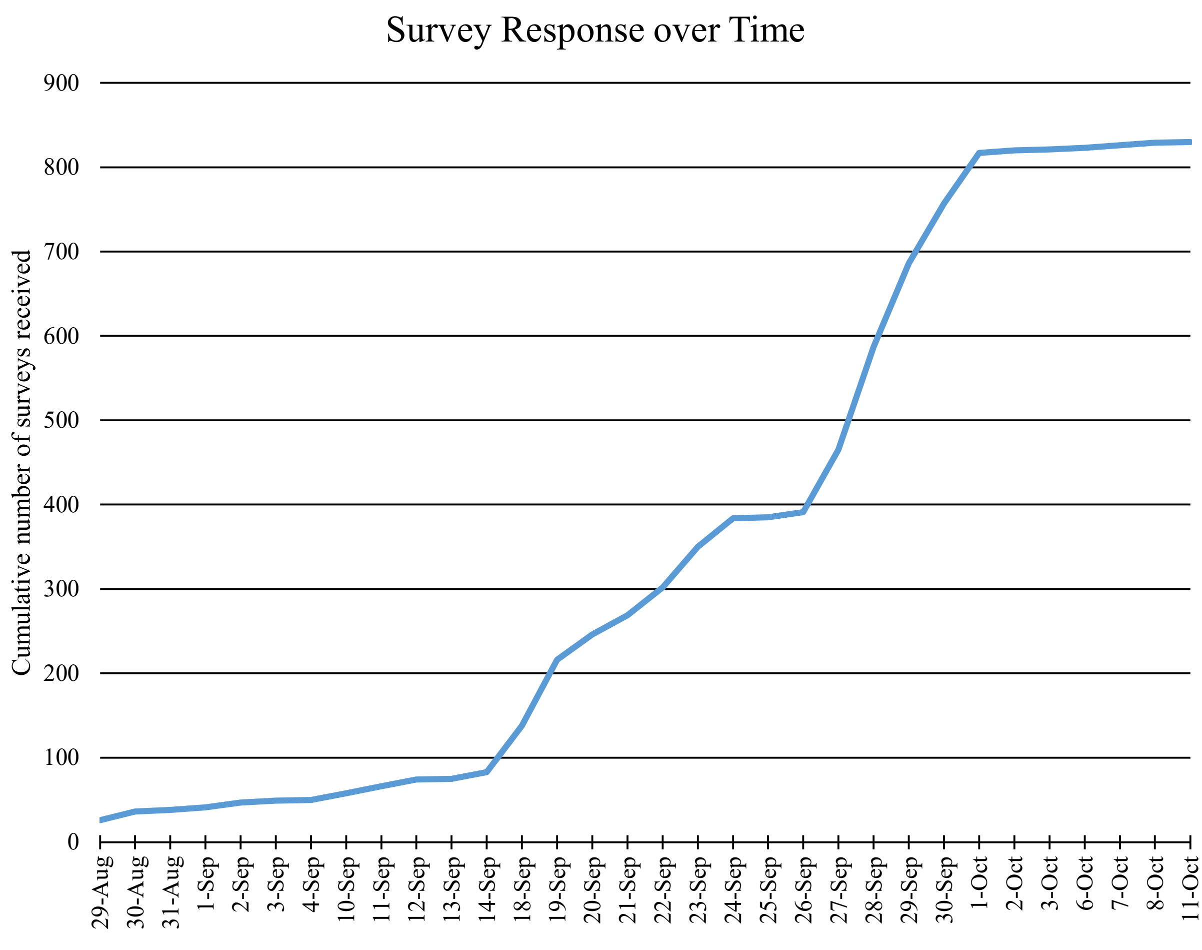


Figure 16. Survey Response Over Time for Online Panel Longitudinal Sample

Mailed Survey – Returned via mail

The paper survey began to be mailed out at the end of June as soon as the first surveys were printed and assembled. The mailing continued for the following month as 30,000 surveys, outgoing envelopes, and return envelopes were printed and assembled by the vendors. It was decided to mail them as they were ready as it would allow for the data collection to begin and not incur an additional cost by storing them for up to a month to mail as one large batch. The first response was received on July 9, 2018 (see Figure 17). As there were large quantities of mail being delivered to the researcher's office, the university's mail room would make one or two large deliveries per week, so the delivery dates are less accurate than the online results which are more disaggregated. 50% of the sample was achieved between 26^{th} and 39^{th} day. As previously mentioned, it is not possible to determine a more precise date given the intermittent delivery of the returned surveys. Even though it was asked for all the surveys to be returned by the end of July with the caveat that they would still be accepted afterward, there has been a long, slow trickle of surveys being returned until October 11, 2018. With the aggregated mail delivery, it is hard to determine if there was a bump in response due to the postcard being sent. There was a noticeable increase in the number of calls received at the 800-number following the delivery of the postcard so I am hopeful that it translated to an increase in the response rate but could not be proven given the previously mentioned constraints imposed by the mail delivery system.

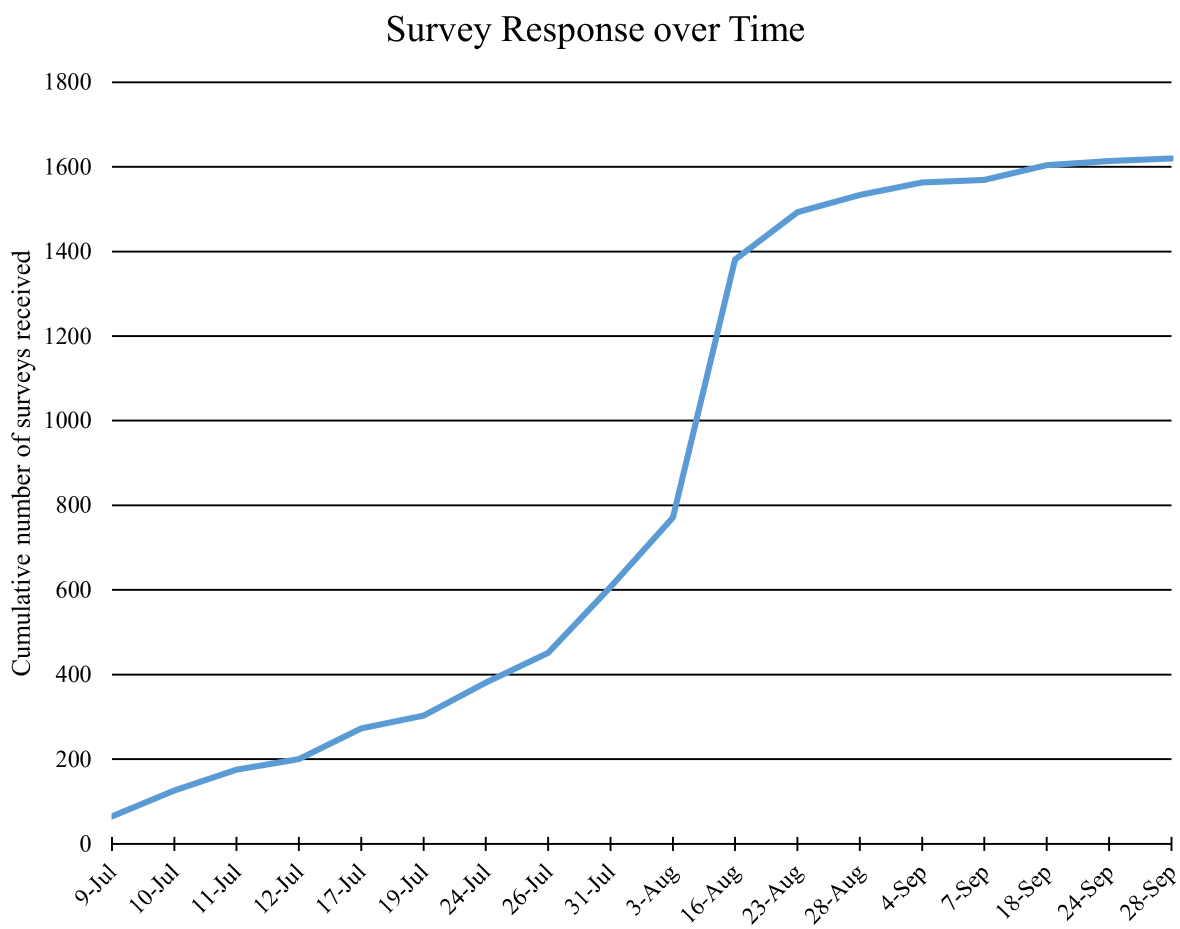


Figure 17. Survey Response Over Time for Mailed Survey Via Mail Reply

Mailed Survey – Completed via online survey system

The mailed survey could be completed by taking the online survey instead. The first response for this was on June 26, 2018 and the last September 19, 2018 (see Figure 18). The survey will remain open for as long as the surveys are still being returned via the mail to not inadvertently prevent someone from completing the survey. This will begin to expose the data to seasonality issues as it is beginning to span multiple seasons and thus will need to be addressed as the research progresses. Half of this sample was achieved by the 36th day.

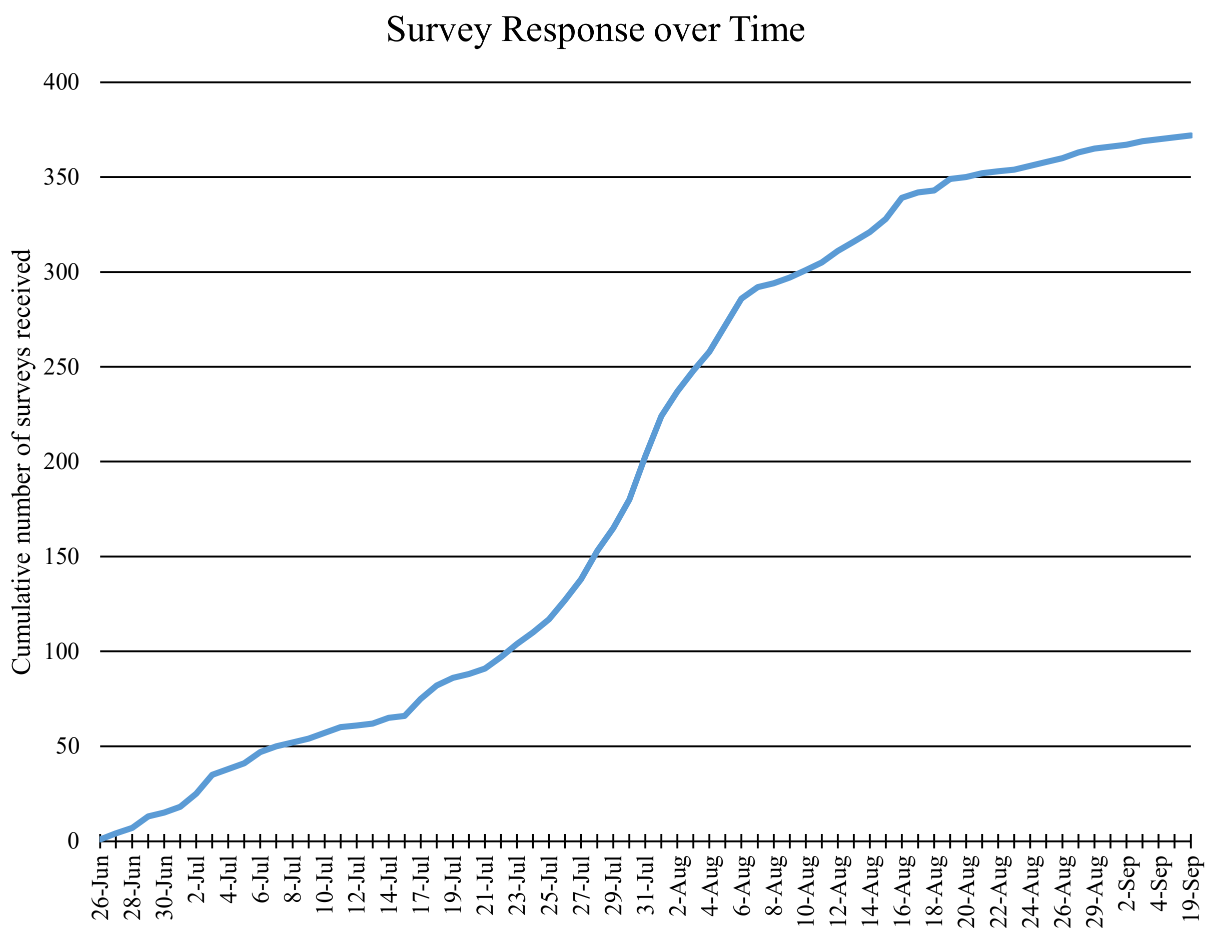


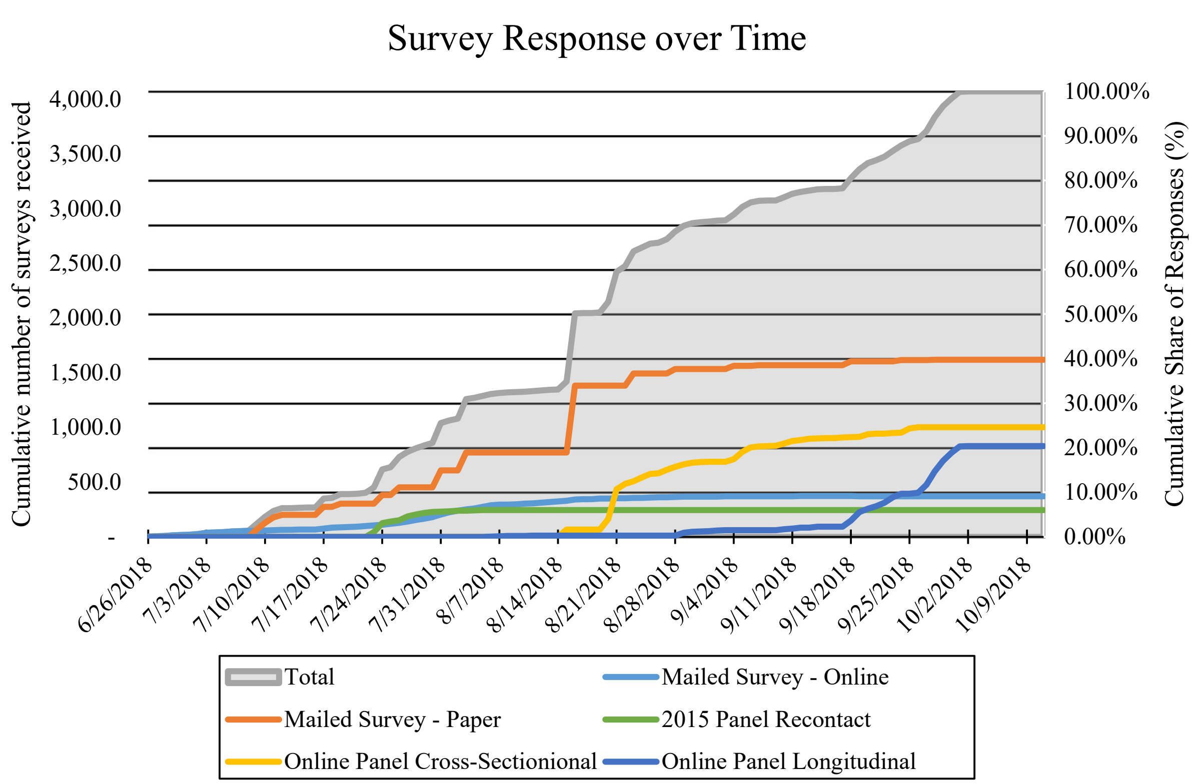
Figure 18. Survey Response Over Time for Mailed Survey Via Online Reply

A Combined View of the Sampling Methods Response over Time

Figure 19 depicts all the sampling methods responses and the total responses received during the data collection period. There are two unique characteristics for the online panel cross-sectional and panel samples that clearly show themselves in the combined chart. They are the delayed start for both samples and the early plateaus in the responses received. The delay start of nearly three weeks for the cross-sectional online panel and then an additional 2 weeks for the longitudinal sample was a result of unexpected personnel changes within the panel vendor. As I was finalizing the data collection launch plans, our primary contact at the vendor took a new position in the company and transitioned over to a new project manager. It then took additional time for the new project manager and research team to review the current project to make sure the deliverable would be achieved and be exactly what we wanted. Both parties were working as quickly as possible but due to challenges with scheduling, it was eventually completed after some delays.

Another interesting feature of these two samples is the early plateaus in the data collection. This was the result of the vendor conducting a "soft launch" of the survey, which was a process of collecting 10% of the final sample to ensure the quota system was functioning correctly, the survey was recording data correctly, and the responses were of high quality. After the review process was complete and any anomalies were addressed, the sampling would resume. No major issues were uncovered during this process, but it did vet our trap question system by terminating multiple cases, which were suspected automated responses, as well as establish the minimum acceptable time to complete the survey (14 minutes, half of the median time).

Figure 19. Survey Response Over Time for All Survey Methods



Data Entry of Mailed Survey

While four of the five sampling subsets were conducted via the online survey platform and thus already properly coded and input in the dataset, the surveys that were mailed back required an extensive data entry operation. Utilizing a team of undergraduate assistants, each survey was double entered into the survey platform. The responses were then checked against each other and any inconsistency between the two rounds of data entry would be flagged for further review. An additional manual check was conducted to review and recode each of the inconsistencies to the final value. The final step was to merge the confirmed correct values with the dataset.

3.4 Data Cleaning

To ensure the quality of the analysis based on the data collected, the survey responses went through a thorough cleaning process. The goal was to identify problematic cases and make sure the data was consistently coded before running the analysis. There were two main actions taken during this process. First identifying cases of such questionable quality that they needed to drop them from the study and second, finding obvious errors to appropriately recode. Please note that this research is based on the dataset as of July 19, 2019 as it had to be locked in to complete the analysis while other team members continued to clean the data for their analysis.

Review of Cases to Remove from Dataset

To identify the potential cases to be reviewed for quality control, a multi-step process was used. The first step was to run a series of logic-based tests on the responses to create flags for potential issues. These tests included:

- Error in sampling
    - If people completed the survey but lived outside the area of study (California) they were flagged to be dropped without further review once their home address was confirmed.
- Failing the trap questions

    - In Section A two questions request the respondent to provide a specific response and were flagged for removal if they did not answer with the requested value.
- Flatlining Section A
    - Section A consists of 35 Likert-type scale attitudinal statements which have the potential for a respondent to provide a single response for nearly the entire section. While this has the potential to be a valid response, the survey was designed to have multiple statements on the same topic, but each phrased positively and negatively, so a person would need to provide a different response (unless it was the middle response) to be consistent within the section. If a respondent was observed to have flatlined the section, they were flagged for further review to determine if a quality case.
- Speeding
    - During the testing and design of the survey, it was determined that it would be questionable for a respondent to finish the survey in less than 14 minutes given the length of the survey and the time needed to complete it. Cases that were below this threshold were flagged for review.
- Inconsistent responses
    - Throughout the survey there were question pairs that were used to check the consistency of the responses and if they were not in alignment the case would be flagged for further review. The following are a few examples of these types of checks that I utilized:
        - Household composition not totaling to the provided household size.
        - Provided commute information but state they are retired/do not work and not a student.
        - Provided telecommuting patterns but state they are retired/do not work and not a student.
        - Stated they work and are retired, which are incompatible statements.
        - If the respondent was recruited via the online opinion panel, their home zip code was asked twice and thus should be the same.
- Poor quality of open response questions
    - Open response questions provided an opportunity to assess the level of engagement with the survey as some questions required responses thus some respondents would enter gibberish, random letters, or nonsensical responses to proceed in the survey. Cases like this were flagged for review as it indicated the respondent may not be taking the survey seriously.
- Questionable travel patterns
    - If the stated commute to work was over 150 miles, it was flagged as it indicated an extreme value which might not be valid and therefore be a bad case.
    - Travel patterns for leisure and commuting were reviewed to identify any that stood out as being unlikely for using all modes listed in the question or a frequency of use that was unreasonably high, *e.g.*, commuting by driving alone in a private car, bus, commuter rail, bike, and walking all 5 or more times a week.
    - Cases that claimed more than 365 long distanced travel trips in a year were flagged as possible being a case that needs to be dropped.
    - Cases with extremely high values for weekly vehicles miles travel (*e.g.*, 5,000 miles) were flagged for review.

Once these checks were completed the total number of issues were tallied for each respondent to determine which were the most problematic. Not solely relying on mechanical checks, the most problematic cases were then manually reviewed in detail on a case-by-case basis to avoid dropping any valid cases that were just outside an expected or typical response. This process was labor-intensive given the number of cases that needed to be reviewed and the amount of data each person provided. However, the researchers needed to ensure the data was reliable. To prevent an individual researcher from imparting their own implicit biases on the dataset, all final decisions were reviewed and finalized by the entire research team to ensure there was consensus on the rationale, thus limiting any biases given the size and diversity of the research team. This process resulted in the identification of 349 cases that were dropped from the dataset. A complete log of all the cases that were dropped and the rationale for each decision is available upon request.

Recoding

The next step in the data cleaning was a thorough review of each case and variable to make any necessary recodes to the provided responses. After dropping the bad cases, the remaining cases were reviewed for issues on individual questions or key pieces of missing information. If a response was an error, such as a typo, and if the actual response could be determined, the response would be recoded as the intended response. For example, if someone said his commute was 100 miles but the distance between his home and office was 10 miles the response for commute distance would be recoded to 10 miles. The other key part of recoding was to establish and implement a system for the missing responses in the survey. Three different types of missing data were coded as described in Table 11.

Table 11. Missing Response Coding

Missing Response Value	Definition
-77777	The question was not displayed or should have been skipped due to the logic in the survey
-88888	Skipped a displayed question
-99999	Invalid response provided

It is worth noting that the missing responses of “skipped a displayed question” (-88888) and “invalid response” (-99999) were predominately associated with the non-online surveys since the online survey platform requires a question to be answered and validated to progress. These values were selected to provide a layer of security in our analysis to highlight missing values to avoid accidental inclusion in the analysis. The extreme negative value would be apparent in the output thus preventing any remaining in the analysis without proper removal.

3.5 Post Processing of the Data

Geocoding for Neighborhood Type

As part of a larger project, this research incorporates many outside data sources to provide a richer understanding of the individual and the individual's built environment. However, this research was limited to using just the geolocation data to determine neighborhood type. While the questionnaire asked for a self-reported neighborhood type of respondent's primary residency, the potential interpretation of the three levels (Urban, Suburban, and Rural) could vary widely. Thus, an impartial analytical approach to determine neighborhood type was implemented. A fellow researcher working on the project, Jai Malik, conducted the initial phases by inputting the home address or nearby street intersection from question C7 into the Google Maps API to retrieve the latitude and longitude of the location. If a respondent did not provide their home location in question C7 one of three approaches was used to determine a location at least to the zip code level of accuracy.

1. If participating via the mailed survey, the mailing address was used.
2. If participating online delete cross-section or longitudinal??, their home zip code was pulled from the screener question (S03).
3. If recalled from the 2015 panel, their prior address would be used if in the 2018 survey they stated in question (C9) that they did not move in the last three years.

Once all the respondents had a geo-location for their home address a two-step cleaning process was conducted. The first round was a high-level review of the results to determine cases that needed to be dropped or recoded. Cases were dropped if the home address was outside of California, which would put it outside the scope of the current study. Recodes were conducted on cases geo-coded by the Google Maps API but were misinterpreted or provided an invalid geo-code. Four members of the team divided the cases and conducted manual checks to determine the actual location.

The next phase of the cleaning was a deep dive into the remaining cases to identify other cases that were incorrectly coded by the Google Maps API. I conducted this part of the cleaning first by identifying cases that did not have consistent zip codes between the screener (S03), their home location (C7), the mailing address if available, and the geo-coded location. If any of these did not match the others, it was flagged for review. To determine if a recode was needed, the manual process of conducting a search on a map for these locations to identify errors was undertaken. The most prevalent type of error was that the location was on the boundary between zip codes. Another common error arose because of the leeway given respondents to pinpoint their home location, which was either the actual address or a nearby intersection. This allowed respondents to have invalid responses for intersections that would not match their other location details and thus get flagged as an error. In these situations, the respondents' response was honored, and the provided response was used even if a more precise location could be determined, typically via the mailing address of the survey.

Please note that while it may seem the Google Maps API provided many inaccurate or unreliable geo-codes, the research team for the 2015 Millennial study conducted a thorough investigation of this topic and found the Google Maps API provided the extremely accurate geo-coding, required the least additional prep work to conduct the geo-coding, and at the least cost. However, being aware that it is not 100% accurate especially given the creative and occasionally incorrect responses provided by the respondents, it was needed to conduct this level of review and cleaning of the geo-coded location data.

The geo-location data was used to determine the census track of the location which was cross-referenced with the neighborhood type analysis conducted by Salon (2015). The classification of all the census tracks in California is based on criteria from the built environment, the composition of the census tracks, and mode use mix for transit and walk/bike (Salon 2015). This was utilized as it was thoughtfully conducted so the veracity of it is not in question. Salon (2015) created a five-tier classification of Central City, Urban, Suburb, Rural-in-Urban, and Rural. These were then aggregated into Urban (Central City and Urban), Suburban (Suburban), and Rural (Rural-in-Urban and Rural) to reduce computational burden and ensure each category has sufficient size for analysis. This has the added benefit of being the same classification method used in the previous 2015 dataset, which will allow for more direct comparisons between the two in other research conducted on these datasets.

4. Methods and Analysis

After data cleaning the dataset consists of n=3,834 cases however due to the inherent limitations of the mailed survey, responding to all questions could not be mandatory as there is no mechanism to enforce this, so cases have missing data in the variables used in the model. To address this the cases with missing data were excluded from the final model as this seemed more prudent than imputing values as there was not a sufficiently accurate means to impute them. There were 916 cases with missing values that were listwise deleted making the final sample size used in the model to be n=2,918. Table 12 presents a summary of the sociodemographics of the sample. The distribution of neighborhood types in the sample is 22.3% rural, 46% suburban, and 31.7% urban; given the land-use makeup of California this is in line with expectations. Females account for 53.9% with Males making up the remainder. Respondents not employed because of either being retired or unemployed account for 36.8% of the sample. The sample is rather highly educated with 90.6% having at least some college for technical education. Household income levels of the sample are approximately evenly distributed across the income categories with the lowest category, <$25,000, being the outlier with 12.7% of the sample. The respondents' self-reported familiarity with AVs was high with 65.7% expressing at least some level of familiarity with the technology.

This data was used to conduct two statistical analyses as the means to address the research questions. The first analysis was a factor analysis on the attitudinal statements to determine the latent constructs. These results were then used in the LCA which segmented the respondents based on a set of key variables related to their stated interest in private and shared AVs, familiarity with AVs, and the likelihood of shared AV use impacting household vehicle ownership levels. The following sections present the details of these analyses.

Table 12. Summary of Sample (n=2,918)

Variable	n	%
Neighborhood Type		
Rural	651	22.3
Suburban	1343	46.0
Urban	924	31.7
Gender		
Female	1574	53.9
Male	1344	46.1
Transgender	0	0.0
Employment Status		
Employed	1844	63.2
Unemployed/Retired	1074	36.8
Education Level		
Some Grade School/High School	45	1.3
Completed High School	280	8.1
Some College/technical School	1485	43.1
Bachelor's Degree	1000	29.0
Graduate Degree	500	14.5
Professional Degree	135	3.9
Age		
18-37 years old	861	29.5
38-53 years old	919	31.5
54-72 years old	894	30.6
73 years old or older	244	8.4
Income		
Less than $25,000	372	12.7
$25,000 to $49,999	545	18.7
$50,000 to $74,999	494	16.9
$75,000 to $99,999	440	15.1
$100,000 to $149,999	552	18.9
$150,000 or more	515	17.6
Familiarity with AVs		
I have never heard of it	134	4.6
I have heard of it but am not familiar with it	866	29.7
I have heard of it and am somewhat familiar with it	1422	48.7
I have heard of it and am very familiar with it	496	17.0

4.1 Factor Analysis

Exploratory factor analysis (EFA) was used to determine the latent constructs present in the response to the attitudinal statements from Section A of the survey. This method was used instead of principal component analysis, which is foremost a data reduction method, as EFA provides the benefits of data reduction but also maintains the correlation between variables by identifying the latent constructs that explain observed variations in responses (Fabrigar, Wegener et al. 1999). EFA is the preferred method for this research as it allows for a data-driven approach that does not rely on any prior assumptions about the factors from the researchers. Also given that there are 30 statements it would make proscriptive methods like confirmatory factor analysis prohibitively time-consuming to test all the possible models and not reinforce assumed relationships in the factors (Malik, Alemi et al. 2021). An additional benefit of this process is it allowed for more efficient estimation of the final model and provided results that are more easily interpretable given its use of a smaller set of latent constructs rather than the full set of how the constructs are expressed. I worked with team members Ali Etezady and Jai Malik to conduct factor analysis of the attitudinal statements using SPSS's dimension reduction functions.

Section A of the survey consisted of 32 statements for which the respondents reported their level of agreement on a 5-point Likert-type scale, with 30 items being attitudinal statements and 2 additional ones designed as attention trap questions. Of the remaining 30 statements, all were included in the factor analysis except the statement associated with respondent's self-reported motion sickness ("I tend to feel sick when I am a passenger in a moving vehicle.") as it is reasonable to keep this as a stand-alone statement, as it is likely not a direct manifestation of (and is not strongly correlated with) an individual's attitudes or preferences.

To guide the exploratory factory analysis, I used maximum likelihood and an oblique rotation (ProMax and kappa less than 4) on the remaining 29 statements to generate a scree plot to guide the determination of the number of factors to use, as shown in Figure 20 below. The scree plot suggests that six factors might be ideal as it is at the elbow of the plot.

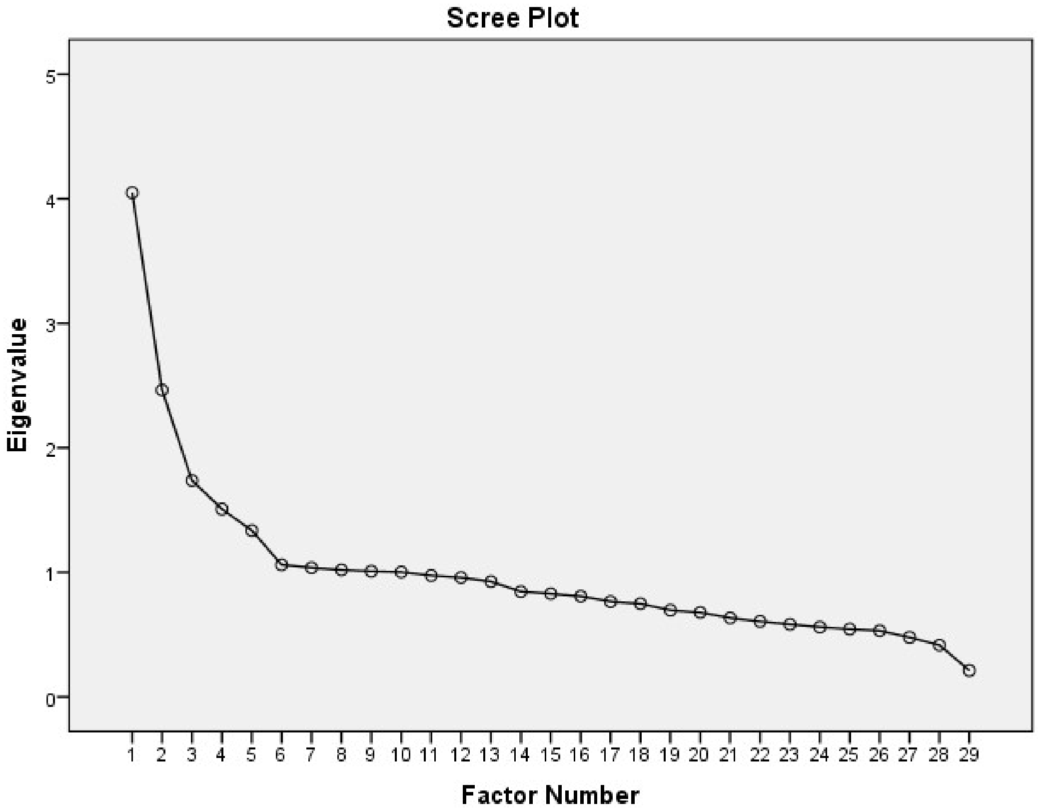


Figure 20. Scree Plot of Attitudinal Statements

To find an upper bound again the initial eigenvalues were considered. They suggested eight factors for this bound, as after eight factors the eigenvalues are smaller than 1, which is the typical threshold used in factor analysis. See Table 13 for a table of initial eigenvalues and the total variance explained for each factor solution.

Table 13. Initial Eigenvalues and Total Variance Explained

Number of Factors	Total	% of Variance	Cumulative %
1	4.38	15.10	15.10
2	2.98	10.27	25.37
3	1.91	6.60	31.97
4	1.63	5.64	37.61
5	1.46	5.02	42.63
6	1.11	3.81	46.45
7	1.03	3.57	50.02
8	1.01	3.50	53.52
9	0.98	3.39	56.91
10	0.89	3.07	59.98
11	0.86	2.96	62.94
12	0.85	2.93	65.87
13	0.80	2.76	68.63
14	0.75	2.57	71.20
15	0.73	2.53	73.72
16	0.69	2.38	76.10
17	0.69	2.37	78.48
18	0.65	2.25	80.73
19	0.63	2.16	82.89
20	0.61	2.09	84.99
21	0.59	2.02	87.01
22	0.57	1.97	88.98
23	0.54	1.87	90.85
24	0.54	1.85	92.70
25	0.52	1.79	94.49
26	0.51	1.75	96.23
27	0.47	1.61	97.84
28	0.41	1.43	99.27
29	0.21	0.73	100.00

These two early tests suggest that the ideal number of factors should be between six and eight. To provide a buffer and ensure that a higher number of factor solution was not preferred, the upper bound was increased to 10. This allows us to focus our effort on the range of 6 to 10 factors to ensure the results were interpretable with clearly defined factors.

Given the exploratory nature of factor analysis, many combinations of extraction methods (maximum likelihood and principal axis factoring), rotations (promax, direct oblimin, and varimax), and numbers of factors (five through ten) were tested to determine the ideal solution. The best solutions were generated when using Maximum Likelihood with a Promax rotation with 8 to 10 factors as it produced results that explained more of the variation (44%) while limiting the number of issues that would need to be addressed, such as poorly loading statements and muddled underlying factor concepts. The first refinement was to change the rotation method to Promax with Kaiser Normalization and a reduction in the Kappa to 3, which produced similar results as the prior method. Since no benefits were seen in the solution the original Promax rotation was kept. The second adjustment was to drop the statement related to the attitude towards strangers ("I am uncomfortable being around people I do not know.") as it was hindering the interpretability of the factor analysis solution. The results were more clearly interpretable when run with solutions for either eight, nine, and ten factors. The eight-factor solution still had mixed underlying concepts in the factors, but nine factors addressed this issue, and the full ten factors, while not having mixed concepts, started to split a previously clearly defined factor. This suggested that nine factors were the ideal solution. A further review of the pattern matrix brought to light the statement on liking bike riding ("I like riding a bike."), which was loading similarly across multiple factors and thus was determined to not be adding sufficiently to the results. The statement was dropped from the factor analysis and used instead as a stand-alone statement. The final 27 statements were then run again to obtain similar results to those from the prior runs but with more clarity in the definition of the factors. The nine-factor solution was selected as the final solution, which is present below in Table 14. To aid in the interpretation of the results, a threshold of .25 was set for the pattern matrix to clearly show which statements highly loaded on each factor by suppressing the results that fell below the threshold. The Bartlett Factor Scores were saved to the dataset as they will be used in the LCA.

Table 14. Statement Loadings for Factors (>.25)

Pattern Matrix									
	Factor								
	1	2	3	4	5	6	7	8	9
We should raise the price of gasoline to reduce the negative impacts on the environment.	0.927								
We should raise the price of gasoline to provide funding for better public transportation.	0.857								
The government should put restrictions on car travel in order to reduce congestion.	0.429								
I like to be among the first people to have the latest technology.		0.624							
Learning how to use new technologies is often frustrating for me.		-0.501							
Having Wi-Fi and/or 4G/LTE connectivity everywhere I go is essential to me.		0.492							
I like trying things that are new and different.		0.407							
I would/do enjoy having a lot of luxury things.		0.286							
I definitely want to own a car.			0.790						
I am fine with not owning a car, as long as I can use/rent one anytime I need it.			-0.421						
I prefer to be a driver rather than a passenger.			0.336						
I prefer to live in a spacious home, even if it is farther from public transportation and many places I go.				0.817					
I prefer to live close to transit even if it means I'll have a smaller home and live in a more crowded area.	0.261			-0.434					
I like the idea of having stores, restaurants, and offices mixed among the homes in my neighborhood.				-0.292					
My schedule makes it hard or impossible for me to use public transportation.					0.782				
Most of the time, I have no reasonable alternative to driving.					0.428				
I am too busy to do many things I'd like to do.					0.388				
My commute is a useful transition between home and work (or school).						0.454			
I like to juggle two or more activities at the same time.						0.377			
I try to make good use of the time I spend commuting.						0.364			
I am committed to an environmentally-friendly lifestyle.							0.581		
I prefer to minimize the material goods I possess.							0.444		
I am willing to pay a little more to purchase a hybrid or other clean-fuel vehicle.	0.326	0.250					0.388		
I am generally satisfied with my life.						0.252		-0.632	
I'm still trying to figure out my career (e.g., what I want to do, where I'll end up).								0.530	
The functionality of a car is more important to me than its brand.									0.619
To me, a car is just a way to get from place to place.									0.556
Extraction Method: Maximum Likelihood. Rotation Method: Promax with Kaiser Normalization.									
a. Rotation converged in 8 iterations.									

Note: Values below .25 have been suppressed for clarity in displaying the key values

Each of the factors was given a name that captured the latent construct of the factor; Table 15 provides a simple, descriptive means to reference the factors.

Table 15. Factor Names with Defining Attitudinal Statements

Factor Number	Factor Name	Attitudinal Statement
1	Pro-Sustainable Policy	We should raise the price of gasoline to reduce the negative impacts on the environment.
		We should raise the price of gasoline to provide funding for better public transportation.
		The government should put restrictions on car travel in order to reduce congestion.
2	Tech Enthusiast	I like to be among the first people to have the latest technology.
		Learning how to use new technologies is often frustrating for me.
		Having Wi-Fi and/or 4G/LTE connectivity everywhere I go is essential to me.
		I like trying things that are new and different.
		I would/do enjoy having a lot of luxury things.
3	Car Enthusiast	I definitely want to own a car.
		I am fine with not owning a car, as long as I can use/rent one any time I need it. (-)
		I prefer to be a driver rather than a passenger.
4	Pro-Suburbia	I prefer to live in a spacious home, even if it is farther from public transportation and many places I go.
		I prefer to live close to transit even if it means I'll have a smaller home and live in a more crowded area. (-)
		I like the idea of having stores, restaurants, and offices mixed among the homes in my neighborhood. (-)
5	Car Dependent	My schedule makes it hard or impossible for me to use public transportation.
		Most of the time, I have no reasonable alternative to driving.
		I am too busy to do many things I'd like to do.
6	Commute Multitasker	My commute is a useful transition between home and work (or school).
		I like to juggle two or more activities at the same time.
		I try to make good use of the time I spend commuting.
7	Eco-minimalist	I am committed to an environmentally-friendly lifestyle.
		I prefer to minimize the material goods I possess.
		I am willing to pay a little more to purchase a hybrid or other clean-fuel vehicle.
8	Life/Career Adrift	I am generally satisfied with my life. (-)
		I'm still trying to figure out my career (e.g., what I want to do, where I'll end up).
9	Car Utilitarian	The functionality of a car is more important to me than its brand.
		To me, a car is just a way to get from place to place.

After running a few exploratory LCA models the results were not providing interpretable results regarding the EFA's latent constructs. To address this a discretization process was conducted to change the output of the continuous variables from the EFA to a categorical variable to convey clear meanings and standardized the results across the factors. While the discrete values of the latent constructs are typically preferred, given the goals of this research it will be sufficient to know the relative magnitude of the EFA variables. The points taken into considerations for this decision were that the final output of this research, *i.e.*, policy recommendations, do not require the level of precision of the continuous variable, the results are more interpretable given their ease of direct comparison between categorical levels, and it reduced the effect of any outliers. There are many methods for discretization from the simplest of Equal Interval Width to complex, supervised methods like Vector Quantization (Dougherty, Kohavi et al. 1995) so it is good to consider its objective, which is "to find a set of cut points to partition the range of into a small number of intervals that have good class coherence" (Kotsiantis and Kanellopoulos 2006). Equal Frequency Interval discretization would provide a robust solution to achieve the goals by reducing a continuous variable into three categories and by standardizing it across the factors it would allow for direct comparisons between the factors. To implement this the factor analysis Bartlett scores were standardized through the software package used to conduct the factor analysis (SPSS). Then this was binned into the three categories of Low, Medium, and High with the thresholds set at +/- one standard deviation from the mean. This was selected as it resulted in the Medium category accounting for ~68% of the cases, which I feel is a fair and accurate assessment of what a middle category should encapsulate when the research outputs are taken into consideration.

4.2 Latent Class Analysis

The modeling methodology used to analyze the data is latent class analysis. This was selected as it provided a robust method to probabilistically segment the data. The conceptual model in Figure 21 was used to identify the groups of users with specific AV-related adoption propensity and study their associated expectations regarding changes in their household vehicle ownership.

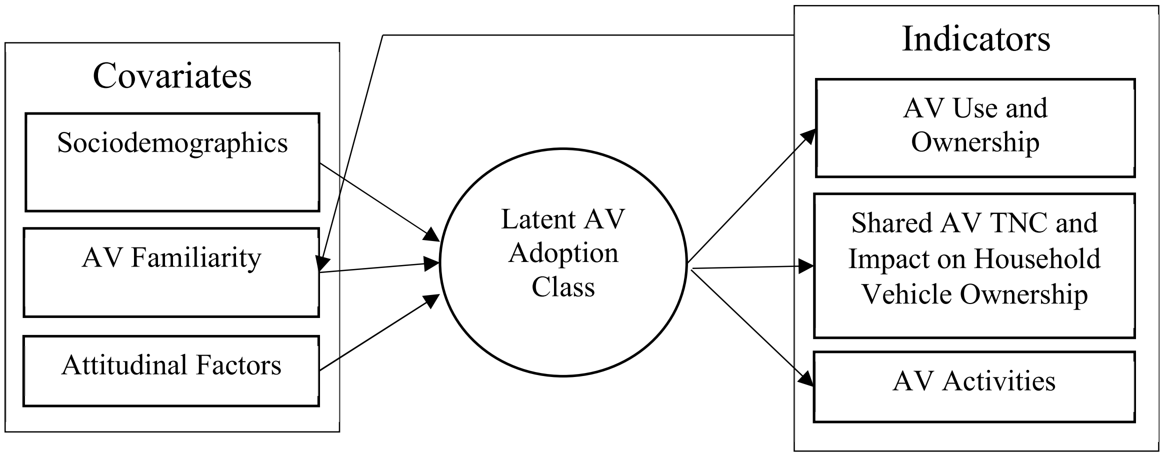


Figure 21. Conceptual Model for Latent Class Analysis

The active covariates of Sociodemographics, AV Familiarity, and Attitudinal Factors were selected to define the latent classes which then are used to estimate the likelihood of key AV-related activities across the indicators of AV Use and Ownership, AV TNC and Impact on Vehicle Ownership, and AV Activities. AV Familiarity plays a unique role as it can be an exogenous variable, but it can also be an endogenous variable as a person's environment may have a level of AV activity that encourages one to gain familiarity with it. In this case, the indicators would affect the level of AV Familiarity. While it is important to capture this in the conceptual model this interaction was not included in the model as the indicators as measured in the survey were measuring hypothetical future scenarios so it could not be the case of them influencing the AV Familiarity as they are not currently widely deployed on the streets.

The general model for the LCA is presented below in Equation 1 (Vermunt and Magidson 2002) and is the foundation for the LatentGold statistical software that was used to conduct the analysis.

Equation 1. General Form of LC Model

$$f(y_i|\theta) = \sum_{k=1}^{K} P_k \ f_k \ (y_i|\theta_k)$$

Where:
y_i is a vector of observed indicator variables for individual i
K is the number of clusters
θ denotes the model parameters
P_k denotes the prior probability of belonging to cluster k

Two quality of fit measures were used to aid in the assessment of an LCA model, Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC), to balance the model's specificity and sensitivity. They consist of a log-likelihood function (i.e., a goodness of fit term) and a penalty function to control overfitting; the general form of these information criteria (IC) is presented in Equation 2 (Dziak, Coffman et al. 2020)

Equation 2. Generalized Information Criterion

$$\mathrm{IC} = -2\ell + A_n p$$

Where:

ℓ denotes the log-likelihood
A_n denotes the penalty weight
n denotes the sample size
p denotes the number of parameters

While there is no single best IC for all scenarios the literature suggests that AIC is preferred when good future prediction is the emphasis while BIC is preferred when emphasizing a parsimonious model (Nylund, Asparouhov et al. 2007, Tihomir Asparouhov 2015, Dziak, Coffman et al. 2020). Is a variation of AIC and is another IC that I used as it more heavily penalizes the additional parameters in the models compared to AIC at a rate of 3 versus 2, respectively. This is beneficial in finding an optimum solution that favors parsimony in the model specification, i.e., not ballooning the number of parameters to artificially improve the AIC score. The penalty weights for the three ICs are present in Table 16:

Table 16. Summary of Information Criterion Penalty Weights

Information Criterion	Penalty Weight
AIC	$A_n = 2$
AIC3	$A_n = 3$
BIC	$A_n = \ln(n)$

Another factor to consider is that each IC tends to a different type of error, with AIC more likely to overfit while BIC is most likely to underfit (Dziak, Coffman et al. 2020). When considering the different emphasis, penalty weight, and likely error of each IC is best to run all of ICs to get a holistic picture of the quality of the model as the strengths and weaknesses of the ICs to offset each other resulting in the best attempt deriving the most valid assessment.

Interpretability was a subjective process where each estimated model's results were assessed based on how clear of a result it provided. The goal was to have classes of clear and differentiated characteristics. It was also important to have the results make sense in the real world as there is potential for results to be counter to observations due to forced unrealistic parameters on the model, *i.e.*, too many or too few classes.

To determine the specific variables among these categories and the number of classes, an iterative process of increasing complexity was used to estimate the model which was assessed for quality of fit and interpretability. The initial round included only the indicator variables related to AV Use and Ownership and AV TNC and Impact on Household Vehicle Ownership were used to provide an estimate of the number of classes to use in later rounds with the results being between 3 and 4. It was limited at this point to establish a baseline on a simplified model. Round 2 added sociodemographic variables of Household Income, Age, Neighborhood Type, Gender, Employment Status, and Education Level as covariates. The result of this estimation indicated that the additional indicators would be needed to add in the interpretability of the results and thus align with the conceptual model. Round 3 went back to a simplified indicator-only model to reestablish the ideal number of classes since all future models would include the full complement of variables. The results indicated that the additional classes should be considered so the following rounds 3-7 classes were estimated. Round 4 reintroduced the sociodemographic variables as covariates: while the results had a high quality of fit, the interpretability was difficult due to the limited number of characteristics available to define the classes. The 7 class results were poor and future rounds did not include an estimation of a 7-class model. Round 5 had the inclusion of a new indicator of expected changes in the level of car ownership in the next 3 years as it was directly related to the other variables. This additional was not impactful as the results followed in lock step with the other covariates and were therefore dropped because interoperability of the results was unaffected. The level of familiarity with AVs was added to the covariates as it made logical sense that a class of current familiarity would impact future expected behaviors. After estimating the results from these 4 levels of classes, an error was noticed in the level of responses for the AV familiarity. It had 5 categories and not the expected 4. This was an error introduced during the data cleaning process and was easy to recode to the correct value for the 2 cases that had the miscoded 5th categorical response. Round 6 was conducted at this point. It was determined that 3 classes would be the final solution as it provides interpretable results and the BIC supported this conclusion. Also, Round 7 was an intermediate step that involved reviewing the covariates and recoding where appropriate. The Age variable was the only variable that needed additional recoding. One category had only 53 responses compared to the others which were 5 to 18 times higher. It was decided to combine the low response category of 18-20 years old with the adjacent category of 21-37 years old to make an 18-37 year-old category. This also had the benefit of creating age categories that were more comparable in size. The 18-20 years old bin included only 3 years because of the artificial boundaries of the data collection process not allowing respondents under 18 to respond. This did not significantly affect the model's statistical results but did greatly improve the interpretability. The two combined age groups closely mirror one another thus simplifying the results while not reducing the meaning. Round 8 was an attempt to make the results even more interpretable by including additional covariates to define the characteristics of the different classes more clearly. The new covariates were the attitudinal factors that were binned (Low, Medium, High) and standardized to allow for comparison while not introducing additional complexity to the interpretability that would come with a continuous variable.

The final vectors of variables used for the indicators and covariates are presented in Table 17 and Table 18, respectively.

Table 17. Summary of Indicators

Variables	Question/Statement	Variable Type	Response Scale
AV Ownership – First to buy	I would be one of the first people to buy a self-driving vehicle.	5-point Likert-type scale	1=Very Unlikely – 5=Very Likely
AV Ownership – Wait until widely accepted	I would eventually buy a self-driving vehicle, but only after these vehicles are commonly used.	5-point Likert-type scale	1=Very Unlikely – 5=Very Likely
Use AV taxi service	I would use a driverless taxi alone or with others I know.	5-point Likert-type scale	1=Very Unlikely – 5=Very Likely
Vehicle Ownership – Keep current vehicle ownership level, not use AV TNC services	I would keep the vehicles(s) that I/my household owns (if any) and not use a driverless taxi or shuttle.	5-point Likert-type scale	1=Very Unlikely – 5=Very Likely
Vehicle Ownership – Keep current vehicle ownership level, use AV TNC services	I would keep the vehicles(s) I/my household owns (if any) and also use a driverless taxi or shuttle, when needed.	5-point Likert-type scale	1=Very Unlikely – 5=Very Likely
Vehicle Ownership – Reduce current vehicle ownership level, use AV TNC services	I would get rid of one (or more) of my household vehicles and use a driverless taxi or shuttle.	5-point Likert-type scale	1=Very Unlikely – 5=Very Likely
AV Activities – Increase travel while tired	I would more often travel even when I am tired, sleepy, or under the influence of alcohol/medications.	5-point Likert-type scale	1=Very Unlikely – 5=Very Likely
AV Activities – Send empty AV for simple errands	I would send an empty self-driving car to do simple errands (e.g., pick up groceries, pick up clothes from dry cleaners).	5-point Likert-type scale	1=Very Unlikely – 5=Very Likely
AV Activities – Send empty AV to pick up/drop off kids	I would send an empty self-driving car to pick up/drop off my child.	5-point Likert-type scale	1=Very Unlikely – 5=Very Likely
AV Activities – Travel more frequently for social/leisure activities	I would travel to social/leisure activities more often (e.g., dining at restaurants, shopping at malls).	5-point Likert-type scale	1=Very Unlikely – 5=Very Likely
AV Activities – Travel farther for social/leisure activities	I would go to more distant social/leisure activities (e.g., visiting friends, shopping).	5-point Likert-type scale	1=Very Unlikely – 5=Very Likely
AV Activities – More long-distance trips by AV, replacing other modes	I would make more long-distance trips by car because it would be less burdensome to travel in a self-driving car.	5-point Likert-type scale	1=Very Unlikely – 5=Very Likely
AV Activities – Work in AV	I would reduce my time at the regular workplace and work more in the self-driving car.	5-point Likert-type scale	1=Very Unlikely – 5=Very Likely

Table 18. Summary of Covariates

Variables	Question/Statement	Variable Type	Response Scale
Neighborhood Type	*Imputed via geolocation data*	Categorical variable	1=Urban 2=Suburban 3=Rural
Gender – Female	What is your gender identity?	Recoded to dummy variable	1=Female 0=Not Female
Employment Status	Are you currently employed?	Recoded to dummy variable	1=Employed 0=Not Employed
Education Level	What is your educational background? Please check the highest level attained.	Recoded to dummy variable	1=Bachelor's Degree or higher 0=Below Bachelor's degree
Age	In what year were you born?	Recoded to categorical variable	1=18-37 2=38-53 3= 54-73 4=73+
Household Income Level	Please check the category that contains your approximate annual household income before taxes.	Categorical variable	1=<$25,000 2=$25,000 to $49,999 3=$50,000 to $74,999 4=$75,000 to $99,999 5=$100,000 to $149,000 6=$150,000 or more
Level of Familiarity with AVs	We are interested in your awareness of or familiarity with the concept of self-driving vehicles before you started taking this survey. Please check the response that best describes you.	Categorical variable	1=I have never heard of it 2=I have heard of it but am not familiar with it 3=I have heard of it and am somewhat familiar with it 4=I have heard of it and am very familiar with it
Attitudinal Factor – Pro-sustainability	*Factor score based on self-reported level of agreement with attitudinal statements.*	Categorical variable	1=Low 2=Medium 3=High
Attitudinal Factor – Pro-technology	*Factor score based on self-reported level of agreement with attitudinal statements.*	Categorical variable	1=Low 2=Medium 3=High
Attitudinal Factor – Pro-car enthusiast	*Factor score based on self-reported level of agreement with attitudinal statements.*	Categorical variable	1=Low 2=Medium 3=High
Attitudinal Factor – Pro-suburbia	*Factor score based on self-reported level of agreement with attitudinal statements.*	Categorical variable	1=Low 2=Medium 3=High
Attitudinal Factor – Pro-car dependency	*Factor score based on self-reported level of agreement with attitudinal statements.*	Categorical variable	1=Low 2=Medium 3=High
Attitudinal Factor – Pro multitasking while commuting	*Factor score based on self-reported level of agreement with attitudinal statements.*	Categorical variable	1=Low 2=Medium 3=High -1 to 1
Attitudinal Factor – Anti-consumerism	*Factor score based on self-reported level of agreement with attitudinal statements.*	Categorical variable	1=Low 2=Medium 3=High
Attitudinal Factor – Life/Career Adrift	*Factor score based on self-reported level of agreement with attitudinal statements.*	Categorical variable	1=Low 2=Medium 3=High
Attitudinal Factor – Pro-car utilitarian	*Factor score based on self-reported level of agreement with attitudinal statements.*	Categorical variable	1=Low 2=Medium 3=High

4.3 Results

As discussed in the previous section, I tested various model specifications with two through five classes and used the criteria of the fit indices plus interpretability of the results to select the final model. Table 19 summarizes the fit indices (BIC, AIC, AIC3) for the two, three, four, and five class solutions.

Table 19. Fit indices for LCA solutions

Model	BIC_{LL}	AIC_{LL}	$AIC3_{LL}$
2-class	94500	93920	94017
3-class	90336	89487	89629
4-class	89117	87999	88186
5-class	89376	87988	88220

The indices suggested a solution of four classes as it is the local minimum for the BIC_{LL}, and $AIC3_{LL}$ fit indices. AIC_{LL} did not result in a minimum, which was expected as this fit index does not heavily penalize for the additional classes (and therefore additional parameters) as AIC3. The fit indices suggested a four-class solution, so it was further scrutinized for real-world interpretability. This solution did not pass this test even though the classes were statistically unique, but two of the classes were close enough that it made interpreting the results impractical. I then reviewed the five-class solution, but the classes continued to provide inconclusive interpretations. Then the three-class solution was reviewed which provided an interpretable result by not having classes that overlapped. The two-class model was examined but it reintroduced the issues of having ill-defined classes as it oversimplified the solution. Thus, the three-class solution was selected as the LCA model for the remainder of the analysis.

Class 1 has 949 members (32.51%), Class 2 has 1,259 members (43.15%), and Class 3 has 711 members (24.35%). Table 20 presents the membership probability for each covariate which is a measure of the likelihood of being in a particular class based on the observations of each variable. Then Table 21 presents the same data for the indicators.

Table 20. Membership Model for 3 Class Solution – Covariates

Covariate	Class 1 *AV Early Adopter* 32.5%	Cluster2 *AV Curious* 43.2%	Cluster3 *AV Hesitant* 24.4%
Neighborhood Type			
Urban	38.9%	41.2%	19.9%
Suburban	34.0%	43.9%	22.1%
Rural	20.3%	44.4%	35.3%
Gender			
Not Female	38.4%	42.0%	19.6%
Female	27.5%	44.1%	28.4%
Employment Status			
Not Employed	23.2%	45.4%	31.4%
Employed	37.9%	41.8%	20.3%
Education Level			
Up to some college/tech school	26.7%	43.0%	30.4%
Bachelor or Higher	37.1%	43.3%	19.6%
Age			
18-37 years old	45.4%	40.6%	14.0%
38-53 years old	33.2%	44.0%	22.8%
54-72 years old	23.6%	43.1%	33.3%
73 years old +	17.2%	49.0%	33.7%
Household Income			
Low (<$50k)	26.7%	45.7%	27.6%
Medium ($50k-100k)	31.4%	40.4%	28.2%
High (>$100k	38.5%	43.3%	18.2%
Familiarity with AVs			
I have never heard of it	25.9%	49.6%	24.5%
I have heard of it but am not familiar with it	24.5%	46.8%	28.6%
I have heard of it and am somewhat familiar with it	32.1%	44.3%	23.6%
I have heard of it and am very familiar with it	49.4%	31.7%	19.0%
Pro-Sustainable Policy			
Low	12.5%	38.7%	48.8%
Medium	31.5%	45.0%	23.6%
High	51.5%	40.4%	8.1%
Tech Enthusiast			
Low	13.3%	43.2%	43.5%
Medium	31.9%	45.3%	22.8%
High	52.6%	34.3%	13.1%
Car Enthusiast			
Low	38.6%	45.8%	15.6%
Medium	31.5%	42.7%	25.8%
High	30.1%	42.0%	27.9%
Pro-Suburbia			
Low	33.9%	40.9%	25.2%
Medium	31.3%	46.0%	22.7%
High	35.5%	34.6%	29.9%
Car Dependent			
Low	30.4%	42.5%	27.1%
Medium	32.2%	45.7%	22.2%
High	35.6%	36.3%	28.1%
Commute Multitasker			
Low	23.7%	46.7%	29.6%
Medium	30.8%	44.7%	24.6%
High	47.8%	33.6%	18.6%
Eco-minimalist			
Low	23.4%	44.4%	32.2%
Medium	32.6%	44.0%	23.4%
High	41.0%	38.2%	20.7%
Life/Career Adrift			
Low	21.3%	40.6%	38.1%
Medium	33.2%	43.3%	23.5%
High	40.7%	45.0%	14.3%
Car Utilitarian			
Low	27.1%	43.5%	29.4%
Medium	33.2%	44.5%	22.4%
High	35.1%	37.3%	27.6%

Table 21. Membership Model for 3 Class Solution – Indicators

Indicator	Class 1 *AV Early Adopter* 32.5%	Cluster2 *AV Curious* 43.2%	Cluster3 *AV Hesitant* 24.4%
AV Use and Ownership			
Be one of the first to buy an AV			
Very Unlikely	12.5%	45.9%	41.7%
Somewhat Unlikely	43.5%	54.0%	2.5%
Neither Unlikely nor Likely	51.8%	47.9%	0.3%
Somewhat Likely	83.6%	16.3%	0.1%
Very Likely	91.7%	8.4%	0.0%
Eventually buy an AV, only after commonly used			
Very Unlikely	2.9%	25.3%	71.8%
Somewhat Unlikely	11.6%	68.4%	20.0%
Neither Unlikely nor Likely	25.1%	68.4%	6.6%
Somewhat Likely	50.9%	47.2%	1.9%
Very Likely	78.9%	20.8%	0.3%
Willing to use an AV TNC service			
Very Unlikely	4.4%	32.8%	62.8%
Somewhat Unlikely	16.7%	75.0%	8.3%
Neither Unlikely nor Likely	33.2%	65.3%	1.5%
Somewhat Likely	65.3%	34.7%	0.1%
Very Likely	91.7%	8.3%	0.0%
Shared AV TNC and Impact on Household Vehicle Ownership			
Keep the same number of vehicles and not use an AV TNC service			
Very Unlikely	67.7%	29.0%	3.4%
Somewhat Unlikely	64.5%	34.8%	0.7%
Neither Unlikely nor Likely	44.8%	49.9%	5.3%
Somewhat Likely	38.9%	55.2%	5.9%
Very Likely	12.3%	38.8%	49.0%
Keep the same number of vehicles and use an AV TNC service			
Very Unlikely	5.1%	21.4%	73.5%
Somewhat Unlikely	14.4%	67.1%	18.4%
Neither Unlikely nor Likely	25.2%	65.0%	9.8%
Somewhat Likely	54.9%	43.2%	1.9%
Very Likely	70.2%	23.6%	6.3%
Reduce the number of vehicles and use an AV TNC service			
Very Unlikely	16.5%	41.4%	42.1%
Somewhat Unlikely	42.3%	53.1%	4.6%
Neither Unlikely nor Likely	40.6%	53.6%	5.8%
Somewhat Likely	70.8%	28.5%	0.7%
Very Likely	75.9%	18.1%	6.0%
AV Activities			
Use AVs to travel more when tired or under influence of alcohol			
Very Unlikely	3.2%	20.5%	76.3%
Somewhat Unlikely	8.8%	70.9%	20.3%
Neither Unlikely nor Likely	16.9%	73.4%	9.8%
Somewhat Likely	46.9%	50.3%	2.9%
Very Likely	78.4%	20.8%	0.8%
Use AVs to do simple errands			
Very Unlikely	5.6%	29.2%	65.2%
Somewhat Unlikely	22.4%	70.5%	7.1%
Neither Unlikely nor Likely	28.2%	67.3%	4.5%
Somewhat Likely	58.3%	40.4%	1.4%
Very Likely	85.5%	13.5%	1.0%

Table 21 (continued)

Indicator	*AV Early Adopter* 32.5%	*AV Curious* 43.2%	*AV Hesitant* 24.4%
Use AVs to pick up/drop off kids			
Very Unlikely	15.3%	44.6%	40.1%
Somewhat Unlikely	45.9%	53.8%	0.3%
Neither Unlikely nor Likely	44.8%	54.2%	1.0%
Somewhat Likely	80.8%	19.2%	0.0%
Very Likely	89.9%	10.1%	0.0%
Use AVs to travel to leisure activities more often			
Very Unlikely	0.1%	15.6%	84.3%
Somewhat Unlikely	4.0%	87.2%	8.8%
Neither Unlikely nor Likely	18.8%	80.7%	0.5%
Somewhat Likely	59.0%	40.9%	0.1%
Very Likely	94.6%	5.4%	0.0%
Use AVs to go to more distant leisure activities			
Very Unlikely	0.1%	12.8%	87.1%
Somewhat Unlikely	2.8%	88.7%	8.5%
Neither Unlikely nor Likely	15.6%	84.0%	0.4%
Somewhat Likely	58.0%	41.8%	0.2%
Very Likely	93.4%	6.6%	0.0%
Use AVs to make more long-distance trips			
Very Unlikely	0.6%	15.0%	84.4%
Somewhat Unlikely	4.2%	88.5%	7.3%
Neither Unlikely nor Likely	15.9%	82.8%	1.3%
Somewhat Likely	50.3%	49.0%	0.7%
Very Likely	88.3%	11.3%	0.4%
Use AVs to work in the car and not at the office			
Very Unlikely	9.7%	32.6%	57.7%
Somewhat Unlikely	26.9%	71.7%	1.4%
Neither Unlikely nor Likely	35.9%	63.1%	1.0%
Somewhat Likely	73.4%	26.3%	0.3%
Very Likely	92.7%	6.8%	0.5%

To aid in the analysis a visualization of the results is prepared in Figure 22 as a profile plot. A profile plot is a useful way to compare the different classes by plotting the class-specific mean magnitude of each indicator for each class rescaled to lie within 0 and 1, so when viewed against each other one can see the distinct profile for the various indicators of each class. The rescaling is "accomplished by subtracting the lowest observed value from the class-specific means and dividing the results by the range" (Vermunt and Magidson 2005). Consistent with the previous description of the three classes, the members of the AV Early Adopter class have a higher average willingness to adopt and use AVs for all activities. Interestingly, they express a stronger propensity towards both buying an AV and using AV TNC service. Interestingly, and somewhat differently from my previous expectation, this class tends to be more "pro-AVs" in general, regardless of the ownership and operational model that is deployed for AVs. This class represents the most interested and engaged AV adopters of all classes, so it has been labeled as “AV Early Adopter”.

The profile of the AV Curious class identifies those who are interested in AVs but are somewhat more hesitant than the members of the first group. They tend to be moderately interested in adopting AVs and use them for several purposes, but also more resistant to being the first one to buy an AV and/or jumping on the automated TNC model and reducing their household vehicle ownership. For this reason, the class is labeled as “AV Curious” to capture their mild interest in AVs but that it is not strongly held.

Members of the AV Hesitant class have little interest in AVs or AV TNC services as suggested by the near 0 values for their profile except for a near 1 value for “Keeping the same number of vehicles and not using AV TNC services”, a clear marker of their adversity to the vehicle automation and their strong preference to maintaining the current status quo. Accordingly, this group is labeled as “AV Hesitant”.

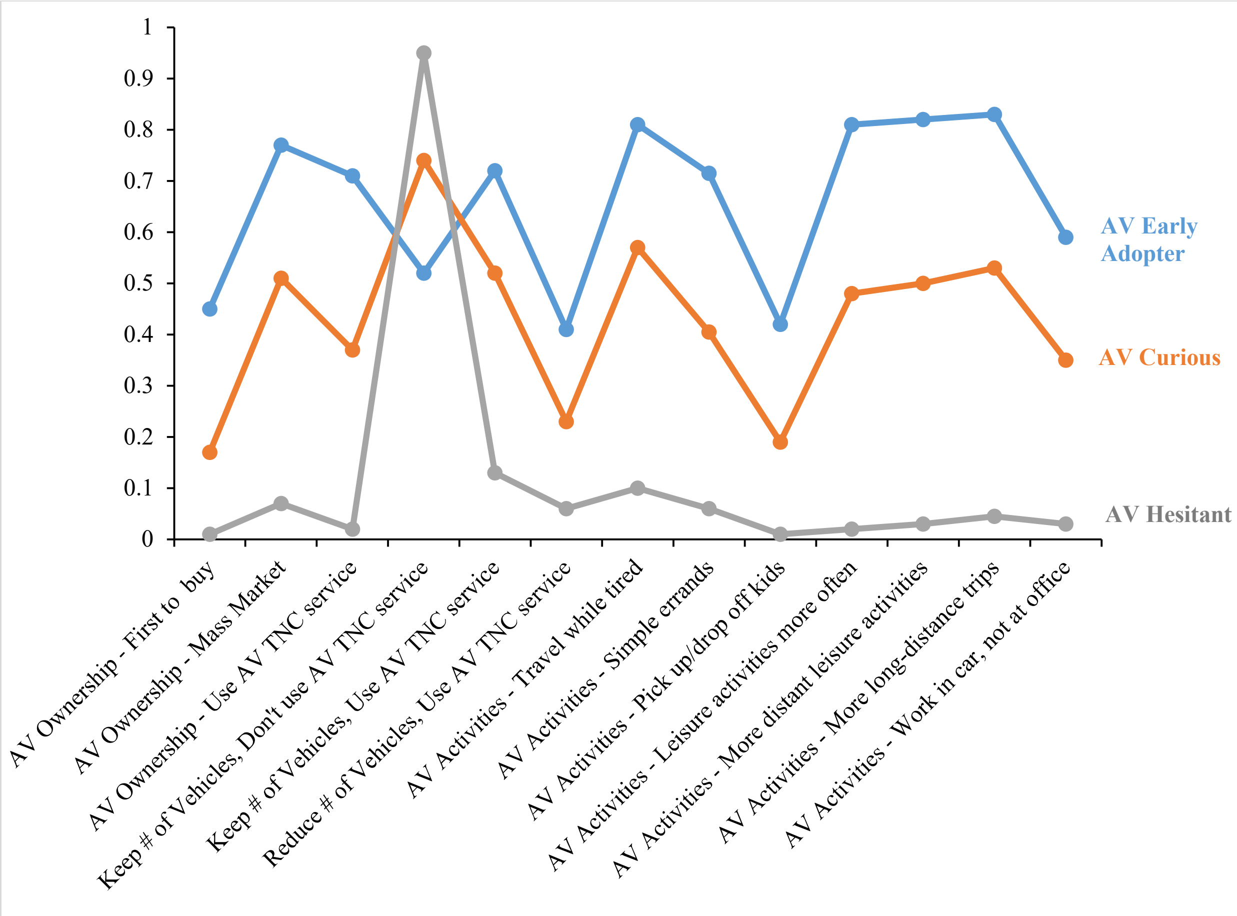


Figure 22. Profile Plot for 3 Class Model

With the classes well defined, it is important to understand who is composing the class. Table 22 lists the membership percent of the sample for each of the sociodemographic characteristics. As AV Early Adopter class is the largest share (43.15) of the sample it closely follows the full sample and will therefore be used as the reference for the comparison of the other two classes. Looking at the neighborhood type, AV Hesitant skew more rural (.32) while AV Curious has the highest likelihood of being in an urban location (.38). AV Hesitant lean more heavily toward being female (.63) and the AV Curious class goes against the full sample and has a bias towards being not female, *i.e.,* males and other non-binary responses. Employment status for AV Early Adopters is in line with the full sample while AV Curious is more likely to be employed while AV Hesitant is more likely to be unemployed. The AV

Curious class is more likely to be more highly educated while the AV Hesitant is more likely to be less highly educated. Regarding the age of the classes, AV Curious skews younger than AV Early Adopters, in contrast to the AV Hesitant who tend to be older individuals. The composition of the classes regarding their household income suggests that AV Curious people have higher household incomes compared to AV Hesitant which skews away from the high level towards medium and low levels of household income. Finally, for the self-reported level of familiarity with AVs before taking the survey AV Curious report a much higher level of being very familiar with AVs (.26) compared to the other classes, AV Early Adopters (.12) and AV Hesitant (.13).

Table 22. Distribution of Covariates by Class

	AV Early Adopter (n=949)	AV Curious (n=1,259)	AV Hesitant (n=711)	Sample Share (n=2,919)
Neighborhood Type				
Urban	30.2%	**37.9%**	25.9%	31.7%
Suburban	46.8%	**48.1%**	41.8%	46.0%
Rural	23.0%	14.0%	**32.3%**	22.3%
Gender				
Not Female	44.9%	54.4%	37.1%	46.1%
Female	55.1%	45.6%	**62.9%**	53.9%
Employment Status				
Not Employed	38.7%	26.3%	**47.4%**	36.8%
Employed	61.3%	73.7%	52.6%	63.2%
Education Level				
Up to some college/tech school	43.8%	36.1%	**54.8%**	44.0%
Bachelor or Higher	56.3%	63.9%	45.2%	56.1%
Age				
18-37 years old	27.8%	**41.2%**	17.0%	29.5%
38-53 years old	32.1%	32.2%	29.5%	31.5%
54-72 years old	30.7%	22.2%	41.9%	30.7%
73 years old +	9.5%	**4.4%**	11.6%	8.4%
Household Income				
Low (<$50,000)	33.3%	25.8%	35.6%	31.4%
Medium ($50,000-$100,000)	30.0%	30.9%	37.0%	32.0%
High (>$100,000)	36.7%	**43.3%**	27.4%	36.6%
Familiarity with AVs				
I have never heard of it	5.3%	3.7%	4.6%	4.6%
I have heard of it but am not familiar with it	32.2%	22.4%	34.9%	29.7%
I have heard of it and am somewhat familiar with it	50.1%	48.1%	47.3%	48.8%
I have heard of it and am very familiar with it	12.5%	**25.8%**	13.2%	17.0%

The attitudinal factors covariates are also important in understanding the composition of classes. As with the sociodemographic covariates, the AV Adopter closely follows the full sample share so it will again be used as the reference point for the analysis of the other two classes. See Table 23 for the full set of results. The AV Curious class has a much higher proportion of people who score high on Pro-Sustainable Policy (.31) than the expected (.18) while AV Hesitant skews the other direction with .30 scoring in the Low category compared to the expected .14. The same pattern is found in Tech Enthusiast with AV Curious have a higher proportion in the High category (.27) while AV Hesitant lean more to the Low category (.27). The attitudes toward being a Car Enthusiast are consistent across the classes with the only notable variance being that the AV Hesitant have a lower share in the Low category and a corresponding shift upwards in the Medium category. The factors for Pro-Suburbia, being a Car Dependent, or being a Car Utilitarian do not suggest anything across the different classes as the differences between them are minimal. The AV Curious class has a notable skew toward the High category for Commute Multitasker (.24) from the expected (.16). AV Hesitant are less likely to be in the High and Medium category for Eco-Minimalist and more likely to be in the Low category (.15) compared to the full sample (.15). The differences are the opposite for the AV Curious, with a lower share in the Low category (.11) and a higher share in the High category (.20) for Eco-Minimalist attitudinal factor. For having the attitude that their Life/Career is Adrift AV Curious people feel this more strongly (.20) than AV Hesitant (.10) which have a higher share in the Low category (.25) than expected (.16). This is likely due to AV Curious people skewing younger than the AV Hesitant and it is reasonable to assume that one’s life might not be fully defined and fulfilling compared to when one is older.

Table 23. Attitudinal Factor Covariates Percent of Sample

	AV Early Adopter (n=949)	AV Curious (n=1,259)	AV Hesitant (n=711)	Sample Share (n=2,919)
Pro-Sustainable Policy				
Low	13.5%	5.8%	30.2%	15.1%
Medium	68.3%	63.4%	63.4%	65.5%
High	18.2%	**30.8%**	6.5%	19.4%
Tech Enthusiast				
Low	15.2%	6.2%	27.2%	15.2%
Medium	71.6%	67.0%	63.9%	68.2%
High	13.1%	**26.8%**	8.9%	16.6%
Car Enthusiast				
Low	16.6%	18.5%	10.0%	15.6%
Medium	76.4%	74.8%	**81.7%**	77.2%
High	7.0%	6.7%	8.3%	7.2%
Pro-Suburbia				
Low	18.0%	19.9%	19.7%	19.0%
Medium	**68.8%**	62.1%	60.1%	64.5%
High	13.2%	18.0%	**20.2%**	16.5%
Car Dependent				
Low	19.9%	18.9%	22.5%	20.2%
Medium	63.3%	59.3%	54.5%	59.9%
High	16.8%	21.8%	**23.0%**	19.9%
Commute Multitasker				
Low	16.0%	10.8%	17.9%	14.8%
Medium	71.3%	65.2%	69.6%	68.9%
High	12.8%	**24.1%**	12.5%	16.4%
Eco-minimalist				
Low	15.8%	11.1%	20.3%	15.4%
Medium	70.3%	69.2%	66.3%	69.0%
High	13.9%	**19.8%**	13.3%	15.7%
Life/Career Adrift				
Low	15.1%	10.6%	25.2%	16.1%
Medium	67.6%	68.8%	65.1%	67.4%
High	17.2%	**20.7%**	9.7%	16.5%
Car Utilitarian				
Low	16.3%	13.5%	19.5%	16.2%
Medium	69.4%	68.7%	61.8%	67.4%
High	14.3%	17.8%	**18.7%**	16.5%

The sociodemographic makeup and attitudinal factors of the classes suggest the following generalization of the classes. The AV Curious class is composed of a highly urban and suburban population that is employed, younger, highly educated, has a Medium ($50,000-$100,000) to High household income (>$100,000), and is somewhat to very familiar with AVs. This class leans more towards being highly

Pro-sustainable Policy, Tech Enthusiast, and Eco-minimalist. The attitudes indicate that they see the potential benefits of the technology but given their current stage in life, *e.g.,* too young, they are hesitant to commit to new technology. Conversely, their enthusiasm for AVs may be hampered by the potential for AVs to be net negative on the environment and society. AV Hesitant is the opposite of the AV Curious with a shift towards being rural, more likely to be unemployed, less highly educated, older, and less likely making a High (>$100k) household income. They are less interested in Pro-Sustainable Policy and do not see themselves as being a Tech Enthusiast (two factors that are the main selling points for the technology) which aligns with not expressing interest in AVs. Also, they reported higher levels of being highly car dependent and seeing a car as a utilitarian device that again reinforces the suggestion from the model that these people see their vehicles as a crucial element in their lives and are hesitant to switch to new, disruptive technology. The AV Early Adopter fits between these two classes but tends to be closer to the AV Curious. AV Early Adopters covers a wide swath of the populace but are predominantly suburban, middle-aged people without strongly leaning attitudes. This is an interesting result as it hints that there might be something else driving this desire that is not captured in the model.

Table 24 shows the beta parameters for the indicators in the 3-class model, which is a "measure of the influence on that predictor" (Vermunt and Magidson 2005). This is a useful way to see the relative loading of each indicator between the three classes. The AV Early Adopters has very strong positive loading on indicators that suggest an early adoption of AV and AV services, "Be one of the first to buy an AV", "Eventually buy an AV, only after commonly used" and "Willing to use an AV TNC service" had parameters of 1.8227, 1.6729, and 2.1196, respectively. AV Early Adopters also exhibits a strong negative loading for "Keeping the same number of vehicles and not use and AV TNC service" (-1.02361) while having a positive parameter for the related indicator of "Keeping the same number of vehicles and use and AV TNC service" (1.2569). This suggests that they are responding to the AV TNC use and not the vehicle ownership levels and therefore continues to build the picture of this class as being very interested in AV use. There was a weaker loading of the "Reduce the number of vehicles and use an AV TNC service" (.8985), which is likely attributed to the reluctance to reduce vehicle ownership more so than the reluctance to use AVs. The remaining indicators, which all related to the potential for use of AVs for different tasks, all loaded strongly.

AV Curious follows a similar pattern as AV Early Adopters but loads more weakly on all the indicators which suggest they are interested in AVs but not nearly as enthusiastically as the first class. The three indicators for AV ownership and use, “Be one of the first to buy and AV” (0.7952), “Eventually but an AV, only after commonly used” (0.3102), and “Willing to use an AV TNC service” (0.6583), are all positive which suggests that this group would be interested in these behaviors but not with the same magnitude of the first class. The indicators related to changes in the vehicle ownership associated with the adoption of automated TNC services suggest that this class would be interested in using AV TNC services as the two indicators for AV TNC services were loading positively (“Keep the same number of vehicles and use an AV TNC service” (0.2433) and “Reduce the number of vehicles and use an AV TNC service” (0.2804)) while the one indicator for not using AV TNC service (“Keep the same number of vehicles and not use an AV TNC service”) was loading negatively (-0.3982). These loadings were again all in the same direction as Class 1 but with a much lower magnitude. The same trend continues with the AV activity indicators by being positively loading on the indicators but rather weakly.

The AV Hesitant is a very different class as it is the inverse of AV Early Adopters. The loadings for AV ownership and use are loading strongly negative, “Be one of the first to buy and AV” (-2.6179), “Eventually but an AV, only after commonly used” (-1.9831), and “Willing to use an AV TNC service” (-2.7779). This suggests a strong disinterest in owning or using AVs. This negative view of AV continues when looked at in relation to their vehicle ownership levels. The only indicator to load in the positive direction is “Keep the same number of vehicles and use an AV TNC service” (1.4242) as it is for the situation where no AV use is expected. Further supporting this clear class characteristic of having little interest in AVs are the indicators for AV activities, which all load very strongly and in a negative direction.

Table 24. Estimates of LCA Parameters for Indicators

	Class 1 *AV Early Adopter*	*Class 2* *AV Curious*	*Class 3* *AV Hesitant*	Wald	p-value
AV Use and Ownership					
Be one of the first to buy an AV	1.8227	0.7952	-2.6179	348.1849	<0.001***
Eventually buy an AV, only after commonly used	1.6729	0.3102	-1.9831	474.9077	<0.001***
Willing to use an AV TNC service	2.1196	0.6583	-2.7779	413.1809	<0.001***
Shared AV TNC and Impact on Household Vehicle Ownership					
Keep same number of vehicles and not use an AV TNC service	-1.0261	-0.3982	1.4242	254.047	<0.001***
Keep same number of vehicles and use an AV TNC service	1.2569	0.2433	-1.5003	405.5033	<0.001***
Reduce number of vehicles and use an AV TNC service	0.8985	0.2804	-1.1789	261.04	<0.001***
AV Activities					<0.001***
Use AVs to travel more when tired or under influence of alcohol	1.7559	0.2009	-1.9568	427.271	<0.001***
Use AVs to do simple errands	1.6505	0.4848	-2.1353	358.324	<0.001***
Use AVs pick up/drop off kids	1.8226	0.9598	-2.7824	298.904	<0.001***
Use AVs to travel to leisure activities more often	3.5722	0.8406	-4.4128	462.8171	<0.001***
Use AVs to go to more distant leisure activities	3.8219	0.9017	-4.7236	385.7579	<0.001***
Use AVs to make more long-distance trips	3.1601	0.6522	-3.8123	317.6634	<0.001***
Use AVs to work in car and not at office	1.9161	0.8878	-2.8038	248.3397	<0.001***

Note: *** denotes statistical significance at p<0.001

5. Conclusions and Next Steps

The following section discusses the conclusions of the research and some of its policy and private sector implications. Then the limitations of the research are discussed and how they might be addressed in future research.

5.1 Conclusions

By conducting a latent class analysis, I determined there were three classes of individuals related to their intentions towards the adoption and use of AVs for various activities. The three classes were defined as AV Early Adopter, AV Curious, and AV Hesitant. The AV Early Adopters were most interested in using and/or owning AVs and were middle-income, tech enthusiasts, and less enthusiastic for car ownership. The AV Curious group members were interested in AVs but were more interested in waiting until the technology matured, and using them to supplement their current vehicle ownership rather than replace them with a shared-AV service. The last segment was the AV Hesitant group which is more rural, older, and lower-income than the other segments. They were less likely to be concerned about environmental or sustainable policy and are enjoying their current vehicle use. This segment was the most reluctant to consider AV use.

The market segmentation suggested by the three-class model provides the groundwork for interesting applications across many cross-sections of the transportation field. The level of interest in AVs is high with the two classes that look at AVs positively, accounting for 75.66% of the market. The two classes that are interested in AVs, AV Early Adopters and AV Curious, seem more interested in TNC services in the early stage of deployment but when AVs are an established mass-market item they shift to a preference of private ownership. This is interesting as it shows that they want to try it before buying it while waiting for the technology to mature which is a prudent approach with such cutting-edge technology. This eventual switch to a preference for private ownership of an AV over their use as part of a TNC service is disconcerting as many of the benefits of lower cost, reduced emissions, and congestion is not realized unless AVs are shared, which is typically expected to be part of a TNC service. A recent study demonstrated via a naturalistic experiment that when AVs are used privately there is a "sizable increase in vehicle miles traveled and the number of trips" with "a substantial proportion of "zero-occupancy" vehicle-miles traveled" (Harb, Xiao et al. 2018). While some of these trips may be beneficial as they are new trips for under-traveled populations, *e.g.,* the elderly, there is still a potential for the benefit to be overshadowed by an influx of less beneficial VMT. The complexity of the situation will require strong policy to encourage AVs to be deployed in a manner that puts shared AVs as a priority over privately-owned AVs, still allowing for private ownership but reducing its negative impacts.

As automobile manufacturers have typically not been overly concerned with the negative externalities of their products, they will be encouraged by these results as they suggest that even with relatively little experience and knowledge of AV, consumers are interested in them. Because of clear demand for AVs, manufacturers will race to deploy AV technology at as many levels of the transportation system as they find profitable. While this is good for the rapid deployment of the technology as companies seek market efficiency by being the leader in a market segment, it will in turn put additional pressure on policymakers to keep pace. The main objects of the policymakers should be to ensure the AVs are deployed safely, equitably, and sustainably. The specifics of how to achieve these goals are outside the scope of this research but this analysis can be used to inform elements of these policy objectives.

If policymakers decide AVs are in the public interest, they should consider the level of external motivation that these segments need to get them to adopt AVs. The AV Early Adopters are already embracing this technology in the limited forms it is currently available in. So little effort should be focused on this group as they do not need any additional incentive. The AV Curious would likely need some policies and initiatives targeted at them but this should not require a massive investment as a small nudge should be enough to get them over their initial skepticism and then they would likely embrace it. As to what their reluctance is grounding in would need additional research but is likely rooted in safety concerns and familiarity/ease of use of new technology. These concerns could be addressed through education campaigns (*e.g.*, demonstrations, informational advertisements, or test drives) and the passive acquisition of experience and familiarity with AVs as the AV Early Adopters begin to normalize the use of AVs. The AV Hesitant group would need the largest and broadest set of policies to shift them into greater acceptance of AVs given their current disinterest in them which begins to illuminate the potential for inequitable AV adoption. The AV Hesitant were more likely to live in a rural neighborhood than the other classes therefore if left for the market to develop naturally these people may be the last to get the services deployed in their areas. While the case can be made from a TNC or automobile manufacturers that building AV services in rural areas are inefficient it should still be encouraged through policy initiatives. To support this, policymakers need to also run education and PR campaigns to inform AV Hesitant of the benefits of AVs to either create the demand for the services or encourage the willingness to use AVs when available. These education campaigns can utilize the attitudinal factors to describe the different classes to steer the messaging to ensure it speaks to the underlying values of each class which would aid the internalization of information that would elicit the desired travel behavior change. For example, a targeted campaign for AV Hesitant could utilize the messaging of its reliability and low cost while not explicitly mentioning the sustainability benefits or going into details on the technology.

Another important result from this study is the clear reluctance from all the classes to reducing their vehicle ownership levels even when presented with the potential for a robust AV TNC service that could replace a personal vehicle. Given California's well-documented love affair with the automobile (Marling 1989, Sachs 1992, Howe 1995, Falconer 2008) it will always be hard to break the cultural norm of personal car ownership even faced with the very real effects of anthropogenic climate change which have materialized more frequency with increasing intensity. As this data was collected in 2018 it does not consider the rise of the COVID-19 pandemic and the crippling effect it has had on shared service due to the requirement to avoid shared spaces and interacting with strangers. While the reluctance to reduce vehicle ownership level is good for the automobile manufacturers, it should be discouraged from a policy standpoint as it is a clear driver of greenhouse gas emissions, increasing congestion, negative health and safety impacts, and relinquishing an ever-increasing portion of public land to infrastructure to support this inefficient mode of transportation. It is hard for policymakers to change an established behavior (Zimbardo and Ebbesen 1970, Lunn 2012) and is even more of a challenge to change a widely accepted cultural behavior (Biglan 1995). This will be a process that will need long-term support both politically and financially as it is not likely behavior can be changed quickly. For lasting behavior change to be achieved according to the Precede-Proceed Model (Green and Kreuter 2005), policy needs to be applied across the three factors of behavior change which are predisposing factors (*e.g.,* attitudes, preferences), enabling factors (*e.g.,* social support, peer influence), and reinforcing factors (*e.g.,* supporting programs and services) (Gielen, McDonald et al. 2008).

This segmentation of the market would also be useful for AV manufacturers and TNC service providers to help them understand the composition of the potential market for AVs and related services. While the three classes are not necessarily the most revelatory by themselves as they follow typical technology adoption types, the attitudinal factors and sociodemographics would be useful in establishing other predictive models for product development, inclusion in forecasting business scenarios, and eventually inform marketing campaigns.

Long-range planning by regional transportation agencies has the challenging task of modeling future scenarios of the impacts of AVs even though many key variables are still not fully understood thus relying on assumptions for adoption, technology development timelines, and impacts on travel behavior. Childress, Nichols et al. 2015 present a good example of this as they created four scenarios for AVs impacts on the Seattle, Washington transportation system in 2040 via the region's activity-based travel model. To achieve this given the uncertainty inherent to this type of modeling, models levels of adoption (30%) for 2 scenarios and full adoption (100%) for the remaining 2 scenarios (Childress, Nichols et al. 2015). While this was a good approach for initial impact assessments of AVs in 2015, now this type of planning, especially in California, should incorporate more precise assumptions as the body of knowledge in the field has expanded. The segmentation presented in this research could be incorporated into long-range planning models to better reflect the likely adoption characteristics of AV users. With the three classes exhibiting different attributes, such as neighborhood type, sociodemographics, and expected adoption preferences, they demonstrate that AV adoption will not be uniform across the general population and the models should reflect this. This would allow planners to anticipate locations where adoption may happen first and ensure infrastructure and policy are in place to encourage responsible adoption of AVs. They would also be able to identify areas or user segments that are not receiving or utilizing the benefits of AVs and work to preemptively address potentially equity concerns that may arise from the imbalance in deployment and adoption.

In the time since the data collection was conducted the COVID-19 pandemic has sent shockwaves through transportation systems, as well as all other aspects of life. Therefore, it would not be reasonable to assume COVID-19 has not impacted the results of this study in at least a few ways. The use of shared transportation has taken a serious setback as most TNC services have dropped that option which will likely slow down the adoption of future shared services, automated or human-driven. Relatedly, during the initial spring 2020 peak of COVID-19 it was observed that non-shared ridehailing active users (used in last 30 days) dropped by as much as 66% (Matson, McElroy et al. Pending Review for Publication - 2021). Conversely, the appeal of a service that does not need a person to drive the vehicle thus providing a trip that adheres to social distancing precautionary measures may have increased the appeal of AV services. As the world continues the arduous task of vaccination, all of this may be eventually moot, but it is unclear on how long the tail will be on this catastrophic event and how long, if ever, it will take to get back to “normal”. So, it will be important to continue to follow transportation research related to the effects of COVID-19 to adapt the findings presented here to the “new normal” we all find ourselves in.

5.2 Next Steps for the Research

There are a few limitations to the research that needs to be acknowledged and can be addressed in any future research that is related to this topic. The most prominent limitation is that the area of study was only the state of California which might impact its generalizability beyond the frame, *e.g.,* at the national level. While great efforts were taken to ensure a broad sampling of the many regions and demographics groups throughout the state of California it is only possible to make inferences on how people in other locations may behave based on how similar they are to the sample, which has the potential to introduce additional error into the results. This has been addressed in subsequent data collection made by the research team and will be incorporated into future extensions of this research.

Another potential limitation of this research is the quality of responses in the AV-related section since it was the second to last in a lengthy survey that covered many topics. The potential for cognitive fatigue by the 25-minute mark is likely to have impacted the quality of responses. The data was reviewed to try and address this and remove any cases with obviously bad responses such as flatlining a section to quickly finish the survey or providing inconsistent responses, but this does leave the data exposed to the risk of a respondent not fully engaging with the question to get their true thoughts and behaviors. One possible solution for this would be to conduct data collection on each topic requiring less than 10 minutes but this would make the survey administration and associated costs increase which would preclude this as a viable solution.

One final limitation that warrants mentioning is the limited depth the survey allowed the researcher to delve into regarding expectations on AV use. The data collection had eight sections covering a broad selection of transportation-related topics with all future mobility topics relegated to one section. This limitation of space required the survey to be designed to provide the data deemed most valuable and only at a high level. Most of the AV questions only scratched the surface of the topic which could support an entire survey. Future research on this topic could begin to address AV use by conducting its own data collection to allow for a more thorough investigation of what was observed in this analysis.

During the time of analysis and writing of this thesis, the research team has conducted additional data collections which can be used to address some of the limitations of this research as well as explore related topics. The first of these new data collections acquired similar data in 8 cities across the country. Then a survey was developed to capture the initial impact of COVID-19 during the peak in the U.S. and Canada on travel behaviors which was an extension of the previous two surveys. There are also two additional data collections planned to collect additional COVID-19 impact data later in the pandemic (Fall 2020) in the Los Angeles region and a full sampling across the nation in Spring 2021. These additional datasets will allow researchers to explore how AV adoption preferences and attitudes change over time as well as explore other aspects of AVs and build upon the research present in this thesis.

Acknowledgments

This thesis was carried out as part of a larger study funded by a grant from the National Center for Sustainable Transportation (NCST) which is supported by the United States Department of Transportation (USDOT) and Caltrans through the University Transportation Centers program. This study also capitalizes on the work developed during three previous research grants funded by NCST and Caltrans. The author would like to thank the NCST, USDOT, and Caltrans for their support of university-based research in transportation, and especially for the funding provided in support of this project. Additional funding to support this research was provided by the 3 Revolutions Future Mobility (3RFM) research consortium of the University of California, Davis. The author would like to thank the many colleagues that contributed to the development of this research (in addition to the members of the thesis committee), including Farzad Alemi, Ali Etezady, Leticia Pineda, Diego Bernal, Rosaria Berliner, Sung Hoo Kim, Junia Compostella, Jai Malik, Niloufar Yousefi, and Kate Tiedeman for their contributions to the work presented in this paper.

References

Acheampong, R. A. and F. Cugurullo (2019). "Capturing the behavioural determinants behind the adoption of autonomous vehicles: Conceptual frameworks and measurement models to predict public transport, sharing and ownership trends of self-driving cars." Transportation Research Part F: Traffic Psychology and Behaviour **62**: 349-375.

Arbib, J. and T. Seba (2017). "Rethinking Transportation 2020-2030." RethinkX, May **143**: 144.

Asgari, H. and X. Jin (2019). "Incorporating Attitudinal Factors to Examine Adoption of and Willingness to Pay for Autonomous Vehicles." Transportation Research Record **2673**(8): 418-429.

Association, A. A. (2017). "AAA Reveals True Cost of Vehicle Ownership (2017 Your Driving Costs)." Retrieved Feburary 9, 2021, 2021, from https://newsroom.aaa.com/2017/08/aaa-reveals-true-cost-vehicle-ownership/.

Bansal, P. and K. M. Kockelman (2017). "Forecasting Americans' long-term adoption of connected and autonomous vehicle technologies." Transportation Research Part A: Policy and Practice **95**: 49-63.

Becker, F. and K. W. Axhausen (2017). "Literature review on surveys investigating the acceptance of automated vehicles." Transportation **44**(6): 1293-1306.

Biglan, A. (1995). Changing cultural practices: A contextualist framework for intervention research, Context Press.

Bösch, P. M., F. Becker, H. Becker and K. W. Axhausen (2017). Cost-based analysis of autonomous vehicle services. ETH IVV seminar, IVT, ETH Zurich.

Charness, N., J. S. Yoon, D. Souders, C. Stothart and C. Yehnert (2018). "Predictors of Attitudes Toward Autonomous Vehicles: The Roles of Age, Gender, Prior Knowledge, and Personality." Frontiers in Psychology **9**(2589).

Childress, S., B. Nichols, B. Charlton and S. Coe (2015). "Using an activity-based model to explore the potential impacts of automated vehicles." Transportation Research Record **2493**(1): 99-106.

Circella, G., G. Matson, F. Alemi and S. Handy (2019). "Panel study of emerging transportation technologies and trends in California: Phase 2 data collection."

Clayton, W., D. Paddeu, G. Parkhurst and J. Parkin (2020). "Autonomous vehicles: who will use them, and will they share?" Transportation Planning and Technology **43**(4): 343-364.

Compostella, J., L. M. Fulton, R. De Kleine, H. C. Kim and T. J. Wallington (2020). "Near- (2020) and long-term (2030–2035) costs of automated, electrified, and shared mobility in the United States." Transport Policy **85**: 54-66.

Cyganski, R., E. Fraedrich and B. Lenz (2015). Travel-time valuation for automated driving: A use-case-driven study. Proceedings of the 94th Annual Meeting of the TRB.

Dougherty, J., R. Kohavi and M. Sahami (1995). Supervised and unsupervised discretization of continuous features. Machine learning proceedings 1995, Elsevier**:** 194-202.

Dziak, J. J., D. L. Coffman, S. T. Lanza, R. Li and L. S. Jermiin (2020). "Sensitivity and specificity of information criteria." Briefings in bioinformatics **21**(2): 553-565.

Fabrigar, L. R., D. T. Wegener, R. C. MacCallum and E. J. Strahan (1999). "Evaluating the use of exploratory factor analysis in psychological research." Psychological methods **4**(3): 272.

Falconer, T. (2008). Drive: A road trip through our complicated affair with the automobile, Penguin Canada.

Gielen, A. C., E. M. McDonald, T. L. Gary and L. R. Bone (2008). "Using the precede-proceed model to apply health behavior theories." Health behavior and health education: Theory, research, and practice **4**: 407-429.

Gkartzonikas, C. and K. Gkritza (2019). "What have we learned? A review of stated preference and choice studies on autonomous vehicles." Transportation Research Part C: Emerging Technologies **98**: 323-337.

Green, L. W. and M. W. Kreuter (2005). Health Promotion Planning : An Educational and Ecological Approach. New York, McGraw-Hill.

Griggs, T. and D. Wakabayashi. (2018). "How a Self-Driving Uber Killed a Pedestrian in Arizona." Retrieved Oct. 15, 2020, from https://www.nytimes.com/interactive/2018/03/20/us/self-driving-uber-pedestrian-killed.html.

Harb, M., Y. Xiao, G. Circella, P. L. Mokhtarian and J. L. Walker (2018). "Projecting travelers into a world of self-driving vehicles: estimating travel behavior implications via a naturalistic experiment." Transportation **45**(6): 1671-1685.

Hardman, S., J. H. Lee and G. Tal (2019). "How do drivers use automation? Insights from a survey of partially automated vehicle owners in the United States." Transportation Research Part A: Policy and Practice **129**: 246-256.

Herrenkind, B., I. Nastjuk, A. B. Brendel, S. Trang and L. M. Kolbe (2019). "Young people's travel behavior – Using the life-oriented approach to understand the acceptance of autonomous driving." Transportation Research Part D: Transport and Environment **74**: 214-233.

Howe, J. (1995). Urban mobility and private car dependency. Urban Forum, Springer.

Hulse, L. M., H. Xie and E. R. Galea (2018). "Perceptions of autonomous vehicles: Relationships with road users, risk, gender and age." Safety Science **102**: 1-13.

Jiang, T., S. Petrovic, U. Ayyer, A. Tolani and S. Husain (2015). "Self-driving cars: Disruptive or incremental." Applied Innovation Review **1**: 3-22.

Johnston, C. and J. Walker (2016). Peak Car Ownership: The Market Opportunity for Electric Automated Mobility Services. www.rmi.org, Rocky Mountain Institute.

Keeney, T. (2017). "Mobility-as-a-service: Why self-driving cars could change everything." ARK Investment Management Research.

Kim, S. H., G. Circella and P. L. Mokhtarian (2019). "Identifying latent mode-use propensity segments in an all-AV era." Transportation Research Part A: Policy and Practice **130**: 192-207.

Kotsiantis, S. and D. Kanellopoulos (2006). "Discretization techniques: A recent survey." GESTS International Transactions on Computer Science and Engineering **32**(1): 47-58.

Krisher, T. (2020). "3 crashes, 3 deaths raise questions about Tesla's Autopilot." Retrieved Oct. 15, 2020, from https://abcnews.go.com/US/wireStory/crashes-deaths-raise-questions-teslas-autopilot-68045418.

Kyriakidis, M., R. Happee and J. C. F. de Winter (2015). "Public opinion on automated driving: Results of an international questionnaire among 5000 respondents." Transportation Research Part F: Traffic Psychology and Behaviour **32**: 127-140.

Lavasani, M., X. Jin and Y. Du (2016). "Market penetration model for autonomous vehicles on the basis of earlier technology adoption experience." Transportation Research Record **2597**(1): 67-74.

Lavieri, P. S., V. M. Garikapati, C. R. Bhat, R. M. Pendyala, S. Astroza and F. F. Dias (2017). "Modeling individual preferences for ownership and sharing of autonomous vehicle technologies." Transportation research record **2665**(1): 1-10.

Litman, T. (2009). "Transportation cost and benefit analysis." Victoria Transport Policy Institute **31**: 1-19.

Litman, T. (2014). "Autonomous vehicle implementation predictions." Victoria Transport Policy Institute **28**(2014).

Litman, T. (2020). "Autonomous vehicle implementation predictions: Implications for transport planning."

Lunn, P. D. (2012). "Behavioural economics and policymaking: Learning from the early adopters." The Economic and Social Review **43**(3, Autumn): 423–449-423–449.

Malik, J., F. Alemi and G. Circella (2021). "Exploring the Factors that Affect the Frequency of Use of Ridehailing and the Adoption of Shared Ridehailing in California." Transportation Research Record: 0361198120985151.

Malokin, A., G. Circella and P. L. Mokhtarian (2021). "Do millennials value travel time differently because of productive multitasking? A revealed-preference study of Northern California commuters." Transportation: 1-37.

Marling, K. A. (1989). "America's Love Affair with the Automobile in the Television Age." Design Quarterly(146): 5-20.

Matson, G., S. McElroy, Y. Lee and G. Circella (Pending Review for Publication - 2021). Longitudinal Analysis of COVID-19 Impacts on Mobility: An Early Snapshot of the Emerging Changes in Travel Behavior. Transportation Research Record.

Mattas, K., M. Makridis, P. Hallac, M. A. Raposo, C. Thiel, T. Toledo and B. Ciuffo (2018). "Simulating deployment of connectivity and automation on the Antwerp ring road." IET Intelligent Transport Systems **12**(9): 1036-1044.

Medina-Tapia, M. and F. Robusté (2018). "Exploring paradigm shift impacts in urban mobility: Autonomous Vehicles and Smart Cities." Transportation research procedia **33**: 203-210.

Milakis, D., B. van Arem and B. van Wee (2017). "Policy and society related implications of automated driving: A review of literature and directions for future research." Journal of Intelligent Transportation Systems **21**(4): 324-348.

Nair, G. S., S. Astroza, C. R. Bhat, S. Khoeini and R. M. Pendyala (2018). "An application of a rank ordered probit modeling approach to understanding level of interest in autonomous vehicles." Transportation **45**(6): 1623-1637.

National Academies of Sciences, E. and Medicine (2016). Between public and private mobility: Examining the rise of technology-enabled transportation services, Transportation Research Board.

Nazari, F., M. Noruzoliaee and A. Mohammadian (2018). "Shared versus private mobility: Modeling public interest in autonomous vehicles accounting for latent attitudes." Transportation Research Part C: Emerging Technologies **97**: 456-477.

Nieuwenhuijsen, J., G. H. de Almeida Correia, D. Milakis, B. van Arem and E. van Daalen (2018). "Towards a quantitative method to analyze the long-term innovation diffusion of automated vehicles technology using system dynamics." Transportation Research Part C: Emerging Technologies **86**: 300-327.

Nishimura, Y., A. Fujita, A. Hiromori, H. Yamaguchi, T. Higashino, A. Suwa, H. Urayama, S. Takeshima and M. Takai (2019). A Study on Behavior of Autonomous Vehicles Cooperating with Manually-Driven Vehicles. 2019 IEEE International Conference on Pervasive Computing and Communications (PerCom.

Nourinejad, M., S. Bahrami and M. J. Roorda (2018). "Designing parking facilities for autonomous vehicles." Transportation Research Part B: Methodological **109**: 110-127.

Nylund, K. L., T. Asparouhov and B. O. Muthén (2007). "Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study." Structural Equation Modeling: A Multidisciplinary Journal **14**(4): 535-569.

Penmetsa, P., E. K. Adanu, D. Wood, T. Wang and S. L. Jones (2019). "Perceptions and expectations of autonomous vehicles – A snapshot of vulnerable road user opinion." Technological Forecasting and Social Change **143**: 9-13.

Pettigrew, S., L. M. Dana and R. Norman (2019). "Clusters of potential autonomous vehicles users according to propensity to use individual versus shared vehicles." Transport Policy **76**: 13-20.

Pourmehrab, M., L. Elefteriadou, S. Ranka and M. Martin-Gasulla (2019). "Optimizing signalized intersections performance under conventional and automated vehicles traffic." IEEE Transactions on intelligent transportation systems **21**(7): 2864-2873.

Rödel, C., S. Stadler, A. Meschtscherjakov and M. Tscheligi (2014). Towards autonomous cars: The effect of autonomy levels on acceptance and user experience. Proceedings of the 6th international conference on automotive user interfaces and interactive vehicular applications.

Sachs, W. (1992). For love of the automobile: Looking back into the history of our desires, Univ of California Press.

Salon, D. (2015). "Heterogeneity in the relationship between the built environment and driving: Focus on neighborhood type and travel purpose." Research in Transportation Economics **52**: 34-45.

Shladover, S. E., D. Su and X.-Y. Lu (2012). "Impacts of cooperative adaptive cruise control on freeway traffic flow." Transportation Research Record **2324**(1): 63-70.

Singleton, P. A. (2018). "How useful is travel-based multitasking? Evidence from commuters in Portland, Oregon." Transportation research record **2672**(50): 11-22.

Sivak, M. and B. Schoettle (2015). Road safety with self-driving vehicles: General limitations and road sharing with conventional vehicles, University of Michigan, Ann Arbor, Transportation Research Institute.

Sivak, M. and B. Schoettle (2016). "Would self-driving vehicles increase occupant productivity." Sustainable Worldwide Transportation, The University of Michigan, Ann Arbor, http://umich.edu/~ umtriswt/PDF/SWT-2016–11_Abstract_English. pdf, accessed **20**.

Sperling, D. (2018). Three revolutions: Steering automated, shared, and electric vehicles to a better future, Island Press.

Stephens, T., J. Gonder, Y. Chen, Z. Lin, C. Liu and D. Gohlke (2016). Estimated bounds and important factors for fuel use and consumer costs of connected and automated vehicles, National Renewable Energy Lab.(NREL), Golden, CO (United States).

Sweet, M. N. and K. Laidlaw (2019). "No longer in the driver's seat: How do affective motivations impact consumer interest in automated vehicles?" Transportation: 1-34.

Tihomir Asparouhov, B. M. (2015). "Residual associations in latent class and latent transition analysis." Structural Equation Modeling: A Multidisciplinary Journal **22**: 169-177.

Vermunt, J. K. and J. Magidson (2002). Latent Class Cluster Analysis. Applied Latent Class Analysis. A. L. McCutcheon and J. A. Hagenaars. Cambridge, Cambridge University Press**:** 89-106.

Vermunt, J. K. and J. Magidson (2005). "Latent GOLD 4.0 user's guide."

Wang, K. and G. Akar (2019). "Factors affecting the adoption of autonomous vehicles for commute trips: an analysis with the 2015 and 2017 Puget Sound Travel Surveys." Transportation Research Record **2673**(2): 13-25.

Webb, J., C. Wilson and T. Kularatne (2019). "Will people accept shared autonomous electric vehicles? A survey before and after receipt of the costs and benefits." Economic Analysis and Policy **61**: 118-135.

Ye, L. and T. Yamamoto (2018). "Modeling connected and autonomous vehicles in heterogeneous traffic flow." Physica A: Statistical Mechanics and its Applications **490**: 269-277.

Zhang, W. and S. Guhathakurta (2017). "Parking spaces in the age of shared autonomous vehicles: How much parking will we need and where?" Transportation Research Record **2651**(1): 80-91.

Zhang, W., S. Guhathakurta, J. Fang and G. Zhang (2015). "Exploring the impact of shared autonomous vehicles on urban parking demand: An agent-based simulation approach." Sustainable Cities and Society **19**: 34-45.

Zimbardo, P. and E. B. Ebbesen (1970). "Influencing attitudes and changing behavior: A basic introduction to relevant methodology, theory, and applications."