Title

A Theory-Based Approach to Travel Behavior Analysis Using Pattern Recognition and Latent Group Identification Techniques

Author
McBride, Elizabeth Callahan

Publication Date
2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Santa Barbara

A Theory-Based Approach to Travel Behavior Analysis Using Pattern Recognition and Latent Group Identification Techniques

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Geography

by

Elizabeth Callahan McBride

Committee in charge:

Professor Konstadinos G. Goulias, Chair

Professor Karen Nylund-Gibson

Professor Daniel R. Montello

March 2022

The dissertation of Elizabeth Callahan McBride is approved.

Daniel R. Montello

Karen Nylund-Gibson

Konstadinos G. Goulias, Committee Chair

December 2021

A theory-based approach to travel behavior analysis using pattern recognition and latent group identification techniques

by

Elizabeth Callahan McBride

ACKNOWLEDGEMENTS

I would first like to thank my advisor Kostas Goulias. Thank you for seeing the potential in me from the very beginning when I was an undergraduate researcher. I will always appreciate your enthusiasm in helping me pursue my overly ambitious research goals. Your openness to learning new skills (like R), your curiosity, your drive, and your advocacy for your students have been inspirations for how I want to live my life. I could not have asked for more from an advisor.

I would like to extend my thanks to my committee members Dan Montello and Karen Nylund-Gibson for your patience, advice, and suggestions along the way. Your insights have been invaluable.

I am grateful for the generous financial support during my years in graduate school from the U.S. Department of Transportation, CalTrans, the University of California Transportation Center, and the Department of Geography at the University of California, Santa Barbara.

Thank you to Adam Davis and Jae Lee for teaching me how to be a graduate student. I can attribute so much of my current skillset to the two of you. Thank you to Adam for introducing me to the world of R programming and for your patience in helping me solve so many problems. To Adam, Jae, Rongxiang Su, and Jingyi Xiao: thank you for being such a warm and supportive group of brilliant researchers. It has been an honor to work with all of you.

Thank you to my mom, my dad, DJ, Brannon, Sari, Rebecca, and Kendra, for your support and pride and for so many conversations that inspired my research. I am also grateful for all the other friends, family, and colleagues that have supported me along the way.

Finally, thank you to Aram. You saw me through the most challenging year I have had as a researcher. Thank you for never judging me on my bad days, for pushing me when I needed it, and for keeping me sane. I am forever grateful for your unwavering support.

VITA OF ELIZABETH CALLAHAN MCBRIDE
December 2021

EDUCATION

Doctor of Philosophy (Ph.D.) in Geography Dec. 2021
University of California, Santa Barbara
Emphasis: Quantitative Methods in the Social Sciences

Master of Arts (M.A.) in Geography Dec. 2017
University of California, Santa Barbara
Thesis: "Enhancing Population Synthesis Using Land Use Indicators"

Bachelor of Arts (B.A.) in Environmental Studies & Geography Dec. 2014
University of California, Santa Barbara
Emphasis: Geographic Information Systems

RESEARCH EXPERIENCE

Co-Primary Investigator (Co-PI), Geography Department, UCSB Aug. 2020–Aug. 2021
Project Title: "A Before-After Intervention Experiment and Survey"
Sponsor: U.S. Department of Transportation

Lead Researcher, Geography Department, UCSB Aug. 2018–Aug. 2019
Project Title: "An Analysis of Accessibility, Social Interaction, and Activity-Travel Fragmentation in California"
Sponsor: U.S. Department of Transportation

Researcher, Geography Department, UCSB Mar. 2018–Mar. 2019
Project Title: "Qatar Activity-Based Model"
Sponsor: Qatar Ministry of Transportation and Communication

Lead Researcher, Geography Department, UCSB Mar. 2015–Mar. 2016
Project Title: "Spatial Transferability Using Synthetic Population Generation Methods"
Sponsor: California Department of Transportation

REFEREED JOURNAL PUBLICATIONS

McBride, E. C., Davis, A. W., & Goulias, K. G. (2020). Exploration of statewide fragmentation of activity and travel and a taxonomy of daily time use patterns using sequence analysis in California. Transportation research record, 0361198120946011.

Su, R., **McBride, E. C.**, & Goulias, K. G. (2020). Pattern recognition of daily activity patterns using human mobility motifs and sequence analysis. Transportation Research Part C: Emerging Technologies, 120, 102796.

Xiao, J., Su, R., **McBride, E. C.**, & Goulias, K. G. (2020). Exploring the correlations between spatiotemporal daily activity-travel patterns and stated interest and perception of risk with self-driving cars. AGILE: GIScience Series, 1, 1-15.

McBride, E. C., Davis, A. W., & Goulias, K. G. (2019). Fragmentation in daily schedule of activities using activity sequences. Transportation Research Record: Journal of the Transportation Research Board.

Davis, A. W., **McBride, E. C.**, & Goulias, K. G. (2018). A latent class pattern recognition and data quality assessment of non-commute long-distance travel in California. Transportation Research Record: Journal of the Transportation Research Board.

Davis, A. W., **McBride, E. C.**, Zhu, R., Janowicz, K., & Goulias, K. G. (2018). Tour-based path analysis of long distance non-commute travel behavior in California. Transportation Research Record: Journal of the Transportation Research Board.

McBride, E. C., Davis, A. W., & Goulias, K. G. (2018). A spatial latent profile analysis to classify land uses for population synthesis methods in travel demand forecasting. Transportation Research Record: Journal of the Transportation Research Board

McBride, E. C., Davis, A. W., Lee, J. H., and Goulias, K. G. (2017) Incorporating land use in synthetic population generation methods and transfer of behavioral data. *Transportation Research Record: Journal of the Transportation Research Board*, (2668), 11-20.

McBride, E., Lee, J. H., Lundberg, A., Davis, A. W., & Goulias, K. G. (2016). Behavioral micro-dynamics of car ownership and travel in the Seattle metropolitan region from 1989 to 2002. EJTIR, 16(4), 735-753.

REFEREED BOOK CHAPTERS

McBride, E. C., Davis, A. W., & Goulias, K. G. (2020). Sequence Analysis of Place-Travel Fragmentation in California. In *Mapping the Travel Behavior Genome* (pp. 371-398). Elsevier.

Goulias, K. G., Davis, A. W., **& McBride, E. C.** (2020). Introduction and the genome of travel behavior. In *Mapping the Travel Behavior Genome* (pp. 1-14). Elsevier.

Davis, A. W., **McBride, E. C.**, & Goulias, K. G. (2020). Environmental correlates of travel behavior from a destination attractiveness and activity timing perspectives. In *Mapping the Travel Behavior Genome* (pp. 83-102). Elsevier.

Davis, A. W., Lee, J. H., **McBride, E. C**., Ravulaparthy, S., & Goulias, K. G. (2019). California Business Establishment Evolution and Transportation Provision. In H. Briassoulis, D. Kavroudakis, & N. Soulakellis (Eds.), *The Practice of Spatial Analysis: Essays in memory of Professor Pavlos Kanaroglou* (pp. 295–323). Cham: Springer International Publishing.

Lee, J. H., Davis, A., **McBride, E.C.**, & Goulias, K. G. (2019). Statewide Comparison of Origin-Destination Matrices Between California Travel Model and Twitter. In C. Antoniou, L. Dimitriou, & F. Pereira (Eds.), *Mobility Patterns, Big Data and Transport Analytics* (pp. 201–228). Elsevier.

PRESENTATIONS AND INVITED LECTURES

Podium Presentation, "Travel Fragmentation among Disadvantaged Groups," International Association for Travel Behavior Research, Santa Barbara, CA, July 2018.

Poster Presentation, "A spatial latent profile analysis to classify land uses for population synthesis methods in travel demand forecasting" Transportation Research Board, Washington D.C., January 2018.

Poster Presentation, "Incorporating Land Use in Synthetic Population Generation Methods and Transfer of Behavioral Data" Transportation Research Board, Washington D.C., January 2017.

Podium Presentation, "Improving synthetic population generation methods," AAG, San Francisco CA, April 2016.

Podium Presentation, "Land use and synthetic population generation methods," UCConnect Student Conference, Riverside CA, February 2016

Poster Presentation, "Behavioral micro-dynamics of car ownership and travel in the Seattle metropolitan region from 1989 to 2002" Transportation Research Board, Washington D.C., January 2016.

Podium Presentation, "Behavioral micro-dynamics of car ownership and travel in the Seattle metropolitan region from 1989 to 2002" 4th Symposium of the European Association for Research in Transportation (hEART), Technical University of Denmark, September 2015.

TEACHING EXPERIENCE

Lead Teaching Assistant, Geography Department, UCSB — Sep. 2019–Dec. 2019
Geography 111A: Transportation Planning and Modeling

Teaching Assistant, Geography Department, UCSB — Sep. 2018–Dec. 2018
Geography 111A: Transportation Planning and Modeling

Teaching Assistant, Geography Department, UCSB — Sep. 2017–Dec. 2017
Geography 111A: Transportation Planning and Modeling

Teaching Assistant, Geography Department, UCSB — Mar. 2017–Jun. 2017
Geography 12W: Maps and Spatial Reasoning

ABSTRACT

A theory-based approach to travel behavior analysis using pattern recognition and latent group identification techniques

by

Elizabeth Callahan McBride

Human behavior is complex: the decisions people make are affected not only by their inner worlds, but also their personal and environmental conditions. In travel behavior research, there is very often a heavy focus on the conditions surrounding behaviors with a lack of focus on the mental processes underpinning the decision-making process. This dissertation presents a novel conceptual model of travel behavior that accounts for these important mental processes and develops data analysis techniques that can help answer research questions based on this framework. In this dissertation, first, the conceptual model is described. Next, an indicator for an important travel behavior concept called fragmentation is established. Common patterns of activities and travel are identified from detailed diaries of travel using sequence analysis, agglomerative nesting, and optimal matching techniques. The relationship between gender and fragmentation in different-sex couples is also analyzed. Then, the design, distribution, and analysis of a survey is presented. This survey looks at travel behavior when COVID-19 stay-at-home orders were in full effect. To the extent possible, the survey used the conceptual model as a guide for what questions to include. A

method using latent class analysis to classify travel behavior is also established. Travel behavior attitude are also classified using latent profile analysis. To assess the relationship between attitudes and behavior, the results of these two classifications are compared using cross-classification. Then, a model is created with the goal of following the conceptual model as closely as possible with the data available in the 2017 and 2019 Puget Sound Regional Council Transportation Study. The previously discussed travel behavior classification methods, index of fragmentation, and activity-travel patterns are all used in this model, and it expands upon the research into the relationship between fragmentation and gender. Finally, there is a discussion of issues in travel behavior survey methodology, and suggestions for how to improve these methods.

TABLE OF CONTENTS

1 Introduction 1

1.1 Background 1

1.2 Theoretical Framework 6

1.3 Dissertation Outline..13

2 An Analysis of Accessibility, Social Interaction, and Activity-Travel Fragmentation in California 16

2.1 Introduction 16

2.2 Data and Data Processing...19

2.3 Methods for Sequences 20

2.4 Small Study in San Louis Obispo and Santa Barbara 25

2.5 Statewide Analysis 44

2.6 Fragmentation within Households 59

2.7 Summary of Findings and Travel Behavior Research 67

3 A Pandemic Survey to Test Key Attitude-Behavior Components 70

3.1 Introduction 70

3.2 Survey Design and Data Collection 72

3.3 Data Collected 75

3.4 Analysis...78

3.5 Conclusions 88

4 Measuring the effect of Exogenous Variables on Latent Classes 90

4.1 Introduction 90

4.2 Methods...91

4.3 Results 105

4.4 Discussion 127

4.5 Conclusion...130

5 Conclusions and Synthesis 132

5.1 Synthesis..137

6 References 139

1 Introduction

Human behavior is a complex, enigmatic phenomenon. Understanding human behavior has been a topic of scientific study since the mid-1800's, yet there is still disagreement about how to explain behavior. The brain is not well-understood, and measuring cognitive processes through self-report questionnaires is susceptible to measurement error due to respondents' inability to recall information, untruthfulness, withholding of information, inability to access cognitive processes to answer questions about them, indirect relationship between attitudes/beliefs and behavior, and individual interpretation of questions (Ben-Akiva et al., 2012; Montello & Sutton, 2013). In the sub-field of travel behavior research, there is very often a lack of focus on the cognitive processes underpinning behavior (Gärling, 2020). This dissertation presents a novel conceptual model of travel behavior that considers these important cognitive processes and develops data analysis techniques that can helps answer research questions based on this framework. Although circumstances of the COVID-19 prevented the planned collection of data that was going to use this model as a guide, in Appendix A there is a suggested set of questions to measure cognitive processes that are theorized to impact behavior. In this dissertation, each chapter aims to investigate different behavioral facets and portions of this conceptual model of behavior.

1.1 Background

California residents' reliance on private cars has wide-ranging consequences across the state. According to the California Air Resources Board, as of the latest report in 2017, 28% of greenhouse gas emissions (GHG) in California come from passenger vehicles (California Air Resources Board, 2019). Compare this to the countrywide average of 17%, and the effects of California's "car culture" become more apparent (Office of Transportation and Air Quality, 2019). Moreover, the region known as the Southern California Association of Governments (SCAG)—which consists of the counties of Los Angeles, Orange, Riverside, San Bernadino, Ventura, and Imperial—is one of the most congested regions in the world (Southern California Association of Governments, 2019a). Of course, the responsibility for these issues does not lie with residents alone. Much of California's infrastructure was built with cars in mind, which means residents face poor public transit access, poor walkability, safety concerns, and long daily trip distances due to urban sprawl. If substantial changes in these areas are desired, it will take serious public investment. The COVID-19 pandemic restrictions' severe impact on travel adds complication to this issue, as the state of "normal" travel as it was previously understood may no longer exist.

The ongoing COVID-19 pandemic is poised to have long-term impacts on travel behavior. For jobs where it is available, telecommuting has become widely used, and online infrastructure has been rapidly created to accommodate this shift to working from home. Millions of people have also been unable to work and have claimed unemployment insurance since the beginning of this crisis. It is plausible that newly formed perceptions, attitudes, and concerns about infection might affect the use of public transportation for an unknown amount of time. However, it is also plausible that reduced traffic on the roads may give Los Angeles residents a "taste" of the possibility of having free-flowing transportation system, and this may affect attitudes and perceptions about the current system. There is also much to be learned from this crisis about people's travel values. This crisis has highlighted the importance of travel for spending time with friends and family, spending time outdoors, going out to bars and restaurants, traveling for fun, *et cetera*.

In the aftermath of COVID-19, although daily travel may not return to the previously normal state, as the country reopens daily travel will increase again and the congestion issues in the Los Angeles area will return. Long-term, there are serious issues with the car traffic in the Los Angeles area that must be addressed. As the metropolitan area population has grown, the road infrastructure has not grown with it. In fact, it would be implausible to expand the road system to such a degree due to the costly and disruptive nature of road expansion projects and the lack of space for such expansions. The phenomenon of **induced demand** also may contribute to this. In the field of transportation, induced demand is the concept that increasing road lanes leads to more cars on the road, and therefore does not cause long-term improvements to traffic. Research into the occurrence of this phenomenon has led to mixed results, with many examples showing its existence (Duranton & Turner, 2011; Hymel et al., 2010), while some studies have had mixed results with a weaker effect than predicted based on past studies (Cervero, 2003), and other research showing that the effect of induced demand is negligible (Mokhtarian et al., 2002). Nonetheless, as mentioned earlier, construction of a road system that could take on the car traffic load of the Los Angeles Area would be implausible simply due to the financial burden and network disruption. Reducing the number of cars on the roads, especially in a notorious area like Los Angeles, will be vital to creating a long-term sustainable system for the movement of people.

Reducing the number of cars on the road will require more use of public transportation, car sharing, carpooling, bicycling, and walking. However, there are numerous barriers to this in a city like Los Angeles, where existing infrastructure is “car-centric,” with unpleasant and unsafe sidewalks and bicycle lanes, long distances between neighborhoods, and poor public transit coverage; infrastructure modifications are costly and slow-moving; and car usage habits are deeply ingrained.

The attractiveness of the private vehicle over other modes of travel, and how to shift this dynamic, is a major topic of discussion in SCAG's newest Regional Transportation Plan (Southern California Association of Governments, 2019b). Although new electric car and self-driving car technologies will be an important part of solving these problems, they are not enough on their own to combat the rising congestion in Southern California. Investment in improvements to the public transportation systems and increased usage of the systems are vital to solve these issues. In large part, this has to do with the carrying capacity of public transportation networks versus individual vehicles. According to the National Association of City Transportation Officials (NACTO), on a single 10-foot lane at peak conditions, assuming one to two passengers per vehicle, private motor vehicles can move 600-1,600 people per hour. Meanwhile, a dedicated transit lane can move 4,000-8,000 people per hour, and on-street transitway, bus or rail can move 10,000-25,000 people per hour (National Association of City Transportation Officials, 2016). In addition, underground subways in most places do not interfere with surface traffic and provide a parallel network of activity and travel for people, thus easing congestion. The benefits of transit are well-established, and the challenge is how policymakers and public entities can influence reductions in personal vehicle usage in favor of public transportation systems. According to a 2018 UCLA report, in the SCAG region, per capita transit ridership was in decline over the preceding ten years, and evidence suggests that the ridership decline has mostly been due to increases in car ownership (Manville et al., 2018). The findings show "a dramatic increase" in car ownership in the region, with a disproportionate increase in vehicle access in the low-income and foreign-born groups that make up the majority of transit riders (p. 68). Manville et al. point out that most transit trips in Southern California are done by a small percentage of the region's population (2% ride transit very frequently, 20% ride occasionally). As vehicle access has important impacts on quality of life (QoL) for disadvantaged populations, the report suggests that instead of attempting to win back the population that has stopped using transit because of recent access to a vehicle (for whom car ownership represents a significant increase in QoL), it is better to target the "untapped potential" of the 77 percent of the population that report rarely or never riding transit (pp. 68-69).

1.1.1 Travel Behavior Theory

Mokhtarian, Salomon, and Singer (2015) stress the importance of looking at intrinsic motivations for travel behavior. In the paper, intrinsic motivation is defined as motivation to travel coming from enjoyment of traveling itself. This contrasts with extrinsic motivation, which is motivation to travel coming from activities being spatially separated. They discuss how the majority of travel behavior research in the past has primarily focused on extrinsic motivators, even though they are hardly sufficient to explain travel behavior. They emphasize that if stakeholders want to better understand and predict behavior, then it is necessary to examine the needs behind behaviors in finer detail than has been done in the past.

Mokhtarian *et al.* (2015) discuss how people with disparate intrinsic motivations would respond differently to policy changes. Examples of intrinsic motivations named in the paper include curiosity, seeking variety, seeking status, and independence. The paper suggests measuring these traits and others in individuals in future surveys so they can be incorporated into travel behavior modeling. When it comes to collecting trip information, they suggest including measurement of the extent to which individual trips are intrinsically versus extrinsically motivated. When asking about trip purpose in a travel diary, this can be accomplished by simply adding the option “just because I wanted/needed to.” It is also necessary to distinguish between travel *to* recreation and travel *as* recreation.

Practical implementation of these concepts presents certain challenges. A survey that thoroughly explores these concepts would likely put too much cognitive load on respondents. However, if the measurement methods are oversimplified, then they may not effectively measure the desired constructs. Balancing these factors is a primary challenge of this dissertation.

Chapter 1 of my dissertation provides a data-driven contribution to travel behavior theory that sets the stage for Chapters 2 and 3. Chapter 2 uses a survey during COVID-19 restrictions to study the correlation between of intrinsic and extrinsic factors and behavior. Chapter 3 repeats some Chapter 2 analyses, but in the context of another behavioral stimulus: the opening of a new Metro line in Los Angeles.

1.2 Theoretical Framework

All the theories outlined above are integrated into a conceptual model (Figure 1.1) for what makes up travel behavior and how attitudes and many other factors combine to lead to specific behaviors. In Figure 1.1, the rectangular boxes are responses to questions and the rounded rectangles are latent variables (called factors in factor analysis) that are used to explain the variation in responses to the items listed in the rectangles. The factors can be a continuous variable or a discrete variable. For example, the latent item “Travel Behavior Clusters” is a discrete variable with categories that represent different behaviors.

1.2.1 Past Research in Behavior

This section reviews the literature informing the conceptual model underlying this dissertation. First is a general discussion of behavioral research and travel behavior research. Then, the conceptual model that forms the framework of this dissertation is presented. The conceptual behavioral model was developed with the intention of testing its effectiveness using the originally planned data collection. The complete descriptions of the validated instruments under consideration to measure each of the constructs in the conceptual model can be found in **Error! Reference source not found.**.

The behavioral research that underlies the data collection and analysis here is based on past research about theoretical and empirical relationships between attitudes and behavior. There are a few important models on this discussed in detail in the following subsections. Behavior here means the amount of travel a person does in a day, the selection of modes for this travel, and any time and money allocation for travel. There are a few fundamental conceptual models that support other more popular conceptual models, and they are presented first below.

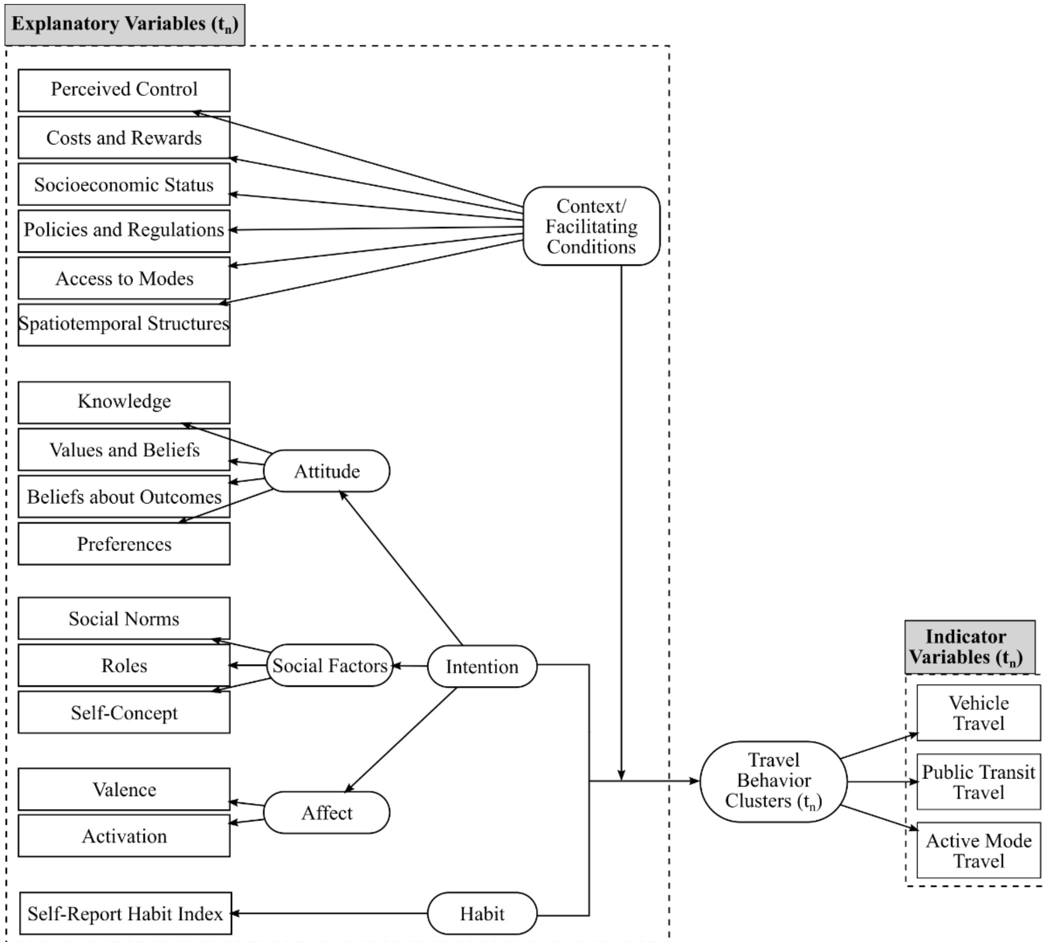
The Knowledge-Attitude-Behavior model suggests that increased knowledge will lead to favorable attitudes, which will in turn lead to favorable changes in behavior (Hungerford & Volk, 1990; Ramsey & Rickson, 1976). Evidence shows there is a moderate relationship between attitudes and behavior, and that the relationship between the two is mediated and moderated by several other variables. Also, it shows that the Attitude-Behavior relationship is not uniformly the same. It varies depending on the type of attitude, characteristics of the attitude (e.g., how long people hold an attitude and how consistent it is over time), and the way of measuring the attitude (Marcinkowski & Reid, 2019).

The Norm Activation Model theorizes that behavior will correspond to personal norms to the extent that someone is aware of the consequences of their behavior and feels some responsibility for those consequences (Bamberg & Schmidt, 2003).

1.2.2 Conceptual Model Used in this Dissertation

The model can be conceptualized as having two distinct kinds of explanatory variables: the cognitive processes that dictate behavior (**Intention** and **Habit** in Figure 1.1) and the conditions that affect how strong/weak the relationship is between cognitive processes and behavior (**Context/Facilitating Conditions** in Figure 1.1). These conditions can be personal characteristics (e.g., age or gender) or external forces (e.g., government policies). More detail about these constructs is provided next.

Figure 1.1 Conceptual model



1.2.2.1 Intention

As discussed in detail earlier, intention is the concept that people have an internal plan to perform certain actions. In this dissertation, intention is built from three latent items: attitudes, social norms, and affect.

Attitude. In this model, attitude is a person's worldview, and it is built from knowledge, values and beliefs, and preferences. Knowledge is a person's understanding of a situation, and it is a part of attitudes because evidence shows that knowing more about something can affect one's attitude about it (Hungerford & Volk, 1990; Ramsey & Rickson, 1976). Values and beliefs are the core morals of a person. They can include political alignment, religious beliefs, and more general principles someone aligns with like altruism. Preferences are a person's likes and dislikes (i.e., "I hate the bus" or "I love riding my bicycle").

Knowledge. Knowledge measures how much people know about the options available to them. As described above, the assumption is that a person's knowledge about the choices they are making will sway their attitude about those choices (Hungerford & Volk, 1990; Ramsey & Rickson, 1976).

Values and beliefs. Travel behavior intention is likely influenced by similar values to those discussed in general pro-environmental value literature. As described by de Groot and Steg (2009), pro-environmental behavior is generally assumed to be influenced by egoistic, altruistic, and biospheric values. These values influence beliefs, which in turn influence intentions. People with strong egoistic values will primarily consider the personal costs and benefits of pro-environmental behavior. People with strong altruistic values will primarily consider the costs and benefits to other people. People with strong biospheric values will primarily consider the costs and benefits for the planet.

Beliefs about outcomes. This comes from the Theory of Interpersonal Behavior (TIB). It is a person's evaluation of the potential outcomes of their decision. This includes perceptions of safety and convenience. For example, a belief that traveling by bicycle is unsafe, or that traveling by car is comfortable.

Preferences. Preferences are based on what people like and dislike, which should influence attitude. The challenge of measuring preferences is that they will be influenced by the other building blocks of intention (beliefs about outcomes, values, and lack of knowledge about the other options). In Random Utility Models (RUM), preferences are considered to be "revealed" from hypothetical choices people make. RUMs are estimated based on these "stated preferences" (Ben-Akiva et al., 1997; Hensher, 1994; Manski, 1977).

Social Factors. Social factors are the external influences on intention. They include the social norms, roles, and self-concept. Social norms are the standard behavioral expectations of the society a person lives in. Roles are the positions a person holds in society and/or the household. A person's behavior will be shaped by the responsibilities and expectations that come with those roles. Self-concept is the way a person sees him or herself. This can be aligned with roles and social norms, but also includes personal views about oneself. For example, someone seeing themselves as an environmentally friendly person, or a moral person.

Affect. Affect is the emotional state of an individual when thinking about a given behavior. Affect toward a behavior could include joy, pleasure, excitement, fear, anger, or anything in between. Russel and Barrett (1999) define "core affect" as occurring on two dimensions called "valence" and "activation." Valence is described as the spectrum of pleasure to displeasure or good to bad mood. Activation is the energy level of the affect, so an example of affect with negative valence and high activation would be anger, and an example of affect with negative valence and low activation would be sadness.

Related to affect are also ideas about Quality of Life (QoL) and subjective well-being that received considerable attention in travel behavior research (Friman et al., 2018). QoL can be divided into "objective" measures (e.g., educational attainment, household income, household wealth, and life expectancy) and in subjective well-being. Subjective well-being has three separable components as identified by Diener (2000): life satisfaction (global judgments of one's life), satisfaction with important domains (e.g., work and marital satisfaction), and experiencing positive and negative effects of emotions and moods (e.g., mental health). The COVID-19 pandemic impacted all dimensions of QoL with amplified impacts on some segments of population, with major differences across cohorts (generations), geographic regions, and groups (Clark et al., 2020; Ferreira et al., 2021; Tai et al., 2021).

1.2.2.2 Habit

Habit is an influence on behavior that is separate from intention. A habit is a pattern of behavior characterized by frequent occurrence, a low level of control over the choice of whether to perform the behavior, and a lack of awareness while performing the behavior. Inclusion of habit in the model comes from the TIB. Habits are repeated behaviors that, with time, become increasingly automatic and take up less mental energy to do. Because of the low cognitive load of choosing the habitual behavior, it can be difficult to choose a different behavior that will take more conscious effort, even if the intention to do so is there. According to Triandis (1977) and Robinson (2010), the strength of the influence of habit on behavior is determined by the novelty of a behavior, frequency of a behavior being performed, how well the conditions reflect the conditions when a behavior was performed in the past, and state of arousal (heightened state of arousal leads to more dependence on habit).

1.2.2.3 Context/Facilitating Conditions

The inclusion of context/facilitating conditions as a moderator for habit and intention is influenced by both the ABC model and the TIB. As explained in the descriptions of the ABC model and the TIB, these are conditions that can either hinder or compel behaviors. Stern (2000) says “Interventions do little or nothing until one of them removes an important barrier to change” (p. 419). These items range from structural conditions (e.g., proximity of a bus stop) to personal conditions (e.g., socioeconomic status). In the model, internal contextual factors include perceived control, costs and rewards, and socioeconomic status. External contextual factors include policies and regulations, access to modes, and spatiotemporal structures. All these conditions mediate whether a person can act in the way that they intend to or are habituated to. The TPB includes restricting factors as perceptions of the individual, while the TIB includes them as objective. Perceptions are important, but there are concrete things like spatiotemporal structure that are objectively present and influential on behavior. However, perceptions about those concrete things are also important (if not more important) than the objective facts.

Spatiotemporal structures are the temporal and spatial constraints to movement in a person's day. These constraints are determined by a person's schedule, which adds certain requirements to a person's day that will limit where they can go and for how long they can do activities. Following Hägerstrand's (1970) seminal keynote speech that has been the foundation of time geography, constraints are characterized as physical (e.g., speed of movement and maximum distance that can be reached), institutional (e.g., opening and closing hours of stores), and coupling (e.g., the need to be contemporaneously at the same place at the same time). These constraints are included in the box of spatiotemporal structures of Figure 1.1.

1.3 Dissertation Outline

In Chapter 2, the goals are to analyze how socioeconomic status variables influence travel behavior, establish an indicator for an important travel behavior concept called fragmentation, and to analyze what patterns of travel emerge when analyzing the daily diaries. Fragmentation of activities and travel is defined here as the sequencing of many short activities and trips that happen in a personal daily schedule. A fragmented schedule is a schedule in which a person switches activities at a high frequency during the day. To study this, first, a smaller study of the diaries in Santa Barbara and San Luis Obispo is performed to initially define indicators for fragmentation. Then, a refined method is applied to the entire state of California and followed by an analysis of patterns of travel that emerge.

Chapter 3 consists of the design, distribution, and analysis of a survey looking at travel behavior of adults in the Greater Los Angeles area in May 2020—at the height of the COVID-19 pandemic when stay-at-home orders were in full effect. Latent class analysis is used in this chapter to get classes of travel behavior from the travel modes respondents report using. Latent profile analysis is used on a set of questions about attitudes to get attitudinal profiles of respondents. These are then cross classified to investigate how strongly attitudes correspond to behavior.

In Chapter 4, 2017 and 2019 Puget Sound Regional Council Transportation Study (PSTS) data is analyzed. The PSTS contains a travel diary, in which respondents report every place they go, when they leave/arrive, how they get there, who they are with, and other traits of their trips over a designated period. The modes of travel used by respondents are used in a latent class analysis to identify underlying profiles of travel behavior. Common patterns in scheduling of activities are identified using similar methods to Chapter 2. These are used along with an index measuring fragmentation, socioeconomic status variables, and attitudinal variables as auxiliary variables to test how they relate to the latent classes of travel behavior. An interaction between fragmentation and gender is tested in this model to expand upon the research in Chapter 2.

In Chapter 5, a summary of the findings from each of the three groups of analysis is provided together with limitations, future work, and ideas to fill gaps found during this research effort. In Appendix A, the survey questions from Chapter 3 are also reported. Appendix B contains regression models that correspond to the work in Chapter 2. Appendix C contains a description of the preliminary steps taken to build a smartphone application to collect travel diaries.

2 An Analysis of Accessibility, Social Interaction, and Activity-Travel Fragmentation in California

2.1 Introduction

The following chapter contains writing presented in two published papers and a report to sponsors (McBride et al., 2020; McBride, Davis, & Goulias, 2019; McBride, Davis, Goulias, et al., 2019).

Fragmentation of activities and travel is defined in this dissertation as the sequencing of many short activities and trips that happen in a personal daily schedule. These are combined with other activities and travel that are much longer to form a complete string of episodes and durations of each episode by each observed individual. Fragmentation is of interest to travel behavior researchers due to concerns raised by Couclelis's (2000, 2004) hypothesis that economic, societal and political developments increase the individual's flexibility in scheduling daily activities. The more recent emergence of disruptive transportation services (e.g., Uber/Lyft) and automation (e.g., self-driving cars) also has the potential of added flexibility to reach places and therefore increased fragmentation.

Fragmentation in a schedule that is made of a sequence of activities means multiple switching between different activities in a day, e.g., the sequence of:

escorting children to schools—go to work—eat meal with colleagues—run errand—go back to work—go to a social event—go back to work—pick up children from schools—go shopping—return home—escort a child to soccer practice—do some work using mobile technologies—escort child back home—work at home

Patterns like this lead to increased transport demand because many activities are no longer bound to specific times and specific places, different people need to be escorted in different activity locations, and work can often be done ubiquitously. This is further enabled by Information and (tele)Communication Technologies (ICT) that release spatial and temporal constraints. The usual mode enabling fragmentation and flexibility in scheduling is the private car. This, however, may change dramatically due to the rapid emergence of ride-hailing services and other shared mobility services that release even more spatio-temporal constraints. These services are labeled as disruptive transportation and—under specific circumstances and for specific demographic segments—using disruptive transportation increases trip making. This theme was explored further in the conference IATBR2018 (see IATBR2018.org). This chapter demonstrates a relatively new method of travel behavior analysis to examine daily patterns in a holistic way. This sets the foundation to understand the potential impact of disruptive transportation by identifying ho*w a*n*d why individuals engage in activity-travel fragmentation.*

Sequence analysis is used here to describe fragmentation and daily patterns. In travel behavior, Wilson (*1998a*, *1998b*) uses biology-inspired sequence alignment methods to study the sequences of activities, Joh et al. (*2001*) explores different techniques to introduce space in the sequence analysis, Goulias (*1999*) uses Mixed Markov Latent class (MMLC) models to study year-to-year and day-to-day variation and transitions in activity-travel patterns, and McBride et al. (*2016*) look at lead and lag effects of land use on travel behavior using MMLC to analyze car ownership and travel as a function of demographic changes and land use.

Sequences of activities and the daily transitioning from one activity to another as well as the amount of time spent in each activity has been an important direction of travel behavior analysis (Auld et al., 2011; Bhat & Pinjari, 2000; Ettema et al., 1995; Přibyl &

Goulias, 2005). Examples include the Dutch diary analysis in Leszczyk and Timmermans (2002), in which gender and age are important determinants of moves from one activity type to another, and the formulation of methods to create daily models of activity participation and travel as in Auld et al. (Auld et al., 2011). Analysis of sequences of activities and travel is of paramount importance in formulating econometric models embedded in activity-based daily simulations of household activity-travel patterns for large-scale travel demand analysis (Bhat et al., 2013).

The following analyses aims to help understand and explain activity sequencing during a day at specific locations, activity duration by type, and their correlation with spatial opportunities as well as social and demographic characteristics. The *sequence analysis* here examines places visited by a person during a day jointly with the duration of activities at each place. It also examines the travel episodes and time spent to reach these places. Entire daily sequences of activities and travel are described by three indicators called *Entropy* (depicting variety in daily schedules), *Turbulence* (depicting complexity in daily schedules), and *Complexity* (based on entropy). These are summary indicators that complement each other to capture daily activity-travel patterns for each individual in a succinct mathematical way. In parallel, representative patterns of daily place-time allocation are derived, and their relationships with these indicators and determinants of travel are examined. All this uses samples of persons in the California Household Travel Survey (CHTS) (NUSTATS, 2013).

The key questions answered are:

1. Is there a clear taxonomy of daily sequences?
2. Are these types of sequences different in their fragmentation?
3. What are the characteristics of the people that use different types of sequences?
4. How different are the behaviors within these patterns?

2.2 Data and Data Processing

The data used in this analysis comes from the 2012 California Household Travel Survey (NUSTATS, 2013). This was a comprehensive travel survey conducted over all of 2012 and some of 2013, including information collected at the household and person level, vehicle ownership information, a one-day travel diary for every respondent, and other information not used here. In total, there were 114,639 respondents in 45,362 households.

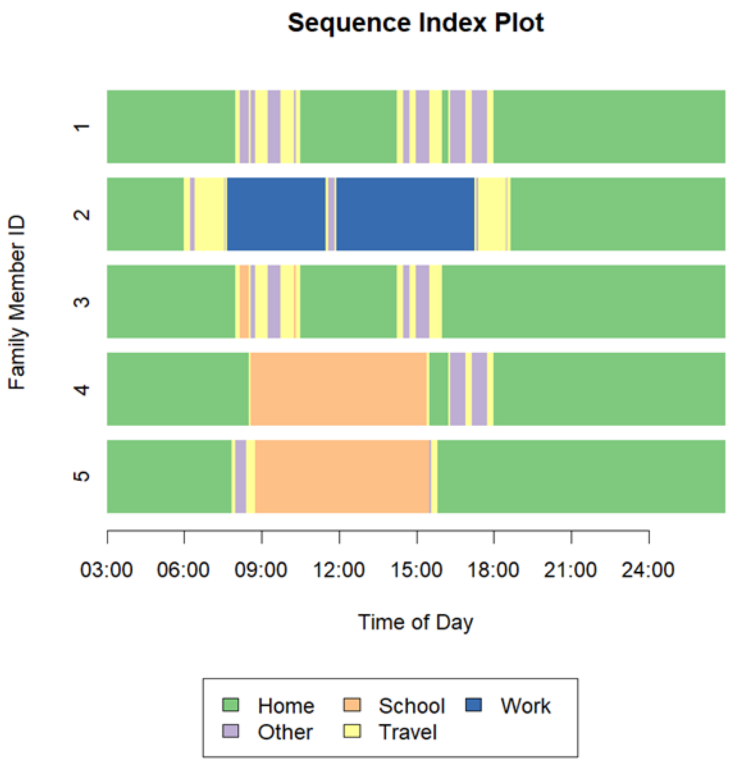
As a test area for the first sequence analysis method, only respondents residing in San Luis Obispo (SLO) and Santa Barbara (SB) counties are used. These counties were chosen for the test area because the researchers' familiarity with the region allows for confirmation of geographic accuracy. In total for this analysis, data from 2,942 persons are used. The Santa Barbara and San Luis Obispo regions comprise an area with a variety of land use types—ranging from high density/urban to low density/rural. Person characteristics, household characteristics, and one-day travel diaries for the respondents are used. The one-day travel diary takes place across the entire year (including holidays, weekends, and weekdays). It covers from 03:00 on the survey day until 02:59 on the following day. Respondents report every place they go, their travel mode, the top three activities performed at each place, and several other characteristics of the places visited and their travel. Land use surrounding each residence is depicted by indicators that are based on a detailed establishments inventory following techniques reported in Chen et al. (Chen et al., 2011) and McBride et al. (McBride et al., 2017). It was necessary to simplify the measurement of activities for this computationally intense analysis, so places people attended throughout the diary day were used, divided into activity at Home, Work, School, and all other places.

The second analysis is a statewide random sample of 5,000 households with a total of 12,704 persons living in many different places in California. Activities in the interview day of these persons were classified in the same way as Central Coast sample (i.e., activity at Home, Work, School, all other, and travel). Then, the same analysis as for the Central Coast is repeated. Male and female fragmentation within the same households are compared using the entire statewide database. The California Household Travel Survey is available at the Transportation Secure Data Center of the National Renewable Energy Laboratory (https://www.nrel.gov/transportation/secure-transportation-data/index.html).

2.3 Methods for Sequences

A sequence is a series of time points at which a subject can move from one discrete "state" to another. In this chapter, these states are based on the types of places people visit and stay during their diary day: Home, Work, School, and Other. Travel between these places is also considered a “state.” People who go through many states in their day are considered to have fragmented schedules. Sequence analysis is used to statistically analyze the fragmentation of respondents’ days using a minute-by-minute time series. Every minute of the day contains a specific state for each person in the study. Figure 2.1 shows an example of the sequences identified from each person’s diary in one family in the study area.

Figure 2.1. One family's sequence in a day (reproduced from McBride et al., 2019)



There are many techniques in the travel behavior field that can be used to measure the duration of episodes within a single state and the transition rates from one state to another (Auld et al., 2011; Bhat & Pinjari, 2000; Ettema et al., 1995; Kroesen, 2014; Přibyl & Goulias, 2005). These define the state of the art in longitudinal data analysis (Kroesen & Goulias, 2016). They can be useful for measuring fragmentation in a person's day but are cumbersome or infeasible when the number of transitions is very high.

We use Entropy, Turbulence, and Complexity that can handle very long sequences. The explanation follows McBride *et al.* (McBride, Davis, & Goulias, 2019) closely, with some important additions.

Entropy is a measurement of "prediction uncertainty" (Gabadinho, Ritschard, Studer, et al., 2010).

$$\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{h}(\boldsymbol{\pi}_i, \ \dots \boldsymbol{\pi}_s) = -\sum_{i=1}^{s} \boldsymbol{\pi}_i \log(\boldsymbol{\pi}_i) \qquad (\textbf{Eq. } 2.1)$$

Where x is the sequence, s is number of potential states and π_i is proportion of occurrences of the *i*th state in the considered sequence. The proportion of minutes allocated to each state over the course of the entire day and the number of distinct states that a person inhabits drive the value of Entropy. For this measure, the number of state changes and the contiguity of states do not matter. It simply uses the proportion of total time spent in each state, regardless of the number of different episodes that time is spread over.

If a person has no state changes during the entire day, for instance if they spend all day at home, their Entropy would be zero. In contrast, someone who moves around a lot will have "high" Entropy. The range of entropy values depends on the number of distinct states. Sequences with more unique states have higher potential maximum Entropy values, and Entropy is at its highest when people spend equal amounts of time in each state. In this study with five distinct states (Home, Travel, Work, School, and Other), the maximum Entropy is 1.61.

The second measure – Turbulence – is a bit more complicated than Entropy in terms of what it uses for its calculations.

$$\boldsymbol{T}(\boldsymbol{x}) = \log_2\left(\boldsymbol{\phi}(\boldsymbol{x})\frac{\boldsymbol{s}^2_{t,max}(\boldsymbol{x}) + \mathbf{1}}{\boldsymbol{s}^2_t(\boldsymbol{x}) + \mathbf{1}}\right) \qquad (\textbf{Eq. } 2.2)$$

- x represents the sequence of activities and travel episodes in one person's diary
- $\phi(x)$ is the number of distinct subsequences in sequence x
- t_i is duration in each distinct state and is used to compute the mean consecutive time and variance below (*i*=1, …, number of distinct episodes)
- s_t^2 is variance of the state-duration for the x sequence
- $s_{t,\,max}^2$ is the maximum value that the variance can take given the total duration of the sequence x

$$s_t, \mathbf{max} = (\boldsymbol{n} - \mathbf{1})\left(\mathbf{1} - \bar{\boldsymbol{t}}\right)^{\mathbf{2}} \qquad (\mathbf{Eq.}\ 2.3)$$

- n is length of distinct state sequence
- $\bar{t}$ is mean consecutive time spent in the distinct states

Turbulence uses the number of distinct subsequences in a given sequence and the number of consecutive time points spent in a given state (Elzinga & Liefbroer, 2007; Gabadinho, Ritschard, Studer, et al., 2010). Consider a person with a daily sequence H-T-W-T-H meaning the person was at home (H) in the morning, traveled (T) to work (W), and after work traveled (T) back home (H). This sequence would contain the following subsequences: an empty sequence; the full sequence itself; subsequences of the type T-W-T-H, W-T-H, and T-W-T; discontinuous subsequences like T-T-H (which skips the work activity); and single activities H, T, and W. Enumerating all these subsequences yields ($\phi(x) = 27$) possible combinations that respect the precedence of activities in the H-T-W-T-H sequence. For a given sequence of activities (x), the measure of Turbulence is a measure of variability in terms of distinct activities, the order of these activities, and the variance of the durations of these activities in a day. All this makes Turbulence a measure of schedule complexity.

Gabadinho et al (2010) define another indicator they call Complexity that is based on Entropy and the transitions within a sequence (s).

$$\boldsymbol{C}(\boldsymbol{s}) = \sqrt{\frac{\boldsymbol{nt}(\boldsymbol{s})}{(\boldsymbol{l}(\boldsymbol{s}) - \mathbf{1})} \frac{\boldsymbol{h}(\boldsymbol{s})}{\boldsymbol{h_{max}}}} \qquad (\mathbf{Eq.}\ 2.4)$$

This is a function of the Entropy and the number of transitions (nt(s) = l(s) -1) in a sequence s, normalized by the maximum theoretical entropy (h_{max}) and the length of the sequence (l(s)). This indicator will have a value between 0 and 1, with zero corresponding to Entropy zero and no transitions.

Table 2.1 provides a few examples of sequences with state duration for each activity, counts of subsequences, Entropy, Turbulence, and Complexity. Person 1 in Table 2.1 stays at home all day and has the sequence of (H, 1440), Entropy zero, and Turbulence 1. The number of distinct subsequences is 2 (i.e., the empty sequence and the sequence itself). Persons 2 and 3 have activity patterns with 5 episodes that are 2 activities at home, 2 trips, and one at work (Persons 2) or some other place (Person 3). Both persons have 27 subsequences, and they are all different in the Entropy of their sequences because the number of minutes allocated to each episode are different between them. Similarly, for the Turbulence, the variance across the durations of activities is different between the two subjects. The Complexity indicator combines the advantages of Entropy (variety of time use) and Turbulence (reflecting the possibility of many subsequences), but in a simpler form than Turbulence, replacing subsequences with the length of the sequence and the number of transitions. Persons 4 and 5 show how the number of subsequences increases dramatically when more activity episodes are added and how this is reflected in the three summary indicators used. These indicators will be used in the summary of findings.

Table 2.1. Examples of Sequences

	(Activity/Place, Duration in minutes)	Pattern	Number of Subsequences	Entropy h(x)	Turbulence T(x)	Complexity C(x)
Person 1	(H,1440)	H	2	0.000	1.00	0.00
Person 2	(H,830)-(T,10)-(W,320)-(T,10)-(H,270)	H-T-W-T-H	27	0.372	6.63	0.0322
Person 3	(H,255)-(T,45)-(O,120)-(T,30)-(H,990)	H-T-O-T-H	27	0.302	6.10	0.0290
Person 4	(H,600)-(T,15)-(O,60)-(T,10)-(O,20)-(T,10)-(H,725)	H-T-O-T-O-T-H	79	0.203	7.87	0.0291
Person 5	(H,485)-(T,5)-(W,169)-(T,2)-(H,10)-(T,14)-(O,70)-(T,10)-(O,25)-(T,15)-(H,125)-(T,15)-(W,321)-(T,14)-(H,160)	H-T-W-T-H-T-O-T-O-T-H-T-W-T-H	9632	0.641	16.01	0.0790

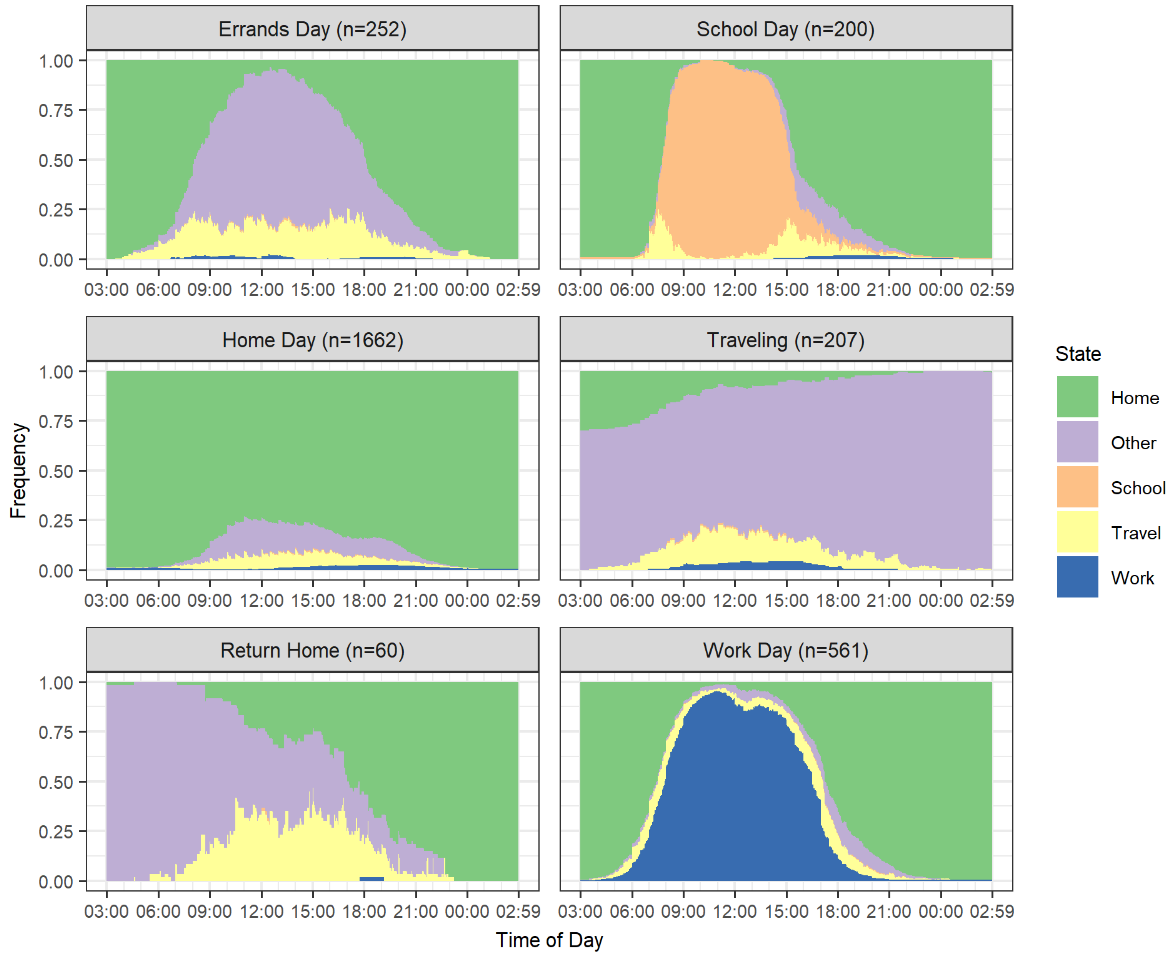
2.4 Small Study in San Louis Obispo and Santa Barbara

Our first objective is to find groups of sequences that are similar using a small sample as a test of the methods. For each of the 2,942 sequences, there is a series of 1,440 bins—one for each minute of the survey day starting at 3:00 AM and ending the next day at 2:59 AM. Each bin is colored by a letter (H, T, W, S, and O).

To identify similar sequences among the 2,942 person-sequences, there must be a rule of comparison. For example, different operations can be performed to reproduce one sequence departing from another and assign penalties to each operation (Wilson, 1998a, 1998b). Measuring the difference between two sequences depends on the number of operations and sum of penalties accumulated in the comparison. The operations applied to this comparison are replacement, insertion, and deletion (indel). In the sequence alignment literature, the measurement of dissimilarity and the number of operations needed to make two sequences the same is called a “distance.” The distance between two sequences is the minimum combination of replacements and indel (Abbott & Tsay, 2000). For ease of interpretation, in this chapter this will be called the *dissimilarity score* between two sequences. The output of an algorithm that does these operations among all the sequences is a matrix of dissimilarity scores. In the analysis here, this is a matrix of 2,942 by 2,942 (=8,655,364) cells containing the dissimilarity scores among sequences for each person in the sample.

This matrix can then be analyzed using clustering techniques to identify a small number of groups of sequences that represent similar time of day activities and travel patterns in the sample. To do this, the agglomerative nesting clustering method is used. Starting with 2,942 sequences, pairs of sequences are grouped based on their dissimilarity scores. Then, all the cluster dissimilarity score averages are compared to each other and clusters with smaller dissimilarity scores are lumped together. The process proceeds until all observations are in one cluster (Kaufman & Rousseeuw, 2009). This process can be thought of as a tree that starts with every sequence as an individual “leaf” and ends with one cluster as the “trunk.” After inspecting the overall time of day patterns, the six-cluster solution was selected because it shows clear representations of time-of-day time allocation patterns to places/activities (Figure 2.2).

Figure 2.2. Area charts: six-cluster solution for daily sequences of places and travel

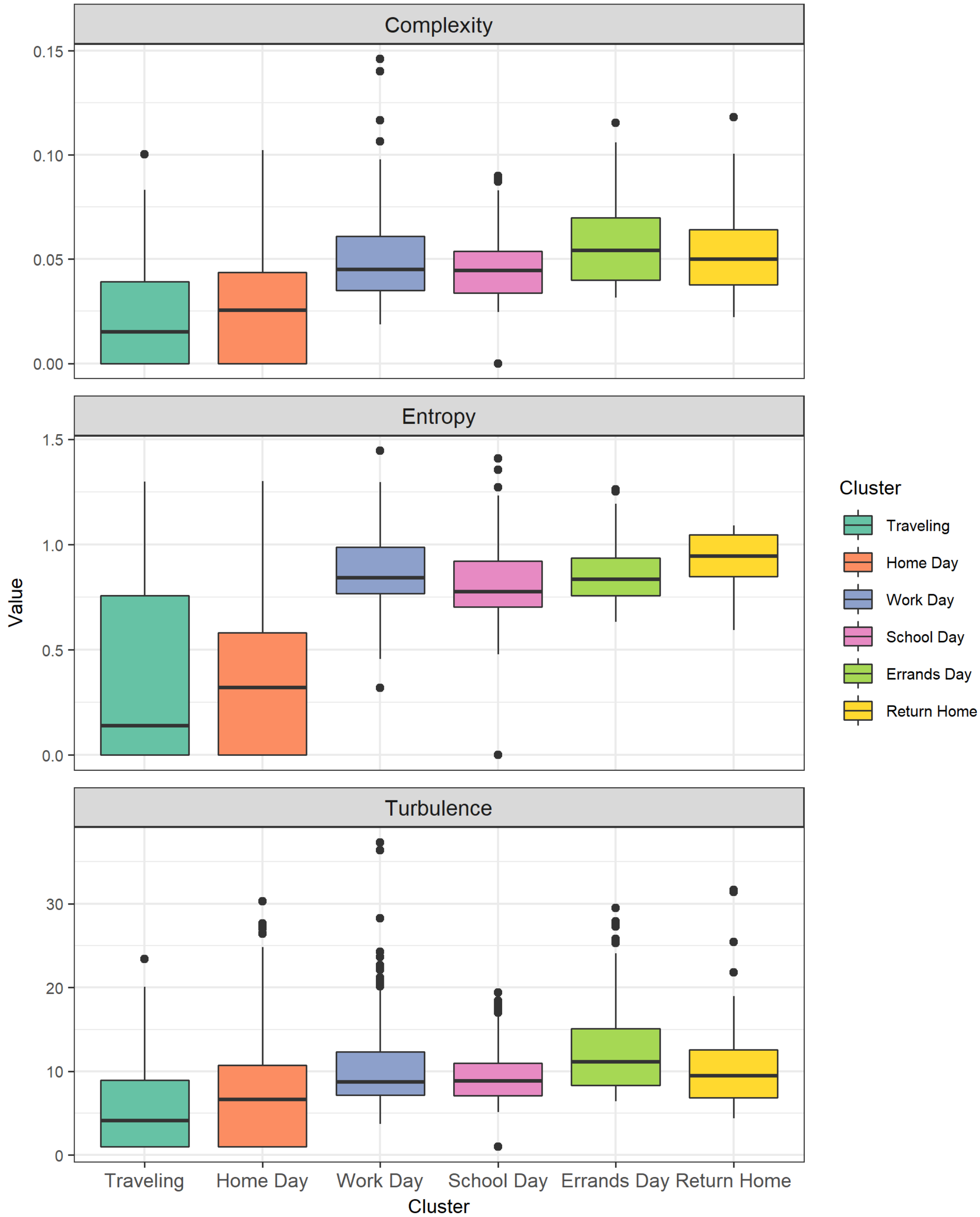


Note. These area charts show the percentage of each cluster's members who are in each state at each time point

The *Traveling* cluster in Figure 2.2 is the 207 persons that were outside the region during the survey or left during the day of the interview and travelled outside the region of SLO and SB. In the *Traveling* cluster, accessibility is set to zero and distance from home when missing is set to 350km. The *Home Day* cluster consists of 1,662 persons that spent most of their time at home with very short trips to places that are classified as Other. It has the smallest median kilometers from home at 5.1 km. This is indicative of the fact that people in the *Home Day* cluster do not venture far from home when they do leave. The *Work Day* cluster, with 561 persons, is the usual working day with travel before and after work and some visits to other places in the after-work hours. The *School Day* cluster is the usual school day for 200 survey participants with substantial travel before and after school and some work activities after school (presumably high school and college students). This cluster has a median kms from home of 7.6, which is the second smallest behind *Home Day* (5.1 km). This shows that students go to school relatively close to home. The *Errands Day* cluster, from 252 participants, shows a typical errands day, where respondents start at home, go to "Other" locations all day and return home in the evening. The *Errands Day* respondents travel throughout the day with no noticeable fluctuations by time of day, indicating that they move from destination to destination all day in many small trips. People in this cluster drive frequently, but around home. The *Return Home* cluster is a group of 60 respondents who are returning home from locations outside the region. They travel more than all other groups. Not all workers are in the *Work Day* cluster, indicating that people with irregular work timing and/or short durations at work locations are in every other type of cluster. In contrast, the place "School" does not appear to be a preeminent sequence state in any other cluster except *School Day*. Locations categorized as "Other" appear in all six patterns, pointing out the need to explore this category further. As these patterns only look at respondents' travel on a single day, it would be expected that depending on the day surveyed, respondents' cluster membership would vary. The composition of the clusters is further examined later with the multinomial logit model.

Figure 2.3 shows the distribution of Entropy, Turbulence, and Complexity values for each of the six clusters. The *Traveling* cluster has the lowest median values for all three indicators due to the way data are recorded for people that are out of town during the interview or leave from home on that day and go far. *Home Day* has the second lowest median and overall distribution for all three indicators among the clusters but shows substantial variation, meaning people in this cluster have a wide variety of sequences. The other four types have similar complexity and variety with *Work Day*. The *Work Day* cluster shows a few outliers in the values of Turbulence and Complexity (recall these two include sequence and subsequence lengths). Figure 2.3 demonstrates that sequences cannot be studied with only clusters or only indicators of Entropy, Complexity and Turbulence. Instead, sequences should be explored using multiple methods.

Figure 2.3. Box plots of within-cluster entropy, complexity, and turbulence



2.4.1 Correlation of Sequences with Person Characteristics

In a parallel analysis (McBride, Davis, & Goulias, 2019), use regression methods are used to study the correlation between Entropy and Turbulence for the entire 2,942 sample of people with social and demographic characteristics of this sample. In summary, people aged 25 to 65 have the most fragmented schedules (particularly as measured by Turbulence), especially when they have children over age 4 in the household. Escorting and joint participation in activities with children is a clear motivation for this. Significant differences among people of different incomes are found, with poverty inhibiting activity fragmentation for a person. Ethnicity and nativity also play a role in distinguishing among sequences. Hispanic people have sequences that are simpler than the US native group, but still more complex than other groups. Gender also emerges as a major covariate for Entropy, but not for Turbulence. People that live in urban and suburban environments, however, tend to have more fragmented schedules most likely due to the mixing of short and long activities in their schedule. Another major factor of fragmentation is the day of the week. Each day of the week appears to have a different composition of activities and durations. Sunday is the day with the least fragmented schedules and Friday shows the highest fragmentation. In addition, older children in the household motivate a more fragmented daily schedules of activities.

2.4.2 Multinomial Logit Model

To better understand the composition of each of these clusters, a multinomial logit model is estimated with categories for the six clusters for each person and identify variables that explain their cluster membership. Table 2.2 shows the results of this multinomial logit model.

Table 2.2. Multinomial logit of cluster membership

	Cluster Type				
	Home Day	Work Day	School Day	Errands Day	Return Home
Constant	1.938 *t = 8.888****	-15.706 *t = -127.915****	-2.681 *t = -6.231****	-0.234 *t = -0.804*	-2.123 *t = -4.596****
Respondent is Worker	-0.275 *t = -1.779**	17.520 *t = 142.692****	-0.910 *t = -2.560***	0.139 *t = 0.700*	1.076 *t = 3.188****
Respondent is Student	-0.844 *t = -3.659****	-2.760 *t = -4.313****	3.766 *t = 10.468****	-0.200 *t = -0.710*	0.514 *t = 1.261*
Respondent is Female	0.091 *t = 0.610*	-0.066 *t = -0.368*	-0.213 *t = -0.854*	-0.218 *t = -1.151*	0.609 *t = 1.984***
Number of Children Under 16 in Household	0.374 *t = 3.551****	0.498 *t = 4.079****	0.576 *t = 4.347****	0.516 *t = 4.395****	0.471 *t = 2.882****
Survey Day: Tuesday	0.232 *t = 0.814*	0.529 *t = 1.655**	0.845 *t = 2.009***	0.324 *t = 0.876*	-0.686 *t = -1.167*
Survey Day: Wednesday	0.353 *t = 1.300*	0.650 *t = 2.111***	1.365 *t = 3.371****	0.373 *t = 1.056*	-0.998 *t = -1.597*
Survey Day: Thursday	0.290 *t = 0.996*	0.592 *t = 1.795**	0.907 *t = 2.052***	0.347 *t = 0.917*	-2.158 *t = -2.011***
Survey Day: Friday	0.466 *t = 1.512*	0.637 *t = 1.858**	0.856 *t = 1.893**	0.924 *t = 2.450***	-0.913 *t = -1.301*
Survey Day: Saturday	-0.245 *t = -0.935*	-2.367 *t = -5.721****	-3.743 *t = -3.510****	-0.049 *t = -0.141*	-0.010 *t = -0.023*
Survey Day: Sunday	0.156 *t = 0.639*	-2.338 *t = -6.306****	-3.399 *t = -4.305****	0.043 *t = 0.132*	-0.251 *t = -0.583*
Akaike Inf. Crit.	5,634.98	5,634.98	5,634.98	5,634.98	5,634.98

Note: *p<0.1; **p<0.05; ***p<0.01
Loglikelihood at convergence = -2,762.488 (degrees of freedom = 55)
Loglikelihood with constants only = -3,818.690; McFadden pseudo R^2 = 0.277

The model tests which sequence cluster people belong to, given the independent variables in the model (worker, student, male, number of children under 16 in the household, and day of survey). All comparisons are made to a reference group that is excluded from the model. In this case, the reference group is the *Traveling* cluster. Across all clusters, workers are more likely to be in the *Work Day* cluster and the *Return Home* cluster. If a respondent is a student, they are more likely to be in the *School Day* cluster. Day of the week of the travel diary is also controlled for in this model and Monday is used as the reference day. CHTS respondents assigned to a Saturday or Sunday have low propensity to be in the *Work Day* and *School Day* clusters. The exact opposite is true for the weekdays. The number of children under 16 in the household has positive and significant coefficients for all five groups in Table 2.2 (also confirmed by the descriptive statistics of Table 2.3). This reflects the fact that the reference group (*Traveling*) has the lowest number of children in the household (Table 2.4). Within clusters, different coefficients are found for the days of the week. For the *Home Day* cluster, none of the coefficients of the days of the week are significantly different than zero, meaning *Home Day* as a pattern of sequences is spread almost uniformly throughout all days of the week for this sample. A similar trend is shown for *Errands Day*, except for Friday.

Table 2.3. Within-cluster sample characteristics of categorical variables

Variable	Traveling	Home Day	Work Day	School Day	Errands Day	Return Home	Overall
Number of females	108 52.17%	916 55.11%	266 47.42%	82 41.00%	117 46.43%	39 65.00%	1528 51.94%
Number of people with disabilities	12 5.80%	149 8.97%	16 2.85%	2 1.00%	11 4.37%	0 0.00%	190 6.46%
Num. of people in hhs with kids aged 00 to 03	16 7.73%	164 9.87%	47 8.38%	42 21.00%	29 11.51%	8 13.33%	306 10.40%
Num. of people in hhs with kids aged 04 to 15	38 18.36%	388 23.35%	125 22.28%	157 78.50%	83 32.94%	21 35.00%	812 27.60%
Num. of people in hhs with kids aged 16 to 18	36 17.39%	163 9.81%	55 9.80%	63 31.50%	31 12.30%	9 15.00%	357 12.13%
Number of students	33 15.94%	177 10.65%	3 0.53%	185 92.50%	47 18.65%	15 25.00%	460 15.64%
Number of weekend responders	81 39.13%	536 32.25%	26 4.63%	3 1.50%	81 32.14%	33 55.00%	760 25.83%
Number of workers	92 44.44%	636 38.27%	561 100.00%	16 8.00%	113 44.84%	38 63.33%	1456 49.49%
Household income at or below poverty line	13 6.28%	118 7.10%	25 4.46%	32 16.00%	26 10.32%	2 3.33%	216 7.30%

Table 2.4. Within-cluster sample characteristics of continuous and count variables

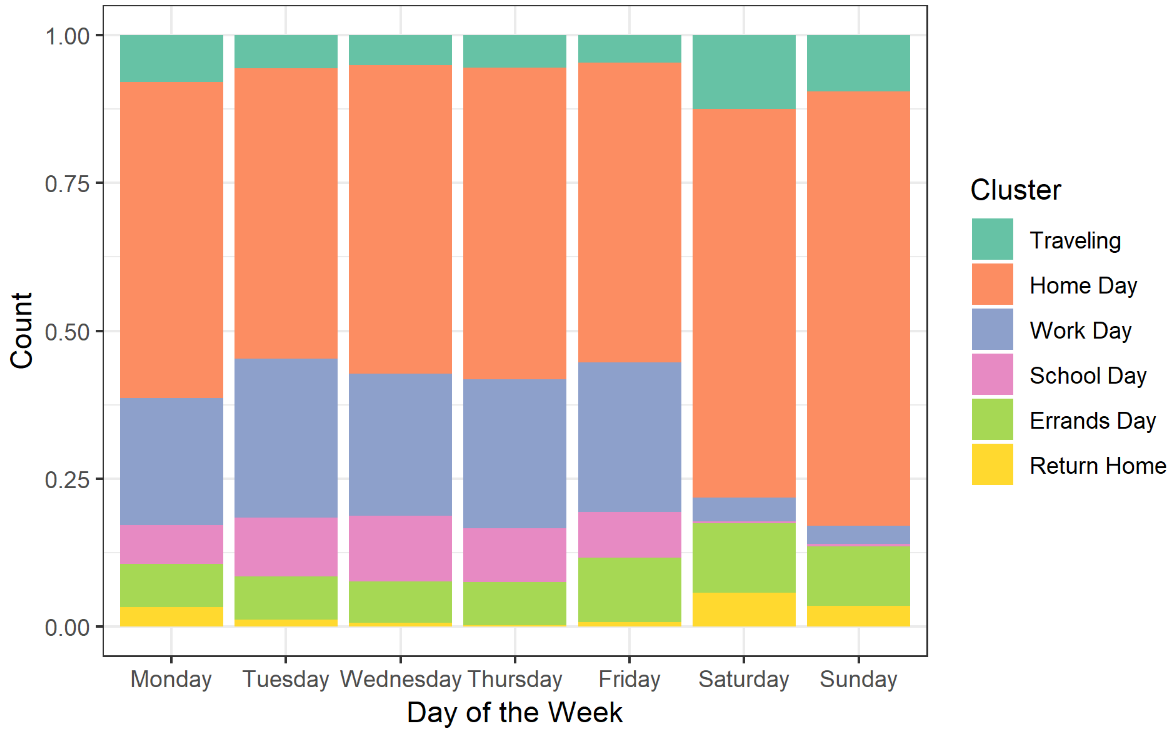
Variable	Cluster	Mean	Minimum	Median	Maximum	St. Deviation
Complexity (C(s), Eq. 2.4)	Errands Day	0.057	0.032	0.054	0.115	0.020
	Home Day	0.026	0.000	0.026	0.102	0.023
	Return Home	0.053	0.022	0.050	0.118	0.021
	School Day	0.046	0.000	0.045	0.090	0.015
	Traveling	0.021	0.000	0.015	0.100	0.024
	Work Day	0.050	0.019	0.045	0.146	0.018
	Total	**0.035**	**0.000**	**0.035**	**0.146**	**0.025**
Customer Service Establishments Near Homes	Errands Day	8.262	0.000	6.490	24.550	6.072
	Home Day	7.754	0.000	6.285	25.240	5.708
	Return Home	8.010	0.000	6.620	24.320	5.729
	School Day	8.504	0.000	7.225	25.240	6.235
	Traveling	2.495	0.000	0.000	24.440	5.193
	Work Day	8.256	0.000	6.500	25.240	6.125
	Total	**7.579**	**0.000**	**6.220**	**25.240**	**5.991**
Maximum Distance from Home (kms)	Errands Day	66.681	0.337	22.215	350.000	105.821
	Home Day	15.283	0.000	3.555	350.000	46.253
	Return Home	197.782	0.116	278.819	350.000	157.157
	School Day	10.972	0.000	5.417	350.000	26.990
	Traveling	302.914	6.915	350.000	350.000	113.791
	Work Day	28.360	0.297	14.592	350.000	52.360
	Total	**45.846**	**0.000**	**8.398**	**350.000**	**99.830**
Number of Children in Household Between 0 and 3	Errands Day	0.123	0.000	0.000	2.000	0.352
	Home Day	0.131	0.000	0.000	2.000	0.423
	Return Home	0.133	0.000	0.000	1.000	0.343
	School Day	0.250	0.000	0.000	2.000	0.519
	Traveling	0.092	0.000	0.000	2.000	0.336
	Work Day	0.102	0.000	0.000	2.000	0.357
	Total	**0.130**	**0.000**	**0.000**	**2.000**	**0.407**
Number of Children in Household Between 4 and 15	Errands Day	0.635	0.000	0.000	5.000	1.038
	Home Day	0.406	0.000	0.000	5.000	0.842
	Return Home	0.583	0.000	0.000	3.000	0.889
	School Day	1.445	0.000	1.000	5.000	1.078
	Traveling	0.242	0.000	0.000	4.000	0.599
	Work Day	0.348	0.000	0.000	4.000	0.731
	Total	**0.477**	**0.000**	**0.000**	**5.000**	**0.889**

Table 2.4 (continued) Within-cluster sample characteristics of continuous and count variables

Variable	Cluster	Mean	Minimum	Median	Maximum	St. Deviation
Number of Children in Household Between 16 and 18	Errands Day	0.143	0.000	0.000	2.000	0.403
	Home Day	0.113	0.000	0.000	3.000	0.365
	Return Home	0.150	0.000	0.000	1.000	0.360
	School Day	0.425	0.000	0.000	3.000	0.705
	Traveling	0.193	0.000	0.000	2.000	0.442
	Work Day	0.107	0.000	0.000	2.000	0.337
	Total	**0.142**	**0.000**	**0.000**	**3.000**	**0.408**
Number of Household Vehicles	Errands Day	2.079	0.000	2.000	6.000	0.983
	Home Day	2.089	0.000	2.000	8.000	1.108
	Return Home	2.333	1.000	2.000	5.000	0.877
	School Day	2.215	0.000	2.000	6.000	1.060
	Traveling	2.222	0.000	2.000	6.000	0.924
	Work Day	2.160	0.000	2.000	7.000	1.028
	Total	**2.125**	**0.000**	**2.000**	**8.000**	**1.064**

Figure 2.4 shows the observed relationship between Day of the Week and cluster types. From this analysis, no major gender differences are detected except for the Return Home cluster in which there are 39 females (65% of this group). Investigation into this revealed that the median of maximum distance traveled from home in the *Return Home* cluster is much lower for females than it is for non-females (80 km vs the imputed maximum of 350km). As it turns out, many more females start from a closer location to home than non-females. They stay in town and run errands before returning to their own homes.

Figure 2.4. Observed cluster membership by day of the week (DOW)



2.4.3 Linear Regression Models of Within-Cluster Complexity

The main objective of this chapter is to explore fragmentation of daily schedules. In this section, within-cluster complexity is analyzed to understand the relationship between fragmentation, person characteristics, household structure, accessibility, and distance travelled to reach places. To study the propensity of persons to fragment their daily activity-travel pattern, six regression models are estimated. For each of the six patterns described above, the complexity indicator (C(s) in Eq. 2.4) computed for each individual in the cluster group is used as the dependent variable. As explanatory variables, person and household characteristics are used. Table 2.5 shows all six regression models.

Table 2.5. By-cluster complexity linear models

	Dependent Variable is Complexity C(s) (Eq. 2.4)					
	Cluster Type					
	Traveling	Home Day	Work Day	School Day	Errands Day	Return Home
Constant	0.025	0.030	0.048	0.045	0.054	0.039
	t = 3.192***	*t = 13.473****	*t = 16.145****	*t = 7.741****	*t = 9.372****	*t = 2.803****
Disability	0.003	-0.006	0.001	-0.015	0.006	
	t = 0.443	*t = -2.741****	*t = 0.310*	*t = -1.281*	*t = 0.874*	
Household Income Near or Below the Poverty Line	-0.011	-0.011	-0.007	-0.009	-0.019	0.018
	t = -1.581	*t = -4.313****	*t = -1.833**	*t = -2.305***	*t = -3.291****	*t = 1.180*
Low to Medium Household Income	0.001	-0.006	-0.003	-0.002	-0.006	-0.004
	t = 0.113	*t = -3.796****	*t = -1.241*	*t = -0.581*	*t = -1.629*	*t = -0.417*
Medium to High Household Income	-0.006	-0.003	0.00001	-0.001	-0.006	0.011
	t = -1.552	*t = -2.141***	*t = 0.006*	*t = -0.532*	*t = -1.744**	*t = 1.651*
Weekend	0.002	-0.005	-0.008	0.002	0.001	0.003
	t = 0.672	*t = -3.894****	*t = -1.954**	*t = 0.263*	*t = 0.336*	*t = 0.490*
Number of Children Under 4 in Household	0.005	-0.004	-0.0003	-0.001	0.006	0.029
	t = 1.104	*t = -3.161****	*t = -0.113*	*t = -0.365*	*t = 1.439*	*t = 2.751****
Number of Children Aged 4 to 15 in Household	-0.007	0.003	0.002	0.001	0.002	0.0003
	t = -2.263**	*t = 3.995****	*t = 2.153***	*t = 0.760*	*t = 1.278*	*t = 0.069*
Number of Children Aged 16 to 18 in Household	-0.013	0.004	0.002	-0.0005	0.005	0.003
	t = -3.342***	*t = 2.798****	*t = 0.903*	*t = -0.268*	*t = 1.611*	*t = 0.327*
Female	-0.003	0.001	0.0004	-0.001	0.005	0.009
	t = -0.852	*t = 0.547*	*t = 0.240*	*t = -0.399*	*t = 1.922**	*t = 1.530*
Worker	0.006	0.003		0.009	0.003	-0.010
	t = 2.014**	*t = 2.691****		*t = 2.195***	*t = 1.189*	*t = -1.468*
Student	0.016	-0.002	0.041	0.006	-0.008	-0.014
	t = 3.118***	*t = -0.925*	*t = 3.869****	*t = 1.338*	*t = -2.113***	*t = -1.471*
Number of Household Vehicles	0.001	-0.002	-0.0004	-0.001	0.0004	-0.002
	t = 0.301	*t = -4.422****	*t = -0.494*	*t = -0.838*	*t = 0.249*	*t = -0.594*
Mean Customer Service Establishments within 10km of Home	0.002	0.0003	0.0003	-0.0001	0.0003	0.001
	t = 6.206***	*t = 2.940****	*t = 1.860**	*t = -0.737*	*t = 1.397*	*t = 0.931*
Maximum kilometers traveled from home	-0.00003	0.0001	0.00001	-0.00004	0.00001	0.00004
	t = -1.697*	*t = 11.809****	*t = 0.700*	*t = -1.078*	*t = 0.933*	*t = 1.931**
Observations	190#	1,543#	526#	195#	240#	58#
R^2	0.328	0.144	0.057	0.090	0.132	0.403
Adjusted R^2	0.274	0.137	0.033	0.019	0.078	0.226
Residual Std. Error	0.020 (df = 175)	0.022 (df = 1,528)	0.018 (df = 512)	0.015 (df = 180)	0.019 (df = 225)	0.019 (df = 44)
F Statistic	6.089*** (df = 14; 175)	18.411*** (df = 14; 1,528)	2.385*** (df = 13; 512)	1.266 (df = 14; 180)	2.440*** (df = 14; 225)	2.283** (df = 13; 44)

Notes: # sample sizes are lower than the cluster membership due to missing values for some of the explanatory variables used here

*p<0.1; **p<0.05; ***p<0.01

2.4.3.1 Traveling Cluster

The number of people in a respondent's household between 4 and 15 years old has a significant negative effect on Complexity. More children in the age range of 4 to 15 corresponds to lower Complexity for members of the *Traveling* cluster. More household members in the 16 to 18 age range also significantly corresponds to lower Complexity in the *Traveling* cluster. The number of children between ages 0 and 3 is not a significant determinant of Complexity in this cluster because there are very few persons in this group with children in this age group. Respondents in the *Traveling* cluster that are employed have higher Complexity. Presumably they are traveling for work and combining their work trip with other activities. A similar effect is found for students.

Accessibility around home measured by average customer serving establishments within 10 km from each home location is highly statistically significant. Considering that 25% of this group started their day at home, it is possible they use access to opportunities around their homes before departing for a long-distance trip. The variable of maximum kilometers a destination was from home captures the interaction between choices in space and timing of trips. Longer distances from home correspond to lower Complexity scores, indicating an inhibiting effect on Complexity when people have longer travel time to far away destinations. However, this analysis does not distinguish among the locations included in *Other,* and this may mask other factors.

2.4.3.2 Home Day Cluster

It is important to note that the *Home Day* cluster does include some movement. People in the *Home Day* cluster are those who mostly spent the day at home, with little other activity. However, a few persons did not stay home all day. The majority of respondents with disabilities are in the *Home Day* cluster (149 of 190). Respondents in the *Home Day* cluster who have disabilities have significantly lower Complexities than those without disabilities. The linear model seems shows that not only do disabled respondents belong mostly to the group that did not leave the house much, but also, they tend to travel less than non-disabled people within the cluster. Disability appears to limit how people travel and therefore has a significant impact on their access to opportunities. People in the poverty group belong to households with an income near or below the poverty line. For members of the *Home Day* cluster, being in the lowest income group has a strong significant negative correlation with Complexity as compared to those that are above the poverty line. This shows how poverty can have a limiting effect on opportunities and movement.

Home Day cluster members who did their travel diary on a weekend day had significantly lower Complexity than those who did it on a weekday. This indicates that, within the *Home Day* cluster, on a weekend day people tend to go to fewer places than they do on weekdays. Having children under 4 in the household significantly reduces Complexity for the *Home Day* cluster. Unsurprisingly, a baby or toddler is a strong anchor to the home. Having more children between 4 and 15 years old has a significant positive effect on Complexity for those in the *Home Day* group. Caretakers who are not working on the diary day are probably responsible for most of the errands or transportation of children. Even if they are spending most of the day at home, there are more likely to be at places they need to go since they have children that need escorting. Having children in a household between 16 and 18 also has a significant positive effect on Complexity for *Home Day* cluster members. Even though in this age group teenagers can legally obtain a driver's license, this does not mean they drive without their parents when they turn 16. Teens may be more independent, but they are not completely self-reliant. The outcome is parents with higher Complexity in their schedules to serve children with a need to be at different places in a day. Workers in the *Home Day* cluster are likely enjoying a day off. Their Complexity is higher than cluster members who are not workers. On days off, workers will have more errands to run that they could not complete on a work day.

In this group, a higher number of vehicles corresponds to significantly lower Complexity. If there are more vehicles available in a household, individuals can go where they want without needing to combine trips with other people in the household and therefore using simpler activity-travel patterns. In contrast, more vehicles in a household also correspond with higher income that increases complexity, and income is controlled for with the presence of the poverty measure in the model. Higher numbers of customer service establishments around the home correspond to higher Complexity in the *Home Day* cluster. People with more establishments around their homes can still have "Home Days," where they spend most of their time at home while still participating in shorter trips. They can probably run errands closer to home more easily. Opposite of the previous group this group of people trade-off time and distance differently. Traveling farther from home means higher Complexity for *Home Day* cluster members. There were lots of people who did not leave their homes at all on their Home Day, but for those that did, the farther they went from home the more fragmented their day was. The respondents whose maximum distance is farther away from home are likely running errands that day, but not enough errands to place them in the *Errands Day* cluster. However, they will still have more complex schedules than those who stay at home or close to their home.

2.4.3.3 Work Day Cluster

For people in the *Work Day* cluster, being below or near the poverty line has a negative effect on Complexity. Respondents with household income below/near the poverty line who are working on the diary day have less fragmented schedules than those earning more money. This could be for several reasons. Respondents might not take a single occupancy vehicle to work, so they are not as mobile. They might not have the monetary freedom to participate in opportunities. It could also be that the type of job they hold does not allow for as much schedule flexibility as the higher-income positions.

On weekend days, people in the *Work Day* cluster tend to have lower Complexity. This could be similar to the effect discussed in the *Home Day* section – that people do not do as much on weekend days. It could also have to do with the types of jobs that are available over weekends. Very few office or manual labor positions are worked on weekends. The jobs would likely be in the service industry. These types of jobs might be more limited in the way they can be performed – they need to be happening in the same place all the time – which reduces opportunities for fragmentation. On work days, Complexity only significantly corresponds to having children aged 4 to 15 in the household. A higher number of children in this age range corresponds to an increase in Complexity. Children in this age range need to be driven around to appointments, activities, etc. They also need to be picked up and dropped off from school. If there are more children in a household, the adult worker will escort the children to various places. A student who is in the *Work Day* cluster is a student who also has a job, and these respondents who do both are statistically significantly more likely to have higher complexity than non-students.

Work Day cluster members tend to have corresponding higher Complexity with the customer service establishment numbers around their home. They have more access to establishments around their homes, so they can more easily access opportunities without having to travel far.

2.4.3.4 School Day Cluster

Our search for significant indicators for this pattern did not yield as many variables as for the other patterns. This is mainly due to the membership in this group that are children and teenagers that are students. However, being at/around the poverty line lowers Complexity, even for students on a school day. This is indicative of the problems in equity for students in poverty. K-12 students in poverty may not have access to the same numbers of after-school activities as their peers. This lack of access to opportunities at an early age sets young people up for fewer opportunities later. *School Day* cluster members who are also workers have higher Complexity than those who are not workers. This has already been addressed in the *Work Day* cluster discussion: students who are also workers have more complex days because they have more places to be and things to do than someone who is just a worker or just a student.

2.4.3.5 Errands Day Cluster

For people in the *Errands Day* cluster, having household income around the poverty line means they have lower Complexity. As in other clusters, this is indicative of the impact of poverty on the mobility of people and access to opportunities. *Errands Day* cluster members in households with higher numbers of children in the 4 to 15 age range have statistically higher Complexity. Adults in this cluster likely spend a lot of their day transporting children, running errands, and performing other general tasks required to manage a household. As will be discussed later, children in this age range are often accompanying parents on errands if it is not a school day. Females in *Errands Day* cluster members have significantly higher Complexity than non-females (males in this sample). This result is in line with past research that has shown that women tend to take on more of the household tasks and is consistent with the household responsibility hypothesis (Crane, 2007; Turner & Niemeier, 1997), even if both partners work full-time. A worker in the *Errands Day* cluster has higher Complexity than a non-worker in the cluster. This was discussed earlier with the *Home Day* cluster: on a worker's day off, they might need to run the errands that they do not have time for on work days. This would result in more fragmentation on their day off. Students in the *Errands Day* cluster have lower Complexity than non-students. This could be because some cluster members are children who come along for errands with their parents, but not for all of them. Younger students might be participating in other activities that are not errands where they do not travel as much (e.g., playing at a friend's house). This student category also includes university students. For that group, irregular schedules allow for them to spread their errands out more, so in a single free day they would not have as much fragmentation as non-student adults. For the *Errands Day* cluster, higher maximum distance traveled from home means higher Complexity. Destinations with more opportunities, like a mall, might be farther away from the home, and while a person is there they go to several different destinations.

2.4.3.6 Return Home Cluster

This group is made up of 65% females, who also show higher Complexity than their counterparts males. The median for maximum kilometers for females and non-females in the *Return Home* differs greatly: for females it is 80.8 km, while it is 350 km (the maximum imputed value) for non-females. The means were also different: 174 for females and 241 for non-females.

So, the *Return Home* cluster in fact contains two separate groups of people: people who travel long distances, and people who stayed around town but did not start at home (and went to places marked “Other” during the day). Women primarily make up the latter group. Presumably, women returning home from a trip far from home when they arrive at home are also running local errands. This may be an indication of the multiple roles played by women in the household responsibility hypothesis discussed above. In this group, having children under 4 in the household increases respondents’ Complexity significantly. Since the *Return Home* cluster consists of people’s travel days, this shows that traveling with young children leads to more fragmentation and therefore complexity in patterns (however, very few persons have children under 4 years old in this group). Workers who are in the *Return Home* cluster have lower Complexity than non-workers.

All together, these results show that clustering of daily patterns using this type of data and then studying the fragmentation characteristics of the cluster members leads to important behavioral conclusions about task allocation and the correlation between fragmentation and income, accessibility, and the trade-offs with distances travelled. In essence, there is heterogeneity in correlation across and within clusters. This has not been analyzed in this depth before.

2.5 Statewide Analysis

This section reports the findings from sequences that are based on a random sample of 12,704 persons in 5,000 households. The analysis is limited to 12,704 persons because the computation of differences in the *dissimilarity score* between pairs of sequences requires a very large matrix that exceeds our local computational facilities. The analysis here uses a matrix that is 12,704 by 12,704 (=161,391,616) cells, containing the dissimilarity scores among sequences for each person in this sample. This matrix is then analyzed using the agglomerative nesting clustering method. The analysis starts with 12,704 sequences and group pairs of sequences based on their dissimilarity scores. Then, all the cluster dissimilarity score averages are compared to each other and lump together clusters with smaller dissimilarity cores. This proceeds until all observations are in one cluster (Kaufman & Rousseeuw, 2009). This process can be thought of as a tree that starts with every sequence as an individual "leaf" and ends with one cluster as the "trunk." After inspecting the overall time of day patterns, select the nine-cluster solution is selected because it shows clear representations of time-of-day time allocation patterns to places/activities. Below, the words "cluster" and "daily pattern" are used interchangeably. Figure 2.5 shows these nine distinct patterns. The cluster names are based on the daily behavior each cluster represents. For each of these nine patterns, the membership is studied in terms of sociodemographic characteristics of respondents using descriptive statistics (Table 2.6 and Table 2.7). These patterns are also expected to be correlated with the day of the week assigned to each respondent, and Table 2.8 shows this correlation.

Figure 2.5. Nine clusters of daily sequences of places and travel statewide

Table 2.6. Cluster/pattern membership in daily patterns by person characteristics

Daily Pattern	Females	Workers	Students	Disabled	Weekend
Home Day	50.55%	36.04%	15.75%	10.72%	36.49%
School Day	46.96%	2.60%	92.96%	2.51%	2.80%
Typical Work Day	45.05%	99.55%	1.78%	1.83%	7.71%
Errands Type 1	49.73%	46.29%	19.17%	5.24%	44.85%
Mostly Out of Home	49.17%	50.09%	18.37%	5.01%	31.54%
Errands Type 2	43.58%	52.20%	14.70%	7.60%	29.73%
Non-typical Work Day	32.97%	100.00%	9.19%	2.70%	19.46%
Leave Home	47.60%	48.50%	22.75%	6.89%	33.23%
Traveling	50.27%	45.60%	32.69%	4.67%	27.20%

Table 2.7. Cluster/pattern membership in daily patterns by age group of respondents

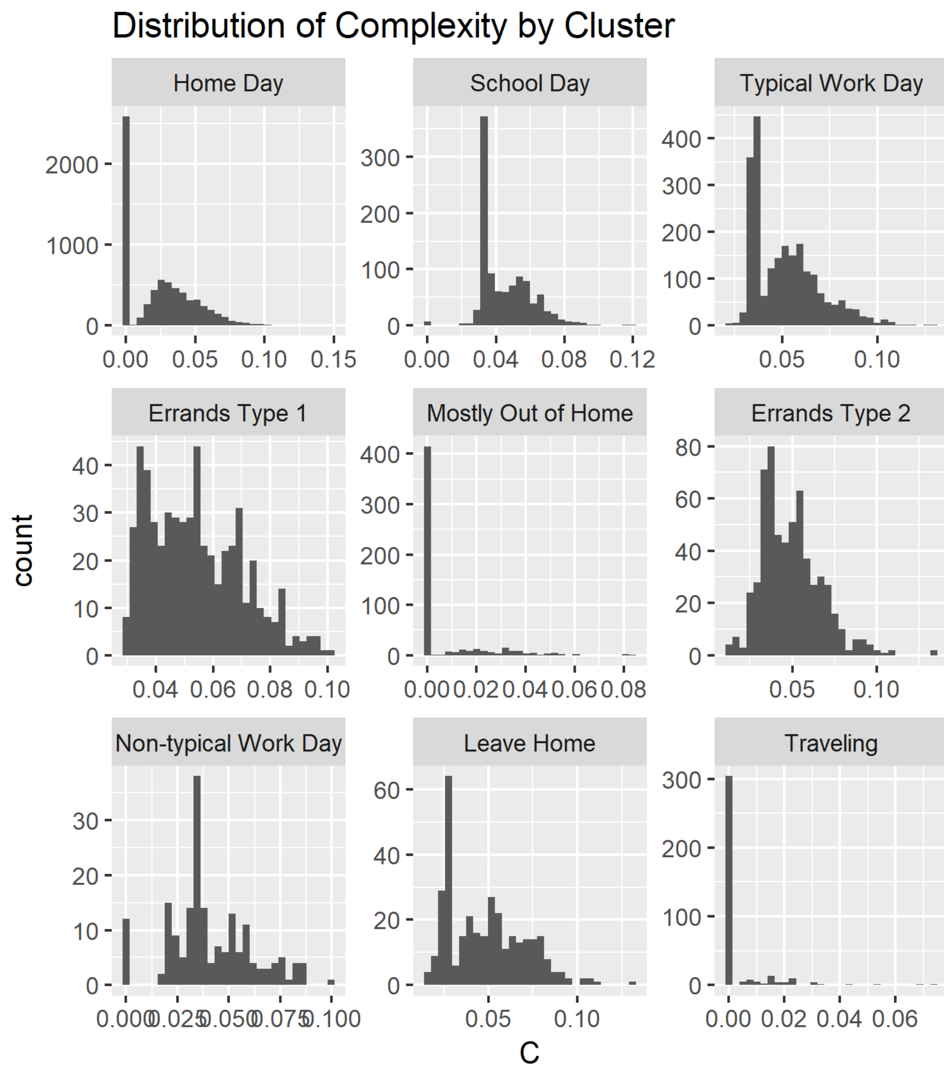
Daily Pattern	Age 00-03	Age 04-15	Age 16-18	Age 19-24	Age 25-34	Age 35-44	Age 45-54	Age 55-65	Age 65+	Did not Tell
Home Day	257	766	212	305	463	654	943	1412	1614	230
School Day	35	740	173	32	19	7	5	2	2	22
Typical Work Day	0	0	9	102	272	408	611	618	141	83
Errands Type 1	15	70	19	29	31	64	87	121	93	24
Mostly Out of Home	16	57	21	31	41	53	80	115	95	30
Errands Type 2	24	68	13	25	37	71	120	136	83	15
Non-typical Work Day	0	0	10	33	26	38	32	34	6	6
Leave Home	6	47	20	32	20	35	56	59	52	7
Traveling	16	87	21	14	31	42	50	52	33	18
Total	369	1835	498	603	940	1372	1984	2549	2119	435

Table 2.8. Cluster/pattern membership in daily patterns by day of the week

Daily Pattern	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday	Total
Home Day	13.19%	12.57%	13.14%	12.43%	12.18%	17.97%	18.52%	100.00%
School Day	16.49%	19.67%	17.36%	22.37%	21.31%	1.45%	1.35%	100.00%
Typical Work Day	16.22%	20.45%	21.30%	19.39%	14.93%	4.28%	3.43%	100.00%
Errands Type 1	11.39%	10.67%	9.76%	10.67%	12.66%	24.95%	19.89%	100.00%
Mostly Out of Home	10.58%	14.47%	13.54%	15.03%	14.84%	14.47%	17.07%	100.00%
Errands Type 2	16.39%	13.85%	11.99%	14.36%	13.68%	10.81%	18.92%	100.00%
Non-typical Work Day	17.30%	14.05%	15.68%	15.14%	18.38%	14.05%	5.41%	100.00%
Leave Home	11.38%	12.28%	12.87%	14.37%	15.87%	19.46%	13.77%	100.00%
Traveling	11.26%	14.56%	15.93%	19.23%	11.81%	17.03%	10.16%	100.00%

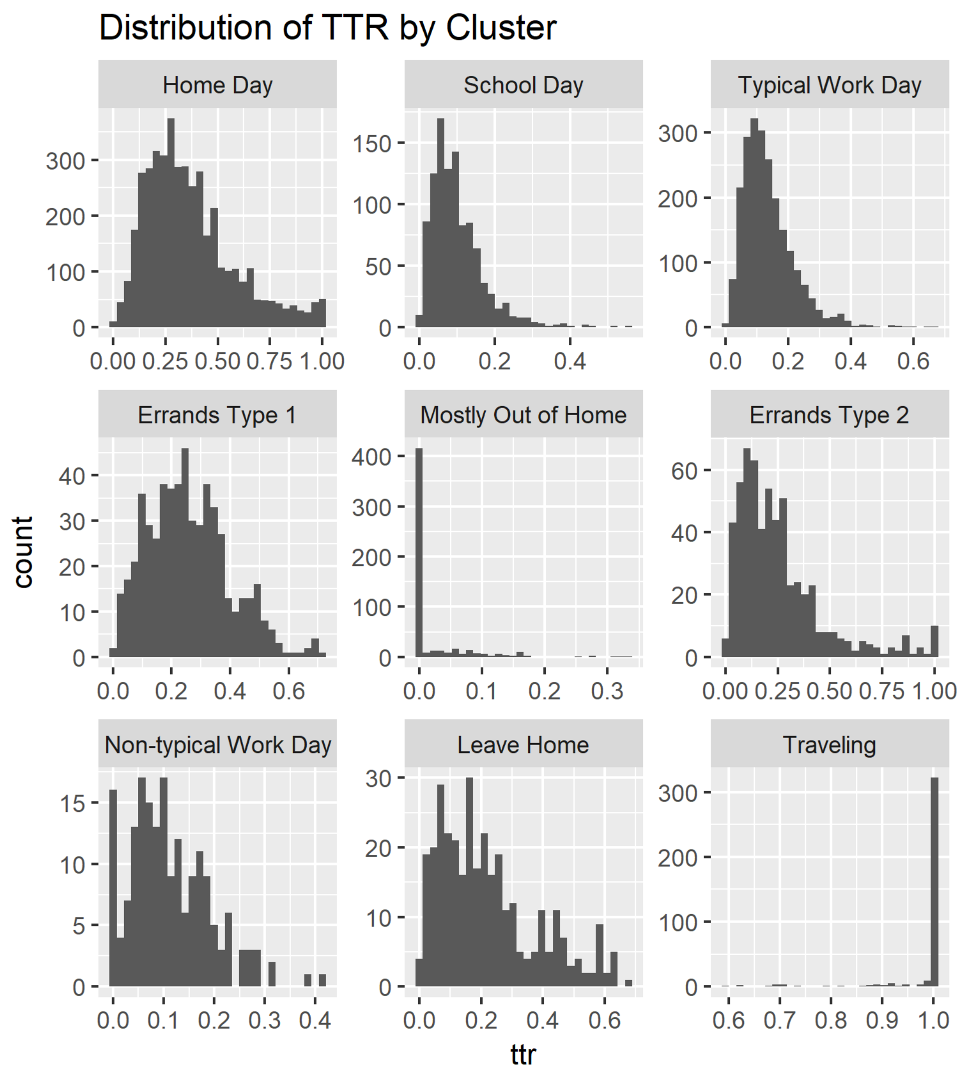
One of the key objectives in this chapter is to explore place-activity-travel fragmentation. As shown in the analysis using the Central Coast data, the indicator named Complexity (C(s) in Equation 2.4) is sufficient as an indicator of sequence fragmentation. Figure 2.6 shows the histograms of the C(s) values for each of the nine daily patterns.

Figure 2.6. Complexity C(s) histograms by daily pattern



This analysis also uses the travel time ratio (Dijst and Vidakovic, 2000, Schwanen and Dijst, M., 2002, Susilo and Dijst, 2009 and 2010, Milakis et al., 2015, Dharmowijoyo et al., 2016, Milakis and van Wee, 2018). Travel time ratio (TTR) is a compact indicator to represent trade-offs of people between travel and activity time. In this chapter, TTR is defined as the total travel time in a day divided by the sum of the total time outside the home plus the total travel time in a day. This allows for studying the percent of time for travel over the total length of Hägerstrand's time-space prism, which is the time elapsed between the first departure from home and arrival at home at the end of the day. This is modified to fit patterns with no home stay and use total amount of time that is not home in the denominator of TTR. Figure 2.7 shows the histograms of the TTR for each of the nine daily patterns derived here.

Figure 2.7. Travel Time Ratio (TTR) by daily pattern



Similar to Table 2.5 for the Central Coast, linear regression models are estimated for the complexity index (Table B.1) and TTR (Table B.2) for each of the nine patterns to explain within-cluster variation in fragmentation and TTR. Table 2.9 contains descriptive statistics of the Complexity and TTR used as dependent variables in the models of **Error! Reference source not found.**.

Table 2.10 shows the number of trips in each cluster and the modal splits (the ratio of travel modes people use) within each cluster. The total number of trips for this sample is 41,175, corresponding to 3.24 trips per person per day.

2.5.1 Home Day Cluster

The most populous cluster is the *Home Day*, with 54% (6,856 persons) of the sample selecting this pattern. These persons spent most of their time at home and a few of them (4,265) travel to other places. This is also the second most popular pattern for weekends (Table 2.6). Notable is the slightly more than half of the persons in this cluster being women, and 11% disabled persons (compared to 7% of total sample being disabled), reflecting movement restrictions for this group. Figure 2.6 shows the composition of this cluster clearly, with a substantial number having zero complexity because they stayed at home all day. This contributes to the average complexity (Table 2.9) being the third lowest among the nine patterns. This pattern also has the second highest travel time ratio indicating that for persons who left home, 36.8% of the time spent outside home was for travel. The average number of trips per person day in this pattern is 2.76 (lower than the overall average) and based on Table 2.10, 83.41% of these trips are by private motorized means with 37.38% driving alone.

In Table B.1, the regression models for complexity (Table B.1) show that the presence of children in the age group 4 to 15 years old contributes positively to higher fragmentation. In contrast, living in the suburbs, exurbs, or rural areas is an inhibitor to fragmentation when compared to the center of the city that has higher density of opportunities. Females also have higher fragmentation than males in this pattern.

The travel time ratio regression (**Error! Reference source not found.**) shows that f rom among the persons that have out of home activities in this cluster, rural residents have a 0.06 higher ratio than center city dwellers. Exactly the opposite happens when the respondent is a child below 15 years old, and students have 5% lower TTR that non-students. **Error! R eference source not found.** also shows the impact of children of any age in the household making the TTR between 0.02 and 0.04 higher than persons with no children in the household.

2.5.2 Typical Work Day Cluster

The second most populous pattern is the *Typical Work Day* pattern (2,244 people representing 17.7% of the sample) that shows usual morning and afternoon peaking of work with a noon break for lunch. As expected, 99.55% of respondents in this pattern are employed persons (Table 2.6), and no children display this pattern (Table 2.7). Weekdays make up most of the days of the week this pattern occurs on, with small percentages on Saturday and Sunday (Table 2.8). This pattern has the second highest fragmentation (Table 2.9), and 13.6% of the out of home time is travel time. **Error! Reference source not found.** shows the presence of children in the age groups 4 to 15 and 16 to 18 years old are correlated with higher fragmentation, but higher car ownership is negatively correlated with fragmentation. This shows the impact of decreased constraints for persons in households with more cars. Senior residents and persons in the high levels of poverty are also more likely to fragment their place-travel less than other groups. The presence of children in the age group 4 to 15 is positively correlated with the travel time ratio, presumably because children need rides to different places, and this increases the amount of time spent traveling for household members who need to provide those rides. Being a student is also positively correlated with TTR, indicating the need for students to travel to work and other non-work activities and therefore having a higher TTR by 0.04 than non-students on a typical work day. People in this cluster make on average 4.24 trips per day, of which 61.44% are driving alone, second only to the non-typical work day discussed later.

Table 2.9. By-cluster complexity and travel time ratio (TTR)

	Mean C(s)	Std. Dev C(s)	Mean TTR	Std. Dev TTR
Home Day	0.024	0.024	0.368	0.213
School Day	0.045	0.015	0.100	0.070
Typical Work Day	0.052	0.017	0.136	0.081
Errands Type 1	0.054	0.016	0.257	0.140
Mostly Out of Home	0.007	0.014	0.019	0.048
Errands Type 2	0.049	0.018	0.253	0.208
Non-typical Work Day	0.041	0.020	0.113	0.080
Leave Home	0.049	0.022	0.220	0.159
Traveling	0.003	0.009	0.984	0.060

Table 2.10. By-cluster number of trips and modal split

	Vehicle Driving Alone Ratio	Vehicle Driving Others Ratio	Vehicle Passenger Ratio	Other Motorized Ratio	Total	Trips per Person
Home Day	37.38%	21.87%	23.45%	0.71%	83.41%	2.76
School Day	7.26%	2.44%	61.00%	2.22%	72.91%	3.52
Typical Work Day	61.44%	13.39%	6.26%	1.83%	82.92%	4.24
Errands Type 1	26.83%	24.69%	33.28%	0.72%	85.52%	5.24
Mostly Out of Home	20.67%	29.92%	30.51%	4.53%	85.63%	1.52
Errands Type 2	39.33%	17.17%	23.84%	1.34%	81.68%	4.41
Non-typical Work Day	64.59%	7.94%	9.33%	1.21%	83.07%	3.13
Leave Home	30.83%	22.28%	31.88%	1.95%	86.95%	3.99
Traveling	27.48%	17.90%	31.04%	2.46%	78.88%	3.24

	Bike Ratio	Walk Ratio	Transit Ratio	Other Non-Motorized Ratio	Other Ratio	Total
Home Day	1.63%	12.11%	2.36%	0.50%	0.00%	16.59%
School Day	2.33%	16.43%	8.03%	0.30%	0.00%	27.09%
Typical Work Day	2.07%	11.05%	3.88%	0.04%	0.03%	17.08%
Errands Type 1	1.07%	9.52%	3.83%	0.07%	0.00%	14.48%
Mostly Out of Home	1.38%	11.22%	0.59%	0.39%	0.79%	14.37%
Errands Type 2	1.15%	11.46%	4.91%	0.15%	0.65%	18.32%
Non-typical Work Day	2.42%	10.71%	3.80%	0.00%	0.00%	16.93%
Leave Home	1.20%	7.80%	2.93%	0.00%	1.13%	13.05%
Traveling	2.97%	13.06%	3.82%	0.17%	1.10%	21.12%

2.5.3 School Day Cluster

The third most populous pattern is the *School Day* (1,037 persons and 8.2% of the sample). Table 2.6 shows that 92.56% of the persons in this cluster are persons classified in the survey as full-time students. Table 2.7 shows the majority of these persons are age 4 to 15 (740 persons) and age 16 to 18 (173 persons) and this pattern is typical of weekdays (Table 2.8) with a very small percentage on Saturday (1.45%) and Sunday (1.35%). Figure 2.6

shows that there are two groups of people in this cluster: (a) a group that has the same low complexity in their schedule and (b) another with high variety. **Error! Reference source not found.** shows that the presence of children in ages 4 to 15 and 16 to 18 increases the need to fragment schedules (presumably, children accompanying other children of the household in different activity locations). The travel time ratio, however, is only 10%, and for children in the age group below 15 years old is even lower. This indicates that both the school location and the other activity locations are most likely in close proximity. Typical of the persons in this cluster is their modal split with 61.00% riding cars as passengers, 8.03% as transit passengers (the highest among all clusters), and 16.43% walking (also the highest among all clusters).

2.5.4 Errands Day Clusters

The next two daily patterns are by persons that visit places classified as *other,* and the are named *Errands Type 1* (553 persons and 4.4% of the sample) and *Errands Type 2* (592 persons and 4.7% of the sample). Both have patterns reaching a peak of visiting places other than home, work, or school, and they both have substantial amounts of traveling. The major difference between the two clusters is the time of day the peak is reached. Both patterns have more men than women. *Errands Type 1* is also the preferred pattern for weekends (44.85% in Table 2.6), with Saturday getting almost a quarter of the persons in this pattern (Table 2.8). Both daily patterns have high fragmentation and high within-cluster average TTR of 25.7% and 25.3% (Table 2.9). The values of fragmentation for both clusters are spread substantially (Figure 2.6). The same is true for their TTR (Figure 2.7). The *Errands Type 1* daily pattern has the highest average number of trips per person (5.24 in Table 2.10), and an almost even spread in the use of private cars, but still 85.52% of trips are made by private motorized modes. In contrast, *Errands Type 2* has a much lower number of trips per person (4.41 in Table 2.10), a higher driving alone ratio (39.33% in Table 2.10), but still high private car modal split.

For *Errands Type 1*, complexity is lower when this pattern happens on weekends, by senior residents, when the household has children 16 to 18 years old, and by students (**Error! Reference source not found.**). In contrast, fragmentation is higher for households t hat have children aged 4 to 15 and live in the exurbs. For *Errands Type 2* there is lower fragmentation in weekends, higher fragmentation for households that have children age 4 to 15 and higher fragmentation when the respondent is female or a worker (**Error! Reference s ource not found.**). The TTR is higher in *Errands Type 1* for households that have children age 4 to 15, and substantially higher for residents in exurbs and rural areas when compared to their counterparts in suburbs and center of a city (**Error! Reference source not found.**). The *E rrands Type 2* TTR variation is not explained by any of the variables tested (**Error! Reference source not found.**).

2.5.5 Mostly Out of Home Cluster

The next daily pattern is *Mostly Out of Home*, with 539 persons (4.2% of the sample). This reflects the definition of places as *other,* which includes second homes, hotels, camping grounds, etc. that could not be assigned as the primary home location. It is also the third most popular pattern for weekends (31.54% in Table 2.6). This pattern has the second lowest complexity and lowest TTR (Table 2.9). **Error! Reference source not found.**. and **Error! Re ference source not found.** show the inhibiting role of very young children in fragmentation for this pattern and the even lower TTR for senior residents in this pattern. This cluster has the lowest number of trips among all patterns and the highest driving others percentage of trips (Table 2.10).

2.5.6 Non-Typical Work Day Cluster

The second working day pattern is the *Non-Typical Work Day*, and it is the least populous with 185 respondents (1.5% of the sample). This is an interesting daily pattern because it is entirely made up of workers that show starting and ending times of work that span the entire 24-hour interview period. It is more populated by men (about 67% in Table 2.6), spread throughout the age groups over 15 years old, and as shown in Table 2.8, a substantial portion of the cluster responded on Fridays and Saturdays (unlike the other typical work day daily pattern that has very few people on Saturday). This pattern shows substantial fragmentation, but low TTR (Table 2.9), presumably due to workers living close to the workplace and/or spending longer hours at work. Differences in fragmentation within this pattern are only due to the presence of very young children in the household and females having higher average fragmentation than males. The TTR ratio is lower for suburban and exurban residents when compared to the center city and rural dwellers. Higher car ownership also decreases the TTR. This pattern shows lower than average number of trips per person at 3.13 trips and is the pattern with the most driving alone trips at 64.59%.

2.5.7 Leave Home Cluster

The next pattern, *Leave Home* (334 persons and 2.6% of the sample), is characteristic of persons that stayed at a location classified as *other* with substantial traveling. Workers make up 48.5% of this group (in fact this pattern contains some activity at workplaces), and students make up 22.75%. This pattern also shows substantial fragmentation and substantial TTR, reflecting the higher likelihood of pattern members traveling far from home. Figure 2.6

shows that there are most likely two groups of people in this pattern: a) a group that leaves home and does not do a lot where they arrive; and b) a group that participates in multiple visits to places. Figure 2.7 shows a wide spread of TTRs within this group. The regression models in **Error! Reference source not found.** show that disability, poverty, and residing in t he exurbs inhibit fragmentation (**Error! Reference source not found.**). The TTR ratio regression in **Error! Reference source not found.** show substantial differences due to the p resence of children 4 to 18 years old increasing TTR by 6%, exurban and rural living increasing TTR by 6% and 9% respectively. In contrast, students have a TTR 11% lower than non-students. Table 2.10 shows this pattern has higher than average number of trips per day and a substantial portion of them by car as a driver or passenger.

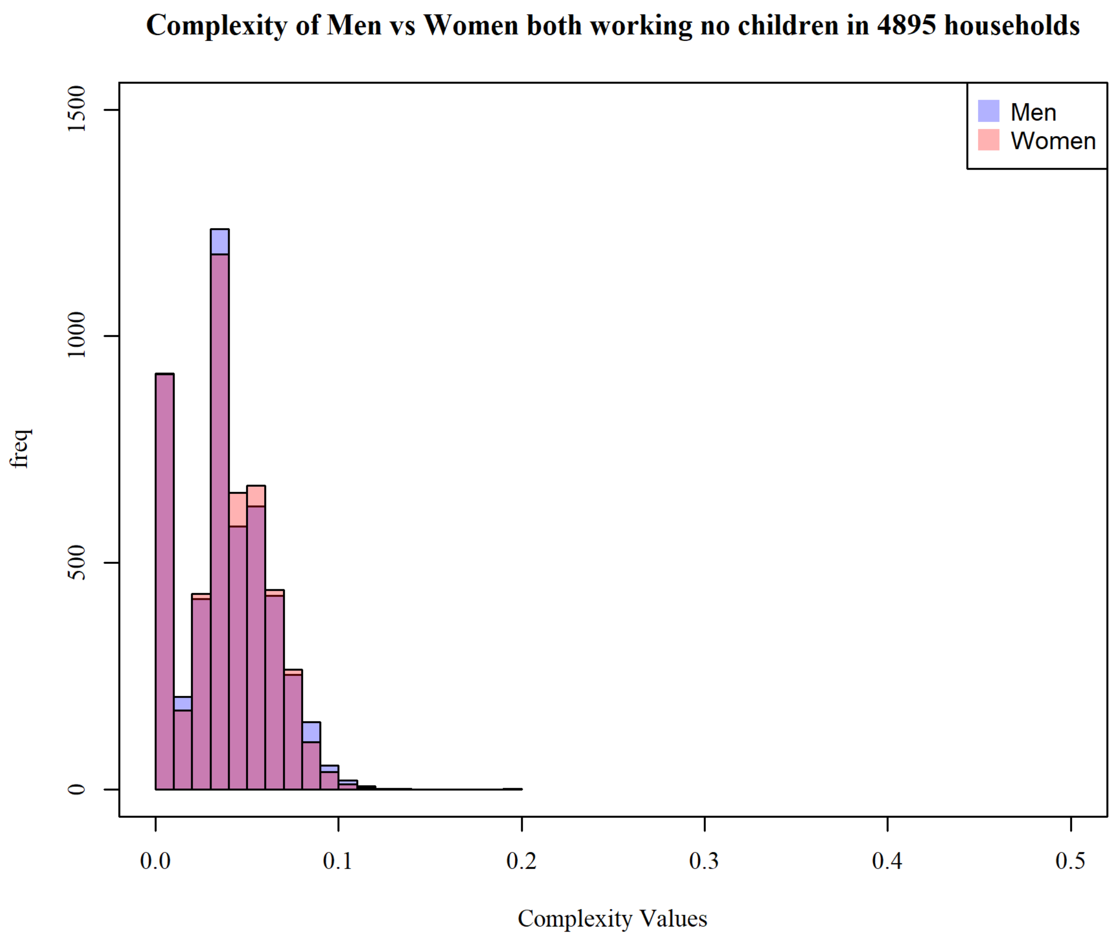
2.5.8 Traveling Cluster

This is a pattern characterized by mostly travelers of all ages and 50.27% females (364 persons and 2.9% of the sample). The pattern is spread throughout the days of the week with the highest percentage on Thursday and lowest on Sunday (Table 2.8). Reflective of this pattern is the lowest fragmentation of 0.003 and highest TTR 98.4% (Table 2.9 and histogram of Figure 2.6 and Figure 2.7). The only variable that increases complexity is if the respondent is a child younger than 15 years old. There is a decrease for persons in households with children 4 to 15 years old. TTR for this pattern is higher on the weekend days and for persons in households with children younger than 15 years old. In contrast, people in poverty and child respondents younger than 15 years old have lower TTR. The number of trips in this cluster is exactly at the overall average number of trips with the highest proportion of trips as passengers in a car and the highest bike share (2.97% in Table 2.10).

2.6 Fragmentation within Households

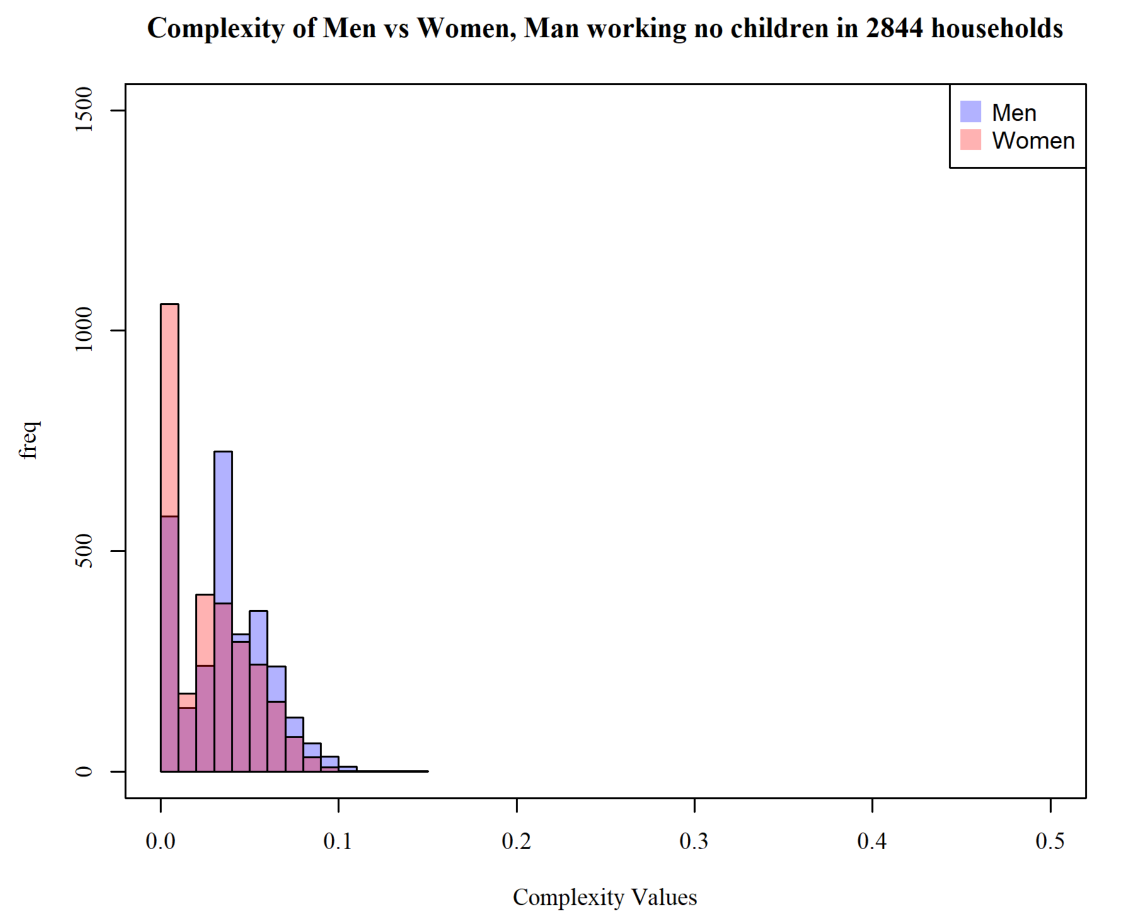
This section explores fragmentation of daily schedules of households that are made by an adult man and an adult woman with and without children. Although in previous sections it was found that men and women have different time of day activities and travel for some patterns, it is necessary to verify if men and women within the same households have different fragmentation in their daily patterns. This will show if the women are different in their daily schedule than men in the same household when they are employed and when they have children. The full CHTS with observations that were complete enough to build place-travel sequences was used. In the sample, there is a total 114,639 persons in 45,362 households. Of these households, 4,895 are adult couples with both spouses working, 2,844 adult couples in which only the man is a worker, and 2,142 adult couples in which only the woman works. Figure 2.8, Figure 2.9, and Figure 2.10 show the three histograms of men versus women for couples with no children. These figures and values of the C(s) demonstrate that working men and women both have high schedule complexity, and couples in which the man works but not the woman, the woman has lower schedule complexity (fragmentation) on average. Exactly the opposite happens when the woman works and the man does (Figure 2.10), displaying a reversal of roles (at least in terms of fragmentation).

Figure 2.8. Couples with both working



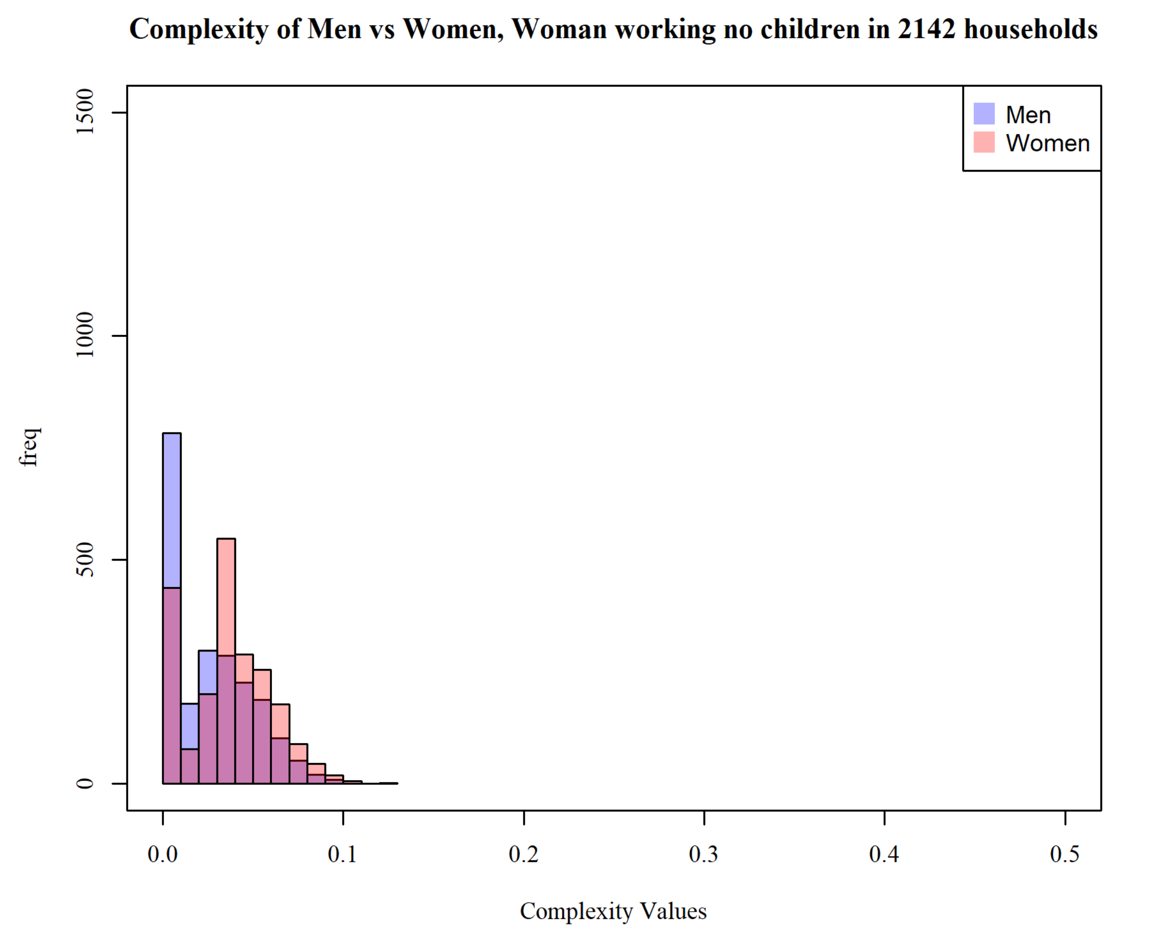
Average C(s) for men (standard deviation) = 0.04 (0.02)
Average C(s) for women (standard deviation) = 0.04 (0.02)
Note. The purple in this histogram is the overlapping area between men and women

Figure 2.9. Couples with only the man working



Average C(s) for men (standard deviation) = 0.04 (0.02)
Average C(s) for women (standard deviation) = 0.03 (0.02)
Note. The purple in this histogram is the overlapping area between men and women

Figure 2.10. Couples with only the woman working



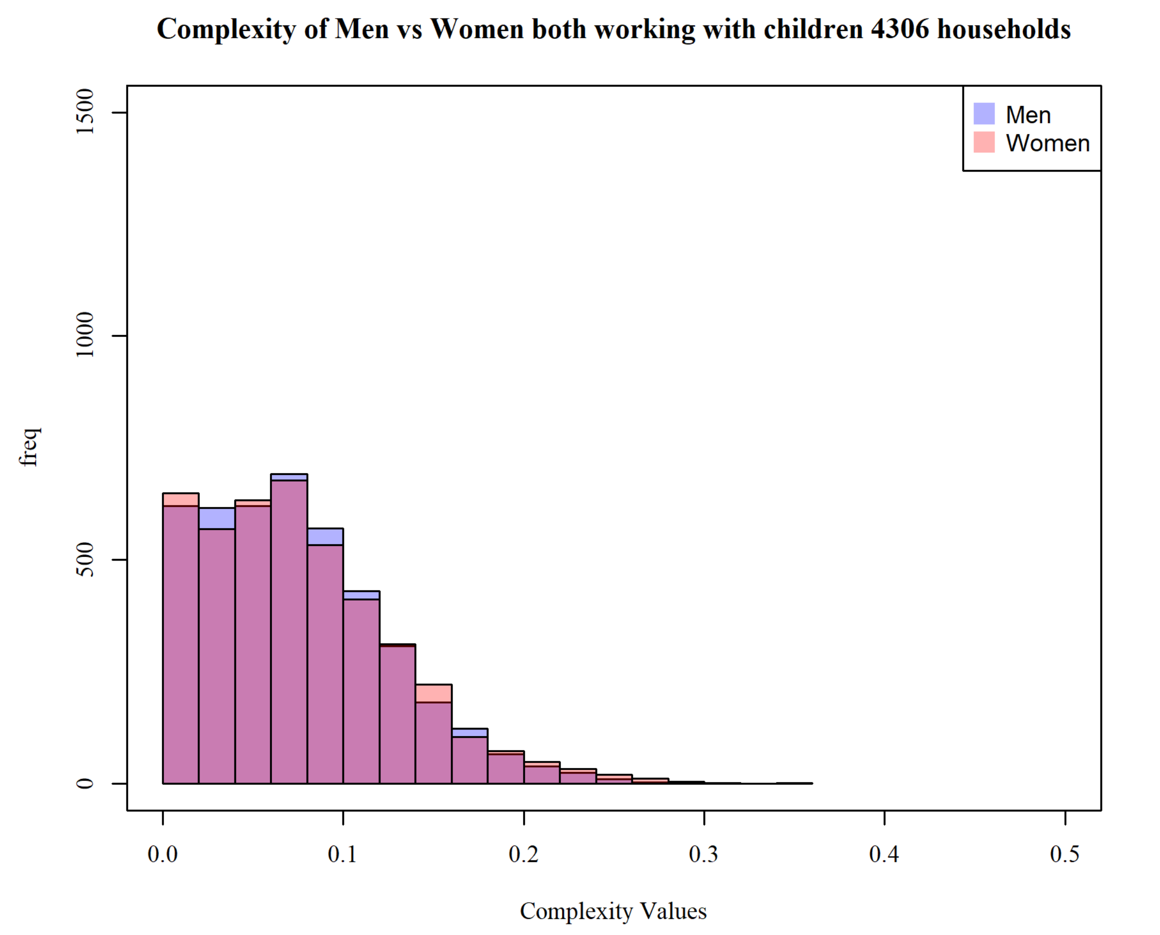
Average C(s) for men (standard deviation) = 0.03 (0.02)
Average C(s) for women (standard deviation) = 0.04 (0.02)
Note. The purple in this histogram is the overlapping area between men and women

The following discussion is of households consisting of different-sex couples with children. Histograms of the families are separated based on whether both adults work (Figure 2.11), only the adult man works (Figure 2.12) and only the adult woman works (Figure 2.13).

Fragmentation of schedules is by far higher for families in which both men and women work (Figure 2.11), and by far more variable (high standard deviation in addition to a wider spread of the histogram). However, in this case, women have on average higher values of C(s) than men. This is an indication of household responsibility hypothesis conforming the findings in the previous sections.

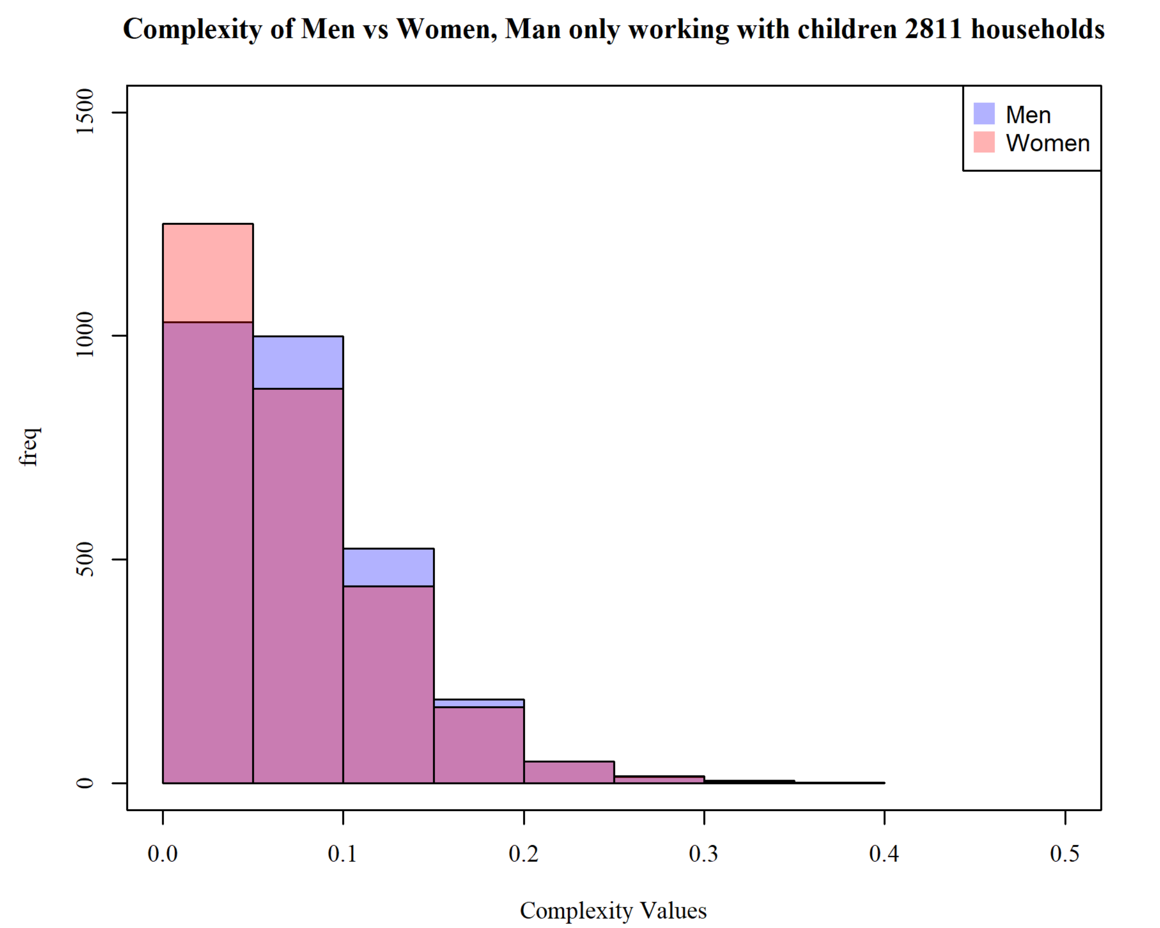
Figure 2.12 shows C(s) for couples with children in which only the man works and Figure 2.13 shows the C(s) for couples with children in which only the woman works. Unlike the couples without children, this time there is no reversal in fragmentation, with women having consistently high fragmentation and variability of this fragmentation, and often higher than men independently of their employment status. This result further strengthens the household responsibility hypothesis, and the role children play in motivating schedule fragmentation.

Figure 2.11. Adult couple with children, both adults work



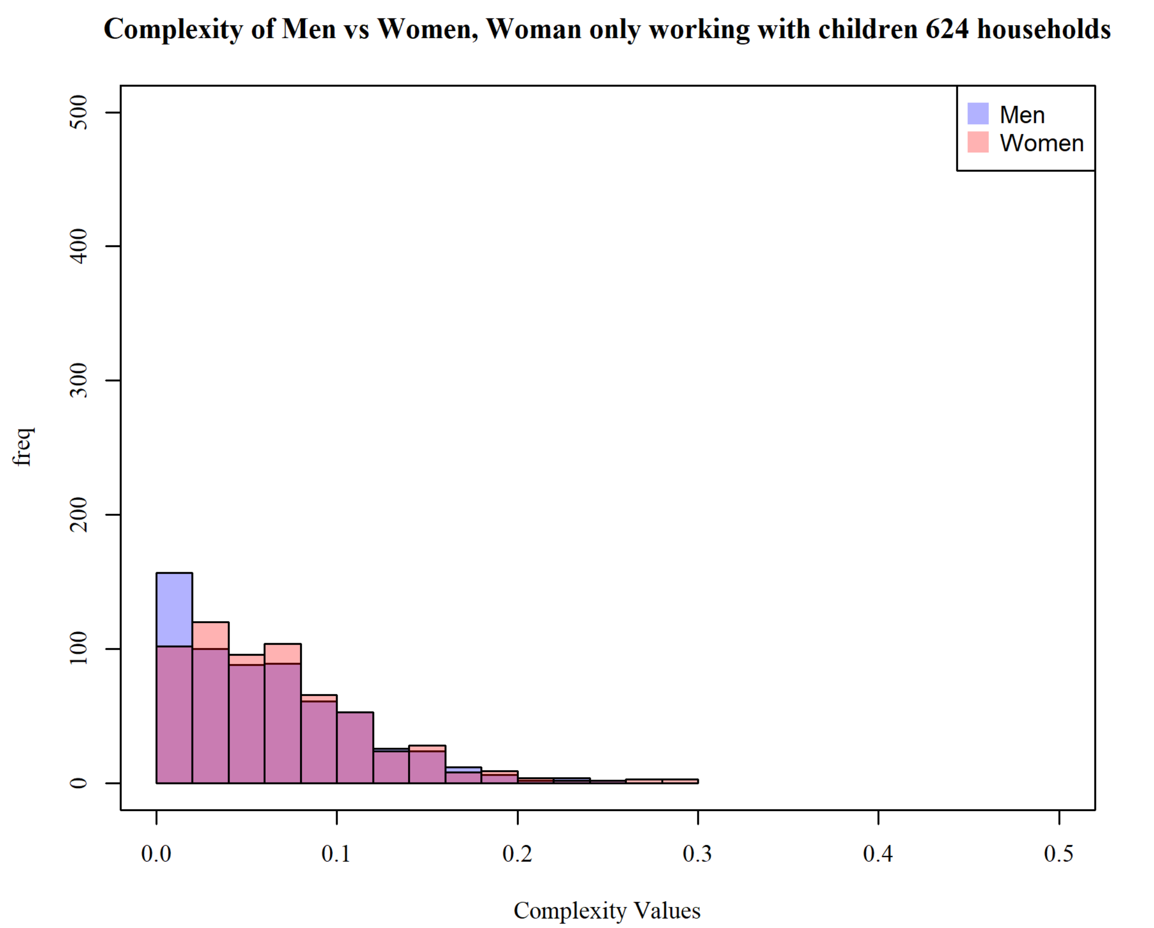
Average C(s) for men (standard deviation) = 0.07 (0.05)
Average C(s) for women (standard deviation) = 0.08 (0.05)
Note. The purple in this histogram is the overlapping area between men and women

Figure 2.12. Adult couple with children, only man works



Average C(s) for men (standard deviation) = 0.07 (0.05)
Average C(s) for women (standard deviation) = 0.07 (0.06)
Note. The purple in this histogram is the overlapping area between men and women

Figure 2.13. Adult couple with children, only woman works



Average C(s) for men (standard deviation) = 0.06 (0.05)
Average C(s) for women (standard deviation) = 0.07 (0.06)
Note. The purple in this histogram is the overlapping area between men and women

2.7 Summary of Findings and Travel Behavior Research

This chapter uses a new type of fragmentation in activity-travel behavior. Place-travel sequences are focused on first, and then activities at each place are reported on. The analysis is first done on data from SLO-SB to test the methods and then expanded using a random sample statewide. The findings in the section using data from SLO and SB are encouraging because they confirm findings using other methods to study activities and their durations, and these new techniques provide new insights about scheduling activities. One key finding is the two clusters of daily activities (*Traveling* and *Return Home*) that are, in essence, absent from contemporary activity-based models. These are not visitors to special travel generators such as events and hotels. These are residents of the study area that either were absent from the study area throughout the period of observation, left the area for a long-distance trip, or came back to the study area from a long-distance trip. This has implications for synthetic population generation. When synthetic population is used to generate the entirety of residents of a study region, it is necessary to to account for the proportion of this population that will have activity and travel behaviors of *Traveling* and *Return Home* clusters. This analysis finds that employment and education status are key determinants of daily schedules. It also shows the number of children at different ages play different roles within each of the six clusters of sequences used here. Overall, however, the presence of children in the household increases the complexity of place-travel daily patterns. Poverty emerges as an important determinant of daily patterns and requires further scrutiny, together with car ownership, car availability, and public transportation services. In addition, the analysis here confirms the household responsibility hypothesis, which is that there are differences between men and women in terms of household responsibilities.

Using a larger sample of 12,704 persons from 5,000 households spread throughout California yields nine distinct daily patterns. These include patterns of people staying at home for long periods in a day, people who follow typical daily working schedules, and people who follow typical school schedules. People that travel for an entire day and people that stay at home in the morning but then travel for the rest of their interview day were found. Two patterns of running errands with different time of day rhythms were also revealed. The ninth pattern is by people that spent most of their time in a day at locations that are not home, work, or school, and travel for very short time. Each pattern also has different memberships in terms of gender, age, and day of the week (in addition to the working and/or student status, as expected).

We also make comparisons between men and women that live in the same household and find that in couples with no children, employment status influences fragmentation of activities in such a way that the employed person has a more fragmented schedule. Men and women that are not employed but their partners appear to have similar fragmentation values. Upon examining adult couples with children, women have consistently more fragmented schedules than men in the same household regardless of employment status. All this further strengthens the household responsibility hypothesis for women who, in addition to work outside the home, also run a variety of errands, and for this reason need to visit multiple places in a day.

From a land use and transportation viewpoint, if more people moved to dense urban environments and adopted similar lifestyles to those observed in this data, they would be more likely to have more fragmented schedules during the *Home Day*, but no major differences for all the other patterns. The added flexibility of Mobility as a Service (MaaS) integrating different services may better serve the higher fragmentation patterns found here (i.e., *Typical Work Day, Errands Type 1, and Errands Type 2*). But, to do this, MaaS will need to become a suitable alternative to and compete successfully with the private car that offers the flexibility to give rides to other people.

3 A Pandemic Survey to Test Key Attitude-Behavior Components

3.1 Introduction

The original aim of this project was to survey residents of a specific area of Los Angeles about their daily travel before and after the opening of the LA Metro Crenshaw rail line. However, COVID-19 and the delay of completion of the rail line by the public transportation agency caused a need to modify our original data collection plan and proposed analysis. Throughout the period of performance of this project, there was continuous uncertainty about how long the virus would disrupt normal operations, so the plan adapted over time, as explained below.

First adaptation: In May 2020, to prepare for the potential of business establishments remaining closed through the duration of this project timeline, we designed a survey and collected data about the effects of COVID-19 on the life of people in the study area emphasizing commuting and going to school. We developed a repeated cross-sectional data collection plan. The first cross-section (Wave 1) would be completed with funding from this project, and the second with UC Santa Barbara funding after the end of this project in approximately May 2022 (Wave 2). Wave 1 data was collected asking people to reflect on their travel behavior before the pandemic, and to report on their current travel behavior during the pandemic. During this time, a detailed theoretical framework was developed for the data collection and analysis according to the original plan. This theoretical framework still guides the final version of the project plan, although the entire conceptual model cannot be tested with the available data.

Second adaptation: Create a smartphone application to collect travel diary data after the Crenshaw line opening. The construction of the Crenshaw line was delayed again due to the virus. The smartphone application was not completed (as explained in the next paragraph), but the process and progress made is described in Appendix C.

Third adaptation: The Crenshaw line was further delayed. It became clear that the pandemic was not going to subside in time to collect travel diary data that would truly be comparable to data collected after the completion of the rail line because the virus disrupted normal travel patterns so severely. All of Los Angeles was still under restrictions. There would be no chance of collecting “after” data to follow up on the COVID survey during the project time, since restrictions remained in place until beyond the end of the funding for this project. For this reason, development of the smartphone application was indefinitely suspended, and it was decided for this project to conduct detailed analysis of the data collected in May 2020 and illustrate the data analysis methods.

As the world reeled from the effects of the COVID-19 pandemic, anyone who could work from home was doing so. In the wake of the crisis, employers set up systems to allow for easier telecommuting, including video chat software, secure remote computer access, and communication tools like Slack. Universities and K-12 schools set up systems to do all teaching remotely. This all raises questions about how these new systems are going to be utilized during and after the crisis recovery period. Will things just go back to business-as-usual, or will people continue to use these systems? Did people see benefits from working from home (e.g., not having to sit in daily commuter traffic, more time with family members, flexible work schedule, healthier work-life balance, etc.)? Moreover, would those benefits outweigh the disadvantages enough that people would continue to telecommute in the future? Although there are People may find that some work activities they have been doing face-to-face actually work better remotely. Companies and institutions may also change their internal procedures to allow telework, mobile work, and flexible work schedules. For example, surgeons may find that in many cases it is not necessary to ask recovering post-operation patients to drive to hospitals for a one-hour meeting. Instead, they offer medical support using telemedicine meetings and decide during those if an in-person meeting would be beneficial. Health care suppliers are already moving to telehealth to enable major changes in the medical system with implications for travel.

The Wave 1 survey conducted examines changes in the following: modes of travel used, employment status, "essential worker" designation, job changes, typical work/school commute travel time and distance, non-commuting travel behavior, perceptions, attitudes, household structure, socio-economic status, and other traits. This includes a retrospective *before* section asking respondents to reflect on their behavior before the pandemic. The *during* section asks about travel behavior of people amid the pandemic, whether people's employment/school status changed, and if they work(ed) from home or took classes from home.

The research process has been documented in this GitHub repository[1]. The data collected in this project are available for other researchers to use on GitHub (see also the Data section in this chapter).

3.2 Survey Design and Data Collection

Responses to this survey were collected from residents of the Greater Los Angeles Metropolitan area in May 2020. SurveyMonkey's proprietary panel was used to recruit 1,002 respondents for this survey. This survey asks respondents about their work, school, and travel behavior before and during the COVID-19 restrictions. There are also a few questions about

[1] https://github.com/e-mcbride/covid19.commuting

people's predictions for how their work behavior might change once restrictions are lifted. For employment, they are asked about employment status, number of workdays, how often they work from home, and how often they participate in online meetings. For schooling, they are asked whether they are in school, what level of school they are in, and how their schooling has been affected by COVID-19 (e.g., have classes been cancelled or moved online, etc.). For travel behavior questions, if respondents work and/or go to school, they are asked what travel modes they use to commute, then they are asked to estimate the distance and time from their homes to work/school by each mode. Everyone, regardless of work/school status, is asked what means of transportation they use for all their trips. They are then asked how many trips they estimate making in a typical week by each of those modes. Respondents were asked to provide the city they live in, ZIP code, household income, gender, and age.

A unique aspect of this survey is the questions about whether people moved because of COVID-19. Many people changed home locations during this pandemic, either temporarily or permanently. Some moved in with friends or family to get their social needs met, others were obligated to move to take care of family members, or to separate households when some members might be exposed to COVID-19. As shown in Table 3.1, about half of respondents who moved plan to move back to their previous residence, while the other half have permanently moved out of that residence.

Table 3.1 Responses from People who Moved

Characteristic	N = 54
Have you moved permanently to the residence you are/were residing in during the Stay-at-Home order?	
No - but I am not going back to my previous residence	7
No - I plan to move back to my previous residence once the order is lifted	23
Yes - I have permanently moved to this residence	24
What influenced your decision to change residences?	
Assist family or friends	14
Protect family or friends	14
Social needs	14
Comfort, access to resources	13
Eviction	8
Necessity (was already moving)	11

3.2.1 Data Cleaning

The data collected required extensive cleaning. All the cleaning was done in R, some using the package “janitor” to identify duplicates, and some using criteria based on the feasibility of travel behaviors. The data started with 1,002 cases, and by the time all cleaning was complete, 202 cases were removed, so the final analysis contained 800 respondents. For this reason, we do not recommend the SurveyMonkey proprietary paid panel for data collection in the future unless proper quality assurance procedures are created, and warranties are provided by the vendor to increase data quality.

Many people who responded multiple times to the survey, giving the exact same answers to almost all the questions, including key questions that made it clear they could only be coming from the same person. Presumably, this is because the panel is paid per survey completed, and some people abuse the system to respond multiple times. It is assumed that SurveyMonkey would have security measures in place to prevent this, but it appears that people are able to get around them. In total, 34 duplicate records were flagged in the data for removal. The survey was only of people aged 18 and older, which was a specification given to SurveyMonkey. 39 people reported that they were under 18 in a survey question. About 12 persons gave “nonsense” responses to at least one important text response question, so their other responses were untrustworthy and had to be removed. 143 people were flagged to remove based on criteria to do with poor (or impossible) responses to travel behavior questions. People whose reported time to get to work or school was greater than 3 hours were also removed. If respondents’ average travel speed was over 80 miles per hour (calculated by dividing reported distance by reported time) they were also removed. This cutoff was chosen because even if people were driving 80 miles per hour (mph) the entire time they were on the freeway, they still would need to get on and off the freeway using side streets with lower speed limits. They were also removed if their walking speed was greater than 6 mph or their biking speed was greater than 30 mph. The walking speed cutoff was based on a study of walking speed for which the maximum was 3.2 miles per hour (Bohannon, 1997). 6 mph was decided as the cutoff point to give some leeway for estimation error from respondents. The biking speed cutoff point was decided based on a study showing that the typical cycling speed in three different municipalities in Sweden was between 7 and 16.5 miles per hour (Eriksson et al., 2019). Leeway was given again for this cutoff. 7 respondents were manually removed based on visual inspection that clearly showed the respondents were not answering truthfully. In total, these add up to 235, which is more than the 202 cases removed because some people ended up flagged in more than one removal category. This led to the final sample size of 800 respondents.

3.3 Data Collected

A summary of the characteristics of the final sample of respondents is shown in Table 3.2. Some comparisons can be made of changes experienced because of the pandemic. 341 respondents reported that they were not working before the pandemic. This jumped up by

112 people to 453 as of May 2020, as shown by the summary question “How many days per week do you typically work now?”. In May 2020, only essential businesses were open in Los Angeles, and the data reflects the impact this had on people’s ability to work. Comparing how many days per week respondents worked from home before versus during the lockdown, the percentage who said they never work from home went from 67% to 39%. This jump in people working from home also corresponds with the effects of the lockdown.

Table 3.2 Summary Statistics

Characteristic	N = 800[1]
In a typical work week (before COVID-19 restrictions) I worked on a...	
...fixed schedule defined by me (start in the morning and end of afternoon/evening)	120 (25%)
...fixed schedule defined by my employer (start in the morning and end of afternoon/evening)	172 (36%)
...flexible schedule defined by me	83 (17%)
...flexible schedule defined by my employer	31 (6.4%)
...shift schedule defined by me	22 (4.6%)
...shift schedule defined by my employer	54 (11%)
Not Working	318
In a typical work week (before COVID-19 restrictions) …	
I just worked from home	26 (5.4%)
I went to work at multiple places designated by others, excluding home (employers, customers, etc.)	49 (10%)
I went to work at multiple places of my own choosing, excluding home	38 (7.9%)
I went to work at the same place every day, excluding home	314 (65%)
I worked from home and other places designated by others	16 (3.3%)
I worked from home and other places of my own choosing	39 (8.1%)
Not Working	318
Before the COVID-19 restrictions how many days did you work from home in a typical week?	
0	304 (67%)
1	25 (5.5%)
2	32 (7.0%)
3	15 (3.3%)
4	15 (3.3%)
5	48 (11%)
6	6 (1.3%)
7	12 (2.6%)
Not Working	343
How many days do you work from home now?	
0	133 (39%)
1	16 (4.7%)
2	24 (7.0%)
3	25 (7.3%)
4	17 (5.0%)
5	101 (30%)
6	14 (4.1%)
7	11 (3.2%)
Not Working	459
Are you a student?	
No	713 (89%)
Yes, full-time	58 (7.2%)
Yes, part-time	29 (3.6%)
What school grade or level do you attend?	
2-year college (community college)	26 (30%)
4-year college or university	30 (34%)
Grade 9 to 12	4 (4.6%)
Graduate school/professional	19 (22%)
Other (please specify)	1 (1.1%)
Technical/vocational school	7 (8.0%)
Not in School	713
Did you move residences during the COVID-19 restrictions, even temporarily?	54 (6.8%)
Do you have a valid driver's license?	697 (87%)
What best describes your gender?	
Female	421 (53%)
Male	379 (47%)
Age	
18-29	169 (21%)
30-44	159 (20%)
45-60	226 (28%)
> 60	246 (31%)

3.4 Analysis

The analysis in this section explores strategically selected aspects of the conceptual model Figure 1.1. This analysis contains two mixture models: a latent class analysis (LCA) to find groups of individuals with similar travel behavior and a latent profile analysis (LPA) to find groups of individuals with similar attitudes about driving cars. The resulting classes from these two analyses are cross-classified and tested with a Pearson's chi-squared.

LCA uses categorical variables to identify underlying unmeasured classes. Grouping people using mixture models like LCA is different from simply using cutoff scores because, unlike cutoff scores, mixture modeling assumes that the class groupings are unknown. It uses probabilities of group membership, where the class with the highest probability is the class that an individual is placed in. Indicator sensitivity is considered, so LCA can look at which indicators are best for differentiating the classes. It also allows for measurement error.

3.4.1 Latent Class Analysis: Mode Choice for Commuting

This LCA identifies the clusters of modes used for commuting to work and/or school based on travel done *before* the coronavirus restrictions. Respondents who reported that they were going to neither work nor school were not included in the LCA and are consolidated into a sixth category. This category will be referred to as *Not in School or Working.*

The LCA was conducted using Mplus 8.6 (Muthén & Muthén, 2012). The order of operations for performing an LCA are as follows: A one-class model is fit, followed by a two-class, *et cetera* until a model is run that is not well-identified (Asparouhov & Muthén, 2012; Masyn, 2013; Nylund et al., 2007). Determining whether a model is well-identified involves inspection of a set of fit statistics which are recorded for each model. These are presented in Table 3.3. Depending on the purpose of the LCA, different fit statistics have greater priority to consider. For a model such as this one that is going to be used in further analysis based on the classification into classes, “within-class homogeneity and across-class separation” are important to consider, which means a greater emphasis placed on high entropy (Masyn, 2013).

Table 3.3 Fit Statistics: Travel Modes Used

Classes	Log likelihood	BIC	ABIC	*p*-value of BLRT	*p*-value of VLMRT	Entropy	BF
1	-1,322.964	2,695.757	2,670.364	-	-	-	0.000
2	-1,212.456	2,530.796	2,476.836	< 0.001	< 0.001	0.712	0.000
3	-1,142.398	2,446.737	2,364.210	< 0.001	< 0.001	0.849	> 15.000
4	-1,125.058	2,468.115	2,357.021	< 0.001	0.003	0.887	> 15.000
5	-1,107.970	2,489.995	2,350.334	< 0.001	0.032	0.899	> 15.000
6	-1,099.065	2,528.242	2,360.013	0.092	0.295	0.965	-

Note. BIC is Bayesian Information Criterion. ABIC is adjusted Bayesian Information Criterion. BLRT is Bootstrap Likelihood Ratio Test. VLMRT is Vuong-Lo-Mendell-Rubin Likelihood Ratio Test. BF is Bayes Factor.

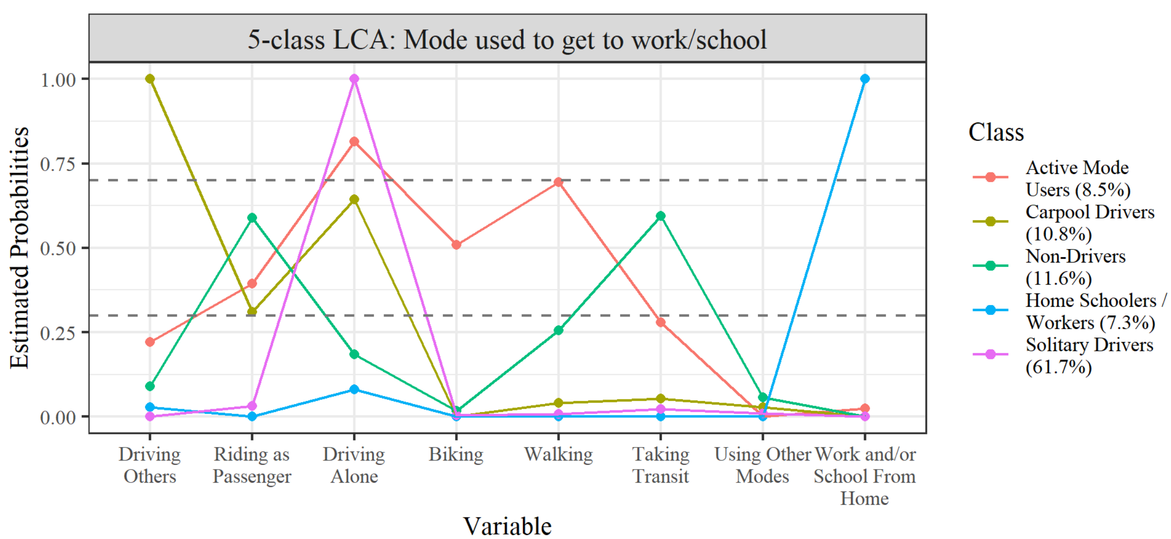
A non-significant *p*-value *for* either the Bootstrapped Likelihood Ratio Test (BLRT) or the Vuong-Lo-Mendell-Rubin Adjusted Likelihood Ratio Test (VLMRT) indicate that there is not a statistically significant improvement in model fit in the *k* class model as compared to the *k-1* class model (Asparouhov & Muthén, 2012; Masyn, 2013; Nylund et al., 2007). As Table 3.3 shows, The VLMRT reached a non-significant *p*-value of 0.295 with the 6-class model.

Based on fit criteria, class sizes, and interpretability, the 5-class model was chosen. An entropy value approaching 1 indicates clear delineation of the classes. So, the entropy value of 0.899 for the 5-class model means the indicators discriminate well between the classes (Celeux & Soromenho, 1996). Based on properties described in further detail below, the 5 classes will be referred to by the names *Active Mode Users*, *Carpool Drivers, Non-Drivers, From-Home Workers,* and *Solitary Drivers*.

Figure 3.1 shows the class-specific propensities for respondents to use each mode to get to work and/or school. There are distinct differences in the mode choices between the classes. Note that the item "Using Other Modes" has low probabilities for all classes. This means it would be an item to consider eliminating if future models were to be run since it does not differentiate well between the classes. Since all classes have low propensity for this item, it will not be examined in the class-specific discussions ahead. All but two classes have a high propensity for driving alone to get to work/school. This is reflective of the "driving culture" of Los Angeles. This is also reflected in the *Solitary Drivers* class containing 61.7% of the sample who work or go to school.

Horizontal lines have been added at y intercepts 0.7 and 0.3. This is to aid in interpretation and class differentiation, as any probabilities above 0.7 or below 0.3 would be considered high or low propensity, respectively (Masyn, 2013).

Figure 3.1 Class-specific item probability profile plot: 5-class LCA of modes used



The *Active Mode Users*, with an estimated proportion of 8.5%, have the highest probabilities of commuting using bicycles and walking of all the classes. This group of individuals has a low propensity for reporting driving others or working/schooling from home. Although the probability of using bicycles is only 0.50, it is significantly higher than the probability of biking in any other class. Class members have a high propensity for reporting walking to work/school, and although their probability of reporting using transit is on the border of the “low” category, it is the second highest probability of any class. Even the class that could be considered the most “eco-friendly” are very unlikely to use transit to commute. They also have a high propensity for driving alone. This is a manifestation of travel culture in Los Angeles and the lack of viable public transit infrastructure. Options for using public transit are limited and slow, and people do not use it unless they must (as is the case for *Non-Drivers*).

The *Carpool Drivers* make up 10.8% of the analyzed sample. They have a 100% probability of choosing driving others as one of their modes of travel. When compared to the low propensities of all the other classes, the difference becomes starker. Basically, almost everyone who drives others to work has been put into this same class. Members of this class have a low propensity for any travel mode that is not in a car: biking, walking, and public transit. They also have a low probability of working from home.

The *Non-Drivers,* with an estimated proportion of 11.6%, are characterized by having a low propensity for driving alone and driving others. Unlike the other classes besides the Home Schoolers / Workers, this group of individuals does not have a high propensity for commuting by any mode that involves driving a car. This class does not have any modes for which the probability of using it is higher than 0.7. The highest propensities this group has are for riding as a passenger and taking transit, which are both around 0.6. This indicates that individuals in this group do not drive a car very often.

The *Home Schoolers/Workers* make up 7.3% of the analyzed sample. Members of this class have a 100% probability of responding that they worked or went to school from home before the pandemic. This is the only class for which people have a low propensity to have reported traveling by any mode to work or school. This means that most of the people in this class mainly work/school from home, and do not ever commute to a work or school location.

Members of the *Solitary Drivers* class are characterized by a high propensity for driving alone and a low propensity for using any other mode or for working/schooling from home. This is the largest class, making up 61.7% of the sample, meaning this is the most typical pattern that would be present in the population. This reflects what is widely considered to be general travel behavior of Los Angeles residents.

3.4.2 Latent Profile Analysis: Attitudes

Latent profile analysis (LPA) is essentially the same as latent class analysis but using continuous instead of categorical variables. All the model interpretation methods described in the previous section still apply. Attitudes towards driving and other modes of transportation were measured using a shortened set of attitudinal questions originally used in the Puget Sound Transportation Panel (PSTP) (Murakami & Waterson, 1990). The full set of 23 questions was previously used by Lee and Goulias (2018) in an LPA, where the Likert scale items were used as continuous variables as done in this analysis. All members of the 800-person sample were used in this analysis. Respondents rated their agreement with the following statements on a scale of 1 to 5, from "Strongly Disagree" to "Strongly Agree":

- "I like the freedom of driving my own car"
- "I won't rely on another person to get to work on time"
- "My schedule is too erratic to be in a carpool"
- "Taking public transit doesn't fit my lifestyle"
- "Driving a car is a relaxing way to commute"
- "I enjoy driving my car even in heavy traffic"

Table 3.4 contains the fit statistics of different models estimated. Typically, for the likelihood ratio tests (VLMRT and BLRT) the *p*-value gradually increases with each increase in number of classes, indicating that the improvement in model fit gets less and less significant with each class increase. With this analysis, the 4-class model has a non-significant *p*-value, but the 5- and 6-class models both show highly significant improvements in model fit. In this type of analysis, it is uncommon for a model with fewer classes to have a less significant *p*-value than the models with more classes. Although according to the non-significant *p*-value of VLMRT for the 4-class model, the 3-class model would be the optimal choice, the 5- and 6-class models both show statistically significant improvements in model fit, and thus are viable candidates for model selection.

Table 3.4 Fit Statistics for LPA of Attitudes

Classes	Log likelihood	BIC	ABIC	*p*-value of BLRT	*p*-value of VLMRT	Entropy	BF
1	-7,652.121	15,384.457	15,346.350	-	-	-	0.000
2	-7,319.193	14,765.394	14,705.058	< 0.001	< 0.001	0.935	0.000
3	-7,126.237	14,426.274	14,343.710	< 0.001	< 0.001	0.985	0.000
4	-6,534.574	13,289.740	13,184.946	< 0.001	0.142	0.999	0.000
5	-6,433.476	13,134.336	13,007.314	< 0.001	< 0.001	0.937	0.000
6	-6,373.291	13,060.759	12,911.508	< 0.001	0.001	0.911	-

Note. BIC is Bayesian Information Criterion. ABIC is adjusted Bayesian Information Criterion. BLRT is Bootstrap Likelihood Ratio Test. VLMRT is Vuong-Lo-Mendell-Rubin Likelihood Ratio Test. BF is Bayes Factor.

Upon visual inspection of the 3- 4- and 5-class Item Probability Plots (Figure 3.2), it becomes clear why the 4-class model does not show significant improvement in fit over the 3-class model, but the 5-class model fits significantly better than the 4-class. While the 4-class model does not appear considerably different from the 3-class model, a unique class emerges in the 5-class model. This class will be referred to as the *Freedom Lovers* and will be described in more detail below. The 5-class model was chosen due to the good fit statistics, good separation of classes (Entropy = 0.937), clear distinctions in the qualities of each class, and parsimony over the 6-class model. The 5 classes will be referred to as *Cars Haters*, *Indifferent Respondents*, *Freedom Lovers*, *Car Users,* and *Car Lovers*.

Figure 3.2 Item Mean Plots for Attitude LPA

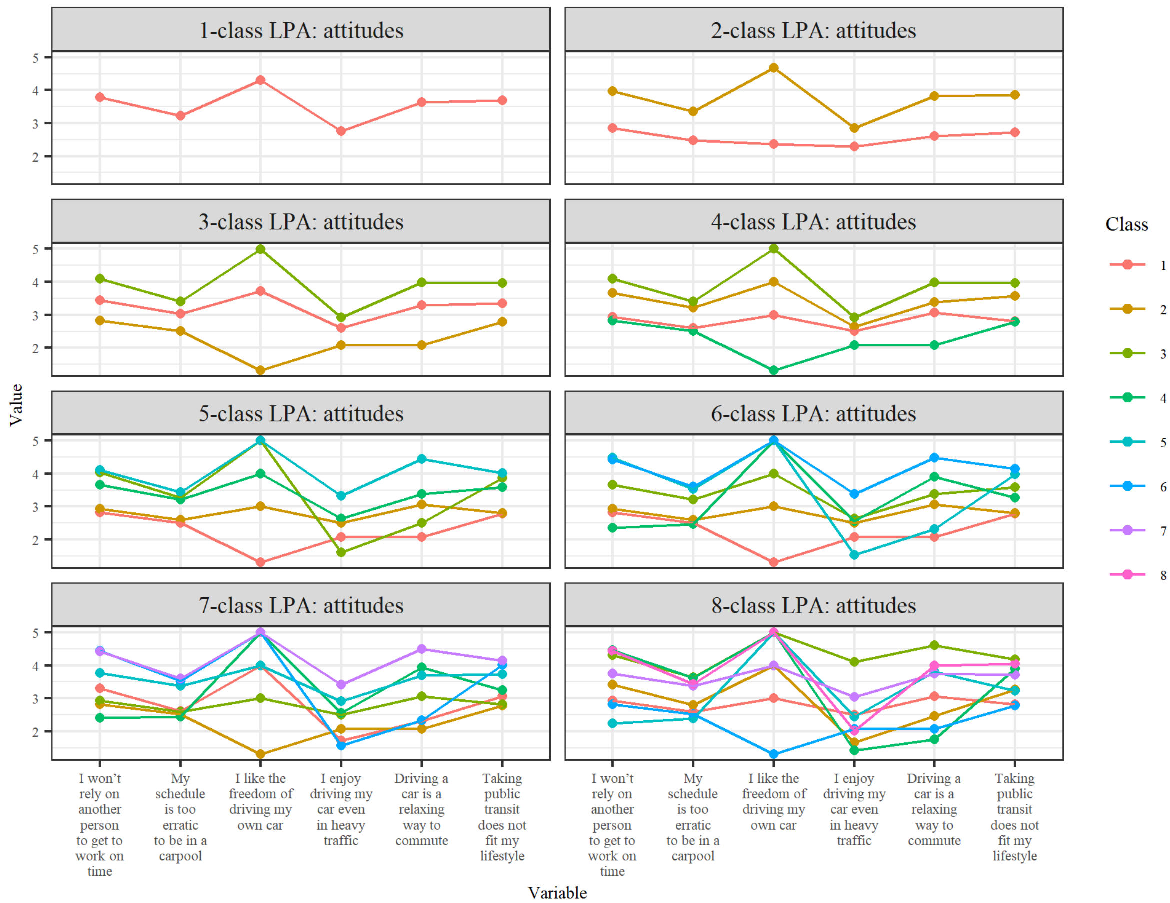
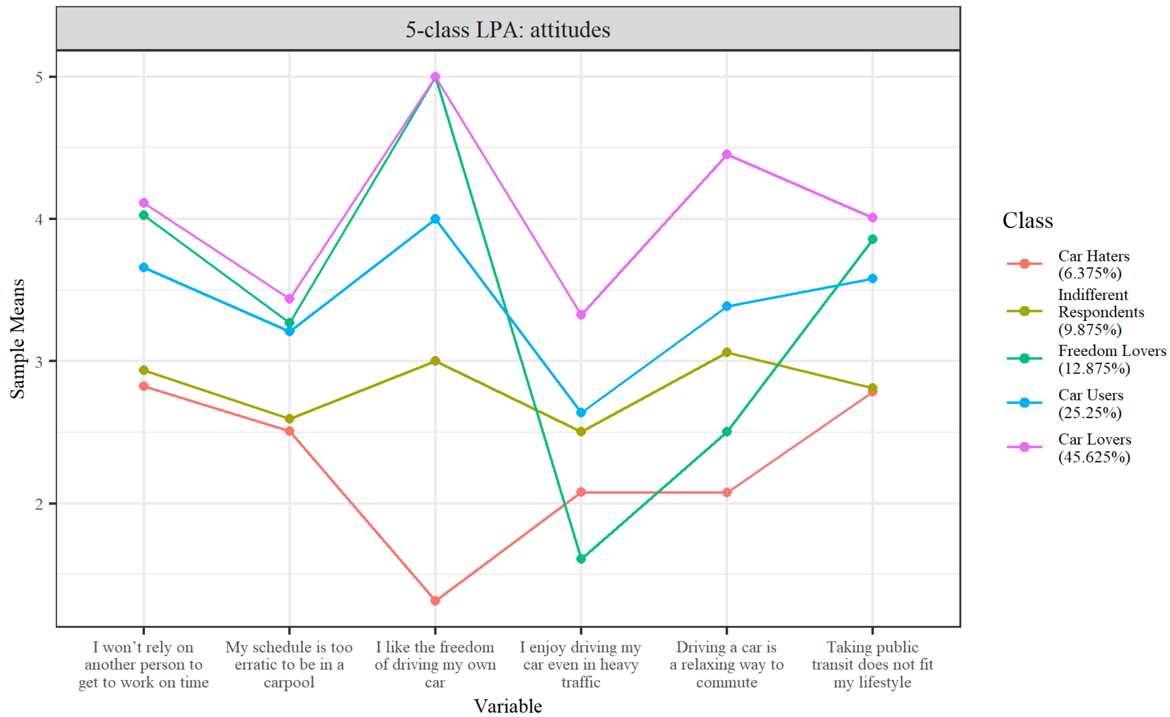


Figure 3.3 shows a larger and more detailed IPP of the 5-class model. On the y-axis, 1 is "strongly disagree", 3 is "neither agree nor disagree", and 5 is "strongly agree". The *Cars Haters* group members, making up 6.4% of the sample, do not think of cars as providing freedom, and they do not like using cars to get around. Unlike any other class, this one has a strong negative response to the prompt "I like the freedom of driving my car." They also have

negative responses to whether they enjoy driving their car in traffic and whether driving a car is a relaxing way to commute.

Figure 3.3 Item Mean Plot: 5-class LPA of Attitudes



Indifferent Respondents make up 9.9% of the sample. They have little opinion either way for most of the questions. The means of their responses to every question came out around the 3, which is "neither agree nor disagree". They have no notably strong opinions about any of the prompts given.

Freedom Lovers, making up of 12.9% of the sample, dislike congestion but love the freedom of driving in their car. They may enjoy driving for pleasure, but not commuting. They are characterized by their strong positive association towards the prompt "I like the freedom of driving my own car" coupled with strong negative association towards the prompt "I enjoy driving my car even in heavy traffic." They also positively endorse the statement that they will not rely on another person to get to work on time, and that public transit does not fit their lifestyle. These responses further emphasizing the importance of freedom of movement to this group.

Car Users make up 25.25% of the sample. Members of this group are fine with using their cars, but they do not feel passionately about it. They do not have extraordinarily strong opinions either way about any of the prompts, but on average they do respond moderately positively to the prompt about enjoying the freedom of driving a car.

Car Lovers is the most common attitude profile, making up 45.6% of the sample. On average, group members respond positively to prompts about driving their cars in all situations. Although they respond less positively to driving in heavy traffic, it is still a more positive response than any other group.

3.4.3 Cross Classification of Travel Mode and Attitude Classes

One way to consider the relationship between the LPA and the LCA model estimates is to use a cross classification of the latent classes extracted from the travel mode and attitude models (Table 3.5).

Table 3.5 Cross-Classification of Mode and Attitude Models

		Attitudes					
		Car Haters	Indifferent Respondents	Freedom Lovers	Car Users	Car Lovers	Row Totals
Commute Modes	Active Mode Users	14.0% (6)	4.7% (2)	7.0% (3)	20.9% (9)	53.5% (23)	100.0% (43)
	Carpool Drivers	5.5% (3)	3.6% (2)	9.1% (5)	30.9% (17)	50.9% (28)	100.0% (55)
	Non-Drivers	10.2% (6)	28.8% (17)	11.9% (7)	30.5% (18)	18.6% (11)	100.0% (59)
	From-Home School / Worker	5.4% (2)	16.2% (6)	24.3% (9)	21.6% (8)	32.4% (12)	100.0% (37)
	Solitary Drivers	3.8% (12)	3.2% (10)	12.8% (40)	26.5% (83)	53.7% (168)	100.0% (313)
	Not in School or Working	7.5% (22)	14.3% (42)	13.3% (39)	22.9% (67)	42.0% (123)	100.0% (293)
Column Totals		6.4% (51)	9.9% (79)	12.9% (103)	25.2% (202)	45.6% (365)	100.0% (800)

Pearson's Chi-squared test

χ-squared = 80.82, *df* = 20, *p*-value = 2.848e-09

For the Attitudes and Commute Modes model cross tabulation, Pearson's Chi-squared test resulted in a chi-squared value of 80.82. The *p*-value is less than the significance level of 0.05, so the null hypothesis is rejected and conclude that the two variables are dependent/correlated. This means the observed relationship between the two is significantly better than chance (i.e., the membership of the categories on one set of groups is not uniformly distributed across the categories of the other set of groups). In other words, attitudes are related to behaviors in a systematic way, as expected. For example, the *Solitary Drivers* are more likely to be *Car Lovers*. *Active Mode Users* make up 14.0% of the *Car Haters* group. In contrast, among the *Solitary Drivers*, only 3.8% are *Car Haters*. Among the *Not in School or Working* group, only 7.5% are *Car Haters*. However, among the *Active*

Mode Users there are 13.9% *Car Haters*. Moreover, 53.7% of *Solitary Drivers* are also *Car Lovers*.

Even using a survey designed to collect information on travel behavior and COVID-19 (and not designed to collect information to fit the theoretical model), the theoretical model shows validity. This illustrates how one goes about testing hypotheses of the relationships between attitudes and behavior using responses in surveys. Adding other factors and testing relationships among factors is one way to study the significance of the relationships in the conceptual model.

3.5 Conclusions

In this project, a few strategic changes of direction have been employed to account for external changes to the project circumstances. The fundamental direction of developing a strong theoretical model accompanied by data collection to test some of its aspects stayed the same as the original intent of the project. In addition, data collection was done to examine COVID-19 impact on the life of Los Angeles Metropolitan area residents. In terms of substantive findings, it was verified that in this region, as in other parts of the US, people experienced loss of jobs, forced relocations, and major changes in working and studying. In terms of the attitude-behavior relationship, this project confirmed the existence of more diversity in attitudinal groups of people with respect to their position towards the private automobile and found that these attitudes are correlated with the use of different modes. The survey design and conceptual model form the foundation for subsequent data collection and analysis. A third undertaking within this project was the design of a smartphone application, which is an ongoing effort at UCSB and will continue beyond the project reported here. More information about the development of the smartphone application can be found in Appendix C.

4 Measuring the effect of Exogenous Variables on Latent Classes

4.1 Introduction

The two goals of this chapter are to synthesize the research from this dissertation's previous substantive chapters into one model and to test the validity of the theoretical model structure shown in Figure 1.1 as much as possible using an already-existing dataset. The analysis includes a latent class analysis (LCA) with similar specifications to those in Chapter 3, resulting in profiles of travel behavior and a sequence analysis with similar specifications to those in Chapter 2 to produce clusters describing common patterns of time allocation in a day. Then, the time allocation clusters along with sociodemographic and attitude variables are tested as auxiliary variables for how they relate to the latent classes.

The results of this analysis, although they do provide supporting evidence for the theoretical model, cannot be used to entirely confirm or reject the model. These results will either add proof to the model validity or result in recommendations for alteration of the model if there are things that do not fit. Since the analysis uses an already-collected dataset, some variables must be adapted or chosen to fill certain roles (as proxies) even though that was not the original purpose of their measurement. They will not represent the strength of the conceptual model nearly as well as analysis of a survey designed to measure the desired constructs would be. The more complex moderation relationships are not plausible to test using this dataset because the survey questions are not designed to test this model structure, and thus the complex interrelatedness of these variables will not be possible to measure nicely without too much error obscuring whatever possible relationship there might be. Thus, a simpler multinomial logistic regression structure is used, and just one interaction term is tested. This interaction term evaluates the relationship between gender and fragmentation of schedules, which builds upon the research done in Chapter 3.

4.2 Methods

The Puget Sound Regional Travel Study was used in this analysis to test a section of the theoretical model and synthesize the research from the first two substantive chapters in this dissertation into one model. First, variables were identified and modified appropriately for use in the analysis. After data cleaning and variable preparation, latent class analysis (LCA) was performed to identify travel behavior clusters using analogous methods to the LCA in Chapter 3. The last phase of the analysis was testing the relationships between the latent classes and a set of auxiliary variables. The auxiliary variables measured time allocation to activities and travel, sociodemographic characteristics, attitudes, preferences, habits, and the interaction between gender and schedule complexity. To use time allocation as an auxiliary variable, sequence analysis was performed on the PSTP travel diaries using analogous methods to those used in Chapter 2. These auxiliary variables were tested in a set of nested multinomial logit models using the latent classes as the dependent variable.

4.2.1 Data

The Puget Sound Regional Travel Study (PSTS) is a cross-sectional travel study collected by the Puget Sound Regional Council with two of the three planned waves completed – one in 2017, one in 2019, and the last collected in 2021 but not yet published (*Regional Household Travel Survey Underway | Puget Sound Regional Council*, n.d.). In this analysis, the 2017 and 2019 data were used together as one dataset. The surveying methods are practically the same for 2017 and 2019, so from this point forward, discussion will only mention the two waves separately when there are important distinctions between the two.

The study covers the four counties that make up the Puget Sound Regional Council: King, Kitsap, Pierce, and Snohomish. According to the 2019 PSTS final report, as of 2017, the region had a total population of over four million people (RSG & Westgroup Research, 2020). The final report explains that, for sample stratification, the consultant team used geographic proportional sampling, “compensatory sampling” (where the target sample size is increased in certain regions based on expected response rates), and “targeted oversampling” (where areas of interest are sampled at higher rates). Population information used to set target sample sizes for each subregion came from the American Community Survey (ACS). Extra funding from the city of Seattle went towards increasing the target sample rate of areas labeled as “Urban Villages,” which are of particular interest to the city.

Responses were collected from April–June in both 2017 and 2019. Data was collected at household- person-, and travel-level (in the form of a travel diary). A travel diary is a type of data collection where respondents report every place visited over the diary period (usually 24 hours). They include details about how they got there, who they went with, the purpose of being there, and several other traits. Respondents reported their travel diaries using either a smartphone application or over a telephone call. For respondents using the smartphone app, seven days of travel diary information was collected. For respondents reporting by telephone call, 24 hours of information was collected (RSG & Westgroup Research, 2020). Because of this discrepancy in diary collection time, only the Tuesday travel diaries from the weeklong survey respondents were used.

Upon assessment of the travel diary data, there were some problematic cases that needed to be repaired or, if repair were not possible, removed before a trustworthy analysis could be performed. In the full 2017 and 2019 dataset, there are 11,940 people in 6,319 households. After the data cleaning process, 10,597 people in 6,029 households remain.

Some cases were repaired and others were not fixable, resulting in the entire travel record having to be removed. When a record was removed, the person and their entire household was removed from the sample. This was done to preserve the reporting of "complete households," meaning households where every member completed a travel diary. Although for this analysis it likely would not affect the outcomes if partial households were included, preserving the consistency of formatting/structure of the dataset (having complete household reporting) ensures that the analysis does not somehow have bias because of that.

4.2.1.1 Identification, repair, and removal of "bad cases"

Whatever the data source, when using travel diary data for analysis, researchers face some common issues. Some issues are due to recording error, while some are due to the nature of the data and how it will be interpreted when read into a data processing/analysis software. All work on the data can be found on GitHub[2]. The issues addressed in this study through repair or removal include the following. The first issue is trips starting on one day and ending on the next (for example, a respondent leaves a friend's house at 23:30 and arrives home at 00:15). In many travel diary datasets, the "departure time" and "arrival time" variables are recorded as simple times without dates attached. This is an issue because, without dates attached to the times, it will look to a computer like these trips ended earlier than they started. See the departure and arrival times for Trip Number 3 in Table 4.1, where even though it is obvious to human eyes that these start/end times make sense, to a computer it is not. The second issue is trips incorrectly reported as evening instead of morning, or vice

[2] https://github.com/e-mcbride/PSRC.analysis

versa. For example, respondent meant 10AM but they reported 10PM. The third issue is trips starting after midnight on the travel diary day. This happens because the travel diary dates typically are from 3am on the diary day to 3am on the following day. The fourth issue is trips ending after 3am, when the diary day is over. In the 24-hour diaries, the end time is "capped" at 3am, so any trip that ends after that is just reported as 3am. However, in the weeklong diaries, it was possible for trips to have reported end times that were past the end of the current diary day. Since only the Tuesday records from the weeklong diaries were used, the end times of these were retroactively capped at 3am.

Table 4.1 Example of travel diary data structure

Unique Person ID	Trip #	Departure Date	Departure Time	Arrival Time	Origin	Destination	Mode used
10	1	09/13/2021	07:30	08:00	Home	Work	Car (alone)
10	2	09/13/2021	18:00	18:45	Work	Friend's Home	Car (alone)
10	3	09/13/2021	23:30	00:15	Friend's Home	Home	Car (alone)
10	4	09/14/2021	00:30	00:40	Home	7-Eleven	Walk
10	5	09/14/2021	00:50	01:00	7-Eleven	Home	Walk

4.2.2 Sequence Analysis

Using the travel diary portion of the PSTS, a sequence analysis was performed. This was to identify the common patterns of travel present in the data. These patterns were then used as variables to classify each observation, and then used in the regression as explanatory variables on the latent classes.

Sequence analysis starts with the status of every observation at every given time point in the sequence. This status is called a "state." For this analysis, the places respondents reported going to were simplified into categories: home, school, work, grocery store, and other. These five place types, along with the travel to get from place to place, were used as the possible states that an observation could be in at any given time in the sequence. The time points of these sequences were minute-by-minute, so in the 24-hour diaries, there were 1,440 sequence time points. So, in the sequence data, at every minute over their 24-hour diary period, a respondent could be at home, at work, at school, at the grocery store, at another location, or traveling.

Using a function in TraMineR, the data were put into a format that the other functions in TraMineR can work with. Many “bad” cases were fixed or removed in a process described in the Data section. After data preparation, the sequence analysis was run using the R package TraMineR Version 2.2-2 (Gabadinho et al., 2011). The steps are the same as in Chapter 1, but to review them and go over what changed:

For every observation, a measure of schedule fragmentation called the complexity index is calculated. The complexity index $C(s)$ of a sequence is a composite measure that considers the number of unique states an observation displays and the number of transitions between states.

$$\boldsymbol{C(s)} = \sqrt{\frac{\boldsymbol{q(s)h(s)}}{\boldsymbol{q_{max}h_{max}}}} \qquad (\textbf{Eq.}\ 4.1)$$

Where s is a single sequence, $q(s)$ is the number of transitions in the sequence, $h(s)$ is the entropy index of the sequence, q_{max} is the maximum number of transitions possible, and h_{max} is the maximum possible value for the entropy index (Gabadinho et al., 2011; Gabadinho, Ritschard, & Studer, 2010; Ritschard, 2021).

The entropy index is otherwise known as Shannon’s entropy.

$$\boldsymbol{h(p_1, \dots, p_a)} = -\sum_{\boldsymbol{i=1}}^{\boldsymbol{a}} \boldsymbol{p_i \log(p_i)} \qquad (\textbf{Eq.}\ 4.2)$$

Where a is the number of possible states and p_i is the proportion of occurrences of the ith state in the considered sequence. A more detailed discussion of the entropy and complexity indices can be found in Chapter 2. The complexity index for each respondent is saved for later use in the analysis of auxiliary variables relationship to the latent classes.

The next step of the sequence analysis was to identify the common patterns within the schedules of respondents. This is a three-step process: first is optimal matching, then agglomerative nesting, then finally deciding on the number of sequences for the solution. The optimal matching and agglomerative nesting processes are computationally costly, so with 1,440 time points for 10,259 sequences (10,259 * 1,440 = 14,772,960 time points), this computation is beyond the capacity of typical computers. To reduce the computational cost, the sequences were reduced by extracting what state each observation was in at every 5-minute interval. This led to 1,440/5 = 288 time points for each respondent. The goal of the optimal matching process is to measure the pairwise dissimilarity scores between sequences. The process is described in detail in Chapter 2, so only a summary will be provided here. First, a substitution cost matrix is created (using function `seqsubm()` from R package `TraMineR`). This uses the observed rates of transition between the various states to calculate the "costs" to move between those states. Next, pairwise dissimilarities between sequences are calculated (using function `seqdist()` from R package ` TraMineR`). These dissimilarity scores are calculated using the number of state changes that would need to happen at each of the 288 time points in a sequence for it to be the same as the sequence it is being compared to. The substitution cost matrix is used in the dissimilarity score calculation, so each substitution that would be required to make two sequences the same has an associated cost, and those are all added together to get the dissimilarity score. To get clusters of similar sequence patterns, these pairwise dissimilarity scores are used in agglomerative nesting (AGNES). The function `agnes()` from the R package `cluster` takes pairwise dissimilarities as input. Once again, because a more detailed explanation is provided of AGNES in Chapter 2, just a brief explanation will be provided here. AGNES is a "tree" method of cluster identification. It starts with every individual score as a "leaf" on the tree (or a single-observation cluster). Step-by-step, it merges clusters until it arrives at one cluster with all observations in it. There are a few methods to decide which clusters are merged at each step, and in this analysis, I use Ward's method. Please see Chapter 2 for an explanation of Ward's method, but in brief it compares every possible pair of clusters at each step and prioritizes reducing within-cluster variance. The result of running AGNES is a set of clusters ranging from one cluster per observation all the way to one cluster with all observations. The next step is to decide the number of clusters.

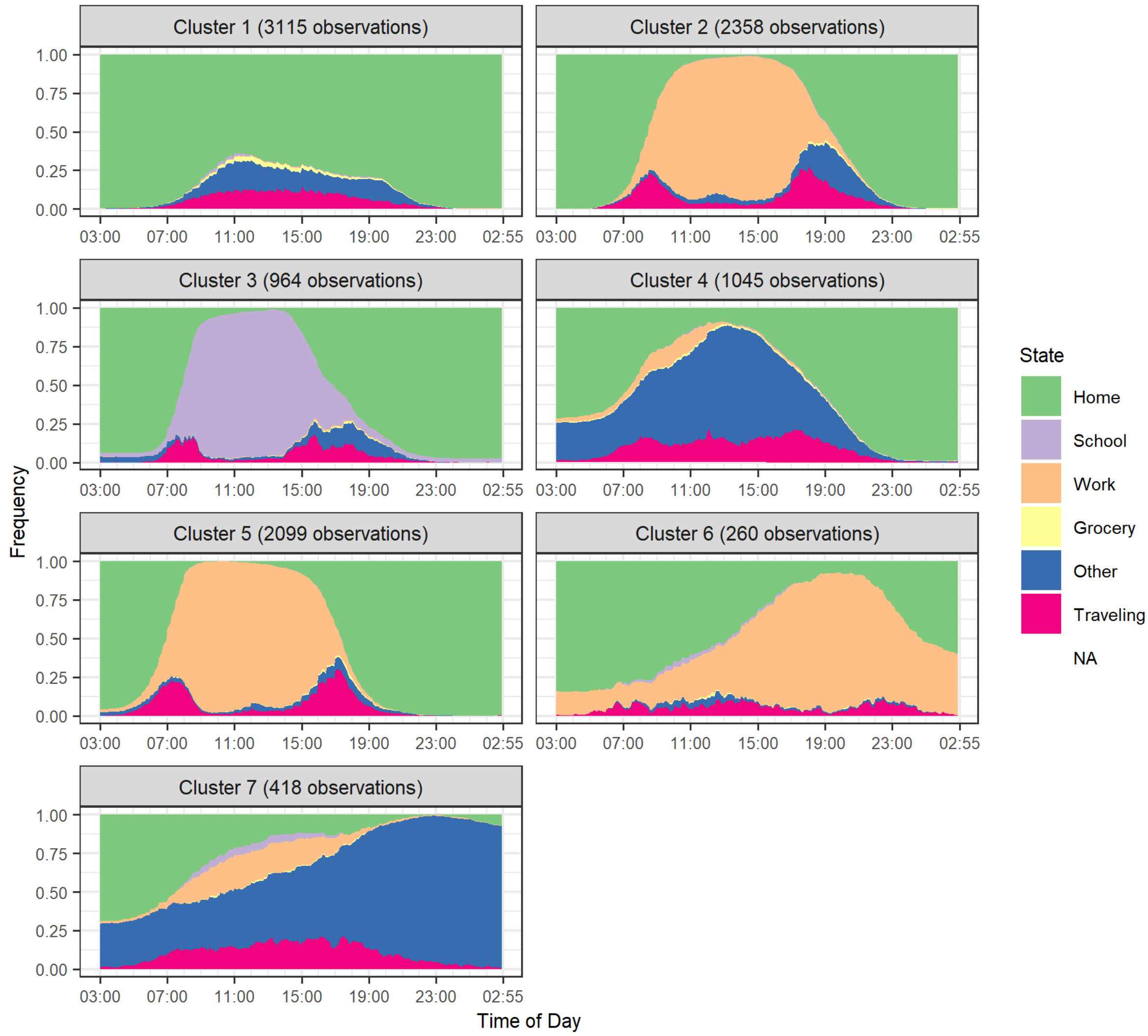
To choose the number of clusters, the two- through eight-cluster solutions were plotted, and the patterns of behavior that emerged were assessed. The traits considered in the decision include the stability of pre-existing clusters as more clusters are added, the distinctness of each cluster from the others, the value of distinguishing the specific new pattern traits from pre-existing patterns when added to the upcoming analysis as auxiliary variables, comparability of patterns to the solution in Chapter 2, and the parsimony of the solution.

The plots (as seen in Figure 4.1, Figure 4.2, and Figure 4.3) are area plots, which are the equivalent of stacked bar charts for continuous data. They show how the composition of each sequence pattern varies over the 24-hour period. At every time point, the proportion of respondents in each state (at home, at work, at school, at the grocery store, at another place, or traveling) is plotted. These plots are divided by cluster.

After inspecting the plots of the two- through eight-cluster solutions, the six-cluster solution (Figure 4.1) was selected as the ideal compromise between parsimony and distinctive activity patterns. Moreover, at six clusters, the set of patterns that emerge is extremely similar to the six-cluster solution chosen in Chapter 1. With each increase in the number of clusters up until the six-cluster solution, more value was added through the emergence of distinct activity patterns. At the seven-cluster solution (Figure 4.2), however, the newest cluster to emerge was nearly identical to the already-existing Typical Work Day cluster of the six-cluster solution.

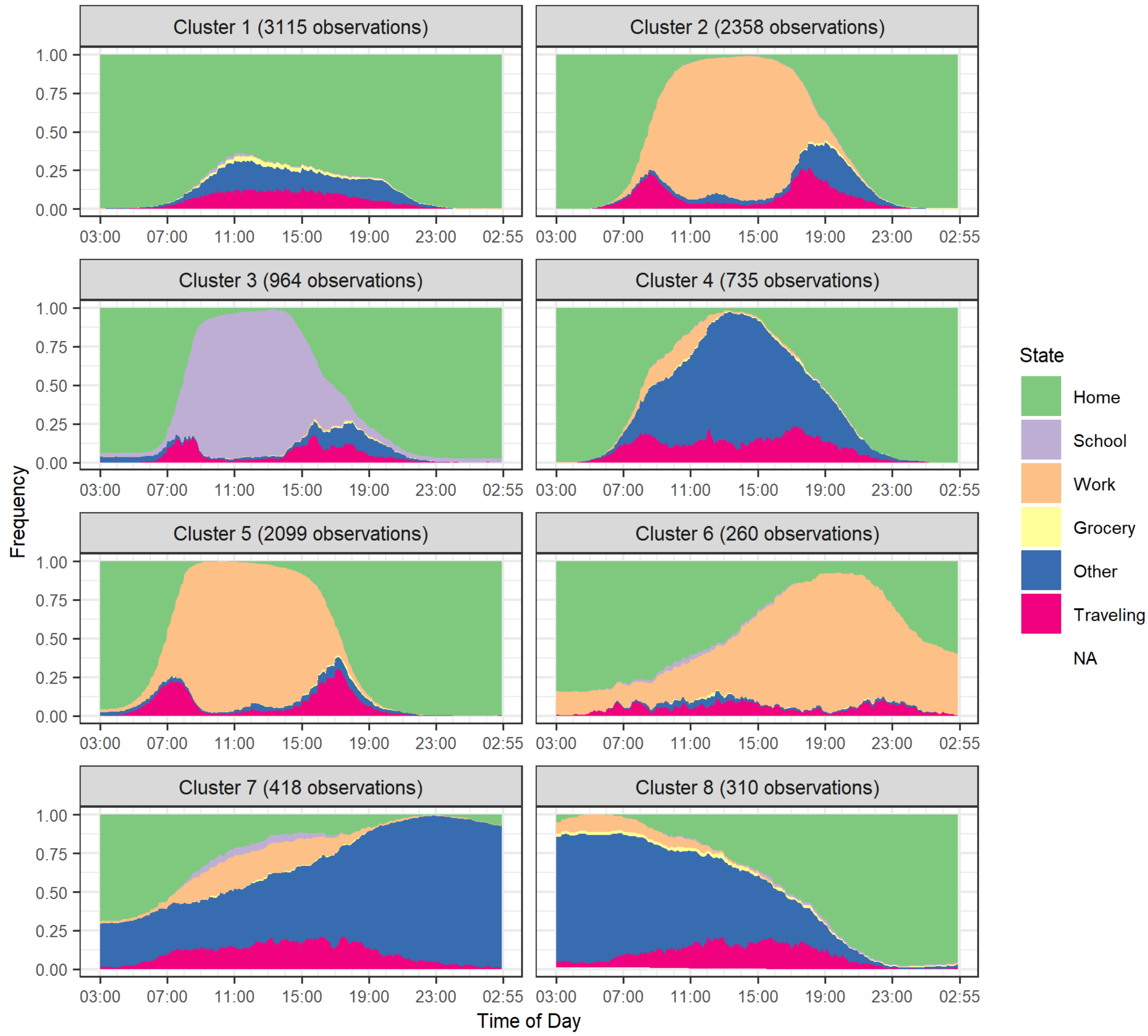
Figure 4.1 Six-Cluster Solution

Figure 4.2 Seven-cluster Solution



The largest number of clusters plotted and inspected was the 8-cluster solution (Figure 4.3). In this solution, a unique activity pattern emerges in Sequence Cluster 8, in which people start their days out of their homes and end up at home by the end of the day. However, the two nearly identical work day clusters were still present. Although this eighth cluster is unique and would be valuable to measure, the presence of the two nearly identical work day classes plus the fact that an eight-cluster solution adds a lot of complexity generally to a model made this solution too complex to be worth it for the loss of parsimony. Thus, the six-cluster solution was decided upon.

Figure 4.3 Eight-cluster Solution



4.2.3 Latent Class Analysis

Latent class analysis (LCA) takes a set of variables for which it is theorized that there is an underlying categorical variable that has not been directly measured that defines the relationship between a set of indicator variables. The goal of LCA is to uncover that latent variable. LCA uses binary variables. For a more in-depth explanation of latent class analysis (LCA), please refer to Chapter 3. An extension of LCA is to explore the relationships between the latent variable and some observed auxiliary variables. In this analysis, auxiliary variables are included. After running a typical LCA, auxiliary variables are added in a multinomial regression while the classes are "held constant" by fixing their measurement error, which takes into account the imperfect assignment of individuals into classes (Asparouhov & Muthén, 2014). Holding the classes constant allows for testing of the auxiliary variables without them impacting class membership.

As of now, the generally-accepted best method to test relationships between latent variables and auxiliary variables is the 3-step method (Asparouhov & Muthén, 2014; Nylund-Gibson et al., 2014; Vermunt, 2010). As described in Asparouhov and Muthén (2014), step 1 is a typical latent class analysis without auxiliary variables to find the most likely classes for each observation. Step 2 is to create a nominal variable from those most likely classes and to calculate the classification uncertainty rates for each class. Step 3 is to run a multinomial regression, with the most likely class variable as the dependent variable and the auxiliary variables as the independent variables. The most likely classes are used as a latent class indicator variable while fixing the classification uncertainty rates to the probabilities found in step 2. By fixing the uncertainty rates, the measurement error that makes the assignment of individuals to classes imperfect is considered, and the classes will remain stable when auxiliary variables are added.

The binary categories used in this LCA to analyze the underlying patterns of behavior were whether a respondent used a mode of travel during their diary day. The number of trips made using each mode did not matter.

4.2.3.1 LCA: Class Enumeration

Although when choosing the number of classes in a LCA there are fit statistics to guide the decision, substantive theory, interpretability, and utility must also be considered (Nylund-Gibson & Choi, 2018). In this case, the fit statistics (Table 4.2) did not definitively point towards a single solution. In the end, the decision was mostly made based on interpretability, utility considering the analysis goals, and the visually identifiable patterns in the item probability plots.

The BLRT and VLMRT are likelihood ratio tests to measure whether there are statistically significant improvements in model fit between a model and the model with one less class (category of the latent variable). Typically, if either the BLRT or VLMRT *p*-values are above 0.05 for a model, it is suggested that the model with one less class should be selected since there was not significant improvement in model fit with the addition of a class (Masyn, 2013; Nylund et al., 2007). However, in this case, these likelihood ratio tests did not ever show non-significant *p*-values. Entropy was not definitive either. Except for a dip for the five-class model, entropy values also continue to improve with more classes added. However, the increase in entropy value between the six and seven class models is smaller than the increase between the five and six class.

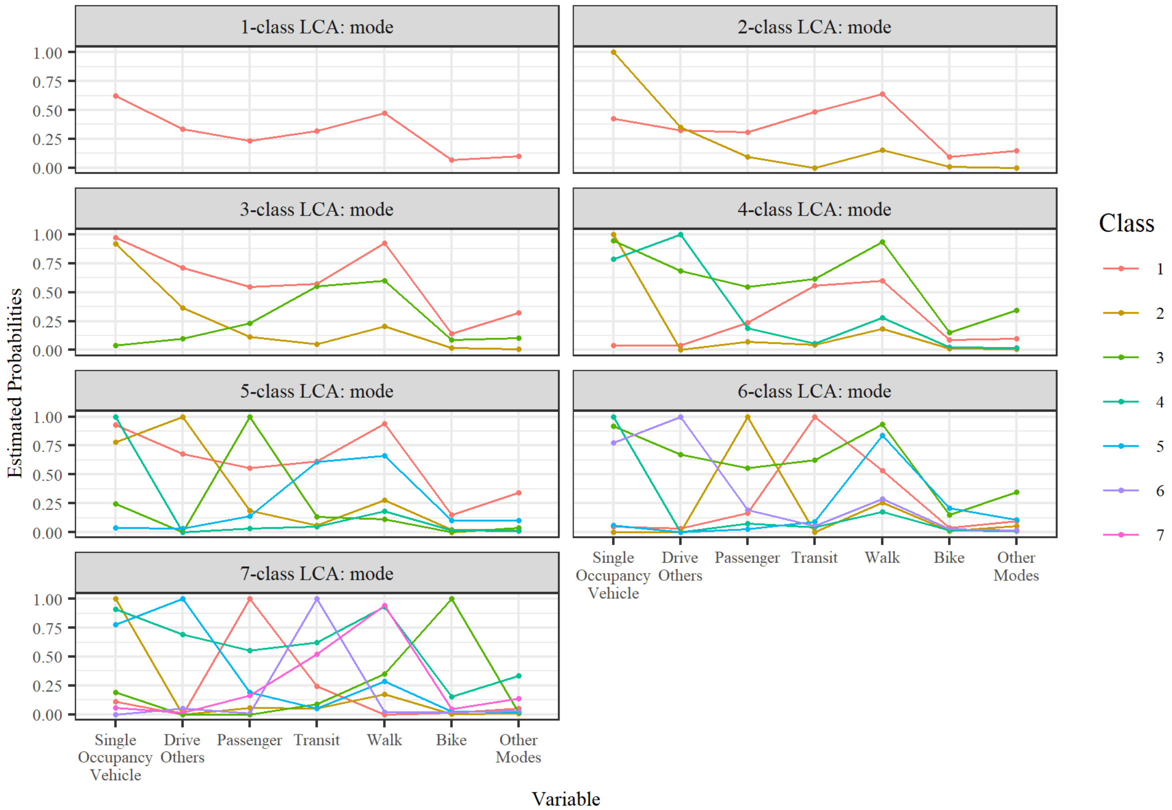
Table 4.2 Latent Class Analysis Fit Statistics

Number of Classes	Loglikelihood	BIC	aBIC	BLRT *p*-value	VLMRT *p*-value	Entropy
1	-34,248.658	68,561.192	68,538.947	-	-	-
2	-32,848.287	65,833.452	65,785.784	< 0.001	< 0.001	0.710
3	-31,764.585	63,739.051	63,665.961	< 0.001	< 0.001	0.772
4	-31,523.506	63,329.893	63,231.381	< 0.001	< 0.001	0.854
5	-31,343.046	63,041.975	62,918.040	< 0.001	< 0.001	0.851
6	-31,164.150	62,757.186	62,607.827	< 0.001	< 0.001	0.870
7	-31,006.649	62,515.186	62,340.405	< 0.001	< 0.001	0.876

Figure 4.4 shows the item probability plots for the one- through seven-class models. Every line on a plot represents the composition of a latent class in terms of the probability of the class members reporting using a certain mode of travel. Because the variables are binary, these probabilities can also be interpreted as the percentages of class members that used each

mode. Because of the indefinite results from the fit indices, the primary way that the number of classes was decided upon was by interpreting these item probability plots, using the analysis goals and theory as guidelines for the decision. The 6-class was decided on because it separates the people who primarily use transit from the people who primarily walk, whereas in the 5-class model these two modes are combined. The 7-class solution was also considered because the people who primarily ride their bicycles are identified, however, parsimony took priority. The size of the classes needed to be considered, because when adding auxiliary variables in the multinomial logit, too small of classes would cause issues. There need to be enough observations in each class so that there are not pairs of auxiliary variable categories and dependent variable categories that have zero observations. Having empty cells can cause a multinomial logit model to become unstable (IDRE Research Technology Group, 2013).

Figure 4.4 Item Probability Plots for All Analyzed Models



4.2.3.2 Addition of auxiliary variables

The latent class model was then tested for the relationships between the set of latent classes and auxiliary variables using a nested multinomial logit model, which is a series of multinomial logistic regression models which progressively add more auxiliary variables. The models are compared to the model that came before them, and each model is tested for loglikelihood improvements. Model 0, or the baseline model, only tests the relationship of sequence pattern clusters to the mode choice classes. Model 1 adds demographic variables, Model 2 adds attitudes/habits/values, and Model 3 adds an interaction effect between gender and the complexity index.

A variable was created to measure income based on the 2017 Washington Self-Sufficiency Standard (SSS) (Pearce, 2017). The SSS calculates cost of living using by-region cost of living information combined with household characteristics including the number of adults, number of children aged 0-2, number of children aged 3-5, number of children aged 6-12, and number of teenagers aged 13-18 in households. For every household in the PSTS, the SSS was calculated using the information available, which was not perfect since age ranges for children did not align perfectly with the SSS age ranges. However, these age ranges were close enough to get estimates of whether each household was "above," "below," or "around," the SSS for their neighborhood.

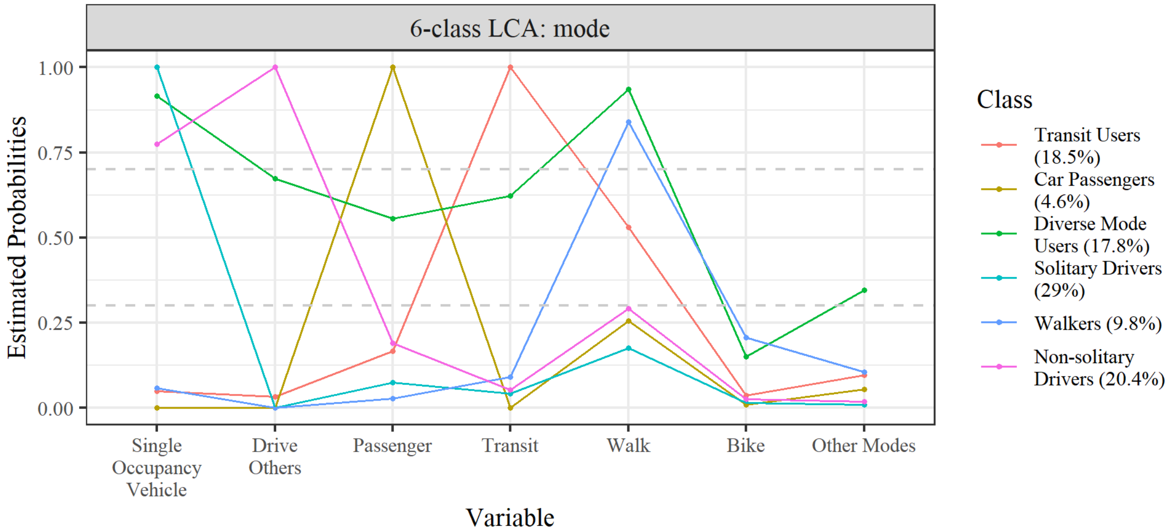
4.3 Results

First, the results of the LCA without covariates are discussed. This includes how the number of classes was decided upon. Then, the results of the LCA with auxiliary variables are presented, looking at one model at a time with progressively more auxiliary variables added. The effect of the newly added variables on the already-present variables from the previous models is also investigated.

4.3.1 Latent Class Analysis

Figure 4.5 is an enlarged plot of the chosen six-class solution, and in the legend, it shows the percentage of respondents that were assigned to each class. The horizontal light-colored lines in this plot are at the 0.3 and 0.7 probability marks to aid in visualizing the strong defining characteristics of each class. Anything above the 0.7 mark can be considered a strong characterization, because that means 70% or more of respondents in a class used that mode of travel. Below the 0.3 mark is strong because it means 30% or less of respondents in a class used that travel mode.

Figure 4.5 Item Probability Plot for 6-class LCA



The first class is described as the *Transit Users* class because it is characterized by a perfect 100% probability of respondents having used public transit during their travel diary day. Respondents in the *Transit Users* class have a very low probability of using single occupancy vehicles, driving others, riding as passengers, riding a bicycle, and using other modes. This class is not characterized by whether members report walking trips (there is about a 50% chance that *Transit Users* report walking trips). The next class is called *Car Passengers* because members had a 100% probability of riding in cars as passengers on their diary day. This class is also characterized by low probabilities of using any other travel mode. *Diverse Mode Users* are people who used multiple modes on their travel day. Membership in this class is primarily characterized by both driving a vehicle on one's own and walking. There is also less strong, but still high, probability of driving others and using transit. Using a bicycle and using "Other" modes is still highly improbable for this class. Multimodal travelers were similarly identified using latent class analysis methods by Molin *et al.* (2016). The *Solitary Drivers* class is characterized by a 100% probability of driving by themselves on their diary day. This class has a low probability of reporting using any other mode of travel. *Walkers* class members have a high probability of reporting walking trips during their diary day and a low probability of reporting any other travel mode. The *Non-Solitary Drivers* class has a high probability of reporting both driving others and driving alone. Class members have low probabilities of reporting using any other mode of travel.

4.3.2 *Latent Class Analysis with Auxiliary Variables*

In all the following models, the reference category for the latent classes is *Solitary Drivers*. The reference category for the sequence pattern clusters is *Travel Day*. For race, the reference category anyone who is not White, Asian, Hispanic, or Black, so they were classified as "Other" for simplicity since individually they make up such small portions of the population.

Table 4.3 Likelihood Ratio Test

Models Compared	LRT	Difference in df	*p*-value
0 vs. 1	3,283.206	80	<.001
1 vs. 2	1,386.554	85	<.001
2 vs. 3	11.334	5	.045

The likelihood ratio test results (Table 4.3) show that every additional model has significantly better explanatory power than the one before it. However, Model 3 is just barely significant. Although the impact of the interaction term added in Model 3 is not very strong, the insight provided by this interaction term makes it valuable to investigate in the model.

4.3.2.1 Model 0 (Baseline Model; sequence pattern clusters only)

The baseline model is a multinomial regression of the sequence pattern clusters on the mode choice classes alone. The reference category for the classes is the *Solitary Drivers*. The reference category for sequences is *Travel Day*. In this table of model results (Table 4.4) and all the following model result tables, unless otherwise noted, a red highlight means a statistically significant and a negative coefficient, while a green highlight means a statistically significant and positive coefficient. Statistical significance in this model table is highlighted at the *p*-value <= 0.05 level. If the *p*-value is <=0.01, the cell is highlighted in a darker shade of red or green. If the cell is just bold, that means a *p*-value between 0.05 and 0.1.

Table 4.4 Model 0 Results (baseline)

	Est. by Class				
Row Labels	**Transit Users (885)**	**Car Passengers (321)**	**Diverse Mode Users (1,244)**	**Walkers (719)**	**Non-solitary Drivers (1,486)**
Sequence: Home Day	**-0.785**	**-0.762**	**-2.239**	-0.266	-0.249
Sequence: Typical Work Day	-0.302	**-2.314**	**-1.848**	**-0.911**	**-0.969**
Sequence: School Day	**0.818**	-0.704	**-0.772**	0.712	-0.664
Sequence: Errands Day	-0.125	**-1.282**	**-1.023**	**-1.115**	-0.246
Sequence: Atypical Work Day	**-0.675**	**-1.909**	**-2.309**	**-0.869**	**-1.260**

Note. Reference class for mode user class is *Solitary Drivers*. Reference class for sequence patterns is *Travel Day*. **Goodness of Fit measures:** Loglikelihood: -11,018.33; AIC: 22,096.65; BIC: 22,301.47; aBIC: 2,206.14; Entropy: 0.815

4.3.2.1.1 Transit Users

The relative log-odds of being in the *Transit Users* class versus the *Solitary Drivers* class would decrease by 0.785 if moving from a *Travel Day* to a *Home Day*. So, in comparison to having a *Travel Day*, having a *Home Day* makes respondents significantly more likely to be in *Solitary Drivers* than to be in *Transit Users*. Having an *Atypical Work Day* instead of a *Travel Day* makes respondents slightly significantly more likely to be in the *Solitary Drivers* class. Having a *School Day* instead of a *Travel Day* makes respondents slightly significantly more likely to be in *Transit Users* than in *Solitary Drivers*. If respondents are having a *Home Day*, they are less likely to be *Transit Users*, but besides that, there are not big differences in the types of days *Transit Users* and *Solitary Drivers* have.

4.3.2.1.2 Car Passengers

Except for a *School Day*, having any type of day instead of a *Travel Day* decreases the relative log-odds of being in the *Car Passengers* class versus the *Solitary Drivers* classes. In other words, in comparison to having a *Travel Day*, having any type of day besides a *School Day* makes respondents significantly more likely to be in *Solitary Drivers* than in *Car Passengers*. Another way to look at it: *Travel Day* and *School Day* people are more likely to be *Car Passengers* than they are to be *Solitary Drivers*. Respondents who are primarily getting around as passengers in a car are significantly more likely to be in a *Travel Day* than they are to be working, mostly staying at home, or running errands. If they are doing those things, they are significantly more likely to be *Solitary Drivers*.

4.3.2.1.3 Diverse Mode Users

Respondents having any type of day that is not a *Travel Day* are significantly more likely to be in the *Solitary Drivers* class than they are to be in the *Diverse Mode Users* class. This means that those who have a *Travel Day* are significantly more likely to be in the *Diverse Mode Users* class than they are to be in *Solitary Drivers*. People who use many different travel modes in a day are most likely to be having a *Travel Day*, and if they are having any other type of day, they are significantly more likely to be driving by themselves than they are to be in the *Diverse Mode Users* class.

4.3.2.1.4 Walkers

Having a *Typical Work Day*, *Errands Day*, Or *Atypical Work Day* makes respondents significantly less likely to be in the *Walkers* class than they are to be in the *Solitary Drivers* class. People who are working or running errands are significantly more likely to be driving by themselves than they are to be walking.

4.3.2.1.5 Non-Solitary Drivers

The relative log-odds of being in *Non-Solitary Drivers* versus *Solitary Drivers* decreases by 0.969 if moving from a *Travel Day* to a *Typical Work Day*. In comparison to having a *Travel Day*, having a *Typical Work Day* or an *Atypical Work Day* makes respondents significantly more likely to be *Solitary Drivers* than to be *Non-Solitary Drivers*. People who are working are more likely to be driving by themselves than to be driving others.

4.3.2.2 Model 1: Adding Socioeconomic Status

In the next model (Table 4.5), demographic/socioeconomic variables are added. Variables added include race, respondent age group, gender, employment status, number of adults in the household, and presence of minors in household divided by age (aged 0–4, aged 5–15, and aged 16–17). Also included was a variable measuring whether household income is below the 2017 Washington self-sufficiency standard (SSS). A description of the creation of this variable measuring SSS can be found in section 4.2.3.2: Addition of auxiliary variables. Travel behavior variables included are whether respondent have a driver's license, whether there is at least one vehicle per adult with a driver's license in the household, and Complexity. Complexity is the measurement of how complex a person's day is using both the number of transitions between activities the person has and the number of unique activities a person participates in. More detail on this calculation can be found in the Methods section of this chapter, and even more detail can be found in the Chapter 2.

Table 4.5 Model 1 Results (baseline + socioeconomic)

Row Labels	Est. by Class Transit Users (885)	Car Passengers (321)	Diverse Mode Users (1,244)	Walkers (719)	Non-solitary Drivers (1,486)
Sequence: Home Day	**-1.035**	**-1.359**	**-1.805**	-0.381	-0.232
Sequence: Typical Work Day	-0.249	**-2.107**	**-2.398**	**-0.651**	**-1.355**
Sequence: School Day	-0.109	**-2.229**	**-1.489**	0.506	**-1.313**
Sequence: Errands Day	-0.022	**-1.234**	**-1.324**	-0.599	**-0.701**
Sequence: Atypical Work Day	**-0.852**	**-1.753**	**-2.544**	**-1.059**	**-1.519**
Has Driver's License	**-6.941**	**-6.300**	**-4.589**	**-6.092**	0.321
$\geq$ 1 Vehicle per Adult w/ Driver's License	**-2.916**	**-1.625**	**-1.922**	**-2.718**	**-0.692**
Race: White	0.001	-0.367	**0.880**	0.042	0.129
Race: Asian	0.028	-0.298	**0.722**	-0.184	0.250
Race: Hispanic	0.025	0.280	**0.910**	0.214	**0.562**
Race: Black	-0.212	-0.283	0.373	**-1.083**	-0.006
Age 18–34	**0.864**	0.116	**2.131**	**1.030**	**0.678**
Age 35–64	**0.479**	-0.110	**1.291**	**0.575**	**0.450**
Number of adults in Household	**-0.897**	**0.370**	**-0.594**	**-0.943**	**0.261**
Female	0.128	**1.223**	**0.162**	**-0.175**	-0.017
Worker	**-0.489**	**-0.978**	**0.354**	-0.163	**-0.486**
Income Below the SSS	0.164	**-0.795**	**-0.779**	-0.066	**-0.435**
Minors Aged 00–04 in Household	-0.422	**0.561**	**0.650**	-0.134	**1.224**
Minors Aged 05–15 in Household	0.164	0.307	**0.845**	**0.414**	**1.491**
Minors Aged 16–17 in Household	0.569	0.276	0.477	0.523	**0.759**
Complexity	1.860	-5.872	**30.202**	**-24.660**	**27.698**

Note. Reference class for mode user class is *Solitary Drivers*. Reference class for sequence patterns is *Travel Day*. **Goodness of Fit Measures:** Loglikelihood of this model: -9,376.727; AIC: 18,973.45; BIC: 19,724.46; aBIC: 19,374.91; Entropy: 0.844

4.3.2.2.1 Transit Users

Considering only p-values below 0.05, for *Transit Users,* all sequence pattern clusters have the same significance as in Model 0. So, people having a *Home Day* are significantly more likely to be *Solitary Drivers* than they are to be *Transit Users*, and besides that there are not significant differences in activity/travel patterns between *Transit Users* and the baseline category.

Having a driver's license makes respondents significantly more likely to be *Solitary Drivers* than to be *Transit Users*, as does having at least one vehicle per adults with a driver's license in the household. Any of the latent classes that are not driving-focused will have lower likelihoods on these variables than classes where people only drive, so this is to be expected.

Race is not significantly different between *Transit Users* and *Solitary Drivers*. However, age is significant: people under age 65 are significantly more likely to be *Transit Users* than they are to be in the baseline category. While holding all other variables constant, increasing the number of adults in a household by 1 is associated with a decrease in the relative log-odds for being in *Transit Users* instead of *Solitary Drivers* by 0.897. Thus, higher numbers of adults in a household corresponds to higher likelihood of being in the *Solitary Drivers* reference class. If respondents work, they are significantly more likely to be *Solitary Drivers*. The Sequence patterns of *Typical Work Day* and *Atypical Work Day* are not highly significantly different, so having a job, whether the diary day is a work day, increases the likelihood of being in *Solitary Drivers* instead of *Transit Users*.

4.3.2.2.2 Car Passengers

The relationships between the sequence clusters and membership in *Car Passengers* versus *Solitary Drivers* are all the same as in the baseline model, except *School Day* has become highly statistically significant, while it is not significant in the baseline model. This is because in the baseline model, something is "muddying" the association between *School Day* and being *Car Passengers*, making the association less negative than it "should" be. According to the correlation matrix of the variables in the model, the variables that have slight correlation with *School Day* are being aged 18 to 34 (0.115) and having household income below self-sufficiency standard (SSS; 0.102). So, younger people and poorer people are more likely to be in school. Once the variable *Age 18-34*, which has a positive coefficient, is added to the model, it accounts for the "positivity" that was obscuring the negative relationship between *School Day* and *Car Passengers*. The income variable has a negative coefficient like *School Day*, so it would not be counteracting the negative relationship.

Like *Transit Users*, respondents holding a driver's license and in households with at least as many vehicles as adults with licenses are less likely to be *Car Passengers* than they are to be in *Solitary Drivers*. An increase of 1 adult in the household increases the probability of being *Car Passengers* instead of *Solitary Drivers* by 0.370. Thus, the higher the number of adults in a household, the more likely a respondent is to be in the *Car Passengers* class instead of the *Solitary Drivers* class. Respondents who live with other adults are more likely to spend time as a passenger because they are likely to perform joint activities with those other adults.

This class contains people who do not have a driver's license, who are in households with fewer cars than adults with driver's licenses, and who are traveling on their diary day. Along with those people, this class has many people in two-parent households who are at home with babies. *Car Passengers* class members are less likely to work, more likely to have household income above self-sufficient levels, more likely to have a higher number of adults in household, and more likely to have babies in household. This indicates that the other adult(s) in the household earn enough money to support the family. Identifying as female is also significantly more likely, which aligns with the general findings that women tend to take on more childcare responsibilities (Goulias et al., 2020). Only presence of babies has a significant effect on class membership, not other child ages. Especially with infant children, it is more common for women to be the ones to stay at home.

4.3.2.2.3 Diverse Mode Users

All but one of the sequence cluster patterns have the same relationship to the *Diverse Mode Users* class as they did in the baseline model. The *School Day* sequence is now significant for the same reasons described in the analysis of the *Car Passengers* class: the inclusion of the variable *Age 18-34*, which has a moderately high correlation with *School Day*. Since *Age 18-34* is also positively associated with the *Diverse Mode Users* class, before it was included, it obscured the relationship between *School Day* and the latent classes, making it less negative than it should have been. Now, some of the variance previously being explained by *School Day* simply because younger people more often attend school is explained by a more direct relationship through the age variable.

As with the previous classes, and for the reasons mentioned earlier, having a driver's license or having at least 1 vehicle per adult in the household makes respondents significantly more likely to be in the *Solitary Drivers* class than in the *Diverse Mode Users* class.

Race being White, Asian, or Hispanic instead of the reference class "Other" makes someone significantly more likely to be *Diverse Mode Users*. People aged 18 to 64 are significantly more likely to be in this class than to be *Solitary Drivers*, thus people age 65+ are more likely to be *Solitary Drivers*. An increase of 1 adult in the household corresponds to a decrease of 0.594 in the log-odds of being in *Diverse Mode Users*, meaning higher numbers of adults in households makes respondents significantly more likely to be in *Solitary Drivers* than to be in *Diverse Mode Users*. Respondents with income below the SSS are more likely to be *Solitary Drivers* than to be *Diverse Mode Users*. Respondents in households with children aged 0 to 15 are significantly more likely to be in *Diverse Mode Users* class than to be in *Solitary Drivers* class.

An increase of 1 in Complexity increases log-odds of being in *Diverse Mode Users* class by 30.202. *Diverse Mode Users* have significantly higher Complexity than *Solitary Drivers*, and this is to be expected since one of the components of the Complexity calculation is the number of shifts between activities. To use many modes of travel in a day, there needs to be frequent shifting between activities, which would increase Complexity.

4.3.2.2.4 Walkers

In the *Walkers* class, compared to the baseline model, *Typical Work Day* went from highly significant determinant of class membership to only slightly significant. *Errands Day* drops from highly significant to not significant at all. *Typical Work Day* is correlated with *Age 18-34* (0.174), being a worker (0.521), and higher Complexity (0.220), and it is negatively correlated with income below the self-sufficiency standard (-0.174). Out of these correlated variables, Complexity was the one that had the most significant association with class membership. Some of the variance being explained by the *Typical Work Day* sequence was indirectly coming from class members having lower average Complexity scores than *Solitary Drivers*. People who are in the *Walkers* class also have lower Complexity than people in the *Solitary Drivers* class.

Only Complexity is correlated with *Errands Day*, so having an *Errands Day* typically correlates with higher Complexity scores. Once Complexity is added to the model, negativity is “taken away” from *Errands Day* because lower Complexity scores correspond to a higher likelihood of being in *Walkers* versus *Solitary Drivers*. *Atypical Work Day* is still significant, and *Typical Work Day* is slightly significant, so it is still less likely for *Walkers* to be having a work day than *Solitary Drivers*.

Compared to the reference class of "Other" races, Black respondents are more likely to be *Solitary Drivers* than to be *Walkers*. All other race categories are not significantly different from "Other." Respondents under 65 are more likely to be *Walkers* than *Solitary Drivers*. More adults in a household means higher odds of being in *Solitary Drivers* instead of *Walkers*. Identifying as female makes someone slightly less significantly likely to be a *Walker*. Children aged 5-15 in households corresponds to respondents having a slightly significantly higher likelihood of being *Walkers*.

4.3.2.2.5 Non-Solitary Drivers

For *Non-Solitary Drivers*, the *Typical Work Day* sequence is still highly significant. Now, *School Day* and *Errands Day* sequences are significant with negative coefficients, where in the baseline model they are not significant. This means one or more variables were making the estimated coefficients less negative than they should have been, and now that those variables are included in the model, the coefficients more accurately represent the direct relationships between *School Day*/*Errands Day* and *Non-Solitary Drivers*. How *School Day* and *Errands Day* sequences are affected by the addition of SES variables has been discussed in the analysis of other latent classes. In short, *Errands Day* is affected by Complexity, which is highly significant and positive for *Non-Solitary Drivers*. Complexity's positive relationship with this latent class removes the counteracting effect from *Errands Day*, and the same is true of the effect of *Age 18-34* on *School Day*.

Unlike all other latent classes being compared to *Solitary Drivers*, the *Non-Solitary Drivers* class did not have significant differences in Driver's License ownership. Of course, since this is another class of drivers, this is expected. *Non-Solitary Drivers* still have a significantly lower likelihood of being in a household with enough cars for the number of adults with licenses. This is also predictable, since some people who are driving with passengers will be with other adults in their household, especially when there are too few cars for everyone to drive their own.

Respondents under age 65 are more likely to be *Non-Solitary Drivers* than to be in the reference class. Increases in the number of adults in the household is associated with higher odds of being *Non-Solitary Drivers*. Workers are significantly more likely to be *Solitary Drivers* than to be in this latent class. Having children of all ages in the household makes respondents significantly more likely to be *Non-Solitary Drivers*. Of course, if there are children in the household, someone must drive them around to school, activities, and appointments, *et cetera*.

4.3.2.3 Model 2: Adding Habit and Attitudes

In Model 2 (Table 4.6), variables added include *Only Uses Car* (measures whether respondents have any travel mode habits besides using their car), three variables measuring what it would take for respondents to use transit more, four variables measuring what it would take for respondents to use bicycles more, and eight variables measuring what was important to respondents when choosing their current homes. With the large number of variables added to the model, a *p*-value between 0.05 and 0.1 is considered statistically significant for Models 2 and 3.

Table 4.6 Model 2 Results (baseline + socioeconomic + attitudes)

Row Labels	Est. by Class				
	Transit Users (885)	Car Passengers (321)	Diverse Mode Users (1,244)	Walkers (719)	Non-Solitary Drivers (1,486)
Sequence: Home Day	-0.700	**-1.320**	**-1.753**	-0.371	-0.237
Sequence: Typical Work Day	0.262	**-2.130**	**-2.290**	-0.566	**-1.378**
Sequence: School Day	-0.253	**-2.551**	**-1.691**	0.300	**-1.400**
Sequence: Errands Day	0.360	**-1.227**	**-1.256**	-0.501	**-0.665**
Sequence: Atypical Work Day	-0.403	**-1.764**	**-2.335**	**-0.942**	**-1.567**
Has Driver's License	**-5.434**	**-5.686**	**-3.429**	**-4.986**	-0.155
At Least 1 Vehicle per Adult with a Driver's License	**-2.094**	**-1.535**	**-1.402**	**-2.214**	**-0.754**
Race: White	0.140	**-0.383**	**0.830**	-0.059	0.138
Race: Asian	0.205	-0.255	**0.782**	-0.095	0.160
Race: Hispanic	0.180	0.278	**0.847**	0.129	**0.475**
Race: Black	-0.197	-0.310	0.383	**-1.104**	-0.065
Age 18–34	0.290	0.212	**1.541**	**0.524**	**0.747**
Age 35–64	0.146	0.004	**0.913**	**0.318**	**0.502**
Number of Adults in Household	**-0.532**	**0.428**	**-0.416**	**-0.771**	**0.227**
Female	0.067	**1.247**	**0.179**	-0.053	-0.045
Worker	**-0.698**	**-0.985**	0.247	**-0.364**	**-0.491**
Income Below the SSS	**0.428**	**-0.785**	**-0.606**	0.136	**-0.495**
Minors Aged 00–04 in Household	-0.154	**0.666**	**0.937**	0.158	**1.152**
Minors Aged 05–15 in Household	0.219	0.238	**1.011**	**0.479**	**1.427**
Minors Aged 16–17 in Household	**1.288**	0.461	**1.097**	**0.966**	**0.730**
Complexity	-3.932	-5.792	**24.910**	**-32.054**	**29.033**
Only Uses Car	**-4.771**	**-0.293**	**-0.849**	**-1.470**	**0.383**
Use Transit More: Safer Ways to Get to Stops	**1.653**	**0.662**	**0.291**	-0.019	0.075
Use Transit More: Increased Frequency	0.309	**-0.578**	0.069	-0.331	0.075
Use Transit More: Increased Reliability	**0.722**	0.201	**0.757**	0.355	0.035
Use Bike More: Shared use path or protected bike lane	0.197	0.202	0.126	-0.036	0.029
Use Bike More: Neighborhood Greenway	-0.078	-0.364	0.157	0.144	0.069
Use Bike More: Bike Lane	0.057	0.026	0.156	0.114	0.101
Use Bike More: Shared Roadway Lane	-0.239	0.294	**-0.401**	**0.537**	-0.118
Use Bike More: End of Trip Amenities	-0.061	-0.327	0.168	0.033	-0.023
Home Choice: Reasonably Short Commute to Work	0.210	-0.205	0.183	**0.279**	0.019
Home Choice: Affordability	**-0.502**	-0.158	-0.007	**-0.783**	-0.103
Home Choice: Being Close to Family or Friends	-0.185	0.025	-0.155	-0.153	0.011
Home Choice: Being Close to the Highway	**-0.582**	-0.123	**-0.387**	**-0.728**	0.018
Home Choice: Quality of Schools (K-12)	**-0.324**	-0.157	**-0.410**	**-0.343**	0.130
Home Choice: Space & Separation from Others	-0.111	**0.277**	**-0.198**	**-0.246**	0.075
Home Choice: Close to Public Transit	**0.945**	-0.031	**0.389**	**0.490**	-0.008
Home Choice: Walkable Neighborhood, Near Local Activities	0.111	0.206	**0.639**	**0.856**	0.047

Note. Reference class for mode user class is *Solitary Drivers*. Reference class for sequence patterns is *Travel Day*. **Goodness of Fit Measures:** Loglikelihood: -8,683.45; AIC: 17,756.9; BIC: 19,088.23; aBIC: 18,468.56; Entropy: 0.862

4.3.2.3.1 Transit Users

Once the attitudes, values, and beliefs variables are added to the model, the *Home Day* sequence is no longer significantly different between this class and the reference class. In Model 2, variance that was previously explained by *Home Day* is now being explained by other variables. The *Home Day* coefficient is less negative than it was (from -1.035 to -0.700), meaning *Home Day* was acting as a proxy for some other negative effect that is now included in the model more directly. In the correlation matrix, *Only Uses Car* has some slight correlation with *Home Day* (0.164). For the *Transit Users* class, *Only Uses Car* is negative, meaning class members are unlikely to only have a habit of using their car, which would be correlated with lower *Home Day* values. The *Use Transit More* variables are all positive, and these are all negatively correlated with having a *Home Day* (Safer Way to Get to Stops: -0.08, Increased Frequency: -0.107, Increased Reliability: -0.11). Although *Only Uses Car* and the *Use Transit More* variables are not very highly correlated with *Sequence: Home Day*, the combination of effects from all these new variables is what made *Home Day* less significant.

The age of respondent variables are also no longer significant determinants of being in *Transit Users* versus *Solitary Drivers*. *Age 18-34* has slight negative correlation with *Only Uses Car* and choosing a home near schools. It has slight positive correlation with *Use Transit More* and *Use Bike More* variables, living within a short commute to work, living near public transit, and living in a walkable neighborhood. These are all traits associated with youth which before were indirectly measured through *Age 18-34*, so their effect on class membership is more directly measured in Model 2. Younger adults tend to be more active, more interested in bicycling, more interested in using public transportation, more interested in living in urban areas with walkable/mixed-use neighborhoods, and less likely to have or want to use a car. *Age 35-64* is also no longer significant for the *Transit Users* class. Although the correlation matrix shows that this age group is not even slightly correlated with most of the "youthful" variables mentioned, it is being compared to *Age 65+*, for whom these variables are much more negatively correlated. *Age 35-64* are more likely than the reference group to express these "youthful" traits.

Income Below the SSS is now significant. Something was making it less positive than it should have been. There is nothing in the correlation matrix that stands out for this one. Correlations between *Income Below the SSS* and other variables are basically non-existent, making analysis of why the income variable is now significant difficult.

Unlike in Model 1, in Model 2, respondents in households with 16 and 17-year-old kids are significantly more likely to be in the *Transit Users* class than to be in the *Solitary Drivers* class. The most notable correlation with *Minors Aged 16-17 in Household*, although it is not very strong, is choosing a home based on quality of K-12 schools (0.157). Now that this is included, its negative effect that was previously being expressed through *Minors Aged 16-17* is more directly represented in the model. With Home Choice variables included Model 2 shows that having older children means parents do not have to be as dependent on cars to get around.

Compared to *Solitary Drivers* (the reference class), those in the *Transit Users* class were significantly more likely to have responded that they would use transit more if there were safer ways to get to stops or increased service reliability. It is important to remember (as mentioned earlier), that people who were already using transit at least 5 days a week were not asked these questions. Those respondents were grouped with the people who said they *would* use transit more under given conditions, since out of the two categories, they were part of the group that was interested in using transit under the right conditions, and the right conditions have already been met for them. So, *Transit Users* class members are more likely to use transit given safer ways to get to stops and increased reliability, or they already do use transit frequently and presumably are satisfied with their choices or are captive to transit—e.g. new immigrants that did not assimilate yet (Beckman & Goulias, 2008).

In terms of residential preferences (used here to account for potential residential self-selection), valuing affordability, proximity to family or friends, proximity to the highway, and quality of K-12 schools significantly decreased the relative log odds of being in the *Transit Users* class versus the *Solitary Drivers* class. However, as would be expected, proximity to public transit was an important determinant of where these respondents chose their homes. This shows that there was already an underlying positive attitude or belief about using public transit in these class members, since proximity to transit was important and they use it almost exclusively.

4.3.2.3.2 Car Passengers

In Model 2, the only variable from Model 1 for which the significance in determining *Car Passengers* class membership changed is *Race: White*. The change in *p*-value is from 0.101 in Model 1 to 0.096 in Model 2, so from just above to just below the threshold to be considered significant at all. This is such a small change that it is not worth discussing further.

Car Passengers were significantly more likely than *Solitary Drivers* to say they would not use transit more if there was increased frequency of buses/trains. In terms of their residential preferences, they were significantly more likely to say that "space and separation from others" was important in choosing their current home. In combination with their responses to the transit questions, this class was likely to contain more people who lived in the suburbs, where the frequency of transit would not matter because there are not stops nearby enough for it to matter. This is further reinforced by the fact that they were significantly more likely to respond that they would increase their transit usage if there were safer ways to get to stops. Access to transit for this group was apparently limited by where they lived. However, these hypotheses cannot be tested in a satisfactory because land use and access to transportation data are not available in this version of the Puget Sound database.

4.3.2.3.3 Diverse Mode Users

For *Diverse Mode Users*, everything that was present in Model 1 has the same relationship with class membership as it does in Model 2, except *Worker* is no longer significant at all and *Minors Aged 16-17* is now a significant determinant of class membership. The decline in significance for *Worker* is not very meaningful, because in Model 1, *Worker* had a *p*-value between 0.05 and 0.1, which was not considered as a statistically significant range. By the terms of this analysis, it was never considered significant. In Model 2, however, a *p*-value between 0.05 and 0.1 is considered statistically significant. Identifying as female in Model 2 stays statistically significant in the 0.05 to 0.1 range between the two models, and because of this it is worth noting. Apparently, identifying as female slightly significantly increases the log-odds of being in the *Diverse Mode Users* class instead of the *Solitary Drivers* class.

Respondents in households with 16 and 17-year-old kids are now significantly more likely to be in the *Diverse Mode Users* class than to be in the *Solitary Drivers* class. The explanation for this change is the same as it is for *Transit Users*. With the Home Choice variables now included, their negative effects are no longer are being expressed by proxy through *Minors Aged 16-17*.

Members of the *Diverse Mode Users* class were more likely to agree that they would use transit more if there were safer ways to get to stops and increased reliability, or they already used transit at least five days a week. They were significantly more likely than *Solitary Drivers* to disagree that they would use bicycles more if there were more shared roadway lanes. Other bicycle questions were not significantly different from the *Solitary Drivers* reference group.

Class members placed significantly less value than the reference class on their home being close to the highway, close to high-quality K-12 schools, and having space and separation from others. They placed significantly more value than the reference class on being close to public transit and living in a walkable neighborhood and near local activities. Class members prioritized living in an area with more mixed-use and prioritize more mobility by modes besides cars.

4.3.2.3.4 Walkers

In Model 2, *Typical Work Day* and *Female* are no longer significant in determining membership in the *Walkers* class instead of the reference class. Despite there not being any strong correlations in the correlation matrix between *Typical Work Day* and any of the variables added by Model 2, the *p*-value of *Typical Work Day* in Model 1 was between 0.05 and 0.1, which is already on the low side to be considered statistically significant. In Model 2, enough variance previously explained by *Typical Work Day* is now explained by new variables to make it no longer statistically significant in determining class membership. Identifying as female is also no longer considered statistically significant in Model 2. This variable is like *Typical Work Day*, in that the level of significance in Model 1 was already low, so it would not take much correlative power shifting away from that variable for it to no longer be considered statistically significant.

Being employed now significantly increases the log-odds of being in the *Solitary Drivers* class instead of the *Walkers* class, where previously it was not significant. *Walkers* were significantly more likely than *Solitary Drivers* to say they chose their homes because they were a reasonably short commute to work, which is somewhat correlated (0.266) with being a worker (as in: not all workers agree that this is important, but it is important to some). There is an interaction effect being expressed here: if they have a job, and they prefer to/must travel on foot, then they also prefer living within a reasonable walking distance from their workplace. *Minors Aged 5-15* is also now statistically significant.

The same pattern observed in *Transit Users* and *Diverse Mode Users* is present in *Walkers*: whereas in Model 1 *Walkers* in households with 16 and 17-year-old kids are not considered significantly different from *Solitary Drivers*, in Model 2, they are considered significantly different. This class shares with the other two the statistically significant and negative relationship between *Home Choice: Quality of Schools (K-12)* and membership in the class. With this variable present, the negative effect on *Minors Aged 16-17* is reduced.

Members of the *Walkers* class are significantly more likely than the *Solitary Drivers* reference class to say they would use bicycles more if there were improvements in shared roadway lanes. For their home choice, they were more likely to have put importance on having a reasonably short commute to work. This makes sense because walking is a more physically taxing mode of transport than driving, so it would be more important to someone who primarily commutes by walking to be able to do so in 30 minutes or less. *Walkers* were more likely to have responded that affordability, proximity to the highway, quality of schools, and space and separation from others were not important to their home choice. They were more likely than *Solitary Drivers* to respond that proximity to public transit, a walkable neighborhood near local activities, and a reasonably short commute to work were important in their home choice. These values show that these were people already more interested in alternative mode use.

4.3.2.3.5 Non-Solitary Drivers

In Model 2, there are no variables that were also present in Model 1 that had significant differences in their relationship to class membership in *Non-Solitary Drivers,* so this discussion will only address what the new variables add to the model. *Non-solitary Drivers* do not show any attitudinal differences from *Solitary Drivers*. They did not have significantly different responses to any of the "use transit more," "use bike more," or "home choice" questions. Just like *Solitary Drivers*, class members drove cars more than they used any other mode. They also had a high probability of driving by themselves during their travel day. The main difference between the two classes was whether they sometimes had passengers in their cars. Thus, no difference in attitudes/values is in alignment with expectations since behaviorally these two groups are not that different. In fact, *Non-Solitary Drivers* were the only class with a higher likelihood than *Solitary Drivers* of only using their car to get around. This class still had a significantly lower chance of having at least as many household vehicles as there are adults. This also makes sense, since they probably are sometimes driving other adults in the household around.

4.3.2.4 Model 3: Adding Interaction

For Model 3, all variables are maintained from previous Models, and an interaction term between gender and the complexity index is added. Testing the relationship between gender and complexity furthers the research in Chapter 2 about how gender and fragmentation relate to each other.

Table 4.7 Model 3 Results (baseline + socioeconomic + attitudes + interaction)

	Est. by Class				
Row Labels	**Transit Users (885)**	**Car Passengers (321)**	**Diverse Mode Users (1,244)**	**Walkers (719)**	**Non-Solitary Drivers (1,486)**
Sequence: Home Day	-0.709	**-1.332**	**-1.760**	-0.371	-0.242
Sequence: Typical Work Day	0.250	**-2.146**	**-2.298**	-0.572	**-1.387**
Sequence: School Day	-0.248	**-2.545**	**-1.684**	0.312	**-1.399**
Sequence: Errands Day	0.358	**-1.232**	**-1.256**	-0.496	**-0.670**
Sequence: Atypical Work Day	-0.413	**-1.794**	**-2.338**	**-0.934**	**-1.574**
Has Driver's License	**-5.597**	**-5.857**	**-3.600**	**-5.142**	-0.383
At Least 1 Vehicle per Adult with a Driver's License	**-2.097**	**-1.548**	**-1.403**	**-2.212**	**-0.758**
Race: White	0.144	-0.372	**0.834**	-0.055	0.142
Race: Asian	0.200	-0.249	**0.777**	-0.101	0.165
Race: Hispanic	0.209	0.312	**0.878**	0.159	**0.487**
Race: Black	-0.193	-0.310	0.399	**-1.101**	-0.062
Age 18–34	0.288	0.215	**1.541**	**0.521**	**0.747**
Age 35–64	0.146	0.004	**0.915**	0.317	**0.503**
Number of Adults in Household	**-0.529**	**0.429**	**-0.415**	**-0.767**	**0.226**
Female	**0.812**	**2.198**	**0.826**	0.495	0.124
Worker	**-0.693**	**-0.981**	0.247	**-0.360**	**-0.488**
Income Below the SSS	**0.433**	**-0.781**	**-0.599**	0.137	**-0.495**
Minors Aged 00–04 in Household	-0.167	**0.644**	**0.930**	0.154	**1.151**
Minors Aged 05–15 in Household	0.212	0.252	**1.010**	**0.476**	**1.428**
Minors Aged 16–17 in Household	**1.289**	0.454	**1.077**	**0.960**	**0.734**
Complexity	4.087	10.134	**31.661**	**-26.069**	**31.021**
Only Uses Car	**-4.837**	**-0.296**	**-0.842**	**-1.469**	**0.384**
Use Transit More: Safer Ways to Get to Stops	**1.658**	**0.665**	**0.296**	-0.016	0.075
Use Transit More: Increased Frequency	0.320	**-0.547**	0.071	-0.317	0.077
Use Transit More: Increased Reliability	**0.708**	0.195	**0.751**	0.343	0.036
Use Bike More: Shared use path or protected bike lane	0.189	0.191	0.124	-0.042	0.025
Use Bike More: Neighborhood Greenway	-0.085	-0.372	0.152	0.139	0.072
Use Bike More: Bike Lane	0.056	0.044	0.153	0.113	0.099
Use Bike More: Shared Roadway Lane	-0.233	0.290	**-0.397**	**0.543**	-0.115
Use Bike More: End of Trip Amenities	-0.052	-0.338	0.171	0.040	-0.023
Home Choice: Reasonably Short Commute to Work	0.209	-0.201	0.186	**0.278**	0.018
Home Choice: Affordability	**-0.497**	-0.159	0.000	**-0.778**	-0.100
Home Choice: Being Close to Family or Friends	-0.187	0.020	-0.157	-0.157	0.010
Home Choice: Being Close to the Highway	**-0.583**	-0.133	**-0.391**	**-0.730**	0.016
Home Choice: Quality of Schools (K-12)	**-0.324**	-0.167	**-0.407**	**-0.345**	0.130
Home Choice: Space & Separation from Others	-0.103	**0.289**	**-0.195**	**-0.242**	0.075
Home Choice: Close to Public Transit	**0.946**	-0.032	**0.380**	**0.490**	-0.009
Home Choice: Walkable Neighborhood, Near Local Activities	0.115	0.218	**0.644**	**0.860**	0.048
Interaction: Complexity & Female	**-16.293**	**-23.599**	**-13.544**	-12.577	-3.989

Note. Reference class for mode user class is *Solitary Drivers*. Reference class for sequence patterns is *Travel Day*.

Goodness of Fit Measures: Loglikelihood: -8,677.783; AIC: 17,755.57; BIC: 19,121.03; aBIC: 18,485.48; Entropy: 0.863

In Model 3, identifying as female corresponds to higher likelihoods of being *Transit Users, Car Passengers,* or *Diverse Mode Users* than of being *Solitary Drivers*. As shown in Table 4.7, the interaction terms between *Complexity* and *Female* for these three classes have large negative coefficients, meaning when women's schedules are less complex, they are more likely to be *Transit Users*, *Car Passengers*, or *Diverse Mode Users*. When women's schedules are more complex, they are more likely to be *Non-Solitary Drivers* or *Solitary Drivers*. It can be inferred that men (and the small proportion of respondents who did not identify as either) with less complex schedules are more likely to drive alone or to drive others. This suggests that women are taking on more complex schedules when they are in a car, which could be tied to having more household responsibilities. To further investigate this relationship between complexity and household responsibilities, more interaction terms with gender and/or Complexity could be added to a future model, such as interactions with presence of children in the household or the *Errands Day* sequence.

4.4 Discussion

It is worth noting a few relationships of particular interest for transportation policies. Car availability captured by the variable *At Least 1 Vehicle per Adult with a Driver's License* clearly shows that driving alone is significantly impacted as the household car fleet size increases. When the impact of this variable is compared with the other two related variables of *Has a Driver's License* and *Only Uses Car,* the policy implication is that unless we find ways to inhibit widespread car ownership (e.g., by adding fuel taxation and registration fees, increasing curbside and garage parking costs, and limiting access to cars in the center of cities), people will continue purchasing cars, that in turn motivates them to use them as single occupancy vehicles (SOVs). The second major finding in this analysis is the relationship between bicycle desired level of service and membership in mode class. In essence, this relationship is inexistent based on this model. There are not enough people riding bicycles in this sample for a unique class of cyclists to emerge until the seven-class LCA model. This means all the improvements in infrastructure for bicycles are less likely to impact daily mode choice and more likely to impact a small number of people that are practicing active transportation such as the *Walkers*.

The classes using transit and walking tend to say that when they chose their homes, it was important to them that they were in walkable neighborhoods and near public transit. This shows that the people in this study who were using more sustainable travel modes intended to use those modes before even living in their current residences. The implication of this is that improving the infrastructure aimed at reducing the need for SOVs in an area may not be enough to convince current residents to change their modes of travel. However, it does imply that improving infrastructure would attract new people who are interested in using it to move there.

What Model 3 adds to the discussion about gender and fragmentation: When women are driving cars, they are taking on more complex schedules than when men are in their cars. Although in this analysis there was not an opportunity to investigate the reason for this further, it does reinforce the results found in Chapter 2 and add nuance to the conversation: Women's schedules are more complex than men's schedules, but only under certain conditions.

This experiment tested a portion of the conceptual model, although direct conclusions cannot be drawn about the viability of the model from the results because there is a vital piece missing from the PSTS: high-quality measurements of the mental processes of respondents that influence behavior. This heavily impacts how well this study can test the hypothesized relationships of the theoretical model. The theoretical model hypothesizes that behavior is directly controlled by the mental processes intention and habit, while conditions are present that affect how well the mental processes can successfully translate into behavior. Conditions include both personal characteristics and external circumstances (e.g., demographics, SES, spatiotemporal structures, access to modes, policies), and they moderate the effects of intentions and habits on behavior. There was no adaptable measurement of affect (mood), nothing that worked for social factors, a poor representation of habit, and only part of attitudes could be measured. Some variables were usable for measuring part of the *Attitudes* construct of the conceptual model, but none were high quality. Most of the available measurements would be considered in the theoretical model to be part of the *Context/Facilitating Conditions*, and these were measured for their direct relationship to travel behavior, while in the theoretical model they would not be considered to directly affect behavior.

With the lack of measuring mental processes, there was not enough to build strong Intention or Habit constructs, so there were not really constructs to apply the moderation to as proposed by the conceptual model. A simpler structure was tested: a multinomial logistic regression tests if there are relationships between the concepts and the travel behavior as it is.

Evidence in favor of hypothesis: Hypothesis is that all these things affect travel behavior. The findings are that basically everything that I tested did influence mode choice to some degree, except for some of the *Use Bike More* variables. This supports the hypothesis that all these things are important for mode choice. Moreover, by testing the models in a nested fashion, it showed how there are interrelationship between the variables, and that direct relationships were obscured because of the unmeasured relationships expressing themselves through the included variables. This supports the use of a model structure that accounts for the interrelationships among explanatory variables, like a moderation.

4.5 Conclusion

This study had two goals: to build upon the research in Chapter 2 about gender and schedule fragmentation, and to test the conceptual model that forms the backbone of this dissertation. For the goal of adding to the research about gender and schedule fragmentation, the study was successful. For the goal of testing the conceptual model, this study was partially successful. The variables that were available and adaptable to represent constructs of the conceptual model were tested for whether they had a relationship with travel behavior, and the tests showed that they did. However, without access to sufficient measurements of mental processes, it was not possible to do any sort of moderation analysis, which is a key component of my conceptual model. The variables used in this study would mostly go into the category called *Context/Facilitating Conditions,* which is supposed to moderate the direct relationships of Habit and Intention on Behavior. Many variables had to be adapted, losing information along the way, because of the structure that was imposed upon this existing dataset. Although some measurement of values and beliefs was possible through the *Home Choice*, *Use Transit More*, and *Use Bike More* questions, this is only a part of what should go into measuring attitude. The variables used to measure habit only satisfy one of two qualities essential to defining a habit: frequent use. The second quality of habits is that they are done without much conscious thought or decision-making (Gardner, 2012, 2015; Verplanken & Orbell, 2003), but there was not access to this information in the PSTS. Because the variables available were limited in a few ways for measuring habit, they were simplified heavily into a binary variable: whether respondents only use their car, or they have other regular modes of travel they use (*Only Uses Car*). Travel behavior variables used to create the latent classes were also adapted with what was available, and they are not a complete measurement of behavior. Only mode(s) chosen on the travel diary day. It does not measure behavior over longer periods besides one specific day, and there is limited information about which modes respondents used over weekly or monthly periods.

In the future, the results of this research would be enhanced by further investigating why women have more complex schedules than men when they are in their cars. This would involve looking at the trip purposes of men versus women in more detail, accounting for whether they are driving cars for those trips. Another future improvement to this research would be to create an effective scoring method for travel behavior. Measuring mode-wise behavior with a score for each mode of travel (e.g., a score for how much of a “walker,” “transit user,” or “driver” someone is) would be an improvement to measuring travel behavior over simple mode choice. This scoring system should incorporate the variables that matter in travel behavior research: mode choice, number of trips by mode, and distance traveled by mode (with different weights for different modes since speed of travel really influences this). Of course, the most valuable future direction for this research would be to conduct a survey that directly measures the variables of interest to fully assess the conceptual model, because measurement of mental processes is nearly always lacking in travel surveys.

5 Conclusions and Synthesis

First, a comprehensive theoretical model was developed. The behavioral determinants in this conceptual formulation were extracted from research in social psychology and applications in travel behavior. These include behavioral intentions (underlined by attitudes, social factors, and affect), habit, and context/facilitating conditions. This model was used as the framework around which the research in this dissertation was done, and many of its aspects are used to answer research questions customized to the problem addressed.

In Chapter 2, fragmentation of activities and travel is defined as the multiple sequencing of many relatively short activities and trips that happen in a person's daily schedule. These are combined with much longer activities and travel to form a complete schedule of activities and travel by each observed individual. Fragmentation of activity-travel schedules may lead to increased transport demand because many activities, enabled by mobile communication technologies and other societal innovations, are no longer bound to specific times and specific places. The main objective in Chapter 2 was to close the research gap in understanding how and why individuals engage in activity-travel fragmentation. Studying the correlation of activity and travel fragmentation with social interaction and accessibility offered by the environment in which people live can close this research gap. This enables distinguishing between people that face social exclusion and the dichotomy between women spending more time in the private sphere, and less in the public one – and vice versa for men. A secondary objective that enabled the analysis was to develop robust statistical methods for fine grained spatio-temporal data to improve travel demand forecasting models.

To study fragmentation, a new method of sequence analysis in activity participation was used. The methods were first tested on participants in the California Household Travel Survey 2012-2013 (CHTS) from the Central Coast of California (San Louis Obispo and Santa Barbara Counties), and the best indicators were identified. The method explored sequences of daily activity and travel, employing techniques from the sequencing of events in the life course of individuals but transferred in the minute-by-minute analysis of activity and travel. Studying sequences of daily episodes (each activity and each trip) are preferable over other techniques of studying activity-travel behavior because sequences include the entire trajectory of a person's activity during a day, while at the same time jointly considering the number of activities, order of activities in a day, and their durations. Substantial fragmentation in activity participation was found among persons with children and in specific age groups (25 to 65) amplified by the presence of children in the household. Poverty was found to play an important inhibiting role. Examinations of the days of the week showed significant and substantial differences among the different days of a week, with both Sundays and Saturdays being different but also with substantial differences among the weekdays.

This analysis was refined and repeated using a statewide sample of 12,704 persons that also participated in CHTS, and nine distinct daily patterns were obtained. These included patterns of people staying at home for long periods in a day and people following typical daily weekday working and school schedules. Patterns of people traveling for an entire day and people staying at home in the morning but then traveling for the rest of their interview day were also found. There were also two patterns of running errands with different time of day rhythms. The ninth pattern was people spending most of the day at locations besides home, work, or school and traveling for a very short time in a day. Each pattern also had different memberships in terms of gender, age, and day of the week (in addition to the working and/or student status as expected). Each of the days of the weekend had a different mix of daily patterns. In regional transportation planning, policy analysis is done using simulation models of the daily behavior of people. To do this, a typical workday is used to represent behavior on any day of the week. The finding here of major differences across all seven days of the week challenges this practice and points to the need for week-long behavioral simulation. This not only would capture the day-to-day variation, but also allow expansion of the analysis to study the shifting of activities from one day to another as policies are implemented. In addition to systematic differences among workers and students, systematic differences were also found in time-of-day patterns between males and females and between age groups. Evidence was also found of higher fragmentation by center city dwellers, but this was different across the daily patterns derived in this analysis. Moreover, to enjoy the same amount of time in activities, rural and exurban residents tended to need to spend longer times travelling than their counterpart suburban and center city dwellers.

The effect of children on schedule fragmentation was substantial, with parents having by far higher fragmentation in their schedules than other adults employed or otherwise. Women in households with children, even when they were not employed, also had fragmented schedules. Employed women in households with children had even more fragmented schedules. All this conforms to the household responsibility hypothesis that when proposed for the first time pointed out that women take on more household and child rearing tasks than men even when they are also full time employed.

In Chapter 3, the conceptual model was used to guide the design of a survey customized for COVID-19 to collect data about its impact and to strategically test selected correlations between attitudes and mode choice using behavioral clusters.

The COVID-19 survey asked respondents about their work, school, and travel behavior before and during the COVID-19 restrictions. There were also a few questions about people's predictions for how their work behavior might change once restrictions are lifted. For employment, they were asked about employment status, number of workdays, how often they worked from home, and how often they participated in online meetings. In terms of travel behavior, respondents were asked what travel modes they used to commute, then they were asked to estimate the distance and time from their homes to work/school by each mode. Everyone, regardless of work/school status, was asked what means of transportation they used for all their trips. They were then asked how many trips they estimated making in a typical week by each of those modes. Respondents were asked to provide the city they live in, ZIP code, household income, gender, and age. A unique aspect of this survey was the questions about whether people moved because of COVID-19. Many people changed home locations during this pandemic, either temporarily or permanently. Some moved in with friends or family to get their social needs met, others were obligated to move to take care of family members, or to separate households when some members might be exposed to COVID-19. Data analysis of this survey and in terms of the attitude-behavior relationship also confirmed the existence of more diversity in attitudinal groups of people with respect to their position towards the private automobile and found that these attitudes were strongly correlated with the use of different modes. The survey design and conceptual model formed the foundation for future data collection and analysis based on the examples of this project.

During this study, an experiment was also completed on the design of a smartphone application. This included guidelines for survey design, survey contents, tradeoffs among different options, and coding options and testing. This was part of an ongoing research effort at UCSB that continues beyond this dissertation to complement the survey described above and support data collection based on the conceptual behavioral model in Chapter 1.

In Chapter 4, the data analysis method of sequence analysis from Chapter 2 is combined with the latent class analysis (LCA) method of Chapter 3 into a final study that further tests a few hypotheses from the conceptual model in Chapter 1. The circumstances caused by COVID-19 did not allow for the collection of the originally planned primary data, so the 2017 and 2019 Puget Sound Regional Council Transportation Study (PSTS) data were used as a proxy of a database that could have been collected. The PSTS has a travel diary, demographic information, and a set of attitudinal questions, so it covers more of the conceptual model than typical travel diary surveys, which rarely include anything about the cognitive processes that might influence travel behavior.

Using the modes PSTS respondents reported using on their travel diary day as the basis for creating classes of travel behavior, the LCA resulted in six latent classes of behavior: *Transit Users*, *Car Passengers*, *Diverse Mode Users*, *Solitary Drivers*, *Walkers*, and *Non-Solitary Drivers*. Using the sequence analysis methods and PSTS diary data, six patterns of behavior were derived: *Home Day*, *Typical Work Day*, *School Day*, *Errands Day*, *Atypical Work Day*, and *Travel Day*.

Then, the six patterns of time allocation and many other variables were explored for their correlations with the six latent classes of travel behavior using multinomial regression. As expected, the modes selected by people and the time allocation patterns are closely related but with substantial heterogeneity among respondents. The Complexity index, used in Chapter 2 as a measure of schedule fragmentation, was also included as an auxiliary variable. To further the research of Chapter 2 looking at the relationship between gender and fragmented schedules, an interaction effect was examined between the complexity index and gender, and the findings showed that the relationship between Complexity and gender has some nuance: only women driving cars to get around showed significantly higher likelihood of having more fragmented schedules than men (and the few respondents who did not identify as either) driving their cars.

5.1 Synthesis

The analysis in this dissertation shows we can identify and measure social exclusion due to roles and responsibilities that lead to time poverty as Lucas (2012) describes in her review and links to transportation service provision. This analysis also demonstrates that many hypotheses can be tested using already collected data and we can even reach conclusions about possible correlations and policy actions. However, it also reveals gaps in our knowledge and limitations in the types of analyses we can do.

The research process of this dissertation has revealed a big problem within the travel behavior research field: the typical travel diary survey, which is the most detailed way to look at travel behavior, rarely includes sufficient investigation of the cognitive processes that influence travel. They usually look almost exclusively at the actual behaviors performed (also called revealed preference in the travel behavior literature), without any probing into *why* respondents behave the way they do. In part, this is because these travel diary surveys are used most often in the branch of the transportation research field looking at trends, making predictions, and making inferences. This is different from the travel behavior research branch focusing on human behavior and exploring not only motivations but also cognitive biases.

A large part of the transportation research field is dedicated to how to solve the "car problem." There are too many cars on the roads, and the volume of cars is getting higher, especially in urban areas. The congestion, pollution, and greenhouse gas emissions from cars are problematic for cities: they affect the health of citizens, access to opportunities (e.g., too much traffic to get to better workplaces), preservation for future generations, and the economy of the city (e.g., if people leave urban areas). If researchers want to help policymakers change travel behavior on a large scale, they need to understand what is going on internally in people that makes those behaviors happen. Understanding why respondents act the way they do requires studying cognition and the conditions that either facilitate or hinder the relationship between cognitive processes and behavior.

6 References

Abbott, A., & Tsay, A. (2000). Sequence analysis and optimal matching methods in sociology: Review and prospect. *Sociological Methods and Research*, *29*(1), 3–33. https://doi.org/10.1177/0049124100029001001

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*. https://doi.org/10.1016/0749-5978(91)90020-T

Asparouhov, T., & Muthén, B. O. (2012). Using Mplus TECH11 and TECH14 to test the number of latent classes. *Mplus Web Notes*, *14*, 1–17.

Asparouhov, T., & Muthén, B. O. (2014). Auxiliary variables in mixture modeling: Three-step approaches using Mplus. *Structural Equation Modeling*, *21*(3), 329–341. https://doi.org/10.1080/10705511.2014.915181

Assemi, B., Jafarzadeh, H., Mesbah, M., & Hickman, M. (2018). Participants' perceptions of smartphone travel surveys. *Transportation Research Part F: Traffic Psychology and Behaviour*, *54*, 338–348. https://doi.org/10.1016/J.TRF.2018.02.005

Auld, J., Rashidi, T. H., Javanmardi, M., & Mohammadian, A. (2011). Dynamic activity generation model using competing hazard formulation. *Transportation Research Record: Journal of the Transportation Research Board*. https://doi.org/10.3141/2254-04

Bamberg, S., & Schmidt, P. (2003). Incentives, morality, or habit? Predicting students' car use for University routes with the models of Ajzen, Schwartz, and Triandis. *Environment and Behavior*. https://doi.org/10.1177/0013916502250134

Beckman, J. D., & Goulias, K. G. (2008). Immigration, residential location, car ownership, and commuting behavior: A multivariate latent class analysis from California. *Transportation*, *35*(5), 655–671. https://doi.org/10.1007/s11116-008-9172-x

Ben-Akiva, M. E., de Palma, A., McFadden, D., Abou-Zeid, M., Chiappori, P. A., de Lapparent, M., Durlauf, S. N., Fosgerau, M., Fukuda, D., Hess, S., Manski, C., Pakes, A., Picard, N., & Walker, J. (2012). Process and context in choice models. *Marketing Letters*, *23*(2), 439–456. https://doi.org/10.1007/s11002-012-9180-7

Ben-Akiva, M. E., Mcfadden, D., Abe, M., Böckenholt, U., Bolduc, D., Gopinath, D., Morikawa, T., Ramaswamy, V., Rao, V., Revelt, D., & Steinberg, D. (1997). Modeling methods for discrete choice analysis. *Marketing Letters*, *8*(3), 273–286. https://doi.org/10.1023/A:1007956429024

Bhat, C. R., Goulias, K. G., Pendyala, R. M., Paleti, R., Sidharthan, R., Schmitt, L., & Hu, H.-H. (2013). A household-level activity pattern generation model with an application for Southern California. *Transportation*. https://doi.org/10.1007/s11116-013-9452-y

Bhat, C. R., & Pinjari, A. R. (2000). Duration modeling. In *Handbook of transport modelling* (6th ed., pp. 91–111).

Bohannon, R. W. (1997). Comfortable and maximum walking speed of adults aged 20-79 years: Reference values and determinants. *Age and Ageing*, *26*(1), 15–19. https://doi.org/10.1093/ageing/26.1.15

California Air Resources Board. (2019). California greenhouse gas emissions for 2000 to 2017. *California Greenhouse Gas Emission Inventory: 2000 – 2017*.

Celeux, G., & Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification*, *13*(2), 195–212.

Cervero, R. (2003). Road expansion, urban growth, and induced travel: A path analysis. *Journal of the American Planning Association*. https://doi.org/10.1080/01944360308976303

Chen, Y., Ravulaparthy, S., Deutsch, K., Dalal, P., Yoon, S. Y., Lei, T., Goulias, K. G., Pendyala, R. M., Bhat, C. R., & Hu, H.-H. (2011). Development of indicators of opportunity-based accessibility. *Transportation Research Record: Journal of the Transportation Research Board*, *2255*(2), 58–68. https://doi.org/10.3141/2255-07

Clark, E., Fredricksid, K., Woc-Colburn, L., Bottazzi, M. E., & Weatherheadid, J. (2020). Disproportionate impact of the COVID-19 pandemic on immigrant communities in the United States. *PLoS Neglected Tropical Diseases*, *14*(7), 1–9. https://doi.org/10.1371/journal.pntd.0008484

Couclelis, H. (2000). From sustainable transportation to sustainable accessibility: Can we avoid a new tragedy of the commons? In *Information, Place, and Cyberspace* (pp. 341–356). Springer.

Couclelis, H. (2004). Pizza over the internet: E-commerce, the fragmentation of activity and the tyranny of the region. *Entrepreneurship and Regional Development*. https://doi.org/10.1080/0898562042000205027

Crane, R. (2007). Is there a quiet revolution in women's travel? Revisiting the gender gap in commuting. *Journal of the American Planning Association*. https://doi.org/10.1080/01944360708977979

de Groot, J. I. M., & Steg, L. (2009). Mean or green: which values can promote stable pro-environmental behavior? *Conservation Letters*, *2*(2), 61–66. https://doi.org/10.1111/j.1755-263x.2009.00048.x

Duranton, G., & Turner, M. A. (2011). The fundamental law of road congestion: Evidence from US cities. *American Economic Review*. https://doi.org/10.1257/aer.101.6.2616

Elzinga, C. H., & Liefbroer, A. C. (2007). De-standardization of family-life trajectories of young adults: A cross-national comparison using sequence analysis. *European Journal of Population / Revue Européenne de Démographie*, *23*(3–4), 225–250. https://doi.org/10.1007/s10680-007-9133-7

Eriksson, J., Forsman, Å., Niska, A., Gustafsson, S., & Sörensen, G. (2019). An analysis of cyclists' speed at combined pedestrian and cycle paths. *Traffic Injury Prevention*, *20*(sup3), 56–61. https://doi.org/10.1080/15389588.2019.1658083

Ettema, D., Borgers, A., & Timmermans, H. (1995). Competing risk hazard model of activity choice, timing, sequencing, and duration. *Transportation Research Record*, *1493*, 101–109.

Ferreira, L. N., Pereira, L. N., da Fé Brás, M., & Ilchuk, K. (2021). Quality of life under the COVID-19 quarantine. *Quality of Life Research*, *30*(5), 1389–1405. https://doi.org/10.1007/s11136-020-02724-x

Friman, M., Ettema, D., & Olsson, L. E. (2018). *Quality of life and daily travel*. Springer.

Gabadinho, A., Ritschard, G., Müller, N. S., & Studer, M. (2011). Analyzing and visualizing state sequences in R with TraMineR. *Journal of Statistical Software*, *40*(4), 1–37. https://doi.org/10.18637/jss.v040.i04

Gabadinho, A., Ritschard, G., & Studer, M. (2010). Indice de complexité pour le tri et la comparaison de séquences catégorielles. *Revue Des Nouvelles Technologies de l'Information*.

Gabadinho, A., Ritschard, G., Studer, M., & Muller, N. S. (2010). Mining sequence data in R with the TraMineR package: A user's guide. *Department of Econometrics and Laboratory of Demography, University of Geneva, Switzerland*.

Gao, Y., Rasouli, S., Timmermans, H., & Wang, Y. (2018). Trip stage satisfaction of public transport users: A reference-based model incorporating trip attributes, perceived service quality, psychological disposition and difference tolerance. *Transportation Research Part A: Policy and Practice*, *118*, 759–775. https://doi.org/10.1016/j.tra.2018.10.029

Gardner, B. (2012). Habit as automaticity, not frequency. *European Health Psychologist*, *14*(2), 32–36.

Gardner, B. (2015). A review and analysis of the use of 'habit' in understanding, predicting and influencing health-related behaviour. *Health Psychology Review*, *9*(3), 277–295. https://doi.org/10.1080/17437199.2013.876238

Gärling, T. (2020). Travel behavior and psychology: Life time achievement 1982–2018. In K. G. Goulias & A. W. Davis (Eds.), *Mapping the Travel Behavior Genome* (pp. 47–61). Elsevier. https://doi.org/10.1016/B978-0-12-817340-4.00004-8

Gkargkavouzi, A., Halkos, G., & Matsiori, S. (2019). Development and validation of a scale for measuring Multiple Motives toward Environmental Protection (MEPS). *Global Environmental Change*, *58*, 101971. https://doi.org/10.1016/j.gloenvcha.2019.101971

Goulias, K. G. (1999). Longitudinal analysis of activity and travel pattern dynamics using generalized mixed Markov latent class models. *Transportation Research Part B: Methodological*. https://doi.org/10.1016/S0191-2615(99)00005-3

Goulias, K. G., McBride, E. C., & Su, R. (2020). Life cycle stages, daily contacts, and activity-travel time allocation for the benefit of self and others. In *Mobility and Travel Behaviour Across the Life Course* (pp. 206–220). Edward Elgar Publishing. https://doi.org/10.4337/9781789907810.00023

Guagnano, G. A., Stern, P. C., & Dietz, T. (1995). Influences on attitude-behavior relationships: A natural experiment with curbside recycling. *Environment and Behavior*, *27*(5), 699–718. https://doi.org/10.1177/0013916595275005

Hägerstrand, T. (1970). What about people in regional science? *Regional Science Association*, *24*(1), 6–21. https://doi.org/10.1111/j.1435-5597.1970.tb01464.x

Harding, C., Faghih Imani, A., Srikukenthiran, S., Miller, E. J., & Nurul Habib, K. (2020). Are we there yet? Assessing smartphone apps as full-fledged tools for activity-travel surveys. *Transportation 2020*, 1–28. https://doi.org/10.1007/S11116-020-10135-7

Hensher, D. A. (1994). Stated preference analysis of travel choices: the state of practice. *Transportation 1994 21:2*, *21*(2), 107–133. https://doi.org/10.1007/BF01098788

Hoehner, C. M., Brennan Ramirez, L. K., Elliott, M. B., Handy, S. L., & Brownson, R. C. (2005). Perceived and objective environmental measures and physical activity among urban adults. *American Journal of Preventive Medicine*, *28*(2 SUPPL. 2), 105–116. https://doi.org/10.1016/j.amepre.2004.10.023

Hungerford, H. R., & Volk, T. L. (1990). Changing learner behavior through environmental education. *Journal of Environmental Education*. https://doi.org/10.1080/00958964.1990.10753743

Hymel, K. M., Small, K. A., & Dender, K. Van. (2010). Induced demand and rebound effects in road transport. *Transportation Research Part B: Methodological*. https://doi.org/10.1016/j.trb.2010.02.007

IDRE Research Technology Group. (2013). *R data analysis examples: multinomial logistic regression*. UCLA Institute for Digital Research and Education. https://stats.idre.ucla.edu/mplus/dae/multinomiallogistic-regression/

Jensen, W. A., Brown, B. B., Smith, K. R., Brewer, S. C., Amburgey, J. W., & McIff, B. (2017). Active transportation on a complete street: Perceived and audited walkability correlates. *International Journal of Environmental Research and Public Health*, *14*(9), 1014. https://doi.org/10.3390/ijerph14091014

Joh, C. H., Arentze, T. A., & Timmermans, H. J. P. (2001). Multidimensional sequence alignment methods for activity-travel pattern analysis: A comparison of dynamic programming and genetic algorithms. *Geographical Analysis*, *33*(3), 247–270.

Kaufman, L., & Rousseeuw, P. J. (2009). *Finding groups in data: an introduction to cluster analysis*. John Wiley & Sons.

Kroesen, M. (2014). Modeling the behavioral determinants of travel behavior: An application of latent transition analysis. *Transportation Research Part A: Policy and Practice*, *65*, 56–67. https://doi.org/10.1016/j.tra.2014.04.010

Kroesen, M., & Goulias, K. G. (2016). Modelling activity-travel behaviour dynamics with panel data: The state-of-the-art. *European Journal of Transport and Infrastructure Research*.

Lee, J. H., & Goulias, K. G. (2018). A decade of dynamics of residential location, car ownership, activity, travel and land use in the Seattle metropolitan region. *Transportation Research Part A: Policy and Practice*, *114*, 272–287. https://doi.org/10.1016/j.tra.2018.01.029

Leszczyc, P. T. L. P., & Timmermans, H. (2002). Unconditional and conditional competing risk models of activity duration and activity sequencing decisions: An empirical comparison. *Journal of Geographical Systems*. https://doi.org/10.1007/s101090200083

Lucas, K. (2012). Transport and social exclusion: Where are we now? *Transport Policy*. https://doi.org/10.1016/j.tranpol.2012.01.013

Manski, C. F. (1977). The structure of random utility models. *Theory and Decision*, *8*(3), 229–254. https://doi.org/10.1007/BF00133443

Manville, M., Taylor, B. D., & Blumenberg, E. (2018). *Falling transit ridership: California and Southern California*.

Marcinkowski, T., & Reid, A. (2019). Reviews of research on the attitude–behavior relationship and their implications for future environmental education research. In *Environmental Education Research*. https://doi.org/10.1080/13504622.2019.1634237

Masyn, K. E. (2013). Latent class analysis and finite mixture modeling. In *The Oxford handbook of quantitative methods (Vol 2): Statistical analysis.* (pp. 551–611).

McBride, E. C., Davis, A. W., & Goulias, K. G. (2019). Fragmentation in daily schedule of activities using activity sequences. *Transportation Research Record*, *2673*(4), 844–854. https://doi.org/0361198119837501

McBride, E. C., Davis, A. W., & Goulias, K. G. (2020). Exploration of statewide fragmentation of activity and travel and a taxonomy of daily time use patterns using sequence analysis in california. *Transportation Research Record*, *2674*(12), 38–51. https://doi.org/10.1177/0361198120946011

McBride, E. C., Davis, A. W., Goulias, K. G., & University of Southern California, L. A. (2019). *An analysis of accessibility, social interaction, and activity-travel fragmentation in California*. United States. Dept. of Transportation. Office of the Assistant Secretary for Research and Technology. https://doi.org/10.21949/1503647

McBride, E. C., Davis, A. W., Lee, J. H., & Goulias, K. G. (2017). Incorporating land use in synthetic population generation methods and transfer of behavioral data. *Transportation Research Record: Journal of the Transportation Research Board*, *2668*, 11–20. https://doi.org/10.3141/2668-02

McBride, E. C., Lee, J. H., Lundberg, A. M., Davis, A. W., & Goulias, K. G. (2016). Behavioural micro-dynamics of car ownership and travel in the Seattle Metropolitan Region from 1989 to 2002. *European Journal of Transport and Infrastructure Research*, *16*(4).

Milfont, T. L., & Duckitt, J. (2010). The environmental attitudes inventory: A valid and reliable measure to assess the structure of environmental attitudes. *Journal of Environmental Psychology*, *30*(1), 80–94. https://doi.org/10.1016/j.jenvp.2009.09.001

Mokhtarian, P. L., Salomon, I., & Singer, M. E. (2015). What moves us? An interdisciplinary exploration of reasons for traveling. *Transport Reviews*. https://doi.org/10.1080/01441647.2015.1013076

Mokhtarian, P. L., Samaniego, F. J., Shumway, R. H., & Willits, N. H. (2002). Revisiting the notion of induced traffic through a matched-pairs study. *Transportation*. https://doi.org/10.1023/A:1014221024304

Molin, E., Mokhtarian, P. L., & Kroesen, M. (2016). Multimodal travel groups and attitudes: A latent class cluster analysis of Dutch travelers. *Transportation Research Part A: Policy and Practice*, *83*, 14–29. https://doi.org/10.1016/j.tra.2015.11.001

Montello, D. R., & Sutton, P. C. (2013). *An Introduction to Scientific Research Methods in Geography and Environmental Studies* (2nd ed.). Thousand Oaks, CA and London: SAGE Publications. https://doi.org/10.1111/gean.12022

Murakami, E., & Waterson, W. T. (1990). Developing a household travel panel survey for the Puget Sound region. *Transportation Research Record*, *1285*, 7.

Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén. https://doi.org/10.1111/j.1600-0447.2011.01711.x

National Association of City Transportation Officials. (2016). Designing to move people. In *Transit Street Design Guide*. Island Press.

NUSTATS. (2013). *2010-2012 California Household Travel Survey Final Report*.

Nylund-Gibson, K., & Choi, A. Y. (2018). Ten frequently asked questions about latent class analysis. *Translational Issues in Psychological Science*, *4*(4), 440–461. https://doi.org/10.1037/tps0000176

Nylund-Gibson, K., Grimm, R., Quirk, M., & Furlong, M. (2014). A latent transition mixture model using the three-step specification. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*(3), 439–454. https://doi.org/10.1080/10705511.2014.915375

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (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. https://doi.org/10.1080/10705510701575396

Office of Transportation and Air Quality. (2019). *Fast Facts U.S. Transportation Sector Greenhouse Gas Emissions: 1990-2017*.

Pearce, D. M. (2017). *The self-sufficiency standard for Washington State 2017*.

Přibyl, O., & Goulias, K. G. (2005). Simulation of daily activity patterns incorporating interactions within households: Algorithm overview and performance. *Transportation Research Record: Journal of the Transportation Research Board*, *1926*(1). https://doi.org/10.3141/1926-16

Ramsey, C. E., & Rickson, R. E. (1976). Environmental knowledge and attitudes. *The Journal of Environmental Education*, *8*(1), 10–18.

https://doi.org/10.1080/00958964.1976.9941552

Regional Household Travel Survey underway | Puget Sound Regional Council. (n.d.). Retrieved November 27, 2021, from https://www.psrc.org/whats-happening/blog/regional-household-travel-survey-underway

Ritschard, G. (2021). Measuring the nature of individual sequences. *Sociological Methods and Research*.

Robinson, J. (2010). *Triandis' Theory of Interpersonal Behaviour in understanding software piracy behaviour in the South African context*. University of the Witwatersrand, Johannesburg.

RSG, & Westgroup Research. (2020). *2019 Puget Sound regional travel study*.

Russell, J. A., & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology*, *76*(5), 805–819. https://doi.org/10.1037/0022-3514.76.5.805

Sandy, C. J., Gosling, S. D., Schwartz, S. H., & Koelkebeck, T. (2017). The development and validation of brief and ultrabrief measures of values. *Journal of Personality Assessment*, *99*(5), 545–555. https://doi.org/10.1080/00223891.2016.1231115

Schwartz, S. H. (2003). A proposal for measuring value orientations across nations. In *Questionnaire package of the european social survey* (pp. 259–319).

Southern California Association of Governments. (2019a). *Connect SoCal: Transportation System Congestion Management*.

Southern California Association of Governments. (2019b). *The 2020-2045 Regional Transportation Plan/Sustainable Communities Strategy of the Southern California Association of Governments: Connect SoCal*.

Stern, P. C. (2000). Toward a coherent theory of environmentally significant behavior. *Journal of Social Issues*. https://doi.org/10.1111/0022-4537.00175

Tai, D. B. G., Shah, A., Doubeni, C. A., Sia, I. G., & Wieland, M. L. (2021). The disproportionate impact of COVID-19 on racial and ethnic minorities in the United States. In *Clinical Infectious Diseases* (Vol. 72, Issue 4, pp. 703–706). Oxford Academic. https://doi.org/10.1093/cid/ciaa815

Triandis, H. C. (1977). *Interpersonal behavior*. Brooks/Cole Pub. Co.

Turner, T., & Niemeier, D. (1997). Travel to work and household responsibility: New evidence. *Transportation*, *24*(4), 397–419. https://doi.org/10.1023/a:1004945903696

Västfjäll, D., Friman, M., Gärling, T., & Kleiner, M. (2002). The measurement of core affect: A Swedish self-report measure derived from the affect circumplex. *Scandinavian Journal of Psychology*, *43*(1), 19–31. https://doi.org/10.1111/1467-9450.00265

Västfjäll, D., & Gärling, T. (2007). Validation of a Swedish short self-report measure of core affect. *Scandinavian Journal of Psychology*, *48*(3), 233–238. https://doi.org/10.1111/j.1467-9450.2007.00595.x

Vermunt, J. K. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis*, *18*(4), 450–469. https://doi.org/10.1093/pan/mpq025

Verplanken, B., & Orbell, S. (2003). Reflections on past behavior: A self-report index of habit strength. In *Journal of Applied Social Psychology*. https://doi.org/10.1111/j.1559-1816.2003.tb01951.x

Wilson, C. (1998a). Activity pattern analysis by means of sequence-alignment methods. *Environment and Planning A*. https://doi.org/10.1068/a301017

Wilson, C. (1998b). Analysis of travel behavior using sequence alignment methods.

Transportation Research Record. https://doi.org/10.3141/1645-07

Appendix A Survey Sample Questions for Public Transportation Interventions

The following write-up contains initial sample questions and their reasoning consistent with the conceptual model of Figure 1.1 (also reproduced as Figure A.1 here). These are the originally planned questions for the survey of residents living near the Los Angeles Crenshaw/LAX Metro Line. For reader convenience, part of the literature review is repeated here to make the appendix self-standing. Recall that the structure of this model is a hybrid of past theories of behavior, primarily influenced by the Theory of Planned Behavior (Ajzen, 1991), the Theory of Interpersonal Behavior (Triandis, 1977), and the Attitude-Behavior-Context model (Guagnano et al., 1995; Stern, 2000). In the main body of this report, a discussion is provided on a shortened version adapted for COVID-19, and provided here is a literature review and survey question examples that are customized to possible behavioral change in favor of public transportation use.

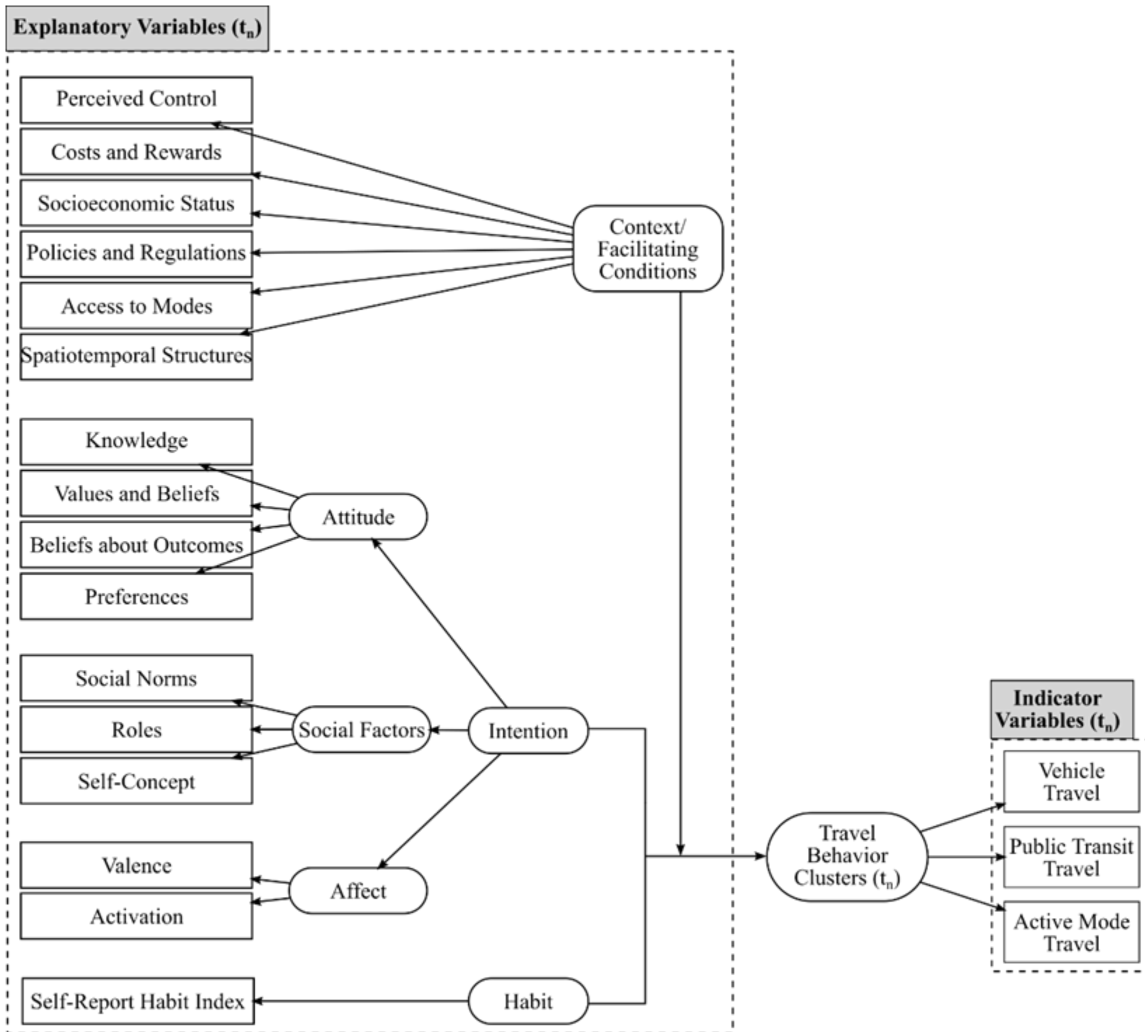


Figure A.1

In the model, the three primary influences on behavior are **intention**, **habit**, and **context/facilitating conditions**. Intention is the primarily conscious motivation to perform a certain behavior. It is shaped by a person's attitudes, social surroundings, and their feelings when they think about performing the action (called "affect"). Habit is a more unconscious driver of behavior. Habit is formed by the repetition of an action and maintained by the low cognitive load of choosing the habitual behavior over a non-habitual option. It can often be at odds with intention. Context/facilitating conditions are moderating conditions that can compel or hinder a behavior. As contextual forces become stronger, their influence over behavior increasingly overpowers the influence of habit and attitudes. Each of these primary constructs have sub-constructs that will be built by direct measurement of relevant variables in the survey. The following write-up discusses the measurement of these sub-constructs, including sample questions for public transportation use.

Intention

The concept of intention as a motivator of behavior originally comes from the Theory of Planned Behavior (TPB). It is also present in the Theory of Interpersonal Behavior (TIB), which built off the TPB. Intention is composed of three sub-constructs: Attitude, Social Factors, and Affect.

Attitude

In the dissertation model, the sub-construct Attitude is built by measuring four key concepts: values, beliefs about outcomes, knowledge about the subject, and preferences. Following is a discussion of each of these concepts with sample questions.

Values. There are two concepts within values that are relevant for measurement: general values and environmental values. Both may contribute to travel behavior intention, but since the Crenshaw/LAX survey will already be difficult to keep short, it is necessary to pare down. A discussion of measuring values in general is included here, including sample questions. However, there will likely only be space to include questions about environmental values, which are more relevant to travel behavior decision-making.

One possibility for measuring values in general is a shortened version of Schwartz's Portrait Value Questionnaire (PVQ), which measures values in ten dimensions: conformity, tradition, benevolence, universalism, self-direction, stimulation, hedonism, achievement, power, and security (Schwartz, 2003). Schwartz later developed and validated two shortened versions of the PVQ: a Twenty Item Value Inventory (TwIVI) and a Ten Item Value Inventory (TIVI) (Sandy et al., 2017). The TIVI, although outperformed by the TwIVI, still meets the acceptable standards for validity and reliability (Sandy et al., 2017).

Values (TIVI)

- Here we briefly describe some people. Please read each description and think about how much each person is or is not like you. Using a 6-point scale from "not like me at all" to "very much like me," choose how similar the person is to you.
- 6 5 4 3 2 1
- very much like me | like me | somewhat | a little | not like | not
- like me | like me | like me | me | at all

- **HOW MUCH LIKE YOU IS THIS PERSON?**

- _____ 1. S/he believes s/he should always show respect to his/her parents and to older people. It is important to him/her to be obedient.
- _____ 2. Religious belief is important to him/her. S/he tries hard to do what his/her religion requires.
- _____ 3. It's very important to him/her to help the people around him/her. S/he wants to care for their well-being.
- _____ 4. S/he thinks it is important that every person in the world be treated equally. S/he believes everyone should have equal opportunities in life.
- _____ 5. S/he thinks it's important to be interested in things. S/he likes to be curious and to try to understand all sorts of things.
- _____ 6. S/he likes to take risks. S/he is always looking for adventures.
- _____ 7. S/he seeks every chance s/he can to have fun. It is important to him/her to do things that give him/her pleasure.
- _____ 8. Being very successful is important to him/her. S/he likes to impress other people.
- _____ 9. It is important to him/her to be in charge and tell others what to do. S/he wants people to do what s/he says.
- _____ 10. It is important to him/her that things be organized and clean. S/he really does not like things to be a mess.

To measure environmental attitudes specifically, it may be difficult to find an instrument that is brief enough to include in full in this survey. All the validated instruments under this topic are too long for our purposes. This is because there are many dimensions to environmental attitudes and values. Most likely, it will be necessary to utilize concepts from these instruments in the creation of a short set of questions.

For specifically measuring environmental attitudes, the Environmental Attitudes Inventory (EAI) from Milfont and Duckitt (2010) may be a good option. There are twelve dimensions measured: enjoyment of nature, support for interventionist conservation policies, environmental movement activism, conservation motivated by anthropocentric concern, confidence in science and technology, environmental threat, altering nature, personal conservation behavior, human dominance over nature, human utilization of nature, ecocentric concern, and support for population growth policies. They give two options: a longer 120-item measure and a shorter, 72-item measure. Both are far too long for the purposes of inclusion in this developing survey; however, it may be possible to utilize the concepts and validated question formatting to create a shortened version of this survey.

Another possibility for measuring this is the scale for measuring Multiple Motives toward Environmental Protection (MEPS) developed by Gkargkavouzi, Halkos, and Matsiori (2019). This is the one in the sample questions. This measures six motives: normative, altruistic, biospheric, egoistic, gain, and hedonic. It also measures constraints to motives. This instrument may be a good option because it integrates some measurement of general values. The MEPS is a 28-item instrument, which is a more reasonable number, but still too large for this survey. This is something that will need to be considered. Perhaps a shortened version can be used without too many issues.

Multi-Motives to Environmental Protection Scale (MEPS)

- **Normative motives**
  - I feel a moral obligation to protect the environment.
  - Don't know; It is not my responsibility to treat nature with respect. (r)
  - The people I care about believe that one ought to protect the environment with his/her actions.
  - Most people who are important to me engage in pro-environmental practices.
- Altruistic motives
  - Good environmental conditions benefit the health of the community and its members.
  - It is urgent to safeguard natural resources for future generations.
  - Environmental degradation has adverse consequences on humanity.
  - Don't know; I am not concerned about the welfare of other people. (r)
- **Biospheric motives**
  - All living organisms have equal intrinsic value.
  - We need to preserve every scrap of biodiversity.
  - Environmental deterioration has adverse consequences on natural ecosystems.
  - Don't' know; I am not concerned about biodiversity loss. (r)
- **Egoistic motives**
  - Nature provides people with food and raw materials.
  - Ecosystems provide recreation and cultural services.
  - A healthy environment is strongly associated with my physical health.
  - Natural areas provide ecosystem services that clean the air and the water.
- **Gain motives**
  - I save money by using public transportation.
  - Government provides monetary subsidies for pro-environmental activities.
  - I gain tax and fees deduction by adopting eco-friendly behaviors.
  - By preserving water and energy, I pay lower utility bills at home.
- **Hedonic motives**
  - I derive pleasure and satisfaction when I engage in environmental behaviors.
  - Don't know; I do not feel any better by protecting the environment. (r)
  - Makes me happy to prevent natural scenery.
  - I enjoy spending time in nature.
- **Constraints to motives**
  - It is expensive to adopt environmental behaviors.
  - It is time-consuming.
  - Needed effort makes the engagement in environmental behaviors difficult. My lifestyle in terms of convenience would change for the worse.

Beliefs about outcomes. This comes from the Theory of Interpersonal Behavior (TIB). It is a person's evaluation of the potential outcomes of their decision. This includes perceptions of safety and convenience. This may be an instrument that need to be designed ex novo, since the measures will be specific to the type of policy explored in the Crenshaw/LAX intervention.

Beliefs about Outcomes

- Traveling by MODE is safe
- Traveling by MODE is comfortable
- Traveling by MODE is convenient
- Traveling by MODE is relaxing

Knowledge. Respondents' knowledge about the options available to them should influence their attitude about those choices. Inclusion of knowledge in this model comes from the Knowledge-Attitude-Behavior (K-A-B) model (Hungerford & Volk, 1990). The sample questions for this section are created by McBride and listed below.

Preferences. Preferences are based on what people like and dislike, which should influence attitude. The sample questions below were written by me to measure this. The challenge of measuring preferences is that they will be influenced by the other building blocks of intention (beliefs about outcomes, values, and lack of knowledge about the other options). Below, MODE consists of "driving my car," "taking the bus," "taking the subway," "riding my bicycle," etc. The main report body also outlines work done in choice modeling on this.

Knowledge

- Please rate your familiarity with the public transportation options accessible to you around your home.
- Please rate your familiarity with the roads around your home.
- Please rate your familiarity with the bicycle lane infrastructure around your home.

Preferences

- **Likes/Dislikes**
  - I prefer to use my car to get around.
  - I love MODE.
  - I hate MODE.
  - If it took a similar amount of time, I would use public transportation to get around.
  - I think MODE is a comfortable way to travel
- **Safety** (this may belong in Perceptions within Context/facilitating conditions)
  - Driving my car is safe
  - Ridng my bicycle is safe
  - Taking public transportation is safe
- **Independence**
  - I do not like to rely on schedules (independence), on others for my movement
  - I prefer the freedom of my own car.

Social Factors

These are the social influences on intention. In the TPB, this is called "subjective norms." They include the social norms, roles, and self-concept. Social norms are the standard behavioral expectations of the society a person lives in. Roles are the positions a person holds in society and/or the household. A person's behavior will be shaped by the responsibilities and expectations that come with those roles. Self-concept is the way a person sees him or herself. This can be aligned with roles and social norms, but also includes personal views about oneself. For example, someone seeing themselves as an environmentally friendly person, or a moral person. This is the construct that has to do with how people see themselves, and the influence of their image. One challenge in measuring these is to separate self-concept from the social norms and roles, especially since these are all going to be self-reported measures. Restructuring the Social Factors section by eliminating the self-image subtopic and integrating it into the remaining two subtopics is under consideration. These sample questions are written ex novo, as validated and succinct instruments to measure these items have not been identified.

Social Norms

- Green self-image as a social norm (Welsch & Kuhling, 2018)
  - The paper does not have the measures used.
- In my neighborhood, people mostly use MODE to get around.
- In my neighborhood, it is normal to use MODE to get around.
- It would be considered weird/odd/abnormal if I used MODE in my neighborhood.

Roles

- I think of myself as a leader in my community.
- I think of myself as a role model for people around me.
- I have people in my life who look up to me.
- There are certain **expectations** I need to live up to.
- I have **responsibilities** that I must maintain.

Self-Concept

- Green self-image (Welsch & Kuhling, 2018)
  - paper does not contain the instruments used
- I see myself as an environmentally-conscious person.

Affect

This is how someone's emotional response to a topic, or their affect, influences their intention. This comes from the Theory of Interpersonal Behavior (Triandis, 1977). Russel and Barrett (1999) define "core affect" as occurring on two dimensions called "valence" and "activation." Valence is described as the spectrum of pleasure to displeasure or good to bad mood. Activation is a person's sense of their energy level: from sleepy to hyperactive. After reviewing validated measures of affect, the Swedish Core Affect Scales (SCAS) (Västfjäll et al., 2002) seems to be an appropriate measure of affect for the purposes of this model. In Västfjäll and Gärling's article validating the SCAS they created, the authors determine that this is a good measure to use for quick assessment of affect in longer surveys, which is perfect for use in a complex model such as this one (Västfjäll & Gärling, 2007). Despite its brevity, the SCAS measures affect on two dimensions: valence (pleasant vs unpleasant affects) and activation (level of arousal). In the sample questions, TRAVELING BY MODE includes "driving your car," "riding in the bus," "riding a subway/rail line," "riding a bicycle," *et cetera*.

Valence

- Please rate your general feelings when you think about TRAVELING BY MODE (scale of 0 to 6 OR -3 to 3)
  - Displeased–Pleased
  - Sad–Glad
  - Depressed–Happy

Activation

- Please rate your general feelings when you think about TRAVELING BY MODE (scale of 0 to 6 OR -3 to 3).
  - Sleepy–Awake
  - Dull–Peppy
  - Passive–Active

Habit

Habit is an influence on behavior that is separate from intention. Inclusion of habit in the model comes from the TIB. Habits are repeated behaviors that, with time, become increasingly automatic and take up less mental energy to do. Because of the low cognitive load of choosing the habitual behavior, it can be difficult to choose a different behavior that will take more conscious effort, even if the intention to do so is there. To measure habit, the Self-Report Habit Index (SRHI) developed by Verplanken and Orbell (2003) seems to be a good option. Their index measures habit based on four attributes: its repetition, lack of awareness and conscious intent, difficulty of avoiding the behavior, and mental efficiency (meaning it has a small cognitive load). Below, where the sample questions say “MODE,” this can be replaced with different modes of travel, like “my vehicle,” “public transit,” or “my bicycle.”

Self-Report Habit Index

- Taking **MODE** to get around is something...
  - I do frequently.
  - I do automatically.
  - I do without having to consciously remember.
  - That makes me feel weird if I do not do it.
  - I do without thinking.
  - That would require effort not to do it.
  - That belongs to my (daily, weekly, monthly) routine.
  - I start doing before I realize I'm doing it.
  - I would find hard not to do.
  - I have no need to think about doing.
  - That's typically "me."
  - I have been doing for a long time.

Context/Facilitating Conditions

The inclusion of context/facilitating conditions as a moderator for habit and intention is influenced by both the ABC model and the TIB. As described in the introduction, these are conditions that can either hinder or compel behaviors. These range from external/structural conditions like proximity of a bus stop or parking policies to personal conditions like socioeconomic status. Briefly, it is worth mentioning that “spatiotemporal structures” are the temporal and spatial constraints to movement in a person’s day. This is a person’s schedule, which adds certain requirements to a person's day that will limit where they can go and for how long they can do activities (Hägerstrand, 1970). See also the explanation in the main body of this report.

In terms of the question design, most of this section was modified and the content switched to COVID-19 issues. Many of the questions above do not require as much background research into validated instruments, as almost all these sub-topics are external/environmental conditions or more objective measures. The exception to this is perceived control over the options available, which comes from the TPB, where it is called "perceived behavioral control." This is people's perception of how much choice they have in their actions. There are a few examples in the literature of active transportation on perception and reality of options attributes (Hoehner et al., 2005; Jensen et al., 2017) and public transportation quality of service (Gao et al., 2018).

Appendix B Correlation with Sociodemographics of Complexity and Travel Time Ratio

Table B.1 By-cluster complexity statewide

	Cluster Type								
	Dependent Variable								
	Complexity C(s) (Eq. 2.4)								
	Home Day	School Day	Typical Work Day	Errands Type 1	Mostly Out of Home	Errands Type 2	Non-typical Work Day	Leave Home	Traveling
Constant	0.03	0.04	0.04	0.06	0.01	0.05	0.04	0.05	0.004
	*t = 27.15****	*t = 15.94****	*t = 8.16****	*t = 23.40****	*t = 4.63****	*t = 17.88****	*t = 9.72****	*t = 13.50****	*t = 2.28***
Disability	-0.01	-0.004	-0.003	-0.005	-0.001	-0.0003	-0.002	-0.01	-0.002
	*t = -8.80****	*t = -1.22*	*t = -0.93*	*t = -1.50*	*t = -0.50*	*t = -0.09*	*t = -0.26*	*t = -1.82**	*t = -0.81*
Household Income Near or Below Poverty Line	-0.002	-0.003	-0.01	-0.002	-0.002	-0.003	-0.01	-0.01	0.002
	*t = -2.26***	*t = -1.74**	*t = -5.10****	*t = -0.81*	*t = -0.81*	*t = -1.07*	*t = -1.03*	*t = -2.15***	*t = 1.05*
Weekend	-0.003	0.001	-0.003	-0.004	0.0001	-0.004	0.002	-0.003	-0.001
	*t = -5.66****	*t = 0.37*	*t = -2.43***	*t = -3.11****	*t = 0.04*	*t = -2.39***	*t = 0.60*	*t = -1.25*	*t = -0.95*
Respondent is Under 15 Years Old	-0.003	-0.001		-0.002	-0.001	-0.002		-0.003	0.003
	*t = -2.92****	*t = -0.69*		*t = -0.62*	*t = -0.29*	*t = -0.67*		*t = -0.59*	*t = 1.71**
Respondent is Over 65 Years Old	-0.002	0.03	-0.003	-0.004	-0.004	-0.001	-0.003	-0.01	-0.001
	*t = -3.09****	*t = 2.87****	*t = -1.95**	*t = -2.21***	*t = -2.34***	*t = -0.25*	*t = -0.30*	*t = -1.55*	*t = -0.45*
Presence of Children Under 4	-0.002	-0.001	0	-0.001	-0.005	0	0.01	-0.003	0.0003
	*t = -1.99***	*t = -0.91*	*t = 0.02*	*t = -0.45*	*t = -2.21***	*t = 0.01*	*t = 1.92**	*t = -0.68*	*t = 0.24*
Presence of Children Aged 4 to 15	0.004	0.002	0.004	0.004	-0.002	0.004	-0.003	-0.001	-0.002
	*t = 5.55****	*t = 0.98*	*t = 5.34****	*t = 2.27***	*t = -1.23*	*t = 1.95**	*t = -0.81*	*t = -0.42*	*t = -1.96**
Presence of Children Aged 16 to 18	-0.001	0.001	0.003	-0.01	-0.003	0.002	-0.002	-0.01	-0.002
	t = -1.15	*t = 0.59*	*t = 2.49***	*t = -2.55***	*t = -1.43*	*t = 0.89*	*t = -0.47*	*t = -1.46*	*t = -1.30*
Female	0.002	0.001	0	0.002	0	0.003	0.01	-0.002	-0.0002
	*t = 3.17****	*t = 1.39*	*t = -0.06*	*t = 1.64*	*t = 0.04*	*t = 2.02***	*t = 2.16***	*t = -0.72*	*t = -0.25*
Worker	0.004	-0.001	0.01	-0.001	-0.0001	0.004		0.003	0.001
	*t = 6.62****	*t = -0.32*	*t = 1.46*	*t = -0.69*	*t = -0.07*	*t = 2.03***		*t = 1.17*	*t = 1.09*
Student	-0.003	0.003	0.005	-0.01	0.0004	-0.003	0.004	0.002	-0.001
	*t = -2.98****	*t = 1.41*	*t = 1.73**	*t = -2.22***	*t = 0.17*	*t = -0.95*	*t = 0.76*	*t = 0.51*	*t = -0.92*
Number of Household Vehicles	0.0002	0	-0.001	0.0001	-0.001	-0.001	-0.001	0.0001	-0.0002
	t = 0.57	*t = 0.05*	*t = -2.92****	*t = 0.11*	*t = -1.09*	*t = -1.09*	*t = -0.43*	*t = 0.08*	*t = -0.37*
Suburban Household	-0.002	-0.0003	0.001	-0.001	-0.001	-0.0001	-0.003	0.004	0.0005
	*t = -3.55****	*t = -0.25*	*t = 0.88*	*t = -0.80*	*t = -0.94*	*t = -0.07*	*t = -0.81*	*t = 1.56*	*t = 0.41*
Exurban Household	-0.003	-0.002	-0.001	0.004	0.001	-0.0004	-0.01	-0.01	0.0003
	*t = -3.92****	*t = -1.35*	*t = -0.93*	*t = 1.83**	*t = 0.40*	*t = -0.17*	*t = -1.22*	*t = -2.24***	*t = 0.17*
Rural Household	-0.01	-0.005	0.0001	-0.0002	0.003	0.001	-0.01	-0.01	-0.002
	*t = -5.92****	*t = -2.13***	*t = 0.09*	*t = -0.07*	*t = 1.31*	*t = 0.22*	*t = -1.12*	*t = -1.45*	*t = -0.77*
Observations	6,853	1,037	2,243	553	537	592	185	334	364
R^2	0.05	0.03	0.04	0.08	0.04	0.04	0.08	0.11	0.04
Adjusted R^2	0.05	0.01	0.03	0.06	0.02	0.02	0.01	0.07	-0.003
Residual Std. Error	0.02 (df = 6,837)	0.01 (df = 1,021)	0.02 (df = 2,228)	0.02 (df = 537)	0.01 (df = 521)	0.02 (df = 576)	0.02 (df = 171)	0.02 (df = 318)	0.01 (df = 348)
F Statistic	24.76*** (df = 15; 6,837)	1.77** (df = 15; 1,021)	6.44*** (df = 14; 2,228)	3.22*** (df = 15; 537)	1.57* (df = 15; 521)	1.63* (df = 15; 576)	1.09 (df = 13; 171)	2.66*** (df = 15; 318)	0.92 (df = 15; 348)

*Note: *p<0.1; **p<0.05; ***p<0.01*

Table B.2 By-cluster TTR statewide

	Cluster Type Dependent Variable Travel Time Ratio (TTR)								
	Home Day	School Day	Typical Work Day	Errands Type 1	Mostly Out of Home	Errands Type 2	Non-typical Work Day	Leave Home	Traveling
Constant	0.42	0.14	0.13	0.28	0.03	0.26	0.17	0.2	1
	*t = 37.59****	*t = 11.98****	*t = 5.02****	*t = 13.08****	*t = 4.10****	*t = 8.27****	*t = 10.08****	*t = 6.96****	*t = 77.39****
Disability	-0.01	0.06	0.02	-0.01	-0.0004	0.05	-0.04	-0.02	0.01
	t = -0.67	*t = 4.65****	*t = 1.40*	*t = -0.47*	*t = -0.04*	*t = 1.34*	*t = -1.00*	*t = -0.70*	*t = 0.77*
Household Income Near or Below Poverty Line	0.01	0.01	-0.01	-0.02	-0.01	-0.03	-0.02	-0.01	-0.02
	t = 1.21	*t = 1.70**	*t = -1.21*	*t = -1.11*	*t = -0.61*	*t = -0.91*	*t = -1.04*	*t = -0.28*	*t = -1.97**
Weekend	-0.04	0.01	0.003	0.01	0.001	0.01	0.02	-0.02	0.01
	*t = -6.44****	*t = 0.82*	*t = 0.53*	*t = 0.50*	*t = 0.27*	*t = 0.65*	*t = 1.48*	*t = -0.84*	*t = 1.99***
Respondent is Under 15 Years Old	-0.06	-0.04		-0.01	-0.01	-0.06		-0.01	-0.03
	*t = -4.40****	*t = -5.42****		*t = -0.29*	*t = -1.18*	*t = -1.60*		*t = -0.18*	*t = -2.55***
Respondent is Over 65 Years Old	0.002	0.09	0.01	-0.01	-0.02	0.03	0.02	0.02	-0.003
	t = 0.24	*t = 1.94**	*t = 1.72**	*t = -0.44*	*t = -2.90****	*t = 1.11*	*t = 0.75*	*t = 0.77*	*t = -0.20*
Presence of Children Under 4	0.04	0.01	0.0004	-0.05	-0.01	-0.01	0.03	0.03	-0.002
	*t = 3.38****	*t = 1.43*	*t = 0.06*	*t = -2.39***	*t = -1.43*	*t = -0.39*	*t = 1.20*	*t = 0.94*	*t = -0.21*
Presence of Children Aged 4 to 15	0.02	0.02	0.01	0.04	-0.01	-0.02	-0.01	0.06	0.01
	*t = 2.54***	*t = 2.03***	*t = 3.23****	*t = 2.48***	*t = -1.14*	*t = -1.11*	*t = -1.00*	*t = 2.72****	*t = 1.79**
Presence of Children Aged 16 to 18	0.02	-0.01	0.01	-0.06	-0.01	0.03	-0.01	0.06	0.02
	*t = 1.92**	*t = -1.66**	*t = 1.35*	*t = -3.35****	*t = -1.91**	*t = 1.17*	*t = -0.56*	*t = 2.22***	*t = 2.65****
Female	-0.02	0.0003	-0.01	0.02	0.0005	0.01	0.002	-0.03	-0.005
	*t = -3.23****	*t = 0.08*	*t = -2.42***	*t = 1.46*	*t = 0.12*	*t = 0.56*	*t = 0.17*	*t = -1.74**	*t = -0.73*
Worker	-0.01	0.05	0.01	-0.02	-0.001	0.03		0.01	-0.01
	t = -0.89	*t = 3.33****	*t = 0.55*	*t = -1.68**	*t = -0.23*	*t = 1.34*		*t = 0.71*	*t = -0.91*
Student	-0.05	-0.02	0.04	-0.04	0.01	-0.05	0.01	-0.11	0.0002
	*t = -4.06****	*t = -1.80**	*t = 3.04****	*t = -1.91**	*t = 1.41*	*t = -1.57*	*t = 0.42*	*t = -3.69****	*t = 0.02*
Number of Household Vehicles	-0.02	-0.01	-0.004	-0.01	-0.003	0.001	-0.02	0.0001	-0.01
	*t = -4.39****	*t = -2.52***	*t = -2.22***	*t = -1.67**	*t = -1.17*	*t = 0.17*	*t = -2.64****	*t = 0.01*	*t = -2.23***
Suburban Household	-0.001	-0.005	-0.003	0.01	-0.0001	-0.01	-0.03	0.02	-0.001
	t = -0.10	*t = -0.98*	*t = -0.63*	*t = 0.73*	*t = -0.03*	*t = -0.37*	*t = -2.34***	*t = 0.81*	*t = -0.07*
Exurban Household	0.003	0.001	-0.01	0.05	0.0004	-0.03	-0.04	0.06	0.01
	t = 0.34	*t = 0.23*	*t = -2.09***	*t = 2.51***	*t = 0.07*	*t = -1.24*	*t = -1.93**	*t = 2.52***	*t = 1.08*
Rural Household	0.06	0.02	-0.003	0.07	0.001	-0.05	0.005	0.09	0.02
	*t = 4.27****	*t = 2.07***	*t = -0.39*	*t = 2.71****	*t = 0.15*	*t = -1.58*	*t = 0.18*	*t = 2.19***	*t = 1.18*
Observations	4,265	1,037	2,243	553	537	592	185	334	364
R^2	0.04	0.1	0.02	0.08	0.04	0.07	0.15	0.14	0.08
Adjusted R^2	0.04	0.09	0.01	0.05	0.01	0.05	0.08	0.1	0.04
Residual Std. Error	0.21 (df = 4,249)	0.07 (df = 1,021)	0.08 (df = 2,228)	0.14 (df = 537)	0.05 (df = 521)	0.20 (df = 576)	0.08 (df = 171)	0.15 (df = 318)	0.06 (df = 348)
F Statistic	12.53*** (df = 15; 4,249)	7.65*** (df = 15; 1,021)	3.02*** (df = 14; 2,228)	3.09*** (df = 15; 537)	1.28 (df = 15; 521)	3.02*** (df = 15; 576)	2.25*** (df = 13; 171)	3.40*** (df = 15; 318)	1.95** (df = 15; 348)

*Note: *p<0.1; **p<0.05; ***p<0.01*

Appendix C Travel Diary Smartphone Application

We will refer to the type of travel behavior data collection done for this project in Chapter 3 as "reflection" data, which would involve asking respondents to reflect on their own behavior over an amount of time. This differs from the travel diary, which is a highly detailed form of data collection from respondents. In essence, the travel diary records every trip a person makes with the mode used, origin and destination, duration in time and distance, mode used and then adding other prompted questions depending on the purpose of the survey. Typically, a travel diary is collected by distributing a travel log and asking respondents to carry this around with them for a day recording their trip details. Respondents must then return this diary to researchers for digitization.

Generally, there are pros and cons to both reflection and travel diary data, as shown in **Error! Reference source not found.**. Although reflection surveys are typically faster to deploy, easier for respondents to fill out, and less work for researchers in terms of data cleanup and prep for analysis, they do not provide as accurate of information, nor as detailed. Respondents have unreliable memories for accurately reporting their activities and travel. The level of detail is also lower with reflection data. The travel diary provides a more accurate representation of a day in the life of a respondent, and it gives researchers a flexible dataset to use for future analysis, since it is at a high level of detail (every trip is recorded in detail for a day). However, a travel diary is typically much more labor-intensive for a respondent, labor-intensive for surveyor, prone to response errors because of the handwritten nature of the diary, labor-intensive for researchers in the digitizing stage, and prone to errors when digitizing these handwritten locations.

Reflection

- **Pros**
  - Easy to collect from respondents
  - Faster to set up and deploy survey
  - Less work for respondents
  - Less data cleanup work
- **Cons**
  - Loss of accuracy
  - Unreliable memories
  - Prone to respondent errors
  - Less detailed information

Travel Diary

- **Pros**
  - More accurate representation of a single day for the respondent
  - Highly detailed information
  - Less issue with memories
  - Flexible for future analysis
- **Cons**
  - More invasive (privacy concerns)
  - Labor-intensive for respondent
  - Labor-intensive for surveyor
  - Prone to respondent errors
  - Prone to digitizing errors

Figure C.1 Comparing Reflection and Travel Diary Data

If it were possible to lessen the downsides of the travel diary, then it would be the preferable option for data collection in most quantitative travel behavior research contexts. One solution to this is to use a smartphone application to collect the diary. Smartphone travel surveys became popular with travel behavior researchers because they decrease the burden of respondent diaries from whom we need to have locations visited and the timing of all their trips in a day. The most attractive feature of smartphones is the accurate location capability and the more accurate recording of travel patterns provided strategic prompts are included in the smartphone app (Harding et al., 2020). Evidence is also starting to accumulate about the positive attitudes of survey participants when the apps are easy to use and are perceived to be useful by the respondents as useful while the perceived risk of loss of privacy does not appear a major issue (Assemi et al., 2018).

Most people already carry a smartphone around with them throughout the day, so respondents would not have to carry the physical diaries around and remember to write every trip down in them. They would not have to mail the diary to the researchers afterwards either.

The smartphone app would just require them to allow location access and provide supplementary information about the trips recorded after the day is over. For researchers, they would no longer need to digitize the handwritten logs, get coordinates from addresses or business names. It would also reduce the need for error checking and data cleanup, and the data would automatically be set up in a database structure.

Although there are a few options available for smartphone applications, most are either proprietary or if they are open source they are no longer maintained and are very out of date. There is not a good option available for a modern, open-source smartphone application framework. The goal is to create an app framework that future researchers can use for a low cost to reduce the typically large monetary barrier to collecting detailed data such as a travel diary. The app designed here would include an online survey, like the survey for COVID-19 described in Chapter 3, with the addition of the smartphone app for the travel diary portion of the overall survey.

Designing the Application

First, the application features must be identified. Then, they must be separated into those that are necessary features must be identified, and separated from the features that would be desirable, but not necessary. **Error! Reference source not found.** details these features and shows the division between the two categories.

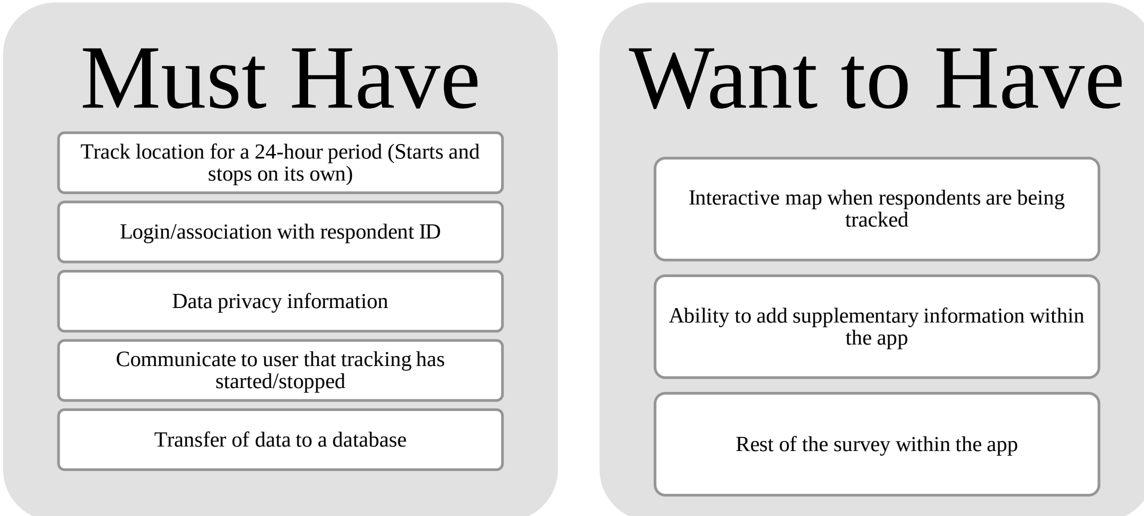


Figure C.2 Necessary and Desired Features

The next step is to create a “wireframe” for each page in the app, as shown in **Error! Reference source not found.**. This provides a visual representation of the application to act as a set of guidelines for the coding process. The wireframe the features of the application and the way it functions, including the relationships between the screens and how navigating between screens will work.

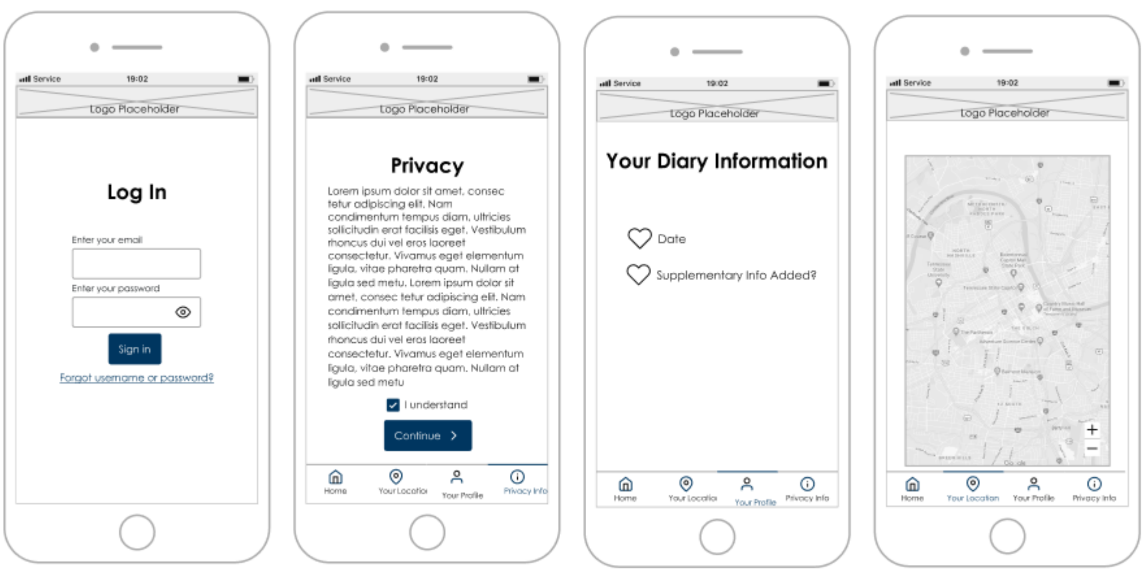


Figure C.3 Wireframe of Smartphone App

Coding the Application

There are two options for building an application for both iOS and Android phones. One option is to build the application separately for each platform, called building a "native" app. iOS uses the language Swift, and Android uses Java. However, maintenance of applications built in this way requires having someone available to bug-check and maintain the app in both languages. The other option is to write the application once for both platforms, called a "cross-platform" app. Some options for doing this include Flutter, Ionic, NativeScript, and ReactNative. ReactNative was chosen, because it has a framework called Expo that makes application development faster and easier for those who have not done it before. ReactNative apps are written in JavaScript, then rendered in native code. Expo makes a simulated version of the app for the developer, where features like as button clicks and even location tracking will work as they would in the final product. It is possible to test the simulated app on both virtual phones that are emulated on a computer and physical phones by downloading the Expo Go app.

Publishing

Publishing the application will involve submitting it to both the Google Play Store and Apple App Store for review. Upon acceptance, the app will be publicly available for download. In the past year or so, the phone companies have been cracking down on background location tracking. Developers must provide a good explanation for why it is needed, and all the privacy measures and details of how the location tracking will work must be made clear to the app users. The justification for using background location in this application is as follows. The locations collected are being used for research purposes, and they will never be sold or used for profit. Their information is kept in a secure database, and it will only be provided to researchers who have received Institutional Review Board (IRB) approval, meaning they have legally agreed to protect the rights, welfare, and privacy of participants. Tracking will permanently stop at the end of a specific, scheduled 24-hour period that participants have agreed to, so they will know exactly when their locations are being recorded. After the tracking period is over, respondents will also have a chance to review the information collected.

Progress

Thus far, the application can track locations with a "start" and "stop" button, and it sends those locations to an Amazon Web Services database. However, it is still not able to start and stop on its own based on a scheduled time. Aiming to have it start and stop on a schedule is a potential barrier, because it can cause some privacy issues that may be flagged by the application stores. The method of tracking the locations for a specific number of hours may need to be revised. Also, there is not yet an interface for respondents to add supplemental information and corrections after locations have been recorded. After these two steps have been completed, the application will be exported from the Expo simulation framework into a standalone application at which point it will be tested on multiple smartphones with different operating systems. Once it has cleared the testing phase, the application will be ready to submit to the app stores for approval. The application in its current state can be found in the GitHub repository linked here. [3]

[3] https://github.com/e-mcbride/travel-diary-app