Title

Traveler Satisfaction Surveys meet Mobile Phone and Vehicle Tracking: Linking Individual Experiences to Travel Habit Changes with Panel Data

Author

Carrel, André Laurent

Publication Date

2015

Peer reviewed|Thesis/dissertation

Traveler Satisfaction Surveys meet Mobile Phone and Vehicle Tracking: Linking Individual Experiences to Travel Habit Changes with Panel Data

by

André Laurent Carrel

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering – Civil and Environmental Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Joan L. Walker, Chair
Professor Raja Sengupta
Professor John F. Canny
Professor Rabi G. Mishalani

Summer 2015

Traveler Satisfaction Surveys meet Mobile Phone and Vehicle Tracking: Linking Individual Experiences to Travel Habit Changes with Panel Data

Abstract

Traveler Satisfaction Surveys meet Mobile Phone and Vehicle Tracking: Linking Individual Experiences to Travel Habit Changes with Panel Data

by

André Laurent Carrel

Doctor of Philosophy in Civil and Environmental Engineering

University of California, Berkeley

Professor Joan L. Walker, Chair

Smartphones are becoming an increasingly interesting survey medium for behavioral research due to their value for collecting long-term panel observations and supplementary data on the choice environment. Thanks to the sensor data, it becomes possible to survey participants based on whether or not a certain activity has been carried out. By fusing the phone-generated sensor data and survey responses with data from outside sources, substantial data sets can be generated which can be used to investigate choices in complex environments. Computational systems for behavior research take advantage of automation and scalability opportunities, thereby building also on pertinent bodies of literature regarding machine learning on large data sets and crowdsourcing. The importance of comprehensive, long-term data sets in understanding behavior has been highlighted in the choice theory literature, specifically with respect to capturing an individual decision-maker's history of choices and personal experiences with those choices. To date, however, relatively few studies have capitalized on emerging technologies to create or analyze such data sets.

Rich data sets which combine panel information on the decision-maker with information on the choice environment can support the study of dynamic phenomena, which is especially important in a rapidly changing world where behavioral adaptation can take place on a relatively small time scale and, once habits are formed, have long-lasting effects. Some examples of pressing questions in the field of transportation involve understanding how travelers are responding to the emerging sharing economy, to new ride sharing services and new information systems, how time use and travel patterns will change due to automated vehicles, and how more sustainable travel behavior can be promoted through incentive or pricing strategies. This dissertation aims to support the adoption of smartphone-based survey technology in travel behavior research in order to lay the groundwork for research

aimed at answering the above questions. It describes the design and implementation of a smartphone-based study, presents a system for fusing smartphone data with externally acquired data, and demonstrates how these ample data sets can be leveraged to generate new behavioral insights. The problem chosen for study is the link between transit service quality, rider satisfaction and ridership retention on public transit. This is motivated by the fact that many transit agencies in the United States continue to see large rates of ridership turnover, and that to date, very little is known about what drives transit use cessation.

The six-week San Francisco Travel Quality Study (SFTQS) was conducted in autumn 2013. It collected a data set that included high-resolution phone locations, a number of daily mobile surveys on specific trip experiences, responses to online entry and exit surveys, and transit vehicle locations. By fusing the phone location data with transit vehicle locations, individual-level automatic transit travel diaries could be created without the need to ask participants. The reduced respondent burden, in turn, facilitated a longer term data collection. Initial recruitment proved to be challenging, with response rates to some of the email and direct mailing lists around 1%, and response rates to in-person recruiting between 8 and 15%. On the other hand, attrition was lower than expected, considering the length of the study: The initial enrollment was 856 participants, of which 555 (65%) participants completed all required surveys and 637 (74%) completed the entry and exit survey as well as at least one daily mobile survey. Interestingly, 36% of participants later stated they would have preferred to fill out mobile surveys more frequently (e.g., one per trip rather than one per day) than what was required in the study.

A central part of the computational infrastructure used to collect the data was the system of integrated methods to reconstruct and track travelers' usage of transit at a detailed level by matching location data from smartphones to automatic transit vehicle location (AVL) data and by identifying all out-of-vehicle and in-vehicle portions of the passengers' trips. This system is presented in detail in this dissertation, where it is shown how high-resolution travel times and their relationships with the timetable are derived. Approaches are presented for processing relatively sparse smartphone location data in dense transit networks with many overlapping bus routes, distinguishing waits and transfers from non-travel related activities, and tracking underground travel in a metro network. While transit agencies have increasingly adopted systems for collecting data on passengers and vehicles, the ability to derive high-resolution passenger trajectories and directly associate them with vehicles has remained a challenge. The system presented in this dissertation is intended to remedy this situation, and it enables a range of different analyses and applications. Results are presented from an implementation and deployment of the system during the SFTQS. An analysis of out-of-vehicle travel times shows that (a) longer overall travel times in trips involving a transfer are strongly driven by transfer times, and (b) median wait times at the origin stops are consistently low regardless of the headway. The latter can be seen as an effect of real-time information, as it appears that wait times are increasingly spent at locations other than the stop and that passengers time their arrivals at the stop. Given these shifts, the traditional assumption that the average wait time at a transit stop of a high-frequency route is half the headway due to random arrivals may need to be revisited.

This dissertation presents two applications to derive new behavioral insights from the SFTQS data set and to demonstrate the power and value of these new types of data. The analyses were based on participants' individual history of transit usage and experiences with service quality. The first analysis used the data from the daily mobile surveys to model the link between participants' reported satisfaction with travel times on specific trips (i.e., their subjective assessment) and objective measures of those travel times. Thanks to the tracking data, it was possible to decompose observed travel times into their in-vehicle and out-of-vehicle components, and to compare the observed in-vehicle travel times to scheduled in-vehicle travel times to identify delays suffered while the participant was on board. The estimation results show that on average, a minute of delay on board a vehicle contributed more to passenger dissatisfaction than a minute of waiting time either at the origin stop or at a transfer stop, and that delays on board metro trains are perceived as more onerous than delays on board buses. Furthermore, the models included participants' baseline satisfaction levels as reported in the entry survey and a daily measure of their subjective well-being. Both variables are relatively new elements in travel surveys, and both are seen to be significant in the estimation results. These results indicate that satisfaction with travel times may be composed of a baseline satisfaction level and a variable component that depends on daily experiences, and that there may be non-negligible interactions between subjective well-being and travel satisfaction. Therefore, it is recommended that future survey designs should include measures for both these variables.

The second application builds on the results of the first to empirically investigate the causes for cessation of transit use, with a specific focus on the influence of personal experiences that users have had in the past, on resulting levels of satisfaction, and subsequent behavioral intentions. A latent variable choice model is developed to explain the influence of satisfaction with travel times, including wait times at the origin stop, in-vehicle travel times, transfer times and overall reliability, and satisfaction with the travel environment on behavioral intentions. The group of variables summarized as "travel environment" includes crowding, cleanliness, the pleasantness of other passengers, and safety. Satisfaction is modeled as a latent variable, and the choice consists of participants' stated desire and intention to continue using public transportation in the future. In addition to the delay types captured in the first analysis, a set of negative critical incidents is included, namely being left behind at stops and arriving late to work, school or a leisure activity. The results of the model and descriptive analysis show that operational problems resulting in delays and crowding are much stronger drivers of overall dissatisfaction and cessation than variables related to the travel environment. The importance of baseline satisfaction, mood and the relatively larger impact of in-vehicle delays are confirmed by this model. Thanks to the framework, the critical incidents can be expressed in terms of equivalent delay minutes. For instance, being left behind at a bus stop is found to cause the same amount of dissatisfaction as approximately 18 minutes of wait time. Furthermore, the effect of delays or incidents on ridership can be quantified, as is demonstrated in a set of simulations using the San Francisco transit network (Muni) as a basis. It is shown that if all passengers were subjected to one hypothetical on-board delay of 10 minutes per person, the resulting loss of riders would account for approximately 9.5% of Muni's yearly ridership turnover.

In summary, the contributions and impact of this dissertation are as follows: It presents a framework and system that allows the researcher to gather detailed information on an individual and on the decision environment through phone-based survey apps in combination with sensor data from the phone and from external sources. In the public transit context, an innovative system is presented to match AVL data with smartphone location data in order to measure the personal experiences of travelers with respect to travel times. With repeated measurements, these data can be used to calculate personalized reliability metrics for individual travelers, reflecting the sum of their travel experiences, or they can be used to derive aggregate travel time distributions across all travelers by time of day, origin-destination pair or location on the network. These metrics can capture the true door-to-door travel times experienced by travelers and can serve as the basis for user-centric performance metrics to supplement system-level performance metrics commonly used by transit agencies. The long-term nature of this data collection and low respondent burden facilitate the observation of behavioral dynamics such as habit formation or lifestyle adjustments in response to changes of the choice set. The low cost and scalability of these data collection methods permits relatively short lead times for studies and frequent data collection. Thus, these systems can be deployed at the early stages of a new product's or technology's emergence (e.g., new ride sharing services or traveler information systems) to gain insights into how they affect consumer choices in a real-world setting, both on short-term and longer-term time scales. Moreover, personalized interactions between the researcher and the participant via the smartphone facilitate behavioral interventions and allow for targeted incentives to change behavior.

By elaborating on the researchers' experiences in designing and implementing the SFTQS, this dissertation provides a powerful example of how these new types of data sets can be harnessed and supports the development of future studies. The application of the methodology and framework to the transit context were motivated by a desire to learn more about the drivers of satisfaction among transit riders and the causes of transit use cessation. The model estimation results underscore the importance of investments into run time stability measures such as transit signal priority systems and dedicated rights of way. Furthermore, thanks to the modeling framework, the cost of holding vehicles with passengers on board and the cost of vehicles not stopping at stops due to overcrowding can be quantified; this can directly impact operating and control policies. On the other hand, the model results provide evidence that investments into stop amenities may be becoming less important as passengers spend less time at stops (except in the case of transfers), but that conversely, the benefits of transit-oriented development may be increasing as passengers can choose to spend their wait times elsewhere thanks to real-time information. The data and modeling framework allow transit agencies to create service quality metrics that can appropriately capture individual users' experiences in real-time, and improve researchers' ability to compare the level of service on public transportation with the level of service of private modes of transportation. Together, this set of methods and results is an important step toward aligning the service provided by transit agencies more closely with the needs of customers.

The framework and methodologies described in this dissertation are useful beyond the specific transit application presented as a case study. Thanks to the flexibility and scalability of the smartphone-based data collection and automated post-processing, they can be used by researchers to quickly gather insights on emerging trends and on travelers' adaptation to new services and new technologies, and to study the dynamics of behavior change over longer time periods. In particular, they can be applied to understand how flexible, shared-ride transportation systems and future automated mobility on demand systems will shape travel demand and how users will interact with those systems. This, in turn, will be a critical input to policy-making and system design.

Contents

List of Figures

List of Tables

Acknowledgments

I would like to express my gratitude to Joan Walker for her excellent guidance, mentoring, patience, and for providing me with financial support to conduct this research. Likewise, I would like to express my gratitude to Raja Sengupta for his excellent guidance and support, and for helping me to see this work from new angles. I am also indebted to Rabi Mishalani, whose guidance and contributions helped perfect the design of the study and evaluation of the data, and to John Canny for his very valuable advice and suggestions.

I would like to thank Sonali Bose and Travis Fox from the San Francisco Municipal Transportation Agency for generously supporting my work and for the trust they placed in me when the outcome of the study was uncertain. The help and advice of Jason Lee were invaluable in making this research study a reality. I am grateful to Elizabeth Sall from the San Francisco County Transportation Authority for generously supporting me during my final year and for her valuable advice along the way. Further thanks go to Kimpreet Puar, David Papas, Belle Yan and Diana Hammons for their work during the implementation of the study.

I am very indebted to Peter Lau for his expertise and companionship, as well as for all the time he spent helping me create much of the infrastructure behind this study. I would like to thank Hoang Nguyen for his brilliant problem solving and Phuc-Hai Huynh for his relentless efforts creating and troubleshooting the software. I am grateful to Andrew Campbell for his help designing and implementing the study, and to Anne Halvorsen, Katie Leung and Venkatesan Ekambaram for their efforts during early design iterations. Further thanks go to Nachiketa Mehta, Stephanie Ku, Joseph John, and Toan Pham for their contributions to the infrastructure.

While designing the study and evaluating the data, I benefited from the expertise of a number of friends and colleagues who provided input and advice, most notably Maya Abou Zeid, Robert Levenson, Aaron Fisher, Leif Nelson, Jillian Anable and Angelo Guevara. Further, I would like to say thank you to Laura Melendy, Carmen Lam, Kathryn Lewis and Lynne Hollyer for their support preparing the study for launch.

I would further like to express my gratitude to the National Science Foundation, the University of California Multicampus Research Programs and Initiatives, the University of California Center on Economic Competitiveness in Transportation, and the University of California Transportation Center for providing me with funding. Last but not least, many thanks go out to all study participants for their time and efforts, and for their interest in this study.

Chapter 1

Introduction

1.1 Motivation

The transportation sector is on the verge of a profound transformation that is driven largely by technology. The way in which transportation services are utilized is being affected by the emergence of the sharing economy and of real-time ride share services in the short run, and will most likely be affected even more by the introduction of drones, autonomous vehicles and mobility on demand systems in the long run. Furthermore, shifts in economic activity patterns due to services moving online, demographic changes, an aging population and new forms of networking mean that the way in which people socialize and access goods and services is poised to evolve rapidly. As transportation is a derived demand, the effects on travel patterns will certainly not be negligible. It is important to gain an early understanding of the behavioral ramifications of these developments in order to make accurate forecasts of demand, plan public investments and design policies aimed at encouraging more sustainable travel behavior. Of particular interest will be the dynamics of travel choices, the adaptation processes of travelers and the formation of new habits.

In his Nobel lecture, McFadden (2001) opened by explaining that the advent of the first computers in the 1960s and the increasing availability of survey data at the time led to a transformation of the fields of economics and behavioral modeling. There is reason to believe that at this point in time, a second such transformation is underway, fueled by mobile technology, ubiquitous sensor networks and the widespread availability of computational power. In the transportation community, technological advancements and the proliferation of mobile devices have begun to allow the collection of increasingly detailed, long-term data on urban systems and travel behavior from mobile phones and other distributed sensor systems; these data are very valuable for shedding light on the behavioral transformations in question. The phone data can be either passively collected (i.e., without any overt interaction with a person), or actively via targeted questions. The use of smartphone-based methods to either supplement or replace traditional, paper-based data collection methods for travel surveys results in a number of new questions that need to be addressed as well as many opportunities that can be explored.

For data collection that is initiated by the researcher, several key advantages are apparent: First, using smartphones for data collection reduces the logistics involved in conducting a study and the infrastructure is typically quite scalable, at least from the technical point of view. Given the low cost of distributing and running a smartphone-based survey, surveys can be conducted more frequently over longer periods of time, which allows for multiple iterations. Because a survey app can operate in the background on a participant's phone over prolonged periods of time without requiring participant interaction, detailed panel data on long-term, dynamic phenomena can more easily be collected. Secondly, the time and location of data collection can be determined by the researcher. It becomes possible to survey participants as a function of their being at a certain location or having carried out a specific activity. This has the potential to reduce the time lag between an event of interest and the time a respondent's assessment of it is collected, which in turn can reduce bias introduced by cognitive distortions caused by time lag. Third, a range of additional information can be collected from phone sensors and by mining information stored on the phone. This can further reduce the respondent burden or help researchers access information that the respondent would not be able to provide. Depending on the study needs, it might not even be necessary to provide a stand-alone survey app to participants; if only sensor data is required, the code can be bundled with other apps used by participants. Fourth, the personalization of surveys becomes more feasible thanks to the lower cost of survey administration and the fact that a smartphone is typically tied to one individual participant.

Aside from data collected for specific survey purposes, there are also increasing numbers of data sets collected for other purposes which can be of value to researchers; a powerful aspect of mobile methods for behavioral data collection lies in the fact that the survey data from mobile phones can be combined with data collected through other channels, either actively or passively. As an example, survey responses can be mapped to information from vehicle positioning systems, payment systems, phone-based accelerometers, microphone data, or weather data. The outside data can be used to augment or to verify survey responses, or to help the researcher better understand the context in which the responses were given. Combined with sociodemographic information on the participant, these expanded data sets can be used to estimate a variety of behavioral models, and the improved understanding of the decision context can then be leveraged for new behavioral insights that would have been very difficult to obtain otherwise.

Data sets where the survey data are combined with data from other sources are particularly useful when the decision environment changes over time or when participants are allowed to carry out the behavior of interest on their own timing, without an imposed time frame from the researcher. Overall, this allows more behavioral data collection to be moved from lab settings into the "real world", resulting in richer data sets and more realistic (natural) choice situations where the influence of the researcher on the participant is kept at a minimum. Of course, the disadvantage that the experimenter does not have the opportunity to manipulate any environmental factors means one must be confident that enough variability can be observed in a natural setting so that the respective choice models can be properly estimated.

On the other hand, mobile technology is surely altering human behavior in ways that have yet to be investigated. Decisions can increasingly be made at a moment's notice and, in many cases, decision makers can rely on real-time information to continuously update their choices and plans. Given the increased number of channels through which the information leading up to those decisions can be acquired, adequately capturing the decision environment has become a challenging task. Furthermore, there are questions that are of particular interest to public agencies conducting surveys and data collection drives like the representativeness of population samples where smartphones were used for data collection, the response rates when compared to paper surveys, and other possible biases that might be introduced by this new medium.

This dissertation explores the points made above in the context of a travel survey. It formulates a research question that has so far received limited attention in the literature, in part due to a lack of suitable data sets, and it illustrates how a smartphone-based survey can be designed to answer said question. In doing so, it highlights the advantages of using a phone-based survey methodology rather than a traditional paper-and-pencil approach. The post-processing of the data and its use for developing choice models underline the value of these new types of data sets and showcase the behavioral insights that can be gleaned from them.

1.2 Methodology

This dissertation presents a case study of using smartphones for conducting a large-scale behavioral research project. Applied in the context of public transportation, this methodology was used to address the question of how personal experiences and customer satisfaction with quality of service can help understand future mode choice behavior of transit riders. Future choice behavior of riders is of importance to transit agencies as it is directly related to ridership retention, an issue that many large agencies in the United States struggle with. Perk, Flynn, and Volinski (2008) report that high levels of ridership turnover are a typical finding in annual customer surveys. Until now, however, relatively little was known about the drivers of transit use cessation. There are indications that, in general, people build their lifestyles around the use of certain modes of transportation, including transit (Vij, Carrel, and Walker, 2013). Shifting away from regularly using a certain mode is often a decision for which the individual has to overcome inertia and break old habits, making it unlikely that the individual would shift back soon after making such a change. A study of transit use cessation would need to observe a sufficient number of such shifts to obtain a meaningful sample size from which to draw conclusions on the causes. However, as these major decisions are made infrequently, the required sample size and duration of a study might be prohibitive; alternatively, the problem can be approached within a smaller time frame by measuring variables that are predictive of future behavior. To understand which variables would need to be collected, the Theory of Planned Behavior and an extension thereof, the Model of Goal-directed Behavior are useful (see section 5.2 for more details).

The design and implementation of a large-scale transportation study, the San Francisco Travel Quality Study (SFTQS), is described in this dissertation, which aimed at investigating the drivers of transit use cessation among regular transit users in San Francisco, with a focus on the causal chain from personal experiences with public transportation to satisfaction to future behavior. This panel study covered 6 weeks from October to December 2013, involved an initial total of 856 participants, and used smartphones as the primary data collection device. Participants were asked to fill out two online entry surveys at the beginning of the study and one exit survey at the end of it. Then, they were asked to download a mobile survey app and keep it running on their smartphone for the six weeks of the study. As an incentive, the participants were supplied with a free one-month transit pass for the city of San Francisco; the requirement of participation was to utilize the transit system on at least five days during the study. Each day, the mobile survey app created a notification on participants' phones, directing them to a survey which asked whether they had used transit on that particular day; if yes, they were asked to provide feedback on that day's experience. In addition to presenting the surveys, the location-aware apps recorded the phone's location every 30 seconds on average. The study also included infrastructure to collect the complete stream of automated vehicle location data from the San Francisco Municipal Transportation Agency (SFMTA) over the course of the six weeks.

Future behavior was captured as participants' intended and desired future travel mode choices. Satisfaction, which has the advantage of being easily communicable to survey participants, was chosen as a primary explanatory variable. The following two measures captured this variable and also provided a means to compare the mobile and online response patterns: an individual's satisfaction with his or her daily travel experience was assessed with the mobile survey which was mapped to the person's actual travel log, and the online entry and exit surveys captured the respondents' overall satisfaction with their aggregate travel experience.

Analyzing the potential link between customer satisfaction and the quality of transit experiences requires an objective measure of what transpired during the respondent's actual trip from point A to point B, including all components of travel time and delays. By establishing quantitative metrics describing what an individual traveler experienced, researchers can begin to flesh out those aspects of experience that appear linked to a person's satisfaction with the service provided. This is the information that could be of most value to the service provider (in this case, the transit agency) who can only influence satisfaction indirectly by controlling the variables that affect it. Without understanding what exact circumstances lead to customer dissatisfaction, a satisfaction survey is of little value to the service provider, as it provides little information on the sensitivity of customer satisfaction to the service quality variables being studied.

Thus, a major strength of smartphone-based data collection is this: External, passively collected data can be leveraged to objectively measure the decision-makers' experiences. In this dissertation, a system is described that linked automatically collected vehicle location data with phone location data and survey answers. By determining a rider's exact route on board a transit vehicle, identifying the precise wait times and in-vehicle travel times from the location data and comparing them to the timetable to identify delays, an objective measurement of the individual travel time components can be made. Additional external data collected during the study also included information on parking pricing, usage of transit payment cards, and accelerometer data. The evaluation of those data, however, is not within the scope of this dissertation and is left to future research projects.

The study resulted in a rich data set which lends itself to a number of different analyses. This dissertation provides a descriptive analysis of the data collected during the SFTQS, followed by the development of two models to explore two questions of interest: The influence of objectively measured travel times on subjective satisfaction, and the influence of satisfaction or dissatisfaction on participants' stated willingness to remain transit riders in the future. An ordinal logit model was used to understand the influence of travel times on satisfaction and a latent variable choice model was developed to help understand the effect on future mode choice behavior.

1.3 Relevant literature

The need to capture personal experiences to understand decision-making has been recognized in the marketing and service studies literature (Woodruff, Cadotte, and Jenkins, 1983). It also finds support with a particular focus on travel decision-making in McFadden (2001), where McFadden describes every choice as being embedded in a lifelong sequence of choices, and where earlier choices affect a decision-maker's current choices via experiences and memories. The complete picture of all influences on a choice also includes the person's motivation and affect, attitudes, perceptions and beliefs and preferences. The choice framework used in this dissertation does not fully capture all those elements, but rather focuses on the influence of personal experiences and satisfaction (which is part of a decision-maker's set of attitudes). Of the remaining variable groups, motivation, affect, perceptions and beliefs have typically been more difficult to capture, especially because they might be specific to individual choice situations and not easily replicated in a stated preference survey. Delivering surveys via smartphones at the particular moment a decision or experience is made provides a new opportunity for collecting information on these variables.

There are several bodies of literature that informed the design of this research project and several others to which it contributes. In this section, only a generalized overview is given, with minimal citations. A more in-depth review with citations is provided in the relevant chapters.

A main area of literature which informed this dissertation is the travel survey literature. To date, multi-day surveys are still relatively rare due to logistical limitations. To ease the respondent burden and enable longer survey periods, a growing number of travel surveys have incorporated GPS data collection, primarily with hand-held GPS loggers. While widespread interest in the use of smartphones for travel data collection is relatively recent, this dissertation reports on the experiences and findings of a large-scale study where the data collection was conducted almost entirely via smartphones. More detailed information can be found in chapter 2.

A second body of literature relevant to this dissertation is the post-processing of automatically collected traveler data to extract information on travel times and travel time distributions. A variety of approaches has been published in the literature that make use of automatic vehicle location, fare card or passenger count data, but the level of detail at which passenger routing and travel times can be extracted is typically limited. In particular, it is relatively difficult to capture out-of-vehicle segments and, depending on whether it is a gated system, to infer alighting stops. Chapter 3 is dedicated to introducing this procedure.

The third body of literature on which this dissertation draws concerns transit rider satisfaction, the subject of chapter 4. This area has been relatively active in recent years, with the majority of contributions aimed at quantifying the importance of different service quality aspects or seeking to develop aggregate satisfaction metrics. With the exception of one study (Friman and Fellesson, 2009), the satisfaction ratings reported by respondents have not been connected to objective measures of service quality. Recognizing that satisfaction is dynamic and changes over time (Mittal, Kumar, and Tsiros, 1999) as a function of the user's personal experiences with transit, it further emphasizes the importance of using panel data for studying this link.

This study was also initially intended to contribute to the literature on travel-related emotions. The interest in the latter is rooted in the study of consumption-related and anticipated emotions in the marketing literature: Researchers studying customer-business interactions in the service industry have reported that experienced positive and negative emotions and anticipated consumption-related emotions are a determinant of a customer's willingness to return. During the design phase of this study, including travelers' emotions was seen as an interesting opportunity to gain insights on a set of variables traditionally not studied. Subjective well-being was included in the surveys in several forms: Participants were asked about their anticipated trip-related emotions in the entry and the exit surveys, and they were asked about the same emotions with respect to their day's transit experiences in the daily mobile surveys. Furthermore, they were asked about their overall mood on the day of the survey. There is a limited body of literature on travel-related emotions that the survey design built on. Literature in psychology has investigated the affective-symbolic relationship between car users and their vehicles, and early experiments in Sweden have indicated that the emotional experience (for instance, whether a trip was pleasing, frustrating, boring, etc.) most likely does have an influence on mode choice and overall satisfaction with transit services. Despite the initial intention of giving a strong focus in this study to travel-related emotions, the primary focus of the data evaluation phase of the study later shifted to satisfaction. A central reason for this was that satisfaction is used as a principal explanatory variable in the choice literature. Furthermore, it was found that due to the simultaneous measurement of satisfaction and emotions, the two were strongly correlated and therefore difficult to include simultaneously in a model. As is mentioned in chapter 6, future research could investigate the difference between models which include emotions variables and models which do not, or could focus on descriptive statistics regarding travel-related emotions. Only one element of the group of subjective well-being variables was included in the final models: the reported mood was successfully included in the models of satisfaction.

The design of the study was further informed by literature on critical incidents, both negative and positive. This term, which also originated in service industry research, describes an experience of a customer that elicits a strong negative or positive emotional response in the customer and remains memorable. Findings in behavioral economics (e.g., Kahneman, Wakker, and Sarin, 1997) indicate that critical incidents might either directly drive behavior or that, when making a decision in the future, the decision-maker may characterize the sum of past experiences as a function of critical incidents experienced as well as the most recent experiences ("Peak-End-Rule"). In the transportation realm, critical incidents have been investigated by Friman, Edvardsson, and Gärling (2001), whose work was a strong motivation for the present endeavor. Friman used hypothetical, simulated situations; more powerful results could potentially be obtained by observing critical incidents and their effects in real-world situations. This, however, necessitates the collection of a panel data set similar to the one collected in this project. In the present study, it was hoped that the effect of negative critical incidents could be observed in the data; however, this was found to be difficult because satisfaction measurements were obtained at irregular intervals and did not necessarily capture all experiences of the study participants. Nonetheless, several types of self-reported negative critical incidents were successfully included in the model presented in chapter 5.

A final body of literature that needs to be mentioned is the travel time reliability literature. For approximately the past 20 years, researchers have been seeking to understand how travel time unreliability affects travel decision-making. In fact, this is the first area where the idea of tracking personal experiences emerged, independently from satisfaction. Due to a lack of appropriate data, many of the studies that have been conducted so far have taken a supply-side perspective where unreliability was characterized by system variables (e.g., a travel time distribution for a particular link or origin-destination pair). However, the experience of variability is highly sensitive not only to the time and location at which a traveler uses a service, but also to personal expectations and norms. Thus, a traveler's decision is not necessarily based on overall average travel times of a transportation service, but rather on the aggregated experiences of that particular traveler, evaluated against his or her norms and expectations. This aggregation of personal experiences was termed the personal travel time distribution, although to date, the data to investigate the formation of this distribution from a user-centric perspective and to identify potential path dependencies and their effects on a person's overall perception of service quality have been lacking. So far, to the best knowledge of the author, there is no pertinent body of literature that proposes user-centric performance measures. Bates et al. (2001b) showed that for public transportation, the personal travel time distribution is much more difficult to derive than for automobile transportation due to scheduled departures, transfers, etc. Because Bates' work was conducted before the advent of real-time information and mobile phones, it is not clear how choice behavior has changed due to new technology. The methodology developed in this dissertation can be used to automatically derive personal travel time distributions from smartphone location data. It is also argued that the capture of personal experience needs to be extended beyond travel time reliability. While reliability may be a powerful component of the personal experience, it is not the only variable attribute of service quality that has an influence on overall satisfaction and behavior; in the case of public transportation, one may think of crowding, cleanliness or the pleasantness of the driver and other passengers, for instance.

1.4 Objectives

This dissertation has four main objectives: First, to demonstrate how combined data from smartphone-based surveys, other automatic data collection systems and smartphone-based sensors can be combined to create rich data sets that allow the investigation of dynamic, long-term processes. A wide range of analyses are possible with data collected in this manner. In the particular case study presented in this dissertation, it is shown how the data so gathered can be postprocessed to provide an unfiltered, objective measure of the quality of the transit rider's experience and of the environmental variables at the time of the experience, all without resorting to asking the participant. This dissertation outlines a framework that is useful beyond the specific transit application; the methodologies and processes can be applied to other situations where customer satisfaction with transportation services is investigated. In particular, these can be used to guide the design of future transportation systems, where emerging automated, shared and flexible on-demand services are combined with "traditional" transportation services or replace the latter. The ability of researchers and practitioners to quickly gather data on customer satisfaction and post-process it with a scalable infrastructure could potentially be used to monitor such new systems, to gather insights into how they are received by customers as they are deployed, and to adjust their design as needed.

Second, this dissertation aims to demonstrate the use of these new data to gain fresh behavioral insights. Specifically, the link between travelers' objectively measured travel times and their subjective satisfaction ratings regarding those travel time components is investigated. Models are developed that quantify this link while controlling for various sociodemographic characteristics of the decision-makers. It is shown that both the baseline (overall) satisfaction with travel times and the person's general mood on the day of the survey contribute significantly to the evaluation given to a specific experience. These are two elements that are typically not found in traditional travel surveys. The modeling frameworks developed in this dissertation are also intended as a basis for moving toward more user-centric performance measurement (e.g., door-to-door travel time), away from purely system-centric metrics (such as on-time departures) as they have been customary in the transportation industry for a long time.

Third, as the research project described in this study used smartphones exclusively for data collection, a part of this dissertation is dedicated to reporting the experiences made and lessons learned in the course of designing and implementing this study.

The fourth objective is to make the link between satisfaction on a daily scale, overall satisfaction at the end of the study period, and future behavior. In the case study at hand, this is modeled with a latent variable choice model. To the best knowledge of the author, this is the first model that relates a traveler's experiences with day-to-day variability in service quality to that person's willingness to remain a transit rider in the future. Thanks to the model, it is possible to quantify the impact of various problems, such as a person being left behind at a stop or a late arrival at work, in terms of minutes of in-vehicle travel time. Via simulations, the contribution of delays and other service quality problems to rider attrition can be determined, which can serve as an input to cost-benefit analyses of improvement projects. In future applications, this model framework can be adapted for use with new transportation services such as mobility on demand systems. This would allow researchers to understand how travelers adapt to such new systems and how traveler behavior is influenced; this information, in turn, can be used to adjust the design of new systems being deployed.

1.5 Contributions

This dissertation contributes to the current state of the literature as well as to the planning and design of future transportation systems. With respect to the literature, there are five main contributions:

1. It demonstrates the power and the value of conducting travel surveys via smartphones and connecting the survey data with other data sources such as phone sensor data and vehicle location data. It argues that this (a) permits the collection of data sets that

capture a more complete picture of the participants' decision environment, and (b) allows researchers to conduct studies on complex, dynamic phenomena.

2. It reports on the researchers' experiences with the design and implementation of a large-scale study with over 800 participants that was conducted primarily via smartphone app, where participants were asked to fill out multiple daily mobile surveys about their travel experiences in addition to online surveys at the beginning and the end. This is intended to inform and support the development of future studies.

3. It presents a computational system to match automatically collected vehicle location data with smartphone location data in order to measure the personal experiences of travelers. With the combined data, it is possible decompose a traveler's overall travel time into its components, i.e., into in-vehicle travel times, wait times at the stop and transfer times.

4. It develops a quantitative model to compare the objective measures of travel times with the subjective satisfaction ratings and to explain satisfaction directly as a function of observed travel times.

5. It develops and presents a model that connects observed experiences and critical incidents with satisfaction and future behavior in one framework. With the help of models like this one, it will be possible for analysts to understand the effects of service quality problems on ridership and to develop more user-centric performance measures.

A major finding in the case study presented in this dissertation is that in-vehicle delays have a stronger impact on customer satisfaction than wait times at stops. This implies that there is a strong need to focus investments on stabilizing vehicle running times, for example with signal priority systems, dedicated lanes, or by upgrading conventional routes to Bus Rapid Transit routes. The results of this study also give a tentative indication that customer satisfaction with travel times, i.e. with operational aspects of the service, has a larger impact on ridership than satisfaction with the a group of variables related to the travel environment (including cleanliness, crowding, the pleasantness of other passengers, and safety). However, it should be noted that at present, the model results do not allow a comparison in monetary terms (e.g., does one dollar spent on operational improvements yield larger gains in customer satisfaction than one dollar spent on travel environment improvements?) since the scales of the associated variables are only in units of customer satisfaction. Further research is required to establish the link between investments and satisfaction. Another finding is that due to real-time information, passengers have begun adapting their arrival times at the bus stops to coincide with predicted bus departures. This has several implications: Relatively speaking, investments into stop amenities such as shelters might become less important in the future, but since passengers can choose to spend their wait time in nearby shops, the case for transit-oriented development around transit routes is strengthened. Based on the results of this dissertation, it can also be argued that since the relative importance of wait times has decreased due to real-time information, increases in service frequency may be a less effective strategy to retain customers relative to improvements of running times. Of course, one needs to bear in mind that the majority of data were collected on high-frequency routes, so the results mainly reflect the marginal benefit of increasing frequency on a route whose typical headways are already 12 minutes or less.

Beyond these immediate contributions to the literature, this dissertation is intended to help shape the future of transportation planning by establishing a framework through which transportation agencies can collect "business intelligence", i.e., insights on customer satisfaction and the drivers thereof. This can serve as a basis for decision-making on an operational level and on a longer-term strategic level, and can directly affect large capital investment decisions. The experiences with the study show that when travelers are given a channel to provide direct feedback about their experiences with transit use, and they know that their feedback will be taken into consideration, one can observe a high willingness among participants to share information. In fact, many participants were willing to do more than they were asked to: The average number of daily mobile surveys filled out by participants exceeded the required minimum by a factor of two, and of the 686 participants who filled out the exit survey, 36% indicated would have preferred to fill out surveys more frequently than they were asked to during the study and 25% indicated they found the surveys too short. The technology used in this study could be further developed into a two-way communication channel between the user and the researcher or transit operator. Potential uses could be to identify the segments of riders that are affected by service disruptions and to offer them personalized compensation or to provide personalized guidance for first-time users. This type of system holds large potential to reduce ridership loss by tracking passengers who have had bad experiences; targeted interventions could be conducted to retain existing riders and reduce attrition. Potential interventions could be built on knowledge of the dynamics of transit use and of habit formation, as can be derived from studies such as this one.

Looking into the more distant future, mobility on demand systems have the potential to replace low-demand routes currently served by traditional fixed-route public transportation. Operators will be looking at how to optimally integrate mobility on demand systems, which may be automated or privately operated, with high-demand bus or rail corridors. Survey systems such as the one presented in this dissertation allow a researcher to obtain a very detailed picture of the door-to-door origin-destination matrices, to segment the ridership market and to understand, via personal surveys, the preferences of individual riders. This could significantly support the design of new mobility on demand systems and the negotiation of contracts between public and private operators. Especially at the point where such systems are relatively new and little is known about rider preferences, smartphone-based surveys could be used to discover user preferences related to issues such as the sharing of the vehicle with other users, or tolerance of detours to pick up or drop off other users. If parts of the service are privately operated, the two-way communication link between the public agency and riders would be a vital element of quality control. Last but not least, if traditional public transportation is to stay competitive compared to these new types of services, more and larger-scale studies such as the one presented in this dissertation will be required to identify the most urgent needs for improvement.

1.6 Organization of this dissertation

The remainder of this dissertation is organized in four chapters: In chapter 2, a description of the design and implementation of the San Francisco Travel Quality Study is given. This chapter describes the motivation behind the design decisions and some of the most important experiences made during implementation. It also gives an overview of key data, including respondent characteristics and response rates. Chapter 3 presents a computational system to match smartphone location data with automatically collected vehicle location data to automatically derive passenger travel times, wait times and transfer times. In an example application at the end of this chapter, some analyses of wait times and overall travel times are given. In chapter 4, the ordinal logit model to match travel times and satisfaction is developed. Chapter 5 then builds on those findings to develop the more complex latent variable choice model which represents future choice behavior as a function of overall satisfaction, daily satisfaction and personal experiences. Estimation results and findings are presented for both models and discussed. Finally, in chapter 6, the dissertation is summarized and a future research outlook is given.

Chapter 2

The San Francisco Travel Quality Study

2.1 Introduction and motivation

Most national or regional household travel surveys are only conducted approximately every 10 years and cover a narrow time frame (e.g., 1-2 days) due to the scale, required logistics, and costs. The data collected from these surveys rarely capture choice behavior that evolves over time or lend themselves to investigating questions that require repeat measurements or specialized survey questions. There is a clear need for researchers and planners to conduct travel surveys more frequently, perhaps covering smaller geographical areas but allowing for specific questions that large-scale surveys omit. Thanks to the proliferation of location-aware smartphones and increasingly cheap cloud computing resources, the costs of conducting such surveys are becoming less prohibitive, such that the regular administration of targeted, smartphone-based surveys is now realistic. Overall, the use of smartphones for data collection presents an opportunity to greatly improve travel behavior research through more specific and personalized questions, direct interaction between researchers and participants, and the collection of long-term panel observations. The importance of collecting such data sets was highlighted by McFadden (2001): An important input to the choice process is the decision-maker's history of past experiences with similar choice situations. The respondent burden associated with collecting such panel data sets with paper and pencil surveys is very high. Smartphone-based survey methods, on the other hand, give researchers the ability to capture a decision-maker's choice history over extended periods of time at significantly lower cost and respondent burden. They also allow for a more accurate and targeted observation of situation-specific variables, such as personal motivations for making a choice, affect and perceptions and beliefs related to the information about available alternatives. It is preferable to capture those variables as closely to the decision-making process as possible.

In this chapter, we describe the design, implementation and outcome of the San Francisco Travel Quality Study (SFTQS), a large-scale, smartphone-based study that ran for six weeks from October to December 2013 with an initial total of 856 participants, 757 of whom downloaded a survey app for Android smartphones (see section 2.5 for demographics and participation rates). While the survey period was six weeks, reporting was not necessarily continuous; participants were only required to report on their transit travel for five days out of those six weeks, and the five days were not required to be consecutive. However, many participants reported on more days. The study was motivated by findings by Perk, Flynn, and Volinski (2008), who reported that many large transit agencies in the United States tend to see high rates of ridership turnover, but that the causes of that turnover are not well understood. Aside from users moving away from the service area or undergoing lifestyle changes due to, for example, having kids, Carrel, Halvorsen, and Walker (2013) found that problems with service quality might be linked to ridership losses. The study was partially funded by the University of California Transportation Center and the National Science Foundation, and its stated goal was to investigate the link between individual travel patterns, transit service quality, personal experiences, satisfaction, subjective well-being and future behavior of transit users. There was a particular interest in observing not only the dynamics of these variables over time, but also the experienced trip-level service quality for every individual participant, thus reducing the need to use aggregate measures in the analyses. To facilitate this and to minimize respondent burden, the survey app continuously collected location and mode use data. To quantify service quality, a complete set of automatic vehicle location (AVL) data from San Francisco's transit service, Muni, was obtained, which could then be combined with the phone location data. Information was also gathered on on-street parking prices in downtown San Francisco during the study via the SFPark system. On the survey side, the study collected a broad range of data on sociodemographics, attitudes, current travel patterns and intended future mode use through an entry and an exit survey. Several daily mobile surveys were conducted (reporting period = 1 day). These surveys were taken on an app installed on the participant's smartphone. They first asked participants whether or not they had used transit on that particular day, and if the participant answered "yes", they were asked to respond to a set of questions regarding their experience using transit. There was a daily survey prompt, i.e. a notification on the respondent's phone asking them to respond to that survey. Two sides of the customers' perspectives were captured: satisfaction and emotions (e.g., whether a travel experience was frustrating, relaxing, etc.). The latter is novel in travel surveys; more information is given in section 2.2. Thus, a rich data set was collected with which a wide variety of questions can be addressed. The study protocol is described in more detail in section 2.4, where a graphical time line is also provided.

The immediate goals of this research were to:

1. Observe the dynamics of transit use in an urban environment at a very detailed level.

2. Develop and test a system to quantify users' personal experiences with service quality via location data. This is embedded in a broader push toward more user-centric performance measures to supplement existing operational measures.

3. Observe the dynamics of transit riders' satisfaction as a function of personal experiences.

4. Model the relationship between objectively measured service quality and users' subjective satisfaction with the experience.

5. Study the role that trip-related emotions play, including their dynamics, the link between emotions and short- and long-term satisfaction, and the influence on future mode choice behavior.

6. Study the effect of users' personal experiences with service quality (in particular reliability) on intentions and desires to remain transit users in the future (i.e., ridership retention).

Beyond these immediate goals, this data set was intended to facilitate future research on other topics. Without smartphone technology and the ready availability of cloud computing resources, it would have been impossible to execute this innovative study at its large scale and low cost.

Section 2.2 gives a brief overview of the literature underpinning the various aspects of the study. Section 2.3 describes the design of the survey instruments and section 2.4 gives an overview of the study protocol, followed by a discussion of key outcomes in section 2.5. Section 2.6 presents the conclusions and gives an overview of research questions that can be investigated with these data.

2.2 Literature review

This section consists of three subsections: First, some previous examples of larger panel travel surveys and smartphone-enabled surveys are discussed, followed by a presentation of travel satisfaction literature and then a subsection on emotions and subjective well-being in travel.

Surveys

Multi-day travel surveys are still relatively rare. As explained by Cherchi and Cirillo (2010), they can be categorized by whether the survey was repeated several times for short periods or whether data were continuously collected over a longer time. Yáñez, Mansilla, and Dios Ortúzar (2010) provide a recent example of research based on the former. Two of the largest panel surveys to date have been the Mobidrive survey, which was conducted in Germany in 1999 (Axhausen et al., 2002) with 317 participants and a survey period of six weeks, and the Stability of Travel Behavior survey which was conducted in Switzerland in 2003 (Löchl, 2005) with 230 participants and a survey period of six weeks. A 12-week panel survey, with 71 participants, was conducted by Schlich, Simma, and Axhausen (2004) to investigate the stability of leisure travel. Shorter panel data sets have been collected since then, including Roorda and Miller (2004); Stopher, Clifford, and Montes (2008). In an overview paper, Ortuzar et al. (2011) review several long-term mobility surveys (of the first type) and argue that they are superior to one-time, large surveys conducted at large intervals. The surveys presented covered between one and seven days in regular intervals, although most of them are not panel surveys. Cherchi and Cirillo (2008) showed that for the purpose of mode choice modeling, one week was sufficient. However, for investigating other dynamic aspects of travel behavior, longer surveys remain valuable, as is evidenced by the large body of research that has been based on Axhausen et al. (2002).

Recently, a growing number of travel surveys have incorporated GPS data collection. Typically, respondents were equipped with hand-held GPS loggers they were asked to carry around. Several researchers have combined the GPS data with manual reports by the study participants; a special case of this are prompted recall surveys where participants are asked to recount their travel with the help of GPS data (Auld et al., 2009; Stopher and Collins, 2005; Wolf et al., 2004; Oliveira et al., 2011). It has been found that participants in manual surveys tend to under-report trips by as much as 20-25% of all trips, particularly shorter ones (Bricka and Bhat, 2006; Stopher, FitzGerald, and Xu, 2007). Schüssler and Axhausen (2008) worked on identifying trip characteristics from the GPS data without further information, based on one week of data from approximately 4,900 participants. Unfortunately, the use of GPS loggers is costly because they need to be distributed to and collected from respondents. Accuracy may also suffer since participants need to remember to carry an additional device. To overcome these limitations, there has been mounting interest in conducting travel surveys via smartphone apps. In a first series of small-scale experiments, Asakura and Hato (2004), Ohmori, Nakazato, and Harata (2005), and Itsubo and Hato (2006) demonstrated the accuracy and effectiveness of this methodology, but battery drainage caused by the GPS sensor was found to be a serious problem. Jariyasunant et al. (2014) conducted a 3-week, 135-subject experiment producing automated travel diaries to derive information on the environmental, financial, health and time impact of participants' travel patterns. The information was presented to participants with the goal of inducing more sustainable travel behavior. The author is aware of two larger smartphone-based travel surveys for which data collection is currently in progress. One is the Future Mobility Survey (FMS) (Cottrill et al., 2013), and the other is the PEACOX 7th Framework Project. The FMS aims to recruit 1000 participants in Singapore who are asked to collect location data for two weeks and manually validate their activities for at least 5 days, and the PEACOX project involves a one-week survey of 400 participants in Basle.

Satisfaction with public transportation

Transit customer satisfaction surveys are the most widely used and most direct way of capturing customers' perspectives on the service being delivered (Hensher, Stopher, and Bullock, 2003a).

An good overview of customer satisfaction survey methods for public transportation is given by Brög and Kahn (2003). The majority of satisfaction surveys described in the literature were conducted only once, asking respondents to rate their satisfaction with overall service. Most researchers have sought to quantify the importance of different service quality aspects or proposed aggregate satisfaction metrics (Eboli and Mazzulla, 2011; Stradling et al., 2007; Swanson, Ampt, and Jones, 1997; Del Castillo and Benitez, 2013). Several important angles that have not received much attention in the past are as follows:

1. Satisfaction is dynamic and can change over time (Mittal, Kumar, and Tsiros, 1999). Given the routine nature of transit use, little is known about the stability of satisfaction ratings in this domain or about the relationship between satisfaction and daily experiences. Friman, Edvardsson, and Gärling (2001) found evidence that negative critical incidents, i.e., memorable negative events, could affect satisfaction ratings even with aspects of the service that were not directly related to the incident. Especially in congested transit networks, service quality can be variable, potentially making critical incidents common.

2. Satisfaction is a function of personal use experience (Anderson and Sullivan, 1993). To understand the relationship, subjective customer satisfaction and objective measures of the individual's experience must be linked (Davis and Heineke, 1998). This necessitates high-resolution data on an individual's interaction with the service. Since transit service quality is inherently stochastic (particularly reliability), it is not predictable when observable negative critical incidents will occur.

3. Satisfaction and behavior are correlated (Klöckner and Matthies, 2004; Abou-Zeid et al., 2012). Behavioral modifications of transit users can occur on different time scales: short-term changes due to immediate service problems (where the users later return to their habitual choices), or long-term changes due to repeated dissatisfying events. The duration of a customer's relationship with a service provider has been modeled as a function of cumulative satisfaction *and* recent experiences in other industries (Bolton, 1998), but little is known about the sensitivity of behavioral intentions to these variables in public transportation. If successful, research on this subject would be able to shed light on the practically unexplored subject of ridership retention.

Thus, a need was identified for a long-term, high resolution data set that observed transit passengers' travel behavior, personal experiences and satisfaction.

Emotions and the travel experience

The influence of emotions, both experienced and anticipated, on the consumption of goods and services has been broadly recognized in marketing research (Hansen and Christensen, 2007). Although emotions are not typically included in customer satisfaction surveys, some authors argue that they are integral to understanding not only customer well-being and satisfaction during the consumption experience, but also future behavior (Oliver, 1993). The model of goal-directed behavior (Perugini and Bagozzi, 2001), a framework extending the theory of planned behavior, includes anticipated emotions as an antecedent of behavioral desire, intention and actual behavior. Smith and Bolton (2002) argue that customers' emotional responses can be particularly important in the case of service failures and negative experiences, though the effects vary across industries. In transportation, this has so far only been investigated in the airline industry (Dubé and Menon, 1998) and to a certain degree in the commute stress literature (Evans, Wener, and Phillips, 2002). Homburg, Koschate, and Hoyer (2006) showed experimentally that the role of cognition in evaluations of repeated service encounters increases and the role of affect decreases over time, but that this was attenuated by inconsistent service quality experiences.

Several publications that have analyzed affective predictors of mode use with focus groups or small, one-time surveys (e.g. Mann and Abraham, 2006; Gatersleben and Uzzell, 2007), either with respect to anticipated emotions or the affective-symbolic relationship of car owners with their vehicles. Carrus, Passafaro, and Bonnes (2008) specifically investigated the influence of anticipated emotions in predicting use of public transportation. The above studies tended to focus on invariant aspects of service, but Gardner and Abraham (2007) noted the importance of negative affect caused by serious service failures in predicting future use of public transportation. Finally, the difference between anticipated emotions (which are the basis of decisions) and remembered emotions (from experiences) in transportation decisions is elaborated on by Abou-Zeid and Ben-Akiva (2010).

Outside of the transportation realm, Kahneman et al. (2004) pioneered a survey method for characterizing daily life experience, including with respect to satisfaction and to emotions. A notable large-scale study on emotions conducted via smartphone is by Killingsworth and Gilbert (2010), and more generally, the Experience Sampling Method (e.g., Csikszentmihalyi and Larson, 1987) has been adapted in a variety of studyes to sample participants with the help of a beeper, cell phone notification, or programmed watch (Connor, 2014).

Conclusions and contribution

In summary, a need was identified for a panel survey that would collect repeated measurements of satisfaction with transit travel, experienced service quality, overall travel patterns and future intended behavior. This was a unique opportunity to investigate the dynamics of travelers' satisfaction and trip-related emotions as well as the links to behavior; these are areas that are recognized to be important in marketing literature but have so far not been studied in the transportation context.

While GPS-supported and smartphone-enabled travel surveys have either been carried out before or are currently underway, the SFTQS represented a new level of a smartphone-based travel survey in going beyond collecting only tracking data. By fusing the location data with time-prompted daily mobile surveys and with vehicle location data, it was possible to collect an additional wealth of information at relatively low cost. The methods and infrastructure used in this project could easily be applied to studying users of other transportation modes as well.

2.3 Design of survey instruments

Participants were asked to fill out an online sign-up survey at recruitment, an online entry survey at the beginning of the study, and an online exit survey at the end of it. In addition, they were asked to use Muni on at least five days during the study and to fill out the corresponding daily mobile surveys. The sign-up survey included sociodemographic information, residential, work and school locations, frequency of travel in San Francisco, and frequency of mode use before the study. The number of participants was capped at 1,000, so if the number of sign-ups had exceeded 1,000, the information from the sign-up survey would have been used to select a sociodemographically representative sample and a mix of transit users and non-users. The entry survey, which was taken in the first week of the study, asked for additional information on length of transit use, duration of residence in San Francisco, household structure, education, occupation and student status, and the cost of parking at the employment or school location. The entry survey was conducted separately from the sign-up survey in order to limit the response burden on participants who did not get selected into the study. It also included a set of questions that were presented in both the entry and exit survey since it was expected that the responses might change during the study:

- Intended and desired mode choice in 2014: Whether participants planned or wanted to use Muni, the car, bicycling, etc. more, the same or less than in 2013.

- Overall satisfaction with a number of service quality aspects.
- Whether the intended/desired changes were related to any of those aspects.
- How often participants expected to encounter transit reliability problems during the study.
- Whether participants avoided any Muni routes due to service quality problems and what they did instead.
- Anticipated emotions when using Muni and when driving in San Francisco.
- Attitudes toward car travel in San Francisco.

Intended and desired future mode use were to be used as choice indicators, i.e., any changes in these variables between the entry and the exit survey were assumed to be attributable to events observed during the study. The exit survey also asked how the participants intended to pay for Muni after the study, whether they could see themselves using Muni in the long run, and what their perceived average wait times were during the study. Lastly, a set of questions was asked about participants' motivation to take part in the study and feedback on various aspects of the study was collected. An ex-ante assessment of the response burden can be found on page 20.

The smartphone-based surveys that were filled out during the study contained a subset of the satisfaction and emotions ratings, and the question prompt focused on the user's experience during that day only. Participants were also asked how they generally felt on that day, and an optional free-form text box was provided for any comments about the travel experience. The question on general feeling was intended as a control variable since subjective well-being and mood can interact with people's assessment of an experience (Ettema et al., 2010).

A major challenge with smartphone surveys is limited screen space which, in combination with the consideration that participants were being interrupted in the middle of their day, meant that only a small number of satisfaction and emotions items could be asked in the daily mobile survey. The selection of those items for the surveys is explained in detail below.

Satisfaction

Many of the satisfaction variables in the entry and exit survey were related to travel times since those were being observed via tracking data. This included satisfaction with overall reliability*, waiting time reliability, in-vehicle travel time reliability, transfer time reliability, travel times on board when there are no delays and frequency of service. A transfer was defined as an event where:

- A passenger used two transit routes consecutively.
- The stops where the passenger alighted from the first route and boarded the second route were no more than 100m apart.

- During the time between alighting and boarding, the passenger remained within a distance of 100m of one of the two stops.
- During the time between alighting and boarding, there was no more than one departure on the second route that the passenger did not take.

Thus, transfer time includes both transfer walk time and transfer wait time. Based on pertinent literature and on personal communication with staff of the San Francisco Municipal Transportation Agency (SFMTA), the following items were added: Crowding*, comfort, cleanliness*, personal safety*, pleasantness of fellow passengers* and the accuracy of real-time information*. Due to participant feedback, staff competence, staff friendliness, the ease of use and reliability of the fare payment system were added in the exit survey. The entry and exit survey asked about overall satisfaction with these aspects. In the daily mobile survey, only 9 satisfaction items could be included. Those that were expected to be most variable were retained (marked with an asterisk above). Satisfaction with travel times was decomposed into three items: wait time, in-vehicle travel time, and transfer time. The question prompt did not specify reliability for those items. Satisfaction was rated on 5-point Likert scales. In the daily mobile survey, the extreme points were labeled with a frowny and a smiley face.

Emotions

The selection of emotions was based on the Consumption Emotion Set (CES) (Richins, 1997), a list of consumption-related emotions widely used in marketing research. The challenge was to adapt the extensive CES to the travel domain, and a maximum of eight items could be included due to space limitations. For this purpose, a small pilot study was conducted: First, emotions that were not considered to be related to service quality on public transportation (e.g., "romantic love") were removed from the list, and the remaining 52 emotions were assembled in a paper-and-pen survey in March 2013. In total, 60 questionnaires were handed out to SFMTA transit riders during a weekday evening rush hour. Respondents were asked to report how they felt after their evening commute trip by rating every emotion on a 5-point Likert scale from "do not feel that way at all" to "feel very much that way". Candidate emotions were first isolated from the results by determining, through visual analysis, which ones showed acceptable variation across responses. An exploratory factor analysis was then conducted to isolate correlated groups, and the descriptors that loaded the strongest on one of the factors were selected as labels. Finally, the list was adjusted with input from UC Berkeley psychology faculty to include positive and negative affect and arousal, as postulated by the circumplex model of emotion. This selection resulted in the following emotions to describe a person's travel experience: Pleased, frustrated (representing the basic emotions of contentment/happiness and anger), relaxed, bored (positive/negative deactivation), impatient, stressed (negative activation), in control (explained below), disgusted. During the survey design, discussions with various volunteers who tested the survey revealed that "disgusted" was often confused with "frustrated", so in the final design, that descriptor was changed to the colloquial term "grossed out". The "in control" variable was found to form its own factor during the exploratory factor analysis. It can be interpreted as positive activa-

	Points (min)	Points (max)	Personal reminder	Estimated time
Sign-up survey	86	86	No	10 min
Entry survey	213	262	Yes	16-18 min
Exit survey	271	318	Yes	20-22 min
Daily mobile survey	30	30	Yes	1 min

Table 2.1: Ex-ante response burden assessment using the point-based system proposed by Axhausen and Weis (2013).

tion. Fear was dropped since it was considered to be strongly correlated with the personal safety satisfaction rating.

Ex-ante response burden measurement

Pre-tests of the survey instruments were conducted informally with university affiliates and SFMTA personnel, as time constraints did not permit a larger-scale pilot study with members of the general public. The ex-ante response burden was assessed using the points-based system proposed by Axhausen and Weis (2013). The results are shown in shown in table 2.1; for the entry and exit survey, a range was estimated to account for optional questions. Recorded response rates and time to complete the surveys are discussed on pages 28 and 32.

2.4 Study design and administration

Study protocol

An overview of the study is shown in figure 2.1. Recruitment began in early October 2013. Principal recruitment channels were email lists maintained by the SFMTA and various large employers and universities in San Francisco as well 23,000 anonymous postcards sent by mail. In addition, a stand was set up at a central location of San Francisco State University during one afternoon and approximately 500 fliers advertising the study were handed out. The fliers directed potential participants to the same web address as the emails and postcards. To be able to participate, a person had to (a) be at least 18 years old, (b) live, work, or go to school in San Francisco, and (c) have an Android smartphone with a data plan. Participants were directed to a website where they received more information on the study and were asked to read the consent form. To sign up, they filled out the online sign-up survey. The list of interested participants was screened for the eligibility criteria and everybody who met the criteria was enrolled and notified.

The study officially commenced on October 21, when participants were asked to fill out the entry survey. On October 27, they received an email with a link to a survey app for their smartphones. Once installed, the app immediately began collecting location and mode use data at approximately a 30-second interval. Participants were asked to continue traveling as normal during the week of October 28. That week, they received a letter in the mail with a free, nontransferable pass valid for unlimited trips on the Muni network. Unfortunately, due to distribution problems, a number of participants did not receive their passes by the end of that week, so the study group was split into two cohorts. Cohort 2 received replacement passes between November 13 and November 15. In the following description, the dates for cohort 1 are given in the text and the dates for cohort 2 in parentheses.

Participants received instructions exclusively by email. Emails were customized with the participant's name, and they were given an email address to contact if they had questions or problems. To the extent possible, every participant who wrote an email received a personal response. It was hoped this would minimize attrition (Dillman, 2011). Participants were instructed not to use the pass before November 4 (November 16). Since it was assumed that the availability of the free pass might cause a change in some participants' travel behavior (Fujii and Kitamura, 2003), the location tracking data for October 28 through November 3 (November 16) could be used to observe their habitual travel choices. The passes were valid through December 1 (December 15). During that time, participants were asked to use Muni on at least five days and to fill out the corresponding daily mobile surveys. They received a daily reminder on their phone at a preset time which they chose in the sign-up survey. The suggested default was 19:30. Once they had reached five daily mobile surveys, they were given the choice of whether or not to continue the daily reminders. After December 8 (December 22), they received an email with a link to the exit survey. Participants were asked to keep the smartphone app and the location data collection running until they took the exit survey, but they were free to uninstall it afterwards.

With the exit survey, two further optional mobile surveys were presented: One that was identical in layout to the daily mobile survey but asked about their experience during the entire study, and one that asked participants to rate their emotions at that very moment. The latter used an extended set of emotions compared to the main survey: Angry, bored, calm, contented, disgusted, excited, frustrated, grossed out, happy, impatient, in control, pleased, relaxed, sad, stressed, worried. The additional items were based on a summary by Laros and Steenkamp (2005), and the purpose was to evaluate other possible emotional descriptors for future research.

Regarding the passes, there was one large mishap: Due to a misunderstanding, the SFMTA canceled all passes on December 1, including those of cohort 2. Although the passes were restored to the cohort 2 cards by December 3, it took several days until card readers on all vehicles would validate the passes again.

Cohorts, duration and control group

Of the original 856 participants enrolled in the study, 18 either withdrew or switched to a phone with a different operating system during the study, 669 were in cohort 1 and 169 in cohort 2. The survey period was driven by the free transit pass that was provided to participants, as the minimum validity of that pass was four weeks, and location data needed to be collected for at least a week before and after that period. 22 participants did not use the mobile app but filled out the sign-up, entry and exit survey and could therefore serve as a control group. It is unknown why they did not install the mobile app.

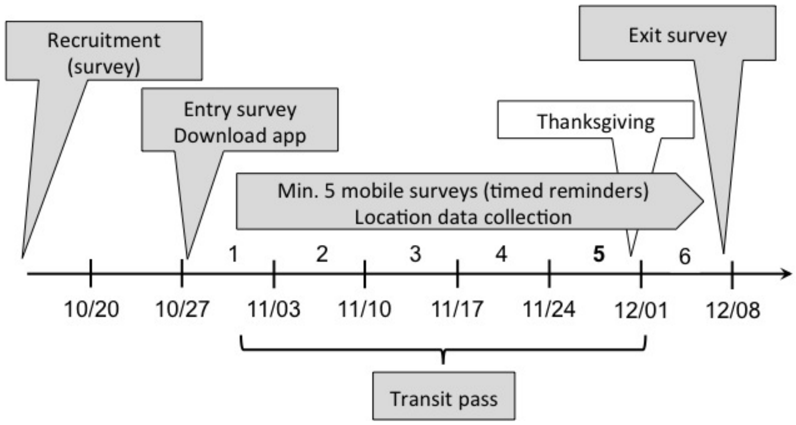


Figure 2.1: Time line for the study.

Remuneration of participants

Incentives for survey participation were the subject of much debate. As was mentioned above, participants received an unlimited transit pass on a smart card, worth US$69 in total. Aside from the passes, two other options were discussed: Cash incentives and a lottery with a large cash prize. A literature review revealed that fixed cash payments have been found to outperform lotteries and donations to charity by many studies. In fact, the latter two have been found to have small effects, as has the amount of money offered in the raffle (Halpern et al., 2011). In addition, ethical concerns have been raised about lotteries (Tooley, 1996), and in some jurisdictions a variety of legal requirements may need to be met. Lastly, it has previously also been found that prepaid fixed cash incentives were superior to postpaid incentives (Simmons and Wilmot, 2004). For the SFTQS, a fixed cash incentive might have been best to also motivate non-transit users to participate, but necessary funds could not be secured. Therefore, the transit pass was chosen as a prepaid, fixed-value incentive. It should be noted that the superiority of prepaid incentives compared to postpaid incentives might be more pronounced for one-time surveys than for longer panel surveys such as the SFTQS. As noted in the following subsection, the fact that multiple reminders needed to be sent out to improve response rates for the exit survey, which was taken after six weeks, suggests that a post-paid incentive or a combination of pre- and post-paid might have been more adequate given the duration of the study.

Challenges

The study faced several challenges:

- The study protocol was relatively complex. There was no personal researcher/subject interaction, and all instructions were sent out by email. It was found that many participants either did not read or forgot the instructions.

- Participants' compliance time in installing the survey app and responding to the entry and exit surveys varied strongly. In particular, the latter required several reminders to be sent out.

- For the app to be able to collect location data, users had to enable location services and allow the app to access them. Despite not using GPS (see page 23), battery consumption was found to be an issue, which caused many participants to deactivate location services at some points during the study. Therefore, location data availability varies across participants.
- Several participants requested a version of the daily mobile survey that allowed them to respond for a previous day if they had forgotten to do so. That version was pushed to participants' phones on November 4. A small group of participants who were experiencing technical difficulties also requested an online version of the daily survey. That was created on November 24.
- Due to proposed statewide legislation in California, the researchers were not able to obtain access to the transaction records of the fare cards used by participants.
- Prior to the experiment, there were concerns about two types of fraud: it was feared that (1) participants might sign up only to receive the free pass, and that once they had received it they would no longer participate in any study activities, or (2), that they would sell the free monthly pass or give it to a friend or family member. Due to those concerns, participants' instructions clearly stated that the pass was nontransferable, that misuse could result in a fine, and that it could be terminated if the participant did not take part in study activities. During the study, no evidence was found of passes being sold or given to friends/family members, and many of the participants who did not participate in all study activities were also found not to have used their passes. However, as they were given until the last day of the study to complete the five daily mobile surveys, in practice it would have been difficult to terminate passes due to non-cooperation. No passes were terminated for that reason.

Infrastructure and data processing

The software infrastructure behind the study included the survey app and two cloud servers. The survey app was a customized version of Open Data Kit Collect, with added location tracking, survey alert functions and automated survey management. The app required users to log in once at the beginning of the study with a user name that was sent via email. Location tracking was done without GPS, only using Wi-Fi and cell tower positioning to reduce energy consumption. It used Google Play services, which also features an accelerometer-based classifier to determine whether the user is walking, biking, in a vehicle, actively using the phone or immobile. That output was used as the mode use data. One of the cloud servers handled the distribution of survey definition forms and the collection of the survey responses, whereas the other server had the following functions: User authentication, updating of survey notification times on the phones and collection of the location and mode use data. Location data were stored in a geospatial database. At regular intervals, the raw data were postprocessed to infer trips and activities with a DBSCAN clustering algorithm, and then the location data were matched with the AVL data to infer transit trips as described

	Excluding passes	Including passes
Per transit response	\$1.63	\$8.11
Per participant-day in study	\$0.30	\$1.47
Per participant-day of tracking data	\$0.55	\$2.75

Table 2.2: Study costs per data unit. Passes were not a direct cost, so the costs are shown including and excluding passes.

in chapter 3. While of high resolution, these data were also completely unlabeled, i.e., the accuracy of the inferences was not verified by the survey participants. The accuracy had previously been tested with ground truth data collected by the researchers. However, manual inspection later showed that the false negative rate (i.e. undetected transit trips) may have been significantly higher than expected due to differences between participants' travel patterns and those tested by the researchers during development: Out of the 8,677 combined days on which participants reported having traveled on Muni, phone location data was missing for 1,002. In the first evaluation run, only 3,010 surveys could be matched to vehicle location data beyond doubt, but future improvements of the detection algorithms might increase that number. Nonetheless, it would have been beneficial to ask the participants to also report ground truth data on their transit use.

Costs

The direct costs of the study were as follows: \$3,500 for recruitment and mailing of the transit passes, \$3,200 operating costs for the server infrastructure, \$8,200 in staff time for the development of the app and the server backend. The study was administered by one full-time researcher with occasional support by SFMTA staff. The total value of the free transit passes was approximately \$59,000. However, those were not a direct cost as they were provided free of charge by the SFMTA. Table 2.2 shows the costs per completed daily mobile survey that related to a transit experience (based on a total of 9,116), per participant-day in the study (856 participants minus 18 withdrawals, for 60 days), and per participant-day of tracking data (26,914 participant-days with 100 or more location points).

2.5 Key data

Overall, the study was considered a success. Out of the 856 participants who signed up, 757 completed the entry survey and 686 completed the exit survey. The survey app was installed on approximately 800 devices, and a total of 13,931 completed daily mobile surveys were submitted by 774 unique users. Of those completed mobile surveys, 7,682 pertained to transit trips made on the day of the survey, 1,434 to transit trips from previous days and 4,815 were reports of not using transit. This amounts to an average of 11.8 transit-related responses and 18 total responses per participant, whereas the minimum required number of responses had been set at 5. A full data set, including sign-up, entry and exit surveys and 5 or more trip-related daily mobile surveys, could be obtained from 604 participants. Location data was collected from 732 unique users.

Demographics

Key demographic information is shown in table 2.3. The table compares the numbers for the SFTQS, the SFMTA 2013 rider survey and the US Census for San Francisco. 48% of the 757 SFTQS participants who signed up and filled out the entry survey were female. The sample was skewed toward younger ages: 19% of participants were ages 18-24, 48% were ages 25-34, 19% ages 35-44, 9% ages 45-54, 4% ages 55-64 and the remaining 1% over the age of 64. Data from the official SFMTA rider survey shows 32% of passengers as being of age 45 or older, compared to 14% in this sample. This skew can in part be explained by the requirement to have a smartphone and in part by the recruitment methods. This is even more pronounced when compared to the US Census numbers, as it can be seen that the SFMTA's ridership itself is skewed toward younger riders. 75% of participants lived and worked or went to school in the city of San Francisco, whereas 25% commuted either into or out of the city for work or school. 79% of participants were either full-time or part-time employed, 19% were full-time students and 6% were part-time students (note that these groups overlap). Yearly household household income followed a somewhat bimodal distribution; the income categories were distributed as follows: 11% under \$20.000, 15% between \$20.000 and \$40.000, 20% between \$40.000 and \$60.000, 11% between \$60.000 and \$80.000, 10% between \$80.000 and \$100.000, 8% between \$100.000 and \$120.000 and 13% above \$120.000. Income was unknown for 12% of participants. Due to an unfortunate communication error during the design of the study, these income categories are different from those of the official SFMTA statistics, which makes a comparison difficult. However, it is clear that lower-income transit riders are underrepresented in our sample; household incomes under \$25.000 make up 38% of the SFMTA ridership. On the other hand, medium-income households (\$40.000 - \$60.000) might be overrepresented in this sample whereas the proportion of high-income households is approximately representative with respect to the SFMTA ridership. While there are certainly other factors that influence this distribution, it might again be a partial consequence of the requirement to have an Android smartphone. Lower income groups might be less likely to have a smartphone, and given the price differences between Android and iOS devices, iPhones tend to be more prevalent in higher-income market segments whereas Android phones cater more to lower and medium income groups. However, a comparison between the SFMTA ridership's income distribution and the US Census shows that in general, the SFMTA's ridership is skewed toward lower-income groups, and that the income distribution in the SFTQS sample might be closer to the general distribution in the population as reported by the US Census. Finally, table 2.4 shows the frequency of prior use (before the study) cross-tabulated against income.

Table 2.5 shows participants' auto access, driver's license status, and ownership of Clipper cards (contactless smart cards used for electronic transit fare payment) by age, gender and income. The columns showing auto access only cover the 89% of participants with a driver's license. 31% had a license but did not have a private car in their household, 29% had a personal car, and 21% shared a car with other household members. 41% of all participants had one car in their household and the remaining 19% had two or more cars in their

Age	SFTQS	**Age**	SFMTA	US Census
18-24	19%	18-24	20%	8%
25-34	48%	25-34	31%	25%
35-44	19%	35-44	17%	19%
45-54	9%	45-54	13%	16%
55-64	4%	55-64	12%	14%
Over 64	1%	Over 64	7%	18%
Income	SFTQS	**Income**	SFMTA	US Census
Under 20,000	11%	Under 15,000	24%	11%
20-40,000	15%	15-24,000	14%	10%
40-60,000	20%	25-34,000	9%	9%
60-80,000	11%	35-49,000	11%	12%
80-100,000	10%	50-99,000	22%	29%
100-120,000	8%	100-149,000	10%	15%
Over 120,000	13%	Over 150,000	10%	14%
Frequency of Muni use	SFTQS	**Frequency of Muni use**	SFMTA	
6-7 days/week	26%	5+ days/week	66%	
4-5 days/week	35%	3-4 days/week	14%	
2-3 days/week	24%	1-2 days/week	7%	
About 1 day/week	8%	1-3 days/month	3%	
1-3 days/month	5%	<1 day/month	6%	
<1 day/month	2%			

Table 2.3: Demographics of SFTQS participants compared to SFMTA rider survey and US Census.

Income (in \$10,000) → Frequency of use	<20	20-39	40-59	60-79	80-99	100-119	≥ 120	N/A
<1 day/month	0	4	3	0	7	1	0	2
1-3 days/month	2	5	9	4	4	6	4	8
1 day/week	7	6	10	12	6	10	10	6
2-3 days/week	31	19	32	23	24	10	34	28
4-5 days/week	29	42	66	29	31	27	36	40
6-7 days/week	24	50	53	22	16	18	21	19

Table 2.4: Frequency of prior use vs. household income (unit: number of participants).

household. Overall, however, 67% used a car regularly in San Francisco before the beginning of the study (this figure includes car sharing vehicles and carpooling), and a full 97% used Muni regularly before the study. 58% typically used Muni between 2 and 5 days per week, 27% more than 5 days per week and the remaining 15% less than 2 days per week. 44% owned a bike in usable condition, and 15% did not have a Clipper card prior to the study.

Table 2.6 shows the self-reported primary mode of travel by trip purpose for all par-

Age	PC	SC	NU	NC	NL	DL	NC	CC	CP
18-24	34	23	14	55	28	126	29	74	51
25-34	108	85	36	142	44	371	53	217	145
35-44	59	36	12	47	10	154	21	93	50
45-54	24	23	5	13	9	65	13	35	26
55-64	18	9	1	9	3	37	8	19	13
Over 64	2	1	0	3	3	6	3	2	4
Gender	PC	SC	NU	NC	NL	DL	NC	CC	CP
Female	118	87	37	129	59	371	79	219	132
Male	124	89	31	140	37	384	48	220	153
No answer	3	1	0	0	1	4	0	1	4
Household Income	PC	SC	NU	NC	NL	DL	NC	CC	CP
Less than $20,000	19	12	4	38	20	73	21	47	25
$20,000-$40,000	23	18	9	53	23	103	20	51	55
$40,000-$60,000	59	26	17	57	14	159	33	81	59
$60,000-$80,000	29	17	6	30	8	82	15	42	33
$80,000-$100,000	30	21	7	23	7	81	10	51	27
$100,000-$120,000	27	20	4	19	2	70	7	42	23
Greater than $120,000	28	38	10	22	7	98	2	69	34
Prefer not to answer	28	25	10	23	16	86	17	53	32
Other	1	0	0	0	0	1	1	0	0

PC	-	Access to personal vehicle
SC	-	Access to shared vehicle
NU	-	Vehicle in household, does not use
NC	-	No motor vehicle in household
NL	-	No driver's license
DL	-	Owns a driver's license
NC	-	No Clipper card
CC	-	Clipper card with cash value
CP	-	Clipper card with pass

Table 2.5: Auto access, driver's license status, and ownership of Clipper cards by study participants (unit: number of participants).

	Work	School	Shopping	Friends/Family	Leisure
Not applicable	87	570	26	30	17
Walk	72	48	143	42	86
Bicycle	47	10	11	13	33
Car or motorcycle	92	31	276	294	223
Muni	361	117	340	330	387
BART plus Muni	101	37	35	88	59
BART	39	8	19	36	18
Shuttle, taxi	46	31	4	14	28
Other	12	6	2	9	5

Table 2.6: Primary mode by trip purpose for all SFTQS participants (unit: number of participants)

.

	Work trips		Non-work trips	
	Overall	SFTQS	Overall	SFTQS
Transit	31%	58%	34%	51%
Car	31%	11%	43%	31%
Not applicable	25%	10%	0%	3%
Walk	7%	9%	13%	11%
Bike	4%	6%	3%	2%
Taxi	1%	5%	5%	2%
Other	2%	1%	1%	1%

Table 2.7: Primary mode for work and non-work trips: comparison of overall San Francisco population with SFTQS population.

ticipants. As can be seen, between 53% and 60% of participants who commuted to work or school identified Muni as their primary mode of transportation for those trips. Table 2.7 compares the primary mode of transportation for work and non-work trips of SFTQS participants with the 2011 San Francisco Mode Share survey; as can be seen, transit was identified as a primary mode of travel by a larger share of SFTQS participants than in the general population.

Response rates

Recruitment proved to be relatively difficult. The overall response rates are unknown since several of the recruitment channels utilized were publicly accessible (social media and press releases), and it is unknown how many of the companies and neighborhood associations that were contacted forwarded the information about the study to their employees or members. Table 2.8 shows the reported recruitment channels of participants who responded to the pertinent question in the exit survey. It should be noted that participants who did not

	Contacted	Responded				
Entry/Exit survey →			No	Yes	No	Yes
Daily mobile surveys →			No	No	Yes	Yes
Recruitment channels						
Unknown		169	97	0	72	0
Postcard in the mail	22000	39	1	6	2	30
Email from UCSF	20000	196	4	16	4	172
In-person (fliers)	500	33	0	5	1	27
Neighborhood assoc.		9	0	1	1	7
Email from SFMTA		41	1	0	1	39
Email from UC Berkeley		63	0	6	2	55
Email within company		56	0	6	2	48
Another email list		21	0	2	0	19
News, social media		74	1	1	5	67
Word-of-mouth		112	0	9	3	100
Other		43	0	3	0	40

Table 2.8: Response rates for different recruitment channels (unit: number of participants).

take the exit survey were not asked that question and therefore are shown in the "unknown channel" row. The table further breaks down the respondents by how much of the study they participated in, i.e., whether they filled out the entry and exit survey and whether they filled out at least five daily mobile surveys.

Only in the case of the postcards, the UCSF email list and the fliers was it known how many people were contacted in the first place. The estimated AAPOR response rate 1 is 0.1% for the postcards, 0.9% for the UCSF email list and 5.4% for the in-person recruiting (fliers). The estimated AAPOR response rate 2 is 0.2% for the postcards, 1.0% for the UCSF email list and 6.6% for the in-person recruiting (fliers). It should be noted that these are lower bounds since a participant might also have been recruited via word of mouth, by somebody else who was initially contacted via one of these channels.

In total, 1401 people responded to the sign-up survey. 856 had an Android smartphone and supplied a valid email address. 94 had an Android smartphone but did not supply a valid email address for further contact. 428 respondents were not eligible because they owned an iPhone, and a further 23 respondents were not eligible because they owned a different type of phone.

All emails with the download link for the app were distributed on October 27. Figure 2.2 shows how many participants filled out their first trip-related daily mobile survey on a given day. Approximately three groups (peaks) can be distinguished: The leftmost are participants who downloaded the app and began using it immediately. The second group began responding around day 5 (Friday) and the third group around days 8-10. This is when the majority of transit passes arrived at their destinations and several additional emails to participants were sent, so this wave likely represents the effectiveness of receiving a physical incentive in motivating participation.

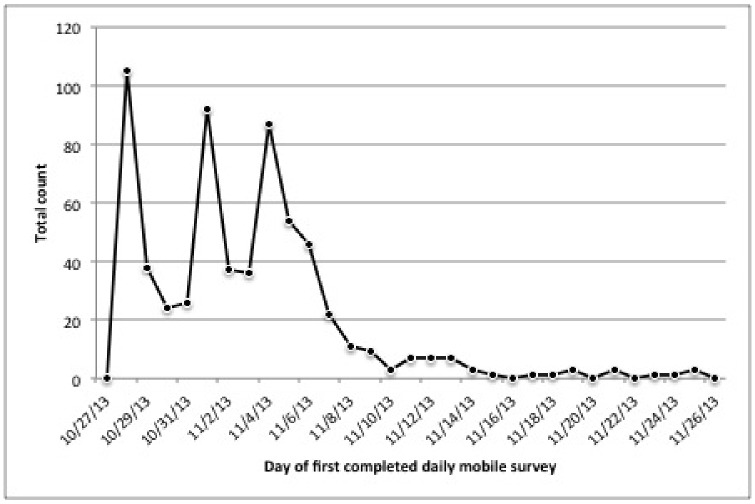


Figure 2.2: Number of first completed travel-related daily mobile surveys by date. Example: On Oct. 28, 2013, 105 participants filled out their first travel-related daily mobile survey.

Figure 2.3 shows distributions of the numbers of completed daily mobile surveys plotted against several characteristics of the respondents. In the left plot, it can be seen that the earlier a participant started participating, the more daily mobile surveys that person was likely to fill out. In the right plot, it can be seen that participants who used Muni frequently before the study (on at least two days per week) submitted, on average, more completed daily mobile surveys than participants who used Muni infrequently before the study (less than two days per week). The middle plot shows that the median number of daily mobile surveys completed by the group that indicated they had taken part in the study primarily for the free Muni pass was approximately the same as the median for groups that had joined for other reasons (excluding wanting to try Muni). The group that said they had wanted to try Muni, which included only 39 participants, had a lower median number of completed daily mobile surveys.

In figure 2.4a, one can see an inverted cumulative density function of the number of days for which location tracking data were collected. 100% corresponds to 838 participants (i.e. all signed-up participants minus withdrawals). 83% of respondents collected 10 days or more of tracking data, 77% 20 days or more and 63% 30 days or more. All 41 days of the study were only covered by about 38% of participants.

Figure 2.4b shows average response times between the time when participants received the notification and the time they started filling out the daily mobile surveys. It should be noted that participants could take the daily mobile surveys at any time they wanted by opening the app and manually starting the survey. The notification only served as a reminder. A large number of participants responded to at least one survey before the alert time, which strongly affected the averages. Therefore, two lines are shown: One excluding responses before the alert time ("average positive response time") and one including them. The former shows that approximately 50% of users had an average response time to the alert of an hour or less, and approximately 85% responded within two hours.

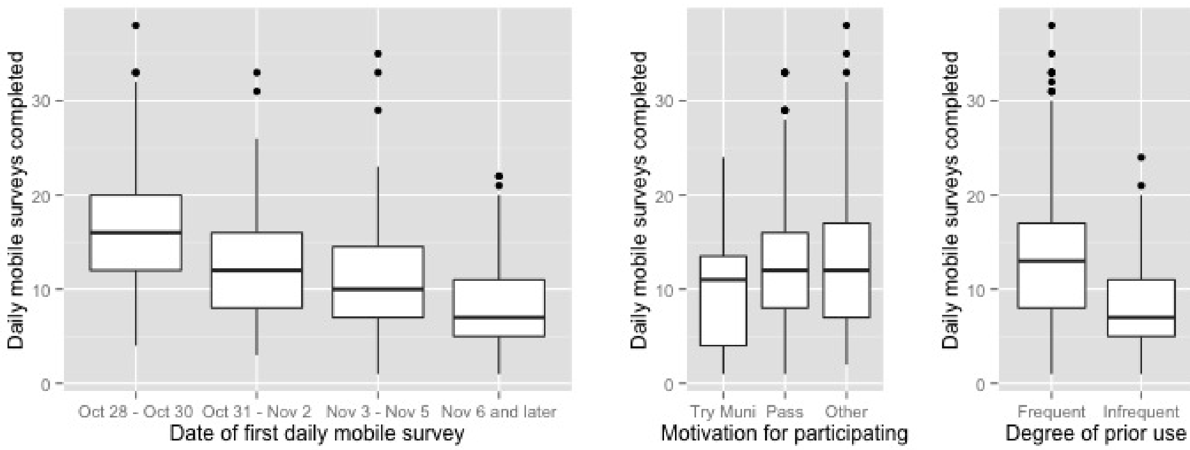
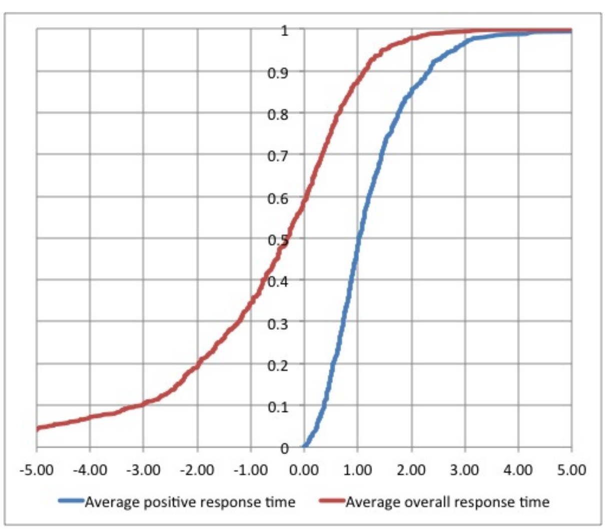
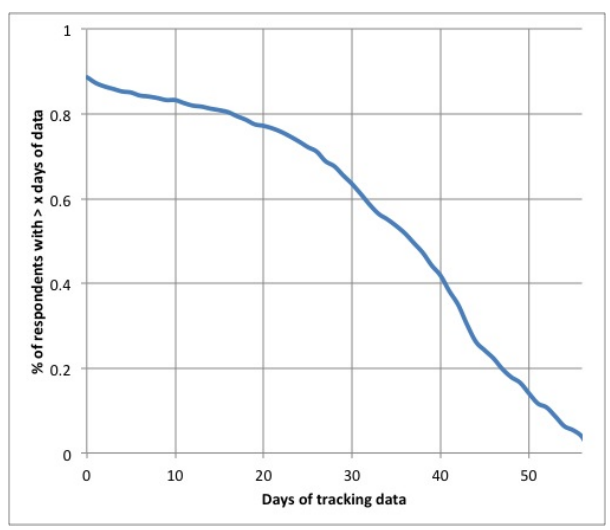


Figure 2.3: Distributions of number of completed daily mobile surveys. The minimum required contribution was 5.



(a) Response times after timed survey prompts, including (left) and excluding (right) responses recorded before the prompt.



(b) Number of days with location tracking data. Example: 77% of participants collected 20 days or less of tracking data.

Figure 2.4: Participant compliance with daily mobile survey prompts and location tracking.

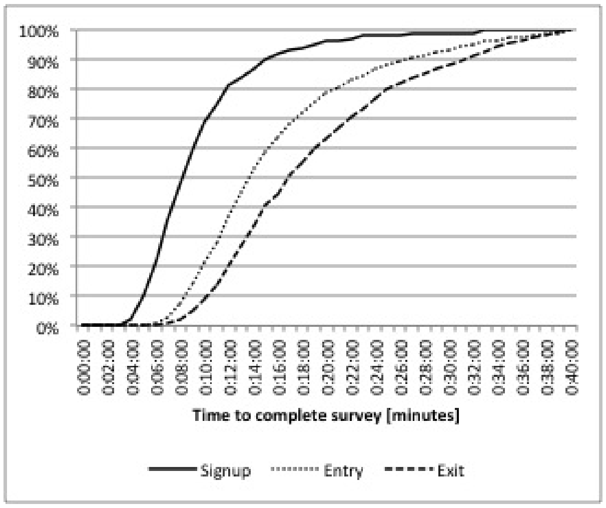
	Coefficient	Std. Error	$p(>\lvert z\rvert)$
Intercept	-0.19	0.44	0.67
Female	0.17	0.16	0.30
Age	-0.01	0.01	0.48
Bike ownership	0.15	0.16	0.36
Income	0.04	0.02	0.07
Unknown income	0.03	0.27	0.90
Frequent car user	-0.70	0.17	0.00
Frequent Muni user	0.70	0.22	0.00
Cohort 2	-1.43	0.19	0.00
Non-Commuter	0.27	0.18	0.13
Knowledge of Muni	0.10	0.10	0.30

Table 2.9: Binary logit model of study participation. Positive coefficients mean a participant was more likely to complete all study-related activities.

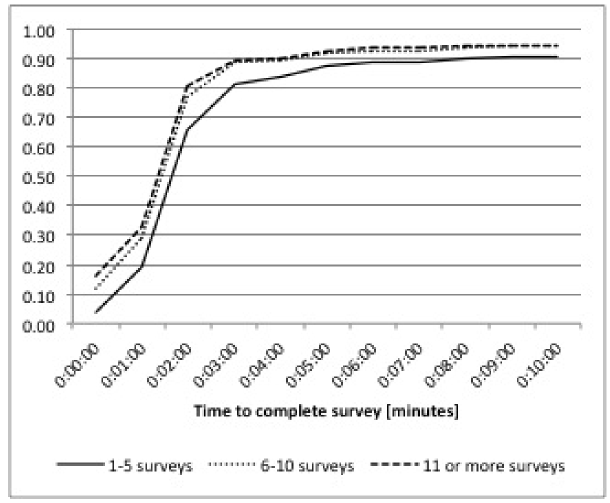
A binary logit model was estimated to determine the factors influencing whether or not a participant would carry out all study-related activities, i.e. fill out the entry and exit surveys and at least five travel-related daily mobile surveys. The results are in table 2.9. With the exception of age and income, all variables were binary. Age was coded as the mean of the participant's age bracket in years (e.g., the mean of age bracket 35-44 is 39), and income was coded as the mean of the participant's income bracket in $10,000. Positive coefficients mean it is more likely for the person to complete all study activities. The single largest influence was whether or not the participant was in cohort 1 or 2. The other two highly significant predictors are whether the person used a car or Muni regularly (defined as 2 days per week or more) in San Francisco before the study. Income also has a moderate effect, as does whether or not somebody commutes into or out of the city for work or school.

Respondent burden

Figure 2.5 shows a set of cumulative density functions of the times to complete the various surveys. The mean, median (50th percentile) and 95th percentile times are shown in table 2.10. The mean combined total burden for the online surveys was 45 min. Response burden for the daily mobile surveys is slightly more difficult to measure as the number of responses per participant varied. Figure 2.5b shows the time taken to respond to the mobile surveys as a function of how many surveys an individual has filled out. A slight learning curve is apparent; the cumulative density function for the first through the fifth survey is a little lower than for the remainder. The minimal contribution was 5 daily mobile surveys, which would amount to a mean total burden of 5 min. In addition, participants were expected to spend an estimated 15 min. across the entire study to read instructions delivered by email and to download and install the app.



(a) Online Surveys.



(b) Daily Mobile Surveys.

Figure 2.5: Time to complete surveys.

	Mean	Median	95th percentile
Signup	10 min.	8 min.	19 min.
Entry	16 min.	14 min.	32 min.
Exit	19 min.	17 min.	34 min.
Daily Mobile	1 min.	1 min.	3 min.

Table 2.10: Mean, median and 95th percentile of the time spent responding to the surveys, rounded to the nearest minute.

Insights from feedback

The 686 participants who took the exit survey were asked for feedback about the study organization. 48% stated that they had joined the study for the free Muni pass, whereas 23% wanted to help improve Muni and 15% said they were motivated because they were given the chance to provide feedback that would be heard. During the design phase of the study, it was feared that the incentive might have the effect of skewing the sample toward lower-income groups. However, based on the discussion of the demographics of the study population, this effect does not appear to have been very large.

Only 10% of participants said the daily mobile surveys were too long, whereas 25% felt they were too short. The remainder thought they were just right. 36% would have preferred to fill out surveys more frequently than once per day (e.g. one per trip), 51% preferred the once-per-day format, and 13% expressed no opinion on the survey frequency. 75% of respondents found the daily mobile survey prompts “very useful” or “somewhat useful”, while 11% found them annoying. 10% did not find them useful or ignored them and 4% experienced technical difficulties with them. On the technical side, 23% of respondents said the battery life of their phone had noticeably decreased.

2.6 Conclusions

This chapter presented the background, design and administration of the SFTQS. This large-scale, smartphone-based transportation study combined the continuous collection of location and mode use data with time-prompted surveys and AVL data to create a rich data set. The focus was on the dynamics of satisfaction and travel-related emotions, which are two areas that have not received much empirical attention in transportation despite their importance being recognized in the literature. One reason for the lack of research on these subjects so far has been the unavailability of a suitable data set, which the SFTQS sought to remedy. Beyond that, this study also provided insights into various aspects of user response behavior in a smartphone-based transportation survey. It was found that more frequent and possibly longer surveys throughout the day or a division into pre-trip and post-trip surveys in future studies would largely be accepted by participants. Given that the incentive offered to participants was a free transit pass, the vast majority of people recruited into the study were regular transit riders. If a non-transit user sample is desired, a different incentive structure would be required. Also, a slight bias was introduced into the sample due to the limitation that users had to have an Android smartphone with a data plan. However, a comparison between the demographics reported by the SFMTA and the study group shows that the bias is relatively small. To obtain a less biased sample, apps for other operating systems would need to be added, as well as an option for users without smartphones or data plans.

The fact that all surveys were online and on the app and all instructions were issued via email brought its own set of challenges. In fact, the most important limitation to scalability was found not to be the technical infrastructure, but rather the researchers' capacity for "customer support". While it is unknown how many emails from participants were received by the researchers in total, the number significantly exceeded prior expectations. There were participants who experienced problems related to the app, but many also wrote to clarify instructions and to receive technical support with installing the app on their phones. The latter is not related to the functioning of the app itself but rather to the operating system of the phone and the app store. It is estimated that in the days following a general email to all 856 participants with new instructions, the typical number of emails received by the researchers was between 50 and 150, depending on the complexity of tasks the participants were asked to carry out. The researchers actively monitored the number of daily mobile surveys completed by the participants and sent individual emails to participants who were slow to complete the required five surveys. On the other hand, incoming location data from the phones was not monitored, which might explain the drop-off seen in figure 2.4b.

Chapter 3

Quantifying transit travel experiences

3.1 Introduction and motivation

Tracking transit passengers as they travel through a transit network can generate data that are useful for numerous applications. In particular, one of the main inputs into long-range planning is the origin-destination (OD) matrix, which is often created using combinations of survey data and automatically collected data. Furthermore, researchers and practitioners have become interested in calculating passenger travel times from automatically collected data with the goal of deriving user-based reliability and performance metrics to complement supply-side metrics that are currently in use. This ties in with research efforts in travel demand modeling, where it has been recognized that travel time reliability, and in particular an individual's *personal experience* with unreliability in the past, plays a major role in travelers' decisions (Fosgerau and Karlström, 2010; Hensher, Stopher, and Bullock, 2003b; Li, Hensher, and Rose, 2010; Benezech and Coulombel, 2013). So far, the data that have been available for these applications were from automatic data collection systems operated by the agency, most notably automatic vehicle location (AVL), fare collection (AFC) and passenger count (APC) data.

The collection of high-resolution, individual travel time data can be challenging. Trips on transit tend to be complex, involving access and egress, wait times and transfers in addition to in-vehicle segments. Passenger-centric measures of travel time distribution are a combination of the travel times experienced on these segments; the overall distribution resulting from the convolution of these individual distributions can be quite complex (Bates et al., 2001b). For the aforementioned applications, being able to capture all these segments of the passenger's trip would be highly beneficial: it would afford a better picture of the users' typical *overall* travel experience and of how the in-vehicle and out-of vehicle travel times compare with the automobile. In the long run, sampling and quantifying the overall travel times of transit users on a large scale can lead to the development of new reliability metrics that better integrate the passengers' perspective.

In fully gated systems, where the fare card needs to be tapped both upon entry and exit from the system, the time spent in the system can be derived, but typically there is limited to no information on out-of-vehicle trip segments. In systems that are not fully gated and do not require passengers to tap their fare cards upon exiting, information on alighting and transfer stops is missing. In open systems without fare gates, such as most bus systems, even the information on boarding stops may not be exact. To further complicate matters, in some systems pass holders are not required to tap their fare card. The incompleteness of stop information for trips in such systems makes determining transit OD matrices and deriving travel time distributions challenging, as one has to rely on inferences and limit oneself to the trip components that are observed.

To derive either passenger-focused reliability metrics or transit OD matrices, passenger trips have to be assigned to transit routes, stops and, if possible, vehicle runs. There is a spectrum of methods published in the literature that make use of AVL, AFC and APC data, in various combinations. As is summarized by Zhao, Rahbee, and Wilson (2007), who uses AFC data with entry tags only, a common assumption that is made when only entry data are available is that the stop where passengers board on one trip is the stop where they alighted on the previous trip. A selection of further related work with AFC data from gated rail systems is by Cui (2006), Chan (2007) and Rahbee (2008). Chapleau, Trépanier, and Chu (2008) and Chu and Chapleau (2010) used AFC data augmented with a geographical information system to derive transit origin-destination matrices. Yuan et al. (2013) works on a similar problem, but from the perspective of tracking individual mobility behavior through smart card transactions. Using smart card data from a fully gated system, (Sun and Xu, 2012) develop a model to infer the various travel time components on an underground network based on travel time distributions. Several authors (including Farzin, 2008; Nassir et al., 2011; Wang, Attanucci, and Wilson, 2011; Munizaga and Palma, 2012; Gordon et al., 2013) have also focused on connecting passenger trips from bus AFC data to vehicle locations observed via AVL data in an effort to better infer boarding locations and times. These were cases where the fare card reader was on the transit vehicle and did not directly record the boarding stop. Frumin and Zhao (2012) used AFC and AVL data in a gated system to infer rail platform wait times and Seaborn, Attanucci, and Wilson (2009) examined distributions of transfer times between a fully gated rail system and a bus system to distinguish pure transfer times from activities carried out at the transfer location.

While these aforementioned contributions have been very valuable, they have in common that due to the coarse resolution of the data, researchers could not obtain exact measurements of every travel time component, including out-of-vehicle travel times, especially in the case of bus travel. Work that attempted to disaggregate total travel time into its individual components did so primarily based on distributions of total travel times. However, thanks to smartphones and other location-enabled devices, it is becoming increasingly feasible to collect individual-level location data over long periods of time and with low respondent burden; with the help of these data, the shortcomings noted above can in many cases be overcome, which is most valuable in ungated systems. Smartphone location data allow a high resolution view of individual trips as long as the traveler remains above ground, including out-of-vehicle segments and exact information on stops. As location-aware devices become ubiquitous, and given the scalability of such systems, planners can find themselves in possession of very large amounts of location data from which user-based performance metrics, personal travel experiences for demand modeling and transit OD matrices can be generated. In addition, such data allow the observation of the *true* origin and destination of a trip.

There has been previous work that utilized passenger smartphone location data, but so far it has mainly been focused on determining the travel mode from data collected through location and other sensors (e.g., accelerometer, microphone). This has been performed by map-matching the location points in a GIS system (Chung and Shalaby, 2005; Gong et al., 2012; Jariyasunant, 2012), by extracting features from location and accelerometer data related to velocity, acceleration or distance traveled and using those as inputs for mode classification algorithms (Zheng et al., 2008; Gonzalez et al., 2008; Parlak, Jariyasunant, and Sengupta, 2012) or a combination of the two (Biagioni et al., 2009; Thiagarajan et al., 2010). Stenneth et al. (2011) included AVL data and extracted the proximity of the phone to a transit vehicle and to transit stops as a classification feature (on a point-by-point basis). Other authors have worked on crowd-sourcing transit arrival times where no AVL data are available. These applications rely on real-time detection of when a smartphone owner is on board transit. Notably, this has been done by Zhou, Zheng, and Li (2012) based on microphone, accelerometer and coarse location, and by Kostakos, Camacho, and Mantero (2013) based on Bluetooth sensors. Lastly, Barbeau, Georggi, and Winters (2010) developed a personal transit travel assistant for cognitively disabled riders and note that AVL data were used in their system, but they do not specify the exact role of those data.

In summary, previous research on has relied on data sets with limited resolution that do not capture all components, especially in ungated systems. The advent of smartphones presents a new opportunity for collecting higher-resolution data and tracking passengers' movements throughout the network. However, so far, research utilizing smartphone data and AVL data has only been focused on mode detection. To the best of our knowledge, no study has combined high-resolution smartphone data to capture the demand side with AVL data to capture the supply side. That is the purpose of the methodology presented in this chapter. Specifically, we present an automated system that is able to extract the elemental "building blocks" of passengers' trips with minimal inferences, i.e., that is able to identify the various parts of the trip that took place on transit as well as transfers, associated travel times and scheduled travel times on a per-segment basis in a careful manner. The output serves as a foundation for the applications described at the beginning of this section to obtain a high-level view of system performance.

The methodology presented here takes as input smartphone location data, automatic vehicle location (AVL) data and static timetable data from the General Transit Feed Specification (GTFS). In developing the system, there were several main objectives. The first was to systematically consider the components of transit travel time from a passenger's perspective, to describe them, and to develop a framework that shows how they can be measured. The second objective was to develop an automated, robust system that (a) can handle low-frequency location data, collected approximately every 30 seconds but sometimes at larger intervals, and (b) will work in dense transit networks where many routes overlap and many vehicles may be in the vicinity of a traveler at any given time. The third objective was to develop a methodology for tracking underground travel in a metro network and combining it with statistics on transit travel above ground. Furthermore, we present an approach for distinguishing wait times at the origin stop and transfer times from legitimate activities. To the best of our knowledge, this is the first time that these issues have been investigated.

The travel behavior of transit passengers is complex, and without simultaneously considering the demand and supply side, much of that complexity would remain unobserved. In this chapter, it is assumed that no additional information is available beyond the location

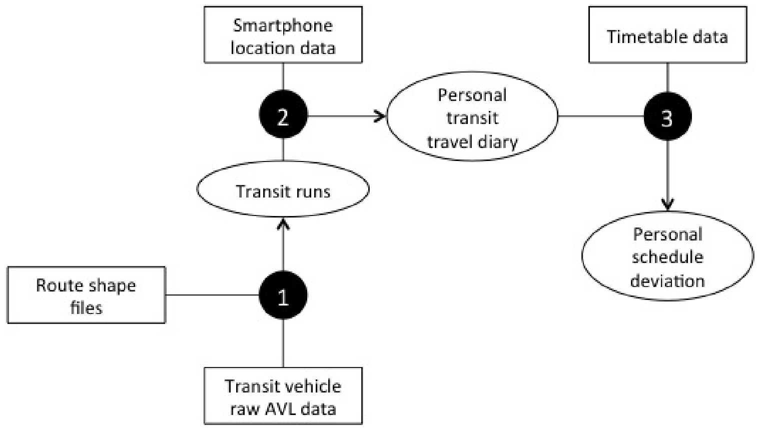


Figure 3.1: The general matching problem.

data, but in future applications, the location readings can be augmented with additional data such as phone-based surveys or user activities on the phone to provide even richer data sets deepening the researcher's understanding of the complexities of human travel behavior.

In section 3.2, we first describe the problem in more detail and provide some definitions that will be used in the remaining sections. Section 3.2 also gives a high-level view of the procedure. Section 3.3 then goes into detail and presents a step-by-step description of the procedure to match phone location points to transit runs. In section 3.4, the comparison of the observed travel times with the scheduled travel times is described. Finally, section 3.5 presents the system validation and first results of a large-scale deployment as part of a travel study in San Francisco.

3.2 Problem description and definitions

For every passenger for whom location data are available, we want to derive that person's *personal transit travel diary*, i.e., exactly when and from where to where the person traveled on transit, on board which vehicle, and exactly how much time was spent on board and waiting at transfer stops or the origin stop. This is the first set of elemental pieces for developing user-based reliability metrics and travel time distributions, and this is also the information used to generate transit OD matrices. The second set of elemental building blocks is extracted by comparing those times to the timetable to obtain deviation measures. For the latter, we are not specifically interested in whether a given transit vehicle was on time; we are interested in how the overall service received by a passenger differed from the service that passenger could expect based on the timetable. Overall, we are dealing with a matching problem as is illustrated in figure 3.1.

In figure 3.1, square boxes represent input data to the system and ellipses represent outputs. The arrows point in the direction in which the processing is made, i.e., from input to output. We begin with two sets of raw data, the phone location data L_p and the AVL data L_v. Each data point includes a time stamp, latitude, longitude and a phone identifier (in L_p) or a vehicle identifier (in L_v). In our case, the test application of our system was in the city of San Francisco; the AVL data were obtained via the public NextBus interface of the San Francisco Municipal Transportation Agency (SFMTA) using the vehicle location command (NextBus, 2013) to obtain raw vehicle location data. The timetable and network information were obtained from the General Transit Feed Specification (GTFS) file published by the agency. The phone location data were obtained from a survey app that collected location points roughly every 30 seconds. The processing server is built on a cloud computing platform and has one application programming interface (API) which receives the series of location points with time stamps. The raw AVL data and the network information are stored in two main geospatial databases using PostgreSQL/PostGIS.

In step 1, which is not described in detail in the remainder of this chapter but included here for completeness, the raw AVL data must be matched to route shape files to generate individual runs. A run is defined as a single transit vehicle in revenue service on a fixed route traveling one way from one terminal to another terminal. The vehicle position and stop locations are measured as the distance from the departure terminal, called the *milepost*, and each run is described by a trajectory that includes a set of time-milepost tuples. They are either derived from the original AVL data (via projection onto the route polygon), or they are the mileposts representing the stops along the route, with times inferred via linear interpolation. In the current application, the matching of raw AVL data was performed as a separate step, using a simple point-to-curve algorithm as described by Quddus, Ochieng, and Noland (2007). Oftentimes, the fitted milepost data are available as a byproduct of commercial real-time arrival prediction systems.

In step 2, the phone location data are matched to transit runs. For any given user trip, it must be determined whether or not transit was used, and if so, which runs the user was on, what the boarding, alighting and transfer stops and corresponding times were. This happens in several steps. First, the user's phone traces are matched to runs and the boarding stop and alighting stop are inferred. This is done in two separate processes: One that finds nearby runs for every phone data point ("Initial vehicle matching"), and one that determines whether any segments of the passenger's trip were underground. The initial matching process on the above-ground (surface) data yields so-called candidate runs. Through further processing, a list of surface trip segments with the respective access, egress and transfer stops is determined. In parallel, a list of underground segments is compiled, and at the end, the two are merged. The time between the boarding and alighting from a single transit vehicle constitutes a *transit segment*. If the traveler cannot be mapped to any surface or underground trips, this step returns an empty result. Next, all segments that are associated with the same trip (i.e., a user traveling from one activity location to another) are assembled to form an uninterrupted, alternating sequence of transit segments and transfers, plus an origin wait at the beginning of the trip. The trip ends if the user stops to carry out an activity somewhere. The set of an individual's trips on transit constitutes the *personal transit travel diary*.

Given only phone location data without additional information, the surface trip matching is based purely on the distance between the phone and a vehicle. This procedure bears the risk of identifying trips with other modes, such as bike or car, as transit trips if the person traveled alongside a transit vehicle for a prolonged time, which might for instance be the case in dense traffic. Please see section 3.3 for a further discussion of this issue.

In step 3, we then compare every trip made on transit to the timetable to derive the *personal schedule deviation metrics* by comparing the observed wait times and travel times to what the subject could have expected from the timetable. This step takes a passenger-centric rather than an operations-centric view of travel time variability, and since no other input is collected from the traveler, it relies on some assumptions about the desired departure time. As previously mentioned, the personal transit travel diary and the personal schedule deviation metrics together constitute the elements from which, at a later stage, user-based performance and reliability metrics, OD matrices, travel time distributions and other measures of interest can be derived.

3.3 Detailed description of the phone-vehicle matching procedure

In the transit network for which this system was developed, the bus system operates entirely above ground, but the rail system operates partially above ground and partially underground. The location of the phones was sampled every 30 seconds, though "holes" in the order of one to two minutes were not uncommon. This complicated the determination of access, egress and transfer stops. In the following discussion, we assume that processing step 1 has already been performed and that the individual vehicle runs have been calculated and stored in a database. In the following subsections, we first describe the matching of phone points to vehicle location points and the inference of the access and egress stop. The inference of transfer stops is more complex and breaks down into several subcases, which are presented in detail. Following that, we describe the underground matching problem and the derivation of the final transit travel diary.

Determining above-ground trips on transit in dense networks with overlapping routes

We first take a user-centric view: given the starting and end point of a trip, and the respective series of location points $L_p = \{(lat_0, lon_0), (lat_1, lon_1), ...\}$, we create a three-dimensional search box defined by $\min(lat_n - \delta)$, $\max(lat_n + \delta)$, $\min(lon_n - \delta)$, $\max(lon_n + \delta)$, t_0, t_{max} and query the database for all vehicle trajectories (runs) that traverse that box (δ is a buffer distance that is added to the search box). Every run has a unique identifier, which is different from the physical vehicle ID. Using the latter may yield false positives if the subject remains at the same location but is passed by the same vehicle going in two opposite directions. We then use Dynamic Time Warping (Müller, 2007) to calculate the similarity between the phone trajectory and the vehicle trajectories, using the phone location time stamps for the time steps. As a cost function we use the absolute distance between the points (the time distance being zero); this has the effect of matching a linearly interpolated vehicle trajectory location to every phone location. We then define a distance threshold and filter the vehicle trajectories to obtain a set of candidate runs for every phone location point.

Next, the transit runs where the user was most likely on board need to be distilled from the candidate list. This can be challenging in dense networks or corridors where many routes overlap: A number of vehicles may be detected in the vicinity of a phone, and if

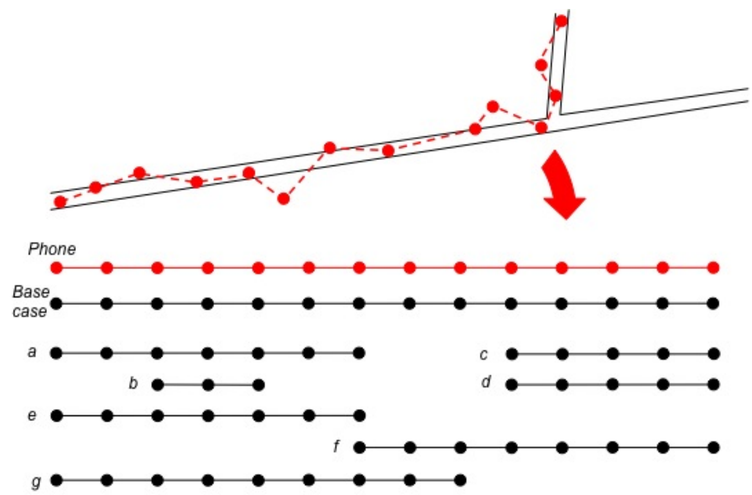


Figure 3.2: Possible relationships between candidate runs (a-g).

several vehicles travel in close proximity, the phone might be mapped to multiple vehicles simultaneously.

Due to positioning errors, the shortest distance between the phone and a vehicle may not necessarily indicate the correct match. Instead, we switch to a run-centric perspective: Given the set of candidate vehicles for every point, we group the data by run ID and create a candidate run list. In practice, this is an associative array with the run ID as key and a second associative array of characteristics as value. The characteristics include the information that was extracted from the candidate vehicle list; most importantly, a list of phone location points where the phone and the vehicle were matched. This is equivalent to projecting the subsets of runs that are close to the phone location onto the one-dimensional trajectory of the phone, as is illustrated in figure 3.2. The red line is the trajectory of the phone, with the dots symbolizing location readings. Below, in black, are the sets of points where the phone location matched a transit vehicle location. As an example, in case (b), the phone was observed to be near this particular transit vehicle in the third, fourth and fifth location point recorded by the phone, but in no others.

The base case, where the phone matched the same candidate run over the entire trajectory, is illustrated at the top. If the phone is matched to different candidate runs, the following relationships are possible:

1. They can be disjoint, as (a) and (c).

2. They can be identical, as (c) and (d).

3. One can be a subset of the other, as (b) is of (a).

4. They can share one location point, as (e) and (f).

5. They can share multiple location points, as (g) and (f).

First, we eliminate subsets (case 3). If two candidate runs are identically matched (case 2), the one with the higher average distance between the phone and the vehicle is eliminated.

We have now retained only runs where the points are either disjoint or overlapping as in 4 and 5. However, this does not allow us to know whether two runs might overlap in reality, so we query the database for every combination of candidate runs in categories 4 and 5 and request the number of possible transfer stops within the overlapping segment. If there is only one possible transfer stop, that pair is reclassified as being disjoint. With only the "truly" overlapping candidate runs remaining in categories 4 and 5, we can now think of the traveler's trajectory as an undirected graph, represented by an adjacency matrix, with an arbitrary set of unconnected trees that each represent a group of overlapping candidate runs.

The groups must now be processed. We traverse the graph in a breadth-first search to extract overlapping runs and sort them by the time stamp of the first location point they matched with the phone, then recursively do the following:

1. Choose r_0 and r_N, the first and last run in the group.
2. Check whether they are disjoint.
3. If no, discard every run in between.
4. If yes, extract the "inner sequence", i.e. the sequence of runs in between r_0 and r_N, and repeat.

In practice, the undirected trees are represented as a list of lists; after eliminating subsets as described above, we can flatten it to a single candidate run list, where every run overlaps with at most one other run at the beginning and the end.

Inferring access and egress stops

This problem may be relatively simple if high-frequency phone location data (e.g. 1 Hz) are available, but if the phone's location is reported at a lower frequency, it becomes more challenging because the time where the passenger boarded or alighted from the vehicle might fall within the gap between two location points, in which case the researcher has to make an informed guess.

To determine the access stop, two data points are useful: l_1, the phone's location at the first point where it was matched with the transit run r, and the last location before that, l_0. The approach chosen was to query the nearest transit stop served by r to each of those two locations, s_1 and s_0; these are usually identical, but if they are not, one of the two is chosen based on two threshold parameters: The maximum difference between time stamps and the maximum time window around the time at which s_0 was served by the run and the time the corresponding phone location was reported, $(t_{s_0} - \delta_1 < t_0 < t_{s_0} + \delta_2)$. The first threshold exists because if the gap is too large, too much unobserved activity may have occurred within it, decreasing the probability that the traveler only waited at that stop to board the run. The second threshold is to guard against inferences based on outliers. If t_0 is within both these thresholds, s_0 is chosen, otherwise s_1. And example is shown in figure 3.3: The phone only generated two location points around the time of boarding, one that was not matched with a bus at l_0 and one that was matched with a bus at l_1. The former is closest to stop s_0, the latter is closest to stop s_1. The inference of the egress stop is identical,

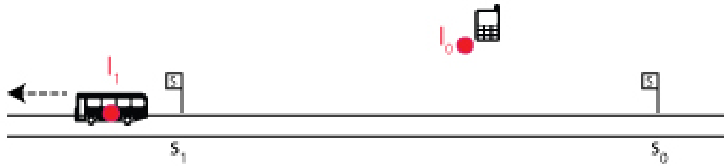


Figure 3.3: The stop matching problem.

except that it is based on the last phone location point that was still matched to the run and the first point that was not matched anymore.

Inferring transfer stops

When trying to infer the transfer stop, we again have the problem that the actual transfer may not be directly observable in the data. We are working with two pieces of information:

- How many possible transfer stops there are between the two runs that the phone was matched to.
- Whether there are any phone location points that were matched to both runs simultaneously.

We first break down the problem according to the latter, then consider the former. We shall call the two runs run A (the first run) and run B (the second run). Then, given that we have identified the stop where somebody alighted from run A and boarded run B, we must determine whether the time spent in between the two runs was a transfer - i.e., time that the user spent waiting for run B - or whether the user stopped at that location to carry out an activity. Only in the former case would the time between the two runs be counted as out-of-vehicle travel time. The following two subsections first discuss the inference of the transfer stop. To aid the discussion, the three discussed cases are illustrated in figure 3.4.

No points matching both runs (disjoint)

We distinguish two possible subcases: (1) If there are location points between the two runs, the transfer was actually observed, and access/egress stops for run A and B can be determined as described previously. This corresponds to case 1 in figure 3.4. (2) If there are no location points in between, the actual transfer location was not captured by the phone tracking data, as illustrated by case 2 in figure 3.4. To infer it, we query the database for all possible transfer stops between the phone's last sighting on run A and its first sighting on run B. We then eliminate stop combinations that were first served by run B and then by run A. If there is only one possible transfer stop remaining after that, it is chosen. If there are multiple possible stops, the first stop served by run B after run A is selected.

One or more points matched both runs

This is case 3 as illustrated in figure 3.4. In this case, the database is queried for all possible transfer stops within the segment of the passenger's trip where the phone points matched

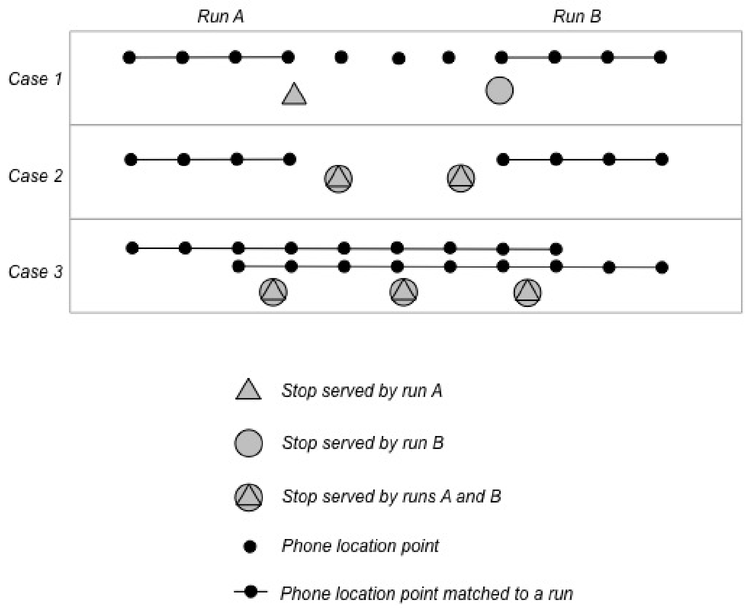


Figure 3.4: Scenarios for transfer stop detection.

both runs, and again, only the stops are retained that were served by run B after run A. If there is only one possible transfer stop, it is chosen. Otherwise, the first stop is chosen where the phone was closer to the vehicle on run B than to the vehicle on run A.

Distinguishing transfers from activities

To determine whether the user stopped at the transfer location for a legitimate activity, we check two conditions. First, a geofence is created that includes the two transfer stops, and it is checked whether the passenger left the perimeter of the transfer location for a prolonged period of time. Second, we query the database for all transit runs that served the passenger's observed transfer stop and destination and determine how many runs passed while the passenger was waiting at the stop without being boarded by the passenger. If either the passenger spent too much time outside of the transfer stop perimeter or more than a certain number of runs passed that were not boarded by the passenger, the time between the two runs is classified as an activity rather than a transfer.

The underground matching problem

Extracting underground travel segments

While AVL data are generally available for underground segments, phone location data are not. However, phones may still generate occasional location readings, either spuriously or during a stop at an intermediate metro station where a Wi-Fi signal can be picked up. The procedure adopted is to create geofences around every subway station as well as tunnel portals in the system. The phone location data are scanned and every point that falls within a geofence is labeled, producing the sequence of geofences traversed. Scanning through this list, for every pair of geofences, the frequency of data collected between them is calculated and compared to an estimated threshold parameter. This step requires knowledge of typical data collection rates when the phone is above ground. If the observed collection frequency between the geofences is above that threshold, the segment is classified as a surface segment, otherwise it is an underground segment. This yields a list of in-tunnel segments. Sequential in-tunnel segments are combined (e.g., $(A \rightarrow B, B \rightarrow C = A \rightarrow C)$, again using a frequency criterion to detect cases where a person exited at a metro station and re-entered it later.

The location points matched to surface runs and those matched to underground segments are not yet guaranteed to be non-overlapping. As a next step, the relationship between every surface segment and every underground segment is tested, and the data are cleaned as follows:

- If the surface segment and the underground segment intersect on one or more points, the surface segment is truncated to match the beginning or end of the underground segment.
- If the tunnel segment is a subset of the surface segment:
  - If the surface segment was on a bus or rail route that never goes underground, split it to exclude the in-tunnel section. It is assumed that the passenger changed from traveling with an above-ground mode of travel to an underground mode and then back to the original one.
  - If the surface segment was on a metro route that runs both above ground and underground, delete the separate underground segment, as it is already included in the surface segment data.

Finally, the routing between the entry and exit point of every underground segment is inferred with a shortest-path algorithm. This assumes there is only one reasonable routing between two stations, as is the case for the underground network in San Francisco. However, this approach would have limitations in denser networks with multiple feasible paths since the assignment does not take into account any additional information on the route chosen by the traveler. In future research, a probabilistic assignment model based on generalized travel time might be required.

Merging underground and surface segments

In the SFMTA's network, rail routes run underground in downtown but emerge to the surface and run at grade in outer districts. Those tunnel portals were included in the list of geofences. To detect trips where the access was at a surface and the egress at an underground stop or vice-versa, the end points of all remaining surface and underground segments are compared; if an underground segment ends at the tunnel portal where a surface segment begins, or vice versa, the two are combined and the independent underground segment is deleted. At the end of this step, two lists are carried forward:

- A list of surface runs matched to the phone location data, including runs where one of the stations used (access or egress) is underground.
- A list of independent underground segments where both the access and the egress station are underground.

Inferring runs from underground-only data

For underground-only segments, we have underground AVL data but lack phone positioning data. We assume that the time between entering an underground station and exiting at another station consists at most of two segments with a transfer, and we have only two data points to match the phone to a run: The last location reported at the origin station, t_{orig}, and the first location reported at the destination station, t_{dest}.

- If there was no underground transfer, we assume that all unobserved wait time was incurred at the origin station. If the user transferred to a surface run *r1* after leaving the underground system, we also know $t_{board,r1}$. After defining a minimum egress time for the given station k, $t_{egress,k}$ (the time for the user to exit the station, and for the phone's location services to locate the phone, typically less than a minute), we create an anchor time $t_{anchor,dest}$ as: $t_{anchor,dest} = min(t_{dest}, t_{board,r1}) - t_{egress,k}$. We query the database for the run between the origin and the destination station that arrived most recently before $t_{anchor,dest}$, then assign the user to it.

- If there were two segments, it is unknown whether the unobserved wait time was incurred at the origin or the transfer stop. We (arbitrarily) assume it was at the transfer stop. The second run is determined as described above for a single segment, and the first run is determined by creating a time anchor at the origin station, assuming $t_{access,j}$ as the minimum time for a user to reach the underground platform. If the user transferred to the underground system from a surface run, *r2*, we know $t_{alight,r2}$, and we define $t_{anchor,dest} = max(t_{orig}, t_{alight,r2}) + t_{access,j}$. We then query the database for the first between the origin and the transfer stop that departed after $t_{anchor,dest}$, and assign the user to it.

Given the uncertainties associated with underground segments, and the fact that phones can sometimes report an old location if they are not able to acquire a new one, the system should contain checks to ensure that the inferred underground routing is feasible.

Deriving the final transit travel diary

First, some final data cleaning might be necessary: The user's transit travel diary may still contain segments identified as being on board transit even though the phone user was inside a car or on a bicycle traveling alongside a transit vehicle. Identifying these false positives is difficult based on location data alone. Ideally, additional sensor data would be needed, for instance from the accelerometer or microphone. In the present case, a simple heuristic approach was chosen based on the distance over which the phone and transit vehicle were continuously matched. This assumes that due to speed differentials between cars, bicycles and transit vehicles, the majority of spurious matches would be short. Using this simple decision rule, transit segments that either span less than a distance d_0 or where less than a fraction f_0 of the location points are matched between the first and the last matching point are discarded. The exact thresholds are parameters that would need to be estimated.

In section 3.2, the matching problem was described as determining which of a user's trips took place on transit, and in section 3.3, we described the classification of location points between transit segments into activities and transfers. Now that the lists of surface and underground segments has been merged, this procedure will also need to be applied to interchanges between surface and underground segments.

The procedures described this far yield an ordered list of trips on board transit, where a trip maps to a series of transit segments and transfers. If the time between two segments is classified as an activity, the trip ends, and a new one begins after that. Thus, for user u and date d, we have: $(tr_{u,d,1}, tr_{u,d,2}, tr_{u,d,3})$, where:
$tr_{u,d,1} = (ow_{u,d,1}, ts_{u,d,1}, tf_{u,d,1,2}, ts_{u,d,2}, tf_{u,d,2,3}, ts_{u,d,3})$,
$tr_{u,d,2} = (ow_{u,d,2}, ts_{u,d,4})$, etc.
The origin wait time, ow, is included above for completeness, but it still needs to be calculated.

3.4 Comparisons with timetable

Now that the personal transit travel diary has been established, we are interested in comparing the travel times experienced by the user with what that person could have expected based on the timetable. The two following pieces of the puzzle are still missing:

1. Inference of experienced wait times at origin stops and transfer stops for surface segments.

2. Comparison of all travel times with what the passenger should have expected based on the timetable.

There are a couple of challenges. First, a passenger's wait time at a stop cannot necessarily be inferred by determining how long that person was in the immediate vicinity of the transit stop before boarding because real-time arrival predictions are widely available and delivered through multiple channels, including ones that the researcher may not observe. Second, time spent by the traveler between transit segments might constitute either transfer time or a legitimate activity for which the traveler stopped at this location. Absent other information, inferences must be made. In the following subsections, we present the calculation of travel times and delays from the passenger's perspective, considering the individual elements of the trip separately - first the origin wait, then the in-vehicle segments, then transfers.

Characterization of the origin wait

Due to the availability of real-time information, the time at which a passenger goes to the stop is not necessarily indicative of the beginning of that passenger's wait time. Investigating the different components of the wait time and schedule delay at the beginning of a trip is beyond the scope of this chapter, but it merits further research. As we are concerned with deriving the elemental pieces of information that can form the basis of such future research, the following discussion aims to show what information can be extracted after combining the AVL and phone location data.

Figure 3.5 gives an overview of the various components of origin wait time that are of interest. Specifically, they are:

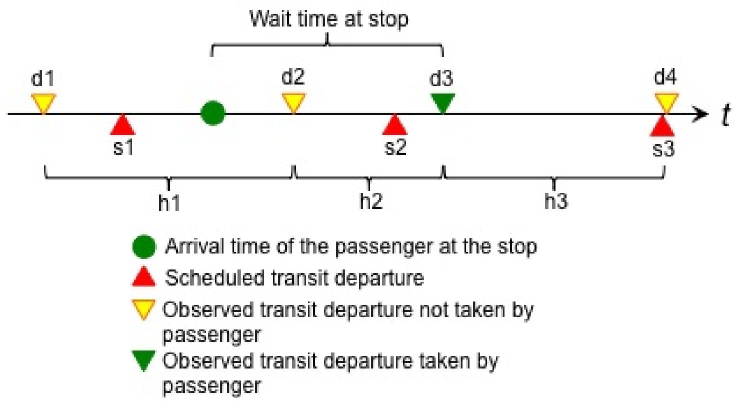


Figure 3.5: Components of the origin wait time from a passenger's perspective.

1. The at-stop wait time.

2. The number of observed departures while the passenger was waiting at the stop (e.g. *d2* in figure 3.5). In congested transit networks, this may be due to the passenger being unable to board a vehicle because of crowding.

3. The observed headways preceding (*h2*) and following (*h3*) the run taken by the passenger (*d3*) as a measure of the observed level of service. If we choose to ignore the departure that the passenger was unable to board, we can consider the headway to be *h1+h2*.

4. The scheduled headways preceding and following *d3* as a measure of the scheduled level of service.

5. The number of scheduled departures while the passenger was waiting at the stop (in this case, *s2*).

6. The number of scheduled departures during the observed headways preceding and following *d3*, i.e., scheduled departures that were not served. Again, if we choose to ignore the departure that the passenger was unable to board, the unserved scheduled departures would include *s1* and *s2*).

The wait time at the stop is calculated by observing how long a person dwelled within a certain radius around the origin stop before the arrival of the run that person took. However, in dense areas, the traveler may have been carrying out an activity near the stop (e.g., shopping), which should not count as wait time. To avoid these erroneous inferences, we proceed as in section 3.3 and query the database for all departures that served the passenger's observed origin and destination stops within a time window before the observed departure time. The time spent near a stop is then only classified as at-stop wait time if:

- Less than a certain number of departures were observed while the person was waiting near the stop. In the system described here, we set the threshold at ≤ 1.

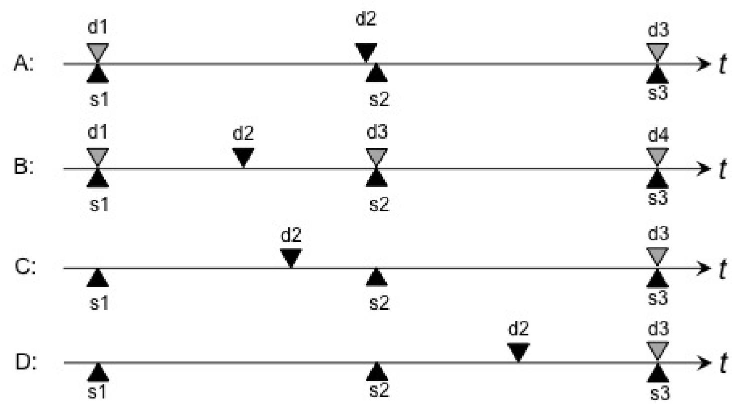


Figure 3.6: Inference of intended departure time (legend in figure 3.5).

- A maximum wait time was not exceeded.

If we assume the passenger is not cognizant of the timetable (for example on routes with frequent service), the passenger's perception of schedule deviation will depend on factors that require further research. Otherwise, we can make an inference on the *intended scheduled departure* based on the observed departure time, and thus calculate a schedule deviation from the passenger's perspective. A suggested procedure is shown in figure 3.6. It should be noted that, to fall into this category, the passenger does not necessarily need to know the published timetable. Many online transit planners, online maps and smartphone apps base their information on GTFS, and thus on the static timetable. Therefore, a traveler who consults one of those products to plan his or her trip can be considered to be cognizant of the timetable.

Several possible cases are illustrated. In case A, departures *s1* and *s3* were served by an observed departure, so we assign the departure taken by the user to *s2*. Thus, the passenger delay with respect to the timetable is *d3-s2*, which can be negative or positive. In case B, on the other hand, *d2* is not assigned to any scheduled run; a passenger who caught *d2* is not assumed to have been intending to catch *s1* or *s2* since both scheduled departures were served. Therefore, we assign a delay of zero to this case. Lastly, consider cases C and D: The observed departure is surrounded by several scheduled departures that were not served, so without additional information, we do not know which of them the passenger intended to take. Because passengers have access to real-time information, they would not go to a stop if they knew that a scheduled departure was not going to be served. As an approximation, we assign probabilities. If historical demand data are available, the probabilities can be derived from it, but failing that, we assign equal probability to all unserved departures before and after *d2*, i.e., for N unserved departures, $\sum_{n=0}^{N}(d2 - s_n)$. Given that these differences may have different signs, the sum may be counter-intuitively small, as in case C where $dev = \frac{1}{2} \cdot (d2 - s1) + \frac{1}{2} \cdot (d2 - s3)$. In future work, when developing personal reliability metrics, we therefore recommend instead summing the absolute values of the deviations weighted by a function for the disutility experienced by the traveler. Lastly, it should be emphasized that this is *not* an operational assignment, i.e., an inference of which scheduled run an observed run was covering, but rather, it is an inference about the scheduled departure the passenger *intended* to take. For further discussion of the problem of assigning observed runs from AVL data to scheduled runs, please see Ji, Mishalani, and McCord (2009).

Deviation of the on-board travel time

For calculating the schedule deviation of the IVTT, we must match d, the observed departure time of the passenger, with one scheduled departure time, s. The passenger would have departed at s if he or she had been at the stop at d.

Based on findings by Carrel, Halvorsen, and Walker (2013), we make the assumption that for infrequent service (headways $\geq$ 12 min.), passengers know the timetable, whereas for frequent service (headways $\leq$ 12 min.), they do not.

If service is infrequent and the departure was one of cases A or B in figure 3.6, this assignment has already been made. In cases C and D, we assign d to the nearest scheduled departure that was not served (in these two examples, *d2* to *s2*). If it is frequent service, we assume a random arrival and define a slack time $t_{w,slack}$ that precedes every scheduled departure to allow for marginally early departures. Then, we assign d to the next scheduled departure s if it is within $s - t_{w,slack}$, and otherwise to the previous scheduled departure. Unlike the calculation of the origin wait time, this assignment has no behavioral significance; we only use it to query the scheduled travel time between the passenger's boarding and alighting stop, $ivtt_{sched}$, and to calculate the *projected arrival time* and the *scheduled arrival time* to compare with the observed arrival time. These are defined as follows:

$t_{arr,obs} = d_i + ivtt_{obs}$	:	The observed arrival time at the destination stop is the observed departure time plus the observed travel time.
$t_{arr,proj} = d_i + ivtt_{sched}$	:	The projected arrival time is the observed departure time plus the scheduled travel time.
$t_{arr,sched} = s_i + ivtt_{sched}$	:	The scheduled arrival time is the scheduled departure time plus the scheduled travel time.

Transfer time variability

Given a transfer between two runs, the task is to calculate:

1. Observed, projected and scheduled transfer times (defined below).
2. Observed departures of the connecting route during delays.
3. Scheduled departures of the connecting route during delays.

We define the transfer times based on our definition of the arrival times in the previous subsection, as illustrated in figure 3.7.

Suppose the phone user first used route A, which was scheduled to depart at *s1*, but was delayed and did not depart until *d1*. The vehicle was again delayed while the passenger was on board, and the observed travel time exceeded the scheduled IVTT. Following the previous subsection, the scheduled arrival time at the transfer stop would have been *a1*, the projected

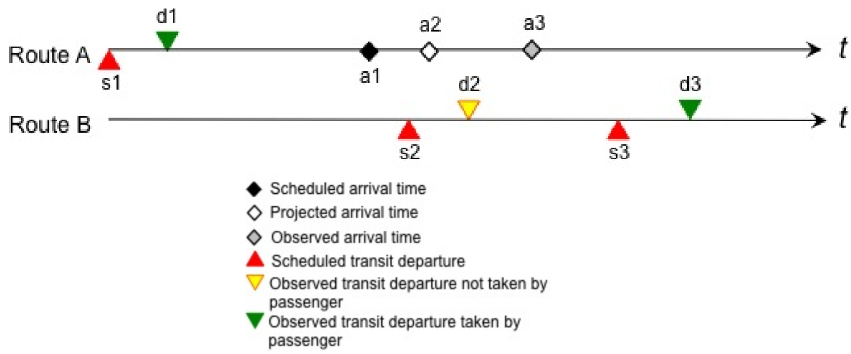


Figure 3.7: Components of transfer delays from a passenger's perspective.

arrival time would have been *a2* and the observed arrival time was *a3*. Note that *d1-s1* = *a2-a1*.

At the transfer stop, the passenger transferred to route B. There are two scheduled departures from the transfer stop on route B, at *s2* and *s3*. Both are delayed, so *s2* departs at *d2* and *s3* departs at *d3*. The passenger catches the connection at *d3*. Thus:

- The *scheduled* transfer time is *s2-a1*.
- The *projected* transfer time is *d2-a2*.
- The *observed* transfer time is *d3-a3*.

The scheduled and projected connections are determined subject to a minimum transfer time constraint. Furthermore, we can calculate:

- Scheduled departures and observed departures during the delay with respect to the scheduled arrival time (*a3-a1*). In the example, this includes *s3* and *d2*.
- Scheduled departures and observed departures during the delay with respect to the projected arrival time (*a3-a2*). In the example, this includes *d2*, but no scheduled departure.
- Scheduled departures and observed departures during the passenger's wait time at the transfer stop (*d3-a3*). In the example, this includes *s3*, but no observed departure.

For every transit segment that follows a transfer, this information is carried forward by determining the scheduled departure time and the projected departure time on route B based on the scheduled arrival time and the projected arrival time calculated on route A. In figure 3.7, this would be *s2* (scheduled) and *d2* (projected).

Overall travel time variability

There are many dimensions to passenger travel time variability, and that a number of different variables can potentially be used to express it. The building blocks that were derived using the system presented here can be used in future research to discuss the calculation of personal travel time distributions and user-based reliability metrics. The most important components of a personal travel time distribution would be the wait time at the stop, the IVTT and the observed transfer times. However, developing an accurate measure for the actual wait time at the origin stop (including wait time not observed by the researcher) requires further research and, for a complete picture, the access and egress times should be added as well. Similarly, for a user-based reliability metric, the most clearly defined components are the deviation of the IVTT from scheduled travel time and the wait time at the stop, but in particular measures for the schedule deviation of the out-of-vehicle components of the travel time at the origin and the transfer stops require further research, and a discussion is needed on how to include missed departures. With the development of this system and the description of the elemental components of travel time from the passenger's perspective, we hope to be able to provide a foundation for such efforts.

3.5 Deployment, validation and example analyses

The system was deployed in a real-world, large scale study as part of the San Francisco Travel Quality Study. In total, approximately 80 million location points were collected. The system described in this chapter was able to identify approximately 7700 transit passenger trips using Muni (SFMTA's light rail and bus service), of which 675 (8.8%) involved a Muni-to-Muni transfer according to the definition introduced in this chapter.

In the remainder of this section, example analyses relating to the out-of-vehicle portions of the transit trips made by participants are presented in order to underscore the importance of capturing the entire trip and all of its components, and to show that an approach based on high-resolution phone location data combined with AVL data can provide richer information and more insights compared to using AFC or APC data.

Validation

To test the inference system described above, ground truth data were collected on a total of 103 transit passenger trips made between March and May 2014. The data consisted of 86 above-ground trips by bus or light rail and 17 light rail trips that were entirely or partially underground. The evaluation showed that 92 (89.3%) of all runs were correctly identified and, for a further 4 underground trips, the origin/destination stations and times were correctly identified, but due to faulty AVL data, the passenger was mapped to a train of a different route. We include those 4 in the list of correctly identified runs since the error was due to the AVL data rather than to the matching algorithm, which allows us to conclude that we obtained a 93% accuracy in identifying the transit vehicle runs a passenger used. Out of those 96 runs, the boarding and alighting stops were identified exactly in 83 (86%) cases, and in 95 (98%) cases, the boarding and alighting stops identified were within one stop of the true boarding and alighting stop. Given that in downtown San Francisco stop spacing can be fairly small, we consider these results to be very good.

It should be noted that these numbers describe the true positive rates. A limitation of this validation procedure is that no ground truth data on non-transit trips was available.

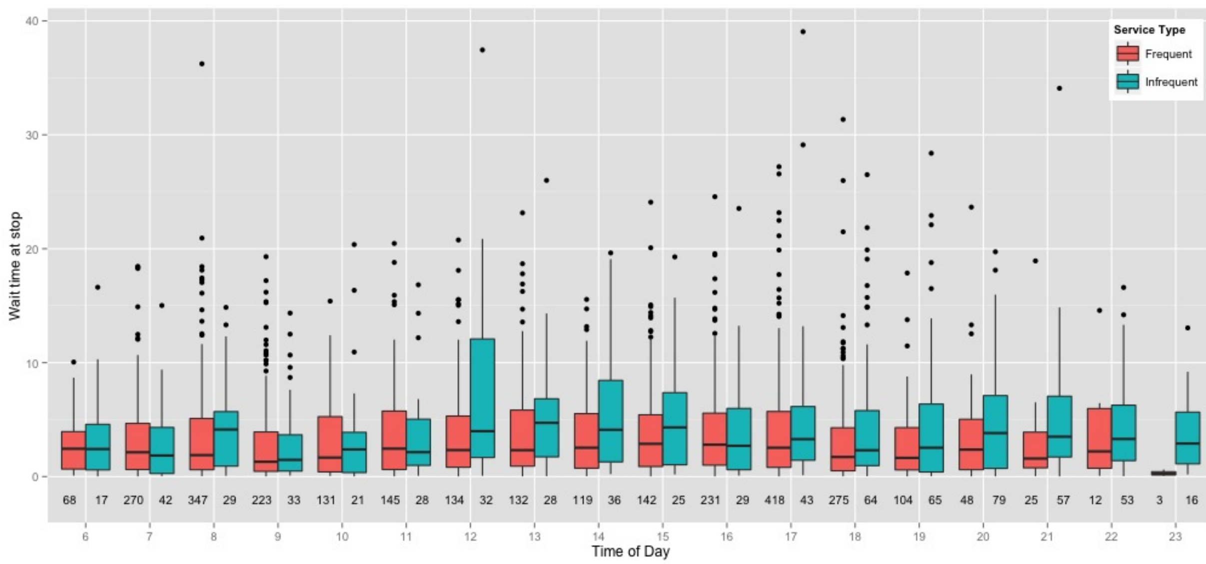


Figure 3.8: Observed wait times at the origin stop.

Therefore, it was not possible to assess the false positive rates, which could be caused by situations as described in section 3.3.

Example analysis: Wait time at the origin stop

The first analysis concerns passenger wait time at the origin stop. The wait time was defined as the time a passenger spent within the area defined by a 100 m. radius around the origin stop before boarding. The upper bound was defined as either 45 min. or the passage of the second observed departure, whichever occurred first. If the passenger was still at the stop after that point, the origin wait time would be marked as unknown since the passenger might have been carrying out an activity in the vicinity of the stop. However, in this practical application, the 45 min. upper bound proved to be a non-binding constraint due to the high service frequencies of the routes used by the study participants. Figure 3.8 shows the results by time of day and by service type. We classified any service with a scheduled headway $\leq$12 min. as frequent. The box shows the range between the 25th and 75th percentiles of wait time, and the whiskers extend to the highest value that is within 1.5 times the interquantile range. The numbers above the horizontal axis show the sample size.

A few notes can be made: Overall, the wait time distributions do not appear to differ strongly and systematically between frequent and infrequent service. The larger variability of the distributions throughout the day for infrequent service may in part be due to smaller sample sizes. The median wait time appears to be at a minimum around 09:00 and 18:00, with a slight increase during midday. Furthermore, the median wait time for frequent service is only around 2-3 minutes and the 75th percentile hovers around 5 minutes. While further research is required to analyze these patterns, it is plausible that the availability of real-time information has an effect in that passengers tend to time their arrivals at the stop, regardless of the time of day or type of service, opting to arrive at a more or less fixed amount of time before the predicted departure.

Until now, the collection of waiting time data in ungated systems has not been feasible on a large-scale basis, save with prohibitively expensive methods. Therefore, waiting time has traditionally been estimated with models and/or assumptions. This example demonstrates the value of the developed matching system in producing actual observations of waiting time on a large-scale basis; such data are critical for a better understanding of the waiting time phenomenon, including, for example, the effect of real-time vehicle arrival information systems as briefly discussed above.

Example analysis: Contribution of origin wait and transfer time to total travel time

A large concern of many users of public transit is arriving at their destination on time, especially on commute trips. Due to variability of transit travel times, users are typically faced with a distribution of arrival times at their destination, and thanks to the combination of phone location data and AVL data, it is possible to decompose that total travel time into its various components in order to understand the source of delays and variability. Such analyses will be able to go into greater detail, but in this section, we limit ourselves to a high-level analysis of the overall travel time variability experienced by the SFTQS participants. For that purpose, we selected the trips containing only one transfer and where the origin wait time had been determined (see section 3.4 for more details). Figure 3.9 shows the contribution of wait time at the origin stop, transfer time and in-vehicle travel time to total travel time for trips containing exactly one transfer. Total travel time in this case is defined as the time between when the passenger first arrived at the origin stop and the time when the passenger alighted at the destination stop. As total travel time increases, the contribution of the origin wait time remains between 8% and 14% of total travel time. On the other hand, the contribution of wait time at the transfer stop appears to increase with total travel time, from approximately 17% in travel times between 10 and 20 min. to 35% for travel times between 50 and 60 min. As an example, on average, a passenger who made a trip that took 45 minutes from the passenger's arrival at the origin stop to alighting at the destination stop, spent 31 min. in a vehicle and 14 min. waiting. On average, around 70% of that wait time was incurred at the transfer stop. Distinguishing the waiting time at the origin stop from the transfer time is particularly important because it is likely that the disutility experienced by the passenger at the transfer stop is greater (Wardman, 2004). There are two principal reasons for this: The traveler has more control over the wait time at the origin stop, and the wait time at the transfer stop is often unknown when the traveler commits to taking transit for that trip.

Finally, we calculated the specific contribution of in-vehicle and transfer travel time variability to the variability of the arrival time at the destination. The deviation of the in-vehicle travel time, the transfer wait time and the arrival time at the destination from the schedule were calculated as described in section 3.4. For the transfer and the arrival time, we used the deviation from the scheduled transfer and arrival time (rather than from the projected time). The variability of wait time at the origin stop had to be excluded since

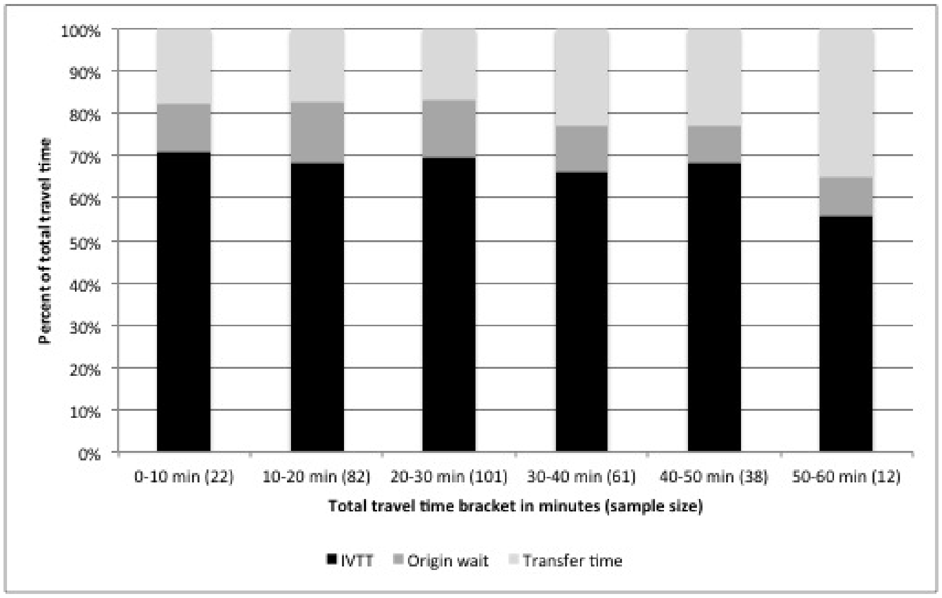


Figure 3.9: Contribution of trip segments to overall travel time.

IVTT	Transfer time	Arrival time	IVTT factor	Transfer factor
+	+	+	%	%
-	-	-	%	%
+	-	+	1	0
+	-	-	0	1
-	+	+	0	1
-	+	-	1	0

Table 3.1: Attribution of arrival time deviations to IVTT and transfer time deviations.

there was no schedule benchmark to compare it to. For this analysis, deviation of the arrival time with respect to the scheduled arrival has to be attributed to these components. This was done as shown in table 3.1: If all three deviations were of the same sign (all positive or all negative), the deviation at the arrival stop was attributed proportionally according to the contribution of each component. This is marked with "%" in table 3.1. If only one of the components had the same sign as the arrival deviation (e.g., the IVTT was lower than scheduled, but there was a delay at the transfer point and the passenger arrived late at the destination), the deviation was fully attributed to the component of the same sign. This is marked with "1" in table 3.1. None of the deviation was attributed to the other component, which is denoted with a "0".

The sample size was $N{=}533$. The average IVTT and transfer factor (as shown in table 3.1) are 39% and 61%, respectively. In determining the average contribution of the IVTT and transfer time deviations to the arrival time deviation, one needs to choose whether to weight each data point or to calculate the average percentages unweighted. We chose to weight each data point by the absolute amount of time by which the final arrival time deviated from the scheduled arrival time, as is shown in equation 3.1. This choice was made for the following three reasons:

- We felt this was most representative of the aggregate experience of the group, attributing more weight to large deviations than to small ones.
- Weighting them by the absolute value of the final arrival time deviation is more conservative than weighting them by the sum of the absolute values of the IVTT and transfer time deviations: If the two deviations were in opposite directions and partially canceled each other out, the arrival time deviation at the final destination is smaller than the latter sum.
- We are only interested in deviations, regardless of whether they are positive or negative (i.e., early or late arrivals). The calculation here is not intended to reflect the disutility experienced by the traveler, where late arrivals would most likely need to be weighted differently than early arrivals.

$$\frac{\sum_{n=1}^{N} min\left(\left|f_i \cdot dev_i\right|, \left|dev_{arr}\right|\right)}{\sum_{n=1}^{N} \left|dev_{arr}\right|} \tag{3.1}$$

In equation 3.1, dev_i is either the transfer time deviation or the origin wait deviation and f_i is the respective factor according to table 3.1. Using this methodology, the contribution of the IVTT to arrival time deviations is calculated to be 26% and the contribution of transfer time variability 74%.

This again demonstrates the value of using phone location data matched with AVL data to track out-of-vehicle travel times. With AFC data, one would almost inevitably need to rely on assumptions: in most gated systems, passengers are not required to tap their fare cards when transferring, and in ungated systems, it is often not known whether the passenger actually alighted at the transfer stop and whether the passenger spent the transfer time waiting or pursuing a non travel-related activity. The former can now be observed exactly, and although the inference of the latter with phone location data may not be perfect, the amount of data available for classifications is significantly richer. Since commonly used reliability metrics often ignore transfer times and focus only on operational measures of vehicle travel times, this also underscores the strong need for future research on the development of additional user-centric reliability and performance metrics that include out-of-vehicle travel times.

3.6 Conclusions

This chapter introduced a system to extract the personal transit travel diary of participants collecting location data with their smartphones by matching their location points to AVL data. It described the various components of a transit passenger's trip and how they can be measured, focusing in particular on the problems that can arise when the phone location data are sparse and when the phone is in the vicinity of multiple vehicles at a given time. Furthermore, this chapter described an approach to detect underground travel on metro networks when AVL data but no phone positioning data are available while the user is underground. The procedure presented in this chapter is of a general nature and can be applied to a variety of different systems; the outlined steps can be implemented by researchers or practitioners and adapted to the specific use cases at hand. Thus, the steps described can be understood as a blueprint for matching smartphone tracking data with AVL data and for dealing with various types of data quality issues. While the described procedure yields exact routing for trips above ground, it relies on path choice inference for underground trips. In simple metro networks, such as the one this system was developed for, the assignment of trips to the shortest path is mostly unproblematic. However, in more complex networks with multiple possible routes between stations, further research may be required to improve path choice inference. It is possible that in the future, if public and private Wi-Fi access points begin appearing in metro stations, underground phone positioning data will become available as well, which would greatly aid in determining underground route choice.

Because the location data can be collected from virtually any app, a researcher could use data from third-party apps such as route planners. The system described in this chapter requires no further input from the participants aside from their consent to allow data collection, which in turn facilitates long-term collection of travel data and the observation of the dynamics of transit use by researchers. Detecting the origin, destination and transfer stops with phone location data is especially useful in systems where passengers are required to tap their fare card either only once or not at all since this information cannot be directly observed from AFC data. This can serve for the construction of OD matrices and the calculation of total travel times. Even in systems where passengers are required to both tap in and tap out, phone location data offers an advantage in that it allows the observation of out-of-vehicle components and of the true origin and destination.

Given the elemental components of a passenger's trip, including wait times, IVTT and transfer times, this chapter then explored how travel time variability with respect to the scheduled service can be measured. Generally speaking, the disaggregate travel time data and personal travel time distributions can be used for system performance monitoring via user-based reliability metrics, for planning applications and for research into passenger preferences and behavior. This can significantly enhance the transit agency's understanding of its passengers. The sample applications presented in this chapter are intended to demonstrate the potential value of such data. Although the system presented does have its limitations, it is hoped that those can be addressed in future research. Notably, additional sensor data would be required to reduce false positive rates caused by a person traveling alongside a transit vehicle and for better distinguishing activities from wait times at stops, and a methodology would need to be devised to determine passenger routing in complex underground networks with multiple possible paths. Overall, however, the approach outlined in this chapter holds great promise for delivering data that will enhance the planning and management of public transportation systems.

Chapter 4

Dissecting Satisfaction

4.1 Introduction and motivation

Performance measurement is a key factor driving policy, planning and operations in public transportation. Customer satisfaction surveys are not only one of the most widely and regularly used performance measurement tools in the industry, but they are also the most direct way of capturing the customers' perspective (Davis and Heineke, 1998; Hensher, Stopher, and Bullock, 2003a). Moreover, satisfaction can be considered an indicator of future choice behavior (Oliver, 2010). Typically, the results of satisfaction surveys stand alongside operational metrics, but with a few exceptions (Friman, Edvardsson, and Gärling, 1998; Morfoulaki, Tyrinopoulos, and Aifadopoulou, 2010), satisfaction has rarely been linked to objective measures of service quality on a large scale. In this chapter, we focus on customer satisfaction with respect to individual travel time components (wait time, in-vehicle travel time and transfer time), which are an important aspect of the transit experience. Operational travel time measures typically capture service provided at an aggregate level, whereas customers respond to satisfaction surveys with respect to their personal experiences. While it can be argued that this difference diminishes with a large enough sample size, much valuable information is lost in the aggregation step.

In order to empirically understand the drivers of dissatisfaction among customers, this chapter presents results from comparing participants' travel times as measured via smartphone location tracking and automatic transit vehicle location (AVL) systems with users' satisfaction ratings from the daily mobile surveys. By connecting these individual-level diary data on personal travel experiences to the survey responses, satisfaction ratings can be understood against the backdrop of objectively measured and quantified service times and delays. As will be demonstrated, these data can provide valuable new insights into customer satisfaction and behavior.

Section 4.2 features a literature review, followed by the presentation of the modeling framework and the explanatory variables in section 4.3. Section 4.4 offers a brief description of the data, and then presents the results of the model estimation, along with an interpretation of the results and a sensitivity analysis.

4.2 Literature and contribution

The majority of satisfaction surveys described in the literature were conducted only once, asking respondents to rate their satisfaction with the entirety of transit service. In fact, the time frame within which respondents were asked to evaluate their experiences with the transit service was often not even specified. Most researchers have sought to quantify the importance of different service quality aspects or proposed aggregate satisfaction metrics (Swanson, Ampt, and Jones, 1997; Stuart, Mednick, and Bockman, 2000; Friman, Edvardsson, and Gärling, 2001; Stradling et al., 2007; Eboli and Mazzulla, 2007; Tyrinopoulos and Antoniou, 2008; Eboli and Mazzulla, 2009; Eboli and Mazzulla, 2010; Eboli and Mazzulla, 2011; Cirillo, Eboli, and Mazzulla, 2011). The inverse relationship, i.e., between aggregate satisfaction ratings and satisfaction ratings with specific attributes of the service, was investigated by Del Castillo and Benitez (2013). However, we argue that satisfaction surveys are most valuable if a link can be made between satisfaction and objective service quality measures (Davis and Heineke, 1998); this was not done by the above studies. Otherwise, an analyst might know that customers are dissatisfied with a particular aspect, but would not be able to know the sensitivity of satisfaction ratings with respect to the delivered service. To the best knowledge of the author, only one study thus far has paired objective service quality data with subjective satisfaction ratings to identify causes of satisfaction from operational service quality data: Friman and Fellesson (2009) investigated satisfaction with transit services in several European cities as a function of aggregate service provision metrics. However, they could not find any significant relationship. The authors concluded that this might be due to several issues, among them the difficulties of transnational comparisons. They also noted that the high-level measures of service provision used may have been too coarse, and that they say little about how well supply is matched with demand or about how the service is delivered (including variables such as staff behavior or fares). Furthermore, they point out that passenger travel behavior (such as trip lengths), which was not included in their models, might be important.

The point by Friman and Fellesson can be elaborated on: Customer satisfaction is, to a large part, a function of *personal use experience* (Woodruff, Cadotte, and Jenkins, 1983; Anderson and Sullivan, 1993), and in transportation, it is generally an aggregate of multiple repeated experiences over time. The larger that time frame is, the more likely a respondent's reported satisfaction might be a function of unobserved events or memory distortions such as the peak-end rule known from behavioral economics (Fredrickson and Kahneman, 1993). Among transit users, this difference between experienced and remembered utility was observed by Pedersen, Friman, and Kristensson (2011). To control for this, the analyst needs to either collect additional data on the respondents' history of transit use or limit the time frame for which the satisfaction survey is being conducted to a recent episode that respondents can remember clearly (Kahneman et al., 2004).

This chapter is aimed at investigating the link between objective, quantifiable measures of travel quality and customer satisfaction at a personal level by using smartphone data to capture respondents' transit travel experiences and connecting them with satisfaction surveys. Similar studies have previously been conducted in other industries where service times are critical, such as emergency medical services (Thompson et al., 1996) or fast food (Davis and Maggard, 1990). An important study was by Davis and Heineke (1998), who found that perception of waiting time was more directly related to customer satisfaction than actual waiting time. The difference between perceived and experienced wait time has become important in the transit realm with the advent of real-time information. Watkins et al. (2011) discovered that without real-time information, riders perceived wait times to be greater than the true wait times, whereas with real-time information, they did not. This chapter considers the direct link between experienced travel times and satisfaction, omitting the moderating effect of perception. However, since a requirement for participating in the study was to own a smartphone with a data plan, it could be assumed that all participants had access to real-time information.

The studies cited above (Thompson et al., 1996; Davis and Maggard, 1990) focused on spatially contained areas where data could be collected more easily. With the advent of location-aware smartphones as data collection tools, spatial limitations are diminishing, and it is becoming possible to study customer behavior in more open, distributed systems such as public transportation. This chapter demonstrates the power and value of connecting subjective passenger satisfaction surveys with other data sources, namely smartphone tracking data and AVL data to obtain a quantitative understanding of what drives customer satisfaction and how dissatisfaction is related to poor performance. In a series of models, we observe the link between wait time, transfer time and in-vehicle travel time (IVTT) on the one hand, and satisfaction with these travel time components on the other hand while accounting for the influence of a series of covariates. Furthermore, the relative importance of the three travel time components in customers' perception of reliability is estimated in a joint model. In previous literature, where in-vehicle delays were not separately identified, the assumption has been made that IVTT is generally less onerous than out-of-vehicle wait time (Wardman, 2004). However, our results show that when in-vehicle delays are entered into the model in addition to IVTT, a more complex picture of how passengers experience travel times emerges: In-vehicle delays appear to be strong drivers of passenger dissatisfaction, and passengers seem to be able to distinguish between scheduled IVTT and deviations from the latter. There is a potential link between our results showing that scheduled IVTT (which can be considered "expected" travel time) has a relatively low disutility and results by Ory and Mokhtarian (2005) and Diana (2008), who investigated the intrinsic, positive value of traveling.

The results of this study also contribute to the general travel survey literature by distinguishing between the baseline satisfaction with transit services (with no specified time frame) and variable daily satisfaction measurements. Furthermore, a measure of subjective well-being (SWB) was included, and it was found to be strongly correlated with satisfaction ratings regarding travel times. SWB is still a relatively rare variable in travel satisfaction surveys, but our findings suggest it would be an important addition. This is in line with findings by Susilo and Cats (2014).

4.3 Data and Modeling framework

Data

We make use of the data set collected during the San Francisco Travel Quality Study, described in chapter 2. A complete survey data set, including the entry and exit survey and at least five daily mobile surveys, could be collected from 604 participants, and partial data sets from a further 148. Data sets of 723 respondents included the entry survey and one or more mobile survey responses; of those, the data sets of 560 respondents could be matched to transit trips identified from tracking data. For 71% of the participants, trips with both bus and rail were detected, whereas for 18% only bus trips were detected, and for 11% only metro trips. No significant differences in age, gender and income were found between users of both modes and users of only one mode.

Table 4.1 shows the distribution of daily mobile survey responses with respect to the entry survey responses in the same category. For all five levels of entry survey satisfaction, there is a spread of responses in the daily mobile surveys, but an interesting pattern can be seen: The entry survey satisfaction levels 4 and 5 ("satisfied" and "very satisfied") predict the largest response group in the mobile surveys. For example, participants who responded with 4 to the question regarding their satisfaction with IVTT in the entry survey were also most likely to respond with 4 to the daily mobile survey question regarding their experienced IVTT. This does not appear to hold true for participants whose ratings in the entry survey were between 1 and 3.

The Ordinal Logit model

Given the data, a series of ordinal logit models was estimated to relate the satisfaction ratings to the observed travel times while controlling for sociodemographics and mode access (Dios Ortúzar and Willumsen, 2011). The ordinal logit model assumes that there is a continuous underlying utility, or in this case, satisfaction. Satisfaction is modeled as a linear function of traveler characteristics and travel times with an i.i.d. extreme value distributed error term. Since the responses are on an ordinal scale, the continuous error distribution is sectioned into as many intervals as there are scale points, and intercept terms τ are estimated as the dividers between intervals. Thus, the probability of choosing response k on the Likert scale is given by:

$$P(k) = \frac{1}{1 + e^{-\mu(V - \tau_{k-1})}} - \frac{1}{1 + e^{-\mu(V - \tau_k)}} \tag{4.1}$$

In the above equation, V is the systematic utility and μ is a scale parameter.

Choice of explanatory variables

A total of four models were estimated. The dependent variables were satisfaction measurements with: (1) IVTT, (2) wait time at the origin stop, (3) transfer time, and (4) overall reliability. These satisfaction measurements relate to the respondent's experience taking transit on one specific day. The independent variables could be divided into three groups:

IVTT	Mobile 1	Mobile 2	Mobile 3	Mobile 4	Mobile 5	Mobile N/A
Entry 1	59	60	89	65	37	0
Entry 2	56	139	259	264	63	0
Entry 3	79	189	533	598	277	0
Entry 4	205	425	977	1694	991	0
Entry 5	70	151	255	490	753	1
Wait time	Mobile 1	Mobile 2	Mobile 3	Mobile 4	Mobile 5	Mobile N/A
Entry 1	118	116	118	208	101	0
Entry 2	253	466	479	714	460	0
Entry 3	182	428	667	943	723	0
Entry 4	138	265	401	895	801	1
Entry 5	13	22	47	78	142	0
Transfer time	Mobile 1	Mobile 2	Mobile 3	Mobile 4	Mobile 5	Mobile N/A
Entry 1	27	22	67	39	16	334
Entry 2	48	66	217	144	82	1220
Entry 3	47	73	285	147	91	1827
Entry 4	92	125	310	395	347	2265
Entry 5	7	7	54	19	68	338
Overall reliability	Mobile 1	Mobile 2	Mobile 3	Mobile 4	Mobile 5	Mobile N/A
Entry 1	82	115	207	296	117	0
Entry 2	173	342	476	740	349	0
Entry 3	92	217	564	996	442	0
Entry 4	78	222	478	1374	975	1
Entry 5	10	12	41	86	294	0

Table 4.1: Distribution of entry and mobile satisfaction responses. 1 = Very dissatisfied, 5 = Very satisfied.

The first group contained sociodemographic and mode access variables, which are invariant for an individual across observations and across models. The second group contained individual-specific variables related to one of the travel time components. These are invariant across observations, but there is little sense in including them in all four models since they are specific to the travel time component being modeled in only one of them. The first two groups of variables were collected with the entry survey that participants filled out once at the beginning of the study. In an effort to maintain comparability across models, the same sociodemographic and mode access variables were included in all models, even if their levels of significance differed. The third group of variables contained the observed travel times on the day that the satisfaction measurements were taken for. The following list provides an overview of the explanatory variables included in the models. Table 4.2 shows an overview of the explanatory variables along with the variable names that are used to present the results later.

Variable name	*Units*	*Explanation*	*Survey*
		Dependent variables	
Daily satisfaction with service (mobile survey)			
sat_ivtt	5-pt. Likert	Satisfaction with IVTT	Daily
sat_wait	5-pt. Likert	Satisfaction with origin wait times	Daily
sat_transfer	5-pt. Likert	Satisfaction with transfer times	Daily
sat_reliability	5-pt. Likert	Satisfaction with overall reliability	Daily
		Independent variables	
Baseline satisfaction ratings (entry survey)			
pre_sat_ivtt	5-pt. Likert	Satisfaction with IVTT	Entry
pre_sat_wait	5-pt. Likert	Satisfaction with origin wait times	Entry
pre_sat_transfer	5-pt. Likert	Satisfaction with transfer times	Entry
no_pre_sat_transfer	Binary	No response to pre_sat_transfer	Entry
pre_sat_reliability	5-pt. Likert	Satisfaction with overall reliability	Entry
Sociodemographics			
age	Years	Age bracket midpoint	Entry
female	Binary	Gender	Entry
income	$10.000	Income bracket midpoint	Entry
unknown_income	Binary	Income is unknown	Entry
employed	Binary	Part-time or full-time employed	Entry
Mode access			
longuser	Binary	Has used Muni regularly for >2 years	Entry
beforestudycar	Binary	Used a car in SF regularly before study	Entry
bike_owner	Binary	Owns a bike in usable condition	Entry
personal_car	Binary	Owns a personal car	Entry
shared_car	Binary	Shares car with other household members	Entry
Mood/Subjective well-being			
general_feeling	5-pt. Likert	"How do you generally feel today?"	Daily
Travel time variables			
sched_ivtt	Minutes	Scheduled IVTT	Daily
ivtt_early	Minutes	Observed IVTT < Scheduled IVTT	Daily
ivtt_delay	Minutes	Observed in-vehicle delay (all modes)	Daily
ivtt_delay_bus	Minutes	Observed in-vehicle delay (Bus)	Daily
ivtt_delay_metro	Minutes	Observed in-vehicle delay (Metro)	Daily
post_headway_diff	Minutes	Difference between headway before observed departure and scheduled headway	Daily
pre_headway_diff	Minutes	Difference between headway after observed departure and scheduled headway	Daily
origin_wait_time	Minutes	Wait time at origin stop	Daily
dep_during_wait	Binary	Denied boarding at the origin stop	Daily
transfertime	Minutes	Observed transfer time	Daily
dep_during_transfer	Binary	Denied boarding at the transfer stop	Daily
unobserved_transfer	Binary	Respondent reported making a transfer	Daily

Continued on next page

Variable name	*Units*	*Explanation*	*Survey*
		but no data were available	
rti_accuracy	5-pt. Likert	"How accurate was NextBus?"	Daily
no_rti_response	Binary	No response to rti_accuracy	Daily
weekend	Binary	Trip took place on a weekend	Daily

Table 4.2: Overview of explanatory variables used in the ordinal logit models.

1. Sociodemographic variables: Among other variables, the entry survey collected respondents' age, gender and household income. Respondents had the option not to respond to the income question, so this information was only available for part of the sample. Age was measured in categories: 18-24 years old, 25-34 years, 35-44 years, and so forth. The highest age category was 65 years and older. 18 was the minimum age for participation. This variable was coded in the models as the midpoint of each age bracket (e.g., 21, 29, 39, and so forth). The highest age bracket was coded as 69. Income was also measured in steps of $20.000, i.e.: $0 - $20.000, $20.001 - $40.000, and so forth. The highest category was $120.000 and over. Income was coded in the model as the midpoints of the income brackets, in multiples of $10.000 so as to scale the variable to a similar magnitude as other variables in the model (e.g., 1, 3, and so forth). The highest category was coded as 13. For observations with missing income, a binary variable was created. Lastly, employment status was entered as a binary variable where 1 indicated somebody who was either part-time or full-time employed. Student status was not included as it was almost complimentary to employment status. A priori, it was expected that higher-income and employed individuals would be more sensitive to delays. There was no a priori expectation about the effects of age and gender.

2. Mode access variables: These were also included in the entry survey. Respondents were asked whether they owned a bicycle or a personal car, or whether they shared a car with other household members. Those were entered as binary variables. Furthermore, respondents were asked whether they had used San Francisco Municipal Transportation Agency services (Muni) and/or a car in San Francisco regularly before the study; if they had used Muni regularly, they were also asked for how long they had been doing so. The binary variable for having regularly used a car was included in the model, but the variable for regularly having used Muni was not, since 98% of participants answered that question affirmatively. The purpose of the last question was to separate more long-term, seasoned transit users from newer users. Because the time steps on which participants indicated the duration of their regular Muni use were not uniform, a number of binary variables were created and the most significant one was retained (which was found to be more/less than 2 years). A priori, it was expected that users with access to a private motor vehicle or a bicycle or who used a car regularly before the study would be more sensitive to delays and wait times since they would be aware of travel times on alternative modes. There was no a priori expectation about the effect of length of Muni use.

3. Satisfaction in entry survey: In the entry survey, participants were asked for their general satisfaction with IVTT, wait times at the origin stop, transfer times, and overall reliability on five-point Likert scales. This question was not in relation to any particular experience, but rather to all their experiences with Muni up to that point. The response to the satisfaction with transfer times question was optional, and participants were only asked to fill it out if they indicated that they transferred between Muni services regularly. Therefore, an additional binary variable was created to indicate a missing transfer satisfaction in the entry survey. The intention was that this group of variables would serve as measurements of a baseline satisfaction. With that in hand, the relationship between these measurements and variations in satisfaction attributable to daily experiences could be explored.

4. Observed travel time variables: These were derived from the location tracking data. Per day, respondents only gave one response about their overall transit experience, but they might have made any number of transit trips on that day. Based on the time of the response, the survey responses were mapped to all transit trips of that day up until the survey response time. All travel time variables used in the equations were averaged over all relevant trips on that day. If sums had been reported, it might have appeared, for example, that a person who made two or more transit trips experienced worse service than a person who made only one trip due to the summation of travel times. All times were measured in minutes. This group of variables included:

   a) For IVTT: The average scheduled IVTT per trip and the average positive or negative deviation from scheduled IVTT. It is worth noting that the average deviation from scheduled IVTT is measured on an individual level and therefore represents the respondent's first-hand experience with travel time variability. The deviation was split up into two variables, depending on whether the observed IVTT was greater (in-vehicle delay) or less than the scheduled IVTT (the trip was faster than scheduled). The in-vehicle delay was further split up by mode, i.e., delay on bus or rail. A priori, we expected increases in IVTT deviation to negatively affect satisfaction. The scheduled IVTT was included since it was assumed that there would be a general disutility associated with IVTT, though we expected that effect to be small.

   b) For the origin wait time: The observed wait time at the stop, and whether or not a departure was observed while the person was waiting at the stop. The latter variable was meant to capture denied boardings (see discussion below). A priori, longer wait times and missed departures were expected to decrease satisfaction, as were positive deviations from scheduled headways. Unlike the IVTT, it was difficult to define a benchmark for wait times from which deviations could be calculated. Many of the routes used by the participants were high frequency routes, and it was not known a priori what the participants considered to be the "expected wait time". Instead, the headways before and after the participant's departure were recorded. These entered into the equation as differences between the observed headway and the scheduled headway in order to capture on-time performance.

c) For transfer time: The observed transfer time and whether or not a departure was observed while the person was waiting at the transfer stop. Again, that variable was meant to capture denied boardings. Not all transfers could be observed with tracking data, so if the person indicated that they had transferred in the daily survey but no tracking data covering the transfer were available, a binary variable for "unobserved transfer" was entered into the model. A priori, longer transfer times and missed departures were expected to decrease satisfaction. As with wait times, it was not known what participants considered to be their "expected transfer time", and therefore it was difficult to define a benchmark based upon which to calculate deviations.

d) In the daily survey, participants were asked to rate their satisfaction with the accuracy of the real-time information system on a five-point Likert scale. Responding to this question was optional. The response to this question was included in the origin wait time and transfer models where available, and if no response was given, a binary variable was entered to indicate a missing response. See section 4.4 for a further discussion of this issue. A priori, decreased accuracy of real-time information was expected to cause decreases in satisfaction.

Since denied boardings were not reported by the participants, they had to be inferred heuristically with tracking and AVL data. This was done as follows: While a person was waiting in the vicinity of a stop before a transit trip, the number of departures from that stop that also served the person's observed destination stop was counted. The fact that the person didn't board those previous departures might have been due to one of three possible explanations: Either that it was a denied boarding, that the person may have been carrying out an activity close to the stop and was not actually looking to board a transit vehicle yet at the time of the previous departure, or that the first feasible service that arrived made a longer route to get to the same destination and thus the user decided to wait for the next one associated to a lower IVTT. These three cases are very difficult to distinguish based solely on the tracking data and the proximity between the user and the stop calculated from it. It was attempted to reduce the number of false positives (i.e., activities identified as denied boardings) with the following heuristic rules: First, a maximum wait time at the stop of 45 minutes was allowed, and second, a maximum of one missed departure was allowed. Third, a departure was only counted as a missed departure if the scheduled IVTT was equal or lower than the scheduled IVTT of the departure the traveler took. If any of those criteria was not met, it was assumed the person was carrying out an activity or waiting for a departure with a shorter IVTT, and the previous departure was not counted as a denied boarding. In practice, the 45-minute cutoff was virtually never a binding constraint. Nonetheless, this heuristic approach most likely does not eliminate all false positives, and it is still likely that only a subset of the departures observed while the person was waiting was true denied boardings.

5. General mood/SWB: This variable was also captured in the daily surveys where satisfaction ratings were collected. Participants were asked how they generally felt that day, on a five-point Likert scale labeled with a frowny face and a smiley face. While studies

have shown that overall SWB and travel well-being and thus satisfaction are correlated, Ettema et al. (2010) note that the exact relationship has not yet been empirically investigated. As the respondents were asked about their general feeling on the particular day of the survey, it is assumed that the survey generally captured the transient experience of positive or negative affect and moods rather than global judgments (life satisfaction) (Diener, 2000). The causal relationship between daily SWB or mood and customer satisfaction ratings is likely bidirectional: Ettema et al. (2010) summarize previous studies that have argued that travel experiences influence SWB by triggering positive or negative affect or stress, and by facilitating engagement in activities which themselves influence SWB. However, the contribution of travel to overall daily SWB compared to other activities is unclear. On the other hand, there is evidence from marketing research that a customer's mood influences satisfaction with service times (Peterson and Wilson, 1992; Chebat et al., 1995; Durrande-Moreau, 1999).

In the mobile survey, satisfaction and mood/SWB were measured simultaneously, and the surveys were taken mostly in the evening. Therefore, in order to interpret the modeling results presented in section 4.4, an assumption is made. We know that the response to the mood/SWB question includes all activities of that day, including ones that are substantially longer than travel, such as work. If we assume without further proof that the influence of travel experiences on daily subjective well-being is comparatively small with respect to all other activities, then we can assume a unidirectional causality, i.e., that SWB affects travel satisfaction but not vice-versa. The model results are interpreted with this assumption in mind, but in future iterations of this type of study, efforts will be made to measure mood/SWB independently of travel satisfaction in order to eliminate the need for this assumption.

4.4 Estimation results

Model estimation

Table 4.4 presents the estimation results for the four models. The models were specified as mixed logit models, i.e., with an additional mixing coefficient σ to account for correlation between the responses given by a single individual. σ is also shown in table 4.2, as well as the adjusted $\bar{\rho}^2$ as a goodness of fit measure. The latter is defined as follows:

$$\bar{\rho}^2 = 1 - \frac{L_f - k}{L_0} \tag{4.2}$$

Where L_0 is the log likelihood when all parameters are zero, L_f is the log likelihood for the final value of the parameters, and k is the number of parameters. As noted in section 4.3, the model assumes that the satisfaction is linear with respect to the variables; in other words, a proportional effect of the variables such as delays on satisfaction is assumed. Positive coefficients indicate that a variable contributes to increased satisfaction. There were fewer observations for origin wait times and transfer times than for IVTT since not every trip included a transfer and not all origin waits could be observed with the tracking data. Generally, there are three cases in which a wait time was considered to be "not observed":

	IVTT	Wait time	Transfer time	Reliability
Total Observations	2403	779	188	741
Observed transfers				38
Reported transfers				212
Unique respondents	529	384	110	373
Avg obs/respondent	4.5	2	1.7	2
Min obs/respondent	1	1	1	1
Max obs/respondent	18	8	7	8
Average value [min.]	11	3	7	
Minimum value [min.]	1	0	0	
Maximum value [min.]	57	14	42	

Table 4.3: Overview of data used for model estimation.

1. If there was insufficient location tracking data available for that portion of the trip. This includes both smartphone and vehicle location data.

2. if the participant was carrying out an activity near the stop (e.g. work) which made it impossible to distinguish activity time from wait time.

3. If the wait time was incurred when the participant transferred from BART (regional rapid transit) to a local metro train inside an underground metro station.

Since BART was not part of the study, the last case was considered an origin wait rather than a transfer, but it was not observable due to it taking place entirely underground and no real-time data being available for BART. Overall, an observed origin wait time could be derived for only 35% of trips. Due to the density of San Francisco and its transit network, the second reason was often the relevant reason for non-identification of wait times. It is not known whether the data contain a systematic bias due to this, and the results should be seen in that light. In future work, it would be recommended to have participants actively report on waiting times rather than to rely entirely on location tracking data. Furthermore, it should be noted that the data used for estimating the models contained variable numbers of observations per respondent, as there was no upper limit to the number of days on which a participants could take the mobile survey. An overview of the data sets is given in table 4.3. For the reliability data set, 212 trips with “unobserved transfers” were included, i.e., trips where the participant filled out the “satisfaction with transfer times” question in the daily mobile survey but no transfer was detected from location tracking data.

The following two subsections discuss the estimation results: First, the baseline satisfaction, sociodemographic, mode access and subjective well-being variables are discussed, followed by the trip-specific variables such as IVTT, transfer and wait times, reliability and the accuracy of real-time information. Finally, the sensitivity analysis shows in graphical form what percentage of participants were satisfied as a function of travel times.

	IVTT		Origin wait		Transfer		Overall reliability	
	$\hat{\beta}$	$p > \|z\|$	$\hat{\beta}$	$p > \|z\|$	$\hat{\beta}$	$p > \|z\|$	$\hat{\beta}$	$p > \|z\|$
Intercept 1	-0.593	0.14	1.280	0.02	0.920	0.33	0.717	0.21
Intercept 2	1.037	0.00	2.930	0.00	2.300	0.00	2.277	0.00
Intercept 3	2.617	0.00	4.110	0.00	3.820	0.00	3.807	0.00
Intercept 4	4.697	0.00	5.850	0.00	5.500	0.00	6.277	0.00
pre_sat_ivtt	0.513	0.00						
pre_sat_wait			0.262	0.00				
pre_sat_transfer					0.266	0.24		
no_pre_sat_transf					0.826	0.26		
pre_sat_reliability							0.461	0.00
age	0.021	0.00	0.000	0.98	0.045	0.05	-0.007	0.46
female	0.114	0.38	0.103	0.55	0.364	0.29	-0.117	0.56
income	-0.006	0.76	-0.012	0.64	-0.077	0.15	-0.043	0.13
unknown_income	-0.019	0.94	-0.578	0.04	-0.858	0.08	-0.724	0.04
employed	-0.188	0.24	0.186	0.34	-0.101	0.79	0.517	0.01
longuser	-0.210	0.16	0.308	0.09	-0.022	0.96	0.015	0.94
beforestudycar	0.396	0.02	0.499	0.03	0.125	0.77	0.230	0.32
bike_owner	-0.197	0.13	-0.136	0.43	0.184	0.57	0.056	0.76
personal_car	-0.013	0.95	0.019	0.94	0.399	0.39	0.071	0.77
shared_car	-0.177	0.37	-0.376	0.11	0.568	0.36	-0.213	0.44
general_feeling	0.532	0.00	0.363	0.00	0.376	0.07	0.484	0.00
sched_ivtt	-0.015	0.04			0.023	0.19	-0.019	0.07
ivtt_early	-0.036	0.59					-0.118	0.21
ivtt_delay_bus	-0.171	0.00					-0.054	0.31
ivtt_delay_metro	-0.283	0.00					-0.089	0.12
post_headway_diff			-0.017	0.09			-0.002	0.77
pre_headway_diff			-0.011	0.26			-0.005	0.64
origin_wait_time			-0.048	0.12			0.007	0.82
dep_during_wait			-0.446	0.08			0.067	0.80
transfertime					-0.065	0.00	-0.014	0.72
dep_during_transf					-0.19	0.33		
unobserved_transf							-0.282	0.14
rti_accuracy			0.881	0.00	0.384	0.04	0.933	0.00
no_rti_response			3.050	0.00	0.708	0.31	2.990	0.00
weekend	0.088	0.35	0.102	0.55	0.362	0.40	0.156	0.34

Continued on next page

	IVTT		Origin wait		Transfer		Overall reliability	
	$\hat{\beta}$	$p > \|z\|$	$\hat{\beta}$	$p > \|z\|$	$\hat{\beta}$	$p > \|z\|$	$\hat{\beta}$	$p > \|z\|$
σ	1.040	0.00	0.662	0.02	0.453	0.26	0.872	0.00
Adjusted $\bar{\rho}^2$	0.358		0.417		0.366		0.484	

Table 4.4: Estimation results for the ordinal logit models of traveler satisfaction.

Baseline satisfaction, sociodemographic, mode access and subjective well-being variables

When interpreting the results, one needs to be cognizant of two factors: First, the sample sizes vary for the four models, which affects the calculated p values. Second, compared to the age and income distribution of the overall SFMTA ridership, younger riders are overrepresented in the study population whereas low-income households are underrepresented.

The respondents' baseline satisfaction from the entry survey is significant at a level of $p \leq 0.1$ in three of the four models. In the fourth model, satisfaction with transfer times, the effect is still positive, though 20 out of the 188 observations did not have a reported entry satisfaction with transfer times. This strong influence of the baseline satisfaction might be an indication of a "hedonic treadmill" effect (Brickman and Campbell, 1971; Lucas, 2007), where individuals may experience temporary increases or decreases in affect due to daily events but revert to their baseline satisfaction levels in the long run. The person's general mood or feeling on a given day is significant at $p \leq 0.1$ and positively associated with satisfaction with all four components of travel times. As previously discussed, the directionality of the effect cannot be conclusively established with these data since measurements were taken simultaneously. If the assumption is made that a person's daily mood is primarily shaped by events other than their daily transit travel, we can interpret these results as saying that the better a person feels on a given day, the more satisfied they tend to be with transportation services. However, this simplification may be too strong since it has been shown in previous literature that commute stress and satisfaction with travel affects overall mood (Bergstad et al., 2011).

The effects of gender and income are not significant at $p \leq 0.1$. This is surprising, as these two variables are typically considered to have moderating effects on satisfaction with service quality (Anderson, Pearo, and Widener, 2008), and income is typically related to the value of travel times. Nothing is known about those who did not indicate their income, other than that they were, on average, less satisfied with wait times, transfer times, and reliability. No data were available on trip purposes, but in future research, it might be possible to impute the purpose from location tracking data to get a more accurate picture. Lastly, one can see that passengers who have regularly used Muni for at least 2 years on average show lower satisfaction with IVTT and higher satisfaction with wait times. The effect of that variable on satisfaction with reliability and transfer times is insignificant at $p \leq 0.1$. Transit use can potentially be thought of as a habitual behavior for long-term users, and as Aarts, Verplanken, and Van Knippenberg (1997) found, habit (and thus, length of use) was inversely correlated with the range of information used by individuals in judging their travel modes. Our parameter estimates may in part reflect this adaptation process, which future research should aim to capture more specifically.

Several variables regarding mode access were included. Interestingly, regularly having used a car in San Francisco before the study has a significant and positive effect on satisfaction with IVTT. Also, none of the bike ownership and auto access variables are significant at $p \leq 0.1$, and the estimated effects are both positive and negative. One would typically expect auto access and use of the automobile to be correlated with lower satisfaction with public transportation services (Wallin Andreassen, 1995; Beirão and Cabral, 2007). Our results may reflect the fact that much of the observed sample consists of choice riders. These riders have presumably already decided which trips to use Muni for, especially if they have a bike or personal car available at all times. Therefore, the trips for which they responded to the daily mobile surveys are likely to be trips for which they already chose to use Muni rather than the car (or bike), and where they are more likely to be more satisfied with the service. Furthermore, car users may be aware that driving in San Francisco - especially during peak hours - can be a stressful experience. However, these are only hypotheses, and their accuracy would have to be verified in future work.

Trip-specific variables

In-vehicle travel time

The question prompt specified only satisfaction with IVTT and did not mention IVTT reliability explicitly, but it appears that participants generally responded with respect to both. First of all, it can be seen that the coefficient for scheduled IVTT is significant at a level of $p \leq 0.05$, showing that in general, satisfaction with IVTT decreases with longer trips, even if the vehicle is on time. It can also be seen that the coefficients for differences between scheduled and observed IVTT differ in significance: If the observed IVTT was less than the scheduled IVTT, the coefficient is not significant ($p = 0.59$), implying that experiencing less IVTT than scheduled does not markedly change satisfaction. On the other hand, if the observed IVTT exceeds the scheduled IVTT, the coefficients for both modes are significant at $p \leq 0.05$ and several times larger than the coefficient for scheduled IVTT. In other words, riders appear to be generally accepting of scheduled IVTT (though longer travel times cause some dissatisfaction per se), whereas in-vehicle delays are the major source of dissatisfaction with on-board travel times. The results also indicate that in-vehicle delays on metro trains were weighted more heavily than in-vehicle delays on buses.

Origin wait time

The estimated wait time coefficient is negative but not significant at $p \leq 0.10$. On the other hand, the coefficient for denied boardings is significant and implies that a denied boarding causes as much dissatisfaction as 9.3 minutes of pure wait time. This number might in fact represent a lower bound due to the difficulties in identifying denied boardings, as explained in section 4.3. Furthermore, there might be a difference in resulting dissatisfaction depending on whether the participant “missed” a departure of his/her own volition, perhaps due to crowding, or whether the driver did not stop or refused to let the participant board. In total, the data set contained 91 missed boardings out of a total of 779 observations.

The coefficient for the headway following the observed departure is larger than the coefficient for the headway preceding the observed departure, and the latter is also insignificant ($p = 0.26$). This might be an indication that, given a desired arrival time and subtracting on-board travel time to obtain an "ideal" departure time, a passenger chooses the first departure before that time rather than afterwards. The larger the headway that time falls into, the more the observed departure time deviates from the ideal departure time. This information is contained in the headway following the observed departure. In general, the satisfaction results for the headway preceding the departure might also be influenced by bus bunching, because the customer would most likely be evaluating the service with respect to the headway preceding the first bus of the bunched platoon, but if the passenger boarded one of the following buses, a small headway would have been identified in the data. While this hypothesis is difficult to investigate with the current data set, the role of observed headways clearly merits further investigation.

The influence of real-time information

We first consider the influence of real-time information with respect to origin wait time satisfaction. One question on the mobile survey asked participants to rate the accuracy of the real-time arrival prediction system. Responding to that question was optional because a traveler may not have used real-time information on a given trip. An optional check box was provided to indicate that the person had not used the real-time information system. Unfortunately, it only became clear later that the question had unintentionally been formulated in a potentially problematic manner by using the brand name of the real-time arrival prediction system, NextBus. The exact question was: "If you used NextBus, how accurate was it?" From emails received from participants, we realized that many participants were not familiar with that brand name. One reason was that numerous participants used third-party smartphone apps that query the NextBus application programming interface but present the real-time arrival information under the app's name. Even if users of those apps knew of NextBus, many were not aware that the information in their third-party app was actually provided by NextBus. Furthermore, it appears that some participants considered the question to pertain only to the NextBus website, but not to the dot-matrix arrival time displays installed at stops. In summary, for missing responses and responses where the check box was checked, it is not clear whether that person did, in fact, not consult real-time information at all or whether the non-response is due to one of these misunderstandings. Therefore, the pertinent variable is coded as non-response.

The estimated coefficient for the accuracy of real-time information is positive and significant at $p \leq 0.05$. In other words, the higher the perceived accuracy of the real-time arrival predictions, the more satisfied customers were with the wait time. The non-response variable is also positive. As explained above, it is not clear what this variable is capturing, and this requires further investigation.

Transfer time

While satisfaction with transfer time is primarily driven by the experienced transfer time itself, it can also be seen that the IVTT plays a role. The positive coefficient for that variable indicates that, all else being equal, passengers appear to be willing to endure longer transfer times on trips with a longer overall IVTT. As with origin wait time, a binary variable indicated whether a departure of the connecting route was observed while the user was waiting at the transfer stop. The estimate for this coefficient is positive and insignificant ($p = 0.23$), which most likely reflects the difficulty of distinguishing true wait time at the transfer stop from activities carried out at that location, as explained in section 4.3. Lastly, the accuracy of real-time information is again found to have a strong positive influence on transfer time satisfaction. Due to the limited sample size, it was not possible to specify the model of transfer time satisfaction with a piecewise linear transfer time term, but future research should aim to do so. Previous results by Rietveld, Bruinsma, and Vuuren (2001) showed that increased transfer times can increase overall trip chain reliability, and it would need to be seen whether this is reflected in passenger satisfaction ratings.

Overall reliability

Aside from the individual components of travel time, respondents were asked to rate the overall reliability of their trip. The importance of this attribute has been noted by a number of authors (e.g. Edvardsson, 1998; König and Axhausen, 2002; Hensher, Stopher, and Bullock, 2003a). Our goal was to decompose an aggregate measure of satisfaction with reliability into the individual contributing factors. This allowed us to compare all components of observed travel time in one equation and to investigate passenger perception of unreliability. This is a first attempt at shedding light onto this question, but given the complexity of the concept of reliability, more research will be needed in the future to fully understand the contribution of various travel time components. It should be pointed out that in this data set, the number of trips with observed transfers was very small (only 38). Therefore, the coefficient estimate related to the transfer time is subject to more uncertainty than the remaining coefficients. All times were expressed in minutes. As can be seen, the estimated coefficients for this model are generally in line with those of the single models, with some exceptions. For instance, the coefficient for delays experienced on board rail vehicles is larger than the coefficient for delays on buses, but the coefficient for deviations from the schedule where the observed IVTT was less than the scheduled IVTT is also negative and larger than both.

The coefficient for scheduled IVTT, though of lesser significance, is much smaller than the coefficients of in-vehicle delays and also negative, suggesting that longer trips are inherently more at risk of unreliability in travel times. As previously mentioned, IVTT is the only travel time component that could be benchmarked against the timetable, and the assumption that passengers' expectations are approximately in line with the scheduled travel time is plausible. For wait times and transfer times, there are several different possible reference points, but it is unknown what the passengers' expectations are. In this combined model, both coefficients for wait time and transfer time are insignificant, as is the coefficient for missed departures at the origin stop; this is not consistent with the other models and needs further investigation.

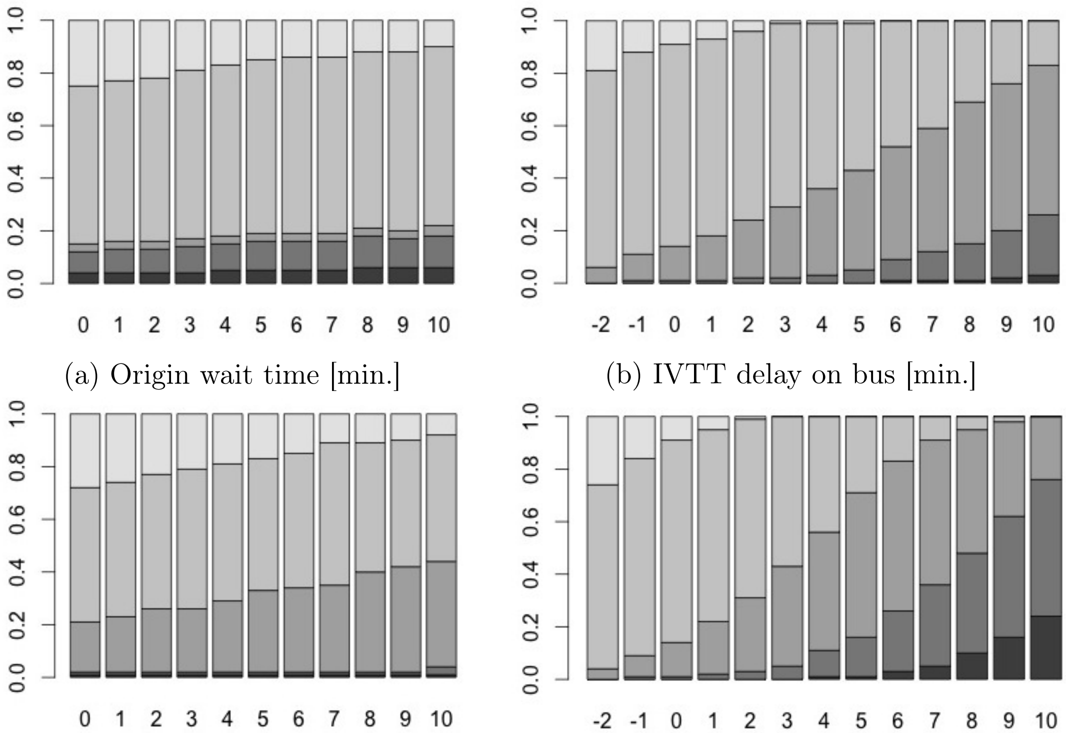
On the other hand, the influence of real-time information is consistent with the findings in the previous models, as are the results for the pre- and post-departure headways.

The fact that the combined reliability model is not fully consistent with the individual models does not invalidate the findings of the individual models. The question prompts for the individual travel time component satisfactions were different from the reliability satisfaction, and the fact that many coefficients are insignificant in the reliability model rather points to the possibility that the data and the model are not fully capturing how the participants perceived reliability. More research will be required to elucidate why this is occurring and to understand what other factors might have been missed.

Sensitivity analysis

To further understand the model results, a sensitivity analysis was conducted. The question was: Everything else being equal, what is the distribution of responses on the five-point Likert scale of satisfaction as a function of travel times? For every participant in the pertinent data set, the IVTT difference, wait times at the origin stop and transfer times were substituted with the simulation values in the respective models. The simulated range was from -2 to 10 minutes in the case of IVTT difference and from zero to 10 minutes in the case of origin wait time and transfer time. All other variables remained unchanged. The model of satisfaction with overall reliability was not included in the sensitivity analysis since it includes all three aforementioned variables and therefore has more than one degree of freedom. This would have required two of the variables to be fixed at arbitrary values in a figure. Figure 4.1 shows the simulated distribution of satisfaction among respondents. As previously found, the strong sensitivity of passengers to in-vehicle delays can be seen, and as expected, delays on metro trains cause more dissatisfaction than delays on buses. A 10-minute in-vehicle delay on bus results in a 69 percentage point decrease of satisfied passengers (defined as the top two categories on the Likert scale) with respect to the zero-minute IVTT difference, and after a 10-minute delay on metro, no more responses in the "satisfied" or "very satisfied" categories are recorded.

A different picture emerges for the origin wait time and the transfer time. In the case of the origin wait time, the main shift that occurs as wait times increase is away from the "very satisfied" category. Between a 0-minute wait and a 10-minute wait, there is a 15 percentage point decrease of "very satisfied" customers and a corresponding 8 percentage point increase of "satisfied" customers. A further 7 percentage points shift to "neutral", "dissatisfied" or "very dissatisfied" customers, which increase from 15 percent at zero minutes wait time to 22 percent at 10 minutes wait time. It is noteworthy that even at zero minutes wait time, 12 percent of customers are still found to be "dissatisfied" or "very dissatisfied". This could potentially be due to a general dissatisfaction with wait times that is expressed by these customers or to the difficulties in identifying wait times as previously explained. On the other hand, the share of customers that are "dissatisfied" or "very dissatisfied" with transfer times is consistently very small (between 2% and 4%) in the range between zero and 10 minutes of transfer time. Here, the shift occurs mainly from the "very satisfied" category, which sees a 21 percentage point decrease between zero and 10 minutes, to the "satisfied" and "neutral" categories.



(a) Origin wait time [min.]

(b) IVTT delay on bus [min.]

(c) Transfer time [min.]

(d) IVTT delay on metro [min.]

Figure 4.1: Satisfaction vs. travel times, from very dissatisfied (darkest) to very satisfied (lightest).

The ordinal logit model assumes that the satisfaction is linear with respect to the individual travel time components. In future research, this assumption could be relaxed in order to uncover potential nonlinearities. For example, it may be possible that only wait times and transfer times above a critical value cause dissatisfaction. However, estimating such a model would require a data set that includes more observations of longer delays and wait times than the present data set. In this data set, the median observed wait time at the origin stop was 2 minutes, and the median observed transfer time was just below 3 minutes. This is significantly below the mean wait times observed with passengers in a similar study, albeit in a lower-density setting (Brakewood, Barbeau, and Watkins, 2014). Due to the overall short wait times, our ability to model passenger responses to longer delay, wait or transfer times was limited.

4.5 Discussion

These results have been able to demonstrate the power of combining personal-level tracking data, which is increasingly available to researchers thanks to the proliferation of location-aware smartphones, with survey data and detailed operational data. Thus, the analyst can connect the survey data to specific, individual-level experiences which can be observed and quantified in an objective manner in order to understand how passengers perceive and conceptualize travel times and unreliability. In this chapter, satisfaction with respect to various travel time components of transit travel was investigated, and the models estimated from the data show a clear relationship between the respondents' reported satisfaction and the various travel time components observed from tracking data. Furthermore, the results illustrate the importance of several other variables that are unrelated to the experience itself, such as the participant's age or how long the participant has been using the system for. It must be noted that since the study predominantly involved participants who were already transit riders, the results only extend to existing transit riders. Nonetheless, the results demonstrate the potential of this general approach. In the future, non-riders could be recruited into a study in which they would try transit for a certain period of time and record their experiences in the same way as was done for this study. That would allow the researchers to observe their learning process and to gain an understanding of how their assessments of service differ from habitual transit riders.

The following list presents some important implications of the results of this study:

- The results of the IVTT satisfaction model give a strong indication that the disutility of scheduled IVTT is much lower than the disutility of in-vehicle delays. To the best of our knowledge, transit planning almost universally still relies on research findings that out-of-vehicle wait time is generally perceived as more onerous than IVTT; summaries can be found in Iseki, Taylor, and Miller (2006) and Wardman (2004). While this may hold true if IVTT is as scheduled, our model results suggest that in the future, trade-offs should be considered separately between out-of-vehicle travel times, scheduled IVTT and in-vehicle delays. While more data would be needed to definitively establish these trade-offs, it is possible that in-vehicle delays may be perceived as worse than out-of-vehicle delays. As summarized by Bates et al. (2001a) with respect to public transportation, two of the most common frameworks for modeling passenger decisions when faced with travel time uncertainty are the mean-variance approach, in which the travel time of a certain choice depends on its mean travel time and the standard deviation or variance (e.g. Jackson and Jucker, 1982; Noland and Polak, 2002), and the scheduling approach (Noland and Small, 1995; Small, 1982), which focuses primarily on the decision-maker's desired arrival time and considers the deviation therefrom. Since we did not collect information on the desired arrival times, it is unfortunately not possible to test these two model specifications, but our results are in line with two main tenets of these models. The scheduling approach assumes different disutilities for early and late arrivals at the destination, which we see evidenced in our results. The mean-variance approach decomposes travel time into an expected value and deviations from it, and assigns different disutilities to the two terms. This can be seen in our results as well.

  Our results further suggest two possible refinements to either the the mean-variance or the scheduling approach. First, the travel time deviation should be further decomposed by where it was incurred, i.e., into wait time variance, IVTT variance, etc. In order to calculate deviations of the wait times and transfer times, data would be required that captures the participant's expectation of those travel time components. For that purpose, participants would need to be asked via the surveys on their smartphones what their expectations of travel times were prior to beginning their trip. Second, it was observed that the disutility of travel time variance differed by mode of transportation, with delays on metro causing more dissatisfaction than comparable delays on buses. There are two possible causes for this: On one hand, traffic is a major factor of bus delays, and travelers may be less dissatisfied if they see the source of delay (Carrel, Halvorsen, and Walker, 2013) or consider it to be beyond the control of the agency. Secondly, experiencing an in-vehicle delay on a bus may be a less unpleasant experience since a passenger might technically be able to get off the idling bus at any time by asking the driver. This is more difficult on board a metro train, and it is impossible when stuck inside a tunnel, causing passengers to feel more "trapped".

If in-vehicle delays were indeed perceived as more onerous than out-of-vehicle delays, this would have policy implications: Projects to reduce IVTT variations would need to be considered at least as important as projects to reduce overall IVTT. This would also put a question mark behind operations control strategies that rely on holding trains and buses, for example to regularize headways. If vehicles are held while passengers are on board, the consequences of dissatisfaction among those passengers should carefully be weighed against the potential benefits. The analysis of satisfaction with IVTT delay could be extended to identify where and how the delays occurred (e.g. due to congestion or delays at stops), which would allow an even more nuanced picture of the type of delays that drive customer dissatisfaction.

- The model also shows that the *accuracy* of real-time information is positively associated with satisfaction with wait times and transfer times. This result is plausible for two reasons: First, if the accuracy of the arrival time prediction was poor (e.g., it showed a ghost bus), then the passenger encountered a different wait time than initially expected. If the actual wait time exceeded the expected wait time based on real-time information, dissatisfaction arises. This result is in line with a large body of prior marketing research, where this experience is termed "disconfirmation" (Oliver, 2010). Furthermore, if the passenger had known the true wait time in advance, he or she may have made a different travel choice. Second, as was discovered by Watkins et al. (2011), real-time information eliminates the discrepancy between experienced wait time and the perception of wait time by the passenger. However, previous research, such as that by Watkins et al. (2011), typically assumed that the real-time information provided was correct. Our results indicate that the accuracy of the real-time information provided should be explicitly considered; if it is noticeably wrong, the passenger's perception of wait time is most likely larger than the actual wait time, as was observed by Mishalani, McCord, and Wirtz (2006), which in turn decreases satisfaction. This underscores the need to continuously invest in improving and upgrading real-time arrival prediction systems.

- This study collected both a long-term baseline satisfaction and a variable daily satisfaction rating from respondents. The model results show the importance of both in understanding how satisfied or dissatisfied customers were with service on a specific day. This may be a manifestation of what is generally known as the "set point theory" of subjective well-being (Brickman and Campbell, 1971; Lucas, 2007), which is sometimes also known as the "hedonic treadmill". How the baseline satisfaction is formed was not revealed by this research, but some indications are given in the marketing literature, which points to three components: Personal use experience, peer opinion and marketing. Furthermore, it can be assumed that some respondents display higher levels of subjective well-being in general and may therefore be more satisfied with the status quo. Based on the experiences with the SFTQS, it is recommended in future research to clearly specify the time frame covered by a satisfaction survey and to collect a baseline satisfaction rating if repeat measurements of trip- or day-level satisfaction are taken. If satisfaction surveys are conducted covering a longer time span (e.g., service over a month or even a year), the researcher should be cognizant of potential memory distortions such as the peak-end rule that might affect responses (Fredrickson and Kahneman, 1993). Therefore, it is suggested to conduct a repeated daily survey over a longer period of time. As demonstrated, this can provide very valuable insights and tangible results to the agency, especially when connected to other data sources, and can help provide a picture that is less affected by biases from memory distortions.

- This survey also collected respondents' general feeling on the day of the survey. This follows previous research that showed a relationship between subjective well-being and travel experiences. Examples include Abou-Zeid et al. (2012) and Friman et al. (2013), as well as a large body of commute stress literature (e.g. Wener et al., 2003; Gatersleben and Uzzell, 2007). In line with those previous findings, the modeling results show, at the very least, a strong association between daily mood and satisfaction with transit services, which has also been observed by Susilo and Cats (2014). As pointed out previously, the direction of causality is subject to debate, and therefore more research is needed in this domain. However, regardless of effects in the opposite direction, it is very likely that mood and subjective well-being on a particular day influence satisfaction ratings (Peterson and Wilson, 1992). Therefore, when conducting surveys, analysts should be cognizant of factors that might influence the general mood of respondents, such as the weather or results of a local sports team. The results in this chapter strengthen the case for conducting repeated satisfaction surveys over an extended period of time in order to better control for this variable. Furthermore, to connect this point with the previous one, it may be advisable to collect baseline information on participants' overall subjective well-being up front as well.

One should of course bear in mind that these results are based on the study population at hand, and that there may be differences between the study population and the overall ridership of the SFMTA. A more detailed comparison of age and income distributions can be found in chapter 2.

There are several additional possible avenues for future research. First of all, there is a need to verify and corroborate the findings of this chapter with larger sample sizes, especially the results of the satisfaction with overall reliability model. This is especially true with respect to origin wait times and headways at the departure stop since it was found in this chapter that it was difficult to capture the impact of those variables with the data at hand. Second, the travel time and wait time variables could be redefined as piece-wise linear in order to capture possible nonlinearities. In future data collection efforts, the SWB should be measured separately from the satisfaction ratings in the mobile survey, as discussed in section 4.3. It could also be useful to collect information on whether the respondents were carrying out some activity while on board (such as checking emails on their phone or reading a book) since that might affect satisfaction with travel times.

In the future, more detailed model results that build on and extend the findings of this chapter may support the direct calculation of the impact of reliability improvements on customer satisfaction. The necessity for this type of research is echoed by Bordagaray et al. (2014), who found that even while accounting for user heterogeneity, reliability remained one of the key elements determining customer satisfaction. With the proper models, it would be possible to calculate the effect of investments that reduce in-vehicle delays, such as bus lanes or improved train signaling systems, on passenger satisfaction, which in turn could be used to build business cases for those investments. In summary, the results of this study clearly reveal new and more nuanced perspectives on the contributors to passengers' satisfactions with their specific experiences not seen at this level of resolution previously, and point to important investigations, with potentially marked implications on service planning, design, and operations.

Chapter 5

Understanding future mode choice

5.1 Introduction and motivation

Public transit is a key element to efficient and sustainable urban transportation, and in the past decades, numerous public policies have been designed to increase its mode share in urban areas through subsidies, service expansions, and land-use zoning. Yet, as is noted by Perk, Flynn, and Volinski (2008), US transit agencies continue to see high levels of ridership turnover; in many cases, a steady influx of new users into the system is offset by similarly high rates of transit use cessation. On the individual level, these shifts are not trivial: As is explained by Vij, Carrel, and Walker (2013), travelers tend to build their lifestyle around the use of certain travel modes, and decisions between, for instance, an auto-oriented lifestyle and a transit-oriented lifestyle are relatively stable. In other words, users who quit using a transit system are relatively unlikely to return unless a major upgrade to the transit system is made.

The fact that little is known about the reasons for which people shift away from transit-oriented lifestyles is primarily due to a lack of suitable data. Authors have generally identified changes in lifestyles associated with events such as marriage, or having children, as causes of transit use cessation. However, Carrel, Halvorsen, and Walker (2013) identified a variety of negative experiences with service quality (e.g., delays, high crowding levels) as further potential drivers. This chapter aims to quantify the effects of negative experiences on transit users' future intentions of transit use. It is based on the SFTQS data set and uses a latent variable choice model to understand the link between their individual experiences, satisfaction and future intentions.

5.2 Literature review

There are very few publications that have investigated transit passenger loyalty. The only two to do so in a generalizable fashion by using smart card data are by Trépanier, Habib, and Morency (2012) and Ma et al. (2013), but they did not link the observed usage patterns to riders' experiences with the service. This chapter is concerned with making that link, and with describing the influence of individual experiences on transit rider satisfaction and on future behavior. The framework laid out in this chapter has several components:

1. The link between individual experiences with transit service quality and reported levels of satisfaction on a daily level.
2. The link between daily satisfaction and overall satisfaction reported at the end of the study.
3. The link between exit survey satisfaction and future behavior.

The first item was the subject of chapter 4 and will be expanded on in this chapter. The pertinent literature is presented in more depth there and is only summarized in this paragraph. Satisfaction surveys are most valuable if a link can be made between satisfaction and objective service quality measures (Davis and Heineke, 1998). So far, this has not been done in transit satisfaction literature with the exception of work by Friman and Fellesson (2009) and the work in chapter 4, and only in the latter was the link made on an individual rather than an aggregate level. In fact, customer satisfaction is a function of *personal use experience* (Woodruff, Cadotte, and Jenkins, 1983; Anderson and Sullivan, 1993), and in transportation, it is generally an aggregate of multiple repeated experiences over time. To control for memory distortions, the analyst needs to be knowledgeable of the subject's usage history and needs to limit the time frame covered by the satisfaction survey (Fredrickson and Kahneman, 1993; Kahneman et al., 2004). In chapter 4, several separate ordinal logit models were estimated, linking satisfaction with individual travel time components to observed travel times. It was found that the disutility of scheduled in-vehicle travel time was much lower than in-vehicle delay time, and that in-vehicle delays appear to be strong drivers of passenger dissatisfaction. Under certain circumstances, the latter might be perceived as more onerous than out-of-vehicle wait time. Furthermore, it was found that the baseline satisfaction with transit services and subjective well-being on the day of the survey were important covariates in the measurement of daily satisfaction.

The second item is the link between daily satisfaction and satisfaction reported at the end of the study. It is recognized in the marketing literature that satisfaction is a dynamic phenomenon and can change over time (Mittal, Kumar, and Tsiros, 1999), and that this change is a function of personal experiences a decision-maker has made with the service or product in question (Anderson and Sullivan, 1993; Davis and Heineke, 1998). This is consistent with the findings of Kahneman et al. (1993), who found that subjects' ratings of a repeated experience were dependent on their history of previous experiences. Bates et al. (2001b) extend this finding to the transportation realm and argue that personal experience is very important in the context of travel time variability, and that travelers will not choose a route based on average travel times on that route, but rather on travel times they have experienced in the past - i.e., their personal travel time distribution. That distribution is updated every time a person makes a trip, so a person's satisfaction reported at the end of a study should be a combination of their satisfaction at the beginning of the study and their satisfaction with experiences during the study. Abou-Zeid et al. (2012) based an experimental design aimed at capturing travelers' subjective well-being before and after a week of transit use on this notion, though satisfaction with daily experiences was not separately measured. Furthermore, Abou-Zeid et al. (2012) argues that travelers may be less cognizant of their well-being and satisfaction when they carry out routine behavior, such as choosing a car for commuting. This is consistent with findings by Lanken et al. (1994). However, we postulate that this effect might be less pronounced in the case of public transportation use where service quality is variable; the outcome of the traveler's decision is therefore less predictable, and travelers are likely to feel like they have less control over their experience. This is supported by work by Friman, Edvardsson, and Gärling (2001), who found a measurable impact of "critical incidents", i.e., memorable positive or negative experiences, on customer satisfaction with public transportation reported in a post-study survey.

The third item is the link between satisfaction with travel modes and future travel behavior. With specific regard to public transportation, Pedersen, Friman, and Kristensson (2011) have investigated the influence of satisfaction on mode choice with a path analysis (a statistical method to describe directed dependencies among variables) and found a positive association between remembered satisfaction and current choices. Satisfaction with travel modes has also been successfully included in discrete choice models of mode choice (Abou-Zeid and Ben-Akiva, 2010; Friman et al., 2013). In the application reported on in this chapter, future choices were not directly observed, so to link satisfaction with behavior, an alternative set of variables is required. The Theory of Planned Behavior (TPB) is a very useful behavioral framework for this purpose (Ajzen, 1991), where satisfaction is considered part of the set of attitudes toward the behavior in question. Along with norms and beliefs, attitudes are linked to a person's intention to carry out a future behavior, which in turn leads to observed behavior. An extension of the TPB, the model of goal-directed behavior (MGB) (Perugini and Bagozzi, 2001), adds anticipated emotions as well as an additional step before behavioral intentions, namely behavioral desire. Therefore, in accordance the TPB and the MGB, we formulated a set of questions regarding participants' future intended and desired mode choice to substitute for measurement of the actual future behavior. Another example in which the TPB has been applied directly to mode choice is Bamberg, Ajzen, and Schmidt (2003).

5.3 Methodology

The data are again derived from the SFTQS, described in chapter 2. The daily mobile survey and the online entry and exit surveys measured satisfaction with nine variables, each on a five-point Likert scale: Overall reliability, in-vehicle travel time, wait time at the origin stop, transfer time (if applicable), crowding, cleanliness, safety, pleasantness of other passengers, and the accuracy of real-time information. An exploratory factor analysis on the results confirmed that there were strong correlations within two groups of variables: The first four can be summarized as satisfaction with operations whereas the following four can be thought of as satisfaction with the travel environment. During the design phase of the study, several outcome variables were considered in order to measure future mode choice. In accordance with the Theory of Planned Behavior, participants were asked whether they intended to use public transportation more, less or the same in 2014 as they did before the study. In addition, they were asked whether, regardless of their intentions, they would like to use public transportation more, less or the same in 2014. These questions were asked both in the entry and the exit surveys in order to capture changes which may be attributable to the experiences made during the study. However, after the beginning of the study, and based on email exchanges with study participants, it was found that the formulation of those questions might not have been optimal. Some respondents were confused by the difference between behavioral intention and desire, and others stated that they did not know in advance what their mode choices during an entire year would be. Therefore, four additional variables were added in the exit survey: Intended and desired mode choice in January 2014, and two statements as follows: "As soon as my circumstances permit, I would like to use public transportation more" and "As soon as my circumstances permit, I would like to use public transportation less". The intended mode choice in January was asked relative to mode choice before the study. The exact prompt was: "Compared to how often you used Muni in the month before the study, you anticipate using Muni in January...". Responses were on an eight-point Likert scale between "not at all anymore" and "much more". Additionally, participants were presented with a list of possible reasons for their response to the question on desired mode choice in January 2014 and were asked to rate the level of influence of every item on their behavioral desire. A descriptive analysis of the results is provided in section 5.4.

Model development

Chapter 4 described the link between experiences and satisfaction in a static context and considering the individual travel time components separately. This chapter extends that work in several ways and embeds it in a broader modeling framework. The primary purpose is to link personal experiences to future behavior via the intermediate construct of satisfaction. In this case, the overall satisfaction with travel times and operations was of interest, and not the satisfaction with the individual components. Following the exploratory factor analysis described previously, we assume that there are two underlying and unobserved latent satisfaction variables - satisfaction with operations and with the travel environment. The observed variables, i.e., the recorded responses, are indicators of that underlying latent variable. To make the links among satisfaction in the entry survey, experiences during the study, reported daily satisfaction, satisfaction in the exit survey and future mode choice behavior, a latent variable choice model (Walker, 2001) was developed, as shown in figure 5.2. In the figure, ellipses denote latent variables, rectangles denote observed variables, and the arrows show the directionality of effects being modeled.

Before introducing the model, we will first consider an interesting observation with respect to the satisfaction reported in the exit survey. All participants were asked to complete the online exit survey, but in addition to that, an optional mobile exit survey was also distributed. Its formatting was identical to the daily surveys, with the exception that the survey prompts asked respondents to indicate their satisfaction with their overall Muni experience during the study. Participants were asked to fill out the mobile exit survey in addition to the online exit survey, but it was made clear that that was not mandatory. The 5-point response scales in the daily mobile surveys were labeled only at the maximum and minimum with a frowny and a smiley face due to space constraints, where as the scale in the online survey was labeled with words since the online survey engine did not allow graphical labels. A distinct difference was noticed between users' responses with respect to their daily satisfaction and their overall satisfaction in the online and mobile exit surveys. In figure 5.1a, the distribution of responses to the nine satisfaction items in the daily mobile survey are shown. For each item, the red colored bar shows the proportion of "very dissatisfied" responses and the dark green bar shows the proportion of "very satisfied" responses, with the remaining bars showing the responses in between. Figures 5.1b and 5.1c show the distributions for the online exit survey and the mobile exit survey, respectively.

Whereas users were willing to state that they were "very satisfied" with their daily experiences on public transportation, it can be seen that:

- Participants were less willing to state that they were "very satisfied" in the exit surveys
- This effect was more pronounced in the online exit survey than in the mobile exit survey.

The sample size for the mobile exit survey ($N = 353$) was smaller than the online exit survey ($N = 482$) since the former was optional. The sets are not mutually exclusive, since all participants who took the mobile exit survey also took the online exit survey. There are four possible reasons for these discrepancies:

1. The service quality experienced by study participants between the end of the daily survey prompts and the time they filled out the exit surveys was markedly worse than the service quality on the days for which daily surveys were filled out. While possible, this explanation is not plausible.

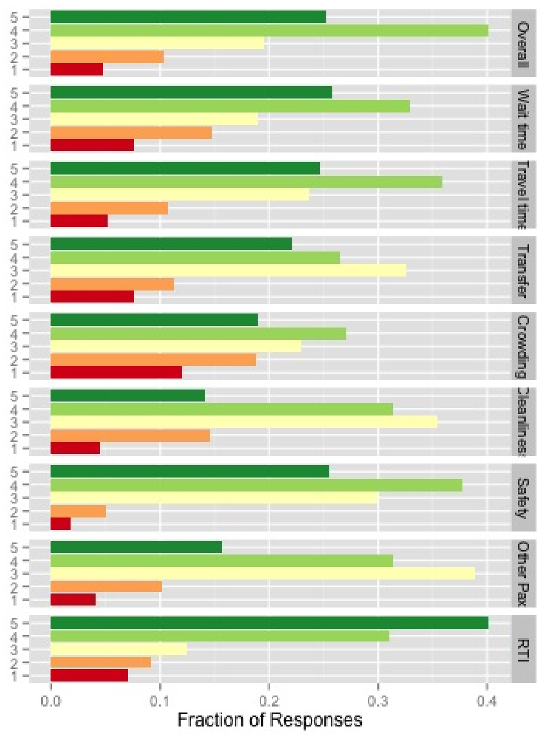
2. The different time frames to which the questions are referring to: when asked about their overall satisfaction, participants may recall negative events that occurred before the study period.

3. The different presentation of the questions, i.e., the fact that the labels differed between the mobile and the online survey versions. While this may account for some of the differences between the online and the daily mobile survey responses, the fact that the exit mobile survey response patterns also differ from the daily mobile survey response patterns indicates that this cannot be the only factor at play.

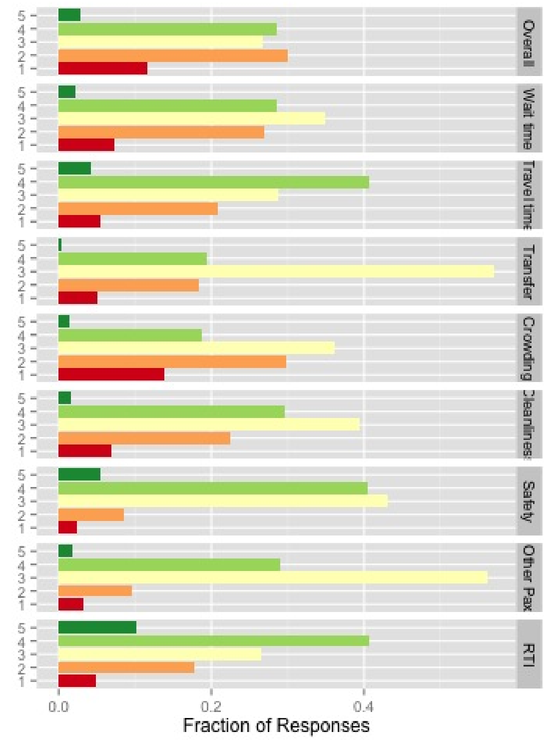
4. The different environments in which the surveys may have been filled out. It is more likely that the online survey was taken by people at home or at the office, whereas the daily mobile survey may have been taken anywhere.

The definitive reason for this discrepancy cannot be elicited without further investigation. Until that is possible, researchers designing future studies should be aware that the medium through which the survey is delivered (in this case, smartphone vs. online survey engine) can have an effect on the response patterns. The fact that the "very satisfied" category in the online exit survey had very few responses would introduce nonlinearity in the measurement model and cause estimation problems since the measurement equations assume a linear relationship between the latent variable and the indicator variables. To avoid these problems, the satisfaction ratings for the online exit survey were re-scaled to a four-point scale where "satisfied" and "very satisfied" were included in one category.

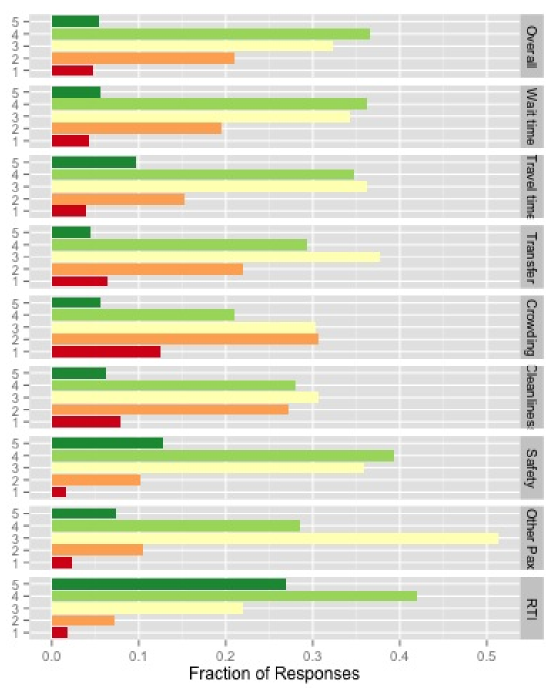
The latent variable model structure is shown in figure 5.2. The latent entry and exit satisfaction are shown as "entry satisfaction with operations" and "exit satisfaction with operations" with the respective indicator variables I_1 through I_6. The indicator variables were the reported satisfactions with the in-vehicle travel time, wait time and overall reliability.



(a) Mobile survey



(b) Online exit survey



(c) Mobile exit survey

Figure 5.1: Distribution of survey responses.

The daily satisfaction was modeled using the same latent variable constructs; those are shown as "d_1 through "d_r. The indicator variables for the daily satisfaction are omitted in the figure due to space limitation, but every daily satisfaction item had four indicator variables which were the aforementioned three satisfaction measurements plus satisfaction with transfer time. Daily satisfaction was measured in five variables: the four most recent responses for which travel time data were available, labeled d_1 through d_4, plus a fifth variable including the average of all remaining daily observations, labeled d_r. This structure was chosen since there were variable numbers of responses per participant. Every participant with at least one daily mobile survey response was included in this data set. The structural model for satisfaction with operations reflects the temporal dependencies between the individual surveys: The daily satisfaction ratings are influenced by satisfaction reported in the entry survey, as well as by travel times experienced on that day and by the user's general feeling on that day. The daily responses in turn feed into the satisfaction reported in the exit survey. The assumption is made that the exit satisfaction depends on the entry satisfaction only by way of the daily satisfaction. All coefficients relating the daily latent satisfaction variables, d_1 through d_4 and d_r, to the exit satisfaction are constrained to be the same. The coefficients of the measurement equations of d_1 through d_4 are also constrained to be the same, but the coefficients for the measurement equation of d_r are allowed to differ to reflect the fact that d_r is averaged over a number of days. In addition to satisfaction with operations, the exit satisfaction with the travel environment is included as a separate latent variable, labeled "exit satisfaction with environment". The indicator variables for the latter are the reported satisfaction with crowding, cleanliness, safety and other passengers in the exit survey. No objective measurements were available for the travel environment. Both latent variables feed into the utility for the choice model.

In addition, several variables on negative critical incidents during the study which were self-reported in the exit survey were included as explanatory variables affecting the exit satisfaction directly. These were:

- The number of times a participant arrived late at work or school ("Late arrival at work").
- The number of times the participant arrived late at a leisure activity ("Late arrival at leisure").
- The number of times a participant reported that he or she wanted to use Muni but was not able to because of a delay on the system ("Could not travel due to delay").
- The number of times a participant was left behind at a stop because the vehicle was full ("Left-behind").

In figure 5.2, the Greek letters denote groups of coefficients corresponding to the notation in table 5.1. Different letters are assigned to different groups of coefficients to improve readability of the model.

As explained at the beginning of this section, the original outcome variables on behavioral intention and desire proved to be problematic; this was unfortunately confirmed by initial data exploration and modeling results. Therefore, three of the alternative variables that were added in the exit survey were used:

- Intended mode choice in January 2014 relative to mode usage before the study, shown in the figure as "change in intentions".
- Desired mode choice in January 2014 relative to mode usage during the study, shown in the figure as "short-range desires".
- Responses to the statement "As soon as my circumstances permit, I would like to use Muni less.". In the figure, this is labeled as "long-range desires".

All outcome variables were reduced to a binary choice between desiring/intending to use Muni less and desiring/intending to use it the same or more. These three variables served as choice indicators (Bollen, 1998). Since only one out of the three outcome variables explicitly references Muni use before the study, a few assumptions must be made:

1. The experiences during the study are representative of the participants' average experiences, and/or
2. the six weeks of the study carry sufficiently large weight compared to previous experiences so that the latter can be disregarded.

We recognize that this is a limitation of the data used, as is further discussed in the following subsection.

In what follows, the model specification is presented. The interpretation of all coefficients used in the model specification is shown in table 5.1. The structural equation for the entry satisfaction with operations was:

$$Sat_{entry,ops} = \gamma_{entrymean\_ops} + \eta_{entry} \cdot \omega_{entry} \tag{5.1}$$

The structural equation for the daily satisfaction with operations (denoted d_i in figure 5.2) was:

$$\begin{aligned} Sat_{daily,ops} = {} & \alpha_{age} \cdot age + \alpha_{income} \cdot income + \alpha_{unknownincome} \cdot unknown\_income + \alpha_{longuser} \\ & \cdot longuser + \alpha_{entrysat} \cdot Sat_{entry,ops} + \alpha_{mood} \cdot mood + \alpha_{ivtt} \cdot ivtt + \alpha_{early} \cdot early \\ & + \alpha_{delay} \cdot delay + \alpha_{waitover5} \cdot waitover5 + \alpha_{waitunder5} \cdot waitunder5 + \alpha_{nowait} \\ & \cdot no\_wait + \alpha_{transfertime} \cdot transfer + \alpha_{unobservedtransfer} \cdot unobserved\_transfer \\ & + \alpha_{notransfer} \cdot no\_transfer + \alpha_{leftbehind} \cdot left\_behind + \eta_{daily} \cdot \omega_{daily} \end{aligned} \tag{5.2}$$

The structural equation for the exit satisfaction with operations was:

$$\begin{aligned} Sat_{exit,ops} = {} & \sum (\beta_{dailysat} \cdot Sat_{daily,ops}) + \beta_{leftbehind\_1\_9} \cdot leftbehind\_1\_9 \\ & + \beta_{leftbehind\_10} \cdot leftbehind\_10 + \beta_{latework} \cdot latework + \beta_{lateleisure} \\ & \cdot lateleisure + \beta_{notravel} \cdot notravel + \eta_{exitops} \cdot \omega_{exitops} + \eta_{errorcorr} \cdot \omega_{errorcorr} \end{aligned} \tag{5.3}$$

The summation term above is the summation over all latent daily satisfaction variables. The structural equation for the exit satisfaction with operations was:

$$Sat_{exit,env} = \eta_{exitenv} \cdot \omega_{exitenv} + \eta_{errorcorr} \cdot \omega_{errorcorr} \tag{5.4}$$

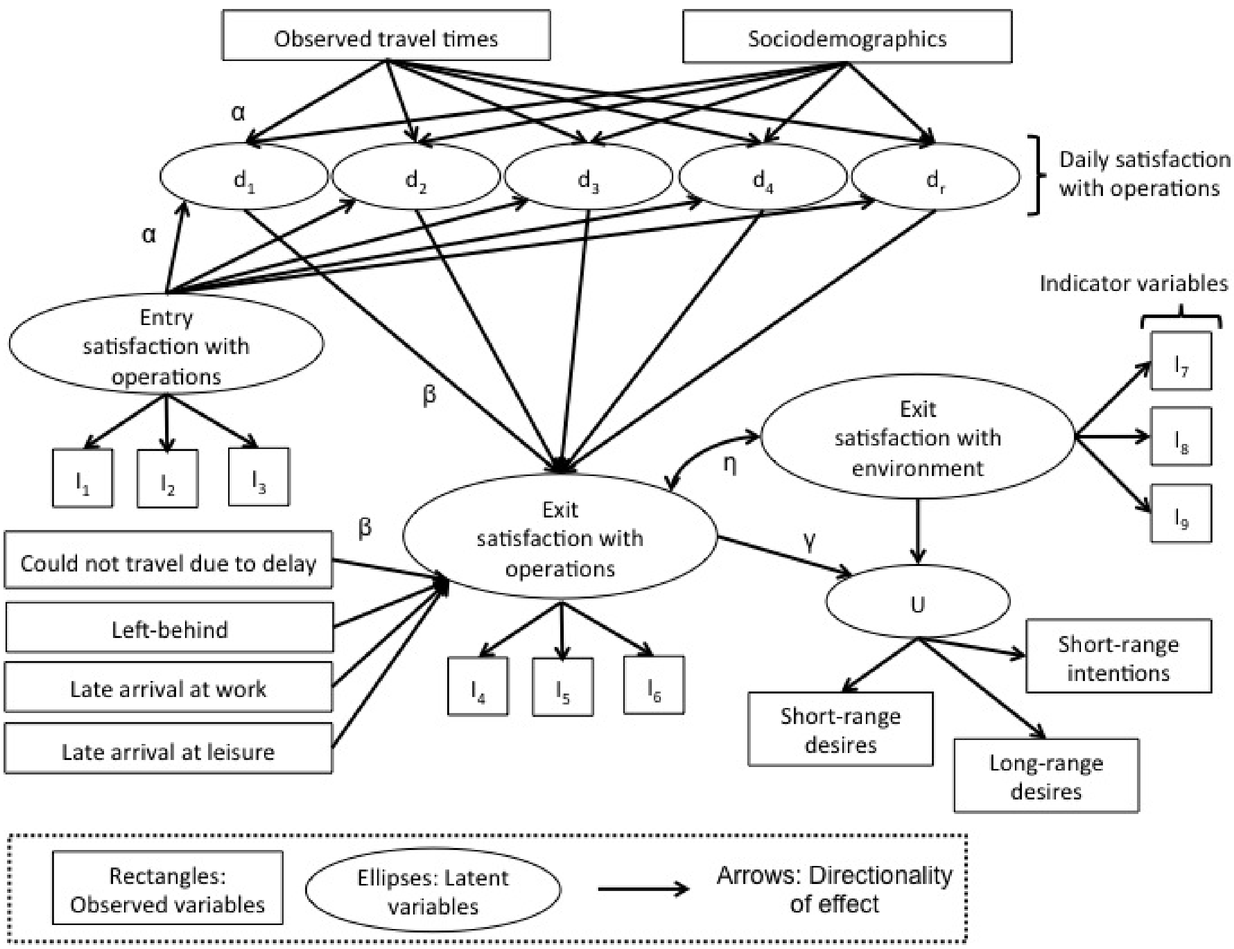


Figure 5.2: Model structure.

And finally, the choice model was:

$$\begin{aligned} V = {} & (\mu_{shortint} + \mu_{shortdes} + \mu_{longdes}) \\ & \cdot (ASC_{shortint} + ASC_{shortdes} + ASC_{longdes} + \gamma_{ops} \cdot Sat_{exit,ops} + \gamma_{env} \cdot Sat_{exit,env}) \end{aligned} \tag{5.5}$$

The μ and ASC terms above are specified such that they only enter into the equation if the choice being modeled relates to the respective outcome variable, and they are zero otherwise. The measurement equations all had the same functional form. For example, the conditional probability for the indicator "satisfaction with IVTT" (IVTTSat_entry) of the entry satisfaction with operations is:

$$\begin{aligned} & P(IVTTSat_{entry} | I_{IVTT}, \delta_{IVTT,entry}, \lambda_{IVTT,entry}) = \frac{1}{\sigma_{IVTT,entry}} \\ & \cdot \Phi\left(\frac{(IVTTSat_{entry} - \delta_{IVTT,entry} - \lambda_{IVTT,entry} \cdot I_{IVTT}}{\sigma_{IVTT,entry}}\right) \end{aligned} \tag{5.6}$$

Coefficient	Meaning
η_{entry}	Error term (entry survey satisfaction with operations)
α_{age}	Age (in year brackets)
α_{income}	Income (10,000 USD brackets)
$\alpha_{unknownincome}$	Unknown income (Binary)
$\alpha_{longuser}$	Long-time user (Binary: System user $>$ 2 years)
$\alpha_{entrysat}$	Entry satisfaction with operations (4 pt. Likert)
α_{mood}	General mood (5 pt. Likert)
α_{ivtt}	In-vehicle travel time (Minutes)
α_{early}	Early arrival at destination stop (Minutes)
α_{delay}	Late arrival at destination stop (Minutes)
$\alpha_{waitover5}$	Wait time greater than 5 minutes
$\alpha_{waitunder5}$	Wait time less than or equal to 5 minutes
α_{nowait}	No wait time inferred from location data (Binary)
$\alpha_{transfertime}$	Transfer time (Minutes)
$\alpha_{notransfer}$	No transfer inferred from location data (Binary)
$\alpha_{leftbehind}$	Denied boardings (Inferred from location data)
$\alpha_{unobservedtransfer}$	Transfer reported but
	not observed in location data (Binary)
η_{daily}	Error term (daily mobile satisfaction with operations)
$\beta_{dailysat}$	Daily satisfaction with operations (5 pt. Likert)
$\beta_{leftbehind\_1\_9}$	Between 1 and 9 denied boardings (Self-reported)
$\beta_{leftbehind\_10}$	10 or more denied boardings (Self-reported)
$\beta_{latework}$	Arrived late at work or school (Self-reported)
$\beta_{lateleisure}$	Arrived late at a leisure activity (Self-reported)
$\beta_{notravel}$	Wanted to use Muni but
	could not due to delay (Self-reported)
$\eta_{exitops}$	Error term (exit survey satisfaction with operations)
$\eta_{errorcorr}$	Error term correlation
$\eta_{exitenv}$	Error term (exit survey satisfaction with the travel environment
$ASC_{shortint}$	ASC intended Muni use short-term
$ASC_{shortdes}$	ASC desired Muni use short-term
$ASC_{longdes}$	ASC desired Muni use long-term
γ_{env}	Coefficient for exit satisfaction travel environment
γ_{ops}	Coefficient for exit satisfaction operations
$\mu_{shortint}$	Scale parameter intended Muni use short-term
$\mu_{shortdes}$	Scale parameter desired Muni use short-term
$\mu_{longdes}$	Scale parameter desired Muni use long-term

Table 5.1: List of coefficients of the structural equations.

Limitations

The model is subject to a few limitations which are discussed here. First, the outcome variables were only measured in the exit survey, and only one of them is relative to the traveler's behavior prior to the study. Therefore, the results can to be understood as indicative of possible links between experiences, satisfaction and future behavior, but there is not definitive proof that the experiences observed during the study were alone responsible for the observed outcomes. Second, the coefficients of the structural model relating the daily satisfaction to the exit satisfaction needed to be constrained to be equal due to estimation difficulties if they were unconstrained. As a consequence, all daily satisfaction ratings were weighted equally. One possible cause is that participants' satisfaction was not measured at the same times with respect to the time of the exit survey, as they were only required to give five responses over the course of the study. By specifying separate coefficients for each of the past satisfaction ratings, it was hoped that a time-dependence would be observable, for example, that the most recent experience would have a stronger influence than more distant experiences. The failure to observe such a time dependence in this model does not mean it is not present, but is more likely due to the data limitations. In future research, satisfaction should be measured at the same time intervals between the daily mobile surveys and the exit surveys for all participants, and the model should be re-estimated with those data to discover potential time-dependencies. In addition, if it is possible to sample satisfaction for all participants on consecutive days, the data would allow the observation of serial dependence between measurements of different days.

5.4 Results

Descriptive analysis

Table 5.2 shows a cross-tabulation of the responses to two questions: "Would you prefer to use Muni less/same/more in January 2014" and "As soon as my circumstances permit, I would like to use public transportation less" (responses to the latter are level of agreement). Out of the 687 participants, 187 stated that they would prefer to use public transportation less in January, and they were subsequently asked for the reasons for that statement. Out of those 187, 96 also agreed with the statement that they would like to use public transportation less as soon as their circumstances permitted. Figure 5.3 shows the stated reasons on a scale from "not at all influential" (1) to "very influential" (5). It can be seen that only 17 out of the 96 participants stated that negative experiences during the study did not influence their desire to reduce their use of transit, and 16 said such experiences were slightly influential. The remaining 63 were split evenly between "somewhat influential", "moderately influential" and "very influential". Among the specific reasons mentioned, the most important ones were overall unreliability, crowding levels, wait time unreliability and unreliability of in-vehicle travel times. Unreliability of transfer times was mentioned less frequently, but that result should be considered with caution since not all participants transferred. Unreliability and crowding levels are of course linked due to bus bunching. It can also be seen that out of the other environmental variables, cleanliness, safety, comfort, the friendliness and competence of staff and the pleasantness of other passengers were reported to be much less influential

		In January, I would like to use transit...			
		less	the same	more	Sum
"As soon as my circumstances permit, I want to use transit less"	Agree	96	81	13	190
	Neutral	42	80	32	154
	Disagree	49	210	84	343
	Sum	187	371	129	687

Table 5.2: Cross-tabulation of two questions regarding cessation of transit use.

than crowding. Travel times and service frequencies when there are no delays were asked about separately, and as can be seen, were reported to be less influential by participants than travel time reliability variables. Lastly, the least influential variables were related to the cost of travel and the fare payment system.

For participants who responded to the question "Compared to how much you used Muni during the study, you anticipate using Muni in January..." either by saying that they were going to increase or decrease their use of Muni, a follow-up question was asked regarding their anticipated mode shifts. Participants who said they anticipated decreasing their Muni use were asked what modes they were going to shift their trips to, and participants who said they anticipated increasing their Muni use where asked what modes they were shifting their trips from. The results are shown in figures 5.4 and 5.5. "None" refers to trips that the participant did not make before or ceased making when the shift to or from Muni occurred. It can be seen that increased Muni use drew mostly from walk trips, followed by trips that were not made before and auto trips. On the other hand, people who decreased their Muni use primarily either ceased making those trips, began using the automobile or walking.

Modeling results

The parameters of the structural model are explained in table 5.1. The estimation results for the structural model are in table 5.3, and the estimation results for the measurement models are in table 5.4. To ensure that the model was identified, several parameters needed to be constrained to 1. Those are marked by asterisks in tables 5.3 and 5.4. In addition, it was observed that the likelihood function was very flat along the dimension of the alternative-specific constant (ASC) for short-term intention, which caused estimation difficulties. Therefore, the latter variable was also constrained to 1.

Coefficient	Value	Robust Std. Error	Robust t-stat	p-value
η_{entry}	1.00	*	*	*
α_{age}	0.01	0.00	1.64	0.10
α_{income}	-0.01	0.01	-0.80	0.42
$\alpha_{unknownincome}$	-0.02	0.65	-0.04	0.97
$\alpha_{longuser}$	-0.25	0.17	-1.47	0.14
$\alpha_{entrysat}$	1.02	0.16	6.54	0.00
α_{mood}	0.38	0.08	5.02	0.00

Continued on next page

Coefficient	Value	Robust Std. Error	Robust t-stat	p-value
α_{ivtt}	0.00	0.01	0.18	0.85
α_{early}	0.01	0.06	0.24	0.81
α_{delay}	-0.20	0.04	-4.88	0.00
$\alpha_{waitover5}$	-0.02	0.03	-0.74	0.46
$\alpha_{waitunder5}$	0.01	0.03	0.55	0.58
α_{nowait}	-0.28	0.09	-3.04	0.00
$\alpha_{transfertime}$	-0.05	0.01	-3.28	0.00
$\alpha_{notransfer}$	-0.21	0.12	-1.83	0.07
$\alpha_{leftbehind}$	-0.12	0.10	-1.20	0.23
$\alpha_{unobservedtransfer}$	-0.51	0.19	-2.70	0.01
η_{daily}	1.00	*	*	*
$\beta_{dailysat}$	0.10	0.01	8.54	0.00
$\beta_{leftbehind\_1\_9}$	-0.10	0.06	-1.85	0.06
$\beta_{leftbehind\_10}$	-0.29	0.12	-2.45	0.01
$\beta_{latework}$	-0.06	0.01	-4.48	0.00
$\beta_{lateleisure}$	-0.04	0.02	-2.75	0.01
$\beta_{notravel}$	-0.07	0.03	-2.59	0.01
$\eta_{exitops}$	0.42	0.07	6.51	0.00
$\eta_{errorcorr}$	0.16	0.18	0.88	0.38
$\eta_{exitenv}$	0.57	0.05	12.56	0.00
$ASC_{shortdes}$	1.07	0.12	9.06	0.00
$ASC_{longdes}$	0.51	0.14	3.68	0.00
$ASC_{shortint}$	1.00	*	*	*
γ_{env}	0.23	0.10	2.23	0.03
γ_{ops}	0.39	0.10	3.89	0.00
$\mu_{shortint}$	1.78	0.18	9.95	0.00
$\mu_{shortdes}$	2.33	0.69	3.38	0.00
$\mu_{longdes}$	1.00	*	*	*

Table 5.3: Estimation results for the structural equations of the latent variable choice model.

	Latent variable	Indicator variable	Value	Robust Std err	Robust t-test	p-value
δ	Entry sat. ops.	Overall reliability	2.90	0.06	47.39	0
δ	Entry sat. ops.	IVTT	2.87	0.05	52.22	0
δ	Entry sat. ops.	Wait time	3.08	0.06	55.68	0
δ	Daily sat. ops.	Overall reliability	2.34	0.12	19.93	0
δ	Daily sat. ops.	IVTT	2.31	0.10	22.21	0
δ	Daily sat. ops.	Wait time	2.24	0.11	19.87	0
δ	Daily sat. ops.	Transfer time	2.22	0.12	19.33	0
δ	Rem. daily sat. ops.	Overall reliability	2.44	0.09	27.87	0
δ	Rem. daily sat. ops.	IVTT	2.23	0.09	24.01	0

Continued on next page

	Latent variable	Indicator variable	Value	Robust Std err	Robust t-test	p-value
δ	Rem. daily sat. ops.	Wait time	2.30	0.08	28.44	0
δ	Rem. daily sat. ops.	Transfer time	2.06	0.11	18.26	0
δ	Exit sat. ops.	Overall reliability	1.87	0.11	17.56	0
δ	Exit sat. ops.	IVTT	2.17	0.08	27.64	0
δ	Exit sat. ops.	Wait time	1.99	0.09	22.83	0
δ	Exit sat. env.	Crowding	1.59	0.04	41.79	0
δ	Exit sat. env.	Safety	2.29	0.03	69.36	0
δ	Exit sat. env.	Other Pax	2.15	0.03	69.20	0
δ	Exit sat. env.	Cleanliness	1.91	0.04	48.37	0
λ	Entry sat. ops.	Overall reliability	0.74	0.06	11.99	0
λ	Entry sat. ops.	IVTT	0.63	0.06	10.70	0
λ	Entry sat. ops.	Wait time	0.64	0.05	13.86	0
λ	Daily sat. ops.	Overall reliability	0.57	0.04	15.85	0
λ	Daily sat. ops.	IVTT	0.51	0.03	18.87	0
λ	Daily sat. ops.	Wait time	0.52	0.03	17.21	0
λ	Daily sat. ops.	Transfer time	0.53	0.04	15.05	0
λ	Rem. daily sat. ops.	Overall reliability	0.37	0.05	7.42	0
λ	Rem. daily sat. ops.	IVTT	0.41	0.04	9.65	0
λ	Rem. daily sat. ops.	Wait time	0.35	0.05	7.07	0
λ	Rem. daily sat. ops.	Transfer time	0.25	0.08	3.34	0
λ	Exit sat. ops.	Overall reliability	1.00	*	*	*
λ	Exit sat. ops.	IVTT	0.73	0.05	14.02	0
λ	Exit sat. ops.	Wait time	0.82	0.03	24.34	0
λ	Exit sat. env.	Crowding	1.00	*	*	*
λ	Exit sat. env.	Safety	0.94	0.11	8.94	0
λ	Exit sat. env.	Other Pax	0.81	0.12	6.96	0
λ	Exit sat. env.	Cleanliness	1.12	0.24	4.61	0
σ	Entry sat. ops.	Overall reliability	0.74	0.03	24.55	0
σ	Entry sat. ops.	IVTT	0.75	0.02	30.87	0
σ	Entry sat. ops.	Wait time	0.78	0.02	41.99	0
σ	Daily sat. ops.	Overall reliability	0.76	0.03	29.10	0
σ	Daily sat. ops.	IVTT	0.88	0.02	47.85	0
σ	Daily sat. ops.	Wait time	0.99	0.02	66.03	0
σ	Daily sat. ops.	Transfer time	0.94	0.04	25.35	0
σ	Rem. daily sat. ops.	Overall reliability	0.58	0.03	18.01	0
σ	Rem. daily sat. ops.	IVTT	0.63	0.03	23.57	0
σ	Rem. daily sat. ops.	Wait time	0.73	0.03	21.25	0
σ	Rem. daily sat. ops.	Transfer time	0.88	0.09	9.81	0
σ	Exit sat. ops.	Overall reliability	0.75	0.05	16.05	0
σ	Exit sat. ops.	IVTT	0.76	0.02	35.25	0
σ	Exit sat. ops.	Wait time	0.72	0.03	21.66	0
σ	Exit sat. env.	Crowding	0.81	0.03	27.19	0

Continued on next page

	Latent variable	Indicator variable	Value	Robust Std err	Robust t-test	p-value
σ	Exit sat. env.	Safety	0.56	0.02	30.34	0
σ	Exit sat. env.	Other Pax	0.58	0.03	21.43	0
σ	Exit sat. env.	Cleanliness	0.68	0.08	8.54	0

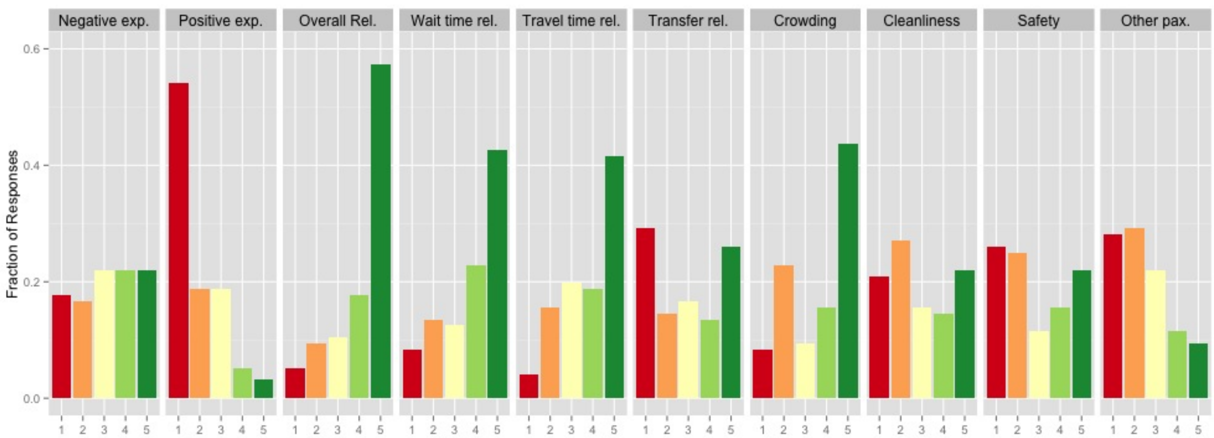
Legend:	
Entry sat. ops.:	Satisfaction with operations in entry survey.
Exit sat. ops.:	Satisfaction with operations in exit survey.
Exit sat. env.:	Satisfaction with travel environment in exit survey.
Daily sat. ops.:	Satisfaction with operations in daily surveys d_1 through d_4.
Rem. daily sat. ops.:	Satisfaction with operations in daily survey d_r.

Table 5.4: Estimation results for the measurement equations of the latent variable choice model. The abbreviations are explained at the end of the table.

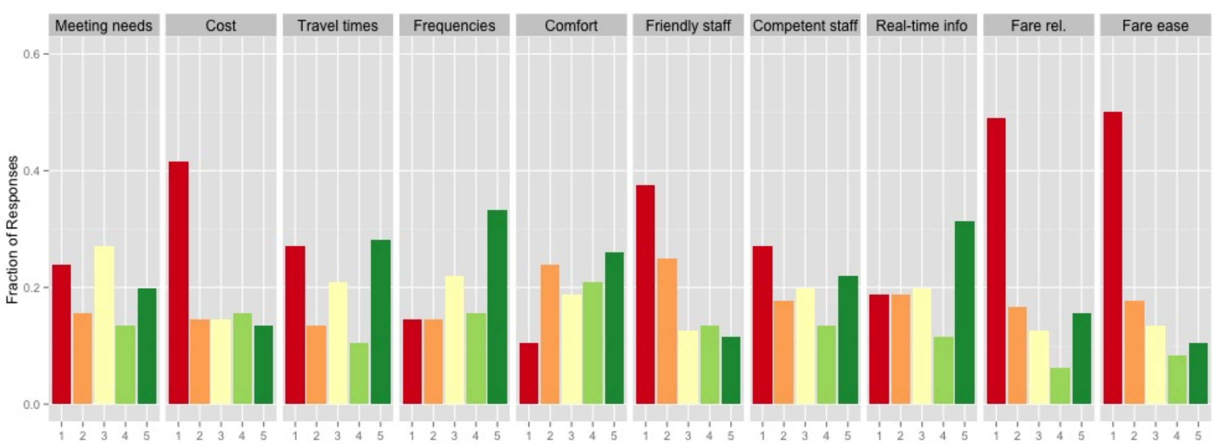
In what follows, the model estimation results are presented and discussed. When comparing these results to those in chapter 4, one should be cognizant of the fact that the assumptions underlying this model are different from those underlying the models in chapter 4. In the latter, the link between the individual travel time components and satisfaction with that component were modeled. No assumptions were made about the relationship between the models, and it was assumed that the reported satisfaction ratings were true measures of the participant's satisfaction. On the other hand, the model results presented here assume that there is an underlying, latent satisfaction with operations (and thus, travel times) variable. The individual satisfactions with travel time components serve as indicators of that latent overall satisfaction.

Effects of sociodemographics, baseline satisfaction and mood

First, we investigate the effect of sociodemographic attributes on people's reported daily satisfaction. It can be seen in table 5.3 that age has a positive effect (0.01 per year) and is significant, whereas having been a long-time user ($\alpha_{longuser}$) has a negative effect (-0.25). The latter, however, is not significant. The effect of income on reported daily satisfaction is also negative (-0.01 per \$10,000), as is the non-response variable to income. The results with respect to income are intuitive, as higher income is associated with a higher value of time. Both effects, however, are not significant at $p = 0.42$ and $p = 0.97$, respectively. The effect of the entry satisfaction ($\alpha_{entrysat} = 1.02$), which again is a latent variable, and of the participant's general mood on the day of the survey ($\alpha_{mood} = 0.38$) are positive and significant at $p < 0.01$. In other words, these two variables have a markedly stronger effect on satisfaction than age, income and the length of Muni use. Of course we would expect the entry satisfaction to depend on sociodemographic variables as well, but it was not possible to include the mood and entry satisfaction in both the daily and the entry survey as this caused estimation problems. These results are in line with findings from chapter 4, though the difference between the age and income variables and the mood and entry satisfaction variables was more pronounced in the latent variable model.



(a) Negative experiences during the study, positive experiences during the study, overall reliability, wait time reliability, travel time reliability, transfer time reliability, crowding, cleanliness, safety, pleasantness of other passengers.



(b) Ability of Muni to meet daily travel needs, cost, on-board travel times when there are no delays, frequencies of service, comfort, friendliness of staff, competence of staff, accuracy of real-time information, reliability of fare payment system, ease of use of fare payment system.

Figure 5.3: Stated reasons for wanting to use transit less or not at all anymore.

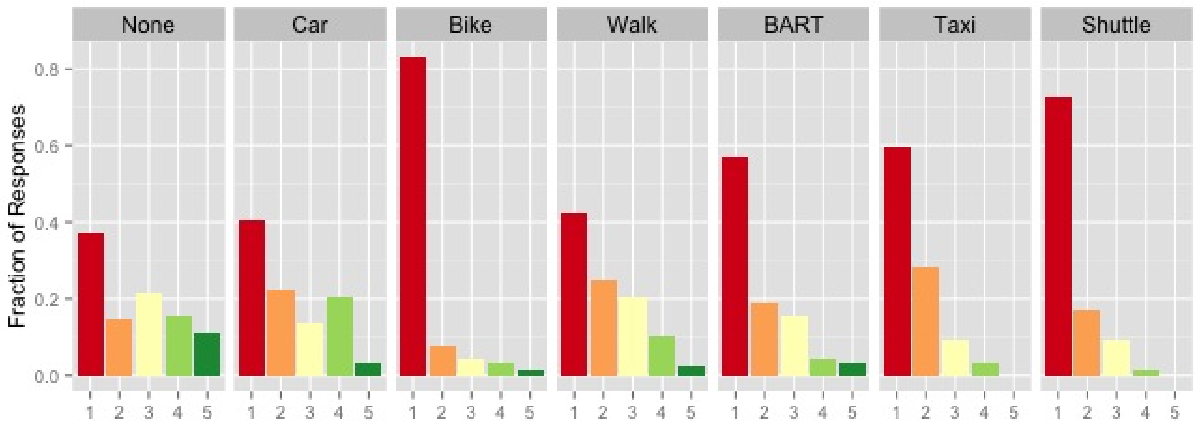


Figure 5.4: Travel modes substituted for Muni trips. (1 - applies to none of the trips, 5 - applies to all trips.)

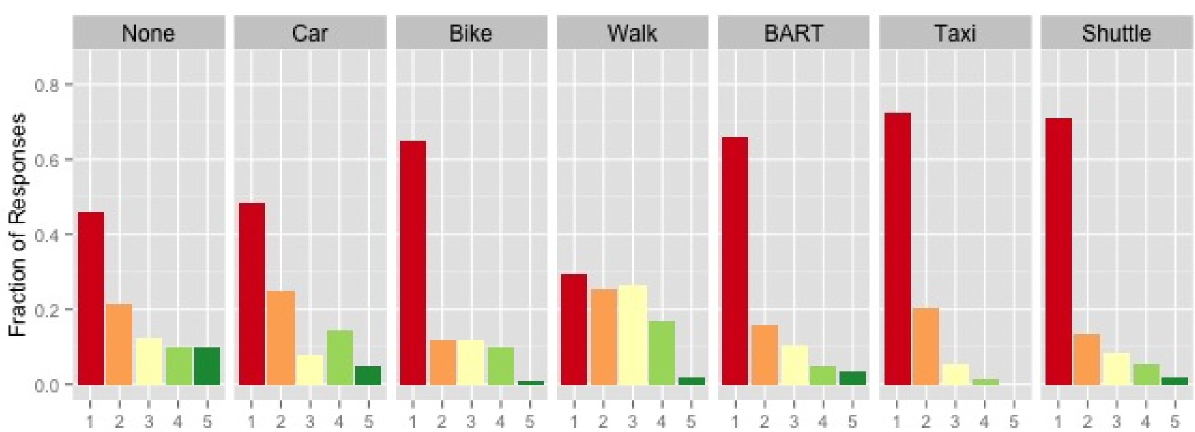


Figure 5.5: Travel modes substituted with Muni trips. (1 - applies to none of the trips, 5 - applies to all trips.)

Travel time variables

The scheduled IVTT is found to have virtually no effect on overall satisfaction with operations (α_{ivtt}). Delays with respect to the scheduled IVTT have a significant negative effect ($\alpha_{delay} = -0.20, p < 0.01$) and earlier arrivals have an insignificant effect on satisfaction ($\alpha_{delay} = 0.01, p = 0.81$). Unfortunately, this model was not able to capture the effect of wait times on a joint, latent satisfaction variable, as both coefficients for wait times below 5 minutes ($\alpha_{waitunder5}$) and wait times above 5 minutes ($\alpha_{waitover5}$) are insignificant. The reason for this merits further investigation, but as discussed in chapter 4, it might be linked to the fact that the majority of wait times was short, with an average around 2 minutes. It must be noted that the observed wait times used in the model estimation only capture time actually spent standing at the origin stop. They exclude additional schedule delay times (i.e., the time between when a person wanted to leave and the time of a transit vehicle departure), which some participants might have chosen to spend elsewhere, such as in their homes. While these might have been perceived as wait times by some participants, it was not possible to automatically identify them with location tracking data from their phones, and therefore, the link between the latent satisfaction with travel times and the wait times could not be established in this case. As passengers rely more on real-time information, the strategy of spending wait time at locations other than the stop and of going to the stop only when an arrival is predicted will most likely become more prevalent.

Wait times could not be identified from the tracking data alone for approximately 50% of observations in the data set. The observations with missing wait times are denoted by a binary variable, the coefficient of which (α_{nowait}) is negative and significant. The possible reasons for missing wait time observations were as follows:

1. If there was insufficient location tracking data available for that portion of the trip. This includes both smartphone and vehicle location data.

2. if the participant was carrying out an activity near the stop (e.g. work) which made it impossible to distinguish activity time from wait time.

3. If the wait time was incurred when the participant transferred from BART (regional rapid transit) to a local metro train inside an underground metro station.

The causes of missing wait time observations are discussed further in section 4.4. It is not known whether the missing observations skewed the distribution of observed wait times in any particular way. The insignificance of the wait time coefficients suggests that the observed wait times were generally in a range that did not significantly affect riders' overall satisfaction with operations. Together with the low sensitivity toward wait times observed in section 4.4, the results of this model suggest that in future research, a different approach should be taken to identifying wait times. There are three possible avenues:

- Adding other sensor data such as accelerometer in order to better identify when a person actually walked to a transit stop.

- Directly asking a participant about the perceived wait time and where it was spent.

- Tracking the use of real-time information on the phone in order to determine when a participant first looked at upcoming departures. This could serve as an indicator of the beginning of a wait.

Unlike the coefficient of the wait time at the origin stop, the transfer time coefficient ($\alpha_{tranfertime}$) is negative and significant at $p < 0.01$. A comparison of the transfer time coefficient and the IVTT coefficient shows that according to the model, one minute of in-vehicle delay causes as much dissatisfaction as four minutes of transfer time. The model further includes two binary variables related to transfer time: $\alpha_{notransfer}$ captures cases no transfer was identified from the location tracking data and the participant did not report a transfer, and ($\alpha_{unobservedtransfer}$) captured cases where the participant reported having transferred but the transfer could not be identified from the location tracking data. Both are negative and significant at $p < 0.10$. While this result is intuitive for the latter coefficient, it is not intuitive for the former, as it suggests that in general, passengers who transfer tend to report a higher satisfaction than passengers who do not transfer. This merits further investigation with a larger data set.

Effect on exit satisfaction and critical incidents

The coefficient $\beta_{dailysat}$ in table 5.3 links daily satisfaction with operations to the exit satisfaction with operations. As expected, it is positive (0.1), and it is also significant at $p < 0.01$, showing a positive correlation between daily satisfaction and exit satisfaction.

Interestingly, all five coefficients related to self-reported critical incidents have negative and significant estimates at $p < 0.1$. The first two are the number of times a person arrived late at work or school ($\beta_{latework}$) due to a transit delay and the number of times a person arrived late at a leisure activity due to a transit delay ($\beta_{lateleisure}$), which were both reported in the exit survey. $\beta_{notravel}$ captures cases where participants reported on their daily mobile surveys that they wanted to use public transportation that day but could not due to a delay and were forced to choose a different mode. However, there was no obligation to report these incidents, so it must be assumed that the reported numbers are a lower bound. Therefore, the estimated coefficient is an upper bound on the "badness" of such incidents. A special case of critical incidents were denied boardings: These were captured both through self-reports in the exit survey and through automated detection. The latter affect the daily satisfaction with operations via $\alpha_{leftbehind}$. The automated detection was only based on location data: If a participant was observed to be at a stop and not board a passing vehicle but board the following one, it was recorded as a denied boarding. There is a risk of misclassification, as the participant may have been carrying out a legitimate activity and may not have intended to board the first vehicle. Therefore, the number of such automatically detected incidents in the data set is an upper bound, and the coefficient estimate is a lower bound for the impact. In table 5.3, $\alpha_{leftbehind}$ is negative but not significant. On the other hand, the effect of self-reported denied boardings was found to exhibit some nonlinearity. There were 10 possible answers to the self-reported question: From 0 to 9, and then "10 or more". The model was found to produce the best fit if these two categories were separated, as in table 5.3. Both coefficients, $\beta_{leftbehind\_1\_9}$ and $\beta_{leftbehind\_10}$, are negative and significant at $p < 0.10$, but the latter is approximately three times larger than the former. It is possible (and plausible) that the larger coefficient for 10 or more denied boardings is capturing protest responses.

Effects on future behavior

The final component is the choice model. It is specified with only two inputs: The participant's satisfaction with operations reported in the exit survey and the participant's satisfaction with the travel environment. All other variables in the model, including the daily satisfaction with operations, the critical incidents, the travel time experiences and the entry satisfaction with operations, affect future behavior through the exit satisfaction with operations. The indicator variables of the satisfaction in the exit survey are on a four-point Likert scale, whereas the choice indicators are binary. Given that the indicator variables have different reference points, one has to make the assumption that the participants did not change their pre-study frequency of transit use during the study in order to interpret the results. Given that assumption, the choice is between (a) continuing to use public transportation at the same frequency as before and during the study or using it more frequently, and (b) using public transportation less frequently or discontinuing it altogether. The choice indicator for the former is 1, for the latter it is 0. Thus, positive coefficients mean there is a positive correlation between the input variable and the participant's willingness to use public transportation the same or more in the future. As can be seen in table 5.3, the effects of both satisfaction with operations and satisfaction with the travel environment in the exit survey are positive (0.39 and 0.23, respectively) and significant at $p < 0.05$. This is intuitive, as higher satisfaction leads to a higher willingness to continue using transit in the future.

Of particular interest here is the relative difference between the two coefficients. The coefficient of the latent satisfaction with operations is approximately 1.7 times the coefficient of the latent satisfaction with the travel environment. Since these are latent variables, their exact values cannot be calculated, but the comparison can be made with the help of two of the indicator variables: A change in the latent satisfaction with operations variable that causes a one-point increase in satisfaction with overall reliability has 1.46 times the effect of a change in the satisfaction with travel environment variable that causes a one-point increase in satisfaction with crowding. This confirms that for the present group, overall satisfaction with operations has a stronger influence on future mode choice decisions than overall satisfaction with the travel environment. This is consistent with the results of the descriptive analysis. In future research, it would be very interesting to add objective measurements to the travel environment variables. Crowding would be of particular interest given the importance reported by participants in figure 5.3a; this could be calculated with data from automatic passenger counting or fare payment systems.

With the help of the final choice model, it is now possible to calculate the relative influence of various experiences on passengers' willingness to remain transit riders in the future. As a calculation example, the effect of one incident of not being able to travel due to a delay on the transit network on the choice utility is $\beta_{notravel} \cdot \gamma_{ops} = -0.07 * 0.39 = -0.027$.

	10 min delay on board	10 min transfer, late to work	Left behind, 10 min wait
Per Person	-0.006	-0.003	-0.002
System-wide	-1628	-745	-504

Table 5.5: Results of the simulation.

Calculation of trade-offs

If the coefficient for wait times were negative and significant, it would be possible to calculate the impact of negative critical incidents in terms of the equivalent amount of dissatisfaction caused by wait times. For instance, if $\alpha_{waitover5}$ and $\alpha_{waitunder5}$ were -0.02 and significant and $\alpha_{leftbehind}$ were -0.12 and significant, one could state that the dissatisfaction caused by one instance of a denied boarding would be equivalent to the dissatisfaction caused by approximately 6 minutes of wait time at the origin stop. Such trade-offs provide very good rules of thumb for transit planning professionals, as is illustrated by the popularity of the rule of thumb that a minute of out-of-vehicle travel time is twice as onerous as a minute of in-vehicle travel time (Wardman, 2004). Therefore, a goal of future research should be to derive significant wait time coefficients in order to calculate such trade-offs.

Simulation

A simulation was conducted to quantify the impacts of various service quality problems on users. The group of subjects from whom the data had been collected served as a convenience sample. The simulation scenarios were based on ridership numbers of the San Francisco MTA, and in total, three scenarios were run:

1. Impact of a ten-minute on-board delay.
2. Impact of being left behind at a stop with corresponding ten minutes of additional wait time.
3. Impact of a 10-minute transfer wait time and arriving late at work.

In every scenario, it was assumed that every rider in the system experienced the delay in question once within a given time period. Table 5.5 shows the results. The base case is the status quo, where on average, participants had a 0.77 probability of remaining transit riders in the future. The second line of the table shows the change in probability due to the incident. The third line is a calculation of the number of lost riders in the SFMTA system, assuming the incident was experienced by all riders and based on a ridership of 280,000. Assuming constant ridership, the SFMTA's rate of turnover is estimated to be approximately 30,000 passengers per year, so the simulated hypothetical events can account for 5.4%, 2.5% and 1.7% of the turnover, respectively.

5.5 Discussion

The results show the link between service quality problems and loss of ridership from two different angles. First, we selected participants who reported a behavioral desire to reduce their use of public transportation on both a short-term and long-term scale, and investigated the self-reported reasons for which they wanted to do so. It was seen that travel time reliability was mentioned as the overall most important factor. In terms of the number of 'very influential' responses, crowding came in second, but in terms of the average of all responses, the second-most important factors were wait time reliability and travel time reliability. Both had an average response of 3.781, compared to 4.125 for overall reliability and 3.635 for crowding. In a broader sense, even crowding can be considered a reliability variable since the crowding of vehicles is related to vehicle bunching and since passengers do not know ahead of time whether they will be able to find a seat. Overall, travel time variables were more influential than travel environment variables. However, it should be noted that this study concerned users who were already regular transit users, and therefore, it may be a self-selected group that might be less concerned with the travel environment than, for instance, a comparison group of auto users.

Second, we presented model estimation results and applied them to a simulation. The model results are generally in line with previous findings presented in chapter 4, showing that the scheduled in-vehicle travel time has little effect on dissatisfaction and that early arrivals at the destination do not have a significant effect on satisfaction, but that in-vehicle delays are an important driver of dissatisfaction. By extension, this means that in-vehicle delays also have a strong impact on satisfaction reported in the exit survey and on passengers' desire to stop using public transportation. The transfer time coefficient is also negative and significant, but the origin wait time coefficients were not significant. It is assumed that this may be partly related to the difficulties associated with properly identifying origin wait times, as explained in section 5.4, and partly with the fact that participants appear to be choosing to spend their wait times at locations other than the stop and rely on real-time information to go to the stop when an arrival is predicted. Therefore, the wait times detected from location data may not necessarily have corresponded to the wait times as define by the user. In future research, these data shortcomings should be addressed in order to solidify our understanding of the impact of various delay times on passengers. The model presented in this chapter goes beyond previous satisfaction models by explicitly linking critical incidents and personal experiences with travel times to future behavioral intentions by way of customer satisfaction, and the significance of the relevant coefficients in table 5.3 demonstrates that this link is present. It is shown that for the present group of participants, which was drawn from current transit users, the satisfaction with travel times and operational aspects is more important in determining their willingness to remain transit riders in the future than satisfaction with the travel environment. Furthermore, even though the influence of wait times relative to in-vehicle delay times and transfer times requires further research, the results clearly demonstrate the value of developing models using participants' *personal experiences* with service quality as a means of understanding future mode choice intentions and the influence of various factors related to service quality. Besides the experiences with travel times, we find that several critical incidents have measurable negative effects on participants' overall satisfaction in the exit survey and thus on their willingness to remain transit riders in the future. The simulation results demonstrate a possible application of these models: They can be used to project how reductions (or increases) in delays and critical incidents can affect rider turnover. In the present application, in can be seen in the simulation results the effects of poor service quality on ridership can be considerable - a single 10-minute in-vehicle delay per passenger per year (or any other arbitrary time frame) can account for over 5% or ridership loss in the case of the SFMTA.

The latent variable modeling framework used in this chapter was valuable as it permitted us to summarize an individual's overall satisfaction with operations in one variable and to determine the influence of a variety of experiences with travel times on that variable. It is flexible and its specification can accommodate variables collected on different time scales, such as the daily satisfaction with operations and the entry and exit satisfaction. Most importantly, it allowed us to account for correlation between the individual satisfaction ratings with respect to the components of participants' experienced travel times.

5.6 Conclusions

In this chapter, we presented an analysis and model results to understand the link between service quality, satisfaction, and transit ridership cessation. This work emphasizes the importance of riders' personal experiences; an innovative procedure was used to map location data from users' mobile phones to vehicle location data in order to automatically identify personal experiences and use them in the estimation of the model. This demonstrates the value and potential of such new data collection methods in answering complex questions and observing phenomena that require panel data. The insights gained from these data help establish the link between travel time variability and other critical incidents, satisfaction and transit ridership loss; this framework makes it possible to directly model the effect of negative personal experiences on future mode choice decisions and thus on ridership loss due to delays and system management strategies. In the future, it can be further refined with additional data in order to form the basis for new operational tools that would enable a move from system-based to person-based performance metrics for transit agencies.

Chapter 6

Final Remarks

6.1 Summary

In this dissertation, it was shown that with the pervasive use of smartphones and mobile sensors, researchers have a powerful new tool at hand to study complex, long-term and dynamic behavioral phenomena. In particular, it is demonstrated how smartphone-based surveys can be combined with sensor data and automatically collected data from other sources to form rich data sets that facilitate such studies. The following research question was formulated: What are the links among transit service quality, riders' experiences therewith, riders' satisfaction and future mode choices? The question was motivated by the fact that many US transit operators see high rates of ridership turnover but that to date, relatively little is known about the factors driving that turnover. While public policies are often aimed at encouraging travelers to move out of their cars and into using public transportation, there is comparatively little focus on retaining riders once they are using transit.

To address these questions, the design and implementation of a large-scale, smartphone-based study is presented. It is argued that capturing personal experiences of riders is of paramount importance to understanding rider satisfaction. Responses to satisfaction surveys alone do not give the analyst indications about the sensitivity of those responses to changes in the real-world variables. For example, an analyst would not be able to know how much improvement in crowding levels would lead to a significant change in passengers' satisfaction with the level of crowding. On the other hand, if the analyst has both objective measures of the quality of the experience and subjective assessments by the decision-maker available, this link can be explicitly modeled, and policy and planning recommendations can be derived from it. It is shown in the context of transit, that in order to objectively measure travelers' personal experiences with travel times, smartphone data can be combined with automated vehicle location data. A computational procedure to automatically achieve this is described in this dissertation.

The experiences that were made with the study and reported in this dissertation can inform future smartphone-based research. Furthermore, the detailed description of the study procedure and data set in this dissertation can support future research undertaken with this data set. While the computational system for matching vehicle and smartphone location data was designed specifically for the system in which the study was undertaken, the overall methodology and logic followed is general and applicable to many other systems.

Using the data, two models were developed to explain the key links from personal experiences to satisfaction and from satisfaction to future behavior. First, a detailed ordinal logit model is developed to explain satisfaction with travel times as a function of the experienced travel times and a number of covariates. A significant relationship between the respondents' reported satisfaction and the various travel time components observed from tracking data is found, and the influence of gender, age and income is captured. Furthermore, it is found that in order to understand travelers' satisfaction with daily quality of service, both baseline satisfaction and the participant's mood on the response day are important to capture. Out of the three travel time components modeled (in-vehicle travel time, wait time at the origin stop and transfer time), passengers were found to exhibit the strongest sensitivity toward in-vehicle delays, although the data only allowed for the analysis of wait times up to 10 minutes.

Informed by these findings, a second model was developed to explain participants' stated future mode choice intentions. Prior to developing the model, descriptive statistics are presented which show that passengers' self-reported reasons for wanting to reduce their use of public transportation are primarily related to travel time reliability and to crowding. The model developed is a latent variable choice model where satisfaction enters through two latent variables: Satisfaction with operations (travel times) and satisfaction with the travel environment. The estimation results confirm the self-reported reasons by the participants in that satisfaction with operations is the more important driver of customer behavior. With the help of the model, the influence of personal experiences with unreliability on future passenger behavior can be quantified, and a variety of negative critical incidents (such as arriving late at work or being left behind) can be expressed in terms of travel time delays. Multiplied by a value of time, this result could also be used to assign monetary values to such events.

The evolution of the study from inception to execution represented a learning process for all researchers involved. Initially, the study was intended to collect data on the effect of negative critical incidents on a rider's emotions and satisfaction with the travel experience, and the link among emotions, satisfaction and behavior. An integral part of the study was the development and deployment of a system to automatically infer travel times from vehicle and phone location data. The latter objective was achieved, with lessons learned for future implementations of similar systems. A further objective was to observe the dynamics of satisfaction and emotions, and to understand how they evolve over time as a function of personal experiences. The original design of the study created a set of variables measuring future behavioral intentions and desires. These were included in identical form in the entry and the exit survey; the intention was to capture differences between the variables and to relate them to experiences made by the participants between the two surveys (i.e., during the study).

While the objective of modeling the link between subjective assessment of experiences and objectively measured travel times was realized, the temporal dynamics of subjective well-being and satisfaction proved difficult to understand from the data. This was mainly due to the fact that participants were required to fill out only five daily surveys out of an unknown number of usage days, and therefore not all days of transit use were captured. Nonetheless, this objective deserves further attention. The data set has not yet been fully analyzed at this point; it might still be possible to extract a subset of the data that can be assumed to be complete and that lends itself to the analysis described above.

Of the satisfaction and trip-related emotions data that were collected, only those regarding satisfaction were used in the final behavioral models. Satisfaction was seen as the more important variable since it is used in the behavior frameworks that underpinned the design of the study (e.g., the Theory of Planned Behavior). Satisfaction and emotions were also found to be strongly correlated and therefore difficult to include simultaneously in a model. Section 6.2 describes some possible future research paths using the data regarding emotions.

6.2 Future research

There are three principal avenues for future research: The further evaluation of the data set collected during the SFTQS, the design of similar studies to validate the results and findings that were presented in this chapter, and work to improve the computational infrastructure that was developed as part of this study.

Further evaluation of the SFTQS data set

The data set collected during the SFTQS allows for more analyses than those presented in this dissertation. In particular, the data on subjective well-being and trip-based emotions was not evaluated; a possible avenue for future research would be to explore the link between experienced emotions, predicted emotions, satisfaction and future behavior. Model specifications could build on those presented in chapters 4 and 5 of this dissertation, or could take different approaches. An interesting question would be whether supplementing the existing satisfaction models with emotions data can increase predictive power, or whether a model based purely on emotional well-being of participants performs better than the satisfaction model. Prior to modeling, a more in-depth descriptive analysis should focus on the prevalence of different trip-related emotions and on the causes of negative or positive emotions.

A second research direction would be to investigate in more depth the dynamics of satisfaction over time, as captured by the series of mobile survey responses given by participants over the course of the six weeks and beyond. A questions that could be posed is, for example, what is the influence of the most recent satisfaction report and of overall high satisfaction or dissatisfaction during the study on final satisfaction. This would help to understand whether the peak-end rule (Kahneman et al., 1993) is observable in the data. Furthermore, research could focus on overall trends during the study and on the question whether any lingering effects of particularly dissatisfying events can be observed in the days following that event. Given the data limitations discussed in section 6.1, this work would need to utilize a subset of the data known or assumed to be complete.

Validating the findings of this research

While the results of the two behavioral models are consistent and were robust over several different specifications, they are nonetheless only first indications, and further research is required to validate the findings. In particular, it would be useful to repeat the study with different samples. Due to the incentive structure, the present data set consisted primarily of habitual transit riders; a more balanced sample that would include many non-transit users could shed light on differences between these groups. It would also permit the observation of adaptation effects as non-transit users grow accustomed to using transit. An overall larger sample size or a study in a different location would also contribute to more generalized findings.

Building on the findings in this dissertation, future research could focus on developing the first user-based performance measurement tools that could be broadly used by transit agencies. This research would need to address questions of representativeness and develop methods for continuous, unobtrusive data collection from a large number of transit riders while being respectful of potential privacy concerns. These new measures could be used alongside existing operational measures in order to gain a more complete understanding of the performance of public transportation when compared to other modes.

Improving the infrastructure

Due to the high service frequencies provided by the SFMTA on their core routes, which were used by many study participants, it was not possible to capture a sufficient number of wait times longer than 10 minutes to adequately model wait time satisfaction beyond 10 minutes. Repeating the data collection with a larger sample and a specific focus on capturing more variation in out-of-vehicle travel times would be valuable in that respect; users of lower-frequency routes might need to be specifically targeted to capture longer wait times.

Furthermore, future research projects would benefit from improvements to the detection of out-of-vehicle travel times. In this work, heuristic rules were used in the automated vehicle location to phone location matching system to separate true wait times at stops from activities near transit stops. This approach was based solely on location data and was limited due to two factors: First, the exclusive use of location data means that it was difficult to detect the walk from an activity location to the stop if the activity was located very close to the stop. The addition of more sensors could prove useful. For example, the accelerometer could be used for the detection of walking. Second, further investigation is needed to understand shifted perceptions of wait times due to real-time information. Since travelers might elect to spend part of their wait time away from the stop, these wait times cannot be detected automatically. At present, the only option is to survey participants in order to capture the entire perceived wait time. The closer this survey can be conducted to the actual wait time, the better.

Both the wait time survey and a future transit experience survey could benefit greatly from real-time detection of transit trips. The ability to detect that a participant is traveling on public transportation in real-time without placing a heavy load on the phone data connection would be a major improvement and is a promising direction for future research.

Bibliography

Aarts, Henk, Bas Verplanken, and Ad Van Knippenberg (1997). “Habit and information use in travel mode choices”. In: *Acta Psychologica* 96.1, pp. 1–14.

Abou-Zeid, Maya and Moshe Ben-Akiva (2010). “A model of travel happiness and mode switching”. In: *Choice modelling: the state-of-the-art and the state-of-practice (proceedings from the inaugural international choice modelling conference)*, pp. 289–306.

Abou-Zeid, Maya et al. (2012). “Happiness and travel mode switching: findings from a Swiss public transportation experiment”. In: *Transport Policy* 19.1, pp. 93–104.

Ajzen, Icek (1991). “The theory of planned behavior”. In: *Organizational behavior and human decision processes* 50.2, pp. 179–211.

Anderson, Eugene W and Mary W Sullivan (1993). “The antecedents and consequences of customer satisfaction for firms”. In: *Marketing science* 12.2, pp. 125–143.

Anderson, Shannon, Lisa Klein Pearo, and Sally K Widener (2008). “Drivers of service satisfaction linking customer satisfaction to the service concept and customer characteristics”. In: *Journal of Service Research* 10.4, pp. 365–381.

Asakura, Yasuo and Eiji Hato (2004). “Tracking survey for individual travel behaviour using mobile communication instruments”. In: *Transportation Research Part C: Emerging Technologies* 12.3, pp. 273–291.

Auld, Joshua et al. (2009). “An automated GPS-based prompted recall survey with learning algorithms”. In: *Transportation Letters* 1.1, pp. 59–79.

Axhausen, Kay and Claude Weis (2013). “Predicting Response Rate: A Natural Experiment”. In: *Survey Practice* 3.2.

Axhausen, Kay W et al. (2002). “Observing the rhythms of daily life: A six-week travel diary”. In: *Transportation* 29.2, pp. 95–124.

Bamberg, Sebastian, Icek Ajzen, and Peter Schmidt (2003). “Choice of travel mode in the theory of planned behavior: The roles of past behavior, habit, and reasoned action”. In: *Basic and applied social psychology* 25.3, pp. 175–187.

Barbeau, Sean J, Nevine L Georggi, and Philip L Winters (2010). “Integration of GPS-enabled Mobile Phones and AVL-Personalized Real-time Transit Navigation Information on Your Phone 2”. In: *Proceedings of the Transportation Research Board 89th Annual Meeting*.

Bates, J. et al. (2001a). “The valuation of reliability for personal travel”. In: *Transportation Research* 37, pp. 191–229.

Bates, John et al. (2001b). “The valuation of reliability for personal travel”. In: *Transportation Research Part E: Logistics and Transportation Review* 37.2, pp. 191–229.

Beirão, Gabriela and JA Sarsfield Cabral (2007). "Understanding attitudes towards public transport and private car: A qualitative study". In: *Transport policy* 14.6, pp. 478–489.

Benezech, Vincent and Nicolas Coulombel (2013). "The value of service reliability". In: *Transportation Research Part B: Methodological* 58, pp. 1–15.

Bergstad, Cecilia Jakobsson et al. (2011). "Subjective well-being related to satisfaction with daily travel". In: *Transportation* 38.1, pp. 1–15.

Biagioni, James et al. (2009). "TransitGenie: a context-aware, real-time transit navigator". In: *Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems*. ACM, 329–330.

Bollen, Kenneth A (1998). *Structural equation models*. Wiley Online Library.

Bolton, Ruth N (1998). "A dynamic model of the duration of the customer's relationship with a continuous service provider: the role of satisfaction". In: *Marketing science* 17.1, pp. 45–65.

Bordagaray, Maria et al. (2014). "Modelling user perception of bus transit quality considering user and service heterogeneity". In: *Transportmetrica A: Transport Science* 10.8, pp. 705–721.

Brakewood, Candace, Sean Barbeau, and Kari Watkins (2014). "An experiment evaluating the impacts of real-time transit information on bus riders in Tampa, Florida". In: *Transportation Research Part A: Policy and Practice* 69, pp. 409–422.

Bricka, Stacey and Chandra R Bhat (2006). "Comparative analysis of global positioning system-based and travel survey-based data". In: *Transportation Research Record: Journal of the Transportation Research Board* 1972.1, pp. 9–20.

Brickman, Philip and Donald T Campbell (1971). "Hedonic relativism and planning the good society". In: *Adaptation-level theory*, pp. 287–305.

Brög, Werner and Thomas Kahn (2003). "Customer Satisfaction as an Element of Strategic Business Decisions". In: *2nd UITP International Marketing Conference, Paris*.

Carrel, Andre, Anne Halvorsen, and Joan L Walker (2013). "Passengers' perception of and behavioral adaptation to unreliability in public transportation". In: *Transportation Research Record: Journal of the Transportation Research Board* 2351.2, pp. 153–162.

Carrus, Giuseppe, Paola Passafaro, and Mirilia Bonnes (2008). "Emotions, habits and rational choices in ecological behaviours: The case of recycling and use of public transportation". In: *Journal of Environmental Psychology* 28.1, pp. 51–62.

Chan, Joanne (2007). "Rail transit OD matrix estimation and journey time reliability metrics using automated fare data". MA thesis. Massachusetts Institute of Technology.

Chapleau, Robert, Martin Trépanier, and Ka Kee Chu (2008). "The ultimate survey for transit planning: complete information with smart card data and GIS". In: *Proceedings of the 8th international conference on survey methods in transport: harmonisation and data comparability*. Annecy, France.

Chebat, Jean-Charles et al. (1995). "Impact of waiting attribution and consumer's mood on perceived quality". In: *Journal of business Research* 34.3, pp. 191–196.

Cherchi, Elisabetta and C Cirillo (2008). "A mixed logit mode choice model on panel data: accounting for systematic and random variations on responses and preferences". In: *87th Annual Meeting of the Transportation Research Board, Washington, DC*.

Cherchi, Elisabetta and Cinzia Cirillo (2010). "Validation and forecasts in models estimated from multiday travel survey". In: *Transportation Research Record: Journal of the Transportation Research Board* 2175.1, pp. 57–64.

Chu, Ka Kee and Robert Chapleau (2010). "Augmenting Transit Trip Characterization and Travel Behavior Comprehension: Multiday Location-Stamped Smart Card Transactions". In: *Transportation Research Record: Journal of the Transportation Research Board* 2183, pp. 29–40.

Chung, Eui-Hwan and Amer Shalaby (2005). "A Trip Reconstruction Tool for GPS-based Personal Travel Surveys". In: *Transportation Planning and Technology* 28.5, pp. 381–401.

Cirillo, Cinzia, Laura Eboli, and Gabriella Mazzulla (2011). "On the Asymmetric User Perception of Transit Service Quality". In: *International Journal of Sustainable Transportation* 5.4, pp. 216–232.

Connor, T.S. (2014). *Experience sampling and ecological momentary assessment with mobile phones*. URL: `http://www.otago.ac.nz/psychology/otago047475.pdf`.

Cottrill, Caitlin D et al. (2013). "Future Mobility Survey". In: *Transportation Research Record: Journal of the Transportation Research Board* 2354.1, pp. 59–67.

Csikszentmihalyi, Mihaly and Reed Larson (1987). "Validity and reliability of the experience-sampling method". In: *The Journal of nervous and mental disease* 175.9, pp. 526–536.

Cui, Alex (2006). "Bus Passenger Origin-Destination Estimation and Related Analyses Using Automated Data Collection Systems". MA thesis. Massachusetts Institute of Technology.

Davis, Mark M and Janelle Heineke (1998). "How disconfirmation, perception and actual waiting times impact customer satisfaction". In: *International Journal of Service Industry Management* 9.1, pp. 64–73.

Davis, Mark M and Michael J Maggard (1990). "An analysis of customer satisfaction with waiting times in a two-stage service process". In: *Journal of Operations Management* 9.3, pp. 324–334.

Del Castillo, JM and FG Benitez (2013). "Determining a public transport satisfaction index from user surveys". In: *Transportmetrica A: Transport Science* 9.8, pp. 713–741.

Diana, Marco (2008). "Making the "primary utility of travel" concept operational: a measurement model for the assessment of the intrinsic utility of reported trips". In: *Transportation research part A: policy and practice* 42.3, pp. 455–474.

Diener, Ed (2000). "Subjective well-being: The science of happiness and a proposal for a national index." In: *American psychologist* 55.1, p. 34.

Dillman, Don A (2011). *Mail and Internet surveys: The tailored design method-2007 Update with new Internet, visual, and mixed-mode guide*. John Wiley & Sons.

Dios Ortúzar, Juan de and Luis G Willumsen (2011). *Modelling transport*. John Wiley & Sons.

Dubé, Laurette and Kalyani Menon (1998). "Why would certain types of in-process negative emotions increase post-purchase consumer satisfaction with services". In: *Advances in services marketing and management* 7, pp. 131–158.

Durrande-Moreau, Agnes (1999). "Waiting for service: ten years of empirical research". In: *International Journal of Service Industry Management* 10.2, pp. 171–194.

Eboli, Laura and Gabriella Mazzulla (2007). "Service quality attributes affecting customer satisfaction for bus transit". In: *Journal of Public Transportation* 10.3, p. 21.

Eboli, Laura and Gabriella Mazzulla (2009). "A new customer satisfaction index for evaluating transit service quality". In: *Journal of Public Transportation* 12.3, pp. 21–37.

— (2010). "How to capture the passengers' point of view on a transit service through rating and choice options". In: *Transport reviews* 30.4, pp. 435–450.

— (2011). "A methodology for evaluating transit service quality based on subjective and objective measures from the passenger's point of view". In: *Transport Policy* 18.1, pp. 172–181.

Edvardsson, Bo (1998). "Causes of customer dissatisfaction-studies of public transport by the critical-incident method". In: *Managing Service Quality: An International Journal* 8.3, pp. 189–197.

Ettema, Dick et al. (2010). "Out-of-home activities, daily travel, and subjective well-being". In: *Transportation Research Part A: Policy and Practice* 44.9, pp. 723–732.

Evans, Gary W, Richard E Wener, and Donald Phillips (2002). "The Morning Rush Hour Predictability and Commuter Stress". In: *Environment and Behavior* 34.4, pp. 521–530.

Farzin, Janine M. (2008). "Constructing an Automated Bus Origin-Destination Matrix Using Farecard and Global Positioning System Data in São Paulo, Brazil". In: *Transportation Research Record: Journal of the Transportation Research Board* 2072, pp. 30–37.

Fosgerau, Mogens and Anders Karlström (2010). "The value of reliability". In: *Transportation Research Part B: Methodological* 44.1, pp. 38–49.

Fredrickson, Barbara L and Daniel Kahneman (1993). "Duration neglect in retrospective evaluations of affective episodes." In: *Journal of personality and social psychology* 65.1, p. 45.

Friman, Margareta, Bo Edvardsson, and Tommy Gärling (2001). "Frequency of negative critical incidents and satisfaction with public transport services." In: *Journal of Retailing and Consumer Services* 8.2, pp. 95–104.

Friman, Margareta, Bo Edvardsson, and Tommy Gärling (1998). "Perceived Service Quality Attributes in Public Transport: Inferences from Complaints and Negative Critical Incidents". In: *Journal of Public Transportation* 2.1, pp. 69–91.

Friman, Margareta and Markus Fellesson (2009). "Service supply and customer satisfaction in public transportation: the quality paradox". In: *Journal of Public Transportation* 12.4, pp. 57–69.

Friman, Margareta et al. (2013). "Psychometric analysis of the satisfaction with travel scale". In: *Transportation Research Part A: Policy and Practice* 48, pp. 132–145.

Frumin, Michael and Jinhua Zhao (2012). "Analyzing Passenger Incidence Behavior in Heterogeneous Transit Services Using Smartcard Data and Schedule-Based Assignment". In: *Transportation Research Record: Journal of the Transportation Research Board* 2274.1, pp. 52–60.

Fujii, Satoshi and Ryuichi Kitamura (2003). "What does a one-month free bus ticket do to habitual drivers? An experimental analysis of habit and attitude change". In: *Transportation* 30.1, pp. 81–95.

Gardner, Benjamin and Charles Abraham (2007). "What drives car use? A grounded theory analysis of commuters' reasons for driving". In: *Transportation Research Part F: Traffic Psychology and Behaviour* 10.3, pp. 187–200.

Gatersleben, Birgitta and David Uzzell (2007). "Affective Appraisals of the Daily Commute Comparing Perceptions of Drivers, Cyclists, Walkers, and Users of Public Transport". In: *Environment and Behavior* 39.3, pp. 416–431.

Gong, Hongmian et al. (2012). "A GPS/GIS method for travel mode detection in New York City". In: *Computers, Environment and Urban Systems* 36.2, pp. 131–139.

Gonzalez, P. et al. (2008). "Automating mode detection using neural networks and assisted GPS data collected using GPS-enabled mobile phones". In: *15th World Congress on Intelligent Transportation Systems*.

Gordon, Jason B. et al. (2013). "Automated Inference of Linked Transit Journeys in London Using Fare-Transaction and Vehicle Location Data". In: *Transportation Research Record: Journal of the Transportation Research Board* 2343, pp. 17–24.

Halpern, Scott D et al. (2011). "Lottery-Based versus Fixed Incentives to Increase Clinicians' Response to Surveys". In: *Health services research* 46.5, pp. 1663–1674.

Hansen, Flemming and Sverre Riis Christensen (2007). *Emotions, advertising and consumer choice*. Copenhagen Business School Press.

Hensher, David A, Peter Stopher, and Philip Bullock (2003a). "Service quality-developing a service quality index in the provision of commercial bus contracts". In: *Transportation Research Part A: Policy and Practice* 37.6, pp. 499–517.

Hensher, David A., Peter Stopher, and Philip Bullock (2003b). "Service quality—developing a service quality index in the provision of commercial bus contracts". In: *Transportation Research Part A: Policy and Practice* 37.6, pp. 499–517.

Homburg, Christian, Nicole Koschate, and Wayne D Hoyer (2006). "The role of cognition and affect in the formation of customer satisfaction: a dynamic perspective". In: *Journal of Marketing* 70.3, pp. 21–31.

Iseki, Hiroyuki, Brian D. Taylor, and Mark Miller (2006). *The Effects of Out-of-Vehicle Time on Travel Behavior: Implications for Transit Transfers*. Tech. rep. 65A0194. Sacramento, CA: California Department of Transportation.

Itsubo, Shinji and Eiji Hato (2006). "Effectiveness of household travel survey using GPS-equipped cell phones and Web diary: comparative study with paper-based travel survey". In: *Transportation Research Board 85th Annual Meeting*.

Jackson, W. and J. Jucker (Nov. 1982). "An Empirical Study of Travel Time Variability and Travel Choice Behavior". In: *Transportation Science* 16.4, pp. 460–475.

Jariyasunant, Jerald (2012). "Improving Traveler Information and Collecting Behavior Data with Smartphones". PhD thesis. University of California at Berkeley.

Jariyasunant, Jerald et al. (2014). "The Quantified Traveler: Travel Feedback Meets the Cloud to Change Behavior". In: *Journal of Intelligent Transportation Systems*. Publication forthcoming.

Ji, Y, Rabi G Mishalani, and Mark R McCord (2009). "Assignment-based Methodology to Match Automated Vehicle Location Transit Bus Trip Trajectories to Schedules". In: *Proceedings of the 11th International Conference on Advanced Systems for Public Transport, Hong Kong*.

Kahneman, Daniel, Peter P Wakker, and Rakesh Sarin (1997). "Back to Bentham? Explorations of experienced utility". In: *The quarterly journal of economics*, pp. 375–405.

Kahneman, Daniel et al. (1993). "When more pain is preferred to less: Adding a better end". In: *Psychological science* 4.6, pp. 401–405.

Kahneman, Daniel et al. (2004). "A survey method for characterizing daily life experience: The day reconstruction method". In: *Science* 306.5702, pp. 1776–1780.

Killingsworth, Matthew A and Daniel T Gilbert (2010). "A wandering mind is an unhappy mind". In: *Science* 330.6006, pp. 932–932.

Klöckner, Christian A and Ellen Matthies (2004). "How habits interfere with norm-directed behaviour: A normative decision-making model for travel mode choice". In: *Journal of Environmental Psychology* 24.3, pp. 319–327.

König, A and Kay W Axhausen (2002). "The reliability of the Transportation System and its influence on the Choice Behaviour". In: *Swiss Transport Research Conference (STRC), Monte Verità/Ascona.*

Kostakos, Vassilis, Tiago Camacho, and Claudio Mantero (2013). "Towards proximity-based passenger sensing on public transport buses". In: *Personal and Ubiquitous Computing* 17.8, pp. 1807–1816.

Lanken, Bas et al. (1994). "Attitude Versus General Habit: Antecedents of Travel Mode Choice1". In: *Journal of Applied Social Psychology* 24.4, pp. 285–300.

Laros, Fleur JM and Jan-Benedict EM Steenkamp (2005). "Emotions in consumer behavior: a hierarchical approach". In: *Journal of business Research* 58.10, pp. 1437–1445.

Li, Zheng, David A Hensher, and John M Rose (2010). "Willingness to pay for travel time reliability in passenger transport: A review and some new empirical evidence". In: *Transportation research part E: logistics and transportation review* 46.3, pp. 384–403.

Löchl, Michael (2005). "Stability of Travel Behaviour: Thurgau 2003". In: *Travel Survey Metadata Series* 16.

Lucas, Richard E (2007). "Adaptation and the Set-Point Model of Subjective Well-Being Does Happiness Change After Major Life Events?" In: *Current Directions in Psychological Science* 16.2, pp. 75–79.

Ma, Xiaolei et al. (2013). "Mining smart card data for transit riders' travel patterns". In: *Transportation Research Part C: Emerging Technologies* 36, pp. 1–12.

Mann, Eleanor and Charles Abraham (2006). "The role of affect in UK commuters' travel mode choices: An interpretative phenomenological analysis". In: *British Journal of Psychology* 97.2, pp. 155–176.

McFadden, Daniel (2001). "Economic choices". In: *American Economic Review*, pp. 351–378.

Mishalani, Rabi G, Mark M McCord, and John Wirtz (2006). "Passenger wait time perceptions at bus stops: empirical results and impact on evaluating Real-Time Bus Arrival Information". In: *Journal of Public Transportation* 9.2, p. 89.

Mittal, Vikas, Pankaj Kumar, and Michael Tsiros (1999). "Attribute-level performance, satisfaction, and behavioral intentions over time: a consumption-system approach". In: *The Journal of Marketing*, pp. 88–101.

Morfoulaki, Maria, Yannis Tyrinopoulos, and Georgia Aifadopoulou (2010). "Estimation of satisfied customers in public transport systems: a new methodological approach". In: *Transportation Research Record, Journal of the Transportation Research Forum* 46.1.

Müller, Meinard (2007). *Information Retrieval for Music and Motion*. Secaucus, NJ, USA: Springer. ISBN: 3540740473.

Munizaga, Marcela A and Carolina Palma (2012). "Estimation of a disaggregate multimodal public transport origin–destination matrix from passive smartcard data from Santiago, Chile". In: *Transportation Research Part C: Emerging Technologies* 24, pp. 9–18.

Nassir, Neema et al. (2011). "Transit Stop-Level Origin-Destination Estimation Through Use of Transit Schedule and Automated Data Collection System". In: *Transportation Research Record: Journal of the Transportation Research Board* 2263, pp. 140–150.

NextBus, Inc. (2013). *Public XML Feed Documentation.* URL: `http://www.nextbus.com/xmlFeedDocs/NextBusXMLFeed.pdf`.

Noland, R. and J.W. Polak (2002). "Travel time variability: A review of theoretical and empirical issues". In: *Transport Reviews* 22.1, pp. 39–54.

Noland, R. and K. Small (1995). *Travel-time uncertainty, departure time choice and the cost of the morning commute.* Tech. rep. Irvine, CA: University of California at Irvine.

Ohmori, Nobuaki, M Nakazato, and N Harata (2005). "GPS mobile phone-based activity diary survey". In: *Proceedings of the Eastern Asia Society for Transportation Studies.* Vol. 5, pp. 1104–1115.

Oliveira, Marcelo G Simas et al. (2011). "Global positioning system-assisted prompted recall household travel survey to support development of advanced travel model in Jerusalem, Israel". In: *Transportation Research Record: Journal of the Transportation Research Board* 2246.1, pp. 16–23.

Oliver, Richard L (1993). "Cognitive, affective, and attribute bases of the satisfaction response". In: *Journal of consumer research*, pp. 418–430.

— (2010). *Satisfaction: A behavioral perspective on the consumer.* ME Sharpe.

Ortuzar, Juan De Dios et al. (2011). "Continuous mobility surveys: the state of practice". In: *Transport Reviews* 31.3, pp. 293–312.

Ory, David T and Patricia L Mokhtarian (2005). "When is getting there half the fun? Modeling the liking for travel". In: *Transportation Research Part A: Policy and Practice* 39.2, pp. 97–123.

Parlak, Siddika, Jerald Jariyasunant, and Raja Sengupta (2012). "Using Smartphones to Perform Transportation Mode Determination at the Trip Level". In: *Transportation Research Board 91st Annual Meeting.* 12-3606.

Pedersen, Tore, Margareta Friman, and Per Kristensson (2011). "The role of predicted, on-line experienced and remembered satisfaction in current choice to use public transport services". In: *Journal of Retailing and Consumer Services* 18.5, pp. 471–475.

Perk, Victoria, Jennifer Flynn, and Joel M Volinski (2008). *Transit ridership, reliability and retention.* Tech. rep.

Perugini, Marco and Richard P Bagozzi (2001). "The role of desires and anticipated emotions in goal-directed behaviours: Broadening and deepening the theory of planned behaviour". In: *British Journal of Social Psychology* 40.1, pp. 79–98.

Peterson, Robert A and William R Wilson (1992). "Measuring customer satisfaction: fact and artifact". In: *Journal of the Academy of Marketing science* 20.1, pp. 61–71.

Quddus, Mohammed A, Washington Y Ochieng, and Robert B Noland (2007). "Current map-matching algorithms for transport applications: State-of-the art and future research directions". In: *Transportation Research Part C: Emerging Technologies* 15.5, pp. 312–328.

Rahbee, Adam B. (2008). "Farecard Passenger Flow Model at Chicago Transit Authority, Illinois". In: *Transportation Research Record: Journal of the Transportation Research Board* 2072, pp. 3–9.

Richins, Marsha L (1997). "Measuring emotions in the consumption experience". In: *Journal of consumer research* 24.2, pp. 127–146.

Rietveld, Piet, FR Bruinsma, and Daniel J van Vuuren (2001). "Coping with unreliability in public transport chains: A case study for Netherlands". In: *Transportation Research Part A: Policy and Practice* 35.6, pp. 539–559.

Roorda, Matthew J and Eric J Miller (2004). "Toronto activity panel survey: demonstrating the benefits of a multiple instrument panel survey". In: *7th international conference on travel methods, Costa Rica*.

Schlich, Robert, Anja Simma, and Kay W Axhausen (2004). "Determinanten des Freizeitverkehrs: Modellierung und empirische Befunde". In:

Schüssler, Nadine and Kay W Axhausen (2008). "Identifying trips and activities and their characteristics from GPS raw data without further information". In:

Seaborn, Catherine, John Attanucci, and Nigel HM Wilson (2009). "Analyzing multimodal public transport journeys in London with smart card fare payment data". In: *Transportation Research Record: Journal of the Transportation Research Board* 2121.1, pp. 55–62.

Simmons, Eleanor and Amanda Wilmot (2004). "Incentive payments on social surveys: A literature review". In: *Social survey methodology bulletin*, pp. 1–11.

Small, K. (1982). "The Scheduling of Consumer Activities : Work Trips". In: *The American Economic Review* 72.3, pp. 467–479.

Smith, Amy K and Ruth N Bolton (2002). "The effect of customers' emotional responses to service failures on their recovery effort evaluations and satisfaction judgments". In: *Journal of the Academy of Marketing Science* 30.1, pp. 5–23.

Stenneth, Leon et al. (2011). "Transportation mode detection using mobile phones and GIS information". In: *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, pp. 54–63.

Stopher, Peter, Eoin Clifford, and Manuel Montes (2008). "Variability of travel over multiple days: analysis of three panel waves". In: *Transportation Research Record: Journal of the Transportation Research Board* 2054.1, pp. 56–63.

Stopher, Peter and Andrew Collins (2005). "Conducting a GPS prompted recall survey over the internet". In: *Transportation Research Board Annual Meeting, 84th, 2005, Washington, DC, USA*.

Stopher, Peter, Camden FitzGerald, and Min Xu (2007). "Assessing the accuracy of the Sydney Household Travel Survey with GPS". In: *Transportation* 34.6, pp. 723–741.

Stradling, Stephen et al. (2007). "Passenger perceptions and the ideal urban bus journey experience". In: *Transport Policy* 14.4, pp. 283–292.

Stuart, Kenneth R, Marc Mednick, and Johanna Bockman (2000). "Structural equation model of customer satisfaction for the New York City subway system". In: *Transportation Research Record: Journal of the Transportation Research Board* 1735.1, pp. 133–137.

Sun, Yanshuo and Ruihua Xu (2012). "Rail Transit Travel Time Reliability and Estimation of Passenger Route Choice Behavior". In: *Transportation Research Record: Journal of the Transportation Research Board* 2275.1, pp. 58–67.

Susilo, Yusak O and Oded Cats (2014). "Exploring key determinants of travel satisfaction for multi-modal trips by different traveler groups". In: *Transportation Research Part A: Policy and Practice* 67, pp. 366–380.

Swanson, J, L Ampt, and P Jones (1997). "Measuring bus passenger preferences". In: *Traffic engineering & control* 38.6, pp. 330–336.

Thiagarajan, Arvind et al. (2010). "Cooperative transit tracking using smart-phones". In: *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems*. ACM, pp. 85–98.

Thompson, David A et al. (1996). "Effects of actual waiting time, perceived waiting time, information delivery, and expressive quality on patient satisfaction in the emergency department". In: *Annals of emergency medicine* 28.6, pp. 657–665.

Tooley, Melissa S (1996). "Incentives and rates of return for travel surveys". In: *Transportation Research Record: Journal of the Transportation Research Board* 1551.1, pp. 67–73.

Trépanier, Martin, Khandker MN Habib, and Catherine Morency (2012). "Are transit users loyal? Revelations from a hazard model based on smart card data". In: *Canadian Journal of Civil Engineering* 39.6, pp. 610–618.

Tyrinopoulos, Yannis and Constantinos Antoniou (2008). "Public transit user satisfaction: Variability and policy implications". In: *Transport Policy* 15.4, pp. 260–272.

Vij, Akshay, Andre Carrel, and Joan L Walker (2013). "Incorporating the influence of latent modal preferences on travel mode choice behavior". In: *Transportation Research Part A: Policy and Practice* 54, pp. 164–178.

Walker, Joan Leslie (2001). "Extended discrete choice models: integrated framework, flexible error structures, and latent variables". PhD thesis. Massachusetts Institute of Technology.

Wallin Andreassen, Tor (1995). "(Dis) satisfaction with public services: the case of public transportation". In: *Journal of Services Marketing* 9.5, pp. 30–41.

Wang, Wei, John Attanucci, and Nigel H.M. Wilson (2011). "Bus Passenger Origin-Destination Estimation and Related Analyses Using Automated Data Collection Systems". In: *Journal of Public Transportation* 14.4, pp. 140–150.

Wardman, Mark (2004). "Public transport values of time". In: *Transport Policy* 11.4, pp. 363–377.

Watkins, Kari Edison et al. (2011). "Where Is My Bus? Impact of mobile real-time information on the perceived and actual wait time of transit riders". In: *Transportation Research Part A: Policy and Practice* 45.8, pp. 839–848.

Wener, Richard E et al. (2003). "Running for the 7: 45: The effects of public transit improvements on commuter stress". In: *Transportation* 30.2, pp. 203–220.

Wolf, Jean et al. (2004). "Applications of new technologies in travel surveys". In: *7th International Conference on Travel Survey Methods, Costa Rica*.

Woodruff, Robert B, Ernest R Cadotte, and Roger L Jenkins (1983). "Modeling consumer satisfaction processes using experience-based norms". In: *Journal of marketing research*, pp. 296–304.

Yáñez, María Francisca, Patricio Mansilla, and Juan de Dios Ortúzar (2010). "The Santiago Panel: measuring the effects of implementing Transantiago". In: *Transportation* 37.1, pp. 125–149.

Yuan, Nicholas Jing et al. (2013). "Reconstructing Individual Mobility from Smart Card Transactions: A Space Alignment Approach". In: *Proceedings of the 13th IEEE International Conference on Data Mining*. Dallas, USA.

Zhao, Jinhua, Adam Rahbee, and Nigel H.M. Wilson (2007). "Estimating a Rail Passenger Trip Origin-Destination Matrix Using Automatic Data Collection Systems". In: *Computer-Aided Civil and Infrastructure Engineering* 22.5, 376–387.

Zheng, Yu et al. (2008). "Learning transportation mode from raw gps data for geographic applications on the web". In: *Proceedings of the 17th international conference on World Wide Web*. ACM, 247–256.

Zhou, Pengfei, Yuanqing Zheng, and Mo Li (2012). "How long to wait?: predicting bus arrival time with mobile phone based participatory sensing". In: *Proceedings of the 10th international conference on mobile systems, applications, and services*. ACM, 379–392.