# In Pursuit of the Happy Transit Rider: Dissecting Satisfaction Using Daily Surveys and Tracking Data

May 25, 2015

Working Paper

Andre Carrel*
Department of Civil and Environmental Engineering
University of California, Berkeley
109 McLaughlin Hall, Berkeley, CA 94720
Tel: +1-510-642-3585, Email: acarrel@berkeley.edu

Rabi G. Mishalani
Department of Civil, Environmental and Geodetic Engineering
The Ohio State University
mishalani@osu.edu

Raja Sengupta
Department of Civil and Environmental Engineering
University of California, Berkeley
rajasengupta@berkeley.edu

Joan L. Walker
Department of Civil and Environmental Engineering
University of California, Berkeley
joanwalker@berkeley.edu

* *Corresponding author*

## Abstract

This paper demonstrates the power and value of connecting satisfaction surveys from public transportation passengers to smartphone tracking data and automatic vehicle location (AVL) data. The high resolution of the smartphone location data allows travel times to be dissected into their individual components, and the connection with AVL data provides objective information on personal-level experiences of the respondents. Analyses show how these data can provide a quantitative understanding of the relationship between planned and provisioned service, and customer satisfaction. The addition of unreliability to the measurement of travel times, which is enabled by the highly detailed tracking data, shows that the relationship between passenger satisfaction and experienced travel times may be more nuanced than has previously been acknowledged. Ordinal logit model estimation results show a strong sensitivity of passenger satisfaction toward in-vehicle delays, and show that delays on board metro trains are perceived as more onerous than delays on board buses. This study also reveals the importance of obtaining a general measurement of satisfaction with transit service when repeated satisfaction measurements are conducted with respect to individual experiences. A baseline satisfaction level and a variable component as a function of experiences can be observed in the model results. Furthermore, the survey data include a measure of subjective well-being, which is a relatively new element in travel surveys. Insights are presented on the importance of this potential new covariate for future survey designs.

# 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, 2003). 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 paper, 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 paper presents results from a large study that measured participants' travel times via smartphone location tracking and automatic transit vehicle location (AVL) systems, while users' satisfaction with the experienced travel times was collected with smartphone-based 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 2 features a literature review, followed by the presentation of the modeling framework and the explanatory variables in section 3. Section 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.

# 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 authors, 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 paper 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 paper 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 paper 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).

# 3 Data and Modeling framework

## 3.1 Data

The data were collected during the six-week San Francisco Travel Quality Study, a large-scale study of transit service quality, which ran from October to December 2013. It involved an initial total of 856 participants recruited from the general public in San Francisco. Recruitment channels were email lists of companies and universities in San Francisco, postcards to holders of residential parking permits and fliers distributed at a large university (San Francisco State University). As an incentive, they received a free one-month transit pass, valid for unlimited travel on the San Francisco transit network. Participants were asked to complete an online entry survey in which sociodemographic and mode access information was collected as well as an exit survey at the end of the study. Furthermore, after completing the entry survey, they were asked to download a survey app for Android phones. The app was location-aware, and participants were instructed to keep location services enabled on their phones. If location services were enabled, the app collected location information from the phone every 30 seconds. Once per day at a time set by the participants (generally in the evenings), they received a survey prompt on their phones. If they responded to the prompt, the app asked them whether they had used transit on that day. If they responded yes, they were presented with a survey on their phone with a total of five screens with the following questions:

1. To what degree did they feel that their transit travel experience was pleasing, frustrating, relaxing, or made them impatient?

2. How satisfied were they with the IVTT, wait times, transfer times (if applicable) and overall reliability experienced during that day?

3. How satisfied were they with the crowding, cleanliness, safety and the pleasantness of other passengers on that day?

4. Did they use real-time information that day, and if yes, how satisfied were they with the accuracy of it? On the same screen, they were also asked how they generally felt that day.

5. Did they have any additional comments? A free-form text box was provided for that purpose.

Satisfaction was measured on a five-point Likert scale which was labeled at the extremes with a frowny face and a smiley face. This paper focuses exclusively on the satisfaction with travel times. Since the daily mobile surveys were filled out only once per day, they referred to all transit trips made by the participant on that day, regardless of the number. Respondents were asked to use Muni on at least five days during the study period and fill out the corresponding mobile surveys. The once-per-day format was chosen to survey participants as closely to the experience as possible while being mindful of the respondent burden. The tracking data were first classified into trips and activities via a clustering algorithm. The location data from trips made by the participants were then matched to the transit vehicle location data to automatically identify whether the participant had used transit on that day and to extract wait times, IVTT and transfer times. This provided objective measurements of the transit travel times experienced by the participants. The measured travel times were then compared to timetable information to identify deviations from the timetable. A complete 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. This corresponds to response rates of 70.5% and 87.9%, respectively. More information on the recruitment channels, study population and the survey app as well as screenshots can be found in Carrel, Sengupta, and Walker (In Review), and a more in-depth description of the data matching can be found in Carrel et al. (2015).

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. More information on the sample sizes can be found in table 3. Overall, the participants were 48% female and 52% male. 19% of them were ages 18-24, 48% ages 25-34, 19% ages 35-44, 9% ages 45-54, 4% ages 55-64 and the remaining 1% over the age of 64. Compared to the general profile of transit riders in San Francisco, the 25-35 age range is overrepresented in this study (the official rider survey reports a share of 31%), and the age ranges above 45 are underrepresented (14% in the study, compared to 32% in the official rider survey). 70% of the participants were either full-time or part-time employed, and 25% commuted into or out of the city regularly, either for work or school. 47% of all participants owned a bike in usable condition, 27% had a personal car, and 22% shared a car with other household members. 11% stated that they did not drive the cars in their household, and the remaining 40% did not have any private cars in their 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). Before the study, 60% typically used Muni between 2 and 5 days per week, 27% more than 5 days per week and the remaining 11% less than 2 days per week. 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 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.

## 3.2 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 $\tau$ are estimated as the dividers between intervals. Thus, the probability of choosing response $k$ on the Likert scale is given by:

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

In the above equation, $V$ is the systematic utility and $\mu$ is a scale parameter.

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

### 3.3 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: 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 2 shows an overview of the explanatory variables along with the variable names that are used to present the results later.

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.

| *Variable name* | *Units* | *Explanation* | *Entry/Daily* |
|---|---|---|---|
| | | ***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 the city 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 a 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 but no data were available | Daily |
| 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 2:** Overview of explanatory variables used in the ordinal logit models. See text for more details.

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 4.3.3 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, 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 Estimation results

## 4.1 Model estimation

Table 4 presents the estimation results for the four models. The models were specified as mixed logit models, i.e., with an additional mixing coefficient $\sigma$ to account for correlation between the responses given by a single individual. $\sigma$ is also shown in table 2, as well as the adjusted $\bar{\rho}^2$ as a goodness of fit measure. The latter is defined as follows:

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

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 3.2, 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":

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 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 sections discuss the estimation results: Section 4.2 focuses on the baseline satisfaction, sociodemographic, mode access and subjective well-being variables, and in section 4.3, the trip-specific variables such as IVTT, transfer and wait times, reliability and the accuracy

| | 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 3:** Overview of data used for model estimation.

of real-time information are discussed. The sensitivity analysis in section 4.4 shows in graphical form what percentage of participants were satisfied as a function of travel times.

## 4.2 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

| | 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 |
| $\sigma$ | 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:** Estimation results for the ordinal logit models of traveler satisfaction.

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.

## 4.3 Trip-specific variables

### 4.3.1 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.

### 4.3.2 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 3.3. 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.

#### 4.3.3 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.

#### 4.3.4 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 3.3. 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.

#### 4.3.5 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, 2003). 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.

### 4.4 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 1 shows the simulated distribution of satisfaction among respondents. As found in section 4, 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 explained in section 4.1. 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.

The ordinal logit model estimated in section 4 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.

## 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 paper, 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

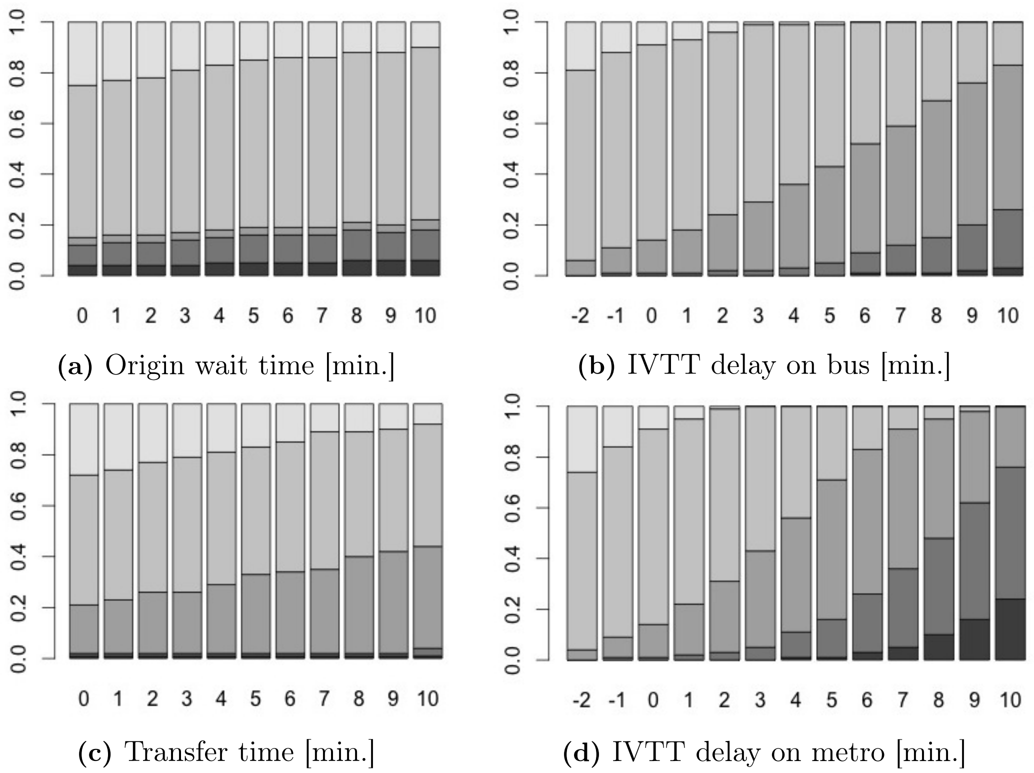


**Figure 1:** Satisfaction vs. travel times, from very dissatisfied (darkest) to very satisfied (lightest).

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. (2001) 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 of the authors, 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, the authors suggest that it is preferable 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 paper 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 Carrel, Sengupta, and Walker (In Review).

There are several additional possible avenues for future research. First of all, there is a need to verify and corroborate the findings of this paper 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 paper 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 3.3. 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 paper may support the direct calculation of the impact of reliability improvementsfon 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.

## Acknowledgments

The authors would like to express their gratitude to the National Science Foundation, the University of California Multicampus Research Programs and Initiatives and the University of California Transportation Center for providing funding for this project. Furthermore, they are grateful to the San Francisco Municipal Transportation Agency and the San Francisco County Transportation Authority for their support and collaboration in organizing the study as well as for providing the incentives to participants.

## References

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

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.

Bates, J. et al. (2001). "The valuation of reliability for personal travel". In: *Transportation Research* 37, 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.

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

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.

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

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* 2351.1, pp. 153–162.

Carrel, Andre, Raja Sengupta, and Joan L. Walker (In Review). "The San Francisco Travel Quality Study: Tracking Trials and Tribulations of a Transit Taker". In: *Transportation.*

Carrel, Andre et al. (2015). "Quantifying transit travel experiences from the users' perspective with high-resolution smartphone and vehicle location data: Methodologies, validation, and example analyses". In: *Transportation Research Part C: Emerging Technologies.*

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.

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.

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.
Dios Ortúzar, Juan de and Luis G Willumsen (2011). *Modelling transport*. John Wiley & Sons.
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.
– (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.
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.
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.
Hensher, David A, Peter Stopher, and Philip Bullock (2003). “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.
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.
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.
Kahneman, Daniel et al. (2004). “A survey method for characterizing daily life experience: The day reconstruction method”. In: *Science* 306.5702, pp. 1776–1780.

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*.

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.

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.

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.

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.

Oliver, Richard L (2010). *Satisfaction: A behavioral perspective on the consumer*. ME Sharpe.

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.

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.

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.

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.

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

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.

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.

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.

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

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

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.

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.