Estimating Demand for New Modes of Transportation Using a Context-Aware Stated Preference Survey

by

Nathanael Christopher James Cox

Bachelor of Actuarial Studies
Bachelor of Economics with First Class Honours
The Australian National University, 2011

Submitted to the Department of Civil and Environmental Engineering
in partial fulfillment of the requirements for the degree of

Master of Science in Transportation

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2015

Author .... Signature redacted ....
Department of Civil and Environmental Engineering
May 21, 2015

Certified by .... Signature redacted .....
Moshe Ben-Akiva
Edmund K. Turner Professor of Civil and Environmental Engineering
Thesis Supervisor

Accepted by .... Signature redacted ....
Heidi Nepf
Donald and Martha Harleman Professor of Civil and Environmental Engineering
Chair, Graduate Program Committee

Estimating Demand for New Modes of Transportation Using a Context-Aware Stated Preference Survey

by

Nathanael Christopher James Cox

Submitted to the Department of Civil and Environmental Engineering
on May 21, 2015, in partial fulfillment of the
requirements for the degree of
Master of Science in Transportation

Abstract

This thesis presents the design of a context-aware stated preference survey that will be used to estimate the demand for new transportation modes and services. It builds on the Future Mobility Survey, a smartphone-based prompted-recall survey that accurately gathers revealed preference information on respondents' travel patterns. By using this GPS data as the context for a hypothetical stated preference survey, we can present realistic travel scenarios to respondents that pivot off their actual behavior. The approach is the first of its kind to combine GPS and external data to generate hypothetical scenarios for a large number of modes. It does this by making use of freely available web services to gather information on travel times and distances on many modes, which then informs the presentation of these modes in the hypothetical scenario. The travel scenario is presented using a web interface that mimics trip-planning software, and the software can be readily applied across different cities and countries.

Thesis Supervisor: Moshe Ben-Akiva
Title: Edmund K. Turner Professor of Civil and Environmental Engineering

Acknowledgments

First and foremost, I am grateful to my supervisor, Professor Moshe Ben-Akiva, for his guidance through these two years, both in teaching and research. Though very challenging at times, I have had the good fortune to be able to work on several different projects during my time here. Each one of them has taught me something new, and each represents a unique, innovative effort to improve the transportation systems of tomorrow. You have always challenged me to think about new and better ways of doing research, and I hope to always keep that determination to improve and innovate as I graduate from MIT.

Many others in the ITS Lab and in the SMART team in Singapore helped me along the way as well: Bilge Atasoy, Jorge Santos, Inês Dias, Maya Abou-Zeid, Fang Zhao, Michael Craig, Tomer Toledo, Peiguang Hu and Haizheng Zhang. Thanks too to Eunice Kim and Katherine Rosa for your encouragement through the process - as we say in the Truckers@MIT promotional video, keep on truckin'![1] Thank you to Cintra and Fujitsu for sponsoring my first year at MIT, and to the Singapore-MIT Alliance for Research and Technology (SMART) for my second year, the result of which is this thesis.

Many of my fellow students gave me the sanity to keep on truckin', providing many experiences along the way: everything from a long-distance bicycle ride to a suburban fried chicken joint, to saving a kayak abandoned by the naval architecture department, to 1.258 assignment parties, to 'sad grad student dinners' at the student centre. Thanks to Emily Gates, Dave Maltzan, Sam Levy, Lauren Tarte, Dan Wood, Rebecca Heywood, Anne Halvorsen, Lauren Alexander, Zach Needell, Dan Fry, Peiguang Hu and Haizheng Zhang for making the experience memorable socially as well as academically.

Thank you my parents, Greg and Donna, who always instilled in me the value

[1] https://www.youtube.com/watch?feature=player_embedded&v=_TDbmYJnTMU

of education and always gave me the freedom to choose my own way in life. You named me after Nathanael Herreshoff, an MIT mechanical engineering graduate of the class of 1870. I don't think any of us ever imagined I would follow in those footsteps and move halfway across the world, but here I am.

Finally, thank you to my fiancé Jesse for your constant support and for always being a telephone call away. I'm very glad the days of Amtrak's Northeast Regional are coming to a close.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In this thesis, we present a way to estimate the demand for new transportation modes and services using a context-aware stated preference survey.

Information technology continues to reshape transportation and spawn new modes that would be unimaginable without the use of web technologies and smart phones. Flexible transit services, once the realm of the mobility-impaired and the outer suburban, are being revisited and repurposed as mainstream urban transportation options. Bike sharing has been made viable by docking stations that communicate with operators and users. And one-way car sharing is beginning to be offered in more and more cities, enabled by an ability to easily reserve a space at the destination via smartphone.

At the same time, resurgent interest and population growth in cities, especially in the developing world, has put pressure on the transport system to accommodate this growth. And in many developed countries, fiscal restraint and liveability concerns constrain the ability to accommodate more cars.

Many of the current new modes of transportation aim to address the significant loss in mobility that results from giving up car ownership. In many cities, people are often forced to choose between two extremes: car ownership, which is expensive but very convenient, and relying on public transportation, which is much cheaper but less convenient, particularly in much of North America. If the mobility gap of car ownership can be filled by other new services that are both reasonably priced and flexible, then the barriers to not owning a car will fall, with positive consequences for urban mobility and liveability.

Knowing how much new modes will change demand and mobility patterns is of great concern to policymakers and transport operators alike. Estimating demand for new modes is typically done using stated preference surveys, which present hypothetical scenarios to respondents. These hypothetical scenarios ask respondents to weigh up the travel times and costs of existing and new modes of transportation, and select the one they would use to travel. Stated preference methods have been in use since the 1980s, and have been used to estimate demand for a wide variety of new products and services, both in transportation and other areas of market research.

Technology has changed the way transportation works in our cities, and it has also influenced the way transportation surveys are conducted. In the 1990s and 2000s, GPS technology began to be used to conduct transportation surveys, providing a highly accurate record of where respondents had been (Greaves et al. 2010). Until very recently, however, the use of GPS has mostly been limited to *revealed preference* surveys, which provide a record of how respondents used existing modes of transportation under existing conditions.

Now, GPS technology is beginning to be used to inform stated preference surveys, by providing hypothetical scenarios that appear more realistic given the context of a respondent's decision (see, for example, Fifer et al. (2014)). The survey in this thesis is another step in this direction. It builds on an existing GPS-based revealed preference survey, picking out trips that form a baseline for a stated preference survey. It takes the trip origin and destination and inputs them into web services, fetching information on distances and travel times for various modes. That information is then used as a baseline to design scenarios that are realistic for the respondent.

This survey makes three primary contributions over existing stated preference surveys that use tables to present static information. First, it is context-aware, meaning it presents scenarios that respondents can engage with. Second, the use of readily available web services means it can be very easily applied simultaneously across multiple cities, regions and even countries. Third, the user interface is able to provide information on a very large number of modes in a way that reduces the cognitive burden on the respondent.

1.1 Thesis Organization

The remainder of this thesis is organized into eight chapters. Chapter 2 discusses the use of models and surveys to estimate the demand for transportation. Chapter 3 discusses how GPS technology has been used to improve travel surveys by increasing the realism of the data collected. Chapter 4 details some newly emerging transportation modes and services that surveys can be used to test demand for. In Chapter 5, we describe the Future Mobility Survey, a smartphone-based prompted-recall survey that collects accurate information on respondents' travel behavior. Chapter 6 presents a stated preference survey that uses the Future Mobility Survey to provide context and enhance realism. The design for a pilot survey is outlined in Chapter 7, and Chapter 8 discusses possible future enhancements that could be added to the pilot. Chapter 9 concludes.

Chapter 2

Estimating Transportation Demand Using Surveys

There are two ways transportation researchers can get information on consumers' preferences for modes: revealed preference (RP) or stated preference (SP) surveys. There is one fundamental difference between the two: an RP survey asks a traveller what they actually did, while an SP survey asks them what they would do in a hypothetical situation. This chapter discusses mode choice models and the two types of survey, with a particular emphasis on SP surveys.

2.1 Mode Choice Models

Before moving to survey methods, it is useful to set out precisely what these methods are trying to achieve, and what we mean when we say we want to understand demand and consumer preferences.

In the standard consumer utility maximization problem (as represented, for example, in Nicholson and Snyder (2007)), there is some collection of available goods, say $\{1, \ldots, N\}$. Letting the quantities consumed be $x_1, \ldots, x_N$, a consumer's preferences are represented by a utility function $U(x_1, \ldots, x_N)$. This utility function is *ordinal*, meaning that its absolute values do not matter, but that the consumer prefers higher values. In other words, if there is a consumption bundle $X^1 = \{x_1^1, \ldots, x_N^1\}$ and $X^2 = \{x_1^2, \ldots, x_N^2\}$, and the consumer prefers X^1 to X^2 ($X^1 \succ X^2$), then the utility is higher: $U(X^1) > U(X^2)$.

The consumer is limited by a fixed amount of income, I, and faces a fixed set of prices, $\{p_1, \ldots, p_N\}$. Their goal is to maximize their utility, subject to not exceeding their budget constraint. Algebraically, the problem becomes:

$$\max_{\{x_1,\ldots,x_N\}} U(x_1, \ldots, x_N)$$
$$\text{s.t.} \sum_{i=1}^{N} p_i x_i \leq I$$

The solution to this problem, $X^* = \{x_1, \ldots, x_n\}$, becomes the consumer's demand for goods 1 through n. If we had a market of many consumers, and knew every consumer's utility function and income, then solved the problem for each consumer, we can add these quantities to find the total market demand.

For transportation mode choice, and indeed many other common consumer situations, the context is very different. Under the classical model in Nicholson and Snyder (2007), only the quantity of goods is what matters. But if a consumer is traveling and chooses different modes, they are ultimately only consuming one trip. Clearly, though, there are differences in utility that lead certain modes to be chosen over others. This led to the idea that, in these cases, it is *attributes* that determine utility, and these attributes come from goods (Ben-Akiva and Lerman 1985).

In transportation, many different attributes can influence the utility of certain modes. The two most prominent of these are the travel time and cost of the mode, but many more exist. In public transportation, for example, the total travel time can be split into walking, waiting and in-vehicle time components, each with its own separate effect on utility. Many more attributes can be used to explain transportation choices, including those that are not easily quantifiable, such as travel time uncertainty and the level of crowding in a transit vehicle.

In a discrete choice framework, a consumer (n) is faced with a choice set $C_n = \{1, \ldots, J_n\}$, where J_n is the number of alternatives, and must select a single alternative from this choice set. Their utility function, U, is now defined over C_n, and depends not only on the attributes of each mode, but also on their own personal socio-economic *characteristics* - for example, higher-income people might have a greater preference for cars over other modes, all else equal. Defining consumer n's utility of alternative i as U_{in}, the consumer chooses the alternative that has the highest utility value, i.e. the problem is $\max_i U_{in}$.

More formally, we can model utility as having a systematic component (V_{in}) and random component (ε_{in}): $U_{in} = V_{in} + \varepsilon_{in}$. The systematic part represents the effects of mode attributes and personal characteristics (k); calling these values x_{ink}, we can use a simple linear specification such as $V_{in} = \sum_k \beta_k x_{ink}$. The random component, ε_{in}, represents the effect on utility of any missing attributes we do not explicitly account for in V_{in}.

For example, with a simple choice set of {Car, Bus} and variables Time and Cost, we can write:

$$U_{\text{Car},n} = \beta_{\text{Car}} + \beta_{\text{Cost}} \times \text{Cost}_{\text{Car},n} + \beta_{\text{Time}} \times \text{Time}_{\text{Car},n} + \varepsilon_{\text{Car},n}$$
$$U_{\text{Bus},n} = \beta_{\text{Cost}} \times \text{Cost}_{\text{Bus},n} + \beta_{\text{Time}} \times \text{Time}_{\text{Bus},n} + \varepsilon_{\text{Bus},n}$$

In this case, the consumer would choose Car if $U_{\text{Car},n} > U_{\text{Bus},n}$, and Bus otherwise. Because the ε_{in}s are random, however, in the model the choices only happen with a certain probability.

Assuming a distribution on the random components, in combination with data on the characteristics, attributes and actual choices, allows us to use maximum likelihood estimation to estimate the values of the β parameters in the model. The most commonly assumed distribution is the extreme value distribution (with location parameter zero and scale parameter μ). This produces the logit model, where

$$Pr(i \mid C_n) = \frac{e^{\mu V_{in}}}{\sum_{j \in C_n} e^{\mu V_{jn}}}.$$

The logit model is part of a family of discrete choice models that are too numerous to describe here. Interested readers should consult Ben-Akiva and Lerman (1985) for further details.

The estimates of β represent the shape of the utility function: these are consumers' preferences. These values can be used to forecast demand through various methods, including aggregate forecasting, sample enumeration and microsimulation (Ben-Akiva and Lerman 1985). To do this requires having some population p with N_p members, where we know every member's individual characteristics and the attributes of the modes they can choose from.

There are two ways travel data can be collected; the difference between them is the context the respondent makes their decision in. Revealed preference surveys collect travel data in 'the world as it is', while stated preference surveys present 'the world as it could be' to elicit choice behavior (Louviere et al. 2000). The next sections discuss the two types of survey, along with their advantages and disadvantages.

2.2 Revealed Preference Surveys

Revealed preference surveys capture the actual travel behavior of a person or household in a fixed time period. By collecting information on the times and costs of every trip alternative and what the respondent actually chose, we can estimate a behavioral model that indicates the underlying preferences of the consumer.

One way of collecting this information is through a diary, where the respondent enters in all their trip information at the end of each day. Alternatively, participants can be interviewed by telephone on a regular basis, which can be helpful in ensuring the responses collected are sensible.

Revealed preference surveys are commonly used to estimate regional travel demand models. One example is the Sydney Household Travel Survey, which covers the metropolitan area of Sydney, Newcastle and Wollongong, Australia (Transport for New South Wales 2015). This survey has been running continuously since 1998 and currently covers around 3,000 households per year. Telephone interviewers collect socio-economic information and collect information on the trips taken in the household over a day. Along with other inputs, such as Journey to Work data from the Australian Census, it is used to estimate and update the Sydney Strategic Travel Model (Bureau of Transport Statistics 2013). Part of this involves estimating a mode choice model, which has seven modes: car driver, car passenger, train/light rail/ferry, bus, bike, walk and taxi.

Morikawa (1989) summarizes the characteristics of revealed preference data. The main advantage of using revealed preference data is that it is based on actual choices made by respondents in real-life situations. However, in a revealed preference survey we typically only observe what the choice was: this means we cannot be sure what the choice set was. Further, attributes (such as travel cost and time) can be strongly correlated, which makes estimating their independent effects on utility less precise. The range of the attributes is also limited to what the range is in reality and may be subject to measurement error, potentially reducing the efficiency of estimation. Finally, intangible attributes (such as reliability and comfort) are difficult to incorporate into the model, because they cannot be easily measured by the researcher.

The biggest disadvantage of revealed preference data for estimating the demand for new modes is respondents make decisions in 'the world as it is' (Louviere et al. 2000): accordingly, it cannot provide information on modes that don't yet exist. We therefore cannot use revealed preference surveys to estimate the demand for new modes, and instead must use stated preference surveys to present the new mode to respondents and observe their behavior.

2.3 Stated Preference Surveys

A stated preference survey asks the respondent what they would do in a hypothetical scenario, as opposed to what they actually did in a real-world scenario. A revealed preference survey is typically one of observation, but a stated preference is more of an experiment: the entire process is controlled by the researcher.

Morikawa (1989) details how this control can be useful to researchers. First, we know the choice set, because the choices available to survey respondents are the choices the researcher presents. Second, we can specify the variable ranges to be wider than might be found in reality. Third, because the combinations of levels shown are chosen by the researcher, we can reduce multicollinearity between variables (say, between travel time and cost). Fourth, we can explicitly include 'soft' variables like comfort and reliability by presenting these to respondents. Finally, we choose what to present to respondents, so all variables are free from measurement error.

The most common shortcoming of stated preference data, however, is that it only records choices made in hypothetical scenarios (Fifer et al. 2014). This can result in biases in the data - for example, Murphy et al. (2005) analyzed 28 valuation studies that use stated preference methods to find willingness to pay. They compared the hypothetical willingness-to-pay values with actual values, finding the median of hypothetical-to-actual ratios to be around 1.35.

Fifer et al. (2014) report on similar biases in a transportation context, examining studies that compare estimates of the monetary value of time using stated and revealed preference data. They find that actual values of time tend to be higher in reality than in hypothetical experiments, but do not find this to be as conclusive as in valuation experiments.

Theis (2011) used a stated preference survey to study the effect of connection times on the utility of flights. The survey asked respondents various statements to gauge their preferences for short connections. It then asked them to choose an airline itinerary, modeling the effects of fares, frequent flyer status, preferred airline, number of connections, total travel time and 'buffer time' (connection time in excess of the minimum required by the airport). They found that connection times do not have a monotonic effect on utility, with longer connection times always being less preferred. Instead, flights with some buffer time were preferred to connections with no buffer time, though long buffer times are still less preferred. Further, the latent variables of risk aversion, rush aversion and lack of trust of the airline's scheduling (generated from the responses) increase this effect.

Many stated preference studies investigate the demand for new modes, since revealed preference surveys cannot be used before a new mode is introduced. One such study was for Swissmetro, a proposed high-speed maglev railway connecting the major cities of Switzerland (Bierlaire et al. 2001). They surveyed rail users and drivers who had taken an intercity trip in the corridor Swissmetro was proposed to run through. They presented three alternatives, driving, current rail and Swissmetro, with various times and costs, and asked respondents to choose one of the three. The final data set consisted of 770 surveys. They used several different model specifications to test the stability of their results, such as multinomial logit, nested logit and cross-nested logit. The value of travel time was estimated at SFR 1.15-1.21 per minute (SFR 69 - 73 per hour). Bierlaire et al. (2001) note this is high, but suggest it is because longer-distance travellers in Switzerland tend to be wealthier than average.

More recently, pivot-style stated preference experiments have become more popular (Hess and Rose 2009). These experiments take into account the survey respondent's knowledge and circumstances. This is typically done in one of three ways. These are: showing a status quo alternative with no attributes or other information; showing alternatives with levels based on the respondent's own experience, but not the exact levels; or including the status quo alternative with all the information. Hess and Rose (2009) note this approach is supported by theories in psychology and economic theory, which support relating experiments to reality as much as possible. Fifer et al. (2010) note that many transportation studies ask respondents to choose a 'typical' trip they make and this is used as a reference to inform the survey design, but this is problematic because people recall their trip details very poorly. However, Train and Wilson (2008) warn that using revealed preference data to construct stated choice experiments creates dependence between the stated preference attributed and unobserved factors. They caution that models estimated using data collected using pivoted designs should account for this dependence, though standard estimation procedures are consistent under certain conditions.

One such pivot-style approach is that of Fifer et al. (2010), which represents one of the first attempts to combine GPS data with a stated preference survey. Their study investigates how motorists would react if a distance-based charging system were introduced that incorporates the risk of driving. This risk of driving is assumed to depend on distance travelled, night-time travel and speeding. The study first monitored participants' driving behavior in Sydney, Australia, through

GPS for five weeks, then introduced the risk-based charging regime. They then answered a stated preference survey once, continued observation for five weeks, and finally filled out the stated preference survey a second time, with different attribute values.

The stated preference surveys in Fifer et al. (2010) show the participant's travel behavior in one column, which includes: the number of travel days; the distance travelled; the percentage of driving time done at night; the percentage of time spent speeding; and the charges that would result for this driving pattern (fixed at $100). They then showed two alternatives, each with its own distance, driving time of day, speeding percentage and charge, as well as the average travel time increase (per trip) if speeding were reduced to the amount shown. All these amounts (except for the travel time increase) were pivoted off the respondent's actual behavior: their values were calculated using percentages of the user's observed behavior. Their results show that drivers are willing to pay some amount not to change their current behavior, in terms of distance travelled, night travel and travel time. Further, they find this willingness to pay depends on the trip type: people are more willing to change their behavior for non-work trips. However, participants were willing to pay $1.68 on average to reduce speeding, though the authors put this down to the sample having negative perceptions of speeding.

2.4 Experimental Design

Once we have selected the variables to include in a stated preference survey, a related question is what combinations of levels should be presented. This section provides a brief overview of the main types of design, drawing on material in Louviere et al. (2000), which provides further details and examples.

Each variable in a stated preference survey has *levels*, which are the values it can take. When all variables are presented together to the respondent, the information presented is called a *profile.*

Suppose our stated choice experiment has factors (variables) $\{X_1, \ldots, X_k\}$ to be presented to survey respondents. Each of these factors has an associated set of levels. For each factor $i \in \{1, \ldots, k\}$, call this set $L_i = \{l_1, \ldots, l_{n_i}\}$, where n_i is the total number of levels for factor i. A profile is some combination of the levels in $\{X_1, \ldots, X_k\}$. Let there be m profiles, and denote profile $j \in \{1, \ldots, m\}$ as $P_j = \{x_1, \ldots, x_k\}$, where for each $i \in \{1, \ldots, k\}$, $x_i \in L_i$. The *experimental design* is simply the set of all profiles: $\{P_1, \ldots, P_m\}$.

For example, if we had two factors, X_1 and X_2, with level sets $L_1 = \{0, 1\}$ and $L_2 = \{2, 3\}$, one possible profile is $P_1 = \{0, 2\}$ and another is $P_2 = \{1, 3\}$. The experimental design is $\{\{0, 2\}, \{1, 3\}\}$.

Our task is to present profiles for the experiment that enable us to elicit the sample's underlying preferences towards the alternatives. Sections 2.4.1 through 2.4.4 detail ways of doing this.

2.4.1 Random Design

One of the simplest approaches to generating profiles is to pick the x_is from each set of levels L_i randomly. Typically, a uniform distribution is used, so that each level has an equal probability of being selected for inclusion in a profile. However, this need not be the case - any distribution can be used to generate profiles, though using a uniform distribution helps to provide a fairly even spread of level values through the design.

2.4.2 Full Factorial Design

Another approach is simply to use every possible combination of levels, which then becomes the set of profiles to use in the experiment. For example, consider a design where there are K factors, each with two levels. Then the full factorial design has 2^K profiles, each representing a unique combination of the levels of the K factors.[1]

Consider, for example, an experiment with three variables: car travel time, bus travel time and bus fare. Each variable can take the values 'high' or 'low'. Then the corresponding full factorial design (called a 2^3 design) is shown in Table 2.1.

	Attributes		
Profile	Car Travel Time	Bus Travel Time	Bus Fare
1	Low	Low	Low
2	Low	Low	High
3	Low	High	Low
4	Low	High	High
5	High	Low	Low
6	High	Low	High
7	High	High	Low
8	High	High	High

Table 2.1: A 2^3 experimental design.

Full factorial designs have two desirable properties: they are *balanced*, in the sense that all possible levels are equally represented across the design. They are also *orthogonal*, in that the sample correlation between any two variables (as presented in the experiment) is zero.

[1]More generally, when the number of levels varies across factors, the total number of possible profiles is $\prod_{i=1}^{k} n_i$.

	Levels per Factor			
Factors	2	3	4	5
1	2	3	4	5
2	4	9	16	25
3	8	27	64	125
4	16	81	256	625
5	32	243	1024	3125
6	64	729	4096	15625
7	128	2187	16384	78125
8	256	6561	65536	390625
9	512	19683	262144	1953125
10	1024	59049	1048576	9765625
11	2048	177147	4194304	48828125
12	4096	531441	16777216	244140625

Table 2.2: Number of profiles in a full factorial design, by number of factors and number of levels per factor.

2.4.3 Fractional Factorial Design

A fractional factorial design is any subset of a full factorial design. By carefully selecting Properly selected fractional factorial designs can dramatically reduce sample size requirements while still allowing parameters of interest to be estimated.

One problem with full factorial designs is that the number of profiles to present becomes infeasibly large very quickly. Table 2.2 illustrates this, showing the required number of profiles for a full factorial design as a function of the number of factors (rows) and the number of levels per factor (columns).

There are two types of effects we can estimate: *main effects* (the effect of variables independently of other variables' values) or *interactions* (where the effect of a variable is a function of other variables' values). For example, Sun (2013) modeled intercity truck drivers' route choices using the following utility function for route i, driver n and choice experiment t:

$$
\begin{aligned}
U_{int} = {} & \beta_{Downtown} + \beta_{Free} + \beta_{Time} \times Time_{int} + \beta_{Toll,n} \times Toll_{int} + \\
& \beta_{TollD,n} \times TollDummy_{int}(1 + \beta_{TollCompany} \times TollCompany_{int}) + \\
& \beta_{Delay} \times Delay_{int}(1 + \beta_{DelayHourly} \times DelayHourly_{int} + \\
& \beta_{DelayTemp} \times DelayTemp_{int}) + \alpha_i \epsilon_n + \epsilon_{int}
\end{aligned}
$$

The full description of the model and its variables can be found in pages 90 and 91 of Sun (2013). The specification includes several main effects of variables on the utility of a route - these show the effect of the variable considered independently of other variables. The terms indicating main effects are:

- $\beta_{Time} \times Time_{int}$: the effect of travel time on utility
- $\beta_{Toll,n} \times Toll_{int}$: the effect of the toll amount on utility
- $\beta_{TollD,n} \times TollDummy_{int}$: the effect of there being any (non-zero) toll on utility
- $\beta_{Delay} \times Delay_{int}$: the effect of the probability of delay on utility.

The terms measuring *two-way* interaction effects (involving two variables) are:

- $\beta_{Delay} \times Delay_{int} \times \beta_{DelayHourly} \times DelayHourly_{int}$: the additional effect of delay on utility, if the driver is paid hourly
- $\beta_{Delay} \times Delay_{int} \times \beta_{DelayTemp} \times DelayTemp_{int}$: the additional effect of delay on utility, if the shipment is temperature controlled.

Typically, the way a fractional factorial design is selected is by ignoring certain higher-order interactions. Consider, for example, an experiment with factors X_1, X_2 and X_3, each of which has possible levels 1 or -1. A full factorial (2^3) design is shown in Table 2.4.3. The yellow highlighted rows and the blue highlighted rows each represent one possible fractional factorial design (called a 2^{3-1} design in this case). Each ignores three-way interactions by having the same values of $X_1 \times X_2 \times X_3$, but halves the minimum sample size and still allows main effects and two-way interactions to be estimated. Sometimes, experiments are divided into *blocks* - groups of profiles to show to a single participant. This is sometimes done based on higher-order interactions, such as three-way interactions.

Profile	Main Effects			Interactions			
				Two-Way			Three-Way
	X_1	X_2	X_3	$X_1 \times X_2$	$X_1 \times X_3$	$X_2 \times X_3$	$X_1 \times X_2 \times X_3$
1	1	1	-1	1	-1	-1	-1
2	1	1	1	1	1	1	1
3	1	-1	-1	-1	-1	1	1
4	1	-1	1	-1	1	-1	-1
5	-1	1	-1	-1	1	-1	1
6	-1	1	1	-1	-1	1	-1
7	-1	-1	-1	1	1	1	-1
8	-1	-1	1	1	-1	-1	1

Table 2.3: A 2^3 full factorial design, showing main effects (columns 2-4) and interaction effects (columns 5-9).

2.4.4 Efficient Design

The efficiency of a design refers to the efficiency of an estimator applied when that experimental data is collected using that design, and used in the estimation process.

In the simple case of a linear model ($y = X\beta + \varepsilon$), the variance of the ordinary least squares estimator (conditional on X) is $\mathrm{Var}(\hat{\beta} \mid X) = \sigma^2(X'X)^{-1}$, where σ^2 is the variance of the error term in the model. More efficient designs will therefore have a 'small' $(X'X)^{-1}$ - the most popular measure of this is D-optimality, which measures $|X'X|$, the inverse of the determinant of $(X'X)^{-1}$. For a linear model, maximizing this measure by changing the composition of X will lead to more efficient parameter estimates.

For a non-linear model, however, this is not so straightforward, because the variance-covariance matrix of $\hat{\beta}$ depends not just on X, but also β. The most common way to address this issue is to assume a prior on β, which can either be a fixed value or a prior distribution with known parameters. Rose and Bliemer (2009) shows that these designs can, under certain circumstances, lead to greater estimation efficiency than orthogonal designs, and therefore reduce sample size requirements.

However, recent research suggests that efficient designs can be risky to use if the researcher does not have good priors for the parameters to be estimated. Walker et al. (2015) performed an experiment that generated different designs for travel time and travel cost. They then assumed a true value of time and a prior for the efficient design, using these and the designs to simulate a real experiment and estimate the value of time. They found that so-called 'efficient' designs do yield parameter estimates with less standard error when the true value of time is close to the assumed prior. But when it is not, these designs perform worse than random and orthogonal designs. Further, assuming priors for only some parameters and not others led to the worse possible design. They argue that a simple random design performs as well as any other design, particularly when it is 'cleaned': that is, profiles where one alternative strictly dominates another in time and cost terms are deleted.

Chapter 3

GPS-Based Travel Surveys

This chapter covers recent developments in GPS-based travel surveys. It begins with an overview of how GPS data can be collected and used in transportation, following up with recent examples of applications in travel surveys, both in an urban passenger context and an intercity freight context. It then describes the Future Mobility Survey, the smartphone-based application that forms the foundation for the stated-preference survey in this thesis.

3.1 Collection and Analysis of GPS Data in Transportation

Once the domain of the armed forces and surveyors, GPS has emerged in the last few decades as an inexpensive and reliable way of capturing location data. The collected information is typically very accurate, forming large data sets showing where the GPS device was in closely spaced intervals (Greaves et al. 2010).

In few industries has GPS data collection become as widespread as it has in transportation. Airplanes and ships use it to locate where they are and where they have been recently. Transit vehicles can be located precisely so that agencies know where they are and can inform customers of delays. Cars with navigation systems can combine the data with a map of the road network to find the quickest route to their destinations. And anyone with a smartphone can locate themselves on a map to see where they are, and use that location to get directions to where they need to go.

As the collection of GPS data grew, so too did interest in using that data to improve transportation systems. Many in-car navigation systems send anonymized location data back to their manufacturers, who can use it to calculate travel speeds and detect congestion. This information is then sent back to individual navigation systems, allowing them to direct drivers away from congested areas. Transit agencies can use GPS location in buses to detect when the bus is about to arrive at a stop, and provide an audible announcement for hearing-impaired passengers. Uber users can summon a ride using their smartphones and see how far away their driver is; they are charged according to a distance calculated by the driver's GPS trace. And GPS data collected by cyclists can be visualized to show what routes are preferred by commuters.[1]

3.2 Previous GPS-Based Travel Surveys

GPS technology has also allowed significant improvements in travel surveys. Traditionally, surveyors have relied on paper-based methods, in which users keep a diary of the locations they visited through the day. Typically, they also record how they got from place to place, who they travelled with and when they departed and arrived. In many cases, this places a significant burden on the respondent, who is not necessarily keeping track of every travel detail. This is especially true of short

[1] For one example, see the Rideable project (www.rideable.org.au), which periodically collects data uploaded by cyclists in Sydney, Australia, and illustrates the routes they chose for their morning commutes.

trips, which might be left out of the diary altogether. Some of this burden can be reduced by allowing survey respondents to fill out their diaries online, but this still requires them to remember everything they did through the day.

One of the first examples of using GPS for trip reporting is documented by Murakami and Wagner (1999). They used GPS-enabled handheld computers to collect and report travel information for the car trips of 100 households in Lexington, Kentucky. These households were then telephoned and asked to recall the trips they had made without the help of the GPS device. They found evidence of underreporting of trips in the telephone survey for around 35% of households, suggesting these households were not able to recall every trip they made. Around 30% households overreported trips, but this was attributed to issues with the equipment. The GPS-reported trips recorded more accurate start and end times than the telephone-reported trips, which tended to be rounded to the nearest half hour. And the self-reported travel distances respondents gave were significantly lower than those calculated using the GPS traces.

Zmud and Wolf (2003) delved into this issue further, identifying factors correlated with trip underreporting using a sample of around 300 households. They compared GPS-matched trips to user-reported trips, finding that 71% of all unreported trips had a duration of under 10 minutes. Underreporting is also correlated with a household's demographics: higher levels were observed in households with more vehicles, low incomes, young people or more workers. They attribute many of these differences to how much the households travel: if fewer trips are made, it is easier to recall all of them accurately.

Ohmori et al. (2005) extended this approach to an activity diary survey using flip phones with GPS functionality. Unlike in Murakami and Wagner (1999), the survey could be used for any type of trip, not just walking, and was activity based, meaning the sequence of trips could be tracked throughout the day. Like Murakami and Wagner (1999), they still maintained a paper survey for some participants to compare the two data sets. They found the GPS data was much easier to process versus a paper diary, and respondents filled out the phone survey both more frequently and sooner after their activities took place. They attribute the greater frequency and lower time lag to respondents being able to fill out the survey while traveling. However, some participants found the phone survey more difficult to use, and battery consumption proved to be problematic, with the survey software shutting off after about 5-6 hours due to low battery.

Stopher et al. (2007) used GPS technology and a stated-recall web survey to identify underreporting in the Sydney Household Travel Survey. Households were given GPS devices for every car they owned, as well as wearable GPS loggers. Once data were collected, households were telephoned and verified their trips verbally. After this, they logged into a web survey, where they were shown their GPS-detected trips for the day and asked to verify them. They found that the overall level of underreporting was about 7%, lower than in other studies. Further, households tended to understate travel distances and overstate travel times compared to the GPS data.

As GPS technology continued to improve and become cheaper to buy, larger-scale surveys became possible. Greaves et al. (2010) conducted a prompted-recall survey similar to that in Stopher et al. (2007), where 29 of the 30 recruits provided 8 weeks of validated trip data with relatively little time burden. In 2009, Giaimo et al. (2010) piloted the use of GPS in the Greater Cincinnati Household Travel Survey, again reporting adequate completion rates. This was followed up by a 2,000-household multi-day survey (in Stopher et al. (2012)), which proved the concept was scalable to large samples. Like other studies, they too found GPS-based trip rates to be significantly higher than in diary surveys. Their study also trialled using the GPS data to impute mode choices and stop purposes. This proved to be quite accurate - the imputed mode was correct 96 per cent of the time, and the imputed activity was correct 90 per cent of the time.

Simas Oliveira et al. (2011) took a similar approach in Jerusalem, Israel, collecting GPS data for around 3,000 households and supplementing this with a prompted recall, computer-based interview. They found that GPS is particularly useful for developing activity-based travel demand models, which analyze an individual's demand for patterns of activity as opposed to viewing individual trips in isolation. By collecting uninterrupted data throughout the day, GPS data captures these activity patterns without gaps or inconsistencies. As in other studies, they found the accuracy of the data and the survey response rates to be high.

3.3 The Future Mobility Survey

The Future Mobility Survey represents the next generation of GPS-based travel surveys. It relies on smartphones instead of dedicated GPS loggers, and forms the basis for the stated preference survey of this thesis.

The initial experience in implementing the Future Mobility Survey is documented in Cottrill et al. (2013). They put forward the case for smartphones as GPS loggers: GPS loggers are now relatively cheap, but giving them to participants and receiving them back is costly. Further, because most people own smartphones and carry them around most places they go, the burden of carrying a separate device just for the survey is eliminated.

The logging of GPS data happens using an application installed on the user's phone; there is no need to open or close the app through the day, as it runs in the background. At certain points in the day, the GPS data is sent back to the Future Mobility Survey server for processing. The server uses machine learning algorithms to infer the user's modes and stops. As in Stopher et al. (2012), the survey is a prompted recall survey, where the user validates their GPS movements and inferred stops at the end of each day. Users can add or delete stops they made.

Cottrill et al. (2013) focussed on the user interface in a small pilot study, with great attention being paid to an intuitive user interface that reflected the chain of trips and stops users made throughout the day. They found that the unobtrusive nature of the smartphone application means users need to be reminded to validate their data, since the application runs in the background and requires no interaction. Battery life tended to be poorer with the application running, due to the need to collect location data, and this varied greatly depending on the type of smartphone the user has. Overall, however, users tended to find the survey relatively simple to use.

Zhao et al. (2015) reports on the findings of the pilot, which has 793 completed users, who collected 22,170 days' worth of data, 7,856 of which were validated. Analyzing the sources of error, they found that while the application accurately reported locations, there were data gaps when smartphones run out of power and users do not properly validate any missing segments. Further, because users are not interviewed but instead fill out a web-based survey, validation errors can arise because the process is controlled by the user.

One potential issue with using smartphones to collect travel information is that only people who own smartphones can participate. This was recognized in Zhao et al. (2015), who found the age distribution of respondents was skewed towards young people. They suggest this can be rectified by sending GPS loggers to users who don't own smartphones and providing help for users who might have difficulty validating their data on a web browser.

Analyzing the validated data, Zhao et al. (2015) found that activity patterns vary greatly from user to user, and from day to day, implying users should be sampled across many days. The Future Mobility Survey provides a very convenient platform to do this: users are not interviewed by a person, so the cost of collecting additional days is minimal. Further, the burden of data validation reduces over time as the system learns travel patterns and users find they need only confirm the inferred information rather than choose from many options.

The Future Mobility Survey is described in greater detail in Chapter 5. This thesis will use the Future Mobility Survey as the basis for a context-dependent stated preference survey that can be used to estimate the demand for new modes.

3.4 Truckers@ MIT: A GPS-Based Route Choice Study for Intercity Trucking

Truckers@MIT is a deployment of GPS technology and a web-based survey for intercity trucking. It is based on a very similar platform to the Future Mobility Survey, with the exception that the data were captured using traditional GPS loggers instead of smartphones. These loggers remained in the participants' trucks for approximately one month, and 20 working days of data needed to be reported to successfully complete the experiment. As with the Future Mobility Survey, the drivers logged in to a web survey periodically to validate their travel behavior. Stops were detected, and drivers verified what they did there - for example, they could be loading or unloading goods, staying overnight (for multi-day trips) or taking a meal break.

The primary motivation behind this survey was to gather revealed preference data for intercity truck behavior in the United States and Canada. The predecessor to this study, documented in Sun (2013), intercepted truck drivers at various truck stops in Texas, Toronto and Indiana. There they conducted a stated preference survey that gave them alternative route options and asked them to choose between them. Sun (2013) found that truckers' route choice behavior was influenced by many more factors than time and cost. Further, they found that their values of time (and willingness to pay tolls) also varied widely: drivers paying the toll out of their own pockets were more sensitive to tolls, but those with temperature-controlled shipments were willing to pay more for time savings. And in some cases, drivers who were paid hourly actually had a negative value of time because they preferred to sit in congestion and be paid more.

In Truckers@MIT, which is documented in Ben-Akiva et al. (2015, forthcoming), we found significant variability in driver tour patterns and driver route choice behavior, information that is made much clearer when collected using GPS and plotted on a map.

For example, Figures 3-1 and 3-2 show a round trip made by a truck driver that originated north of Austin. In Figure 3-1, the driver is driving south towards Austin during the morning rush hour, and uses a circumferential toll road (Texas State Highway 130) to avoid traffic. In Figure 3-2, the driver returns north on a weekend morning, but instead chooses to avoid the toll and take the free route back through Austin: Interstate 35.

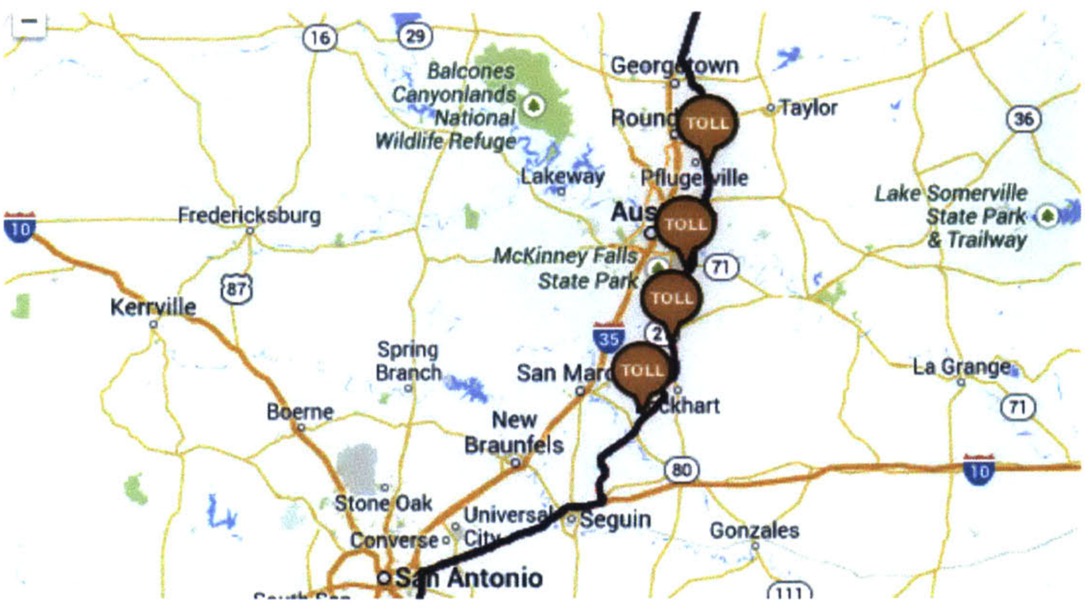


Figure 3-1: GPS trace of truck trip, moving south through Austin in morning rush hour. The driver chooses a tolled circumferential route, Texas State Highway 130.

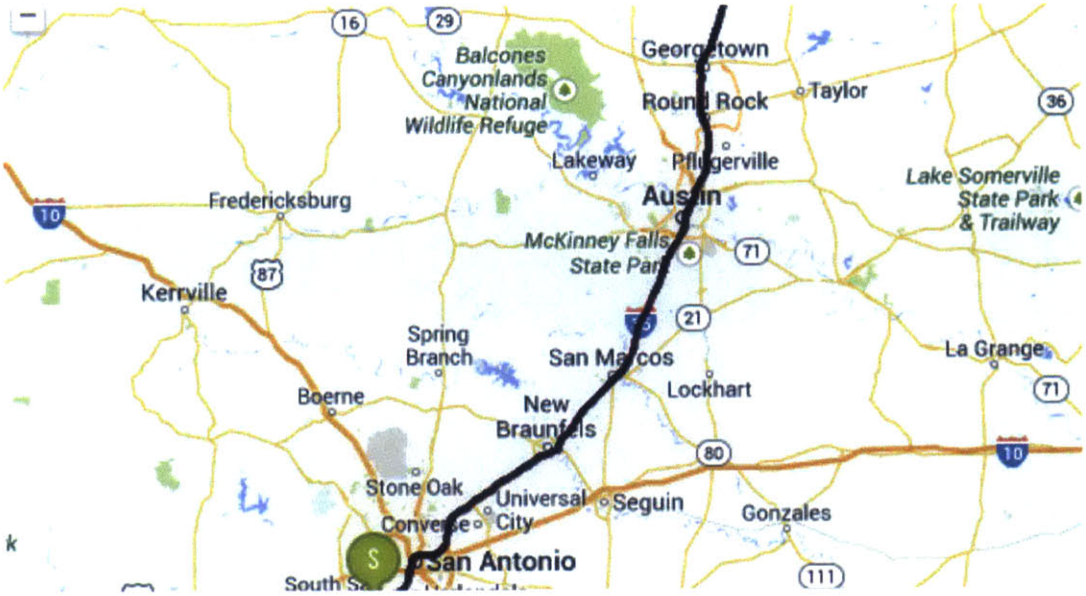


Figure 3-2: GPS trace of return truck trip, moving north through Austin in a weekend. This time, the same driver chooses to avoid the toll due to the lack of traffic congestion on the free route.

In other cases, different decisions can be made by the same driver on the same day. Figure 3-3 shows this for a trip in the Chicago area: different tolled routes are used for the incoming and outgoing trips. These two examples ilustrate the richness of the information that can be collected using GPS technology. By being able to visualize the actual routes taken, we can gain insights into the complexity of routing behavior that would not be possible in a standard survey.

Figure 3-3: GPS trace of a truck trip originating from Gary, Indiana, round trip to Milwaukee, Wisconsin. Different routes are used for the outgoing and incoming legs of the tour.

On a longer-term scale, we observed very different patterns in driver behavior. Figure 3-4, for example, represents a pattern of a 'long tour' driver who makes regular, long tours, each lasting multiple days and traveling distances in the thousands of miles. This driver returns to their home location for a break, then continues on a similar pattern as before. There are relatively few loading and unloading stops on these long tours.

In contrast, Figure 3-5 shows a pattern of a 'short tour' driver who makes short, regular tours and typically returns home at the end of each day. Most of the tours these drivers make are very similar, with the same unloading points. Finally, there is the 'gypsy trucker', illustrated in Figure 3-6. This driver has no predictable tour pattern and no apparent home location, simply driving around looking for work from city to city.

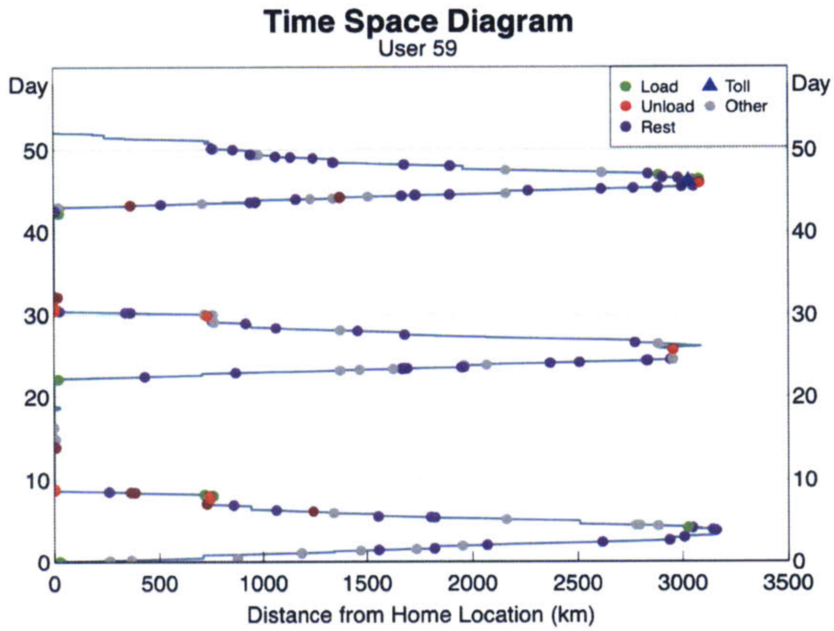


Figure 3-4: Long-tour driver

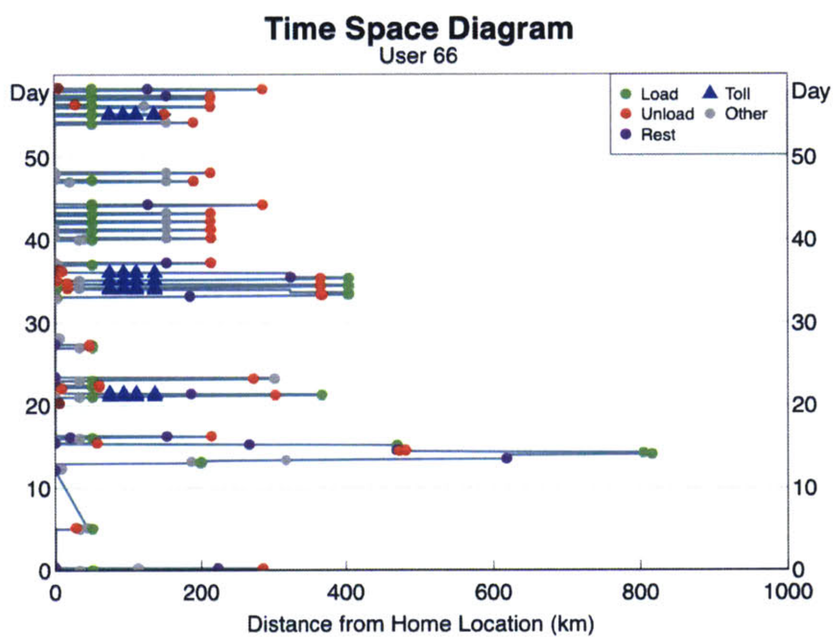


Figure 3-5: Short-tour driver

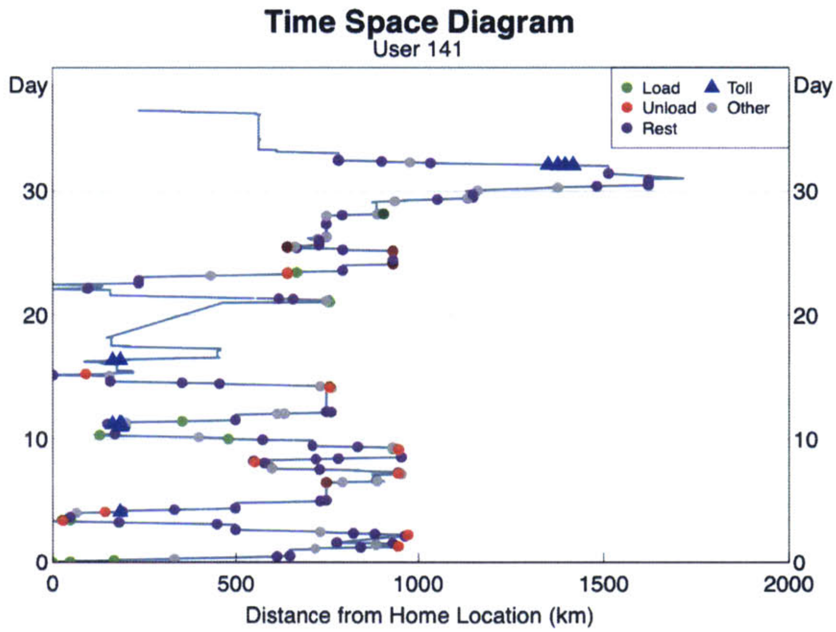


Figure 3-6: 'Gypsy' driver

There was also a stated preference survey component of Truckers@MIT, which was completed at the very end of the experiment, after the 20 days of data had been validated. This survey picked out eight trips in total that consisted of the journey from a loading stop to an unloading stop. It then offered them a new alternative that was potentially longer, faster and/or had an additional toll. An attempt at being context-specific was made by specifying the levels in terms of miles longer, minutes faster and additional toll compared to the existing route they took. The choice task was to select the original route or the alternative presented. The wording of the question is presented below:

On <day of the week, date and time> you departed from <name of origin> with a <truck configuration + special services> carrying <cargo type>. You were scheduled to transport this load to <name of destination> at <schedule>.

The map shows the route for this trip. Compared to it, the alternative route is:

- <distance> miles longer
- <travel time> minutes faster
- Has an additional toll of <toll> USD
- <Toll bearing method and terms based on real conditions for the trip>

Which route will you take?

☐ Original route

☐ Alternative route

Table 3.1 shows the levels presented for the experiment. The design was a random design - the value of each attribute was chosen randomly (with equal probability) from the sets of levels. The toll bearing method and terms were based on the real conditions for the driver and for the trip. Table 3.2 shows the possible values - for example, if we noticed the driver was paying cash and they said elsewhere that another company does not pay their toll costs, they would be shown the statement: 'You will pay cash at the booth, and you are responsible for the toll cost'.

Attribute	Levels
Distance (additional for alternative, miles)	[5, 10, 15, 20]
Travel time (saving for alternative, minutes)	[0, 10, 20, 30]
Toll (additional for alternative, $)	[0, 5, 10, 15]

Table 3.1: Truckers@MIT Stated Preference Survey: Levels for Experiment

You have an electronic toll tag, and the cost will be paid directly by the company or shipper
You have an electronic toll tag, and the cost will be reimbursed by the company or shipper
You have an electronic toll tag, and you are responsible for the toll cost
You will pay cash at the booth, and the cost will be reimbursed by the company or shipper
You will pay cash at the booth, and you are responsible for the toll cost

Table 3.2: Truckers@MIT Stated Preference Survey: Toll Payment Method and Toll Bearing Terms

The stated preference survey was filled out by 95 drivers overall. Unfortunately, however, even after the data were cleaned to remove irrational observations, we were not able to use them to estimate a meaningful discrete choice model. One possibility is that drivers did not care much about filling out the survey given it was their last task before receiving their compensation. Another is that reliability was not included as a variable, so routes with no travel time savings, but with a toll and a longer distance might have been preferred because they were perceived to be more reliable. Alternatively, the setup of the attributes may not have been appropriate for longer-distance trips: a 30-minute saving is unlikely to mean much in the context of a trip spanning thousands of miles.

In this thesis, we plan to use a similar prompted-recall methodology to estimate the demand for new modes. This will be achieved by selecting a trip the user made, and carefully selecting the levels to present for all modes, in a way that respondents perceive as realistic given the context. Chapter 4 outlines some examples of new transportation modes we might like to model the demand for, and Chapters 5 onwards present our survey strategy.

Chapter 4

New Modes of Transportation and Mobility Services

This chapter describes several emerging modes of urban transportation, all of which have the common theme of providing alternatives to car ownership in cities. One way this can be achieved is through bicycle and car sharing programs, which allow travelers to use these modes without having to own them. Another trend is of new paratransit modes that fill the gap between inflexible fixed-route public transportation and more expensive taxis. Finally, services are emerging that package together these new modes of transportation with existing modes, providing a subscription that can be used flexibly. If car ownership is to be reduced in cities, with the subsequent benefits to traffic congestion and parking availability, consumers need to be given alternatives that efficiently fill the gaps in their mobility that not having a car would create. These new modes and packages are all ways of filling those gaps. Stated preference surveys can be used to estimate the demand for these modes and packages, which in turn can help to build the case for their introduction.

4.1 Bike Sharing

Bicycle sharing is an old concept that has very recently been revitalized by information technology. Shaheen et al. (2012) characterizes bike sharing into three generations. In the first generation, bike sharing was simply a system where bicycles were painted one color and left unlocked around cities. The first, in Amsterdam in 1965, released fifty 'white bikes' into the streets. Similar programs were implemented elsewhere in Europe and North America, but were thwarted by bicycle theft. Later, second-generation systems used fixed docking stations where bicycles could be picked up and returned. Theft was less of a problem because these systems used a coin deposit, but the anonymity of the system meant it continued as an issue.

The third generation of bike sharing is what would be familiar to today's users. Third-generation systems maintain the fixed docking stations of their predecessors, but rely on information technology systems to dispense and collect bicycles. Users need a credit card to sign up and are charged fees if they steal the bicycles or do not return them within a set time period.

As of 2012, there were 136 bike share programs operating in 30 countries around the world, with over 200,000 shared bicycles and 13,000 stations. The largest was in Hangzhou, China, which had 60,600 bicycles connecting over 2,400 stations (Shaheen et al. 2012). Further expansion is planned for many cities: by 2018, the Citi Bike system in New York City will be expanded from 6,000 bikes and 330 stations to 12,000 and over 700 stations. The system will extend further up Manhattan and into Brooklyn, and will make a debut in Queens (New York City Department of Transportation 2014). And in the San Francisco Bay Area, the operators of Bay Area Bike Share have proposed to increase the size of the system from 700 to 7,000 bicycles, with a greatly expanded reach across more parts of the Bay Area (City and County of San Francisco: Office of the Mayor 2015).

With the recent growth in bike share systems, bike share's impact on urban travel behavior is now beginning to be studied in detail. In London, for example, a survey found 60% of bike share users only began cycling for transportation purposes in the last six months, and 50% never use a bicycle of their own to get around London. Another study from China found that 80% of respondents replaced walking, own bicycle and public transport with a bike share trip (Fishman et al. 2013). Another study, Martin and Shaheen (2014), looks at whether bike share is complementary to or competing with public transport: on one hand, shared bicycles can expand the reach of public transport, but could also be used instead of public transport. They find that this depends on the user's location: bike share increased public transport trips in outer areas because it can be used to access bus and rail stations. But in the urban center, it was being used to replace public transport trips.

4.2 Car Sharing

Similarly to bike sharing, car sharing began to be implemented in the 1970s in Europe, and by the late 1980s it became a commercially viable operation (Steininger et al. 1996). Under a car sharing system, a user signs up to the system and makes a reservation to rent a car on a short-term basis, typically less than one day. Traditionally, most car-sharing services have been round-trip, meaning users had to drop off the car at the same place as they picked it up in. The largest of these is Zipcar, which has over 900,000 members and 10,000 cars in North America and Europe. An annual membership costs $70, and rental rates start at $8.25 per hour, which includes fuel, insurance and a guaranteed parking spot at the pick-up location (Zipcar 2015). More recently, one-way car sharing has been made possible by services like Car2Go, where a car is found by a smartphone application, driven to the destination and dropped off (Car2Go 2015). The growing popularity of these services has attracted the attention of larger rental car companies: Zipcar was acquired by Avis and Hertz offers Hertz 24/7, which operates on a similar basis to Zipcar (Avis Budget Group (2015), Hertz (2015)).

Much of the research on car sharing focusses on its impact on vehicle miles traveled (VMT) and car ownership. Lane (2005), for example, studied the first year of operation of car sharing in Philadelphia, Pennsylvania. They found that each car share vehicle replaced 23 privately owned vehicles on average. The actual switch that was made away from driving depended heavily on whether car sharing members reduced their vehicle ownership. Those who did tended to replace car trips with transit, walking and taxis, while those who kept their cars tended to use car sharing to substitute for other car-based trips, such as traditional car rentals, borrowing friends' cars or taxis. Martin et al. (2010) conducted a larger-scale survey of over 6,000 car sharing users across North America, and found that the average number of vehicles per household dropped once car sharing was used, from 0.47 to 0.24. Further, the largest shift in ownership was from one to zero cars. They suggest car sharing has reduced overall vehicle ownership by 90,000 to 130,000 vehicles.

4.3 Flexible Mobility on Demand (FMOD)

Flexible Mobility on Demand (FMOD) is a paratransit concept that aims to fill the gap that currently exists between flexible, but expensive, taxi services and inflexible, but cheap, transit services.

The FMOD system, which is documented in Atasoy et al. (2015, forthcoming), comprises a fleet of minibuses with a passenger capacity of about eight. The system uses these minibuses flexibly to offer three types of services to users:

- *Taxi*: door-to-door private ride
- *Shared Taxi*: door-to-door ride, shared with other passengers

- *Minibus*: pick up and drop off at pre-designated bus stops, with fixed route and flexible schedule.

These three services offer a spectrum of services, allowing the user to choose between different degrees of flexibility and costliness, with taxi being the most convenient and expensive and minibus being the cheapest but least flexible.

The FMOD system is accessed using a smartphone application. The user gives the application their trip origin and destination, a preferred arrival or departure time window, and the number of passengers. The system then determines a feasible set of transport options with various schedules. This feasible set includes options that are in the passenger's preferred time window, as well as ones slightly outside of it. A choice model is then used to select the best options for passengers, which need not include all three of the service types above. The aims of this choice model can differ: the assortments can be chosen to maximize the operator's profit, maximize consumer surplus or a combination of the two.

Atasoy et al. (2015, forthcoming) show the benefits of this system for both users and operators. For users, the service is attractive because it provides an alternative to public transport that is less inflexible, and the variety of different options allows users to efficiently trade off price and convenience. For operators, the dynamic allocation of vehicles to different services allows them to use a limited number of vehicles more efficiently and target different transport sub-markets. Atasoy et al. (2015, forthcoming) show this through a simulation in Hino City, Tokyo, which has an area of about 9×8 kilometers. With 60 vehicles serving 5,000 requests a day, they showed that allocating vehicles flexibly delivers higher operator profit and higher consumer surplus, compared to having vehicles fixed as taxis, shared taxis or minibuses.

The FMOD concept is an umbrella transport concept that covers various services already operating. Uber, for example, operates similarly to the FMOD taxi service. UberPOOL offers an option to share a ride with another person going the same way, for a discount of 10-50% (Uber 2015). Via is a similar service in New York City that offers shared rides in Manhattan between 14th and 110th Streets for a flat fare of $5. The vehicles it uses are SUVs, which are a similar size to the minivans in FMOD. But it only picks up and drops off passengers at the ends of blocks to avoid making diversions (Via 2015). Services similar to the FMOD minibus are less widespread. One is Bridj, which offers express rides in Boston on a set schedule for a cost a little more than a regular public transport trip (Bridj 2015). Another is Leap, which provides service in San Francisco between Lombard Street and the downtown area, for a flat fee of $6 per ride. The buses are billed as being more comfortable than regular buses, with WiFi, power outlets and food available for purchase on board (Leap Transit 2015).

4.4 Mobility as a Service

Another emerging trend is that of Mobility as a Service (MaaS), a concept that allows households to purchase packages of mobility that provide an alternative to car ownership. These packages bundle the use of multiple different modes, such as car sharing, public transportation and bike sharing, into one package, similarly to how text messages, data and voice calls are bundled into cell phone plans (Hietanen 2014).

One example of this is UbiGo, a MaaS 'transport broker' in Gothenburg, Sweden (Sochor et al. 2014). In a trial involving about 100 households, UbiGo offered a monthly subscription for five services: public transport, car sharing, car rental, bike sharing and taxis. Each month, participants bought credit for each of the five services, and could access them via a smartphone application and a smartcard. The system also offered guaranteed rides if public transport was delayed by 20 minutes, cheaper public transportation to outlying areas and a rewards system for sustainable travel.

Sochor et al. (2014) surveyed the trial households about their use of transportation modes before and after using UbiGo. They found a modal shift away from private cars and towards public transportation, walking and cycling. Participants also held more negative perceptions about driving than before they used the service. Overall, UbiGo proved to be popular with the trial group, with over 80% saying they would be interested in continuing to use the service.

The trial of UbiGo was conducted with the support of local industry and government organizations (UbiGo 2015). Another, SwissPass, is operated by SBB, the Swiss national railway (SBB 2015). This new version of an SBB season pass now enables the use of complementary mobility services all across Switzerland. The card can access Mobility Carsharing at 1,400 locations, PubliBike, which is a bike sharing provider that also allows electric bicycle rental, and ski passes. It also provides a trip planner for walking, cycling and other outdoor activities.

SHIFT, a startup in Las Vegas, takes a different approach again, offering a combination of shared bicycles, small cars (the Smart Fortwo) and large cars (the Tesla Model S). The system is currently in beta testing, but a variety of plans are likely to be offered, with monthly bike-only access at $25, 20-30 trips at $250 and unlimited trips for $500 (Ferenstein 2014). Unlike in other systems, all the vehicles are owned by SHIFT. Within downtown Las Vegas, the system guarantees that users will be using a car or bike within five minutes of making a request via the SHIFT mobile application. And users do not make a choice about which mode to choose from a menu of options: instead, the system pairs them with an option that will be most convenient for the trip (SHIFT 2015). For example, short trips in the downtown area are more likely to be paired with a bicycle, while longer round trips to outlying suburbs would be paired with a larger car.

Chapter 5

The Future Mobility Survey

This chapter will present the structure of the Future Mobility Survey's GPS data collection and validation. As explained in Section 3.3, the Future Mobility Survey collects data using a smartphone application called FM-Sensing. This application is available on the iPhone and Android platforms.[1]

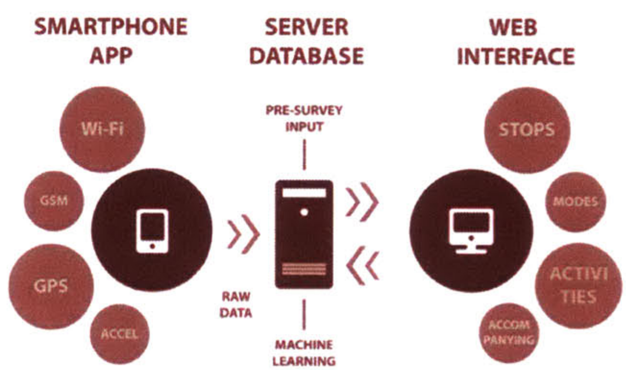


Figure 5-1: Future Mobility Survey System

[1]These applications can be found at https://play.google.com/store/apps/details?id=edu.mit.smart.fmsurvey.android&hl=en and https://itunes.apple.com/sg/app/fm-survey/id604011160.

5.1 Registration

The URL of the FMS web site is www.happymobility.org. The home page, illustrated in Figure 5-2, gives a brief description of the survey.

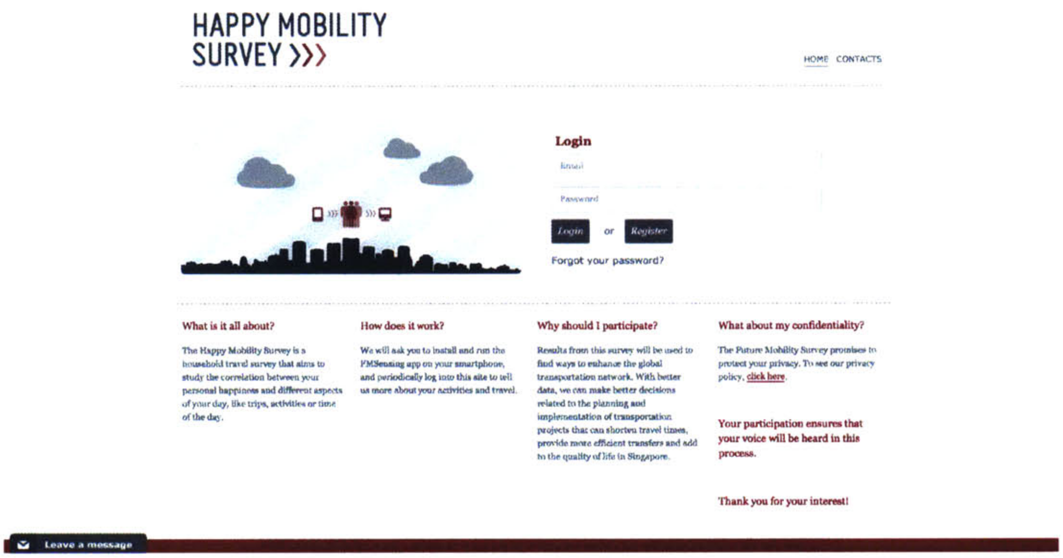


Figure 5-2: Future Mobility Survey Homepage

5.1.1 Registration

Upon registering, users are asked for their home address, work address and school address (if applicable). They are also asked to enter the address of any second homes they use as residences.

These addresses are used to automatically infer information at stops. When a stop is close enough to the address of the user's home, work or school, the system automatically selects this option during validation. This allows the user to simply confirm the inferred information instead of having to choose from all possible stop options.

5.1.2 Pre-Survey

Upon registering for FMS, the user answers a pre-survey with questions about themselves, their household and their mobility options. These questions are listed in Tables 5.1, 5.2, 5.3, 5.4, 5.5 and 5.6 below:

Information	**Question**	**Possible Responses**
Age	What is your age?	(Whole number)
Gender	What is your gender?	• Female • Male • Prefer not to answer
Marital Status	What is your marital status?	• Single, never married • Married/living with partner • Divorced/separated • Widowed • Prefer not to answer
Employment Status	Which of the following best describes your current employment situation?	• Employed full-time • Employed part-time • Self-employed • Temporarily on leave from a full or part-time job • Retired • Unemployed and seeking work • Not employed and not seeking employment

Table 5.1: Individual Questions: FMS Pre-Survey (Part 1)

Information	Question	Possible Responses
Ethnicity	How would you describe your racial background?	• African American/Black • Asian • American Indian, Alaskan Native, Native Hawaiian/Other Pacific Islander • Hispanic/Latino • White/Caucasian • Multiracial • Other • Prefer not to answer • Don't know

Table 5.2: Individual Questions: FMS Pre-Survey (Part 2)

Information	Question	Possible Responses
Years at Residence	For how many years (to the nearest year) have you lived at your current address?	(Whole number)
Dwelling Type	Which of the following best describes the building in which you currently live?	• Single-family detached house • Single-family attached house • A building with 2 or more apartments or condos • A mobile home or trailer • College dormitory, fraternity or sorority house • Other • Don't know

Table 5.3: Household Questions: FMS Pre-Survey (Part 1)

Information	Question	Possible Responses
Ownership Status	What is the ownership status of your current residence?	• Own • Rent • Occupy without owning or paying rent • Other • Don't know • Prefer not to answer
Household Income	Last year, what was your total household income from all sources before tax?	• (Whole number) • Prefer not to answer • Don't know

Table 5.4: Household Questions: FMS Pre-Survey (Part 2)

Information	Question	Possible Responses
Access to Car	How many vehicles are owned, leased or regularly available to the members of your household? *Include cars, motorcycles, mopeds, trucks, RVs, etc.*	0, 1, 2, 3, 4, 5+
Access to Bicycle	How many bicycles do you have at your home?	0, 1, 2, 3, 4, 5+
Bike Sharing	Do you currently use or are you a member of a bicycle sharing service?	• Yes • No • Don't know
Driver's License	Do you currently hold a driver's license?	• Yes • No
Car Sharing	[Only asked if participant has license] Do you currently use or are you a member of a car sharing service?	• Yes • No • Don't know

Table 5.5: Mobility Questions: FMS Pre-Survey (Part 1)

Information	Question	Possible Responses
Transit Pass	Do you currently have any type of transit pass?	• Yes • No • Don't know • Prefer not to answer
Mobility Issues	Do you have any mobility issues that make it difficult for you to go outside your home unassisted or unaccompanied?	• Yes • No • Don't know • Prefer not to answer
Mobility Aids	Which, if any, of the following aids do you use for moving outside your home? Please select all that apply.	• I do not use any aids • Walking stick/crutches for body support • Walker • Non-Motorized wheelchair • Stick/cane/guide dog to prevent obstacles to movement • Others • Prefer not to answer

Table 5.6: Mobility Questions: FMS Pre-Survey (Part 2)

5.2 Collection and Validation of GPS Data

Once a user finishes the pre-survey, they are directed to install the application on their smartphone. Screenshots from the smartphone application, FM Sensing, are shown in Figure 5-3. The left-hand image shows the start-up screen, and the right-hand image shows what the user sees once the application is open. The mobile application is very minimal, only showing a trace of where the user has been that day, with no indication of inferred stops or travel modes. The trace is stored on the phone, and is then uploaded to the Future Mobility Survey server, where stop and trip inference is made using machine learning algorithms.

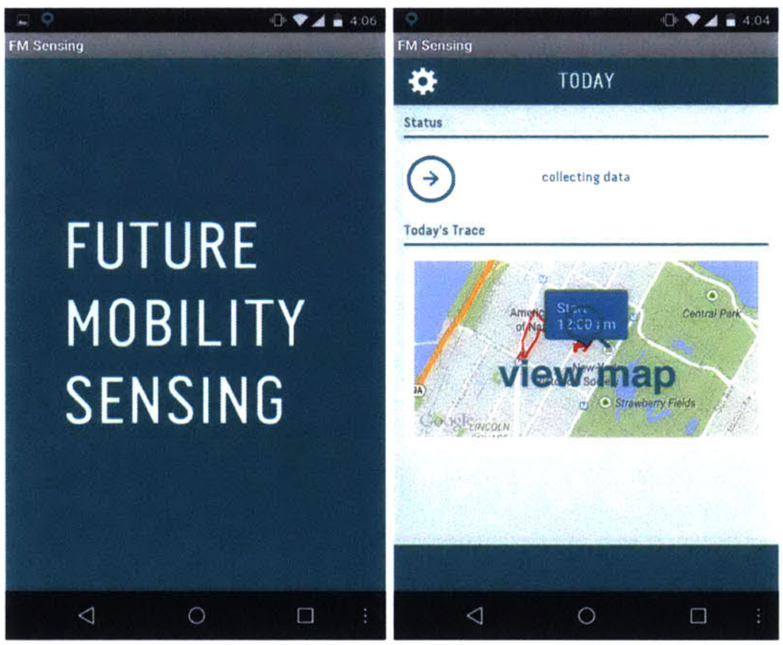


Figure 5-3: FM Sensing Application, Android Version

Periodically (ideally, at the end of each day), users log onto the Future Mobility Survey web site to validate their data. Upon logging on, they are shown the screen in Figure 5-4. From there, they can access support in validating their data and the Future Mobility Survey privacy policy.

Figure 5-5 shows a typical activity diary a user might see for one day. The GPS trace is shown for stops, as are stops the system inferred for that day. The activity performed at each stop might be inferred based on the user's past responses. To the right is the user's sequence of activities and trips, which they in turn validate for every stop and trip. Users may delete or add stops into their activity schedule.

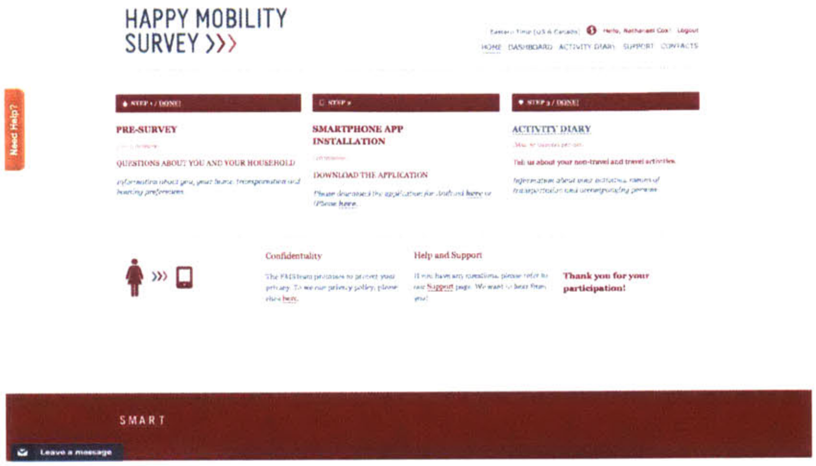


Figure 5-4: Future Mobility Survey: Welcome Screen

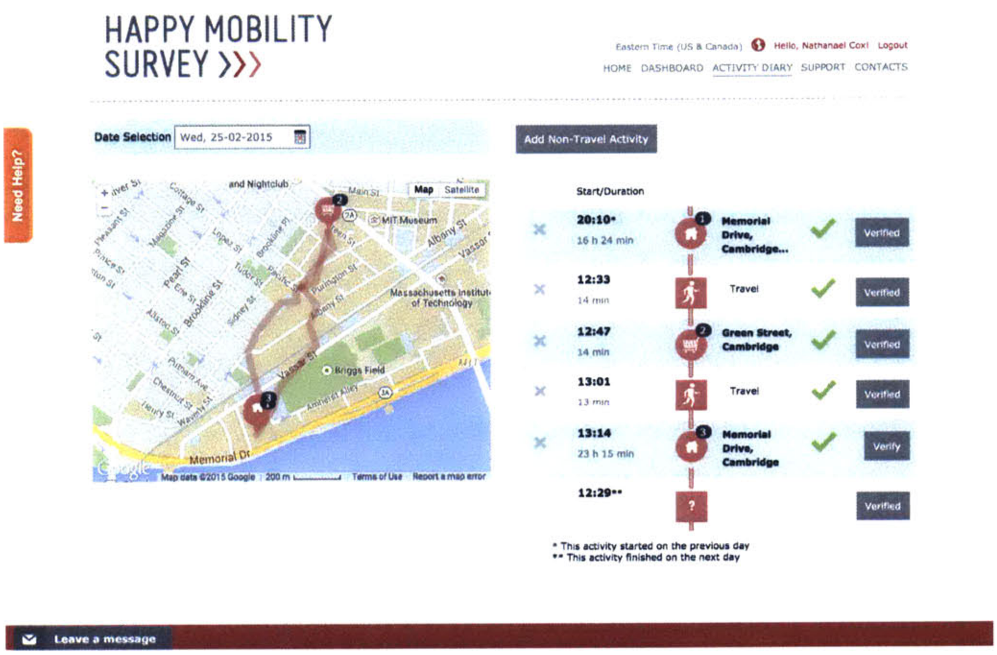


Figure 5-5: Future Mobility Survey: Validation Screen

Figure 5-6 shows the process of validating a stop. Users can choose from the 16 stop types listed in Table 5.7, or enter a text description if none of these is suitable. More than one activity type can be selected, which prompts another question asking what the main activity type was. The start of the activity can also be changed; the end time can be as well, but this can only be done by changing the start time of the next trip, which by definition is the end of the preceding activity.

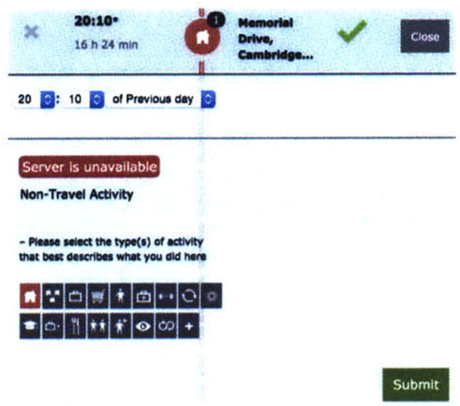


Figure 5-6: Future Mobility Survey - Validating a Stop

Stop Type	Symbol	Stop Type	Symbol
Home		Education	
Other Home		Work-Related Business	
Work		Meal/Eating Break	
Shopping		Social	
Personal Errand/Task		To Accompany Someone	
Medical/Dental (Self)		Entertainment	
Sports/Exercise		Pick Up/Drop Off	
Change Mode/Transfer		Other (Please specify)	
Recreation			

Table 5.7: Future Mobility Survey - Stop Types

Between every pair of stops is a period of travel: a trip. The interface for validating trips is shown in Figure 5-7. Users select one mode of transportation from the symbols given in the interface; a full list is shown in Tables 5.8, 5.9 and 5.10. They also specify how many other people travelled with them (with options 0, 1, 2, 3, 4 and 5+), and if this is more than one, they are asked if these are household members, non-household members, or a combination or both. Some modes also come with supplemental questions asking how the mode was used. For the car mode, for example, there are supplementary questions about the type of car, who was driving it and how it was parked. A full list of these supplemental questions is given in Tables 5.8, 5.9 and 5.10.

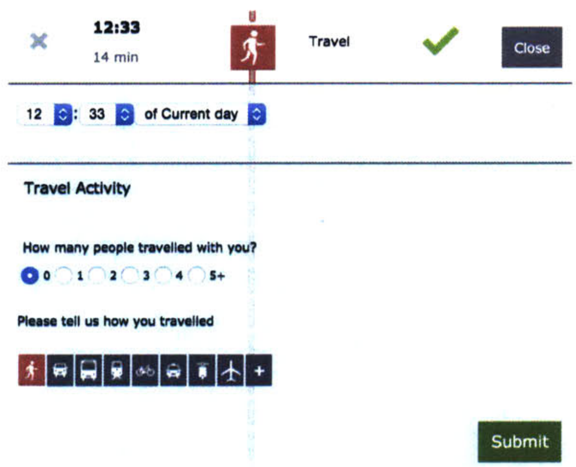


Figure 5-7: Future Mobility Survey: Validating a Trip

Mode	Symbol	Supplemental Questions	Possible Answers
Foot		None	
Car/Van		Vehicle Type	• Car • Lorry • Van
		Were you the driver?	• Yes • No

Table 5.8: Future Mobility Survey - Supplemental Mode Questions (Part 1)

Mode	Symbol	Supplemental Questions	Possible Answers
Car/Van		Parking Place	• Street • Personal Garage • Driveway • Parking Lot
		Parking Payment Type	• One-off Payment • Monthly/Season • No Charge
Bus		Bus Type	• Public bus • School bus • Company bus • Shuttle bus
		Bus Number	(User-entered route number)

Table 5.9: Future Mobility Survey - Supplemental Mode Questions (Part 2)

Mode	Symbol	Supplemental Questions	Possible Answers
LRT/MRT		None	
Bicycle		None	
Taxi		Taxi Fare	(Dollar amount)
		Was your fare reimbursable?	• Yes • No
Motorcycle/Scooter		Were you the driver?	• Yes • No
Air		None	
Other		Please specify	(User-entered text)

Table 5.10: Future Mobility Survey - Supplemental Mode Questions (Part 3)

Chapter 6

The Stated Preference Survey

The stated preference survey component is the primary contribution of this thesis. Its goals are threefold. First, it aims to be context-dependent, in the sense that any travel options presented seem realistic to the user, given where and when they are traveling. Second, it aims to present large amounts of information for many modes, but in a way that is easy to navigate and understand. Third, it should be able to be applied anywhere in the world without the use of skims specific to a certain region.[1]

These three goals are achieved by the survey described in this thesis. To satisfy the first goal, the survey design relies on web directions services, such as Google and Bing Maps. These services take in origin and destination points (and, in some cases, departure or arrival times) and deliver a set of directions with an estimated distance and travel time. These services provide a rough guide as to what a reasonable travel time to present is. Similarly, they can also be a useful source of mode availability - if a user makes a trip in a rural area with no transit service, no directions will be picked up by Google Maps, and a transit option will

[1] In travel demand modeling, *skims* are matrices that represent values of travel-specific variables between all possible origin and destination zones. Typically these are travel times, say by car, from the center of one zone to another. But skims can also represent times on other modes, such as transit and bicycling, or costs, such as a transit fare.

not be presented, because it is most likely not realistic for the user.

The second goal is achieved by constructing a user interface that is able to present a large amount of information. Instead of a traditional, table-based approach, we use a menu-based approach that allows users to view only the information they want to view.

We achieve the third goal of an interface that can be applied anywhere by using mapping services to find travel times and distances. We can find driving, walking and bicycling times for almost anywhere in the world using Google or Bing Maps. For public transportation, we can find route information in most major urban areas using Google Transit, which is available in over 70 countries (Google 2015).

This chapter illustrates the overall structure of the presentation of the stated preference survey, with information on the user interface and attributes presented. Chapter 7 details the experimental design that determines the values shown to the user, and Chapter 8 discusses how the survey could be extended and adapted to more new modes and new scenarios.

6.1 Traditional Stated Preference Surveys

Traditionally, stated preference surveys have relied on large tables of information to convey the attributes of the alternatives. An example is given in Table 6.1 for a hypothetical scenario, where the respondent chooses between car and bus for the commute trip.

This approach is very straightforward, but is not suitable for analyzing a large number of modes with a potentially large number of variables. If the survey had ten modes and perhaps many more variables, it is likely to be very difficult for users to take in and interpret realistically.

	Car	Bus
Total travel time (min)	28	45
Fuel cost ($)	5	-
Parking cost ($)	15	-
Bus fare ($)	-	2.75
Number of transfers	-	0

Table 6.1: Example Stated Preference Survey Profile

Lang (2013) adopts the table approach for a large number of modes in Lisbon. There, the survey is sequential: the respondent is asked to select their preferred option from three groups (car-based, public transport and multimodal). Each of these three groups has its own screen. Then, a summary screen is displayed, which shows the three preferred modes from each group, and the user selects one of these to make their final decision. If a car-based mode is selected, a fourth screen appears, giving the user the option to change their departure time to avoid congestion and a proposed congestion charging regime.

One way to avoid the potential perception issues of using a table is to use pictorial representation, which reduces information overload, makes users' perceptions of modes more homoegeneous, makes the task more interesting and increases realism (Morikawa 1989).

Another approach adopted in the market research literature is information acceleration. Marketers of new products use information acceleration by placing consumers in a virtual environment similar to that a consumer might experience when actually purchasing the good. For example, Urban et al. (1997) describe this process for an electric vehicle, where a virtual showroom was created that allowed consumers to walk around and view cars, sit in them and talk about them with a sales representative. They find this greatly increases realism compared to traditional marketing surveys, potentially leading to better sales forecasts.

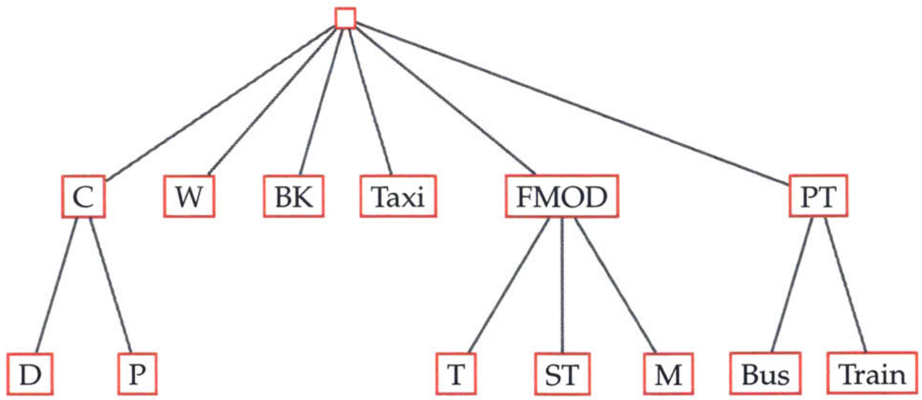
6.2 A Menu-Based Approach

Figure 6-1 shows the mode choice structure for the pilot version of this survey. There are ten modes in total, a number too large to represent in a table. Our alternative approach is menu-based: it relies heavily on pictorial representation to reduce the task burden, as suggested by Morikawa (1989). It also uses the process of information acceleration in Urban et al. (1997), by mimicking common trip planning applications. This increases the realism of the survey, as users feel like they are planning an actual trip.

We start by selecting the trip from home to the respondent's *primary destination* for that day, which is the main place the respondent visited in their day's travels. For simplicity, the primary destination is defined as the longest stop the user made that day, though there are other ways to define this (see, e.g., Bowman (1998)).[2]

We then present the hypothetical scenario to the user, which is to consider a trip made from the home to the primary destination without stops. We show the trip without stops to simplify the choice setting. If respondents had to plan a trip that had stops (say, the same stops that they actually made), they would need to select a mode for each leg of the half-tour, which could become very complex. Users would need to think about how their use of one mode impacts on their potential to use other modes. For example, if the first trip of a three-trip half tour was made by car and then another mode was used in the second leg, users would need to think about how they would eventually collect the car. From a survey design perspective, to present a parking charge we would need to know how long the user would have parked at the first stop, which further increases the complexity of the survey. And even if this was possible, it is not obvious that all stops would even be made under certain circumstances. For

[2]It is possible that the user does not have a primary destination, because they never left home that day. In these cases, the survey would not be shown for that day's data validation.



Abbreviation	Translation
C	Car
D	Car Driver
P	Car Passenger
W	Walk
BK	Bicycle
FMOD	Flexible Mobility on Demand (from Chapter 4)
T	FMOD Taxi
ST	FMOD Shared Taxi
M	FMOD Minibus
PT	Public Transport

Figure 6-1: Mode Choice Structure: Pilot Survey

example, if driving were presented as being prohibitively expensive and the user's revealed preference data shows they drove, users might use public transportation instead, forgoing stops that are easy to reach by car but difficult to access via public transportation. Instead, when the trip is one with no stops from home to the primary destination, these issues can be avoided.

6.3 Pilot Version: User Interface

The steps in the survey are:

1. Collect and validate GPS data using the FM Sensing smartphone application

(see Section 5.2 for details)

2. Enter the stated preference part of the survey
3. If this is the first time the user has seen the stated preference survey, present information on new modes
4. Present information on the trip (from home to primary destination, without intermediate stops)
5. Show information for review only
6. Allow selection of preferred travel mode.

Once the user has completed Step 1, they will be directed to start the stated preference survey via a hyperlink. This immediate redirection is a very important part of the survey process: the longer the time lag between data collection and the stated preference survey, the less the user is likely to remember about the trip. In Truckers@MIT, this proved to be an issue, because the stated preference part of the survey occurred at the very end of data validation, which in many cases took over four weeks. We suspect this was one of the factors that prevented us from getting meaningful results, because the truckers may have had difficulty remembering the trips and the conditions they encountered when they were on them. We only present the survey once per validation, which should not be any longer than once per day. Because the survey is reasonably complex, we would like to avoid presenting too many profiles to reduce the cognitive burden on respondents, as argued in Louviere et al. (2000).

In Step 3, we present the new modes, which in this case is FMOD. This screen, shown in Figure 6-2, describes the mode and shows its pictorial representation, which will be carried through to the rest of the survey to jog the user's memory. However, to minimize the burden on respondents, the new modes will not be shown again once the survey has been completed the first time. Users will, however, be able to access this information at a later time if they choose - see the representation in Figure 6-11. If, as we anticipate, other new modes, such as car and bicycle sharing, are presented, they will be shown here as well, and their order of presentation will be randomized.

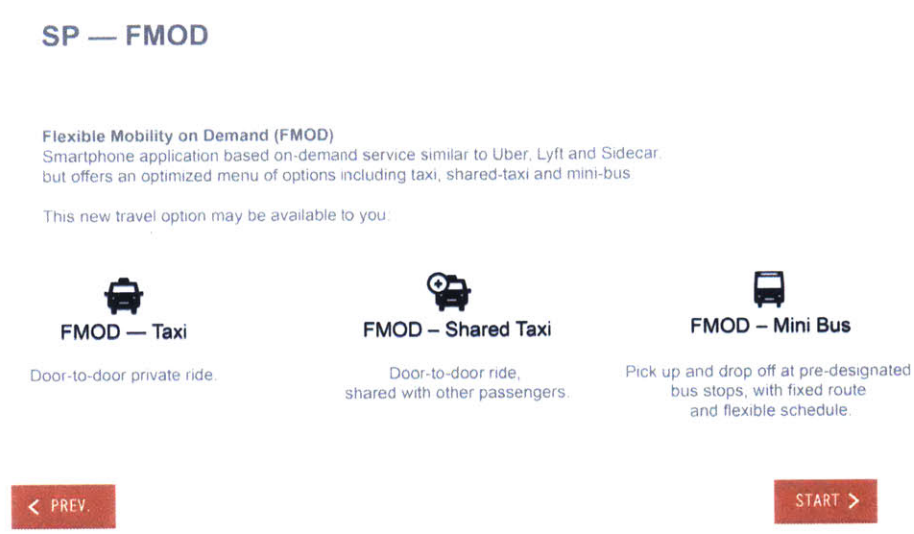


Figure 6-2: Stated Preference Survey: Initial Presentation of FMOD

Figure 6-3 shows Step 4: the presentation of the trip from home to the primary destination. To refresh the user's memory about the trip they will be considering, a map of the trip is shown, along with markers showing the home and primary activity locations. The activity at the primary destination is also shown so the user is aware of what they did at the primary destination, which might impact on their travel preferences.

The description of the task has been kept relatively short to minimize the respondent burden. It has been broken up into three steps to make sure the user understands all three points. The six icons at the top represent the six groups of modes the user will eventually choose from; they are the symbols from Figure 6-1. Their locations will also be randomized, so that car is not always presented

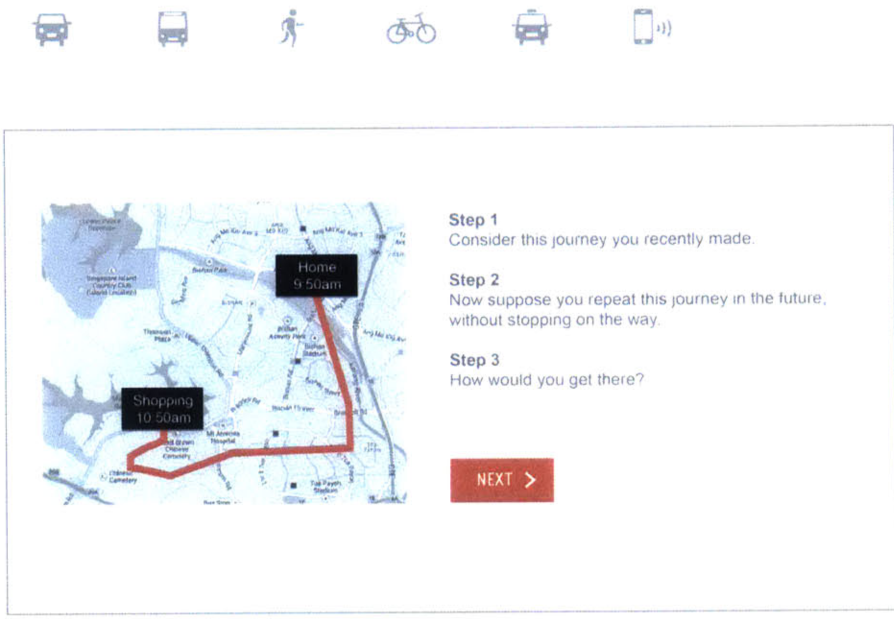


Figure 6-3: Stated Preference Survey: Presentation of Trip

first and FMOD is not always presented last. Table 6.2 illustrates which symbols relate to which modes.

Icon						
Mode	Car	Transit	Walking	Bicycle	Taxi	FMOD

Table 6.2: Stated Preference Survey: Mode Group Icons

6.3.1 Mode Review and Selection

Once the user has been reminded of the trip they made, they proceed to Steps 5 and 6, where they review the information for all the mode groups and proceed to a selection screen. There are separate stages of the process for review and selection; this is to encourage the user to explore the different options instead of just clicking on any mode and selecting that to avoid having to put in more effort.

The next parts show the user interface for each of the options. The interface is very similar between the review and selection stages - the only difference is that the option to select a mode is greyed out in the review process. However, during the selection process, the user will have the freedom to view (or not view) any mode group they choose. They can click on any of the mode group icons in Table 6.2 and make their selection from there.

The six mode groups are shown to the respondent in a random order, and the user can move between them by clicking on the 'previous' and 'next' buttons. Not all mode groups will be presented to the user - if a mode is not available, it will not appear. For example, if the user does not own a car, this will not be presented as a travel option. The availability criteria for modes are described in more detail in Chapter 7.

Car

The review and selection screens are shown in Figures 6-4 and 6-5. Where possible, two routes will be shown: one that has a (hypothetical) toll and one without a toll. These two routes are presented on a map, and the total trip time is displayed, along with the departure and arrival times. The travel time is displayed, and is broken down into four components: access time (walking time from trip origin to car); drive time; parking time; and egress time (walking time from parking place to destination). Toll, parking and fuel costs are displayed separately.

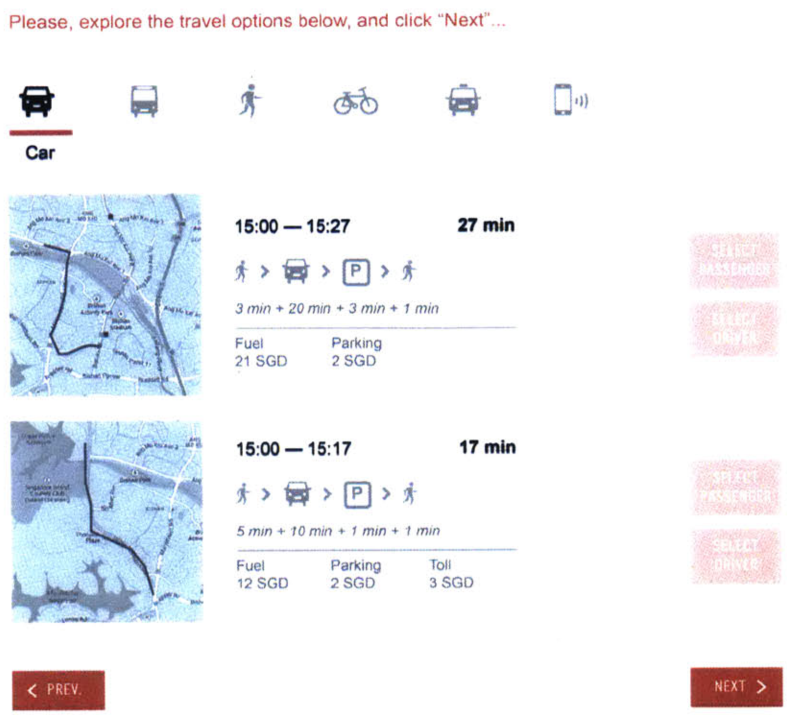


Figure 6-4: Stated Preference Survey: Car Mode - Review Screen

Figure 6-5: Stated Preference Survey: Car Mode - Selection Screen

Transit

The review screen for the transit mode group is shown in Figure 6-6; the selection screen is not shown here, but is identical to the review screen, except for the select buttons not being greyed out.

The transit mode group presents two mode groups: bus and train. As with the car mode, relevant travel times and costs (in this case, the fare) are displayed. The sequence of the trip is slightly different from the car, however. It is broken into:

- Access time: walking time from trip origin to the first bus stop or railway station
- Waiting time
- In-vehicle time
- Egress time: walking time from the last bus stop or rail station to the trip destination.

Transit trips may well involve more than one transfer, and this is built into the experimental design in Chapter 7. For each leg of the journey, then, there is an associated waiting time and in-vehicle time. This is accommodated in the survey by showing two different waiting times and in-vehicle times in sequence.[3]

Figure 6-6: Stated Preference Survey: Transit Mode Group - Review Screen

[3]To simplify the interface, we did not consider cases where transfers require walking a non-trivial distance from one transit stop to another.

Walking

The review screen for walking is shown in Figure 6-7. As before, the relevant times are shown, and there is no segmentation of different types of travel time. The walk distance is shown, as is the type of terrain. We hope to be able to retrieve the type of terrain from Google Maps, which can display this information for some queries. As with the car option, two routes are shown if possible, with maps that will enable users to identify them.

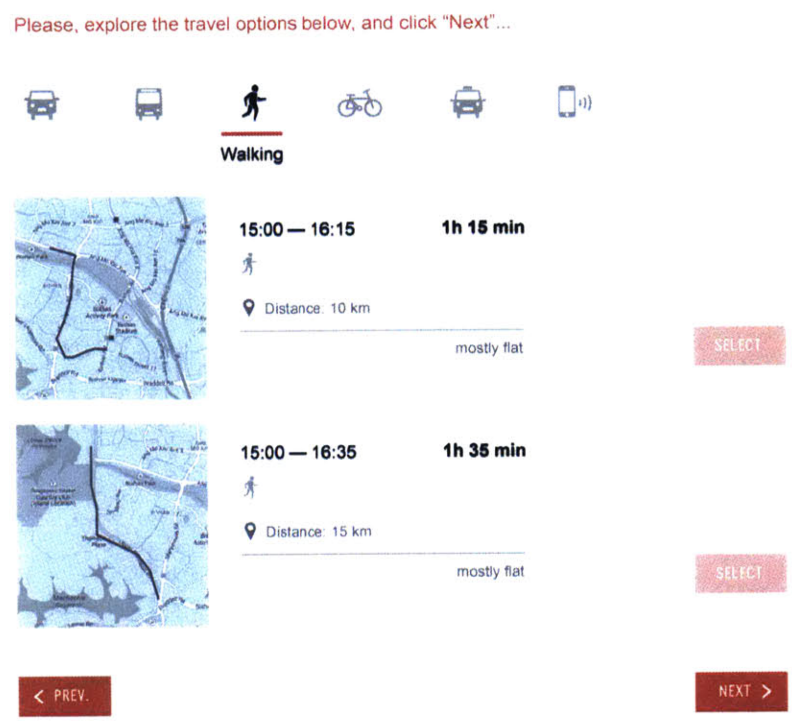


Figure 6-7: Stated Preference Survey: Walking Mode - Review Screen

Bicycle

Figure 6-8 shows the review screen for bicycles. The information shown is very similar to walking, with travel times, distances, terrain and route maps. The only addition is the percentage bicycle lane: where possible, the interface will show two routes. One is a safe route (shown in Figure 6-8 as the second option), with more bicycle lane but a longer travel time. The direct route (the first option shown in Figure 6-8) is quicker, but has less bicycle lane.

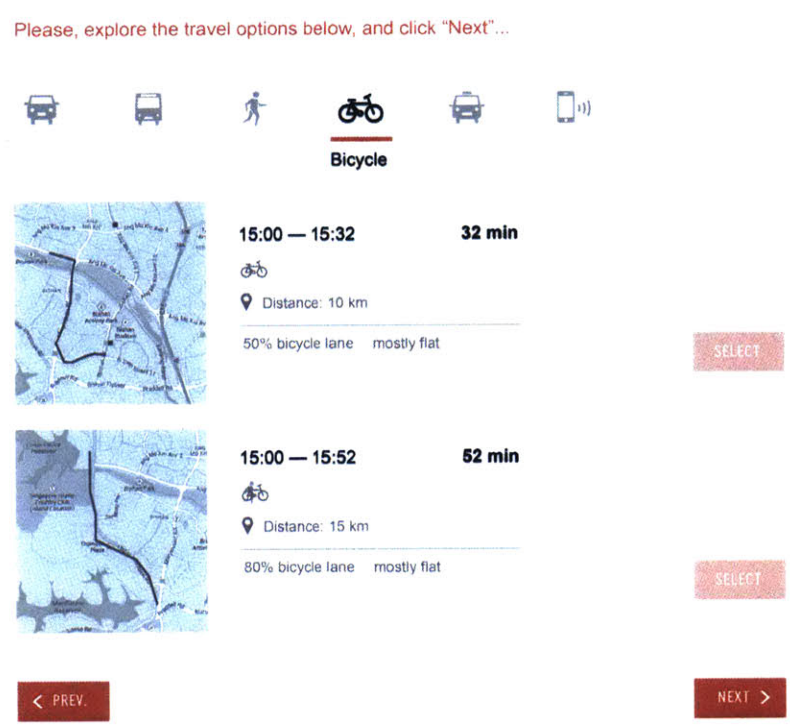


Figure 6-8: Stated Preference Survey: Bicycle Mode - Review Screen

Taxi

The review screen for the taxi mode is shown in Figure 6-9. Similarly to the car mode, there are two routes, and the relevant times and costs are illustrated. The total travel time is broken down into waiting time and in-vehicle time components; we assume there is no access or egress component, because taxis are a door-to-door service.

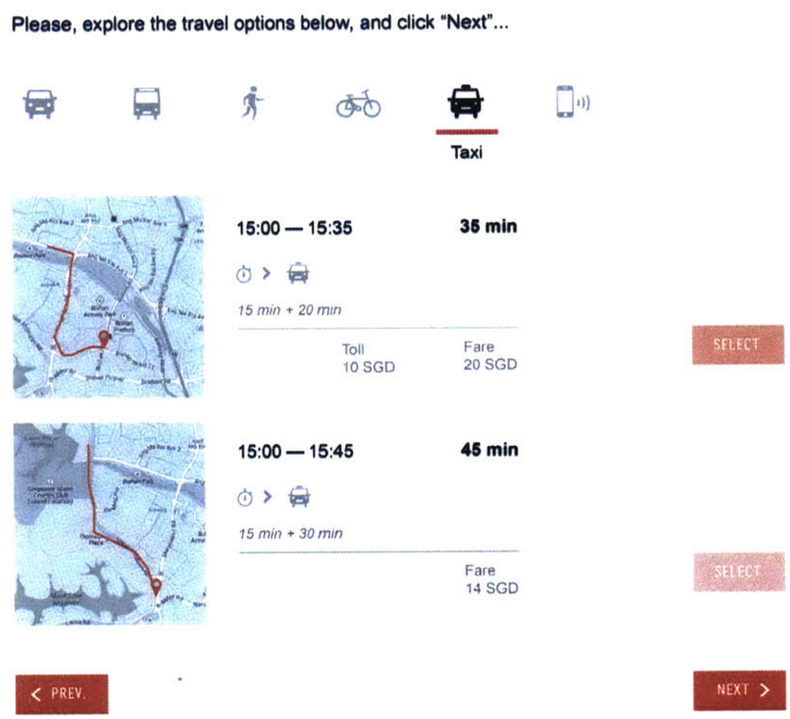


Figure 6-9: Stated Preference Survey: Taxi Mode - Review Screen

FMOD

Figure 6-10 shows the FMOD review screen. This shows the three separate FMOD alternatives, which will be presented in a random order and may not all be available (for more details, see Chapter 7). The presentation of the FMOD taxi and shared taxi options are the same as for a regular taxi. The fixed route bus alternative, FMOD minibus, is presented in a similar way to a regular bus, but we assume no transfers will be made. Figure 6-11 shows the pop-ups that are presented if the user hovers over the information icons. These descriptions are the same as those presented to the user the first time they take the stated preference survey.

Figure 6-10: Stated Preference Survey: FMOD Mode - Review Screen

Figure 6-11: Stated Preference Survey: FMOD Mode - Review Screen, with explanatory captions

6.3.2 Concluding Screens

Once the user has completed the survey, they are asked to confirm their selection and are presented with a completion page, as shown in Figure 6-12.

This chapter has presented the user interface of the survey, showing the steps the user will take to select a travel mode. Chapter 7 describes in detail how we determine which modes to show and how the values shown to the user are generated. Chapter 8 discusses ways the survey can potentially be extended, with more new modes and other applications.

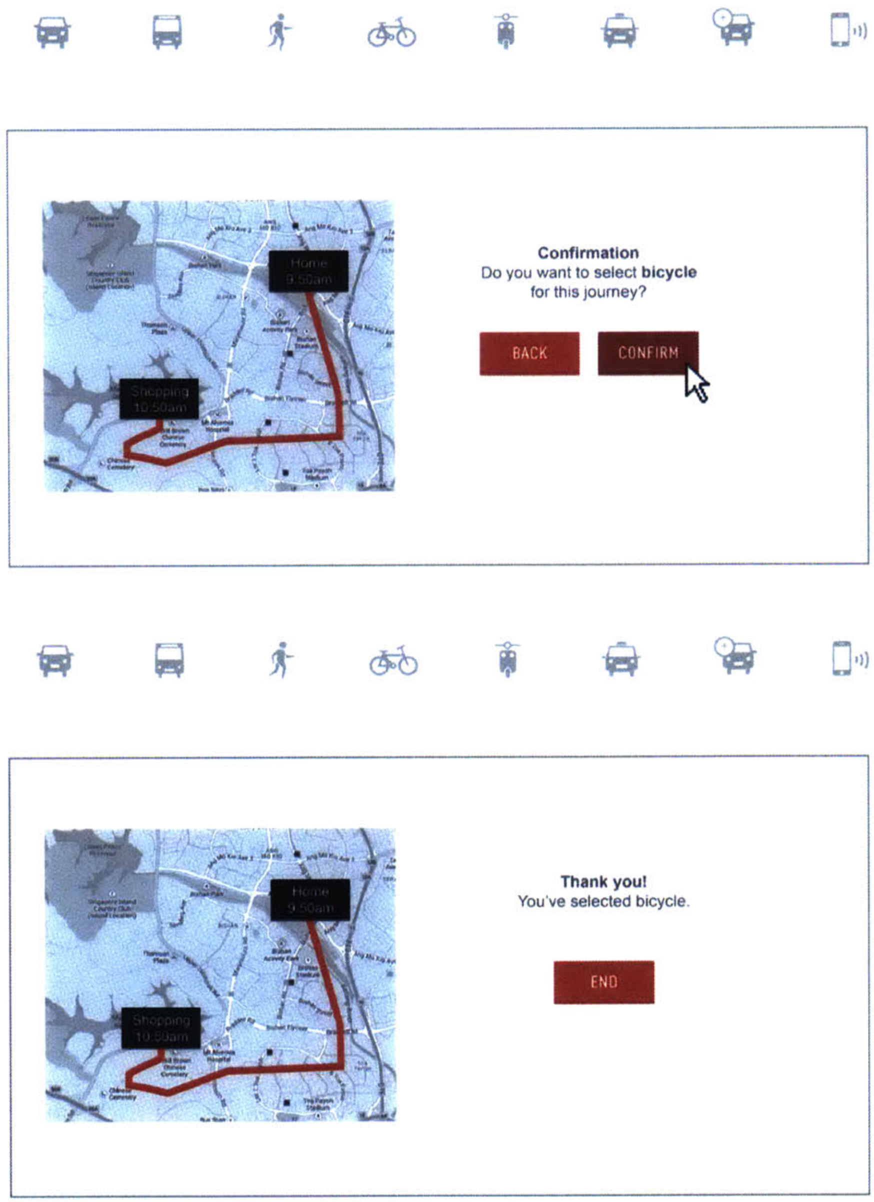


Figure 6-12: Stated Preference Survey: Confirmation Screens

Chapter 7

Experimental Design

Chapter 5 described the general structure of the survey. It showed the different modes and the user interface, but not how we generate the values for travel times and costs for different trips, or whether these modes should even be available to users. This chapter provides those answers.

7.1 Fetching Context-Aware Travel Information

As described in chapter 6, the trip shown is a trip from the user's home to their primary destination for the day, made without stops. The trip origin is the respondent's home location, and the trip destination is the location of the primary destination, which is the stop where the user spent the longest period of time.

Once we have the origin and destination of the trip, as well as the arrival time, we can use this information to gather information on travel times and distances from web services, such as Google and Bing Maps. This includes:

- *Driving information*: the travel time and distance, without traffic.
- *Walking information*: walking time, walking distance and terrain type.

- *Bike*: bicycling time, bicycling distance and terrain type.
- *Transit*: There are two queries to make to the Google Directions API - one for bus and another for train.
  - *Bus*: this query is made by selecting 'bus' as the preferred option in Google Maps, with the desired arrival time as the time the user actually arrived at the primary destination. We use these values:
    * Walking time from trip origin to first transit stop
    * Walking time from last transit stop to destination
    * Total time spent in a transit vehicle (either bus or train)
    * The arrival time for the suggested transit trip

    For short trips (such as walking two blocks to a store), it is likely that Google Directions won't show a transit trip - it will instead show walking. In these cases, we will not show buses or trains as options. We also do not use transfer information Google Directions; instead, the number of transfers will come from the experimental design.
  - *Train*: We make and store the same information as in the query for the bus, but with 'train' selected as the mode option.

  In many cases, the results from the bus and the train queries could be the same - in an area where there are no trains, selecting train as the preferred option will still give a bus. We keep these results (and expect them in many cases) - when we present a rail option, for example, this will be purely hypothetical, but the travel time will pivot from the existing bus travel times.

7.2 Mode Availability

The mockups in Chapter 6 assume that all modes are available; however, this is not likely to be the case for all respondents. Some, for example, might not own cars, while others may live in areas with limited or no public transport service.

We therefore apply the following criteria to determine whether to offer modes to respondents in the survey. In some cases, this information comes from the Future Mobility Survey pre-survey (in Chapter 5), but in other cases it relies on trip information we fetch from web services.

- *Car Driver*: only available if the respondent's household owns at least one car and has a valid driving license.
- *Car Passenger*: always available.
- *Walk*: always available.
- *Bicycle*: available if the respondent's household owns at least one bicycle.
- *Transit - Bus*: only available if a Google Maps search for the trip (from origin and destination, with 'bus' selected as the preferred mode) returns a trip with an arrival time within two hours of the actual arrival time.
- *Transit - Train*: only available if a Google Maps search for the trip (from origin and destination, with 'train' selected as the preferred mode) returns a trip with an arrival time within two hours of the actual arrival time.
- *Taxi*: always available.
- *FMOD Taxi*: always available.
- *FMOD Shared Taxi*: always available.
- *FMOD Minibus*: only available if the minimum driving distance from origin to destination is less than 2 kilometers.

7.3 Survey Design

The design approach we use for this experiment is a cleaned random design. According to Walker et al. (2015), 'the random design (which is the easiest to generate) performs as well as any design, and it (as well as any design) will perform even better if data cleaning is done to remove choice tasks where one alternative dominates the other'. Because the pilot survey includes new modes of transportation, we do not have good priors on the parameter values, so efficient design is not likely to be the most efficient choice. Further, of the standard fractional factorial designs, it is difficult to generate an orthogonal design with such a large number of variables. And even if we managed to generate such an orthogonal design, the large number of pairwise comparisons that would have to be made to clean profiles (see Section 7.4) would make the design far from orthogonal. The random design performs similarly to the orthogonal design in Walker et al. (2015) and it is much more straightforward to generate, which is why we employ it here.

The design for this experiment is a pivot-style design, which was described in Section 2.3. We combine four sources of information to generate the profiles shown to users:

- *Demographics*: to determine mode availability (Section 7.2)
- *Revealed preference (GPS)*: to find trip information (origin, destination, arrival time; Section 7.1)
- *Web services inputs*: to find trip times and distances (Section 7.1)
- *The experimental design*: to vary travel times and costs for presentation to the user, based on the rough guide given to us by web services.

The next sections describe how profiles are generated for each of the modes in the stated preference survey.

7.3.1 Attributes and Levels: Common to All Modes

There are certain attributes and relationships presented to the user that are common across all modes. We list them here so they are not repeated for every mode:

- *Arrival time*: the user's actual arrival time at the primary destination, from the GPS data. The only exception is for public transportation, where the arrival time is the arrival time from Google Directions.

- *Total travel time*: the sum of all travel time sub-components, which may include:
  - Access time
  - Egress time
  - In-vehicle time
  - Parking time
  - Waiting time
- *Departure time*: the arrival time, less the total travel time.

7.3.2 Attributes and Levels: Car

The experimental design for the car mode is shown in Table 7.1.

Variable Code	Design Variable	Levels
C-1	Travel time ratio: route 1, peak	1.2, 1.3, 1.35, 1.55, 1.7, 2, 2.5
C-2	Travel time ratio: route 1, off-peak	1, 1.05, 1.15, 1.2, 1.5
C-3	Travel time ratio: route 2, peak	1.2, 1.3, 1.35, 1.55, 1.7, 2, 2.5
C-4	Travel time ratio: route 2, off-peak	1, 1.05, 1.15, 1.2, 1.5
C-5	Fuel cost (per kilometer)	0.1, 0.25, 0.3, 0.5, 0.8, 1.0, 1.2, 1.5
C-6	Free parking	True, False
C-7	Parking cost (Hourly rate, $/hr)	0.5, 1, 1.5, 3, 4, 8
C-8	Parking cost (Daily rate, $/day)	2, 3, 5, 10, 15, 30
C-9	Toll (if only one route available)	True, False
C-10	Toll cost ($)	0.5, 1, 3, 4, 7, 15
C-11	Parking time (min)	1, 2, 3, 4, 5, 8, 10
C-12	Access time (min)	1, 2, 3
C-13	Egress time (min)	1, 2, 3, 4, 5, 8, 10, 15

Table 7.1: Design Variables: Car Mode

The attributes shown to users for the car mode are:

- *Access time (min)*: comes directly from design (variable C-12). Same time shown for both routes. The access time levels are generally shorter than the egress time levels, because the trip starts from home, where a car is likely to be parked very close by.

- *Travel time (Route 1, min)*: If the trip ends during 7-9am or 4-7pm on a weekday, show the driving time from the mapping service for route 1, multiplied by the peak travel time ratio for route 1 (variable C-1). Otherwise, multiply the driving time by the off-peak ratio for route 1 (variable C-2). There are different sets of peak and off-peak levels to reflect the greater possibility of traffic congestion during peak times.
- *Travel time (Route 2, min)*: Only shown if the mapping service returns multiple routes. If the trip ends during 7-9am or 4-7pm on a weekday, show the driving time from the mapping service for route 2, multiplied by the peak travel time ratio for route 2 (variable C-3). Otherwise, multiply the driving time by the off-peak ratio for route 2 (variable C-4).
- *Parking time (min)*: comes directly from the design (variable C-11). Same time for both routes.
- *Egress time (min)*: comes directly from the design (variable C-13).
- *Fuel cost (Route 1, $)*: multiply the distance travelled in route 1 by the per-kilometer fuel cost rate (variable C-5).
- *Fuel cost (Route 2, $)*: multiply the distance travelled in route 2 by the fuel cost rate (variable C-5).
- *Parking cost ($)*: different parking costs are shown for short and long trips. For trips of under three hours, a per-hour rate is used, and the parking cost is the primary activity duration (in hours), multiplied by the per-hour rate (variable C-7). Otherwise, a flat daily rate is shown (variable C-8). The parking cost shown is the same for both routes, since the trip destination is still the same regardless of the route.
- *Toll ($)*: what is shown depends on how many driving routes are returned by the web service.
  - If only one driving route is returned by the mapping software, the toll for this route will be zero (and the word 'toll' will not be shown at all)

if the toll dummy variable (variable C-9) is false. If it is not, then the toll will be a nominal dollar amount (variable C-10).

  - If two driving routes are returned, then the route with the longest driving time is given a toll of zero, and the word 'toll' is not shown. The shorter route will have a toll given by variable C-10.

7.3.3 Attributes and Levels: Bus

The experimental design for the bus mode is shown in Table 7.2.

Variable Code	Design Variable	Levels
B-1	Access time ratio	0.75, 0.9, 1, 1.05, 1.25
B-2	Egress time ratio	0.75, 0.9, 1, 1.05, 1.25
B-3	Number of transfers (if total trip time is between 10-20 minutes)	0, 1
B-4	Number of transfers (if total trip time exceeds 20 minutes)	0, 1, 2
B-5	Headway, Leg 1 (min)	3, 5, 8, 10, 15, 30
B-6	Headway, Leg 2 (min)	3, 5, 8, 10, 15, 30
B-7	Headway, Leg 3 (min)	3, 5, 8, 10, 15, 30
B-8	Waiting time ratio, Leg 1	0.1, 0.35, 0.5, 0.65, 0.85, 1
B-9	Nominal waiting time, Leg 1 (min)	1, 2, 3, 4, 5
B-10	Waiting time ratio, Leg 2	0.1, 0.35, 0.5, 0.65, 0.85, 1
B-11	Waiting time ratio, Leg 3	0.1, 0.35, 0.5, 0.65, 0.85, 1
B-12	In-vehicle travel time ratio	0.75, 0.9, 1, 1.05, 1.25, 1.5, 2
B-13	Proportion of in-vehicle time on leg 1, if one transfer (%)	20, 40, 60, 80
B-14	Proportion of in-vehicle time on leg 1, if two transfers (%)	15, 25, 35, 40
B-15	Proportion of in-vehicle time on leg 2, if two transfers (%)	15, 25, 35, 40
B-16	Fare ($)	0.5, 1, 1.5, 2.5, 3, 5, 7, 8, 10

Table 7.2: Design Variables: Bus Mode

The attributes shown to users for the bus mode are:

- *Access time (min)*: this is the walk from the trip origin to the first transit stop.

Multiply the access time from the web service by the ratio from the design (variable B-1). The variability in access and egress times reflects possibilities where the closest bus stop is either closer to or further away from the user's home location.

- *Egress time (min)*: this is the walk from the final transit stop to the trip destination. Multiply the egress time from the web service by the ratio from the design (variable B-2).
- *Number of transfers*: this depends on the total in-vehicle time, which is the total in-vehicle time from the web service, multiplied by a ratio (variable B-12). The rationale here is that very short trips are unlikely to involve much transferring.
  - If the total in-vehicle time is under 10 minutes, we will not show any transfers because the trip is very short.
  - If the total in-vehicle time is between 10 and 20 minutes, we can either have zero or one transfers; this is given by variable B-3.
  - Otherwise, if the total in-vehicle time is over 20 minutes, we can show either zero, one or two transfers; this value is given by variable B-4.
- *In-vehicle times (min)*: to get these, we first find the total in-vehicle time (by multiplying the web service value by variable B-12), and then divide it up depending on how many legs of the transit trip there are. The number of legs in the transit trip is equal to the number of transfers, plus one.
  - *Leg 1*: if there are no transfers, this is just the total in-vehicle time. If there is one transfer, this is the total in-vehicle time, multiplied by the proportion of in-vehicle time spent on leg 1 (variable B-13). If there are two transfers, this is the total in-vehicle time, multiplied by the proportion of in-vehicle time spent on leg 2 (variable B-14).
  - *Leg 2*: if there are no transfers, there is no leg 2 and this value is not shown. Otherwise, if there is one transfer, this is equal to the total in-vehicle time, less the in-vehicle time in leg 1. If there are two trans-

fers, this is the total in-vehicle time multiplied by the proportion of in-vehicle time spent on leg 2 (variable B-15).

- *Leg 3*: only shown if there are two transfers, because if there are less than two transfers, leg 3 does not exist. This is the total in-vehicle time, less the in-vehicle times for the first two legs.

- *Waiting times (min)*: the calculation of waiting times differs for the first and subsequent legs.

  - *Leg 1*: if the headway of service is more than 10 minutes, we use a nominal value, because people are more likely to rely on the schedule and control when they get to the bus stop. The 10-minute threshold is suggested by Frumin (2010), who examined Oyster card data from London to calculate waiting times as a function of headway. In this case, the waiting time is just the variable B-9. Otherwise, the service is a turn-up-and-go service where the waiting time is the headway multiplied by a factor. The idea here is that if a bus runs every 5 minutes (and every bus is on time), you could be waiting anywhere from zero up to 5 minutes. The waiting time then becomes the headway of leg 1 (variable B-5) multiplied by the waiting time ratio for leg 1 (variable B-8).
  - *Leg 2*: in subsequent legs, we cannot control our arrival time at the transit stop, because it is determined by the first leg. Therefore, regardless of headway, the waiting time is the headway of leg 2 (variable B-6) multiplied by the waiting time ratio for leg 2 (variable B-10).
  - *Leg 3*: similarly to leg 2, this is the headway of leg 3 (variable B-7) multiplied by the waiting time ratio (variable B-11).

- *Fare ($)*: comes directly from the design (variable B-16).

7.3.4 Attributes and Levels: Train

The design for the train is exactly the same as for the bus mode - the only difference is in the inputs from the mapping services. Whereas the bus design relies on the inputs of bus in-vehicle time, access and egress times, the train design relies on these same inputs, but for a web service query where train is preferred. Further, different values will come out of the design (say, for example, the fare) compared to the bus design values.

7.3.5 Attributes and Levels: Walk

The only design variable for walking is a walk-time ratio, which is applied equally for both routes. The levels for this ratio are [0.75, 0.9, 1, 1.05, 1.25]. This reflects variability in walking speeds, and means the walking times shown are the walk time from the web service muliplied the walk time ratio. The walking distances for both routes come directly from the web service, as does the terrain.

7.3.6 Attributes and Levels: Bicycle

The experimental design for the bicycle mode is shown in Table 7.3.

Variable Code	Design Variable	Levels
BK-1	Bicycling time ratio	0.75, 0.9, 1, 1.05, 1.25
BK-2	Bike lane, route 1 (%)	0, 20, 50, 80, 100
BK-3	Bike lane, route 2 (%)	0, 20, 50, 80, 100

Table 7.3: Design Variables: Bicycle Mode

The attributes shown for the bicycle mode are:

- *Bicycling time (Route 1, min)*: the bicycling time from the web service for route 1, multiplied by the bicycling time ratio (variable BK-1). The bicycling time ratio represents the different speeds people ride bicycles at.

- *Bicycling time, (Route 2, min)*: the bicycling time from the web service for route 2, multiplied by the bicycling time ratio (variable BK-1).
- *Terrain type (Route 1)*: as returned by the web service, for route 1.
- *Terrain type, (Route 2)*: as returned by the web service, for route 2.
- *Distance, (Route 1)*: as returned by the web service, for route 1.
- *Distance, (Route 2)*: as returned by the web service, for route 2.
- *Per cent bike lane (Route 1)*: comes directly from the design (variable BK-2).
- *Per cent bike lane (Route 2)*: comes directly from the design (variable BK-3).

7.3.7 Attributes and Levels: Taxi

The experimental design for the taxi mode is shown in Table 7.4.

Variable Code	Design Variable	Levels
T-1	Waiting time (min)	1, 2, 3, 5, 8, 10, 15
T-2	Travel time ratio, Route 1	0.6, 0.75, 0.85, 1, 1.1
T-3	Travel time ratio, Route 2	0.6, 0.75, 0.85, 1, 1.1
T-4	Fare ratio	0.5, 0.7, 1, 1.4, 2

Table 7.4: Design Variables: Taxi Mode

The attributes shown for the taxi mode are:

- *Waiting time (min)*: comes directly from the design (variable T-1).
- *In-vehicle time (Route 1, min)*: first multiply the travel time from car route 1 from the car mode (including the ratio from the car design) with the travel time ratio for car route 1 (variable T-2). Then we take the maximum of this time and the time from the web service for route 1. We do this because, in congested conditions, a taxi might be able to travel more quickly than a car by taking advantage of carpool lanes. But by taking the maximum, we ensure the taxi does not travel faster than the without-traffic travel time returned by the web service.

- *In-vehicle time (Route 2, min)*: as for the in-vehicle time for route 1, but using the associated values for route 2.
- *Fare ($)*: multiply the Singapore taxi fare that would apply for this trip (about $3, plus 55 cents per kilometer) by the fare ratio (variable T-4). The distance taken here comes from the distance of the quickest route, as returned by the web service.
- *Toll ($)*: this value is the same as the car toll. It is applied to the route with the shortest in-vehicle time. The longer route is assumed to have no toll.

7.3.8 Attributes and Levels: FMOD Taxi

The experimental design for the FMOD taxi mode is shown in Table 7.5.

Variable Code	Design Variable	Levels
FMOD-T-1	Waiting time (min)	1, 2, 3, 5, 8, 10, 15
FMOD-T-2	Travel time ratio	0.6, 0.75, 0.85, 1, 1.1
FMOD-T-3	Fare ratio	0.5, 0.7, 1, 1.4, 2

Table 7.5: Design Variables: FMOD Taxi Mode

The attributes shown for the FMOD taxi mode are:

- *Waiting time (min)*: comes directly from the design (variable FMOD-T-1).
- *In-vehicle time (min)*: first multiply the quickest car travel time from the car mode (one of the two routes, and including the ratio from the design) with the travel time ratio (variable FMOD-T-2). Then we take the maximum of this time and the time from the web service, similarly to in the taxi case.
- *Fare ($)*: multiply the taxi fare from Section 7.3.7 by the fare ratio (variable FMOD-T-3).

7.3.9 Attributes and Levels: FMOD Shared Taxi

The experimental design for the FMOD taxi mode is shown in Table 7.5.

Variable Code	Design Variable	Levels
FMOD-ST-1	Waiting time (min)	1, 2, 3, 5, 8, 10, 15
FMOD-ST-2	Travel time ratio	1, 1.2, 1.5, 1.7, 2
FMOD-ST-3	Fare ratio	0.4, 0.5, 0.65, 0.85, 0.9

Table 7.6: Design Variables: FMOD Shared Taxi Mode

- *Waiting time (min)*: comes directly from the design (variable FMOD-T-1).
- *In-vehicle time (min)*: multiply the FMOD taxi in-vehicle time by the travel time ratio (variable FMOD-ST-2). This ratio is always more than one, because a shared ride should never be quicker than a taxi.[1]
- *Fare (\$)*: multiply the FMOD taxi fare from Section 7.3.8 by the fare ratio (variable FMOD-ST-3).

7.3.10 Attributes and Levels: FMOD Minibus

The experimental design for the FMOD minibus mode is shown in Table 7.7.

Variable Code	Design Variable	Levels
FMOD-M-1	Access time (min)	2, 4, 5, 7, 8, 10, 15
FMOD-M-2	Waiting time (min)	2, 4, 5, 7, 8, 10, 15
FMOD-M-3	Travel time ratio	1, 1.2, 1.5, 1.7, 2
FMOD-M-4	Egress time (min)	2, 4, 5, 7, 8, 10, 15
FMOD-M-5	Fare ratio	0.4, 0.5, 0.65, 0.85, 0.9

Table 7.7: Design Variables: FMOD Minibus Mode

- *Access time (min)*: comes directly from the design (variable FMOD-M-1).

[1] It can, however, be the same: if the passenger is the last to be picked up and the first to be dropped off, then the travel time will be very similar to a taxi because there are no intermediate stops.

- *Waiting time (min)*: comes directly from the design (variable FMOD-M-2).
- *In-vehicle time (min)*: multiply the FMOD shared taxi in-vehicle time in Section 7.3.9 by the travel time ratio (variable FMOD-M-3).
- *Egress time (min)*: comes directly from the design (variable FMOD-M-4).
- *Fare ($)*: multiply the FMOD shared taxi fare from Section 7.3.9 by the fare ratio (variable FMOD-M-5).

7.4 Profile Cleaning

In a random design such as ours, some profiles convey much less useful information about users' preferences than others, because they don't involve any trade-offs in time, cost or other attributes. Take, for example, a profile that presents a car as being both quicker and cheaper than public transportation. In a large majority of cases, the car would be chosen. Compared to a second profile where a car was quicker but more expensive, we do not learn as much, because the second profile requires respondents to trade off time and cost between the two modes. Keeping these profiles in an experimental design reduces the efficiency of estimation, because it does not require 'hard choices' on the part of the respondent.

This gain in efficiency is shown in Walker et al. (2015), who test experimental designs with and without the presence of these profiles. They simulate a choice experiment with two alternatives, with a model specified as:

$$U_{in} = \beta_{Time} \times Time_{in} + \beta_{Cost} \times Cost_{in} + \varepsilon_{in}$$

They generate designs for the time and cost variables, and compare designs that have been 'cleaned' of profiles with strictly dominated alternatives (higher time and cost) to designs that retain these profiles. They find that cleaning the design improves estimation efficiency, as measured by D-error.

Mode	ASC
Bus	-2.08
MRT (Subway)	-2.19
Drive alone	0
Motorcycle	-8.51
Walk	-2.23
Taxi	-5.57

Table 7.8: Alternative-Specific Constants for Usual Work Tour (Li 2013)

One caveat to their work, however, is that they assume there is no alternative-specific constant in either alternative. This means that, excluding the error term, any alternative that has both a higher time and cost is strictly dominated in utility terms. In reality, however, alternative-specific constants do vary significantly. If we have a strong belief that one mode has a significantly higher alternative-specific constant, then it may be chosen even if it takes longer and is more expensive. This may well happen in many urban situations: in rush hour, it is possible that bicycling is quicker and cheaper than driving. But many would still choose to drive anyway, and this comes down to the alternative-specific constant likely being higher for car than bicycle.

Li (2013), in estimating a tour mode choice model for home-to-work trips using data from Singapore's Household Interview Travel Survey (HITS), found the alternative-specific constants in Table 7.8. There is clear variation between modes: all else equal, car is the most preferred, then walking and public transportation, then taxi, then motorcycle.

We use these values to inform our strategy of eliminating profiles with strictly dominated alternatives. Under this strategy, one mode is strictly preferred to another if it has a higher alternative-specific constant, *and* it has a lower total cost and lower total travel time. We do not have motorcycle as a mode in the survey (though Chapter 8 discusses its eventual inclusion), but we assume the alternative-specific constant for bicycle is similar to for motorcycle. The two public transportation modes have similar alternative-specific constants, as does walking. Therefore, public transportation should not be slower than walking, because walking would likely be preferred in utility terms. Further, between the two public transportation modes (bus and train), there should not be one that has a lower time, cost and number of transfers, because the alternative-specific constants are very similar.

These rules are summarized in Table 7.9, which is a series of pairwise comparisons between modes. Any profile that does not pass all these tests is considered to have strict domination, and would not be included in the survey. The tests are only applied if both the modes in the test are available (as determined using the rules in Section 7.2). Within the three FMOD products, there is no need to test strict domintion, as it is built into the design: FMOD taxi is always quicker than FMOD shared taxi, which is in turn cheaper than FMOD minibus. The costs work in the reverse way, such that none of the three products is cheapest and quickest. For different types of bicycle routes, a valid profile should not have a bike route that is both shorter and has more bike lanes.

We also remove profiles that have excessive differences in time and cost. If these differences are very large, then even though certain modes (say, those with an extremely high cost compared to others) might not be strictly dominated for any possible combination of time and cost parameters, they are likely to be strictly dominated for most plausible values of time.[2] Therefore, a passing profile must satisfy these two tests:

- The longest travel time (excluding walking and bicycling) cannot exceed 5 times the shortest travel time.[3]
- The costliest mode that isn't free cannot exceed 15 times the cheapest mode

[2]The actual values of time that would cause a profile to be excluded depend on the total travel times and costs, which in turn depend on the parameters from Google Maps. If we let the minimum and maximum times be t_{min} and t_{max} and the costs be c_{min} and c_{max}, and assuming the costliest mode is also the fastest, then the value of time that would lead to the costliest mode to be chosen over the cheapest mode is $(c_{max} - c_{min})/(t_{max} - t_{min})$.

[3]Walking and bicycling are excluded from this comparison because significant differences in walking and bicycling times are likely to occur for longer trips, and are realistic.

Test	Mode 1	Mode 2	Test fails if:
1	Car	Bus	Car is cheaper and quicker[a]
2	Car	Train	Car is cheaper and quicker[a]
3	Car	Taxi	Car is quicker and cheaper[a]
4	Car	FMOD Taxi	Car is quicker and cheaper[a]
5	Car	FMOD Shared Taxi	Car is cheaper and quicker[a]
6	Car	FMOD Minibus	Car is cheaper and quicker[a]
7	Taxi	FMOD Taxi	Taxi is either cheaper and quicker, or slower and more expensive
8	Walk	Bus	Walking is quicker than bus[b]
9	Walk	Train	Walking is quicker than train[b]
10	Walk	Taxi	Walking is quicker than taxi[b]
11	Walk	FMOD Taxi	Walking is quicker than FMOD Taxi[b]
12	Walk	FMOD Shared Taxi	Walking is quicker than FMOD Shared Taxi[b]
13	Walk	FMOD Minibus	Walking is quicker than FMOD Minibus[b]
14	Bike (direct route)	Bike (safe route)	Direct route has more % bike lane than safe route[c]

[a] Within each test involving the car mode, each route is tested separately, and the test will pass only if both routes pass.
[b] We do not need to compare travel costs in this case, because the cost of walking and bicycling is assumed to be zero.
[c] By definition, the direct route always has a shorter travel time, so we need only test the percentage of bike lane.

Table 7.9: Tests for Strict Domination of Profiles

that has a monetary cost. Walking and bicycling are assumed to be free, and are not used in this comparison.

7.5 Monte Carlo Simulation of Design

One test of the usefulness of an experimental design is its ability to replicate assumed parameters of a model. We do this by using Monte Carlo simulation. This process involves assuming certain values for the parameters of the model, using those parameters to generate hypothetical choices, and then using the synthetic data to estimate the original model. A good design is one that is able to reproduce reasonably well the parameters that were originally assumed.

Unlike a typical experimental design, the actual variables a survey respondent sees are a product of more than just a pre-determined design matrix. They are derived from the pre-survey (to determine mode availability) and the characteristics of the trip that they fill out later (which provides the context for the experiment), as well as the pivot-style experimental design.

Therefore, we took a random sample of trips from several users who had validated data from the Future Mobility Survey. We used those trips, along with the users' demographic information, to simulate the choice experiment and estimate a mode choice model.

Due to difficulties in obtaining public transportation data from Google Directions, the Monte Carlo experiment does not include the public transportation mode. Further, for simplicity, only one route is shown for car, bicycle and walking, and the choice of car driver and car passenger is removed. Figure 7-1 shows the mode choice structure for the Monte Carlo simulation; for simplicity, the nesting structure for the FMOD modes has been removed.

For each observation, 2,000 experimental design rows were prepared, drawing

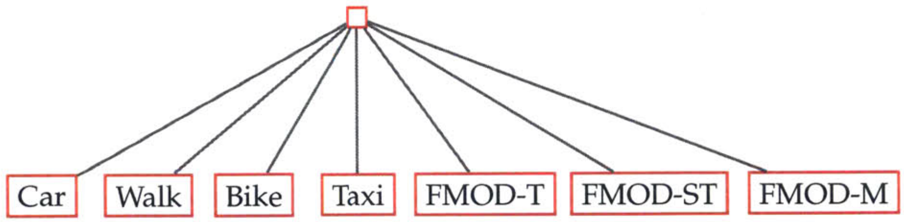


Figure 7-1: Mode Choice Structure: Monte Carlo Simulation

randomly from each design input variable. Similar tests for strict dominance and excessive differences (presented in Section 7.4) are used. The tests accounted for mode availability, in the sense that they did not reject profiles with problems based on unavailable modes. If a profile failed any one of the tests, the next design row was tried. Each trip was therefore given a profile that passes these tests, and the row of the design is removed so that subsequent profiles do not share the same experimental design levels.

Once the profiles are generated for each trip, we then calculate the systematic part of the utility, which involves assuming certain values on the model parameters. The assumed systematic utility functions are loosely based on Li (2013) and are shown below:

$$
\begin{aligned}
V_{Car} &= -0.5 \times TotalTime_{Car} - 0.07 \times TotalCost_{Car} \\
V_{Walk} &= -2.5 - 2 \times TotalTime_{Walk} \\
V_{Bike} &= -3 - 2 \times TotalTime_{Bike} \\
V_{Taxi} &= -3 - 0.45 \times TotalTime_{Taxi} - 0.03 \times Fare_{Taxi} \\
V_{FMODTaxi} &= -3 - 0.45 \times TotalTime_{FMODTaxi} - 0.03 \times Fare_{FMODTaxi} \\
V_{FMODSharedTaxi} &= -3.5 - 0.45 \times TotalTime_{FMODSharedTaxi} - 0.03 \times Fare_{FMODSharedTaxi} \\
V_{FMODMinibus} &= -4 - 0.45 \times TotalTime_{FMODSharedTaxi} - 0.03 \times Fare_{Minibus}
\end{aligned}
$$

The estimation results of the Monte-Carlo simulation are presented in Tables 7.10 and 7.11. We first use a sample of 1,124 trips from the Future Mobility Survey to find profiles and simulate the experiment. The results are shown in Table 7.10; the t-statistics and p-values are reported as against the true value, not against zero. The estimated parameters are reasonably close to the true values with this relatively small sample size. We then expand the sample size to 11,240 observations to test whether the estimated parameters converge on the true parameters for a sufficiently large sample size.[4] The estimation results, presented in Table 7.11, suggest that they do. The parameters are generally closer to the true values and are estimated with a lower standard error.

Parameter	True Value	Est. Value	Std. Error	t-statistic[a]	p-value[a]
ASC - Walk	-2.5	-2.26	0.365	0.65	0.51
ASC - Bicycle	-3.0	-3.33	0.307	-1.07	0.28
ASC - Taxi	-3.0	-3.24	0.220	-1.09	0.28
ASC - FMOD Taxi	-3.0	-3.15	0.214	-0.70	0.48
ASC - FMOD Shared Taxi	-3.5	-3.24	0.219	1.18	0.24
ASC - FMOD Minibus	-4.0	-4.01	0.253	-0.03	0.97
β_{Time} - Car	-0.5	-0.404	0.373	0.25	0.80
β_{Time} - Walk	-2.0	-1.98	0.353	0.06	0.95
β_{Time} - Bike	-2.0	-1.55	0.594	0.76	0.45
β_{Time} - Taxi and FMOD	-0.45	-0.255	0.232	0.84	0.40
β_{Cost} - Car	-0.07	-0.0631	0.00662	1.04	0.30
β_{Cost} - Taxi	-0.03	-0.0274	0.00841	0.29	0.77

[a] Measured with a null hypothesis that the parameter is equal to the true value, not zero.

Table 7.10: Estimation Results: Monte-Carlo Simulation (1124 observations)

[4]This consists of the original 1,124 trips replicated 10 times each; the profiles were re-generated for each replication.

Parameter	True Value	Est. Value	Std. Error	t-statistic[a]	p-value[a]
ASC - Walk	-2.5	-2.37	0.144	0.90	0.37
ASC - Bicycle	-3.0	-3.08	0.087	-0.92	0.36
ASC - Taxi	-3.0	-3.16	0.068	-2.34	0.02
ASC - FMOD Taxi	-3.0	-3.06	0.068	-0.88	0.38
ASC - FMOD Shared Taxi	-3.5	-3.52	0.077	-0.26	0.80
ASC - FMOD Minibus	-4.0	-4.16	0.084	-1.90	0.06
β_{Time} - Car	-0.5	-0.573	0.115	-0.63	0.53
β_{Time} - Walk	-2.0	-2.02	0.152	-0.13	0.90
β_{Time} - Bike	-2.0	-1.70	0.167	1.80	0.07
β_{Time} - Taxi and FMOD	-0.45	-0.283	0.077	2.16	0.03
β_{Cost} - Car	-0.07	-0.0676	0.0021	1.14	0.26
β_{Cost} - Taxi	-0.03	-0.0297	0.0025	0.12	0.90

[a] Measured with a null hypothesis that the parameter is equal to the true value, not zero.

Table 7.11: Estimation Results: Monte-Carlo Simulation (11240 observations)

Chapter 8 discusses improvements that can be made to the stated preference survey in the future. These primarily focus on the addition of modes, both existing and new. We also discuss ways to extend the survey to estimate demand for Mobility as a Service packages, and how driving times could be made more realistic through the use of time-dependent travel times from Google Directions.

Chapter 8

Extensions

This chapter presents several ways the stated preference survey could be enhanced to include a greater variety of modes, model the demand for subscription services and present more realistic car travel times. The pilot survey in Chapters 6 and 7 has been kept relatively simple to enable us to test it with a small group of users and estimate a mode choice model with relatively few parameters. If, however, the Future Mobility Survey is deployed on a very large scale, such as for a metropolitan area, the functions of the stated preference component could be expanded.

8.1 Adding Modes

8.1.1 Motorcycle

Figure 8-1 shows how the motorcycle mode can be implemented in the survey. The presentation is similar to that of the car mode, in Figure 6-4, in that the four steps in a trip are: access, riding, parking and egress. As with the car design, there are potential fuel, parking and toll costs. In the design, these costs would be offered as percentage discounts from the car version, since motorcycles do not consume as much fuel, are typically cheaper to park and usually pay lower tolls than cars. Similarly, motorcycles may be able to use transit lanes, so a design that provides some time saving from the car travel time (similarly to the taxi design in Table 7.4) may also be appropriate. Even if we choose to lower both time and cost compared to a car, the car is unlikely to be strictly dominated by the motorcycle, because the alternative-specific constant for motorcycle is much lower than for car (see Table 7.8).

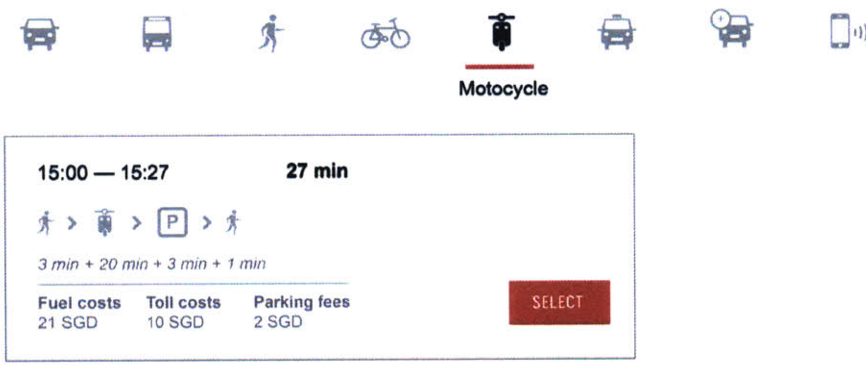


Figure 8-1: Mockup for motorcycle mode in extended survey.

8.1.2 Car Sharing

Figure 8-2 shows a possible way car sharing could be presented in the survey. In this case, the user first walks to where the car sharing vehicle is located, drives it to a car sharing space close to their destination, and then walks from there. The in-vehicle time presented here should be similar to the car in-vehicle time, with perhaps a small variation to account for the car being driven to and from slightly different locations. To be able to use car sharing, the user pays a per-hour cost, which includes fuel costs. A free, reserved parking space is available at the end of the trip, so there is no parking time (unlike the car mode), and no separate parking cost.

Figure 8-2: Mockup for car sharing mode in extended survey.

8.1.3 Bicycle Sharing

In Figure 8-3, bike share is presented under the bicycle mode group. Similarly to car share, the user walks to their closest bike share station, rides to the station nearest to their destination, and then walks the rest of the way. There is also a nominal fee for use of the bicycle.

8.1.4 Access and Egress Modes for Public Transportation

The presentation of transit in the pilot (Figure 6-6) assumes that walking is the only access and egress option for all forms of transit. However, many other modes can also be used to access bus stops and railway stations, such as car, taxi or bicy-

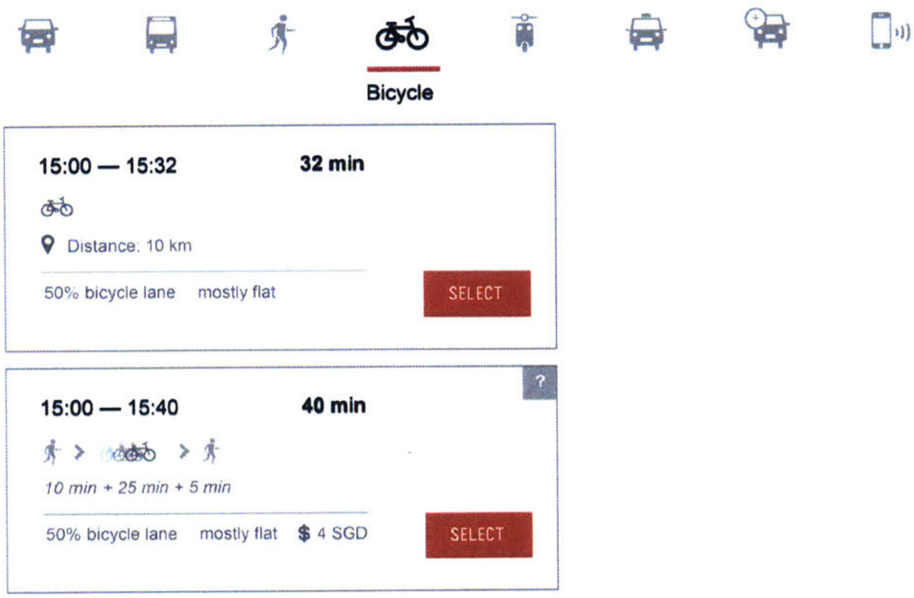


Figure 8-3: Mockup for bicycle sharing mode in extended survey.

cle. Alternative access options are particularly important in suburban locations, where there are relatively few public transportation options, and services to the city center (such as commuter rail) make relatively widely spaced stops.

Access options are also especially relevant for modes that perform a 'last mile' option, such as bike sharing and taxis. Figure 8-4 shows how access options can be selected by the user, with a drop-down menu. Selecting different options changes the sequence of times and costs. For example, the lower image illustrates a park-and-ride option at a railway station. The driving time is included, as is a parking time and an access time to the railway station. And fuel, toll and parking costs have also been included. Egress options can also be included by providing a second drop-down menu. These drop-down menus can include any number of modes we reasonably think could be used to access transit, such as taxis, bicycle sharing and FMOD.

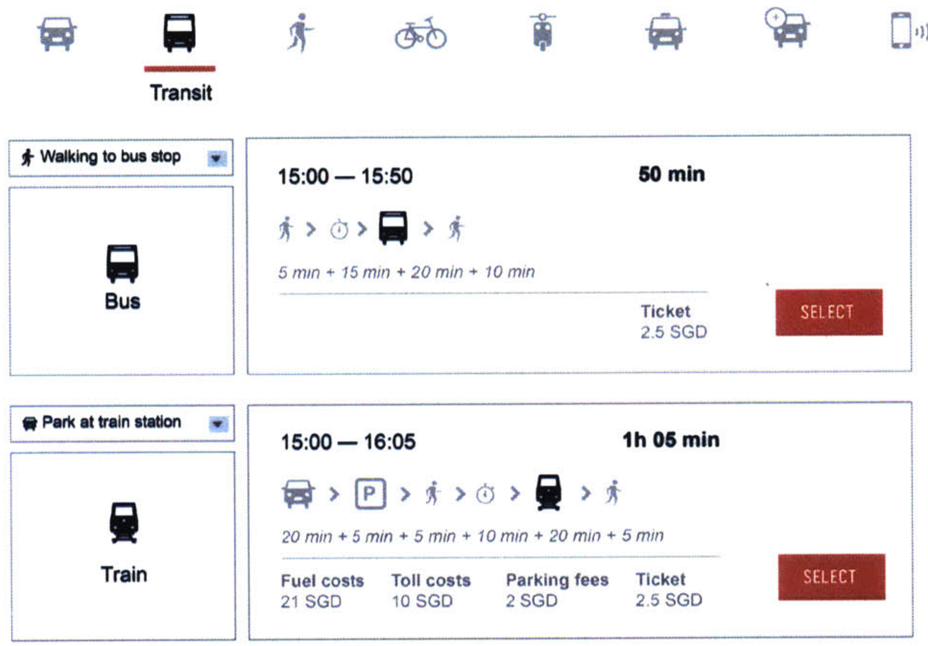


Figure 8-4: Mockup for transit access modes in extended survey.

8.2 Long-Term Decision Survey for Subscriptions

In its current form, the survey presents the marginal cost of using different travel options. That is, even if the user has a travel pass that gives them free transit for a month, the values shown do not reflect this. This is also true for other modes, such as FMOD, which could feasibly be subscribed to, providing a certain number of free rides for a month. Bicycle sharing often operates on a similar principle - frequent users typically buy annual memberships to the system, which allow unlimited rides of up to 30 minutes for one year.

There are a number of ways subscription services could be represented in the survey. We could ask users in the pre-survey about whether they own any such passes, and how much they cost. If they owned a pass, the cost would be shown as zero, which is more realistic for the user. But a new approach is needed for new transportation services that don't yet exist, such as a subscription to FMOD.

Decisions about subscriptions are likely to be made on a longer-term basis, and not for a single trip. Including the option to buy them as an option within a trip-based stated preference survey is not a sensible approach, because the context the decision is being made in is a long-term one, not a short-term one.

What is needed, therefore, is a longer-term decision survey, where the user makes decisions about packages prior to making decisions about trips. The packages presented could be:

- Transit passes
- Bike share memberships
- Car sharing memberships
- Monthly parking at the usual work or school location
- A subscription to FMOD, allowing a certain number of trips per month.

A separate design would be needed for these subscriptions, and if the user chooses any of them, the corresponding marginal cost should be zero for that mode. Importantly, we can also include Mobility as a Service subscriptions that combine multiple modes. Different packages could be offered, with different amounts of free trips on taxi, transit, FMOD, bicycle sharing and so on.

Another approach would be to first monitor the user's travel behavior over an initial period, and then use this information to present a longer-term survey to inform the rest of the data collection. For example, we could monitor a user's use of transit, taxis and bicycle sharing for a week, present this information to them and then use it to inform their decision about whether to buy a subscription service at a discount from the pay-as-you-go rate. Conducting the longer-term decision survey in the middle of the data collection period would also allow us to see how different people change their mobility once they purchase a subscription.

8.3 Time-Dependent Travel Times

One new feature recently introduced into Google Maps is the ability to set arrival or departure times, and obtain back a travel time range. This is shown in Figure 8-5, which illustrates directions from suburban Boston to MIT, which is located close to the city center. The departure time has been set to 8 a.m. on a Monday; as a result, there is high variability in the travel time range shown, from 18 to 35 minutes. In Figure 8-6, the departure time was changed to 9 p.m., and the range given narrows significantly, because this is an off-peak period.

Currently, the experimental design in Chapter 7 handles this in a fairly simple way, applying different sets of travel time ratios for peak and off-peak periods. This has undesirable side-effects for its application across different parts of cities and different regions. For example, even in the peak time, a reverse commute would likely have significantly less variability, but this is not reflected in the levels presented. Further, for smaller regions and towns, in many cases there may well be almost zero variability, even in peak commuting times. Take, for example, a commute in Cape Cod, Massachusetts, at Monday at 8 a.m. - as shown in Figure 8-7, no travel time variability is shown at all. In these situations, showing variability beyond a few minutes is not likely to be perceived by the respondents as realistic.

If we can find a way to obtain these travel time ranges for trips users make, we could use the range provided by Google Directions as the basis for the travel times we present. This allows the survey to be highly adaptive to different driving environments in space and time.

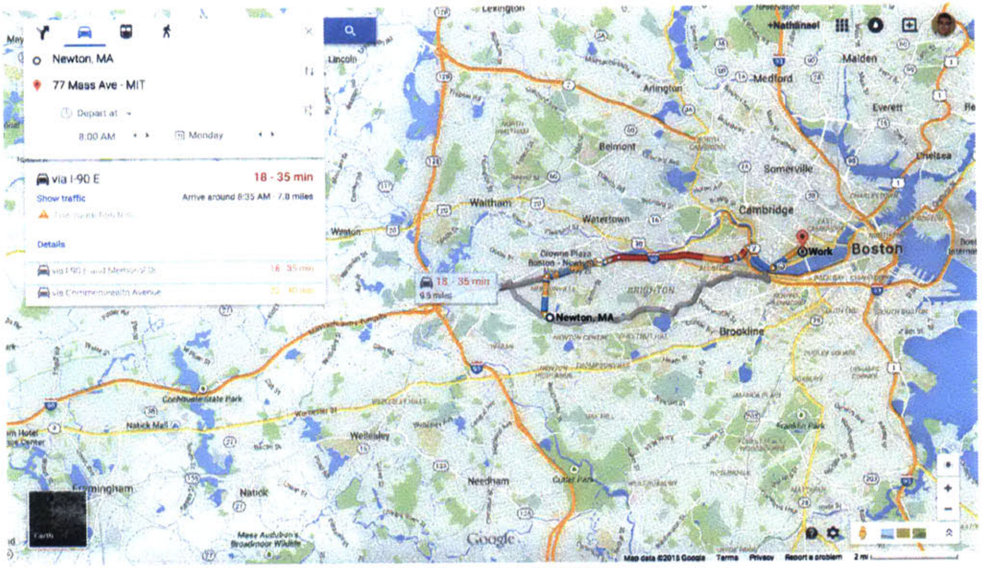


Figure 8-5: A commute from Newton, MA, to Cambridge, MA, shown by Google Maps, with a departure time of 8 a.m. on Monday. Source: Google Maps (maps.google.com).

Figure 8-6: A commute from Newton, MA, to Cambridge, MA, shown by Google Maps, with a departure time of 9 p.m. on Monday. Source: Google Maps (maps.google.com).

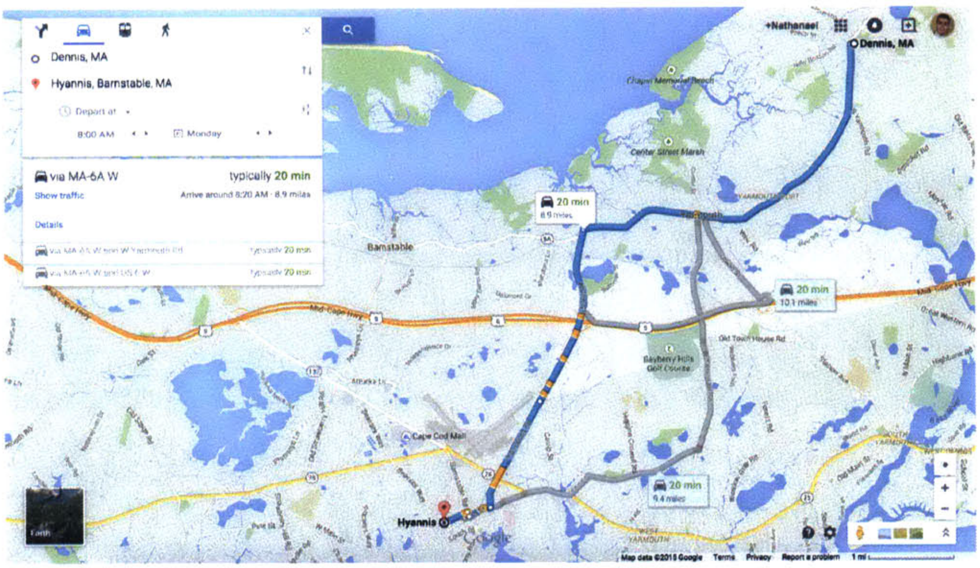


Figure 8-7: A commute from Dennis, MA, to Hyannis, MA, shown by Google Maps, with a departure time of 8 a.m. on Monday. Source: Google Maps (maps.google.com).

Chapter 9

Conclusion

There are two main ways behavioral data can be collected for transportation: through a revealed preference survey, which collects information on choices in the world of today, or a stated preference survey, which collects information on choices in a hypothetical world. While revealed preference data suffers from less potential bias, stated preference data is necessary to estimate demand for new modes that do not yet exist. It also allows a wider variety of policy scenarios to be tested, because we can control the levels of the experiment and widen their ranges from outside what they currently are.

GPS technology has greatly improved the quality of revealed preference data. Survey respondents typically underreport trips when they are asked to fill out diaries, especially for short trips that are easier to forget. GPS has resolved this issue, but until the introduction of the Future Mobility Survey data was collected using dedicated loggers, which either lived in cars or were forgotten by respondents when they took trips. The Future Mobility Survey solves this problem by making the smartphone the GPS logger; respondents have an incentive (beyond the incentive of completing the survey) to being the logger along with them.

This main contribution of this thesis is a context-aware stated preference survey that uses revealed preference data from the Future Mobility Survey as a reference point to construct hypothetical scenarios. It does this by getting the origin and destination of the trip from home to the primary destination for the day, fetching information from web services to get travel times and distances for various modes, and then altering that information to present to the users.

The survey includes a wide variety of modes, and with this comes the challenge of presenting this information in a way that is intuitive to respondents. We achieved this through the use of an interface that mimics a trip planner, with extensive use of symbols to represent a large number of modes and, within those modes, the different stages of a trip a user can experience. This is particularly important given the survey occurs straight after another survey: the validation of GPS-inferred trips and stops.

The stated preference component of the Future Mobility Survey represents the beginning of a new generation of stated preference survey that is realistic and intuitive to respondents. It can be efficiently deployed using smartphones and web-based validation, potentially to many thousands of users, in any city in the world. In an environment where information technology is rapidly spawning new types of transportation services, the need for surveys that can accurately and efficiently estimate their potential demand is becoming more critical. Just as technology has transformed existing transportation services and created new ones, so too can it create surveys that help us to better understand mobility patterns and behavior.

Bibliography

Atasoy, B., Ikeda, I. and Ben-Akiva, M. (2015, forthcoming), 'Optimizing a Flexible Mobility on Demand System', *Transportation Research Record: Journal of the Transportation Research Board* .

Avis Budget Group (2015), 'Avis Budget Group :: Zipcar'.
URL: *http://www.avisbudgetgroup.com/company-information/our-brands/zipcar/*

Ben-Akiva, M. and Lerman, S. (1985), *Discrete Choice Analysis: Theory and Application to Travel Demand*, 1st edn, MIT Press.

Ben-Akiva, M., Toledo, T., Santos, J., Cox, N., Zhao, F., Marzano, V. and Lee, Y. (2015, forthcoming), 'Freight Data Collection Using GPS and Web-Based Surveys: Insights from U.S. Truck Drivers' Survey and Perspectives for Urban Freight', *Case Studies on Transport Policy* .

Bierlaire, M., Axhausen, K. and Abay, G. (2001), The Acceptance of Model Innovation: The Case of Swissmetro, *in* '1st Swiss Transport Research Conference'.

Bowman, J. (1998), The Day Activity Schedule Approach to Travel Demand Analysis, PhD thesis, Massachusetts Institute of Technology.

Bridj (2015), 'Bridj'.
URL: *http://www.bridj.com/*

Bureau of Transport Statistics (2013), Sydney Strategic Travel Model (STM): Modelling Future Travel Patterns, Technical report, Transport for New South Wales.

Car2Go (2015), 'What is Car2Go?'.
URL: *https://www.car2go.com/en/austin/what-is-car2go/*

City and County of San Francisco: Office of the Mayor (2015), 'New Bay Area Bike Share Proposal Expands Across Five Bay Area Cities'.
URL: *http://www.sfmayor.org/index.aspx?recordid=827&page=846*

Cottrill, C., Pereira, F., Zhao, F., Dias, I., Lim, H., Ben-Akiva, M. and Zegras, P. (2013), 'Future Mobility Survey: Experience in Developing a Smartphone-Based Travel Survey in Singapore', *Transportation Research Record: Journal of the Transportation Research Board* (2354), 59 – 67.

Ferenstein, G. (2014), 'Armed With a Fleet of Teslas, Shift Takes Aim at Uber, Public Transit'.
URL: *http://www.washingtonpost.com/blogs/innovations/wp/2014/11/07/armed-with-a-fleet-of-teslas-shift-takes-aim-at-uber-public-transit/*

Fifer, S., Greaves, S., Rose, J. and Ellison, R. (2010), 'A Combined GPS/Stated Choice Experiment to Estimate Values of Crash-Risk Reduction', *Journal of Choice Modelling* **4**(1), 44 – 61.

Fifer, S., Rose, J. and Greaves, S. (2014), 'Hypothetical Bias in Stated Choice Experiments: Is It a Problem? And If So, How Do We Deal with It?', *Transportation Research Part A* **61**, 164 – 177.

Fishman, E., Washington, S. and Haworth, N. (2013), 'Bike Share: A Synthesis of the Literature', *Transport Reviews* **33**(2), 148 – 165.

Frumin, M. (2010), Automatic Data for Applied Railway Management: Passenger Demand, Service Quality Measurement, and Tactical Planning on the

London Overground Network, PhD thesis, Massachusetts Institute of Technology.

Giaimo, G., Anderson, R., Wargelin, L. and Stopher, P. (2010), 'Will It Work? Pilot Results from First Large-Scale Global Positioning System–Based Household Travel Survey in the United States', *Transportation Research Record: Journal of the Transportation Research Board* (2176), 26 – 34.

Google (2015), 'Transit - Google Maps'.
URL: *http://www.google.com/landing/transit/cities/index.html*

Greaves, S., Fifer, S., Ellison, R. and Germanos, G. (2010), 'Development of a Global Positioning System Web-Based Prompted Recall Solution for Longitudinal Travel Surveys', *Transportation Research Record: Journal of the Transportation Research Board* (2183), 69 – 77.

Hertz (2015), 'Hertz 24/7 - Hourly Car Rental. Rental Cars by the Hour, Day or Week'.
URL: *https://www.hertz247.com/*

Hess, S. and Rose, J. (2009), 'Should Reference Alternatives in Pivot Design SC Surveys be Treated Differently?', *Environmental and Resource Economics* **42**, 297 –317.

Hietanen, S. (2014), ''Mobility as a Service' - The New Transport Model?'.

Lane, C. (2005), 'PhillyCarShare: First-Year Social and Mobility Impact of Carsharing in Philadelphia, Pennsylvania', *Transportation Research Record: Journal of the Transportation Research Board* (1927), 158 – 166.

Lang, Y. (2013), Modeling Preferences for Innovative Modes and Services: A Case Study in Lisbon, Master's thesis, Massachusetts Institute of Technology.

Leap Transit (2015), 'Leap - Your daily commute. Redesigned.'.
URL: *http://leaptransit.com/*

Li, S. (2013), Technical Report for 2.3.1 Mode Choice for Work Tour to Usual Location. Unpublished.

Louviere, J., Hensher, D. and Swait, J. (2000), *Stated Choice Methods: Analysis and Application*, Cambridge University Press, Cambridge, United Kingdom.

Martin, E. and Shaheen, S. (2014), 'Evaluating Public Transit Model Shift Dynamics in Response to Bikesharing: A Tale of Two U.S. Cities', *Journal of Transport Geography* **41**, 315 – 324.

Martin, E., Shaheen, S. and Lidicker, J. (2010), 'Impact of Carsharing on Household Vehicle Holdings: Results from North American Shared-Use Vehicle Survey', *Transportation Research Record: Journal of the Transportation Research Board* (2143), 150 – 158.

Morikawa, T. (1989), Incorporating Stated Preference Data in Travel Demand Analysis, PhD thesis, Massachusetts Institute of Technology.

Murakami, E. and Wagner, D. (1999), 'Can Using Global Positioning System (GPS) Improve Trip Reporting?', *Transportation Research Part C* **7**, 149 – 165.

Murphy, J., Allen, P., Stevens, T. and Weatherhead, D. (2005), 'A Meta-Analysis of Hypothetical Bias in Stated Preference Valuation', *Environmental and Resource Economics* **30**, 313 – 325.

New York City Department of Transportation (2014), 'NYC DOT, Alta and Citi Announce Agreement to Expand and Enhance Citi Bike Program in New York City'.
URL: *http://a841-tfpweb.nyc.gov/dotpress/2014/10/citi-bike-program-in-new-york-city/*

Nicholson, W. and Snyder, C. (2007), *Microeconomic Theory: Basic Principles and Extensions*, 10th edn, Thomson/South-Western, Mason.

Ohmori, N., Harata, N. and Nakazato, M. (2005), 'GPS Mobile Phone-Based Activity Diary Survey', *Proceedings of the Eastrn Asia Society for Transportation Studies* **5**, 1104 – 1115.

Rose, J. and Bliemer, M. (2009), 'Constructing Efficient Stated Choice Experimental Designs', *Transport Reviews: A Transnational Transdisciplinary Journal* **29**(5), 587 – 617.

SBB (2015), 'SBB: SwissPass - the key to your mobility.'.
URL: *http://www.sbb.ch/en/travelcards-and-tickets/railpasses/swisspass.html*

Shaheen, S., Guzman, S. and Zhang, H. (2012), Bikesharing Across the Globe, *in* J. Pucher and R. Buehler, eds, 'City Cycling', MIT Press, chapter 9, pp. 183 – 210.

SHIFT (2015), 'SHIFT - Car Sharing, Rides and Bikes in Las Vegas'.
URL: *https://shiftconnects.com/*

Simas Oliveira, M., Vovsha, P., Wolf, J., Birotker, Y., Givon, D. and Paasche, J. (2011), 'Global Positioning System–Assisted Prompted Recall Household Travel Survey to Support Development of Advanced Travel Model in Jerusalem, Israel', *Transportation Research Record: Journal of the Transportation Research Board* (2246), 16 – 23.

Sochor, J., Strömberg, H. and Karlsson, M. (2014), Travelers' Motives for Adopting a New, Innovative Travel Service: Insights from the UbiGo Field Operational Test in Gothenburg, Sweden, *in* '21st World Congress on Intelligent Transportation Systems'.

Steininger, K., Vogl, C. and Zettl, R. (1996), 'Car-Sharing Organizations: The Size of the Market Segment and Revealed Change in Mobility Behavior', *Transport Policy* **3**(4), 177 – 185.

Stopher, P., FitzGerald, C. and Xu, M. (2007), 'Assessing the Accuracy of the Sydney Household Travel Survey with GPS', *Transportation* **34**, 723 – 741.

Stopher, P., Wargelin, L., Minser, J., Tierney, K., Rhindress, M. and O'Connor, S. (2012), GPS-Based Household Interview Survey for the Cincinnati, Ohio Region, Technical report, Ohio Department of Transportation Office of Research and Development; Federal Highway Administration.

Sun, Y. (2013), Decision Making Process and Factors Affecting Truck Routing, Master's thesis, Massachusetts Institute of Technology.

Theis, G. (2011), Incorporating Attitudes in Airline Itinerary Choice: Modeling the Impact of Elapsed Time, PhD thesis, Massachusetts Institute of Technology.

Train, K. and Wilson, W. (2008), 'Estimation on Stated-Preference Experiments Constructed from Revealed Preference Choices', *Transportation Research Part B* **42**, 191 – 203.

Transport for New South Wales (2015), 'Sydney Household Travel Survey'.
URL: *http://www.bts.nsw.gov.au/Statistics/Household-Travel-Survey/default.aspx*

Uber (2015), 'Announcing UberPool, Carpooling With Uber | Uberblog'.
URL: *http://blog.uber.com/uberpool*

UbiGo (2015), 'About (English) | UbiGo'.
URL: *http://web.viktoria.se/ubigo/las-mer/about-english/*

Urban, G., Hauser, J., Qualls, W., Weinberg, B., Bohlmann, J. and Chicos, R. (1997), 'Information Acceleration: Validation and Lessons from the Field', *Journal of Marketing Research* **34**(1), 143 – 153.

Via (2015), 'Via | New York City's Smartest Rideshare App - Just $5 A Ride'.
URL: *http://ridewithvia.com/*

Walker, J., Wang, Y., Thorhauge, M. and Ben-Akiva, M. (2015), D-Efficient or Deficient? A Robustness Analysis of Experimental Designs in a VOT Estimation Context, *in* 'Transportation Research Board 94th Annual Meeting'.

Zhao, F., Pereira, F., Ball, R., Kim, Y., Han, Y., Zegras, P. and Ben-Akiva, M. (2015), 'Exploratory Analysis of a Smartphone-Based Travel Survey in Singapore', *Transportation Research Record: Journal of the Transportation Research Board* **Forthcoming**.

Zipcar (2015), 'Zipcar Overview | Zipcar Press Center'.
URL: *http://www.zipcar.com/press/overview*

Zmud, J. and Wolf, J. (2003), Identifying the Correlates of Trip Misreporting - Results from the California Statewide Household Travel Survey GPS Study, *in* '10th International Conference on Travel Behaviour Research', Lucerne.