The Potential of Online Respondent Data for Choice Modeling in Transportation Research

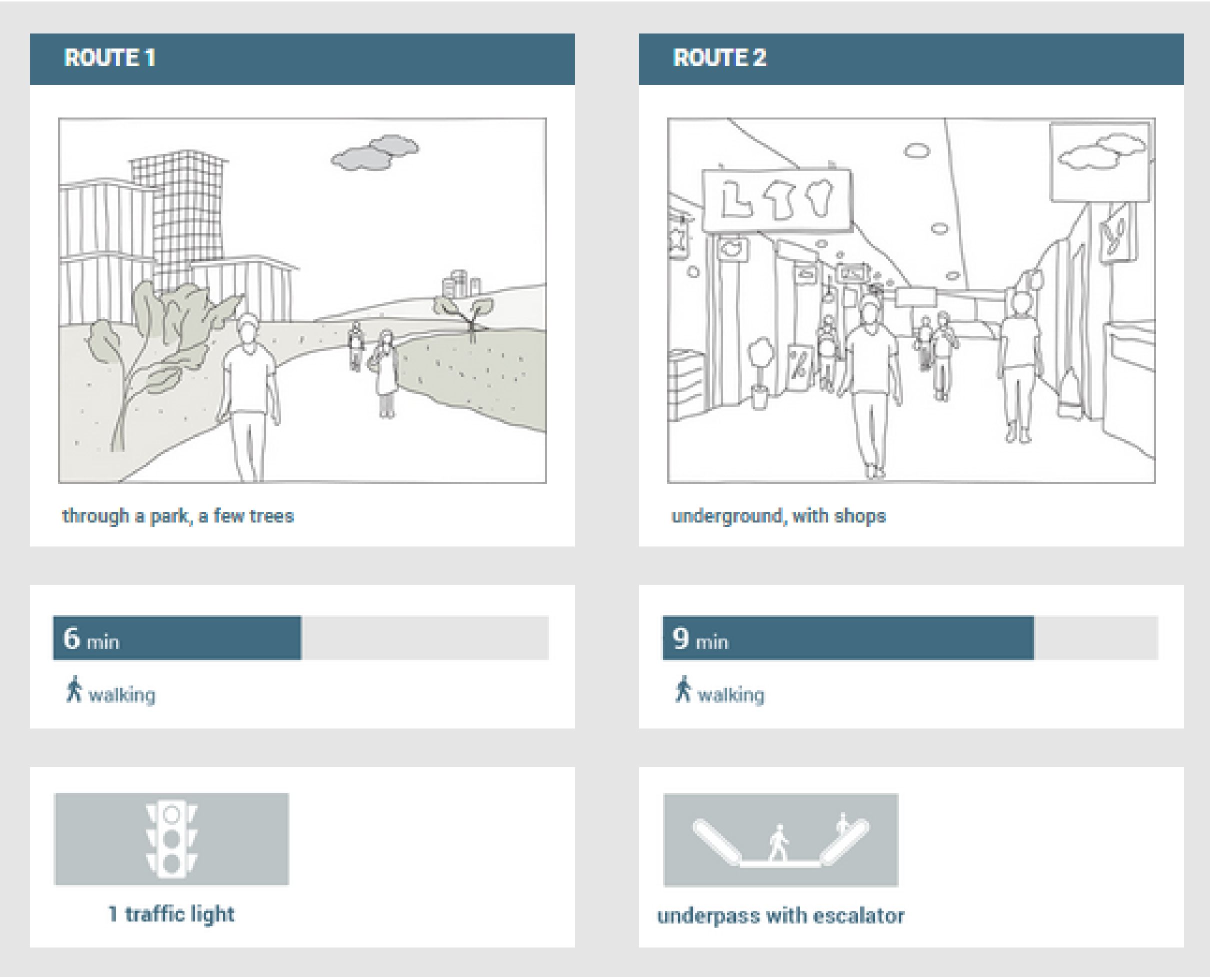
Evidence from Stated Preference Experiments using Web-based Samples

Master Thesis

Author(s):
Hoffer, Brice

Publication date:
2015

Permanent link:
https://doi.org/10.3929/ethz-a-010526606



The Potential of Online Respondent Data for Choice Modeling in Transportation Research:

Evidence from Stated Preference Experiments using Web-based Samples

Brice Hoffer

Supervision:
Prof. Dr. Kay W. Axhausen
Dr. Alexander Erath
Michael Van Eggermond, MSc

Master Thesis
Spatial Development and Infrastructure Systems **June 2015**

(SEC) SINGAPORE-ETH CENTRE 新加坡-ETH研究中心
(FCL) FUTURE CITIES LABORATORY 未来城市实验室

Acknowledgement

First, I would like to thank Prof. Kay Axhausen for his guidance and for giving me the chance to complete my master thesis at the Singapore-ETH Centre. I am also grateful to Dr. Alexander Erath who supervised the totality of my work in Singapore. Especially, I appreciated his support and knowledge in the computation of the choice models. I thank a lot Michael Van Eggermond for his great help in the generation of choice sets and in the online implementation of the surveys. I am grateful to Basil Schmid for his support in the calculation of the response burdens of the questionnaires. I also thank all the members of the transportation planning module of the Future Cities Laboratory who were always available to give a hand.

I thank Nathalie Wandel for her understanding and support during these four months abroad. Finally, I am extremely grateful to my parents Claudine and Bernard who allowed me to study what I like and who supported me during my whole education time.

Table of contents

List of tables

List of figures

Abbreviations

ACQ	Attentiveness Check Question
CBD	Central Business District
FCL	Future Cities Laboratory
HIT	Human Intelligence Task
IIA	Independence from irrelevant alternatives
IID	Independently and Identically Distributed error-terms
MNL	Multinomial Logit
MTurk	Amazon Mechanical Turk
RP	Revealed Preference
SP	Stated Preference
SSI	Survey Sampling International
URA	Singapore Urban Redevelopment Authority

The potential of Online Respondent Data for Choice Modeling in Transportation Research: Evidence from a Stated Preference Experiment using Web-based Samples

Brice Hoffer
Bordgeais 13
2800 Delémont

+41 79 213 48 16
brice.hoffer@gmail.com

June 2015

Abstract

The aim of this thesis is to analyze the potential of online survey services for conducting stated preference experiments in the field of transportation planning. Several web-products for hosting questionnaires are evaluated considering important features required when conducting a stated preference survey. Based on this evaluation, the open-source platform LimeSurvey is the most appropriated for this kind of research. A stated preference questionnaire about pedestrians' route choice in a Singaporean urban context is submitted to three different samples: a commercial sample of Survey Sampling International (SSI), workers of Amazon Mechanical Turk (MTurk) and a convenience sample of urban planning specialists. Samples' quality is investigated based on the answers' collection time, the survey completion rate, the answer consistency, the representativeness of the Singaporean population and the reliability of the choice behavior. The last criterion is assessed by computing multinomial choice models. After evaluation, the SSI sample is elected as the best online sample investigated. Especially, it provides the only one valuable choice model. The reliability of the choice behavior of this sample is comparable to the one of a Singaporean field survey used as reference. Finally, effects of stated preference design variations on answer consistency and reported choice behavior are investigated. It is observed that respondents who answer surveys with pictorial representations perceive better the experiment attributes, especially the walking time.

Keywords

Choice modeling; Stated preference; Pedestrian; Walkability; Online samples; Internet samples; Amazon Mechanical Turk; Transportation; Singapore; QuestionPro; LimeSurvey; Biogeme.

Preferred Citation Style

Hoffer, B. (2015). 'The Potential of Online Respondent Data for Choice Modeling in Transportation Research: Evidence from a Stated Preference Experiment using Web-based Samples', *Master Thesis*, Future Cities Laboratory, ETH-Singapore Centre for Global Environmental Sustainability (SEC), Singapore.

1 Introduction

1.1 Background

In addition to conventional face-to-face interviews and mail surveys, increasing internet usage and coverage offer researchers new opportunities for gathering answers to questionnaires. Online surveys especially might be very interesting in terms of quick collection time (Sheehan and McMillan 1999) and worldwide respondent pool (Kavanaugh et al. 2013). However, it is questionable if the quality of the data collected by this remote approach equals the one of in-person experiments or of mail questionnaires (Horton and Chilton 2010). The socio-demographic representativeness of internet sample might also be of concern (Olsen 2009). This thesis has to investigate these questions for the specific field of stated preference experiments in transportation planning. It is closely related to a research project of the Future Cities Laboratory (FCL) of the Singapore ETH-Centre for Global Environmental Sustainability. Conducted in collaboration with the Urban Redevelopment Authority of Singapore (URA), the aim of FCL project is to explore “what people value when walking in Singapore’s dense city centre and tropical climate, to enable planners to better plan for walkability”(Future Cities Laboratory 2015). A stated preference questionnaire designed in the context of the FCL project is used to conduct the present study. This experiment investigates the decision behavior of the respondents about the choice of walking routes in the urban context of Singapore.

1.2 Research Questions

The aim of this thesis is to analyze the potential of online survey services for conducting stated preference experiments in the field of transportation planning. First, a survey platform is required to implement a questionnaire on internet. Commercial and open-source services are available. These products have to be evaluated using a list of criteria accounting for the main features required to build a stated preference survey.

Secondly, internet offers different sample types. This study considers three sample types: a sample from a commercial panel, workers of Amazon Mechanical Turks and a convenience sample of FCL researchers and URA members. The quality of these different samples has to be assessed. Researchers are interested in collecting answers quickly and with a few respondents dropping out the surveys, for limiting risks of non-response bias (Christensen et al. 2014;

Heerwegh and Loosveldt 2008). These completion characteristics have to be investigated in the context of the descriptive statistics. Respondents might also provide irrational answers, which reduces the reliability of the gathered data (Hess et al. 2010). To identify this misbehavior, a consistency test is implemented in the survey. Researchers always consider a targeted population when conducting behavioral researcher. However, the representativeness of online respondents might be challenging (Olsen 2009). Questions asking for personal characteristics of the respondents are implemented in the questionnaires to assess this representativeness. Characteristics of the Singaporean population are used as a reference. Since the stated preference experiment of this study is dedicated to the walking behavior of Singaporeans, the reported choices have to be also representative. Hence, the choice behavior of the respondents of the different samples is compared to the one of a Singaporean field survey, which has been conducted by FCL researchers. Multinomial Logit models are computed with the consistent data of the different samples. Attribute valuations are compared with the values of a choice model computed with the data of the FCL study.

Thirdly, design variations of the stated preference experiments might have an influence on the answers of the survey takers (Jansen et al. 2009; Rizzi et al. 2012; Vriens et al. 1998). Three different designs are implemented in the MTurk survey to investigate this question: attributes in text form, attributes in pictorial form and attributes in both text and pictorial form. Potential effects of these different designs are assessed in the analysis of the answer consistency and of the choice behavior.

1.3 Thesis Structure

Chapter 2 provides the reader with fundamental know-how, on which this thesis relies. Methodology and approach are introduced in chapter 3. Results of the online services' evaluation are presented in chapter 4. Chapter 5 is dedicated to the descriptive statistics of the different samples. Outcome of the computed discrete choice models is presented in Chapter 6. Study results are discussed in Chapter 7. Finally, conclusions and recommendation are stated in Chapter 8.

2 Literature Review

The literature review is segmented into four sections. The first section gives a short introduction on the concept and science of walkability and its application to the case of Singapore (section 2.1). Principles of discrete choice modeling are introduced in the second section (section 2.2). The third section focuses on stated preference experiments (section 2.3). Finally, pro and contra of different survey approaches are presented in section 2.4.

2.1 Walkability in Singapore

Section 2.1.1 presents walking environment attributes, which might influence the route choice of pedestrians. Section 2.1.2 gives then a brief insight into the Singaporean urban context.

2.1.1 Relevant Attributes

In comparison to other transportation modes, the literature about pedestrian choice modeling and its potential influencing factors is quite sparse (Agrawal et al. 2008; Muraleetharan and Hagiwara 2007). Also, most of the relevant studies interested in which attributes of the urban environment tend to encourage walking, for instance benefits of land use and building density (Greenwald and Boarnet 2001; Sun et al. 2014). For the present study, it is rather important to figure out which factors might affect pedestrians' route choice. Trip purpose might have a great influence on the significance of influencing factors (Greenwald and Boarnet 2001; Owen et al. 2004; Saelens and Handy 2008; Seneviratne and Morrall 1985). Indeed, the perception and expectations are not the same when somebody walks as an exercise, to get to work or to buy groceries. Walking time is the most sensitive variable influencing route choice (Muraleetharan and Hagiwara 2007; Agrawal et al. 2008; Seneviratne and Morrall 1985); other attributes are of less relevance. Traffic safety is another attribute of significant influence (Saelens and Handy 2008). Presence of sidewalk or safe crossing infrastructure such as traffic lights increase the probability of a route being chosen (Rodríguez and Joo 2004; Agrawal et al. 2008; Ariffin and Zahari 2013). Indeed, such amenities effectively contribute to pedestrian safety (Pucher and Dijkstra 2000). Also, environment aesthetic might have a slight effect on the walking propensity (Saelens and Handy 2008; Ball et al. 2001). Seneviratne and Morrall report also a positive correlation between route choice and shopping amenities (Seneviratne and Morrall 1985). Routes passing through a park tend to be related with higher walking likelihood (Adkins et al. 2012; Guo 2009; Rodríguez and Joo 2004). Also, Adkins et al. report in their work that green streets are preferred than conventional road by surveyed people (Adkins et al. 2012). However, other studies indicate that environmental attributes mainly influence recreational walking and less transportation walking (Owen et al. 2004; Saelens et al. 2003; Agrawal et al. 2008). Thus, the effect of environmental variables on walking choice behavior is controversial amongst the literature. Of course, meteorological variables might also affect walking behavior. Ariffin and Zahari report that people tend to avoid walking when extreme weather conditions occur, e.g. high precipitations or extreme temperatures (Ariffin and Zahari 2013). This study is quite relevant for the present work since it focusses on behavior of Malaysian citizens who live in a similar climate as in Singapore. Concerning the socio-demographic characteristics of pedestrians, Humpel et al. indicate that gender might have an influence on walking behavior but without systematic effects (Humpel et al. 2004). However, Börjesson notes that for the specific topic of perceived insecurity, women tend to valuate more open spaces with visibility than men (Börjesson 2012). Concerning age influence, Seneviratne and Morrall indicate that elderly people tend to prefer routes with small crossings' number (Seneviratne and Morrall 1985). Cultural differences might also lead to different walking behavior. In a worldwide study about pedestrians' satisfaction, Martínez and Barros report that Europeans perceive walking comfort characteristics as central. In contrast, Latin Americans put more emphasis in safety and urban design attributes. In addition to different cultural perceptions, differences in the urban environment of Europe and Latin America probably play an important role in these findings. For this reason, a description of the Singaporean walking environment is provided in next section.

2.1.2 Singaporean Urban Context

Singapore counts about five and half million of inhabitants (Singapore Department of Statistics 2014). Despite of car ownership restrictions, Singapore is characterized by high traffic flow. A relatively high density of wide arteries (four lanes or more) can be observed. Beside of common traffic light systems, overhead bridges either equipped with lift, escalator or only stairs are available to cross roads. These infrastructure have typically a height of 5 m. Underground crossings are also present, especially close to metro stations. Lifts, escalators or stairs might be provided for accessing undergrounds. Based on field observations, the cycle length of traffic lights lasts about 2 minutes in Singapore. The quasi constant hot and humid weather of Singapore might make walking quite uncomfortable, especially between 10:00 AM and 5:00 PM. Also, heavy rains occur very often. Because of these meteorological conditions, covers are sometimes provided on the sidewalks. These covers create a rain protection and provide some shade for sunny times. Walking inside building might be very appreciable because of air conditioning presence. Living up to its reputation, Singapore is a very green city. Large parks are present at various locations and trees often adorn sidewalks. Sometimes, this greenery can provide some shade and rain protection as well. Finally, pedestrians' facilities usage varies depending on week- and daytime. During weekdays, walking infrastructures around metro stations, working places (Raffles Place, etc.) and schools might be crowded at peak hours (between 08:00 and 09:00 AM, respectively 06:00 and 07:00 PM). During the week-end, important crowds of pedestrians might be observed close to metro stations, main shopping amenities or recreational zones (e.g. Sentosa Island).

2.2 Discrete Choice Modeling

In various fields such as marketing, policy making or public infrastructures' planning, it is necessary to analyze the choice behavior of the population targeted by a product, respectively a policy. Especially, understanding the impact of an item single characteristic on whole choice making process is crucial to design efficiently a product or a service. For instance, if a government wants to achieve higher public transportation use, it is important to identify the factors influencing mobility choice in favor to private or public transportation. Hence, analyzing a choice behavior or an individual preference does not only enable the policy maker to understand the actual situation, but it let him also predict future states after variations of the item characteristics. Dedicated to this topic, the econometric field of choice modeling allows researchers to investigate individual choice behaviors and preferences. Depending on item nature, continuous and discrete choice behaviors might be observed (Hanemann 1984). In the context of this study, only discrete choices are considered. Next section introduces the theoretical basis of discrete choice modeling (section 2.2.1). The multinomial logit model, which is a standard model type for analyzing discrete choice behaviors, is detailed in section 2.2.2.

2.2.1 Generalities

The concept of discrete choice describes a choice behavior where the decision maker faces a group of alternatives called "choice set". According to Train, a choice set is compatible with a discrete choice modeling approach if three conditions are met (Train 2009):

- **Alternatives of the choice set are mutually exclusive:** the decision maker can only choose one alternative of the choice set.
- **The choice set is exhaustive:** it contains all existing alternatives.
- **Finite number of alternatives:** the number of alternatives can be count and is not infinite.

In particular, the third condition makes this kind of model unsuitable for continuous variables. A further important concept of discrete choice modeling is the definition of utility. Indeed, discrete choice modeling typically relies on the assumption of an rational utility-maximizing decision maker (Train 2009). Utility is an immaterial measurement describing the benefit of an individual buying a certain good or service, or completing a certain activity. The utility value is deprived of scale. This does not influence the choice behavior since utility difference between alternatives is the only relevant value (Train 2009). In the context of choice modeling, the utility of an alternative is assumed to be explained by characteristics of this alternative, like in a hedonic approach (Equation 1).

$$V_{A1,i} = V_{A1,i}(x_{1,A1,i}, x_{2,A1,i}, \ldots, x_{ni}) \qquad (1)$$

$V_{A1,i}$: deterministic utility of the alternative A_1 for an individual i [-]
$x_{1,A1,i}$, $x_{2,A1,i}$, …, $x_{n,A1,i}$: values of characteristics of A_1 for an individual i

However, deterministic Utility V is only a part of the whole Utility U on which a decision maker bases his choice. Indeed, a model might not consider all relevant attributes (missing variables), respectively account for individual taste variations as well as measurement errors. Thus, an error term ε is integrated to describe a random utility (Louviere et al. 2000). Generally, a linear attribute combination is assumed for defining a random utility model (Train 2009):

$$U_{A1,i} = \beta_{1,A1} \cdot x_{1,A1,i} + \beta_{2,A1} \cdot x_{2,A1,i} + \cdots + \beta_{n,A1} \cdot x_{n,A1,i} + \varepsilon \qquad (2)$$

$U_{A1,i}$: utility of alternative A_1 for an individual i [-]
$x_{1,A1,i}$, $x_{2,A1,i}$, …, $x_{n,A1,i}$: values of the characteristics of alternative A_1 for an individual i
$\beta_{1,A1}$, $\beta_{2,A1}$, …, $\beta_{n,A1}$: taste coefficients of alternative A_1
ε: error-term

Since utility definition has a random character, decision maker's choices need to be estimated in a probabilistic way. Hence, the choice probability becomes greater as an alternative's utility increases. The probabilistic expression of discrete choice varies the assumed error-term distribution (Louviere et al. 2000). Next section presents the multinomial logit model (MNL), which is considered a basic and well-proven discrete choice model.

2.2.2 Multinomial Logit Model

Theoretical Basis

Multinomial Logit Model assumes a logistic distribution of the error term (Wooldridge 2012). This means that unobserved effects of the model are thought to be similar to distributed extreme values (Train 2009). Thus, it reduces the error-term to so a called white noise. Considering a choice set containing *n* alternatives, following choice probability for an alternative A_1 and an individual *i* is derived:

$$P_{A1,i} = \frac{\exp(V_{A1,i})}{\sum_{j=1}^{n} \exp(V_{Aj,i})} \tag{3}$$

$P_{A1,i}$: Probability of choice A_1 for an individual *i*
$V_{Aj,i}$: deterministic utility of alternative *j* for an individual *i* [-]
n: total number of choice set alternatives

According to equation 3, the choice probability resulting from the multinomial logit model depends only on the deterministic part of the utility (*V*). This model definition with limited dependent variable ensures that the estimated choice probability lies into the unit interval (Wooldridge 2012). Also, the sum of all choice probabilities must be equal to 1. Also, MNL model relies on the axiom of independence from irrelevant alternatives (IIA). This states that the ratio of choice probabilities of two specific alternatives is independent of the presence or the probability variation of a third alternative of the choice set (Louviere et al. 2000). This axiom arises from the assumption of independently and identically distributed error-terms amongst choice set alternatives, also known as IID. However, IIA assumption is rarely met in real choice situations. Hausman test, which compares coefficient estimates before and after removing one alternative of the choice set, might be an option to check IIA validity (Louviere et al. 2000). If IIA is violated, MNL model might lead to biased choice probabilities. In order to accurately estimate choice probabilities while relaxing IIA assumption, several more sophisticated discrete choice models such as Nested Logit or Multinomial Nested Logit exist (Train 2009). However, this study is restricted to the use of MNL models.

Model's Interpretation

In contrast to linear probability models, the interpretation of coefficient estimates is not straightforward for logit models. In addition to being logistically distributed, the error-term of MNL models must have a variance of value $\pi^2/6$. However, the original variance of utility's error-term is usually equal to a multiple of $\pi^2/6$, e.g. $\varsigma \cdot \pi^2/6$ (Train 2009). Dividing the utility expression (see equation 2) by factor ς, also called scale factor, normalizes the error-term.

However, this normalization scales as well the coefficient estimates. Hence, only scaled coefficient estimates β / ς are estimated. Coefficient magnitudes cannot be interpreted directly. However, dividing the coefficient estimates among them removes the scale factor and allows an interpretation. A famous attributes' ratio is the willingness to pay. It consists in the division of any coefficient estimate by the cost coefficient estimate (Louviere et al. 2000). Equation 4 gives an example of willingness to pay for the presence of air conditioning in a transport mode.

$$WTP_{AC} = \frac{\beta_{AC}}{\beta_{cost}} \qquad (4)$$

WTP_{AC}: Willingness to pay for air conditioning in a transport mode
β_{AC}: taste coefficient of air conditioning in a transport mode
β_{cost}: taste coefficient of transport mode cost

Scaling does not affect typical values of the marginal rate of substitution such as direct and cross elasticities (Train 2009). Direct elasticity describes the percentage change of the choice probability of an alternative in relation to a percentage change of an attribute of the same alternative. In contrast, cross elasticity measures this variation with respect to the percentage change of an attribute of another alternative of the choice set (Louviere et al. 2000). Equation 5 and 6 provide the mathematical expression of these both measurements when considering an attribute x_1 and two alternatives A_1 and A_2 of the same choice set.

$$E(P_{A1,i}, x_{1,A1,i}) = \frac{\partial P_{A1,i}}{\partial x_{1,A1,i}} \cdot \frac{x_{1,A1,i}}{P_{A1,i}} = \beta_{1,A1} \cdot x_{1,A1,i} \cdot (1 - P_{A1,i}) \qquad (5)$$

$E(P_{A1,i},x_{1,A1,i})$: direct elasticity of the choice probability of A_1 for an individual i with respect to $x_{1,A1}$
$P_{A1,i}$: Probability of choosing A_1 for an individual i
$x_{1,A1,i}$: Attribute x_1 of alternative A_1 for an individual i
$\beta_{1,A1}$: taste coefficient of attribute x_1 of alternative A_1

$$E(P_{A1,i}, x_{1,A2,i}) = \frac{\partial P_{A1,i}}{\partial x_{1,A2,i}} \cdot \frac{x_{1,A2,i}}{P_{A1,i}} = -\beta_{1,A2} \cdot x_{1,A2,i} \cdot P_{A2,i} \qquad (6)$$

$E(P_{A1,i},x_{1,A2,i})$: direct elasticity of the choice probability of A_1 for an individual i with respect to $x_{1,A1}$
$P_{A1,i}$: Probability of choosing A_1 for an individual i
$x_{1,A2,i}$: Attribute x_1 of alternative A_2 for an individual i
$\beta_{1,A2}$: taste coefficient of attribute x_1 of alternative A_2

The researcher might be interested in choice characteristics of a whole population and not only of a single individual. Indeed, discrete choice models primary describe the choice behavior of a unique person. One approach might be the calculation of choice probabilities, respectively of marginal rates of substitution, using aggregated values such as sample averages (Train 2009). However, this method is not indicated as discrete choice models are not linear. Another approach called sample enumeration is more suitable (Louviere et al. 2000). It consists in calculating each individual estimate separately and then averaging the values. Following this framework, a better approximation of the aggregate effects might be obtained.

2.3 Stated Preference Experiments

Generalities of stated preference methodology are introduced in section 2.3.1. Potential issues related to stated preference experiments are outlined in section 2.3.2. Finally, challenges and opportunities of using media when conducting stated preference experiments are detailed in section 2.3.3.

2.3.1 Generalities

While choice modeling is a methodology applied for analyzing and estimating choice behavior, stated preference analysis refers to the kind of data used in the models, as well as the experimental framework applied to the data collection. There are indeed two main data types which can be used to model choice behavior: revealed preference (RP) and stated preference (SP) data. The first type, which is also known as *market data*, refers to behavior measurements in the field (Louviere et al. 2000). Data gained through a RP methodology describe an existing behavior. In contrast, SP data arise from experiments where a decision maker has to choose among hypothetical alternatives. This approach presents some advantages in comparison to the RP method. Even if RP data are closely related to effective choice behavior, they might be very limited in terms of the alternatives they are able to describe. Indeed, RP data can only report actual situations and do not give information about hypothetical alternatives, for instance the effect of the building of a new road on a regional modal split. Since SP data result from experiments where choice alternatives are defined by the researcher, they do not suffer any alternatives' limitation (Hess et al. 2010). Another advantage is that SP data are less likely subject to measurement error bias compared to RP data. Also, SP data might be used to understand choice behavior in presence of non-market goods, such as environmental services (Hess et al. 2010). Finally, SP data might be advantageous in terms of monetary costs. Indeed, RP data are usually expensive to record (Louviere et al. 2000). However, RP data are more suitable to provide reliable choice behavior's estimations when dealing with existing situations, since the behavior is reported in real time and real context. Also, SP data might have some bias arising from the experimental framework. Indeed, a realism gap likely exists between the content of a survey and the real world. Further issues related to SP data are introduced in section 2.3.2. It should be noted, that the pooling of RP and SP data might provide enhanced model estimations. Indeed, a wider range of information is collected by using this approach which combines data describing actual and hypothetical situations (Adamowicz et al. 1997). Since this method is not applied in this study, it is not further detailed. The reader can refer to Louviere et al. for more details (Louviere et al. 2000).

2.3.2 Potential Issues

This section introduces some common limitations of stated preference experiments. It focuses mainly on issues encountered in written-form surveys, since this approach is used in this study.

Irrational Respondent Behavior

Beside of model misspecification, data quality issue is the main source of model bias (Hess et al. 2010). When building a discrete choice model such as MNL model, the researcher assumes that the respondent behaves in a rational and fully compensatory[1] way. Also, axioms of the consumer preference theory such as *transitivity*[2] should hold. However, this is usually not the case either because humans are sometimes not rational or because of information processing and understanding issues related to the context of SP experiments. This might be concerning regarding the internal validity[3] of the data collected in SP experiments. There are three main behavioral issues: lexicographic behavior, non-trading behavior and inconsistent behavior. These limitations are detailed in the following subsections. Data suffering of irrational respondent behavior usually do not appear as outliers in a large dataset and have to be tracked.

Lexicographic Behavior

Considering the way the respondents use the information provided by the researcher to base a decision, SP experiments relies on the assumption of a so called single homogenous information processing strategy (Rose et al. 2013). More specifically, the survey taker should con-

[1] Fully compensatory behavior: "… gains in one attribute are traded against losses in another attribute" (Rose et al. 2013)

[2] Transitivity: if an alternative A is preferred to an alternative B, while B is preferred to an alternative C, then A has to be preferred to C (Tversky and Kahneman 1986).

[3] "External validity is an assessment of whether the causal estimates deduced from experimental research would persist in other settings and with other samples. Internal validity pertains to the question of whether causal estimates appropriately reflect the effects of the experimental manipulation among the participants in the original setting." (Berinsky et al. 2012)

sider each attribute provided. However, some respondents might show an answering pattern known as lexicographic behavior. It means that the decisional process is based on a subset or a single alternative's characteristic (Blume et al. 1989). A lexicographic behavior can be the source of biased estimations. Indeed, the person surveyed might simplify his choice because of a bad understanding of the study design. This misbehavior is called apparent lexicographic behavior. However, it might also be a rational decision strategy based on one or a few specific attributes. For example, the respondent might be only interested in the monetary costs of an alternative. This is known as natural lexicographic behavior. Differentiating natural from apparent lexicographic behavior is however not straightforward (Hess et al. 2010). The general lexicographic behavior of a respondent respective to a certain attribute might be identified. For instance, data of a survey might be analyzed considering the respondent behavior about the explanatory variable “costs”. In a preventive approach, the limitation of the number of attributes displayed might reduce the propensity of apparent lexicographic behavior. Indeed, Bradley and Daly state that individuals might more likely base their decision on a subset if many attributes are provided (Bradley and Daly 1994). Also, the authors advise that avoiding dominant[4] alternatives when building experiments might reduce the tendency of observing lexicographic behavior. According to Rose et al., data resulting from lexicographic behavior might have a high biasing potential (Rose et al. 2011). In contrast, removing such respondents might potentially lead to a self-selection bias of the choice model if a certain demographic group becomes excluded (Saelensminde 2002). Indeed, Saelensminde indicates that lower education is usually related to a higher tendency of having a lexicographic behavior (Saelensminde 1998).

Non-Trading Behavior

Non-trading behavior describes the case where a respondent always chooses the same option across an experiment (Hess et al. 2010). For instance, a survey taker always elects the car mode to travel, regardless on choice set variations in the context of a transportation SP experiment. Also, a respondent might only elect the alternative situated at the left of the survey sheet. This behavior sometimes results from a strong affinity of the respondent for an alternative (Hess et al. 2010). Also, the survey taker might employ a non-utility maximizing strategy if he is bored, suffers of fatigue or does not understand well the experiment (Hess et al. 2010). Finally, this behavior might result from policy bias where the respondent is extremely for or against an alternative if the experiment topic is sensible (Hess et al. 2010). The two last kind

[4] A dominant choice describes a choice set, where one alternative is obviously the most advantageous (Bradley and Daly 1991).

of non-trading behaviors might be problematic for the reliability of a study. Like for the lexicographic behavior, it is very difficult to differentiate the anomaly source. It is as well difficult to remove this kind of data from an answered questionnaire without creating a self-selection bias. Hence, researchers should avoid survey designs that might likely cause this biasing behavior, such as long surveys (fatigue). Furthermore, results have to be carefully interpreted when dealing with a politically sensible thematic. It should be noted that the effects resulting from fatigue might be investigated if a so called scaling approach is applied to the model estimation. A scaling approach is a statistical estimation method allowing to estimate different error-term distributions when combining different data types, such as RP and SP together (Bradley and Daly 1994). To account for respondent fatigue, the data might be separated into for instance pairwise question groups (Bradley and Daly 1994). Finally, the likelihoods of the conventional and scaling approach estimation can be compared to see if they differ a lot.

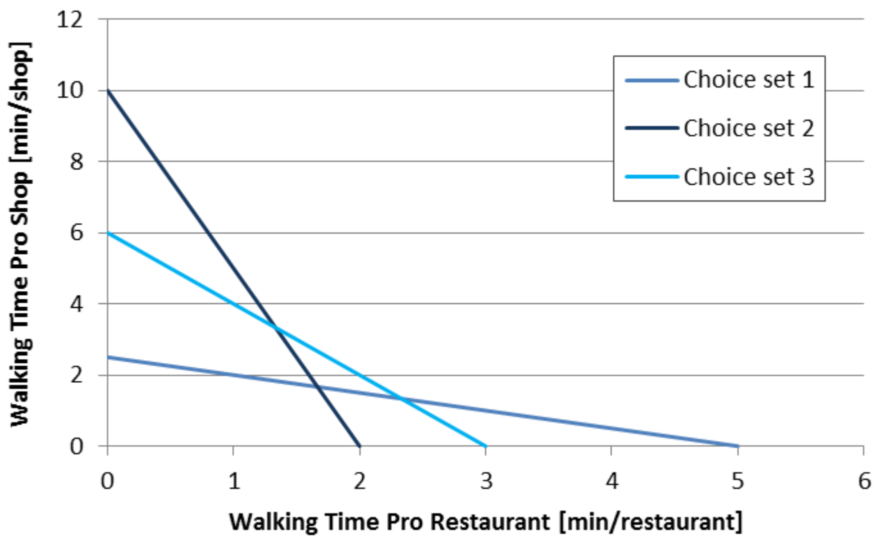
Inconsistent Behavior

The definition of inconsistent behavior is mainly based on the transgression of the transitivity axiom. For instance, a respondent accepts a high cost increase in a choice set and refuses a marginal price change later in the survey (Saelensminde 2002). Such anomalies might arise from small respondent errors, fatigue, boredom or misunderstanding of the experiment (Bradley and Daly 1994). As for other behavioral issues discussed above, discriminating the sources of the anomalies is not straightforward (Tversky and Kahneman 1986). The inclusion of a dominant choice in the experiment might perform a certain cleaning of inconsistent choices. Indeed, respondents who do not identify the dominant alternative are thought to not understand the experiment and are removed of the responses. Another possibility is to use of a so called *ray diagram* (Saelensminde 2002). This tool allows the researcher to check if the choices of a respondent are mutually consistent. Considering a hypothetical choice experiment with three attributes (Table 1), the ray diagram of Figure 1 shows graphically the "willingness to walk" for the number of restaurants and respectively the number of shops.

Table 1 Hypothetical choice experiment

		Choice Experience 1		**Choice Experience 2**		**Choice Experience 3**	
		Alternative 1	Alternative 2	Alternative 1	Alternative 2	Alternative 1	Alternative 2
Walking Time	[min]	5	10	10	20	4	10
Restaurants	[-]	4	5	5	10	3	5
Shops	[-]	3	5	5	6	4	5

Figure 1 Ray diagram for an hypothetical choice experiment



If the respondent chooses the alternative 1 in the first choice set (Table 1), it would mean that his answering domain lies on the left side of the displayed line for choice set 1 (Figure 1). Considering the further choice sets, it can be observed whether the choice domain remains consistent across the experiments. However, this methodology is quite time consuming since each respondent answer has to be analyzed individually. This consistency test is then not appropriate for checking a questionnaire which requires lots of different attributes.

Summary

Identifying and analyzing irrational behavior is very challenging. It requires a long process involving the observation of each respondent survey separately. The researcher should be very careful when deleting respondent data in order to avoid self-selection bias. Removal should only be performed when an anomaly has been clearly identified.

2.3.3 Use of Media

In order to simulate precise choice situations, the respondent has to be provided with a certain number of characteristics. Nevertheless, the survey taker may become unable to properly visualize and combine information if a large number of attributes are displayed in text form (Heiner 1983). Indeed, the respondent would rather base his choice on an incomplete set of characteristics or even make unfounded decisions when facing such situation, as outlined in section 2.3.2. Arentze et al. report a loss of accuracy of the coefficient estimates if five attributes are included rather than three in a SP experiment about transportation mode choice (Arentze et al. 2003). The use of media (e.g. images or videos) might be an option to overcome the trade-off between sufficient information and overloading of the cognitive ability of the respondentt. The present section primary focusses on the use of pictorial media. According to Green and Srinivasan, pictures enable the visualization of many attributes as well as their interactions at the same time. This perception would be difficult if displayed solely in textual form (Green and Srinivasan 1978). Another advantage of using images in stated preference analysis is the representation of qualitative characteristics, for instance the aesthetic of a pedestrian route (Hurtubia and Donoso 2014). In fact, such qualitative attributes might be difficult to perceive when designed in pure text form. However, Hurtubia and Donoso note that the creation of pictures for the purpose of choice modeling might be very complex, especially in the field of urban study. Indeed, urban areas are characterized by many different features, whose pictorial representation should be as close as possible to real world perception. Scales of street elements might be for instance under- or overrated. Gaber and Gaber point out that the correct representation of spatial element relationship in picture is difficult since three-dimensional objects are visualized in two dimensions (Gaber and Gaber 2004). Also, images may include information that is irrelevant for the conducted study (Orzechowski et al. 2005). Jansen et al. state that the presence of accidental details in pictures may non-systematically attract the attention of the surveyed persons (Jansen et al. 2009). According to this study, artistic visualizations or computer-generated images should be preferred as real photography in order to better control the attributes represented. Nevertheless, Hurtubia and Donoso state that photo-based pictures might lead to a perception which would closer to real-life feelings (Hurtubia and Donoso 2014). Considering the influence of images on SP analysis, several authors agree that attributes gain in relative importance when represented in pictorial form rather than verbal (Jansen et al. 2009; Rizzi et al. 2012; Vriens et al. 1998). However, some other studies report no evidence of differences between coefficient estimate values when the attributes are displayed in text or in image form (Arentze et al. 2003; Harline and Burris 2014; Orzechowski et al. 2005). Furthermore, Hurtubia and Donoso state in their work that tested attributes show a slightly lower magnitude when visualized in pictorial rather than in text form, as well as a lower statistical significance (Hurtubia and Donoso 2014). Considering choice model performances, the study of Arentze et al. suggests no evidence of better model fitness when using pictures for displaying attributes (Arentze et al. 2003). In the field of transportation research, Harline and Burris share this opinion (Harline and Burris 2014). In the framework of housing preferences estimation, Orzechowski et al. note that the choice model performs slightly better when using pure textual data rather the ones coming from a pictorial representation (Orzechowski et al. 2005). Hence, no real consensus exists about a positive or negative effect on the data quality resulting from image inclusion into SP experiments.

2.4 Survey Methodology

Section 2.4.1 introduces different surveying methods as well as their related advantages and disadvantages. Section 2.4.2 focusses on the comparison of different commercial online samples. Factors which might influence survey completion rate are presented in section 2.4.3.

2.4.1 Comparison of Survey Modes

The most traditional way to collect respondent data is the face-to-face survey, where the researcher and the respondent are in direct contact. Thanks to its high level of human interaction, face-to-face surveys are usually characterized by a very high response rate and thus a limited non-response bias (Christensen et al. 2014; Heerwegh and Loosveldt 2008). Non-response bias occurs if the people who do not take part to the survey differ regarding their choice behavior from the people which complete the survey (Vaske 2008). On the other hand, face-to-face data tend to show a certain social desirability bias (Yun and Trumbo 2000; De Leeuw 2005). This means that the respondents might avoid less politically correct answers and do not present their true opinion in the presence of an interviewer. This aspect might have a great impact on the data when dealing with sensible topics. More generally, the interviewer might have a certain influence on the respondent, potentially leading to an interviewer bias (Berrens et al. 2003). Respondents surveyed in an interview context are susceptible to give more extreme positive answers than people surveyed in a written form (Marta-Pedroso et al. 2007). This effect is also observed in telephone surveys, where a direct interaction between the researcher and the survey takers also occurs (Dillman et al. 2009). Telephone surveys are not further discussed in this section since the cognitive burden arising from a stated choice experiment is too high for an aural survey approach. In contrast, mail and internet surveys are suitable for this kind of complex experiment. The second one experiences a growing use because of increasing internet coverage level as well as the related cost reduction (no printing, no respondent movement, etc.) (Olsen 2009). In comparison to face-to-face technic, mail and internet surveys are characterized by quite low response rate (Heerwegh and Loosveldt 2008). Indeed, the impersonality of these approaches likely increases the chance that people do not take part to a survey (Griffis et al. 2003). It is commonly thought that mail surveys tend to be related with a higher response rate compared to online surveys (Weible and Wallace 1998; Yun and Trumbo 2000). However, several studies show higher response rate when using an online mode rather than mail distribution. Cobonaglu et al. report for instance a response rate of 44 % for emails compare to 26 % for mail (Cobanoglu et al. 2009). Olsen notes an higher response rate when using internet as survey mode in the context of his study (59 % versus 50 % for mail) (Olsen 2009). Other works report similar response rates for mail and online surveys (Raj Mehta 1995; Weible and Wallace 1998). Hence, response rate range to surveys is wide and there is no evidence a higher value of a mail approach compared to online surveys. In contrast, internet surveys have a clear advantage over mail surveys concerning the response speed. Indeed, the survey distribution through internet is immediate. There is as well no waiting time due to the sending time back of the completed questionnaire to the researcher. In their work, Sheehan and McMillan compare several studies using both mail and internet survey modes (Sheehan and McMillan 1999). In all cases, the average response time was quicker for the internet surveys. Another advantage of the internet surveys compared to the mail surveys is the possibility to integrate multimedia material, enabling the creation of more sophisticated questionnaires (Brandon et al. 2013). Furthermore, the integration of logic features make internet surveys more efficient (Olsen 2009). As stated before, mail and internet data might be biased because of non-responses. Also, a principal/agent problem might not be excluded since the questionnaire completion occurs without direct supervision of the researcher (Horton and Chilton 2010). In contrast, mail and internet data are related with a higher self-disclosure level compared to face-to-face data (Tourangeau 2004). In the case of internet surveys, some bias might occur because the population fraction which does not have internet connection (Olsen 2009). This aspect should be considered if a general population is targeted in a region with poor internet coverage. Also, the perception of the experiment in a digital form might have some impact on the respondent choice behavior (Christian et al. 2007). However, there are some indications that data collected in an online survey have similar characteristics as the data of mail surveys (Griffis et al. 2003). Finally, the World Wide Web enables an easy and quick access to an international pool of respondents, which is not the case with other survey modes. Hence, internet surveys might provide very diverse samples, especially when compared to face-to-face surveys which traditionally involve college samples (Brandon et al. 2013; Ipeirotis 2009). In summary, data from internet survey seem to be mostly representative and statistically reliable for conducting research. Furthermore, this survey mode shows many advantages in terms of costs as well as design possibility or respondent accessibility. However, it should be kept in mind that internet samples are mostly non-random, with all related issues.

2.4.2 Comparison of Online Sampling Frames

Online sampling might occur through various channels. This section focuses on web services providing answers from rewarded respondents. Such platforms are particularly interesting as they have access to pools of hundred thousand to millions of respondents of the entire world (Kavanaugh et al. 2013). A few studies analyze the socio-demographic characteristics of the respondents provided by the commercial online platform Qualtrics. In the context of a nutrition survey, Mohr et al. report that the Qualtrics sample is about gender balanced with 53 % of female subjects (Mohr et al. 2012). A similar observation is made by Spielmann and Delvert when conducting a marketing study with a women percentage of 52 % (Spielmann and Delvert 2014). About the quality of Qualtrics data, both works show a very high validity rate of 90 % and respectively 84 %. Considering a sample provided by another commercial panelist called Survey Monkey, Cook and Loraas report also very high validity level of 90 % (Cook and Loraas 2013). However, the researcher buys respondents from a black-box system when using such services. Indeed, the chosen platform delivers respondents they pooled themselves. Then, non-random sampling issues might not be excluded. Nevertheless, Gligor et al. do not note significant differences in the answers from a Survey Monkey sample and from more traditional mode such as a university sample or targeted peers when comparing them with an analysis of variance (Gligor et al. 2013).

Compare to commercial panels, the literature is very prolific about the characteristics and quality of the respondent of Amazon Mechanical Turk (MTurk). MTurk is a web-based market place where requesters and workers can register (Paolacci et al. 2010). The requester publishes a so called Human Intelligence Task (HIT) that workers can complete. HIT might be for instance translation, pictures’ sorting or surveys. The researcher might build the survey on another platform or use a rudimentary survey tool provided by MTurk (Paolacci et al. 2010). A main difference of Amazon Mechanical Turk compared to commercial samples is that MTurk does not interact directly in the survey process. The survey is simply posted on the web platform and the workers might take part to it. Another major difference is the average price of an answered survey on MTurk in comparison to commercial panels. Indeed, MTurk is extremely inexpensive with an average hourly wage of about 1.40 $. In comparison, the researcher has to pay several dollars for a survey of about 20 minutes when using commercial services (Brandon et al. 2013; Cook and Loraas 2013). Furthermore, the billing system of MTurk is very straightforward and allows the researcher to review single worker results before proceeding to payment (Paolacci et al. 2010). Amazon charges a 10 % commission on the promised salary (Chandler et al. 2013). MTurk is known to be very quick in providing answers (Berinsky et al. 2012). Indeed, the data collection of a survey typically takes less than three days on MTurk (Wang et al. 2011). Of course the questionnaire length and the reward level also influence the completion (see section 2.4.3). Like commercial panel providers, MTurk allows the researcher to target a population, for instance by setting a geographic constraint (Chandler et al. 2013). Workers who do not correspond to the restriction criterion cannot see the advertised study in the HIT pool (Paolacci et al. 2010). However, restrictions about socio-demographic attributes such as age or gender are not possible in MTurk.

Doubts rose concerning the validity of the data collected through MTurk. About the internal validity, a risk of multiple accesses to the survey by the same respondent might be considered. However, it is possible to restrict the access to the HIT to one single attempt. Also, a similar restriction might be implemented in the survey logic itself, so that in practice this issue remains marginal (Berinsky et al. 2012). Furthermore, an open-source program called Turkgate can be included in the survey process. This software improves the access security of the survey and prevents survey previews by the worker (Goldin and Darlow 2013). However, it is thought that MTurk respondents mostly complete HITs to kill time and not earn money (Ipeirotis 2008). A second concern about the internal validity of the MTurk data might be the inattentiveness of the participants. Nevertheless, researchers might use the respondent approval rate to restrict the access to their survey. Typically, researcher target workers, who have at least 95 % of approved previous HITs (Peer et al. 2013). Hence, using this kind of high-quality workers should reduce the risk of inattentiveness. But considering data external validity, some of these workers might participate very frequently to surveys. This could be a potential issue regarding the potential non-naïveté of the participants (Glinski and Glinski 1970). Indeed, people who are used to a certain experimental framework might react differently than a person facing a survey for the first time. This might be especially problematic if a same respondent attempts two or more experiments of the same researcher (Chandler et al. 2013). However, MTurk accounts for the grouping of HITs so that a respondent can only take part to a single experiment of this group. The software TurkGate also prevent this kind of issue when using an external survey link (Goldin and Darlow 2013). To ensure the external reliability of the answers, the socio-demographic characteristics of the worker pool should be close to the ones of the population targeted. The MTurk service counts around 100’000 workers worldwide (Pontin 2007). However, the majority of them are either Americans or Indians (Mason and Suri 2011). The proportion of US workers should be rising since MTurk does not allow the registration of new workers outside of the USA since 2012 (Ipeirotis 2013). Concerning gender representation, several researchers report more female samples with typically 55 to 60 % of women (Berinsky et al. 2012; Gosling et al. 2004). If a gender balanced survey is required, a quota question should be implemented into the questionnaire logic. Concerning the education and income level of the American workers, Paolacci notes that Amazon’s respondents are in tendency more educated but earn less than the average US population

(Paolacci et al. 2010). However, Berinsky notes in his work that MTurk respondent have a lower level of education compared to high-quality internet samples (Berinsky et al. 2012). Finally, both authors state that MTurk respondents are in average younger than the average US population, respectively than the high-quality internet sample studied.

2.4.3 Influence Factors of Survey Completion

Several questionnaire characteristics might influence the response rate and the completion time of a survey. The response rate might likely drop if the length and the complexity of the survey increase. These survey complexity and length are also known under the concept of response burden. However, it is of interest to understand the relationship existing between the response rate and the response burden in order to estimate the effective number of responses to a survey campaign. A unique methodology enabling this forecast for mail surveys has been proposed by Axhausen and Weis (Axhausen and Weis 2010). First, the response burdens of former surveys are calculated using a scoring system which rates the burden of each question based on its type and length. The ex-ante response burdens are then regressed on the effective response rates to derive linear prediction models. The models take into account if a prior recruitment respectively a motivation call occurs. As expected, the results of the study indicate that the response rate diminish with the extent of the response burden. An update of the study occurred in 2015 (Axhausen et al. 2015). While the first version was restricted to general survey questions, an improvement of the scoring system allows the consideration of stated choice experiments. Furthermore, a greater number of surveys are taken into account enabling the computation of non-linear models. Here also, a clear relation between an increase in the response burden and the diminution in the response rate is assessed. The effect of incentives is also analyzed. The results indicate that the presence of an incentive might lead to an augmentation of the participation rate. Considering MTurk data, several studies agree with these findings (Buhrmester et al. 2011; Berinsky et al. 2012). Furthermore, the incentive level would only affect the collection time and the answers quantity, but not the answer quality in the context of MTurk (Mason and Watts 2009). Schmid and Axhausen also state, that incentives might sometimes be counterproductive by attracting people which are not interested by the questionnaire topic (Schmid and Axhausen 2015). These respondents might then drop out of the survey or worse, provide answers of bad quality. Hence the interest and motivation for the survey thematic is also a crucial point influencing the response rate and quality. However, leverage-saliency theory states that the exclusion of some respondents because of their lack of commitment might result to a potential bias (Groves et al. 2000). Offering incentive could then be a solution to attract certain types of people which would not take part to a study in other conditions. Hence, it is not really clear which approach between providing incentives and not is the most suitable, especially because the saliency effects are difficult to quantify (Axhausen et al. 2015).

3 Methodology

This chapter presents the methodological framework of the present study. The evaluation approach of online survey services is described in section 3.1. Section 3.2 introduces the surveying approach. Questionnaire's characteristics are as well presented in this section. Finally, section 3.3 is dedicated to choice modeling.

3.1 Evaluation of Online Survey Services

One research question of this thesis is to find out if online survey platforms are suitable for conducting behavioral research (see section 1.2). Furthermore, a commercial sample is required for investigating the online respondents' quality. For this purpose, an evaluation framework is defined in section 3.1.1 for the commercial platforms, respectively in section 3.1.2 for the commercial samples.

3.1.1 Evaluation of Online Platforms

The Worldwide Web hosts many different commercial survey platforms allowing the implementation of questionnaires on internet. In the context of this study, five commercial survey services are considered: Qualtrics (Qualtrics 2015), QuestionPro (QuestionPro 2015), FluidSurveys (FluidSurveys 2015), Survey Analytics (Survey Analytics 2015) and SurveyMonkey (Survey Monkey 2015). An evaluation is conducted for identifying the pro and contra of these platforms. Furthermore, FCL researchers have had successful experience in working with the open-source tool LimeSurvey (LimeSurvey 2015) for conducting stated preference experiments. Hence, this platform is also part of the evaluation to enable a comparison between commercial and free services. The most advantageous commercial platform is then used with LimeSurvey in the study's survey campaign for conducting a real surveying experiment. To evaluate these various tools, a simple criteria list is built. This list does not detail every service the platforms might offer. Emphasis is put on advanced settings that are helpful when conducting a survey, either in a general purpose or in the specific field of transportation planning. Elements related to stated preference experiments are carefully studied. Criteria are organized in five categories:

- Media Support
- User Interface's Layout
- Advanced Features

- Survey Distribution and Accessibility
- Platform Cost

A detailed description of these categories follows in the next subsections. A summary of all the criteria is provided in Table 2. Evaluation results are presented in section 4.1.

Media Support

Criteria of this category check the possibility of including images, graphics and videos in questionnaires. Furthermore, it verifies if these media can be combined with a multiple choice question.

User Interface's Layout

Aim of this criteria group is to assess the availability of some important layout features, e.g. answering progression bar or inclusion of an institution logo in the questionnaire's footer or header.

Advanced Features

Survey platforms might offer advanced features which are useful for conducting stated preference experiments. First, the availability of a stated preference analysis (also called conjoint analysis) question type is reported. The possibility of integrating graphics and images into the conjoint analysis question is assessed. External experiment plan support, e.g. a self-defined combination of experiment attributes, is also investigated. Tracking and limitation of answering time are assessed. The possibility to randomly display questions' sequence (sequence randomization) as well as to randomly assign a certain group of questions to certain respondents (block randomization) is investigated. Availability of branching logic is reported. Branching logic means, that respondents might be redirected to different question groups based on previous answers. Finally, the availability of an incentive question type, which allows the researcher to provide easily a reward to survey takers, is considered.

Survey Distribution and Accessibility

In order to obtain answers easily and quickly, a respondent pool provided by the survey platform might be very helpful. It may also be useful to identify the IP address of the respondents. It might for instance give some information about survey taker's localization. Depending on the experiment scope, it might be necessary to fix quotas based on respondent characteristics. Automated answering by for instance web-robots might be of concern. Hence, the availability of a captcha question type, preventing automatic access, is assessed. Multiple access of a single respondent (ballot box stuffing) should be avoided as well. Finally, the availability of a question type allowing the respondent to forward the survey link to acquaintances is reported. This feature might be interesting for distributing a survey in a viral way to conduct for instance convenience sampling.

Platform Cost

Platform usage fees are considered. The cost might be reported either monthly or annually. It should be noted that the price level depends on the platform edition.

Table 2 Description of the criteria of the survey platform evaluation

Characteristic	Definition
Media Support	
-Image/Graphic	Image/graphic support
-Video	Video support
-Combination Flexibility	Combination of a media with a multiple choice question
User Interface's Layout	
-Progression Bar	Display of a progression bar
-Footer/Header Editing	Display of a logo in the header/footer
Advanced Features	
-Conjoint Analysis (CA)	Conjoint analysis question
-Design Flexibility of CA	Image support in the conjoint analysis question
-External Experiment Plan for CA	Support of an external experiment plan for the conjoint analysis
-Time Tracking	Tracking answer time
-Time Countdown	Time limit to answer
-Sequence Randomization	Randomization of the questions' sequence
-Block Randomization	Random assignation of respondents to different question groups
-Branching	Assignation of respondents to different question groups based on previous answers
-Incentive Question	Availability of an incentive question
Survey Distribution and Accessibility	
-Respondent Pool	Availability of a respondent pool
-Respondent Tracking	Identification of the IP adress of the respondent
-Quota Setting	Definition of respondent quotas based on respondent characteristics
-Captcha Question	Availability of a captcha question
-Ballot Box Stuffing Prevention	Prevention of multiple access to the survey
-Recursive invitation	Possiblility to forward the survey link to acquaintances
Platform Cost	
-Usage fees	Usage fees per month / per year [USD]

3.1.2 Evaluation of Online Panels

An analysis of commercial online respondent panels is conducted to select the commercial sample used in the survey campaign of this study. The evaluation considers panels provided by the commercial survey platforms presented in section 3.1.1. The online marketplace Amazon Mechanical Turk (MTurk) is also part of the analysis (Amazon Mechanical Turk 2015). Finally, a commercial panel provider called Survey Sampling International (SSI) is considered (Survey Sampling International 2015). The evaluation takes into account five criteria (Table 3). The cost per respondent is first reported. This price mainly depends on the population targeted, the questionnaire length and the number of answers requested. 100 answers to a 20 minutes long survey for a general population without any location constraint are considered. Geographic coverage of the sample is also investigated. Also, the panel respondents' source is analyzed. In other words, it is sought whether the provided panel is managed by the considered company or whether it comes from an external respondents' provider. Indeed, the control of the survey platform company on the respondents' quality might be questionable if the panel management is outsourced. The possibility of asking for private contact data to the respondent is assessed. This might likely be restricted by panel providers to prevent a direct contact between respondents and scientists. Finally, respondent pool protection against the automated answering is examined. Evaluation results are presented in section 4.2.

Table 3 Criteria of online panels' evaluation

Characteristic	**Definition**
Respondent Cost	Cost per answered survey [USD]
Geographic Coverage	Geographic extent of the respondents' locations
Private Contact of Respondent	Possibility of asking for respondent private contact data
Prevention of Automated Answering	Protection against web robots
Panel Source	Source of the respondent Panel [Own/Mixed/Extern]

3.2 Survey Approach

Section 3.2.1 describes the process of survey implementation and distribution on internet. Content and structure of the questionnaire used in this study are described in section 3.2.2. Answer consistency tests are introduced in section 3.2.3. Design variations of the stated preference design are presented in section 3.2.4. Section 3.2.5 summarizes the different survey types tested in this study.

3.2.1 Hosting Platforms and Distribution Method

Answers are collected in three ways. First, a questionnaire built on a commercial platform gathers answers using an appropriated commercial sample. These are elected using the evaluation frameworks defined in section 3.1. Two other surveys are implemented online with the help LimeSurvey. One of them collects answers in a convenience sampling approach. FCL researchers and URA members are here targeted. The third survey is answered by respondents of MTurk. Table 4 summarizes these combinations. A detailed description of the different survey approaches is provided in following subsections. Surveys are named according to their sample type.

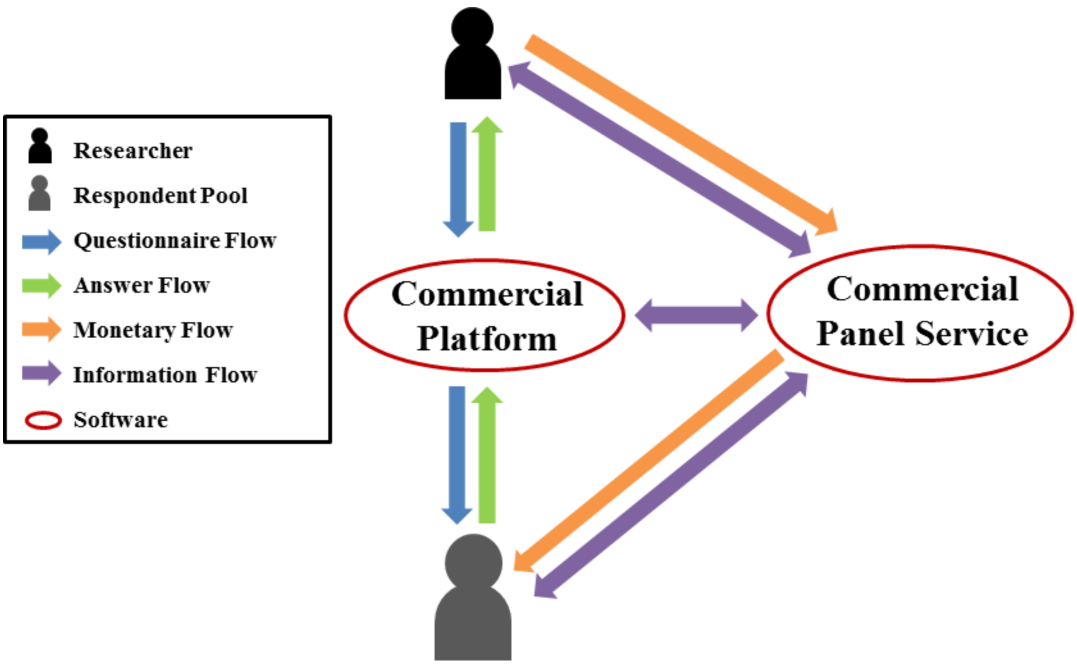
Table 4 Different platform and sample types of the study

	Sample		
Platform	Commercial	MTurk	Convenience
Commercial	X		
LimeSurvey		X	X

Commercial Survey

In the case of the commercial survey, the questionnaire is implemented on an online platform after paying usage fees. A commercial panel provider is contacted to collect answers from a targeted population. Because of the Singaporean character of the present experiment (see section 3.2.2), a Singaporean population is targeted. The survey link is provided to the panel company, which forward it to a respondent pool. The answers are directly collected by the commercial platform. The panel company remunerates the respondents. Figure 2 summarizes this process.

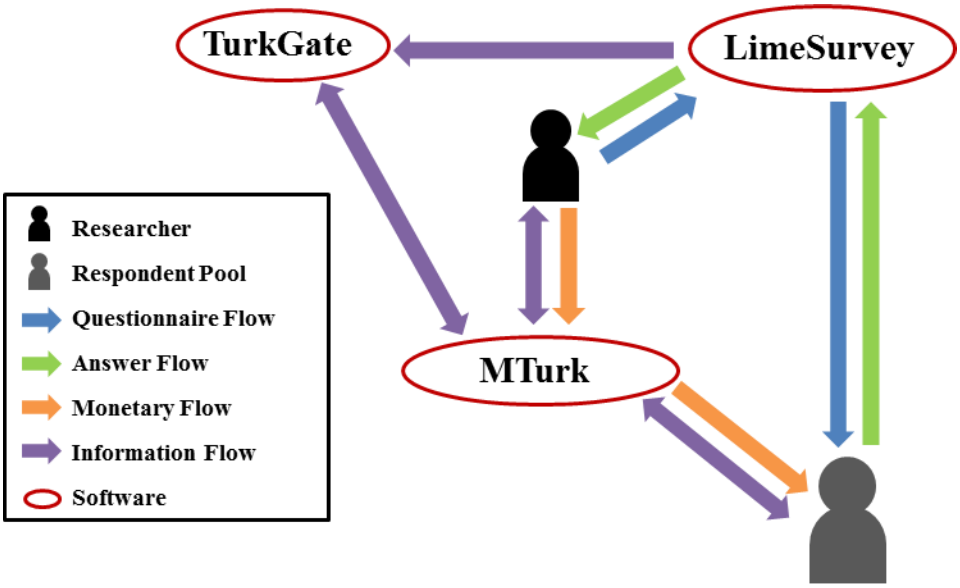
Figure 2 Flowchart of commercial survey process



MTurk Survey

MTurk survey process is very similar to commercial sample one. The difference is the involvement of a further program, TurkGate, which ensures that MTurk workers have only a single access to a published task (Goldin and Darlow 2013). Furthermore, TurkGate provides a convenient interface allowing the researcher to check the effective survey completion before rewarding the respondents in MTurk. Detailed process is presented in Figure 3. One should note that although the rewarding of the workers occurs on MTurk, the researcher has to check himself answer quality and to pay the respondents. Pre-test revealed that the Singaporean population of MTurk is too small to be surveyed in this work. Indeed, no answer was collected after 24 hours. A further pre-test targeting a Southern Asian population gave similar results. Considering literature findings, it has been decided to collect answers from American and Indian respondents because of their high presence (Ipeirotis 2013; Mason and Suri 2011). Respondents having an approval rate of 95 % or more are targeted for ensuring a certain answer reliability (Peer et al. 2013). It should be noted that the origin of MTurk respondents might have a negative effect on answer quality, since the stated preference experiment of this study is related to a Singaporean urban context. This aspect carefully considered in the further steps of the study.

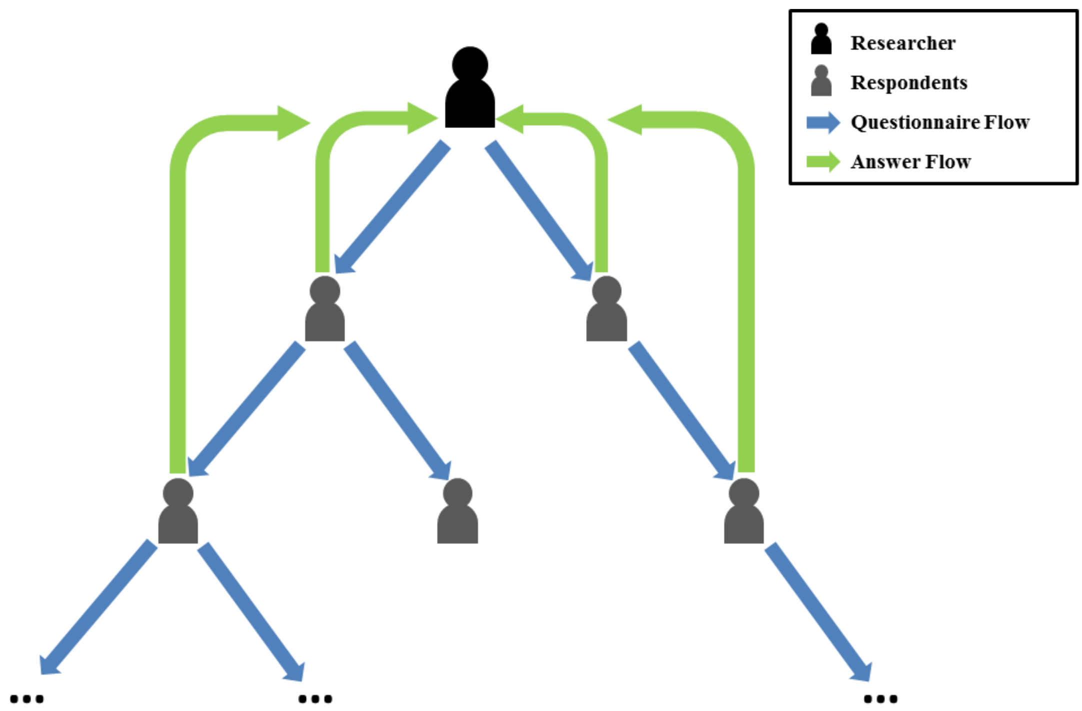
Figure 3 Flowchart of MTurk survey process



Convenience survey

Convenience survey process differs greatly from the two others, especially because of the absence of reward. The researcher contacts seed respondents, which might forward the survey link to acquaintances after completion. Although Singaporean seeds of FCL and URA are targeted, the researcher cannot ex-ante ensure that the whole sample will only comprise Singaporean respondents. Figure 4 displays the process in details.

Figure 4 Flowchart of the convenience survey process



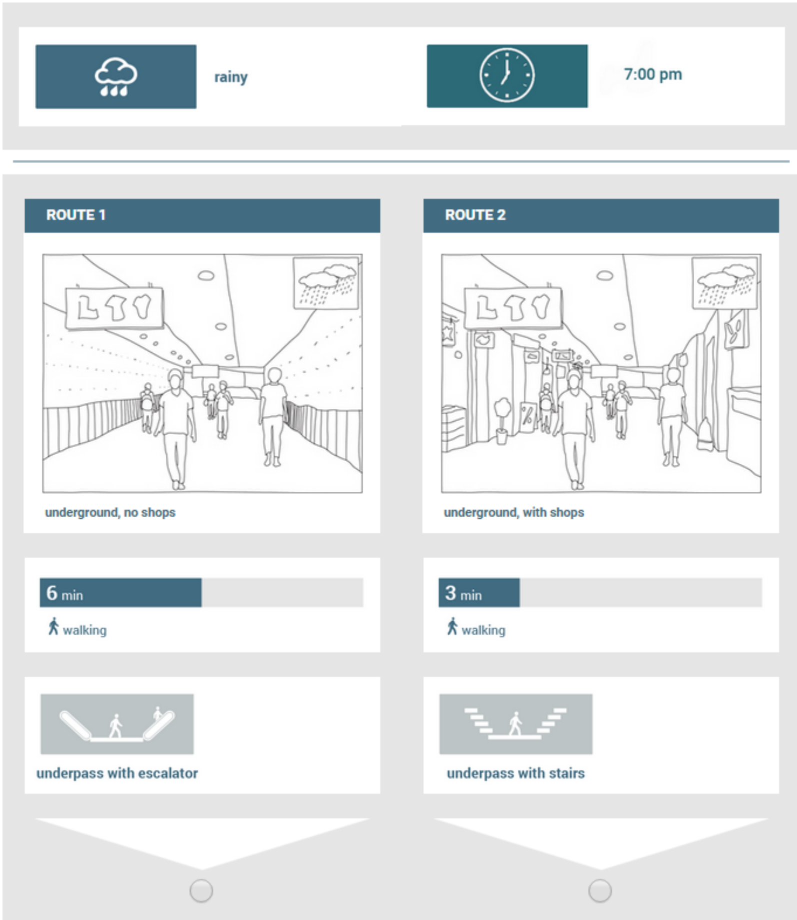
3.2.2 Questionnaire

Core of the study’s questionnaire is a stated preference experiment about pedestrians’ route choice. This experiment is related to a FCL project about walking habits in Singapore (see section 1.1). Aim of this experiment is to identify the effects of route characteristics on walking time perception. Furthermore, questions about respondent’s localization and socio-demographic characteristics are included in the questionnaire. This information must enable a comparison of the samples’ populations with Singaporean references. Walking characteristics of the respondents are also surveyed. These different question types are described in following subsections. Each question is displayed in Annex A 1.

Stated Preference Experiment

The stated preference experiment consists in a set of eight questions showing to the respondent two hypothetical walking situations related to the urban context of Singapore (Figure 5). Beforehand, the respondent is provided with detailed explanations of experiment attributes and framework. These explanations are presented in Annex A 1.

Figure 5 Example of a stated experiment question of t questionnaire



Choice sets are constructed using the stated choice generator software Ngene (ChoiceMetrics 2015). Both walking situations show similar weather conditions and day time. In contrast, they might differ in terms of route type, presence of shops or greenery, crossing characteristics, number of traffic lights and of course walking time. Alternative's attributes and levels are presented in Table 5.

Table 5 Alternative's attributes with levels

Attributes	Levels
Weather	sunny, cloudy, rainy
Day Time	9 am, 1 pm, 7 pm
Route Type	minor road, major road, park, inside building
Shops	shops, no shop
Greenery	trees, no tree
Traffic Light	1 traffic light, 2 traffic lights, no traffic light
Crossings	overpass with stairs, overpass with lift, underpass with stairs, underpass with escalator, jaywalking 2 lanes, jaywalking 4 lanes, no crossing
Walking Time	[min]

Walking Behavior

To identify the walking patterns of the different samples, a few questions concerning daily walking activities are asked. Respondents report their average walk frequency and duration for different trip goals. Table 6 describes the attributes of these questions.

Table 6 Attributes of the walking behavior

Trip Goal	Attribute
Buying Groceries	Frequency
	Walking Time
Eating Out	Frequency
	Walking Time
Leisure Activities	Frequency
	Walking Time

Respondent Location

Different questions about location are asked to the respondents, depending on whether they live in Singapore or not. In the first case, respondents should provide the district number where they reside, as well as postal number. This information is of interest for FCL in the context of a long-term study with a large sample. However, this is not analyzed in this study. For the respondents who do not live in Singapore, only the country of residence is required. Since MTurk survey is answered only by American and Indian respondents, two additional questions about the urban character of their location (e.g. town center, suburbs, country side) as well as municipality size are asked. Aim of this is to identify, if some MTurk respondents might live in urban conditions close to Singaporean ones.

Socio-demographic Information

Various questions about socio-demographic characteristics of the respondents are asked. These provide information allowing a comparison of respondents' characteristics with Singaporean population ones. The content of the socio-demographic questions is summarized in Table 7. In the convenience survey, an additional question about survey origin (e.g. from friends, social media, etc.) is included. Concerning the MTurk survey, some changes occur to make the survey compatible with American and Indian populations. Singaporean question group is removed, personal income is displayed in US dollars and a self-rating of English proficiency is required rather than the main spoken language in household. As for the questions about respondent location, only a subset of the information is used to assess population representativeness. Additional answers might be used in a more detailed study of FCL.

Table 7 Resume of the main socio-demographic questions

Question Group	Topic
General Information	gender date of birth
Singaporean*	residential status ethnicity nationality travel frequency to CBD
Education	highest education level achieved
Job	work situation personal income
Mobility	transport mode to get to work/school transport mode for leisure purpose
Household	car ownership in household number of people in household language spoken in household

*Only for people living in Singapore

3.2.3 Consistency Tests

Among stated preference experiments, an additional question is included in order to answer consistency. Two different consistency tests are used. The first type is a dominant alternative, where one of both route is obviously more suitable than the other one (e.g. time really shorter and route type more convenient). The choice set with dominant alternative is presented in Annex A 1. The second type of consistency test consists in displaying two times the same choice sets, once at the beginning and once at the end of the questionnaire. However, the alternative which was displayed before on the right side is on the left side in the redundant choice set and vice versa.

3.2.4 Experimental Variations

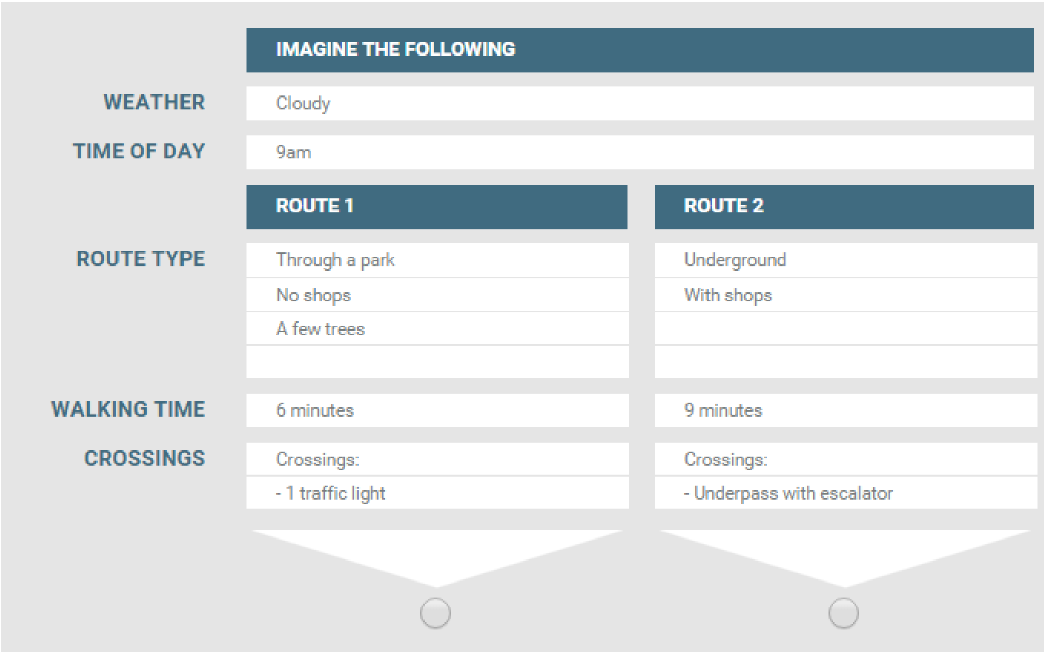
To investigate whether a different presentation of the stated preference experiments influences answer quality and reliability, two further stated preference designs are experimented in addition to the standard layout of Figure 5. These variations are tested in the MTurk survey only, since this platform ensures a large number of respondents at a moderate cost (see section 4.2). First variation tested is a slight modification of the original design where the description of environmental route characteristics (e.g. shops or trees) is removed (Figure 6).

This approach investigates if an image alone is enough for the respondents to perceive route characteristics.

Figure 6 Stated preference variation without description route characteristics (image only)

The second variation is a complete new design without any pictorial representation. The route attributes are represented in table form with different sections (Figure 7). In this case, it is experimented if respondents can understand the present stated preference experiment without pictorial support.

Figure 7	Stated preference variation in text form (text only)



3.2.5 Summary

Considering the different samples considered as well as the implemented design variations, five different survey types are conducted (Table 8). Commercial and convenience surveys use both the standard version of the questionnaire only, while the MTurk survey tests the three types of experimental variations.

Table 8	The different survey types of the study

	Design		
Sample	Text and Image	Text only	Image only
Commercial	X		
MTurk	X	X	X
Convenience	X		

3.3 Discrete Choice Modeling

Using data the different sampling frames (see section 3.2.5), MNL models are computed using BIOGEME, a software specialized in the estimation of discrete choice models (Bierlaire 2003). First, simple models are built with MTurk data in order to investigate the potential effects of the survey design type on model outcome. Interaction terms based on survey type or are tested for the different model attributes. Equation 7 presents an example for shops' presence interaction. Furthermore, scaling factors related to survey design groups are implemented. Hence, this approach enables the calculation of different error-terms distributions depending on respondents' groups. The significance of these factors is then analyzed. Finally, some simple choice models are computed using only a certain survey design type only and results are compared. Since MTurk counts American and Indian respondents, a similar approach is conducted for investigating a potential influence of the country of origin on model results.

$$\beta_{shops} \cdot (1 + \beta_{text,shops} \cdot type_{text} + \beta_{image,shops} \cdot type_{image}) \cdot shops \qquad (7)$$

β_{shops}: coefficient estimate of shops
$\beta_{text,shops}$: interaction coefficient between shops and survey design in text only
$\beta_{image,shops}$: interaction coefficient between sops and the survey design with image only
$type_{text}$: variable of survey design text only
$type_{image}$: variable of survey design image only
shops: variable of shops

In a second step, extended choice models accounting for time interaction of environment variables and route characteristics are built using one single sample source. Effect of gender and age of the respondents are also investigated. Models' outcomes are then analyzed to identify which sample shows the most reliable choice behavior. In addition to goodness of fit such as adjusted R^2, time interaction coefficients and attributes ratios are compared. Time interaction attributes route attributes interacted with walking time. An example considering the presence of shops is presented in equation 8. This measure is interpreted as reduction/increase of perceived walking time perception due to the presence of shops.

$$\beta_{time} \cdot (1 + \beta_{t,shops} \cdot shops) \cdot time \qquad (8)$$

β_{time}: coefficient estimate of walking time
time: walking time variable
$\beta_{t,shops}$: interaction coefficient between walking time and shops
shops: variable of shops

Attribute ratios are used interpret the effect of crossing infrastructures and traffic lights presence. Crossings' coefficient estimates are divided by the walking time coefficient estimate.

This measure can be interpreted as a willingness to walk for avoiding this kind of crossing infrastructure. Equation 9 shows an example for the presence of an overpass with stairs. It should be noted that this kind of valuation is also applied to environmental variables or route characteristics if the walking time interaction approach cannot be computed. To assess choice reliability of the different samples, time interactions attributes and attribute ratios of are compared with the values of a choice model computed with high quality data from a former field survey of FCL. The data collection of this study included a first contact with respondents on the streets and then the completion of the stated preference survey on the web. Hence, this process was really demanding compared to the online approach investigated in this work.

$$WTW_{overpass,stairs} = \frac{\beta_{overpass,stairs}}{\beta_{time}} \quad (9)$$

β_{time}: coefficient of walking time
$\beta_{overpass,stairs}$: coefficient estimate of the presence of one traffic light
$WTA_{overpass,stairs}$: willingness to avoid one traffic light [min]

4 Evaluation of Online Survey Services

Section 4.1 is dedicated to evaluation results of online platforms. The evaluation outcome of the online samples is described in section 4.2. Finally, section 4.3 presents the commercial platform and sample which have been elected for the further surveying campaign.

4.1 Evaluation of Online Platforms

Following subsections present the results of the evaluation defined in section 3.1.1. A summary of these findings is provided in Table 9.

Media Support

All the services of this category are available for in any platform tested.

User Interface's Layout

Only FluidSurveys is not able to display an institution logo in the footer or header.

Advanced Features

Time tracking, branching logic, block and sequence randomization settings can be found in any software analyzed. However, the level of complexity supported differs among platforms. For instance, Qualtrics allows a time tracking at a page level while Survey Monkey can only measures answering time at a survey level. Also, Survey Monkey just supports simple branching. Then, conjoint analysis question type is only available for Qualtrics and Survey Analytics (also with pictures). In contrast to the open-source software LimeSurvey, none of the commercial platform allows researcher to use an external experimental plan. Finally, QuestionPro is the only one platform which provides an incentive question type. Indeed, respondents can be rewarded with online vouchers of various commercial companies on this platform.

Survey Distribution and Accessibility

Most of the settings of this group can be found in any online platforms tested. However, LimeSurvey does not offer a respondent panel in contrast to commercial platforms. All services provide a captcha question type, except Survey Monkey. A question type designed for recursive invitation of respondents can only be found in QuestionPro, Survey Analytics and LimeSurvey.

Platform Cost

Subscription fees differ greatly among commercial platforms. QuestionPro offers the lowest price: the subscription fees amount to $99 per month. On the other hand, Survey Analytics charges 1'000 $ per month. As opposed to commercial tools, it should be noted that the availability of LimeSurvey's features might depend on programming skills of the researcher.

Table 9 Evaluation of survey platforms

	Qualtrics	**QuestionPro**	**FluidSurveys**	**Survey Analytics**	**Survey Monkey**	**LimeSurvey**
Edition considered	Upgraded	Corporate	Ultra	Conjoint Analysis	Platinium	-
Media Support						
-Image/Graphic	Available	Available	Available	Available	Available	Available
-Video	Available	Available	Available	Available	Available	Available
-Combination Flexibility	Available	Available	Available	Available	Available	Available
User Interface's Layout						
-Progression Bar	Available	Available	Available	Available	Available	Available
-Footer/Header Editing	Available	Available	Missing	Available	Available	Available
Advanced Features						
-Conjoint Analysis (CA)	Available	Missing	Missing	Available	Missing	Available
-Design Flexibility of CA	Missing	Missing	Missing	Available	Missing	Available
-External Experiment Plan for CA	Missing	Missing	Missing	Missing	Missing	Available
-Time Tracking	*Page Level	*Survey Level	*Question Level	*Survey Level	*Survey Level	*Group Level
-Time Countdown	*Page Level	*Survey Level	*Question Level	Missing	Missing	*Group Level
-Sequence Randomization	Available	Available	Available	Available	Available	Available
-Block Randomization	Available	Available	Available	Available	Available	Available
-Branching	*Complex Logic	*Complex Logic	*Complex Logic	*Complex Logic	*Simple Logic	Available
-Incentive	Missing	Available	Missing	Missing	Missing	Available
Distribution and Accessibility						
-Respondent Pool	Available	Available	Available	Available	Available	Missing
-Respondent Tracking	Available	Available	Available	Available	Available	Available
-Quota Setting	Available	Available	Available	Available	Available	Available
-Captcha Question	Available	Available	Available	Available	Missing	Available
-Ballot Box Stuffing Prevention	Available	Available	Available	Available	Available	Available
-Recursive invitation	Missing	Available	Missing	Available	Missing	Available
Platform Cost						
-Cost of the Platform Use	- / $2,500**	$99 / $899	- / $840	$1,000 / -	- / $1,200	- /0 $

Legend: Available (green) / Missing (pink) / *Comment

**Indication from AboutAnalytics (2015)

Other sources: direct contact with companies

4.2 Evaluation of online samples

Table 10 presents the results of the online samples' evaluation defined in section 3.1.2. All the commercial survey platforms analyzed in section 4.1 as well as MTurk and the specialized panel provider SSI are evaluated. As expected, asking for respondents' contact details is not permitted by any panel provider. All companies state that they carefully control the quality of their respondents, respectively workers. In particular, automated responses by means of web-robots should not be of concern. The evaluation results are based on these declarations. However, this fact cannot be verified by the author. Five of the analyzed providers have an own panel. In contrast, QuestionPro uses a mix between own and external respondents. The panel source of FluidSurveys is completely outsourced. All the commercial panel providers have samples which cover more or less the entire world. This does not mean that every country of the world is represented, but every continent or region has some representatives enabling global studies. In contrast, MTurk might reasonably provide enough respondents only for the USA and India as stated in section 3.2.1. These findings are consistent with main literature statements (Ipeirotis 2013; Mason and Suri 2011). Finally, costs of the panels from commercial survey platforms are all in the same range of 4.50 to 5.00 $ per answer. In contrast, the average price for an MTurk answer is almost ten times cheaper (0.55 $).

Table 10 Evaluation of online samples

	Qualtrics	QuestionPro	FluidSurveys	Survey Analytics	Survey Monkey	MTurk	SSI
Private Contact of Respondent							
Prevention of Artificial Answering							
Panel Source	Own	Mixed	Extern	Own	Own	Own	Own
Geographic Coverage	Global	Global	Global	Global	Global	USA/India	Global
Respondent Cost	$5.00*	$5.00	$4.50	$4.50	$5.00	$0.55**	$4.50

Legend: Available | Missing

*Indication from AboutAnalytics (2015)

**Indication from Horton et al. (2011)

Other sources: direct contact with companies

4.3 Selection of the Commercial Online Services

Considering the results of section 4.1, it is clear that conducting stated preference experiments using an external experimental plan is not realizable with any of the analyzed commercial survey platforms. Nevertheless, another strategy using fix survey images and block randomization logic might be used. In order to build an efficient survey, complex branching logic is required to allow respondent to skip some socio-demographic questions. Based on these conditions, any of the commercial platforms except Survey Monkey could be elected. Also, all the commercial panels show quite similar characteristics. Hence, the final choice of the commercial online service relies on the cost of the platform use. Since this study only requires two months of platform use, QuestionPro is definitely the cheapest option. Indeed, it is with Survey Analytics the only one platform where a monthly abonnement is provided and subscription fees of QuestionPro are from far the cheapest with 99 $ a month (Table 9). The logic would be to use QuestionPro as well as a panel provider in order to reduce administration efforts. However, panel costs for a Singaporean sample diverges a lot from the price of a general sample for this company (9.99 $ per answer for 200 respondents).[5] In contrast, SSI offers a Singaporean sample for 7.50 $ the answer.[6] Thus, the commercial survey is implemented on the QuestionPro platform and answers are collected from a SSI panel. Since surveys of the study are identified by their sample name, the commercial survey is called SSI survey in the following sections.

[5] Email conversation with QuestionPro company, 29-04-2015.

[6] Email conversation with SSI company, 29-04-2015.

5 Descriptive Statistics

Section 5.1 analyses the survey process, especially in terms of collection time, completion time and rate. Data quality of the different samples is assessed in section 5.2. Population's representativeness is described in section 5.3.

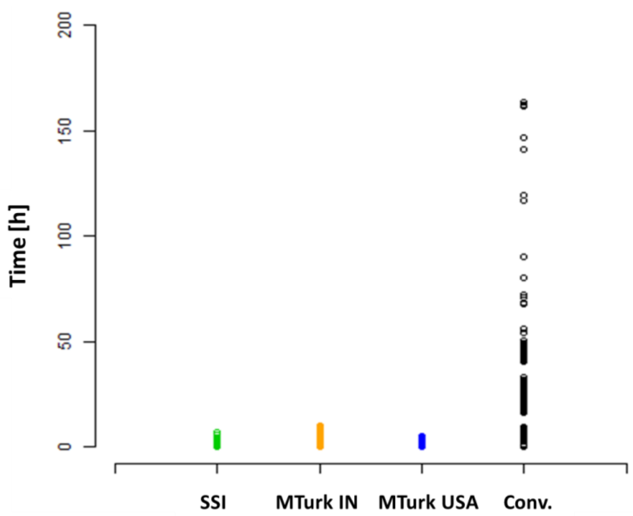
5.1 Completion Characteristics

Table 11 presents the total number of answered surveys collected during the online survey campaign, the cost per answer, the total and median collection time. MTurk data is separated into an Indian and an American group since the survey was posted into two distinct tasks: one for each country of origin. The analysis focuses on the median time, as we expected the collection time to resemble a skewed distribution. It can be observed that both MTurk surveys and the SSI survey have a median collection time of a few hours (1.5 to 3.1 hours). In the case of the convenience survey, it took approximately 8 times as long to collect a smaller sample, namely 23 hours. Figure 8 highlights this duration difference in a graphical way. These findings are consistent with the literature stating that rewards accelerate response time (Mason and Watts 2009). However, the collection approach is very different in the case of the convenience sampling compared to the other surveys. Indeed, the researcher relies on the willingness of seed respondents to forward the questionnaire link to friend and acquaintances (see section 3.2.1). In the case of MTurk, the results show a smaller collection time duration for American workers in comparison to Indian workers. This might indicate a larger number of American respondents than Indian respondents, which is consistent with the literature and makes sense considering the access restriction to non-American works since 2012 (Ipeirotis 2013; Mason and Suri 2011). The cost per answer differs per survey type. The SSI sample costs 7.50 $ per answer while the convenience sample does not cost anything. MTurk on other hand, costs 0.55 $ per completed survey. Respondent characteristics are very diverse. The SSI sample is a one hundred percent Singaporean sample. In contrast, the nationality of the respondents in the convenience sample cannot be controlled ex-ante because of the viral distribution approach. As expected, no Singaporean respondent completed the questionnaire on MTurk.

Table 11 Number of answers, cost and collection time by sample type

Sample	**Answers** [-]	**Cost** [USD/answer]	**Collection Time (median)** [h]	**Collection Time (total)** [h]
SSI	194	7.50	1.5	7.2
Convenience	159	0.00	22.9	163.7
MTurk India	200	0.55	3.1	9.9
MTurk USA	200	0.55	1.2	5.0

Figure 8 Collection time sequence by survey type



The SSI and both MTurk samples have a completion rate of 91 % (Table 12). More people drop out of the convenience survey (completion rate of 49 %). Before to launch the surveys, an ex-ante completion rate has been estimated (see section 2.4.3 and Axhausen et al. 2015). In the time estimation no distinction is made between Indian and US respondents for the MTurk survey, since this identification requires the completion of the nationality question. Models accounting for incentive and/or recruitment are chosen. Estimates are then compared to the effective completion rates, which correspond to the proportion of respondents that achieved completely the questionnaire. Even if the applied model was designed for mail surveys, a quite good match can be observed between the estimated and effective completion rates of the SSI and MTurk surveys (6 to 7 % difference). In contrast, the completion of the convenience survey is not well estimated by the two regression models tested. Model with recruitment overestimates the completion rate of 35 % while the prevision without recruitment is 38 % to low. Detailed calculation and more results of the estimation of the ex-ante response rate are presented in Annex A 2.

Table 12 Comparison of the ex-ante and effective completion rate depending of sample and regression type

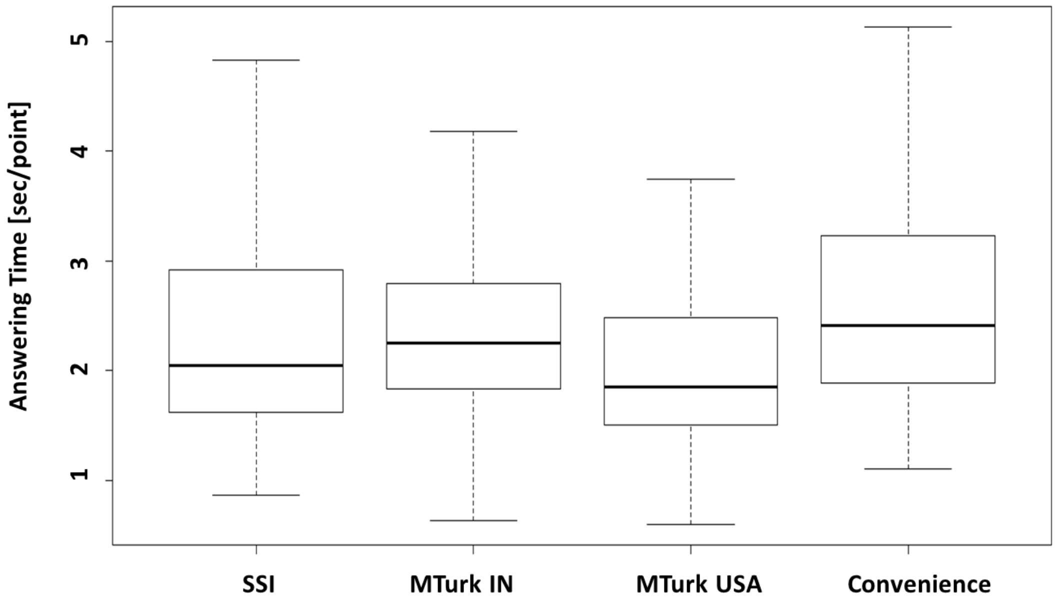
Sample	**Regression Type**	**Completion$_{estimated}$**	**Completion$_{real}$**	**Difference***
SSI	Recruiting + Incentive	85%	91%	-7%
MTurk	Recruiting + Incentive	85%	91%	-6%
Convenience	Recruiting Only	66%	49%	35%
Convenience	No Recruiting + No Incentive	30%	49%	-38%

*Completion$_{real}$ as reference

Regression model's source: Schmid and Axhausen (2015)

The surveys differ slightly in terms of length, especially because the MTurk questionnaire contains less socio-demographic questions. Hence, a direct comparison of the survey completion times is not possible. Using the results of the response burden models of Axhausen et al. (Axhausen et al. 2015), the answering time per burden point is then calculated. Figure 9 shows that the respondents from the convenience sample require the longest answering time per burden point with a median value of 2.4 seconds, followed by the Indian MTurk respondents with a median of 2.3 seconds. Finally, the SSI and the American MTurk samples have the lowest median answering times per burden point: less than 2 seconds. However, the distribution of the SSI answers is larger than for MTurk USA.

Figure 9 Completion time per burden point for the whole survey



In LimeSurvey platform, the answering time can be tracked a group level. In contrast, QuestionPro only reports the total answering time. Hence, Figure 10 and Figure 11 display the answering time per burden point in the SP experiments, respectively in the other survey questions and text introductions for the convenience and MTurk samples only. It can be observed that both pictures are quite consistent with results of Figure 9. However, the median answering time per burden point of the convenience sample is shorter than for the Indian MTurk sample when considering the socio-demographic questions and text information (2.8 versus 3.2 seconds). Furthermore, the median answering time per burden point of the SP experiments is about 50 % smaller than for the other questions or text introductions. It might indicated that the burden of the SP answers have been overestimated by the model, respectively the burden of the other survey parts underestimated.

Figure 10 Completion time per burden point for the stated preference questions

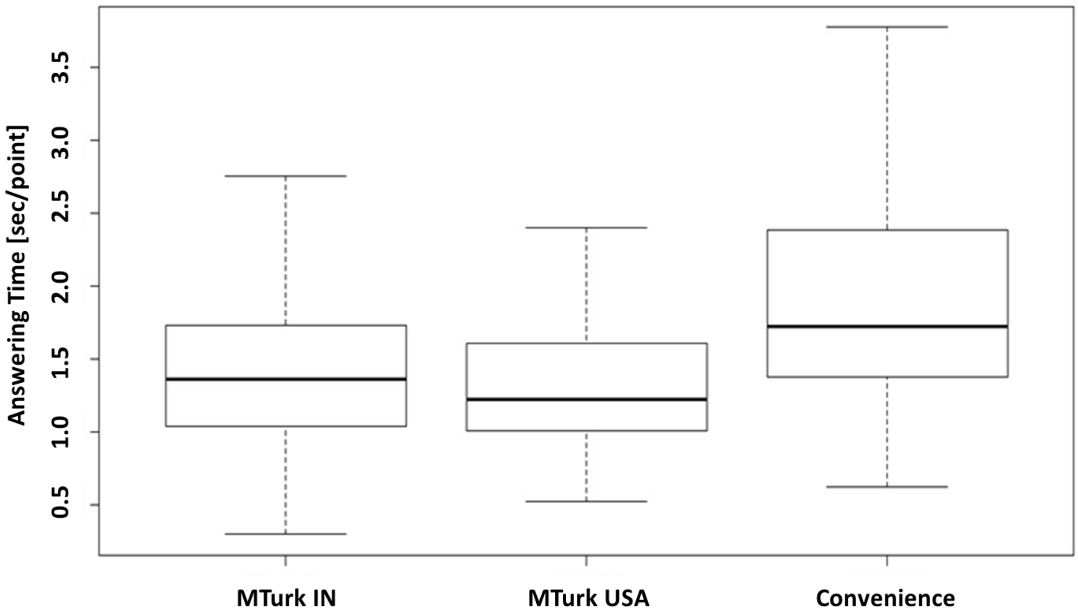
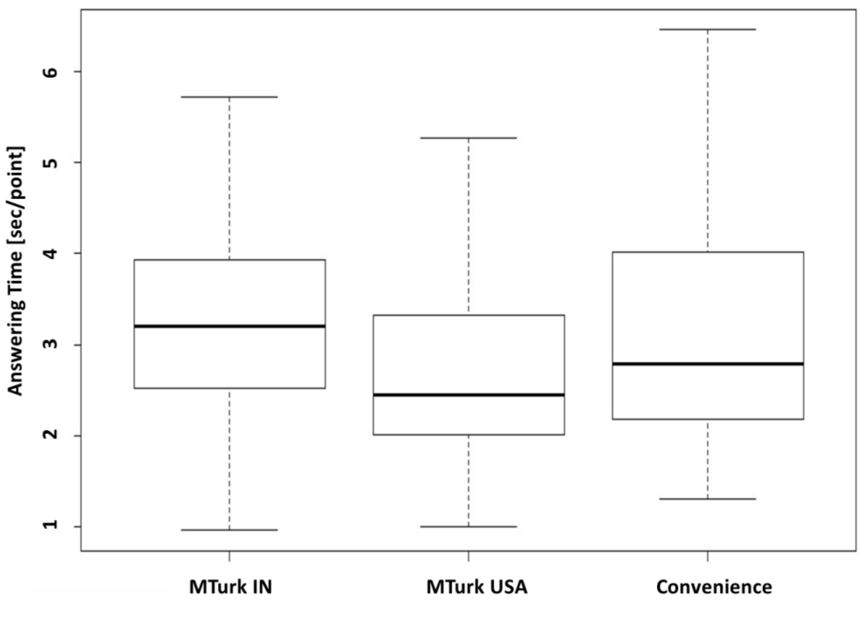
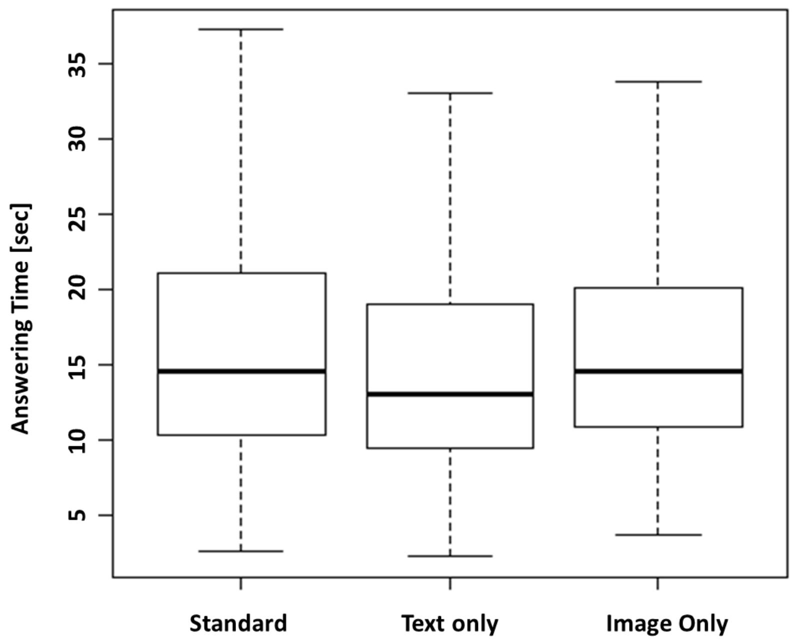


Figure 11 Completion time per burden point for the personal questions and text introductions



Finally, the completion time needed to answer the stated preference questions depending on design form is also investigated (Figure 12). Since it is assumed that the SP questions of all surveys have the same response burden, the answering time can be directly interpreted. In general, the three types of designs display quite similar median completion time ranging from 13 (text version) to 14.5 seconds for (standard and image only version). Low answering time of the text version might be somehow surprising, as this design seems to be a priori less intuitive to understand than the ones with images. This effect is not related to the respondent origin, since 49.7 % of respondents who answered the text version where Indians. It might indicate that the people facing a survey with images spend more time to analyze the pictures. Furthermore, the survey with text design might bore the respondents. Thus, they answer the question more quickly, maybe with responses based on an attributes' subset only.

Figure 12 Completion time per stated preference question by design type



5.2 Data Quality

Data quality is tested in several ways. First, the results of the dominance and redundant questions described in section 3.2.3 are used to check the answer consistency of the different samples. Lexicographic behavior related to walking time is also investigated. Walking time is indeed the only one characteristic, which is present in every choice set. A non-trading test is also conducted. It investigates if some respondents systematically choose the first answer up of the personal questions, respectively the right or left answer of the stated preference experiments. However, this last test did not identify any non-trading respondent so that the non-trading analysis is not further described in the following subsections.

5.2.1 Consistency

Table 13 presents the proportion of respondents who failed to the redundancy and dominance test. First, it can be observed that a higher failure rate occurred for the redundancy question compared to the dominance question in all samples. The Indian MTurk sample shows a high proportion of non-consistent respondents of 36 %. The American sample of Amazon has almost 50 % less inconsistent answers. Hence, the quality of Indian and American MTurk workers is not equal with regard to data consistency. The convenience sample is the most consistent with less than 13 % failed answers. Finally, SSI sample has a significant number of inconsistent answers of about 25 %

Table 13 Proportion of inconsistent respondents based on redundancy and dominance test for the different samples

Sample	**Redundancy** [% error]	**Dominance** [% error]	**Total** [% error]
SSI	35.3	11.5	25.2
MTurk India	39.6	30.4	35.7
MTurk USA	20.8	15.2	18.6
Convenience	17.6	3.5	12.6

Considering the survey design variation, there is some indication that the standard version with both images and text is related to a slight higher percentage of inconsistent answers. However, the difference is not very large (31 versus 24 and 25 % respectively).

Table 14 Total proportion of inconsistent MTurk respondents depending on the survey design type

Design Type	**Inconsistency** [%]
Standard	31.1
Text only	23.7
Image only	25.2

5.2.2 Lexicographic Behavior

Lexicographic behavior with regard to walking time has been also investigated. Respondents are thought to be lexicographic on walking time if they systematically choose the quickest route. It can be observed that the respondents of SSI sample are by far the most lexicographic with a proportion of almost 7.8 % (Table 15). On other hand, the convenience sample is almost free of lexicographic behavior (0.6 %). These observations are consistent with the findings of the consistency checks. Because of the small extent of the lexicographic behavior in relation to walking time, no socio-demographic influences are investigated.

Table 15 Proportion of walking time lexicographic respondents per sample type

Sample	**Lexicographic** [%]
SSI	7.8
MTurk India	1.1
MTurk USA	1.0
Convenience	0.6

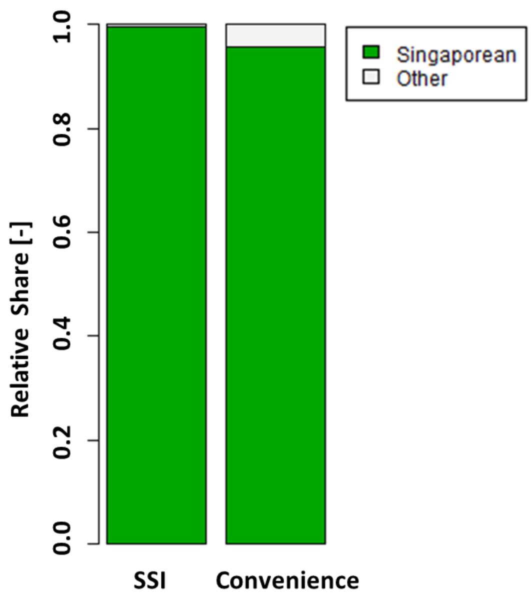
5.3 Population's representativeness

This section presents some personal characteristics of the people surveyed. Section 5.3.1 is dedicated to the respondents' location. Socio-demographics characteristics are introduced in section 5.3.2. Finally, walking habits of the different samples are detailed in section 5.3.3.

5.3.1 Respondent Location

In SSI and convenience samples, a question asked the survey takers if they live in Singapore. Results to this question are displayed in Figure 13. It can be noticed that a very number of respondents of both samples were located outside Singapore. In the case of the SSI sample, only 2 respondents reported to live outside of Singapore. Then, the convenience sample has a proportion of about 5 % of people who do not live in Singapore. This is surprisingly low considering the viral distribution approach of this sample.

Figure 13 Proportion of Singaporean inhabitants in SSI and convenience samples



In the case of the MTurk survey, the urban characteristics of the Indian and American respondents were investigated. The urbanity type and size of the Indian MTurk group are simultaneously displayed in Figure 14. It can be observed that the majority of these respondents live either in city centers of suburbs (85 %). In contrast, the population is distributed among municipality sizes ranging from village to metropoles. Less than 20 % of the inhabitants of city centers and suburbs live in municipalities of one million inhabitants or more. Hence, there are about 13 % only of the Indian MTurk respondents who live in an urban environment, which could approach Singaporean conditions. Considering American respondents (Figure 15), a great part of the surveyed population lives either in towns of middle size or greater cities. 80 % of them reside in municipalities of 10’000 inhabitants or more. However, only 16 % of the American MTurk respondents live in cities of one million inhabitants or more. Here again, a very small portion of the surveyed people live in similar urban conditions as Singapore's inhabitants.

Figure 14 Cross proportions of the urbanity type size [inhabitants] of the locations of Indian MTurk respondents

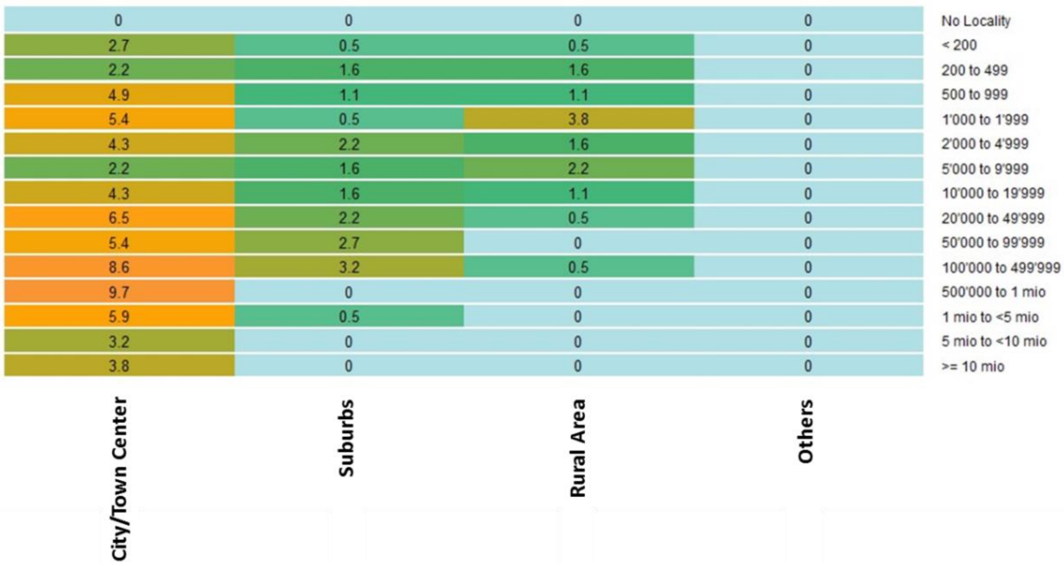
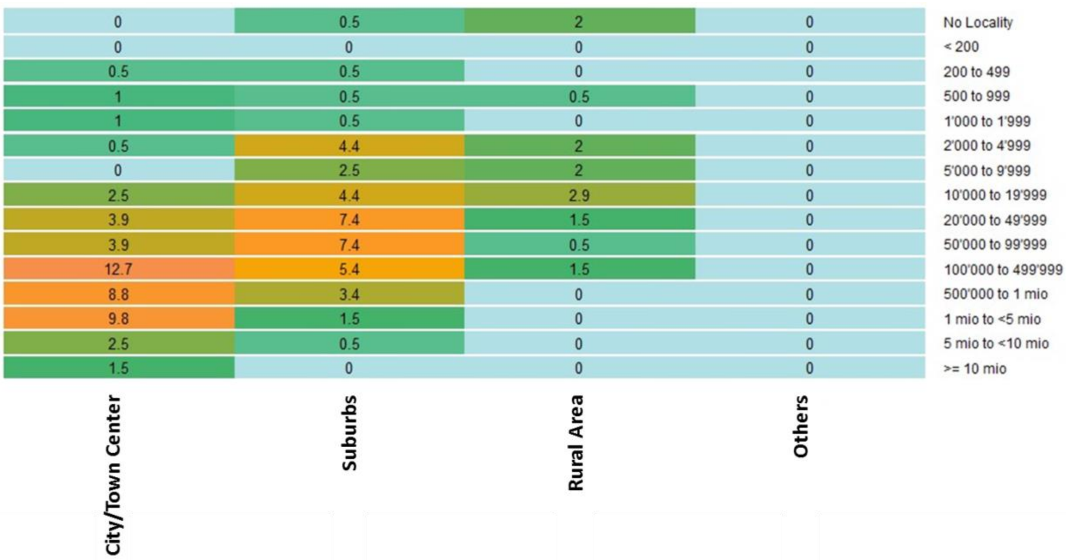


Figure 15 Cross proportions of the urbanity type and size [inhabitants] of the locations of American MTurk respondents

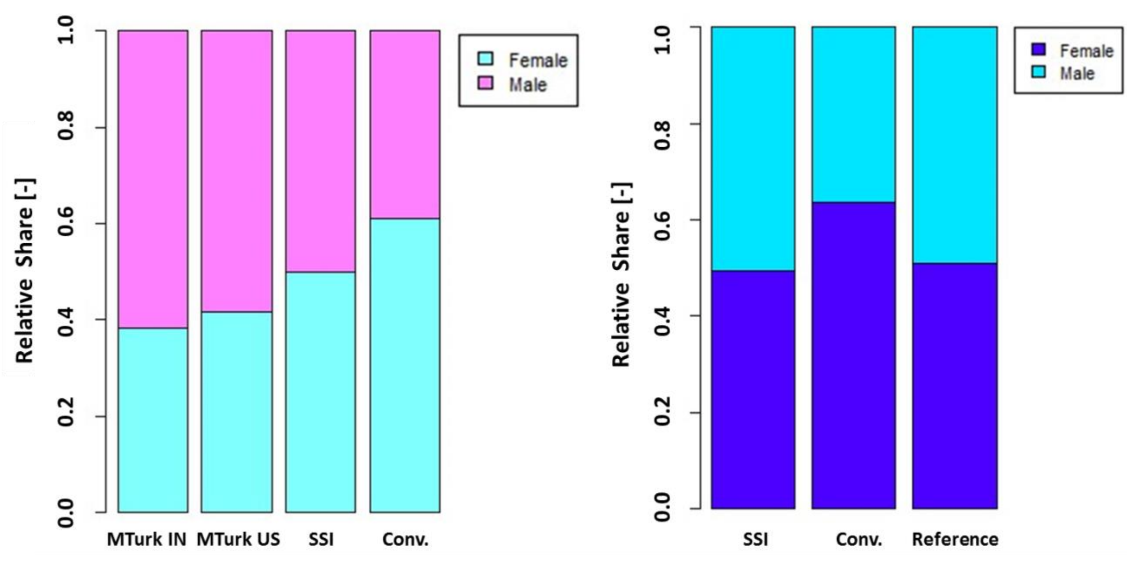


5.3.2 Socio-Demographic Characteristics

This section introduces some socio-demographic characteristics of the different samples. When Singaporean statistics are available, values of the Singaporean respondents of the SSI and convenience samples are investigated compared to it. However, it should be noted that Singaporean statistics only take into account Singaporean citizens and permanent residents. This means that about one 1 million and 700’000 inhabitants of Singapore are not surveyed (Singapore Department of Statistics 2014). This might limit the comparison reliability.

Gender distributions of the different samples are displayed in Figure 16. The SSI sample shows an equal proportion in men and women. The convenience sample is slightly more female (about 60 % female). In contrast, both MTurk samples show a higher proportion of male respondents (about 60 % male). This finding differs from the main literature results stating that MTurk has a slight higher percentage of female workers (Berinsky et al. 2012; Gosling et al. 2004).

Figure 16 Gender statistics of the whole population per sample type (left) and of the Singaporean population with reference value (right)

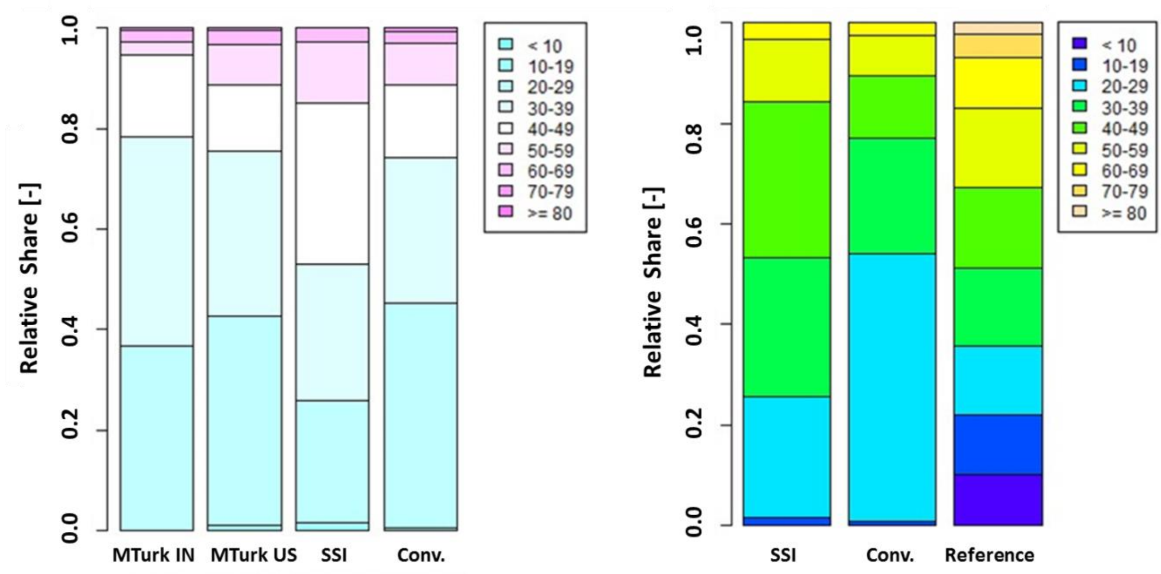


Reference source: Singapore Department of Statistics (2014)

The age statistics shows that almost all the respondents are between 20 and 50 years old (Figure 17). For both MTurk samples and the convenience sample, about 80 % of the surveyed people are in the categories 20 to 29 and 30 to 39 years old in about equal proportions.

In contrast, SSI sample is related with a higher percentage of older respondents (35 % between 49 and 50 years old). Considering the Singaporean respondents only, both the SSI and convenience samples lack of young (less than 20 years old) and elderly people (70 years old and more) in comparison to the reference value. Especially, the convenience sample overrepresents the age category 20 to 29 years old. In contrast, the SSI sample shows a more balanced distribution.

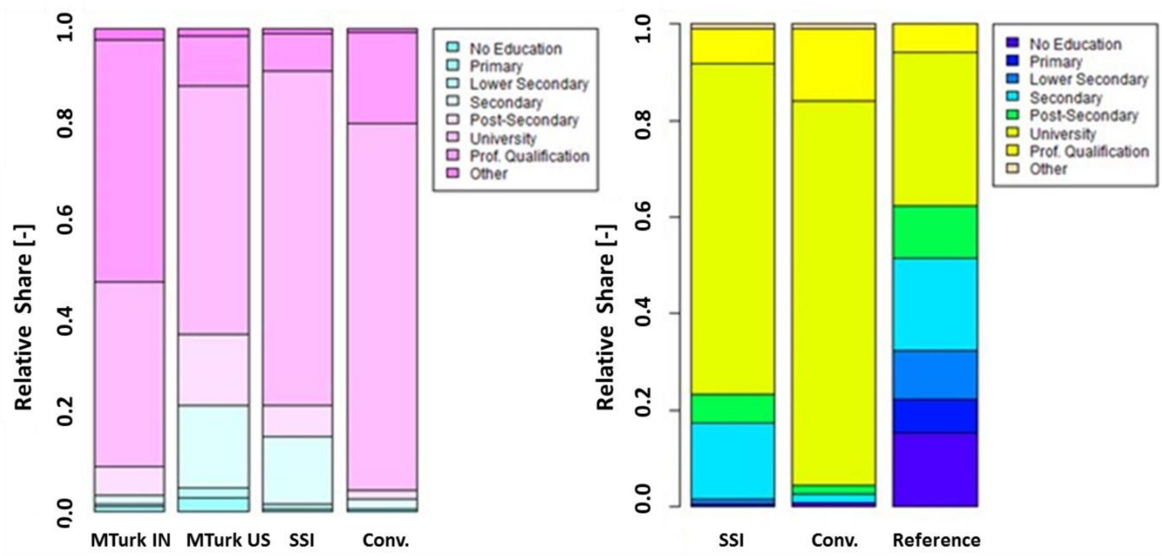
Figure 17	Age statistics [years old] of the whole population per sample type (left) and of the Singaporean population with reference value (right)



Reference source: Singapore Department of Statistics (2014)

Both the convenience and SSI samples are related to a high percentage of respondents who received a university qualification , e.g. 80 and 70 % of the population (Figure 18). Compared to the Singaporean reference, these proportions are drammatically too high (less than 30 % university qualification). All other education type are about absent of these two samples. MTurk samples present as well a quite low divesity of educational achievements. MTurk USA shows also a high representation of respondents having a university certificate (50 %). In contrast, a great part of Indians achieved a professional qualification (50 %). The proportion of university qualified respondents of is group is about 30 %.

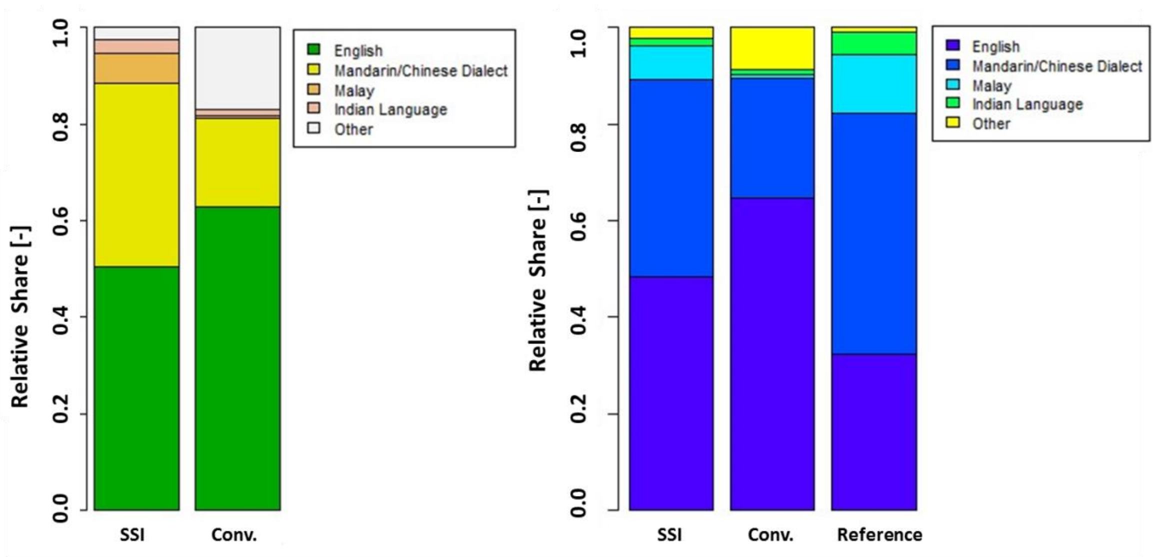
Figure 18 Statistics of the highest education level achieved of the whole population per sample type (left) and of the Singaporean population with reference value (right)



Reference source: Singapore Department of Statistics (2010)

Figure 19 reports which language the respondent mostly speaks at home. The results of SSI sample has quite similar distribution to the reference value, excepting a slight overrepresentation of English compare to Chinese (45 % of English rather than 35 %). In contrast, English has a mucher higer weight in the convenience sample (about 65 %). This is not surprising, since a main respondents' source this survey was researchers from the Future Cities Laboratory, who are in majority academic foreigners evolving in a global athmosphere.

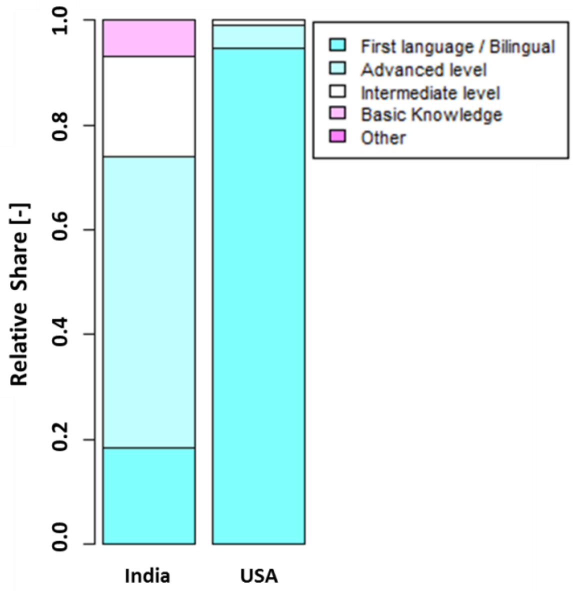
Figure 19 Proportions of the language most spoken at home of the whole population per sample type (left) and of the Singaporean population with reference value (right)



Reference source: Singapore Department of Statistics (2014)

In MTurk survey, the respondents were asked to rate their English level (Figure 20). Of course, almost all the American respondents state to be native speakers. In contrast, only 20 % of the Indian respondents make this statement. About 55 % of them rate their English as advanced and the 25 % remaining as lower. Further socio-demographic statistics can be consulted in Annex A 2.

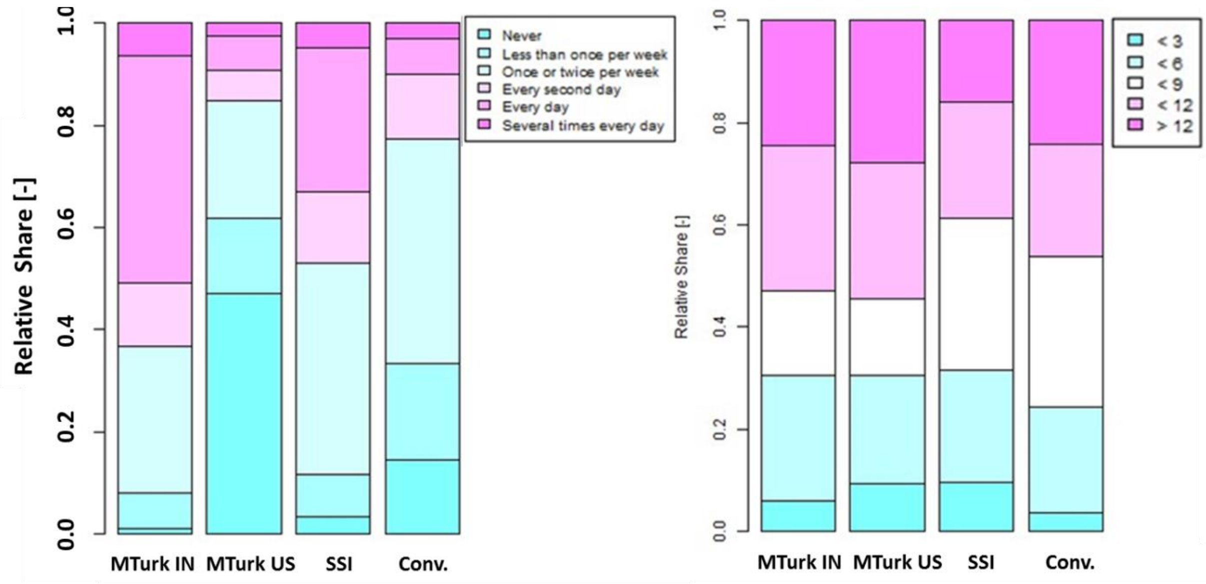
Figure 20 Self-rated English level for the Indian and American MTurk samples



5.3.3 Walking Behavior

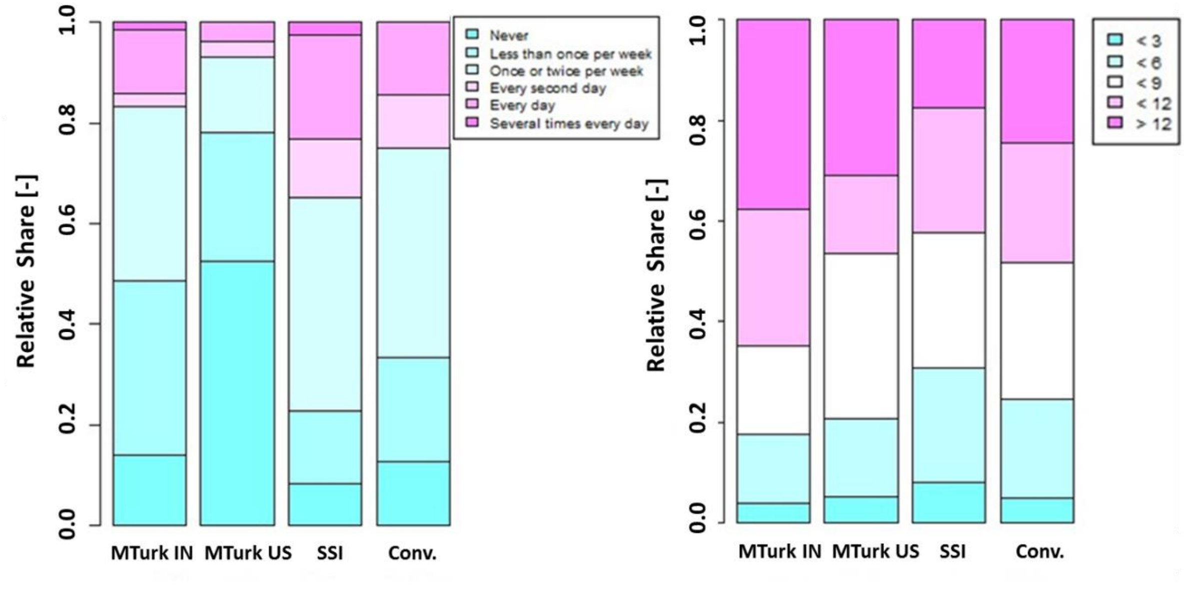
The questionnaire requested some information about the respondent walking habits. Figure 21 displays the distribution of the walking frequency and time of the survey takers to buy groceries. It is not surprising that American respondents of MTurk show a quite low walking tendency to buy out since the United States are a car oriented country. Indian respondents of MTurk show the highest walking tendency with very few people who never walk to buy groceries and a large proportion (40 %) of respondents doing it every day. 40 % of the SSI respondents only walk once a week to buy groceries. The convenience sample shows close values to SSI sample ones, but with a higher proportion of people walking less than once week or never to buy groceries. Considering walking time, the distribution is quite similar among the different samples. A very few proportion of people has to walk less than three minutes to buy groceries. The other time categories are quite balanced.

Figure 21 Distribution of the walking frequency (left) and time [min] (right) to buy groceries per sample type



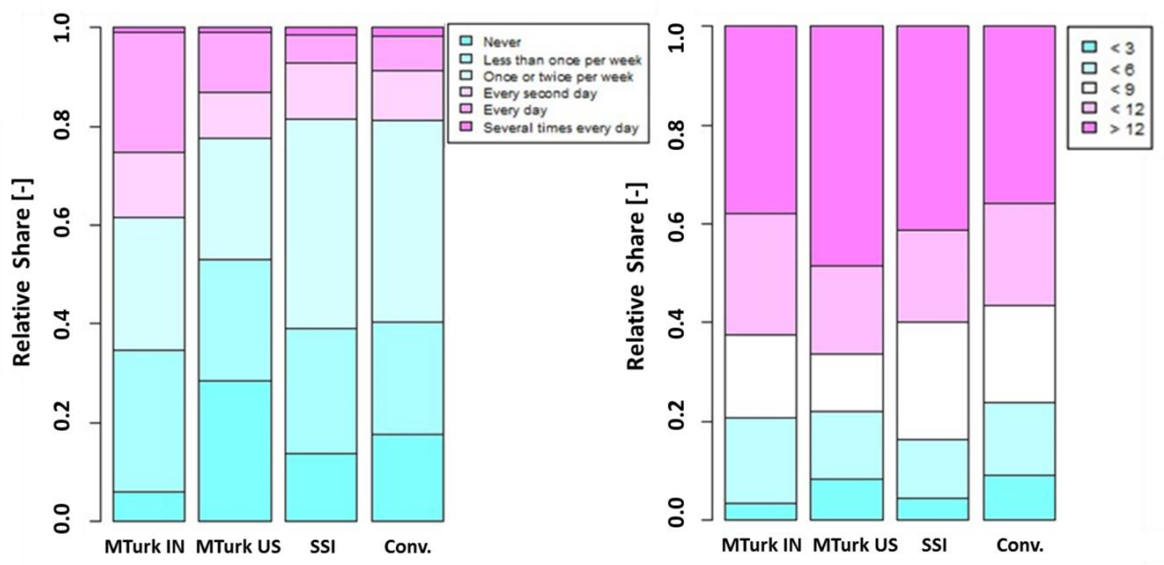
The feature of the walking statistics to eat out is different (Figure 22). In this case, the SSI and convenience sample have very similar walking proportions. They both show the highest walking frequency with 70 to 75 % of the respondents who walk to eat out at least once or twice a week. Based on the author experience, this finding seems to be consistent as it is common in Singapore to eat out at food courts located in the neighborhood. Again, US respondents of MTurk show a very high percentage of people who never walk to eat out. In this case, it might also be related to different cultural habits and not only to a high level of car usage. Considering the walking time, it can be noticed that about 80 % of the SSI and convenience walker have trip times of less than 12 minutes to eat out. This is consistent with the previous results of the walking frequency and the food court culture hypothesis. Both Indian and American MTurk samples have higher proportions of people who need to walk more than 12 minutes to a restaurant (40 and 35 %).

Figure 22 Distribution of the walking frequency (left) and time [min] (right) to eat out per sample type



Finally, Figure 23 displays the walking statistics to leisure activities. As in the previous analysis, the SSI and convenience samples show very similar distributions. About 65 % of these respondents walk once or twice per week to leisure activities and about 10 % never walk for this purpose. Indian MTurk sample has close results, but with fewer people who never walk (5 %). Again, American respondents of MTurk display the highest proportion of people who never walk (30 %). The walking times are quite the same for all the samples. According to these results, it is very common to walk more than 12 min to reach leisure activities (about 40 % for all samples). It seems logical that people might be more willing to walk far when being in a relaxed leisure context. In this case, American MTurk has the population who tends to walk the most. In addition to the statistics presented in this section, heat maps displaying cross-proportions of frequency and walking time the different samples are provided for the different samples in Annex A 2.

Figure 23 Distribution of the walking frequency (left) and time [min] (right) to leisure activities per sample type



6 Model Results

This chapter presents the results of the different discrete choice models computed. It should be noted that inconsistent data (see section 5.2.1) were removed before the estimation process. By contrast, lexicographic data have not been removed. They represent a very small share of the population and it is not clear if they arise from natural or apparent lexicographic behavior. All the estimated models accounts for panel effects, since a respondent provides several observations. During the model estimation process, it came out that the crossing type underground with stairs and underground with escalator were closely related to the road type inside building. Indeed, the underground crossing types only appeared in the experimental design with this road type together. This misspecification leads to a high correlation between the underground crossings and the road type inside building. Hence, the underground crossing types have been removed of the models in order to evaluate correctly the route type effects. In addition to the model results presented in the following sections, utility functions and tested insignificant variables are described in Annex A 3. Models' specification is introduced in section 6.1. Results of the survey design analysis are presented in section 6.2. Differences amongst Indian and American respondents are reported in section 6.3. Finally, extended choice models with and without walking time interaction are presented in sections 6.4 and 6.5 respectively.

6.1 Model specification

In order to estimate the models, one attribute of each category has to be excluded. Coefficient estimates are then interpreted in relation to this reference category. These reference attributes are introduced in Table 16. Also, the specification of all relevant variables included in the choice models is provided in Table 17.

Table 16 Reference Attributes per category

Attribute Category	Reference
Crossings	No Crossing
Day Time	9:00 AM
Country of Origin	USA*
Route Type	Minor Road
Survey Type	Standard*
Traffic Light	No Traffic Light
Weather	Cloudy

*MTurk only

Table 17 Specification of the model variables

Variable	Definition
asc_left	Left alternative specific constant
b_1trafficlight	One traffic light
b_2trafficlights	Two traffic lights
b_cover_sidewalk_time	Interaction of covered sidewalk with walking time
b_cover_sidewalk_female	Interaction of b_cover_sidewalk if female
b_cover_sidewalk_time_rain	Interaction of b_cover_sidewalk_time if rainy
b_cover_sidewalk	Covered sidewalk
b_cover_sidewalk_rain	Interaction of b_cover_sidewalk if rainy
b_cover_sidewalk_image	Interaction of b_cover_sidewalk if survey type is image
b_cover_sidewalk_sun	Interaction of b_cover_sidewalk if sunny
b_greenery	Greenery presence for any route type
b_inside	Route type inside building
b_inside_india	Interaction of b_inside if respondent origin is India
b_inside_rain	Interaction of b_inside if rainy
b_inside_sun	Interaction of b_inside if sunny
b_insideRain	Route type inside building when rainy weather
b_inside_time	Interaction of road type inside building with walking time
b_inside_time_rain	Interaction of b_inside_time if rainy
b_jaywalk_2lanes_age50	Interaction of b_jaywalk_2lanes if age greather or equal than 50 years
b_jaywalk_2lanes_india	Interaction of b_jaywalk_2lanes if respondent origin is India
b_jaywalk_2lanes	Jaywalking of 2 lanes
b_jaywalk_4lanes_india	Interaction of b_jaywalk_4lanes if respondent origin is India
b_jaywalk_4lanes	Jaywalking of 4 lanes
b_major	Route type major road
b_major_time	Interaction of road type major road with walking time
b_overpass_lift	Overpass with lift
b_overpass_stairs	Overpass with stairs
b_overpass_stairs_age50	Interaction of b_overpass_stairs if age greather or equal than 50 years
b_overpass_stairs_female	Interaction of b_overpass_stairs if female
b_park_1pm	Interaction of b_park of daytime is 1 pm
b_park_rain	Interaction of b_park if rainy
b_park_sun	Interaction of b_park if sunny
b_park	Route type park
b_park_time	Interaction of road type major road with walking time
b_shops	Shopping amenities
b_shops_1pm	Interactionof b_shop if daytime is 1 pm
b_shops_time	Interaction of shopping amenities with walking time
b_time_text	Interaction of b_time if survey type is text
b_time	Walking time
b_2trafficlights_text	Interaction of b_2trafficlights if survey type is text
b_tree_road	Tree presence on road
b_tree_road_time	Interaction of trees on the road with walking time

6.2 Survey Design Analysis

Table 18 presents the results of a basic model accounting for survey design variations. It should be noted that overpass variables are not included in the model since they both are in-significant when modeling with MTurk data. According to this analysis, the survey design with image only does not show fundamental differences compared to the standard version of the questionnaire in terms of experiment perception. In contrast, walking time and the variable of two traffic lights interact significantly with the text design term. Respondents facing text design weight the importance of the walking time half less than other surveyed people. In the case of the two traffic lights variable, this becomes insignificant if the survey is in text form. Considering the computed scaling factors, it can be observed that both text and image version are not significantly related with different error-term distributions than the standard version (Table 19).

Table 18 Results of basic MTurk data model with survey design type interactions

Variable	Value	SE robust	p-value robust
asc_left	-0.02	0.045	0.68
b_1trafficlight	-0.33	0.079	0.00
b_2trafficlights	-0.30	0.087	0.00
b_2trafficlights_text	-1.00	0.236	0.00
b_cover_sidewalk	0.25	0.094	0.01
b_greenery	0.17	0.074	0.02
b_inside	0.15	0.125	0.22
b_jaywalk_2lanes	-0.37	0.152	0.01
b_jaywalk_4lanes	-0.77	0.218	0.00
b_major	0.06	0.067	0.35
b_park	0.94	0.202	0.00
b_shops	0.11	0.048	0.02
b_time	-0.12	0.027	0.00
b_time_text	-0.47	0.223	0.03
$N_{observations}$	**2424**		
R^2 ajdusted	**0.044**		

Table 19 Scale factors per survey design type of the MTurk sample

Scale	Value	SE robust	p-value robust
Standard	1.00	-	-
Only Text	1.14	0.277	0.61
Only Image	1.10	0.29	0.74

To better investigate the effect of survey design differences, the basic model of Table 18 is computed again after removing the data from the text survey. Aim of this step is also to identify, if data collected with the text design have to be included or not to estimate an extended MTurk model with environmental interactipns. According to the results of Table 20, a model computed with all the collected MTurk data seems more promising. Indeed, more variables are significant in the model estimated with the whole MTurk dataset. Except for the variable of two traffic lights, all coefficient estimates are of similar signs and amplitudes in both models. Hence, all the MTurk data are used to estimate an extended model. However, interaction terms of the text design version are included to enable a better interpretation of the results.

Table 20 Results of basic MTurk data models with all data and without the text design data

	All Data			Without Text Design		
Variable	**Value**	**SE robust**	**p-value robust**	**Value**	**SE robust**	**p-value robust**
asc_left	-0.03	0.045	0.48	-0.04	0.058	0.49
b_1trafficlight	-0.33	0.079	0.00	-0.38	0.100	0.00
b_2trafficlights	-0.30	0.087	0.00	-0.48	0.114	0.00
b_cover_sidewalk	0.27	0.073	0.00	0.29	0.093	0.00
b_greenery	0.19	0.073	0.01	0.13	0.097	0.17
b_inside	0.21	0.101	0.04	0.21	0.129	0.10
b_jaywalk_2lanes	-0.35	0.140	0.01	-0.39	0.180	0.03
b_jaywalk_4lanes	-0.80	0.173	0.00	-0.82	0.217	0.00
b_major	0.08	0.070	0.25	0.06	0.088	0.49
b_park	1.03	0.126	0.00	0.97	0.162	0.00
b_shops	0.12	0.046	0.01	0.05	0.057	0.39
b_time	-0.10	0.017	0.00	-0.12	0.022	0.00
	$N_{observations}$	**2424**		**$N_{observations}$**	**1436**	
	R^2 ajdusted	**0.045**		**R^2 ajdusted**	**0.043**	

6.3 Country of Origin

As for the survey design type (see section 6.2), effects of the country of origin of MTurk respondents on model outcome is investigated using a simple choice model (Table 21). It can be observed that Indian respondents perceive jaywalking differently than American respondents. Indeed, the burden of jaywalking two traffic lanes is reduced of more than 90 % when being Indian and crossing four traffic lanes is even related with a slight utility compared to no crossing. This basically means that jaywalking is insignificant for Indian MTurk respondents. The scale factors do not show significant differences concerning the error-term distribution of both populations (Table 22).

Table 21 Results of basic MTurk data model with respondent origin interactions

Variable	Value	SE robust	p-value robust
asc_left	-0.03	0.045	0.49
b_1trafficlight	-0.35	0.083	0.00
b_2trafficlights	-0.32	0.088	0.00
b_cover_sidewalk	0.29	0.073	0.00
b_greenery	0.20	0.079	0.01
b_inside	0.21	0.104	0.05
b_jaywalk_2lanes	-0.60	0.199	0.00
b_jaywalk_4lanes	-1.52	0.237	0.00
b_jaywalk_2lanes_india	-0.92	0.316	0.00
b_jaywalk_4lanes_india	-1.09	0.174	0.00
b_major	0.08	0.072	0.27
b_park	1.06	0.149	0.00
b_shops	0.12	0.050	0.01
b_time	-0.11	0.020	0.00
$N_{observations}$	**2424**		
R^2 ajdusted	**0.053**		

Table 22 Scale factors per country of origin of the respondents of MTurk sample

Scale	Value	SE robust	p-value robust
USA	1.00	-	-
India	0.97	0.197	0.89

To complement the investigation, simple models have been estimated using American, respectively Indian MTurk data only. Models' outcomes are presented in Table 23 with the results of a model estimated with all MTurk data as well. According to these results, the model computed with American answers only shows the best goodness (adjusted R^2 of 0.064). Indian data model shows the poorest fitting quality with an adjusted R^2-value of 0.033. Also, lots of variables of this model are insignificant. The model including all respondent data shows a fitting quality ranging between the two other models (adjusted R^2 of 0.045). It has the highest number of significant variables. Since the aim of the extended is to test the effect of route characteristics on walking perception, all the MTurk data are hence used to compute the extended choice model. Furthermore, using more data might enable to test more socio-demographic effects. Interaction terms related to the respondent's country of origin are not included in the extended MTurk. This inclusion would make the comparison of interacted attributes with values of other models impossible. Indeed, the American reference category is not present in the SSI and convenience samples.

Table 23 Results of the basic MTurk data models with all data, US data only and Indian data only

	All Data			Data USA			Data India		
Variable	**Value**	**SE rob.**	**p-value rob.**	**Value**	**SE rob.**	**p-value rob.**	**Value**	**SE rob.**	**p-value rob.**
asc_left	-0.03	0.045	0.48	0.02	0.059	0.71	-0.10	0.067	0.12
b_1trafficlight	-0.33	0.079	0.00	-0.32	0.111	0.00	-0.37	0.113	0.00
b_2trafficlights	-0.30	0.087	0.00	-0.25	0.114	0.03	-0.43	0.136	0.00
b_cover_sidewalk	0.27	0.073	0.00	0.19	0.102	0.06	0.40	0.106	0.00
b_greenery	0.19	0.073	0.01	0.26	0.099	0.01	0.10	0.115	0.38
b_inside	0.21	0.101	0.04	0.20	0.143	0.16	0.20	0.142	0.16
b_jaywalk_2lanes	-0.35	0.140	0.01	-0.55	0.206	0.01	-0.11	0.188	0.57
b_jaywalk_4lanes	-0.80	0.173	0.00	-1.52	0.243	0.00	0.15	0.266	0.58
b_major	0.08	0.070	0.25	0.08	0.091	0.38	0.08	0.115	0.50
b_park	1.03	0.126	0.00	1.03	0.174	0.00	1.09	0.182	0.00
b_shops	0.12	0.046	0.01	0.17	0.062	0.01	0.05	0.070	0.47
b_time	-0.10	0.017	0.00	-0.11	0.023	0.00	-0.09	0.027	0.00
	$N_{observations}$		**2424**	**$N_{observations}$**		**1368**	**$N_{observations}$**		**1056**
	R^2 ajdusted		**0.045**	**R^2 ajdusted**		**0.064**	**R^2 ajdusted**		**0.033**

6.4 Models with Walking Time Interaction

During the estimation process extended models, it came out that significant time interaction terms could be only computed with SSI data. Indeed, all the tested walking time interactions where insignificant for the convenience and MTurk samples. The results of the SSI data model are presented in Table 28. To discuss the outcome of this model, data from a previous survey campaign conducted by FCL researchers are used to estimate the model again. The same utility function as for the SSI model is considered. The results of the FCL model can be consulted in Table 25.

Table 24 Results of the SSI model with walking time interaction

Variable	**Value**	**SE robust**	**p-value robust**
asc_left	0.19	0.066	0.00
b_1trafficlight	-0.45	0.139	0.00
b_2trafficlights	-0.99	0.154	0.00
b_cover_sidewalk_time	-0.36	0.049	0.00
b_cover_sidewalk_time_rain	1.29	0.373	0.00
b_inside_time	-0.21	0.061	0.00
b_inside_time_rain	3.28	1.310	0.01
b_jaywalk_2lanes	-0.41	0.236	0.08
b_jaywalk_4lanes	-2.08	0.235	0.00
b_major_time	0.06	0.044	0.14
b_overpass_lift	-0.65	0.248	0.01
b_overpass_stairs	-1.65	0.308	0.00
b_overpass_stairs_age50	-0.85	0.283	0.00
b_overpass_stairs_female	-0.47	0.163	0.00
b_park_time	-0.31	0.056	0.00
b_shops_time	-0.11	0.037	0.00
b_time	-0.34	0.034	0.00
b_tree_road_time	-0.12	0.040	0.00
$N_{observations}$	**1255**		
R^2 ajdusted	**0.198**		

Table 25 Results of the FCL study's model with walking time interaction

Variable	**Value**	**SE robust**	**p-value robust**
asc_left	0.03	0.047	0.51
b_1trafficlight	-0.04	0.094	0.67
b_2trafficlights	-0.28	0.074	0.00
b_cover_sidewalk_time	-0.29	0.041	0.00
b_cover_sidewalk_time_rain	1.81	0.418	0.00
b_inside_time	-0.34	0.059	0.00
b_inside_time_rain	1.87	0.582	0.00
b_jaywalk_2lanes	-0.43	0.162	0.01
b_jaywalk_4lanes	-1.12	0.150	0.00
b_major_time	0.09	0.034	0.01
b_overpass_lift	-0.48	0.153	0.00
b_overpass_stairs	-0.80	0.171	0.00
b_overpass_stairs_age50	0.51	0.602	0.40
b_overpass_stairs_female	0.11	0.293	0.70
b_park_time	-0.43	0.051	0.00
b_shops_time	-0.17	0.030	0.00
b_time	-0.31	0.021	0.00
b_tree_road_time	-0.22	0.030	0.00
$N_{observations}$	**1255**		
R^2 ajdusted	**0.124**		

6.5 Models with Direct Effects

Since the computing of models with walking time interactions was not successful for the MTurk and convenience samples, choice models with direct effects of route characteristics are estimated. In order to compare attribute ratios amongst all samples, a model without walking time interaction is also computed using SSI data. The results of these models are presented in Table 26, Table 27 and Table 28. It should be noted that text design interaction terms are included in the MTurk model, since this design type influences significantly the walking time and two traffic lights variables (see section 6.2).

Table 26 Results of the SSI model with direct effects of the route characteristics

Variable	Value	SE robust	p-value robust
asc_left	0.14	0.068	0.04
b_1trafficlight	-0.43	0.139	0.00
b_2trafficlights	-0.75	0.153	0.00
b_cover_sidewalk	0.88	0.170	0.00
b_cover_sidewalk_rain	2.05	0.683	0.00
b_cover_sidewalk_sun	0.96	0.441	0.03
b_inside	0.73	0.202	0.00
b_inside_rain	2.29	0.911	0.01
b_jaywalk_2lanes	-0.40	0.272	0.14
b_jaywalk_4lanes	-2.08	0.238	0.00
b_major	-0.23	0.099	0.02
b_overpass_lift	-0.62	0.262	0.02
b_overpass_stairs	-1.62	0.277	0.00
b_overpass_stairs_age50	-0.80	0.275	0.00
b_overpass_stairs_female	-0.44	0.162	0.01
b_park	1.20	0.243	0.00
b_park_1pm	-0.74	0.250	0.00
b_shops	0.52	0.123	0.00
b_shops_1pm	-0.70	0.319	0.03
b_time	-0.30	0.028	0.00
b_tree_road	0.27	0.106	0.01
$N_{observations}$	**1255**		
R^2 ajdusted	**0.207**		

Table 27 Results of the convenience model with direct effects of the route characteristics

Variable	**Value**	**SE robust**	**p-value robust**
asc_left	-0.12	0.079	0.15
b_1trafficlight	-0.50	0.133	0.00
b_2trafficlights	-0.72	0.134	0.00
b_cover_sidewalk	1.01	0.281	0.00
b_cover_sidewalk_female	-0.35	0.141	0.01
b_cover_sidewalk_sun	-0.85	0.246	0.00
b_insideRain	0.72	0.270	0.01
b_jaywalk_2lanes	-0.67	0.235	0.00
b_jaywalk_4lanes	-0.80	0.251	0.00
b_major	0.03	0.113	0.76
b_overpass_stairs	-1.80	0.423	0.00
b_overpass_stairs_age50	-1.59	0.735	0.03
b_overpass_stairs_female	-0.68	0.172	0.00
b_park	2.47	0.439	0.00
b_park_rain	-0.83	0.119	0.00
b_park_sun	-0.61	0.124	0.00
b_shops	0.32	0.078	0.00
b_time	-0.07	0.027	0.01
$N_{observations}$	**1112**		
R^2 ajdusted	**0.081**		

Table 28 Results of the MTurk model with direct effects of the route characteristics

Variable	Value	SE robust	p-value robust
asc_left	-0.05	0.043	0.28
b_1trafficlight	-0.39	0.088	0.00
b_2trafficlights	-0.54	0.139	0.00
b_2trafficlights_text	-0.88	0.202	0.00
b_cover_sidewalk	0.36	0.127	0.00
b_cover_sidewalk_sun	-0.89	0.344	0.01
b_inside	0.27	0.118	0.02
b_inside_sun	-2.02	0.707	0.00
b_jaywalk_2lanes	-0.54	0.165	0.00
b_jaywalk_2lanes_age50	3.08	1.390	0.03
b_jaywalk_4lanes	-0.73	0.214	0.00
b_major	0.10	0.065	0.12
b_park	1.66	0.337	0.00
b_park_rain	-0.76	0.110	0.00
b_park_sun	-0.50	0.128	0.00
b_time	-0.10	0.025	0.00
b_time_text	-0.59	0.250	0.02
$N_{observations}$	**2424**		
R^2 ajdusted	**0.054**		

7 Discussion

Section 7.1 is dedicated to the experience gained by using online survey platforms. Outcome of the data analysis is discussed in section 7.2. Choice model results are analyzed in section 7.3.

7.1 Web Survey Platforms

To conduct the different surveys of this study, the commercial platform QuestionPro and the open-source service LimeSurvey have been used. Even if both experiences were successful, LimeSurvey provides the researcher with a much higher flexibility than QuestionPro. Indeed, LimeSurvey allows the researcher to customize a survey with very high freedom. LimeSurvey supports for instance the generation of choice sets based on an external experiment plan. It should be noted that this feature was not supported by any commercial platform considered in this study (see section 4.1). In QuestionPro, the choice sets had to be implemented as images. Furthermore, answering time is reported on a group level in LimeSurvey. QuestionPro only reports the total interview time without further details. Based on these findings, QuestionPro does not justify 99 $ of subscription fees. Since the evaluation of section 4.3 elected QuestionPro as the most suitable commercial platform to conduct this stated preference experiments' survey, it is doubtful that other commercial platforms might provide researchers with notable advantages in comparison to open-source tools. This holds especially when dealing with complex questionnaire designs such as stated experiments. However, it should be noted that some programming skills might be required to implement advanced settings in open-source software, e.g. external experiment plan for choice sets' generation.

7.2 Data Analysis

Results of the data analysis are discussed in this section. Section 7.2.1 is dedicated to the completion characteristics. Data quality is discussed in section 7.2.2. Finally, findings about the population's representativeness of the different samples are presented in section 7.2.3.

7.2.1 Completion Characteristics

In terms of collection time, both SSI and MTurk samples have an indisputable advantage compared to the convenience sample, e.g. less than 10 hours versus about one week for collecting less answers. A much higher completion rate is reported for the paid respondents, e.g. 91 % for SSI and MTurk samples versus 49 % for the convenience sample. The risk of non-response bias is hence lower for the SSI and MTurk samples, since less respondents drop out of the survey (Christensen et al. 2014; Heerwegh and Loosveldt 2008). In the present case, a reward effect on the completion rate cannot be assessed since convenience sampling was conducted in a different distribution approach than the two other samples (viral distribution). In addition, the high response rates of SSI and MTurk samples contradict the findings of Heerwegh and Lossveldt (Heerwegh and Loosveldt 2008). Indeed, these authors state that internet surveys are related to a low response rate. It is observed that people of convenience and MTurk India samples needed more time to answer the questionnaire (see Figure 9, section 5.1). In the case of the convenience sample, this might be interpreted as a slow and careful survey completion, since convenience respondents show a high level of answer consistency (see Table 13, section 5.2.1). In contrast, it is possible that some respondents of MTurk India spent more time because of low English proficiency (see Figure 20, section 5.3.2). English level is here considered as low if the respondent did not state to have a bilingual or advanced level. However, results of Table 29 indicate that the majority of the people having a low English proficiency answered more quickly than the median completion time of the population of MTurk India. An explanation might be that many Indian respondents reported a higher level of English than it is in reality. It is also possible that some of these respondents needed more time to answer because they were not used to stated preference experiments. More people of American MTurk and SSI samples had maybe faced this kind of complex questionnaire before and are hence quicker. Naiveté might also concern the convenience sample, whose respondents were the slowest to answer.

Table 29 Proportion of Indian MTurk respondents with smaller, respectively equal or higher survey answering time than population's median answering time depending on English proficiency

English Proficiency	< Median Time	>= Median Time
High	48%	52%
Low	55%	45%

7.2.2 Data Quality

As described in section 5.2.1, the convenience sample shows the highest data consistency level, e.g. less than 13 % error. It is followed by American respondents of MTurk with about 19 % failure rate to consistency check questions. SSI sample is in third position (25 %) and the Indian MTurk respondents are the most inconsistent with 36 % of error. It should be noted that the consistency check with redundant choice sets was much more selective than the dominant alternative test. Considering survey design variations, people who faced the survey in standard form with both images and text description reported more inconsistent answers than others. Nevertheless the difference is not very high: 30 % failure for the standard version compared to 25 % for the version in text form and with images only. Lexicographic behavior on walking time concerns only a small share of respondents. Non-trading behavior cannot be assessed in any samples. Table 30 presents the proportion of inconsistent respondents depending on highest education level achieved. In all samples except for MTurk India, the respondents with university education have a smaller failure rate than other ones. This effect is important for the SSI and convenience samples but slight for MTurk USA. The highly educated respondents of MTurk India show in contrast a slightly higher inconsistency level compared to less educated people of this population (38 versus 34 %). It should be noted that the samples are quite different in terms of education level distribution (see Figure 18, section 5.3.2). Convenience and SSI samples face an important overrepresentation of respondents with university education (80 and 70 %). The presence of highly educated people is more moderate amongst American and Indian MTurk populations (50 and 30 %). Hence, the low educated populations of the convenience and SSI samples are maybe too small to be representative. Considering these findings, an education effect on answer inconsistency cannot be assessed with certainty. If it exists, this effect might be only relative within a same population. Indeed, the highly educated respondents of the SSI sample have the same inconsistency level as the less educated people of the convenience sample.

Table 30 Proportion of inconsistent respondents per sample type depending on the highest education level achieved Change comments

Sample	University	Other
SSI	21.7%	33.3%
MTurk India	38.0%	34.2%
MTurk USA	16.2%	21.2%
Convenience	9.9%	21.1%

Results of Table 31 report the share of inconsistent respondents of the American and Indian MTurk samples depending on their English proficiency. As outlined in section 7.2.1, English proficiency is considered high if the respondent states to have a bilingual or advanced level. In both MTurk samples, respondents with lower English proficiency show a higher failure rate. However, there are only 1 % of surveyed MTurk Americans who report an English level lower than advanced (see Figure 20, section 5.3.2). Hence, the American MTurk sample is not suitable for investigating potential effects of literacy level on data quality. In the case of MTurk India, about one quarter of the respondents have a low English proficiency, which is more representative. The effect of low English proficiency on answer consistency is quite slight for the sample of MTurk India (37 versus 29 %). Hence, no clear evidence of low English proficiency influence on answer consistency can be assessed based on these findings.

Table 31 Proportion of inconsistent respondents of the MTurk sample depending on English proficiency

	English Proficiency	
Sample	**High**	**Low**
MTurk India	29.2%	37.2%
MTurk USA	18.3%	50.0%

The main language spoken at home had to be reported in the SSI and convenience surveys. No relationship can be assessed between people who do not speak English as a first language at home and answer inconsistency (Table 32). Respondent age effects have been also studied, but without any clear findings (Table 33).

Table 32 Proportion of inconsistent respondents in SSI and convenience samples depending on the main language spoken at home

Sample	**English**	**Other**
SSI	25.0%	25.5%
Convenience	12.0%	13.6%

Table 33 Proportion of inconsistent respondents per survey type depending on the age category of the respondent

	Age			
Sample	**< 30 years old**	**30 to 39 years old**	**40 to 49 years old**	**>= 50 years old**
SSI	18.9%	33.9%	28.8%	16.1%
MTurk India	41.2%	29.9%	36.7%	40.0%
MTurk USA	18.1%	16.0%	19.3%	12.1%
Convenience	8.3%	15.2%	30.4%	22.2%

The relationship between answering time and answer quality is also investigated. Table 34 reports the shares of inconsistent answers if the answering time is smaller, respectively greater or equal to the median answering time of the considered population. Since QuestionPro recorded the answering time for the whole survey and not per question group (e.g. stated preference experiments' group), SSI sample is not considered in this analysis. For the Indian respondents of MTurk, a clear relationship between quick answer and inconsistency can be assessed, e.g. about 42 % versus 26 % inconsistent answers. Hence, a certain share of the inconsistent respondents of MTurk India might have been first interested in the reward and answered the survey quickly with low commitment. In contrast, both the American MTurk and convenience samples do not show any clear answering time effect on answer inconsistency.

Table 34 Proportion of inconsistent respondents of MTurk and convenience samples depending on the answering time per burden point of stated preference questions

Sample	< Median Time	>= Median Time
MTurk India	41.8%	25.5%
MTurk USA	19.6%	16.5%
Convenience	11.4%	13.8%

7.2.3 Population's representativeness

The SSI panel provided the only one gender balanced sample (see Figure 16, section 5.3.2). Indeed, convenience sample is slightly more female (60 %) while both MTurk samples have about 60 % of male respondents. All samples underestimated the share of younger and older respondents (see Figure 17, section 5.3.2). Also, ages ranging from 20 to 29 years old are overestimated. However, SSI sample was here again the most representative than the other samples, with a better balanced age distribution the age categories from 20 to 60 years old. Considering the reported walking behavior, the respondents of the SSI and convenience samples show very similar habits for all investigated trip purposes. Especially, the tendency of these two populations to walk a relative small distance quite frequently to eat out is consistent with the food court culture of Singapore (see Figure 22, section 5.3.2). In contrast, both Indian and American MTurk respondents show different walking characteristics. The difference is especially large for the Americans since the USA are a very car oriented country. Indeed, these respondents reported low walking frequencies for all trip purposes tested. Variations in walking characteristics are probably due to differences in the cultural as well as in the urban context of the respondents. In particular, it comes out from the analysis of the urbanity form and size of MTurk respondents' locations that in both Indian and American samples, only 15 % of these populations live in urban conditions which might be close to the Singaporean situations (see Figure 14 and Figure 15, section 5.3.2). Of course, urban specificities and climate conditions of Singapore are not considered in this statement.

7.3 Discrete choice Models

The outcome of the discrete choice models is discussed in this section. Results of the survey design analysis are analyzed in section 7.3.1. Findings about the effects of the country of origin on choice behavior are presented in section 7.3.2. Hypotheses about the sign and magnitude of the coefficient estimates of the extended models are introduced in section 7.3.3. The results the models with walking time interaction, respectively with direct effects, are discussed in section 7.3.4 and 7.3.5.

7.3.1 Survey Design Analysis

The analysis of survey design effects on respondent answers shows that stated preference experiments displayed in text form provide different coefficient estimates for the walking time and the two traffic lights variables compared to the standard version and the version with images only (see section 6.2). Indeed, walking time burden is about 50 % less valuated in text form. The perception of the walking time is important for the experiment understanding, since this variable should be decisive for the pedestrian route choice (Muraleetharan and Hagiwara 2007; Agrawal et al. 2008; Seneviratne and Morrall 1985). It should be noted that in the standard and image only designs, the walking time is displayed with a proportional bar (Figure 24). This image might help the respondent to perceive walking time. The insignificance of the two traffic lights variable when displayed in text form is difficult to explain. Indeed, the one traffic light variable has been similarly rated in all three survey types. Hence, it seems to have no reason to interpret the presence of two traffic lights like as an absence of crossing. Maybe some respondents did not consider this attribute level while answering the text survey. Analysis outcome is consistent with the findings of several authors who report a better perception of attributes presented in pictorial form, rather than solely in text form (Jansen et al. 2009; Rizzi et al. 2012; Vriens et al. 1998).

Figure 24 Display of walking time attributes in the standard and image only design versions

7.3.2 Country of Origin

During the investigation of country of origin influences on MTurk model outcome, it has been observed that Indian respondents are insensible to jaywalking (see section 6.3). This might be related to a different perception of traffic safety of Indian people compared to Americans. No significant differences were observed concerning the scaling factors of the Indian and American samples. Because of its highest goodness of fit (adjusted R^2 of 0.064 versus 0.033) and a highest number of significant variables, the simple choice model computed with American MTurk data is of higher quality than model of MTurk India. Combining these modeling results with the consistency findings (see section 7.2), it can be stated that the collected data from MTurk USA are more valuable than MTurk India ones. However, considering the low goodness of fit of both MTurk models, the general quality of all MTurk data gathered during study seems doubtful.

7.3.3 Hypotheses

The following hypotheses are based on literature findings of section 2.1.1 and on common sense. The Singaporean context described in section 2.1.2 is also considered. Table 35 introduces the hypotheses concerning the sign and amplitude of the extended models' coefficient estimates. It should be noticed that remarks are based on the absolute values of the coefficient estimates. Since the crossing category has no crossing as reference (see Table 16, section 6.1), all crossing types are thought to be of negative value. This holds for the traffic light presence. Since traffic light cycles of Singapore last about 120 seconds (see section 2.1.2), it is assumed that the waiting time of one traffic light is of about 1 minute. Hence, the coefficient estimates of one traffic light and of walking time should have similar amplitudes. Also, two traffic lights should be equal to two times one traffic light. The burden of an overpass equipped with lift should be smaller than the one of an overpass with stairs. Further, jaywalking four traffic lanes is assumed to be the crossing type with the highest burden since it is very dangerous. Indeed, safety tends to be highly valuated by pedestrians (Saelens and Handy 2008). Route type inside building is thought to be more appreciated than minor road since it is free of traffic noise. Furthermore, places inside buildings are often air conditioned in Singapore, which might be appreciable for walking in a tropical climate. Because of higher traffic fluxes, the route type major road should be negatively valuated compared to minor road. By contrast, several authors state that park routes are appreciated by surveyed pedestrians (Adkins et al. 2012; Guo 2009; Rodríguez and Joo 2004). The coefficient estimate of this route type should hence be positive. Covered sidewalk should also be perceived as advantageous since they provide the walker with shade and protection against rain. Trees on sidewalk provide as well some shade and are related with a certain aesthetic value. Finally, the walking time is generally reported in the literature as being the most important attributes when making route decisions and should then be negative (Muraleetharan and Hagiwara 2007; Agrawal et al. 2008; Seneviratne and Morrall 1985).

Table 35 Hypotheses about sign and amplitude relationship of the coefficient estimates

<table>
<tr><th>Variable</th><th>Sign</th><th>Remark</th></tr>
<tr><td>b_1trafficlight</td><td>-</td><td>~ | b_time |</td></tr>
<tr><td>b_2trafficlights</td><td>-</td><td>~ 2 · | b_1trafficlight |</td></tr>
<tr><td>b_cover_sidewalk</td><td>+</td><td></td></tr>
<tr><td>b_inside</td><td>+</td><td></td></tr>
<tr><td>b_jaywalk_2lanes</td><td>-</td><td></td></tr>
<tr><td>b_jaywalk_4lanes</td><td>-</td><td>> 2 · | b_jaywalk_2lanes |
> | b_2trafficlights |
> | b_overpass_stairs |</td></tr>
<tr><td>b_major</td><td>-</td><td></td></tr>
<tr><td>b_overpass_lift</td><td>-</td><td>< | b_overpass_stairs |</td></tr>
<tr><td>b_overpass_stairs</td><td>-</td><td></td></tr>
<tr><td>b_park</td><td>+</td><td></td></tr>
<tr><td>b_shops</td><td>+</td><td></td></tr>
<tr><td>b_time</td><td>-</td><td></td></tr>
<tr><td>b_tree_road</td><td>+</td><td></td></tr>
</table>

Weather conditions and daytime might affect the magnitude of the coefficient estimates (Table 36). It can be assumed that the utility of a covered sidewalk or of routes inside building should increase if the weather is rainy or sunny rather than cloudy. Indeed, these both attribute levels are synonymous with rain protection and shade. In contrast, the utility of the park routes should be diminished, respectively the burden of a major road increased in such weather conditions. Using an overpass with stairs might be more tedious if the weather is sunny, considering the tropical weather of Singapore. The daytime 1:00 PM is thought to be generally hotter than 9:00 AM. Hence, a similar influence as for the sunny weather is expected with exception of the attribute covered sidewalk. Indeed, the 1:00 PM daytime is here interpreted as temperature and not as sun intensity increase. Finally, it is assumed that the presence of shops might be considered as more utile at the end of the day, when people come back from work.

Table 36 Hypotheses about the effect of weather conditions and daytime on the amplitude of the coefficient estimates

Variable	**Sign**	**rainy**	**sunny**	**1:00 PM**	**7:00 PM**
b_1trafficlight	-				
b_2trafficlights	-				
b_cover_sidewalk	+	Inc	Inc		
b_inside	+	Inc	Inc	Inc	
b_jaywalk_2lanes	-				
b_jaywalk_4lanes	-				
b_major	-	Inc	Inc	Inc	
b_overpass_lift	-				
b_overpass_stairs	-		Inc	Inc	
b_park	+	Dim	Dim	Dim	
b_shops	+				Inc
b_time	-				
b_tree_road	+				

Inc: increasing // Dim: diminishing

Gender effects are also investigated (Table 37). According to Humpel et al., some gender differences in walking choices might be observed but they are usually not systematic (Humpel et al. 2004). Nevertheless, it is here assumed that women might be more concerned by walking safety and would hence stronger perceive the burden of jaywalking two or four lanes. If the female respondent tends to use high-heels shoes, she might dislike more walking long stairs such as in overpass without lift equipment than male respondents.

Table 37 Hypotheses about gender effecst on the amplitude of the coefficient estimates

Variable	**Sign**	female
b_1trafficlight	-	
b_2trafficlights	-	
b_cover_sidewalk	+	
b_inside	+	
b_jaywalk_2lanes	-	Inc
b_jaywalk_4lanes	-	Inc
b_major	-	
b_overpass_lift	-	
b_overpass_stairs	-	Inc
b_park	+	
b_shops	+	
b_time	-	
b_tree_road	+	

Inc: increasing / / Dim: diminishing

Finally, hypotheses are stated about the influence of being 50 years old or older (Table 38). Since older people might be more susceptible to hot weather, it can be thought that they appreciate more routes inside building than a general population. It can also be assumed that people of 50 years old or older might walk slower than younger ones and hence face more risks when jaywalking. Considering these potential physical issues, walking stairs of an overpass might be more difficult for them.

Table 38 Hypotheses about age effects on the amplitude of the coefficient estimates

Variable	Sign	age >= 50 years old
b_1trafficlight	-	
b_2trafficlights	-	
b_cover_sidewalk	+	
b_inside	+	Inc
b_jaywalk_2lanes	-	Inc
b_jaywalk_4lanes	-	Inc
b_major	-	
b_overpass_lift	-	
b_overpass_stairs	-	Inc
b_park	+	
b_shops	+	
b_time	-	
b_tree_road	+	

Inc: increasing // Dim: diminishing

7.3.4 Models with Walking Time Interaction

The SSI data provided the only one model with walking time interaction amongst all samples tested. This indicates that most the respondent of SSI understood well the experiment framework since they could perceive route characteristics in terms of walking time increase or reduction, in contrast to the respondents of the convenience and MTurk samples. However, it should be recalled that this model is computed with consistent answers only and that one quarter of the SSI data had to be removed beforehand (see section 7.2.2). The model computed with the SSI data has a much higher goodness of fit than the model estimated with the FCL study's sample (adjusted R^2 of 0.198 versus 0.124). It can be observed that the SSI model has a significant coefficient for the left alternative specific constant (see Table 24, section 6.4). This means, that the left alternative tends to have a certain utility for the respondents. It could be an indication of non-trading behavior. However, this issue has been investigated in the framework of the data analysis and no respondent having this misbehavior could be identified (see section 7.2.2). Table 39 reports the effects of the route characteristics on perceived walking time. Weather interactions might increase or diminish the effect of the route characteristics (Table 40). Except for major road, it can be observed that the estimated interactions are similar in both models. Indeed, SSI respondents do not perceive a difference in the utility of major and minor roads. All estimated signs are consistent with hypotheses of Table 35. Amplitudes are as well very similar in the two models. However, FCL respondents tend to valuate more covered sidewalk than SSI respondents when the weather is rainy (+181 % effect versus +129 %). In contrast, people from SSI sample appreciate more to walk inside buildings when it rains than FCL sample ones (+328 % effect versus +187 %). Finally, it can be observed that the FCL sample shows a higher affinity for trees on sidewalks (- 22 % time perception versus - 12 %). All other potential weather influences assumed in Table 36 could not be assessed during the modeling process.

Table 39 Comparison of the effects of the route characteristics on perceived walking time in SSI and FCL study's models with walking time interaction

Attribute	**SSI**	**FCL Study**
Covered sidewalk	-36%	-29%
Inside building	-21%	-34%
Major road	-	+9%
Park	-31%	-43%
Shops	-11%	-17%
Trees	-12%	-22%

Table 40 Weather interactions in SSI and FCL study's models with walking time interaction

Interaction	**Effect on**	**SSI**	**FCL Study**
Rain	Covered sidewalk	+129%	+181%
	Inside building	+328%	+187%

The attribute overpass with stairs is the only one to show a significant interaction with socio-demographic characteristics in the SSI model (Table 41). Indeed, female respondents valuate about 50 % less the burden of using an overpass with stairs than males. This result is not consistent with the stated hypotheses of Table 36. However, it is possible that women do not rate the disutility of overpass with stairs so high because of the safety aspect of this crossing type. Also, 50 years old and older respondents perceive only 15 % of the burden of an overpass with stairs compared to other people of SSI sample. Despite of physical issues, it might be that older Singaporeans have less repulsion to use theses infrastructures with stairs only since they experienced a certain time of their life without lift infrastructures. It should be noted that these both socio-demographic interactions were insignificant in the case of the FCL model. All other socio-demographic hypotheses of Table 37 and Table 38 are insignificant in both SSI and FCL models.

Table 41 Socio-demographic interactions in SSI and FCL study's models with walking time interaction

Interaction	Effect on	SSI	FCL Study
Age 50 years old or older	Overpass with stairs	-85%	-
Female	Overpass with stairs	-47%	-

Table 42 presents the ratio of crossings' and traffic lights' attributes with walking time attribute. These values can be interpreted as a willingness to avoid to crossing infrastructures in minutes walking. In other terms, the average respondent of the SSI sample prefers to walk 1.2 minutes rather than encountering one traffic light. It should be noted that the following ratios are representative for a global population, in order to enable a comparison. For instance, the ratio of overpass with stairs of the SSI sample has been corrected with the coefficient estimates of age and gender, using the corresponding population's proportions of the FCL sample. For all crossing categories, SSI respondents perceive a higher burden than FCL ones. All the ratios of SSI sample confirm the hypotheses of Table 35. Indeed, one traffic light is perceived like about 1 minute walking time. Two traffic lights are almost equal to two times one traffic light (2.6 minutes versus 2.4 minutes). In contrast, the respondents of FCL sample are not sensible to the presence of one traffic light. Two traffic lights would correspond to 0.9 minutes of walking time for these respondents. This is not consistent with the stated hypotheses. However, they recognize the burden of jaywalking. As assumed in Table 35, the coefficient estimate of jaywalking four lanes is much higher than for a 2 lanes road (3.6 minutes versus 1.4 minute). SSI answers report a significant disutility for jaywalking four lanes only. For both analyzed samples, jaywalking four lanes is the highest perceived crossing burden, which is consistent with the hypothesis of Table 35. Finally, crossing an overpass with stairs is related with a higher disutility than taking an overpass with lift in both samples. While the willingness to avoid an overpass with lift is somehow similar for the two respondent groups, FCL respondents valuate the burden of an overpass with stairs slightly less than SSI respondents (2.9 minutes versus 3.4 minutes).

Table 42 Comparison of the willingness to avoid of the SSI and FCL study's models with walking time interaction

Attribute	SSI [min]	FCL Study [min]
1 trafficlight	1.2	-
2 trafficlights	2.6	0.9
Jaywalk 2 lanes	-	1.4
Jaywalk 4 lanes	5.7	3.6
Overpass with lift	1.8	1.6
Overpass with stairs	3.4	2.9

Considering its high goodness fit and its very consistent attribute valuations, it can be stated that the choice behavior of the SSI respondents seems to be as reliable as the one of the high quality FCL study. Moreover, people from SSI samples perceived the burden of traffic lights infrastructures even better than the ones of the field survey.

7.3.5 Models with Direct Effects

Choice model estimated with the convenience and MTurk dataset are both of very poor fitting quality with adjusted R^2 of 0.081 and 0.052 (see Table 27 and Table 28, section 6.5). In contrast, the SSI sample provides a model with direct effects of quite high goodness of fit (adjusted R^2 of 0.207), as in its walking time interacted version (see Table 26, section 6.5). The left alternative constant is insignificant in both convenience and MTurk samples (see Table 27 and Table 28, section 6.5). In contrast, respondents of SSI seem to have preferred the left alternatives (see Table 26, section 6.5). Nevertheless, this should not be related to non-trading behavior as described in section 7.3.4. In this new version of the SSI model, an interaction between sunny weather and covered sidewalk could be computed (Table 43). Indeed, a covered sidewalk shows a higher utility when the sun shines: 96 % augmentation. This relation was insignificant in the model with walking time interaction (see Table 24, section 6.4), but was assumed to exist because of shade benefits (Table 36). Daytime interaction terms have been also estimated in the SSI model. Both the route type park and the shops' presence variables lose 70 % of their utility when the daytime is 1:00 PM. In the first case, this could be due to hotter weather which makes outside walking less attractive at noon. This effect is consistent with the stated hypotheses (Table 36). Also, respondents of the SSI sample have maybe no time to shop during lunch. It should be noted that daytime influences could only be assessed in this model. In contrast to the time interacted model of SSI, the route major road is significant in the version with direct effects (see Table 26, section 6.5). Indeed, it shows a disutility compared to minor road, as predicted in Table 35. All the other attributes of the present SSI model are very similar in terms of signs and ratio values to the version with walking time interaction (see section 7.3.4). Hence, these attribute values are not further discussed but used as references for interpreting the results of convenience and MTurk models. Several meteorological interactions could be computed for the convenience sample (Table 43). First, the respondents perceive 85 % less utility of a covered sidewalk when the weather is sunny. This finding is contrary to the hypothesis of Table 35 and is difficult to explain. In contrast, no significant effect of rainy weather on coverage utility could be computed. However, the route type inside building is only significant and perceived as utile if the weather is rainy (see Table 27, section 6.5). According to the hypotheses of Table 36, this route type should be more appreciated than a minor road in any weather conditions as it should be more comfortable to walk (air conditioning in Singapore). In contrast, the utility diminution of a park route when the weather is sunny (- 61 % utility) or rainy (- 83 %) is consistent with the stated hypotheses (Table 35). The MTurk model shows similar weather influences as the convenience model (Table 43). A utility reduction of the sidewalk coverage when it is sunny is as well computed (- 89 % effect). Since MTurk respondents do not live in Singapore, they might not be aware of the local climatic conditions. Nevertheless, half of the respondents of this sample are Indian who might also face hot and humid weather. Although the route type inside building is significant in the MTurk model, it shows a disutility if the weather is sunny (- 202 % of former utility). This is in contradiction to the benefits that an inside place offers in hot weather (Table 36). Such perception might be related to the ignorance of the Singaporean context, the air conditioning of buildings for instance. Finally, park routes are also less appreciated in sunny (- 50 % utility) and in rainy (- 76 %) weather conditions. It can be noted that these two values are really close to the ones computed in the convenience model.

Table 43 Weather and day time interactions in SSI, convenience and MTurk models with direct effects

Interaction	Effect on	SSI	Convenience	MTurk
Rain	Covered sidewalk	+205%	-	-
	Inside building	+229%	-	-
	Park	-	-83%	-76%
Sun	Covered sidewalk	+96%	-85%	-89%
	Inside building	-	-	-202%
	Park	-	-61%	-50%
1:00 PM	Park	-74%	-	-
	Shops	-70%	-	-

The convenience model shows similar gender and age effects on the disutility of overpasses with stairs as the SSI model (Table 44). A hypothesis about the higher affinity of female respondents for this crossing type is stated in section 7.3.4. However, 50 years old and older respondents of the convenience sample perceive these infrastructures as more advantageous than a route without crossing. This finding cannot be explained by the author. Also, female respondents seem to appreciate covered sidewalk in smaller extent than the male population. Here again, the author identifies no clear reason for this behavior. In the case of the MTurk model, the effect of being 50 years old or older on the jaywalking of two lanes is the only one socio-demographic interaction that could be computed. Indeed, this age category perceives a four times higher burden than the other ones (+ 308 % of burden). This effect is consistent with the hypothesis of Table 38.

Table 44 Socio-demographic interactions in SSI, convenience and MTurk models with direct effects

Interaction	Effect on	SSI	Convenience	MTurk
Age 50 years old or older	Overpass with stairs	-80%	-159%	-
	Jaywalk 2 lanes	-	-	+308%
Female	Overpass with stairs	-44%	-68%	-
	Covered sidewalk	-	-35%	-

Table 45 presents a comparison of the attribute ratios of the SSI and of the convenience models. Like for the models with walking time interactions, the ratios have been corrected with weather or socio-demographic interactions and considering the respective population’s share. As for the crossing burdens in section 6.4, the ratios are interpreted as willingness to avoid something in minutes of walking time. If a ratio is negative, then the respondents would walk more time to encounter this attribute during their trip. It can be observed that all the ratios of the convenience sample are consistent in terms of sign with the ones of the SSI sample. However, they all have greater magnitudes. The presence of one traffic light is perceived like 6.8 minutes walking. This is much higher than the effective 1 minute of the traffic lights of Singapore (see section 2.1.2). Also, the burden of two traffic lights is only 50 % higher than for one traffic light. It should be two times this value. An overpass with stairs is identified as the less appreciable crossing facility by the convenience respondents. In contrast, the assumptions of Table 35 would have designated the crossing of four lanes as being the less convenient. Furthermore, there is only a small difference between the burden of crossing four lanes and four lanes (9.2 minutes versus 11.0 minutes). However, it can be considered that jaywalking of four lanes is much more dangerous and that a higher difference should be reported by the respondents. It should be noted that overpasses equipped with lift are perceived as no crossing. This fact in surprising since the Singaporean people of the convenience sample should know that it takes a while to cross overpasses, even with a lift. No aesthetic or coverage utility has been related to the presence of trees on sidewalk. Concerning the route types, parks are valuated with an extremely high benefit. In other words, respondents of the convenience sample would perceive the same utility of walking 20 minutes more through a park than in a minor street. Such a high difference does not make sense. The presence of shops has a four times higher advantage in the convenience than in the SSI sample (- 4.3 minutes versus - 1.3 minutes). According to these findings, it seems that the convenience respondents valuated too much most of the environmental and crossing variables in comparison to the walking time. It should be recalled that this sample is mainly constituted of urban planning and architecture specialists (see section 3.2). Hence, they might put too much emphasis in urban environment and infrastructures. However, it should be considered that some important variables such as overpass with lift have not been valuated. Because of its low goodness of fit and inconsistent values of several attribute ratios, the convenience sample shows a poor quality in terms of choice behavior. The computed model of this sample cannot be used for conducting behavioral research. These findings are surprising considering the high answer consistency of this sample (see section 7.2.2).

Table 45 Comparison of the willingness to avoid of the SSI and convenience models with direct effects

Attribute	**SSI** [min]	**Convenience** [min]
1 trafficlight	1.4	6.8
2 trafficlights	2.5	9.8
Covered sidewalk	-6.4	-9.1
Jaywalk 2lanes	-	9.2
Jaywalk 4lanes	6.9	11.0
Major road	0.8	-
Overpass with lift	2.1	-
Overpass with stairs	3.9	17.5
Park	-2.9	-20.3
Shops	-1.3	-4.3
Trees	-0.9	-
Inside building	-4.0	-
Inside building if rainy	-8.0	-9.9

According to the results of Table 46, several ratios are insignificant in the MTurk model. First, walking close to a major road is considered to be similar to walking in the smaller street. It should be noted that the SSI model with time interaction and the convenience model reported a similar result. However, the presence of shops or of trees on sidewalks is also considered insignificant. Benefits of trees were also perceived in the convenience model. However, both the SSI and convenience models reported a positive utility for shops. Both overpass categories are insignificant in the MTurk and are hence perceived like an absence of crossing. It might be that this kind of infrastructure is less common in the USA and in India than in Singapore. However, there is likely a misunderstanding of the experiment framework, since it is clear that using a bridge to cross a street requires time. Despite of the removal of inconsistent answers, a lack of commitment of the remaining MTurk respondents might not be excluded. As in the convenience model, traffic lights are perceived with a high burden: 4 minutes for one traffic light. However, the burden of two traffic lights is only slightly greater than the one of one traffic light (5.6 minutes). Logically, two traffic lights should have the same willingness to avoid as two times one traffic light, as in the SSI model. Jaywalking four lanes is perceived as the less appreciable crossing type, which is consistent with the hypothesis of Table 35. However, the ratio of jaywalking two lanes is almost of same magnitude as the one of four lanes (7.0 minutes versus 7.6 minutes). A higher difference should be observed since jaywalking four lanes is much more dangerous. As in the convenience model, the road type park shows an excessively high utility compared to minor road: a willingness to walk almost 11 minutes more for the same utility. Finally, covered sidewalk and the road type inside building are perceived as advantageous compared to their reference attributes, as expected in the hypotheses (Table 35). As for the convenience sample, the choice model computed with the MTurk data is of very poor quality because of its low goodness of fit and of several inconsistent attribute ratios. This might be due to a bad understanding of the experiment or to ignorance of the Singaporean urban context. Also, a low commitment of the MTurk respondents cannot be excluded. Indeed, important infrastructures such as overpasses have not been perceived as a walking time burden. However, it should be noticed that the survey process conducted with MTurk is not ideal since three different survey designs and two groups of respondents of two different continents were used to collect answers.

Table 46 Comparison of the willingness to avoid of SSI and MTurk models with direct effects

Attribute	**SSI** [min]	**MTurk** [min]
1 trafficlight	1.4	4.0
2 trafficlights	2.5	5.6
Covered sidewalk	-6.2	-2.6
Jaywalk 2 lanes	-	7.0
Jaywalk 4 lanes	6.9	7.6
Major road	0.8	-
Overpass with lift	2.1	-
Overpass with stairs	4.1	-
Park	-2.9	-11.2
Shops	-1.3	-
Trees	-0.9	-
Inside building	-3.9	-1.5

8 Conclusion

This chapter presents the final conclusions of this study. Section 8.1 is dedicated to the knowledge gained through the realization of this work. Recommendations for further research are stated in section 8.2.

8.1 Gained knowledge

The results of this work indicate that it is possible to collect reliable answers to a stated preference survey on internet. Commercial platforms such as QuestionPro can be used to implement a questionnaire online. However, it comes out from this experience that open-source platforms such as LimeSurvey are at least as suitable as commercial platforms or even more appropriated for conducting complex surveys, e.g. with stated preference experiments (see section 7.1). Indeed, LimeSurvey could support an external experiment plan to generate choice sets in contrast to QuestionPro. Furthermore, open-source tools are free of charge. By contrast, QuestionPro was the cheapest commercial platform evaluated in this work with a monthly subscription fee of 99 $.

200 answers could be collected in a few hours using a commercial panel of SSI and the marketplace of Amazon Mechanical Turk (see section 7.2.1). Convenience sampling required in contrast about one week for gathering 160 answers. Both SSI and MTurk samples showed a very high response rate of 90 %, while 50 % of the respondents of the convenience sample left before the end of the questionnaire. Hence, SSI and MTurk are able to provide samples very quickly and with few respondents dropping out of the survey. By contrast, convenience sampling is quite slow to collect answers and shows a risk of non-response bias (Christensen et al. 2014; Heerwegh and Loosveldt 2008).

While non-trading and lexicographic behavior have not been identified as an important issue, answer inconsistency needs to be tested when working with online sampling. The inconsistency test using two times the same choice sets (redundancy) was more selective than the display of a dominant alternative (see section 7.2.2). Respondents of the convenience samples provided the most consistent answers (13 % error). By contrast, 36 % of the answers of MTurk India had to be removed because of inconsistency. Access limitation to workers with at least a 95 % approval rate was apparently not sufficient for ensuring answer quality as stated by Peer et al. (Peer et al. 2013). MTurk USA and SSI samples showed an answer consistency ranging between these two percentages (19 and 25 % error). Several potential socio-demographic influences on answer consistency were investigated, such as respondent education or age. However, no clear relationship could be assessed.

SSI sample was the most representative of the Singaporean population for most of the investigated socio-demographic characteristics (see section 7.2.3). By contrast, the convenience sample contained mainly young and highly educated people and was hence not very diverse. Both MTurk samples were not comparable to the population of Singapore since their respondents were either Indian or American. Analysis of the walking characteristics showed many similarities between SSI and convenience samples. This highlights the importance of location and culture on the walking behavior.

The choice model computed with SSI data had a quality at least as good as a model estimated with answers collected in the fastidious field survey of FCL (see section 7.3.4). Respondents of SSI perceived even better the burden of traffic lights than the respondents of the FCL study. By contrast, the MTurk and convenience samples provided choice models of very poor quality which could not be used for conducting behavioral research (see section 7.3.5). Especially, no walking time interaction could be computed with these datasets. The models with direct effects of the route characteristics computed with these both samples had very low goodness of fit and many of attribute valuations (willingness to avoid) were not consistent. In the case of the convenience sample, it might be that the respondents paid too much attention to route characteristics because of their urban planning background. They hence neglected the most important attribute: the walking time (Muraleetharan and Hagiwara 2007; Agrawal et al. 2008; Seneviratne and Morrall 1985). In the case of the MTurk sample, ignorance of the urban environment of Singapore might have played an important role in the experiment misunderstanding of these respondents. However, there are some indications such as the insignificance of overpasses infrastructures that some MTurk respondents had a lack of commitment while answering the survey.

Survey design analysis showed that the walking time and two traffic lights attributes were less valuated in a stated preference experiment without pictorial information (see section 7.3.1). The display of a proportional bar next to the walking time values in survey versions with images obviously helped the respondent to perceive this attribute. By contrast, it cannot be explained why two traffic lights have been interpreted as a no crossing by the people who faced stated preference experiments in pure text form. However, these findings indicate that a better perception of experiment attributes is achieved when using pictorial representations, as reported by many authors (Jansen et al. 2009; Rizzi et al. 2012; Vriens et al. 1998). A higher failure rate was observed in the standard version of the stated preference experiments (text and images) than in other designs (see section 7.2.2). However, this difference was slight such as no direct causality can be assessed between stated preference designs and answer inconsistency.

Because of its higher answer consistency (see section 7.2.2) and higher quality of its choice model (see section 7.3.2), American sample of MTurk was of better quality than the one of MTurk India. Furthermore, Indian respondents were insensible to all kinds of jaywalking. This finding might be related to certain cultural aspects of these respondents, e.g. a perception of traffic safety which is different in India than in the USA.

8.2 Recommendations

Because of clear advantages in terms of collection time, flexibility and geographic coverage, researchers should consider online surveys as a very powerful research tool in comparison to traditional interviews and mail surveys. To implement questionnaires on internet, open-source platforms should be used rather than commercial tools since they provide equivalent or better support. Because of positive literature findings about the data quality of several commercial panels (Cook and Loraas 2013; Gligor et al. 2013; Mohr et al. 2012; Spielmann and Delvert 2014) as well as the present successful experience with SSI, commercial respondent pools can be used to collect reliable answers. By contrast, convenience sampling with viral distribution should be avoided because of risks of social clustering. Considering the positive experience of many researchers who worked with MTurk samples (Berinsky et al. 2012; Paolacci et al. 2010; Peer et al. 2013) as well as the low cost of this respondents' source, further research should be conducted to better identify the real potential of Amazon Mechanical Turk for conducting research in transportation planning. Especially, an experiment context which is familiar to the respondent should be used to investigate this question. Since the American sample was more reliable than the Indian one in the present study, the potential of US MTurk respondents should be considered in priority. When conducting stated preference experiments with online sample, consistency checks should be implemented to ensure answer reliability. Based on the findings of this work, a consistency check with a redundant choice set should be preferred to a test using a dominant alternative. However, consistency tests only provide a first selection of the data. They cannot ensure that the respondent really understood the experimental framework. Finally, pictorial representations of attributes should be included in the design of the stated preference experiments, since they help the respondents to perceive these attributes.

9 References

AboutAnalytics (2015). Qualtrics Research Suite, AboutAnalytics, accessed on 04-06-2015, *http://www.aboutanalytics.com/qualtrics-inc/qualtrics-research-suite*.

Adamowicz, W., J. Swait, P. Boxall, J. Louviere and M. Williams (1997). 'Perceptions versus Objective Measures of Environmental Quality in Combined Revealed and Stated Preference Models of Environmental Valuation', *Journal of Environmental Economics and Management* 32(1): 65–84.

Adkins, A., J. Dill, G. Luhr and M. Neal (2012). 'Unpacking Walkability: Testing the Influence of Urban Design Features on Perceptions of Walking Environment Attractiveness', *Journal of Urban Design* 17(4): 499–510.

Agrawal, A., M. Schlossberg and K. J. Irvin (2008). 'How Far, By Which Route, and Why? A Spatial Analysis of Pedestrian Preference', *Journal of Urban Design* 13(1): 81–98.

Arentze, T., A. Borgers, H. Timmermans and R. Del Mistro (2003). 'Transport Stated Choice Responses: Effects of Task Complexity, Presentation Format and Literacy', *Transportation Research Part E: Logistics and Transportation Review* 39(3): 229–244.

Ariffin, R. N. R. and R. K. Zahari (2013). 'Perceptions of the Urban Walking Environments', Asia Pacific International Conference on Environment-Behaviour Studies, University of Westminster, United Kingdom, September 2013.

Amazon Mechanical Turk (2015). Amazon Mechanical Turk, accessed on 26-06-2015, *https://www.mturk.com/mturk/welcome*.

Axhausen, K. W., B. Schmid and C. Weis (2015). 'Predicting response rates updated', *Arbeitsberichte Verkehrs- und Raumplanung*, 1063, IVT, ETH Zürich.

Axhausen, K. W. and C. Weis (2010). 'Predicting Response Rate: A Natural Experiment', *Survey Practice* 3(2).

Ball, K., A. Bauman, E. Leslie and N. Owen (2001). 'Perceived Environmental Aesthetics and Convenience and Company Are Associated with Walking for Exercise among Australian Adults', *Preventive Medicine* 33(5): 434–440.

Berinsky, A. J., G. A. Huber and G. S. Lenz (2012). 'Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk', *Political Analysis* 20(3): 351–368.

Berrens, R. P., A. K. Bohara, H. Jenkins-Smith, C. Silva and D. L. Weimer (2003). 'The Advent of Internet Surveys for Political Research: A Comparison of Telephone and Internet Samples', *Political Analysis* 11(1): 1–22.

Bierlaire, M. (2003). *BIOGEME: A free package for the estimation of discrete choice models [Software],* 3rd Swiss Transportation Research Conference, Ascona, Switzerland.

Blume, L., A. Brandenburger and E. Dekel (1989). 'An Overview of Lexicographic Choice under Uncertainty', *Annals of Operations Research* 19: 231–246.

Börjesson, M. (2012). 'Valuing perceived insecurity associated with use of and access to public transport', *Transport Policy* 22: 1–10.

Bradley, M. A. and A. J. Daly (1991). 'Estimation of Logit Choice Models using Mixed Stated Preference and Revealed Preference Information', Les méthodes d'analyse des comportements de déplacements pour les années 1990 - 6ème Conférence Internationale sur les comportements de déplacements, Château Bonne Entente, Québec, Canada, Mai 1991.

Bradley, M. A. and A. J. Daly (1994). 'Use of the logit scaling approach to test for rank-order and fatigue effects in stated preference data', *Transportation* 21(2): 167–184.

Brandon, D. M., J. H. Long, T. M. Loraas, J. Mueller-Phillips and B. Vansant (2013). 'Online Instrument Delivery and Participant Recruitment Services: Emerging Opportunities for Behavioral Accounting Research', *Behavioral Research in Accounting* 26(1): 1–23.

Buhrmester, M., T. Kwang and S. D. Gosling (2011). 'Amazon's Mechanical Turk A New Source of Inexpensive, Yet High-Quality, Data?', *Perspectives on Psychological Science* 6(1): 3–5.

Chandler, J., P. Mueller and G. Paolacci (2013). 'Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers', *Behavior Research Methods* 46(1): 112–130.

ChoiceMetrics (2015). Ngene, ChoiceMetrics Pty Ltd, accessed on 06-06-2015, *http://www.choice-metrics.com/features.html*.

Christensen, A. I., O. Ekholm, C. Glümer and K. Juel (2014). 'Effect of survey mode on response patterns: comparison of face-to-face and self-administered modes in health surveys', *European Journal of Public Health* 24(2): 327–332.

Christian, L. M., D. A. Dillman and J. D. Smyth (2007). ‘The Effects of Mode and Format on Answers to Scalar Questions in Telephone and Web Surveys’, *Advances in Telephone Survey Methodology*, 250–275, John Wiley & Sons, Inc.

Cobanoglu, C., B. Warde and P. J. Moreo (2009). ‘A comparison of mail, fax and web-based survey methods’, *International Journal of Market Research* 43(4): 441–452.

Cook, K. and T. Loraas (2013). ‘The Influence of Source Credibility on Framing and Resistance Effects: Evidence from the Federal Estate Tax’, Auburn University and Texas Tech University.

De Leeuw, E. D. (2005). ‘To Mix or Not to Mix Data Collection Modes in Surveys.’, *Journal of Official Statistics* 21(5): 233–255.

Dillman, D. A., G. Phelps, R. Tortora, K. Swift, J. Kohrell, J. Berck and B. L. Messer (2009). ‘Response rate and measurement differences in mixed-mode surveys using mail, telephone, interactive voice response (IVR) and the Internet’, *Social Science Research* 38(1): 1–18.

FluidSurveys (2015). FluidSurveys, accessed on 16-06-2015, *http://fluidsurveys.com/*.

Future Cities Laboratory (2015). Measuring Walkability, accessed on 16-06-2015, *http://www.fcl.ethz.ch/project/measuring-walkability/*.

Gaber, J. and S. L. Gaber (2004). ‘If you could see what I know: Moving planners’ use of photographic images from illustrations to empirical data’, *Journal of Architectural and Planning Research* 21(3): 222–238.

Gligor, D. M., M. C. Holcomb and T. P. Stank (2013). ‘A Multidisciplinary Approach to Supply Chain Agility: Conceptualization and Scale Development’, *Journal of Business Logistics* 34(2): 94–108.

Glinski, R. J. and B. C. Glinski (1970). ‘Nonnaivety Contamination in Conformity Experiments: Sources, Effects, and Implications for Control’, *Journal of Personality and Social Psychology* 16(3): 478–485.

Goldin, G. and A. Darlow (2013). *Turkgate (Version 0.4.0) [Software]*.

Gosling, S. D., S. Vazire, S. Srivastava and O. P. John (2004). ‘Should We Trust Web-Based studies? A Comparative Analysis of Six Preconceptions About Internet Questionnaires’, *The American Psychologist* 59(2): 93–104.

Green, P. E. and V. Srinivasan (1978). ‘Conjoint Analysis in Consumer Research: Issues and Outlook’, *Journal of Consumer Research* 5(2): 103–123.

Greenwald, M. and M. Boarnet (2001). ‘Built Environment as Determinant of Walking Behavior: Analyzing Nonwork Pedestrian Travel in Portland, Oregon’, *Transportation Research Record: Journal of the Transportation Research Board* 1780: 33–41.

Griffis, S. E., T. J. Goldsby and M. Cooper (2003). ‘Web-Based and Mail Surveys: A Comparison of Response, Data, and Cost’, *Journal of Business Logistics* 24(2): 237–258.

Groves, R. M., E. Singer and A. Corning (2000). ‘Leverage-Saliency Theory of Survey Participation: Description and an Illustration’, *The Public Opinion Quarterly* 64(3): 299–308.

Guo, Z. (2009). ‘Does the pedestrian environment affect the utility of walking? A case of path choice in downtown Boston’, *Transportation Research Part D: Transport and Environment* 14(5): 343–352.

Hanemann, W. M. (1984). ‘Discrete/Continuous Models of Consumer Demand’, *Econometrica* 52(3): 541–561.

Harline, C. and M. Burris (2014). ‘Impact of Traffic Images on Route Choice and the Value of Time Estimates in Stated Preference Surveys’, *Transportation Research Record: Journal of the Transportation Research Board* 2405: 24–32.

Heerwegh, D and G. Loosveldt (2008). ‘Face-to-Face versus Web Surveying in a High-Internet-Coverage Population’, *Public Opinion Quarterly* 72(5): 836–846.

Heiner, R. A. (1983). ‘The Origin of Predictable Behavior’, *The American Economic Review* 73(4): 560–595.

Hess, S., J. M. Rose and J. Polak (2010). ‘Non-trading, lexicographic and inconsistent behaviour in stated choice data’, *Transportation Research Part D: Transport and Environment, Specification and interpretation issues in behavioural models used for environmental assessment* 15(7): 405–417.

Horton, J. J. and L. B. Chilton (2010). ‘The Labor Economics of Paid Crowdsourcing’, 11th ACM Conference on Electronic Commerce, New York, USA.

Horton, J. J., D. G. Rand and R. J. Zeckhauser (2011). ‘The online laboratory: conducting experiments in a real labor market’, *Experimental Economics* 14(3): 399–425.

Humpel, N., N. Owen, D. Iverson, E. Leslie and A. Bauman (2004). 'Perceived environment attributes, residential location, and walking for particular purposes', *American Journal of Preventive Medicine* 26(2): 119–125.

Hurtubia, R. and P. Donoso (2014). 'Using images to measure qualitative attributes of public spaces through SP surveys', 10th International Conference on Transport Survey Methods, Leura, Australia.

Ipeirotis, P. (2008). Why People Participate on Mechanical Turk, Now Tabulated, accessed on 29-06-2015, *http://www.behind-the-enemy-lines.com/2008/09/why-people-participate-on-mechanical.html*.

Ipeirotis, P. (2009). Turker Demographics vs Internet Demographics, accessed on 29-06-2015, *http://www.behind-the-enemy-lines.com/2009/03/turker-demographics-vs-internet.html*.

Ipeirotis, P. (2013). Mechanical Turk account verification: Why Amazon disables so many accounts, accessed on 29-06-2015, *http://www.behind-the-enemy-lines.com/2013/06/mechanical-turk-account-verification.html*.

Jansen, S., H. Boumeester, H. Coolen, R. Goetgeluk and E. Molin (2009). 'The impact of including images in a conjoint measurement task: evidence from two small-scale studies', *Journal of Housing and the Built Environment* 24(3): 271–297.

Kavanaugh, M. L., D. Bessett, L. L. Littman and A. Norris (2013). 'Connecting Knowledge about Abortion and Sexual and Reproductive Health to Belief about Abortion Restrictions: Findings from an Online Survey', *Women's Health Issues: Official Publication of the Jacobs Institute of Women's Health* 23(4): 239–247.

LimeSurvey (2015). LimeSurvey, accessed on 16-06-2015, *https://www.limesurvey.org/en/*.

Louviere, J. J., D. A. Hensher and J. D. Swait (2000). *Stated Choice Methods: Analysis and Applications*, Cambridge, UK: Cambridge University Press.

Marta-Pedroso, C., H. Freitas and T. Domingos (2007). 'Testing for the survey mode effect on contingent valuation data quality: A case study of web based versus in-person interviews', *Ecological Economics* 62(3-4): 388–398.

Mason, W. and S. Suri (2011). 'Conducting behavioral research on Amazon's Mechanical Turk', *Behavior Research Methods* 44(1): 1–23.

Mason, W. and D. J. Watts (2009). ‘Financial Incentives and the ‘Performance of Crowds’’, *ACM SIGKDD Explorations Newsletter* 11(2): 100–108.

Mohr, G. S., D. R. Lichtenstein and C. Janiszewski (2012). ‘The Effect of Marketer-Suggested Serving Size on Consumer Responses: The Unintended Consequences of Consumer Attention to Calorie Information’, *Journal of Marketing: A Quarterly Publication of the American Marketing Association* 76(1): 59–75.

Muraleetharan, T. and T. Hagiwara (2007). ‘Overall Level of Service of Urban Walking Environment and Its Influence on Pedestrian Route Choice Behavior: Analysis of Pedestrian Travel in Sapporo, Japan’, *Transportation Research Record: Journal of the Transportation Research Board* 2002: 7–17.

Olsen, S. B. (2009). ‘Choosing Between Internet and Mail Survey Modes for Choice Experiment Surveys Considering Non-Market Goods’, *Environmental and Resource Economics* 44(4): 591–610.

Orzechowski, M., T. A. Arentze, A. W. J. Borgers and H. J. P. Timmermans (2005). ‘Alternate methods of conjoint analysis for estimating housing preference functions: Effects of presentation style’, *Journal of Housing and the Built Environment* 20(4): 349–362.

Owen, N., N. Humpel, E. Leslie, A. Bauman and J. F. Sallis (2004). ‘Understanding environmental influences on walking; Review and research agenda’, *American Journal of Preventive Medicine* 27(1): 67–76.

Paolacci, G., J. Chandler and P. G. Ipeirotis (2010). ‘Running Experiments on Amazon Mechanical Turk’, *Judgment and Decision Making* 5(5).

Peer, E., J. Vosgerau and A. Acquisti (2013). ‘Reputation as a sufficient condition for data quality on Amazon Mechanical Turk’, *Behavior Research Methods* 46(4): 1023–1031.

Pontin, J. (2007). Artificial Intelligence, With Help From the Humans, *The New York Times*, accessed on 29-06-2015, h*ttp://www.nytimes.com/2007/03/25/business/yourmoney/25Stream .html?_r=0*.

Pucher, J. and L. Dijkstra (2000). ‘Making Walking and Cycling Safer: Lessons from Europe’, *Transportation Quarterly* 54(3): 25–50.

Qualtrics (2015). Qualtrics, accessed on 16-06-2015, *http://www.qualtrics.com/*.

QuestionPro (2015). QuestionPro, accessed on 16-06-2015, *http://www.questionpro.com/*.

Raj Mehta, E. S. (1995). 'Comparing response rates and response content in mail versus electronic mail surveys.', *Journal of the Market Research Society*: 429–439.

Rizzi, L. I., J. P. Limonado and S. Steimetz (2012). 'The impact of traffic images on travel time valuation in stated-preference choice experiments', *Transportmetrica* 8(6): 427–442.

Rodríguez, D. A. and J. Joo (2004). 'The relationship between non-motorized mode choice and the local physical environment', *Transportation Research Part D: Transport and Environment* 9(2): 151–173.

Rose, J. M., S. Hess and A. T. Collins (2011). 'Impact of Violations of Model Assumptions: The Role of Lexicography, Inconsistent Choices and Changing Preferences in Model Estimation', Transportation Research Board 90th Annual Meeting, Washington D. C., USA, January 2011.

Rose, J. M., S. Hess and A. T. Collins (2013). 'What if My Model Assumptions are Wrong? The Impact of Non-standard Behaviour on Choice Model Estimation', *Journal of Transport Economics and Policy* 47(2): 245–263.

Saelens, B. E. and S. L. Handy (2008). 'Built Environment Correlates of Walking: A Review', *Medicine and Science in Sports and Exercise* 40(7): 550–566.

Saelens, B. E., J. F. Sallis and L. D. Frank (2003). 'Environmental correlates of walking and cycling: findings from the transportation, urban design, and planning literatures', *Annals of Behavioral Medicine: A Publication of the Society of Behavioral Medicine* 25(2): 80–91.

Saelensminde, K. (1998). 'The impact of choice inconsistencies on the valuation of travel time in stated choice studies', World Congress of Environmental and Resource Economics in Venice, Venice, Italy, June 1998.

Saelensminde, K. (2002). 'The Impact of Choice Inconsistencies in Stated Choice Studies', *Environmental and Resource Economics* 23(4): 403–420.

Schmid, B. and K. W. Axhausen (2015). 'Post-Car World: Survey Methods and Response Behavior in the Pre-Test', 15th Swiss Transport Research Conference, Ascona, Switzerland.

Seneviratne, P. N. and J. F. Morrall (1985). 'Analysis of factors affecting the choice of route of pedestrians', *Transportation Planning and Technology* 10(2): 147–159.

Sheehan, K. B. and S. J. McMillan (1999). ‘Response Variation in E-Mail Surveys: An Exploration’, *Journal of Advertising Research* 39: 45–54.

Singapore Department of Statistics (2010). Census of Population 2010, accessed on 21-04-2015, *http://www.singstat.gov.sg/publications/publications-and-papers/population#census_of_population_2010.*

Singapore Department of Statistics (2014). Population Trends 2014, accessed on 21-04-2015, *http://www.singstat.gov.sg/publications/publications-and-papers/population-and-population-structure/population-trends*.

Singapore Ministry of Manpower (2014a). Labour Force in Singapore, 2014, accessed on 21-04-2015, *http://stats.mom.gov.sg/iMAS_PdfLibrary/mrsd_2014LabourForce.pdf#page=17*.

Singapore Ministry of Manpower (2014b). Report on Wage Practices 2014, accessed on 21-04-2015, *http://stats.mom.gov.sg/Pages/Report-On-Wage-Practices-2014.aspx.*

Spielmann, N. and M. Delvert (2014). ‘Adapted or standardized copy: Is non-cultural English the answer?’, *Journal of Business Research* 67(4): 434–440.

Sun, G., N. M. Oreskovic and H. Lin (2014). ‘How do changes to the built environment influence walking behaviors? a longitudinal study within a university campus in Hong Kong’, *International Journal of Health Geographics* 13(1): 28.

Survey Analytics (2015). Survey Analytics, accessed on 16-06-2015, *https://www.surveyanalytics.com/.*

SurveyMonkey (2015). SurveyMonkey, accessed on 16-06-2015, *https://www.surveymonkey.com*.

Survey Sampling International (2015). Survey Sampling International, accessed on 26-06-2015, *www.surveysampling.com*.

Tourangeau, R. (2004). ‘Survey Research and Societal Change’, *Annual Review of Psychology* 55(1): 775–801.

Train, K. E. (2009). *Discrete Choice Methods with Simulation*, Cambridge , Cambridge University Press.

Tversky, A. and D. Kahneman (1986). ‘Rational Choice and the Framing of Decisions’, *The Journal of Business* 59(4): 251–278.

Vaske, J. J. (2008). *Survey Research and Analysis: Applications in Parks, Recreation and Human Dimensions*, Venture Publishing.

Vriens, M., G. H. Loosschilder, E. Rosbergen and D. R. Wittink (1998). 'Verbal versus Realistic Pictorial Representations in Conjoint Analysis with Design Attributes', *Journal of Product Innovation Management* 15(5): 455–467.

Wang, J., S. Faridani and P. Ipeirotis (2011). 'Estimating the Completion Time of Crowdsourced Tasks using Survival Analysis Models', *Crowdsourcing for search and data mining* 31.

Weible, R. and J. Wallace (1998). 'Cyber research: The impact of the Internet on data collection', *Marketing Research* 10(3): 18–31.

Wooldridge, J. M. (2012). *Introductory Econometrics: A Modern Approach 5th Edition*, Michigan State University.

Yun, G. W. and C. W. Trumbo (2000). 'Comparative Response to a Survey Executed by Post, E-mail, & Web Form', *Journal of Computer-Mediated Communication* 6(1).

A 1 Walkability Survey

Explanations

Figure 25 Explanation for the stated preference experiment (Part 1)

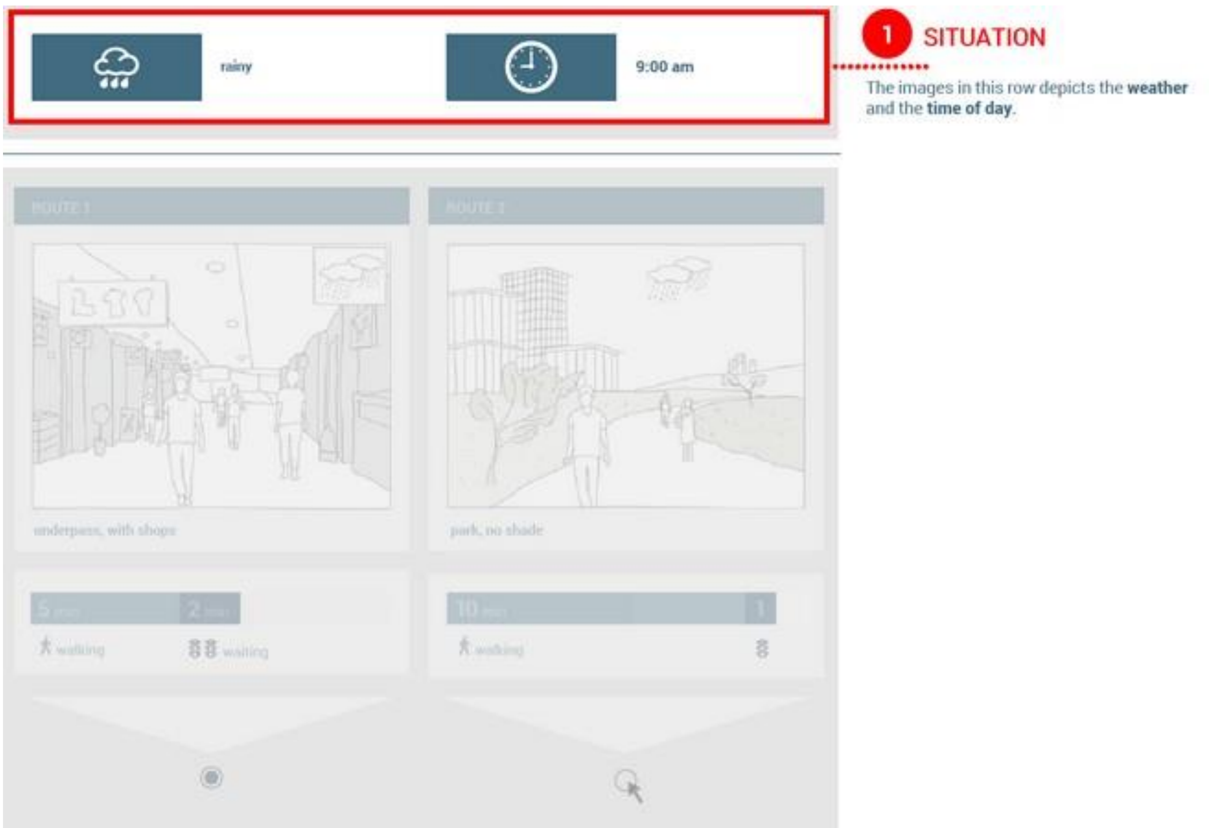


Figure 26 Explanation for the stated preference experiment (Part 2)

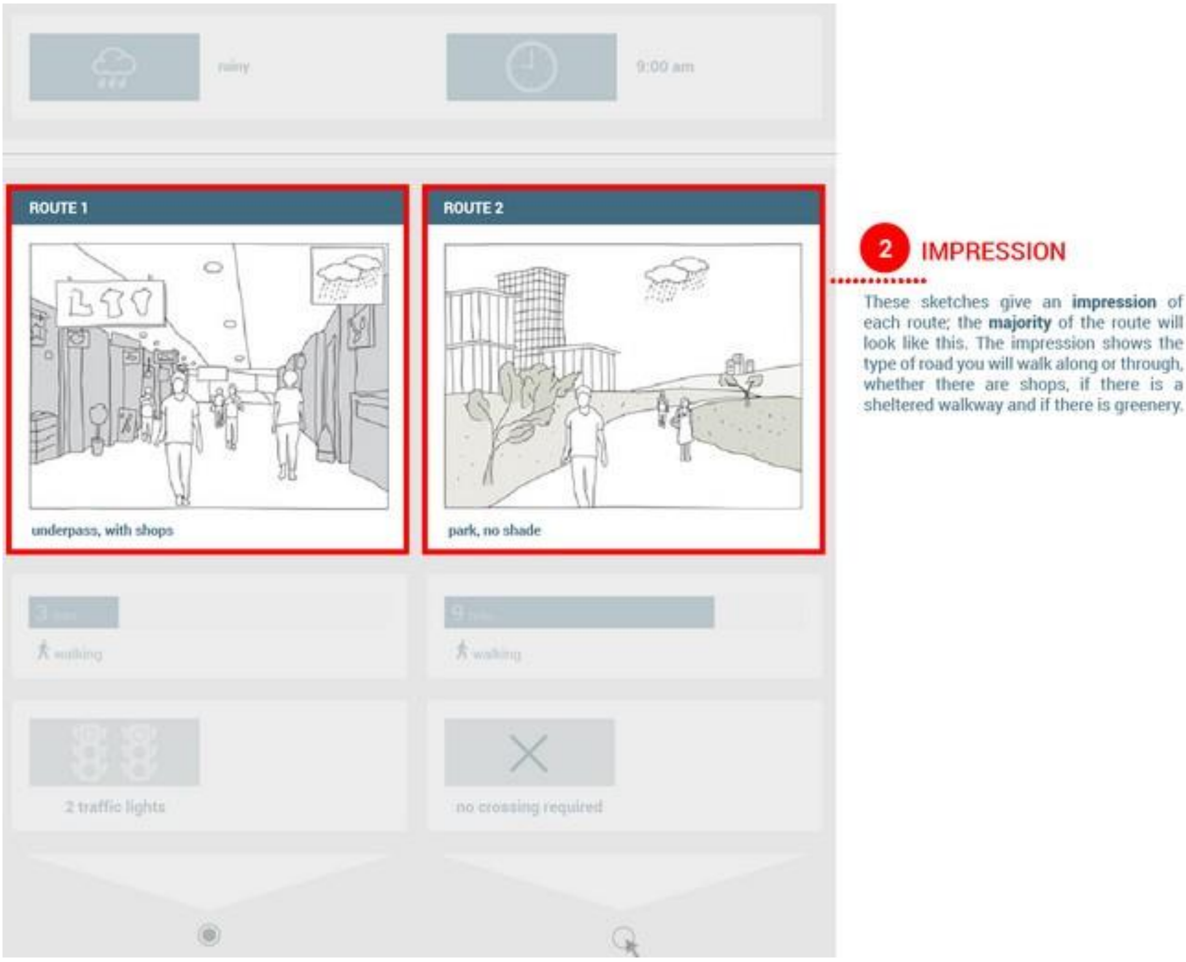


Figure 27 Explanation for the stated preference experiment (Part 3)

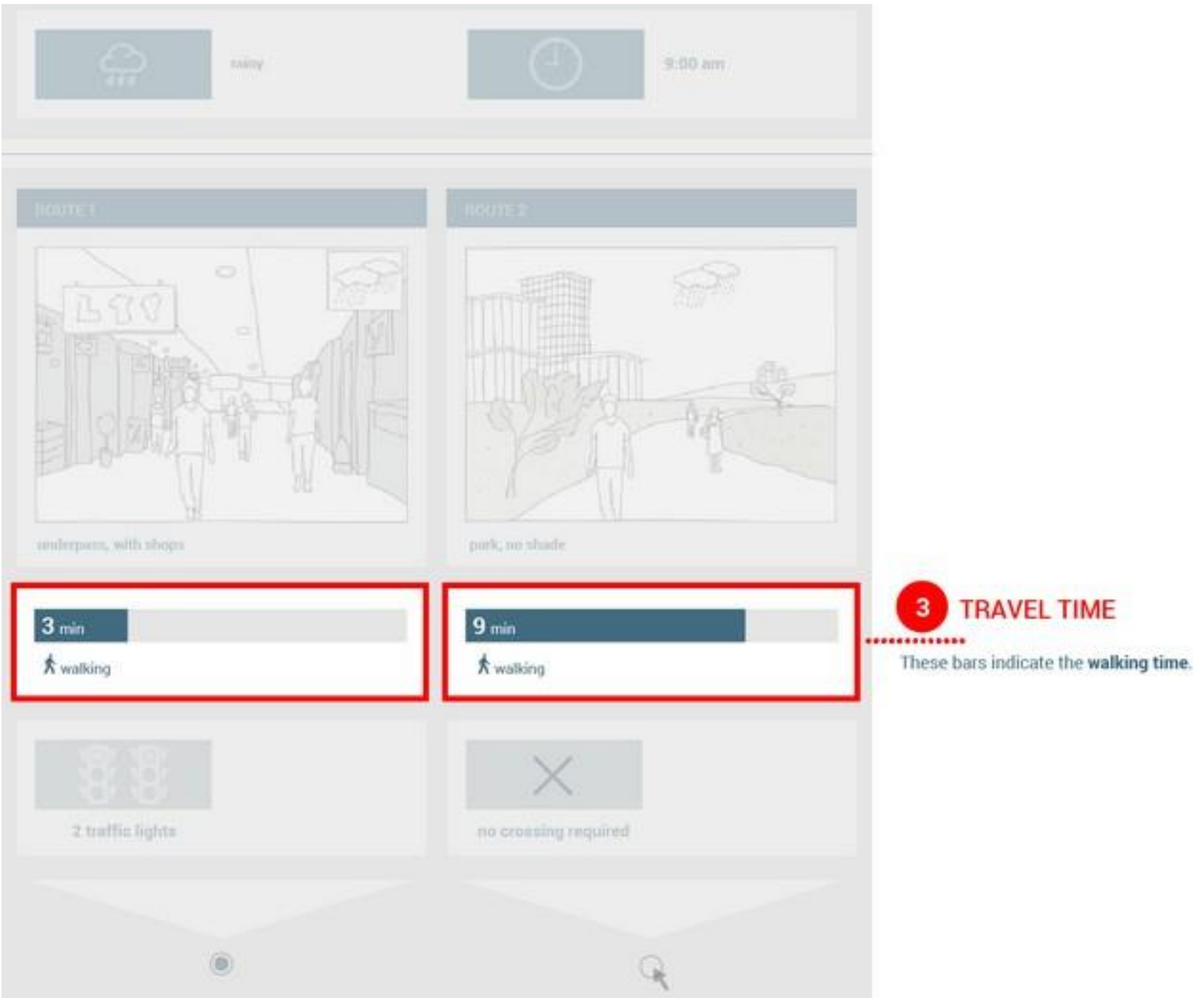


Figure 28 Explanation for the stated preference experiment (Part 4)

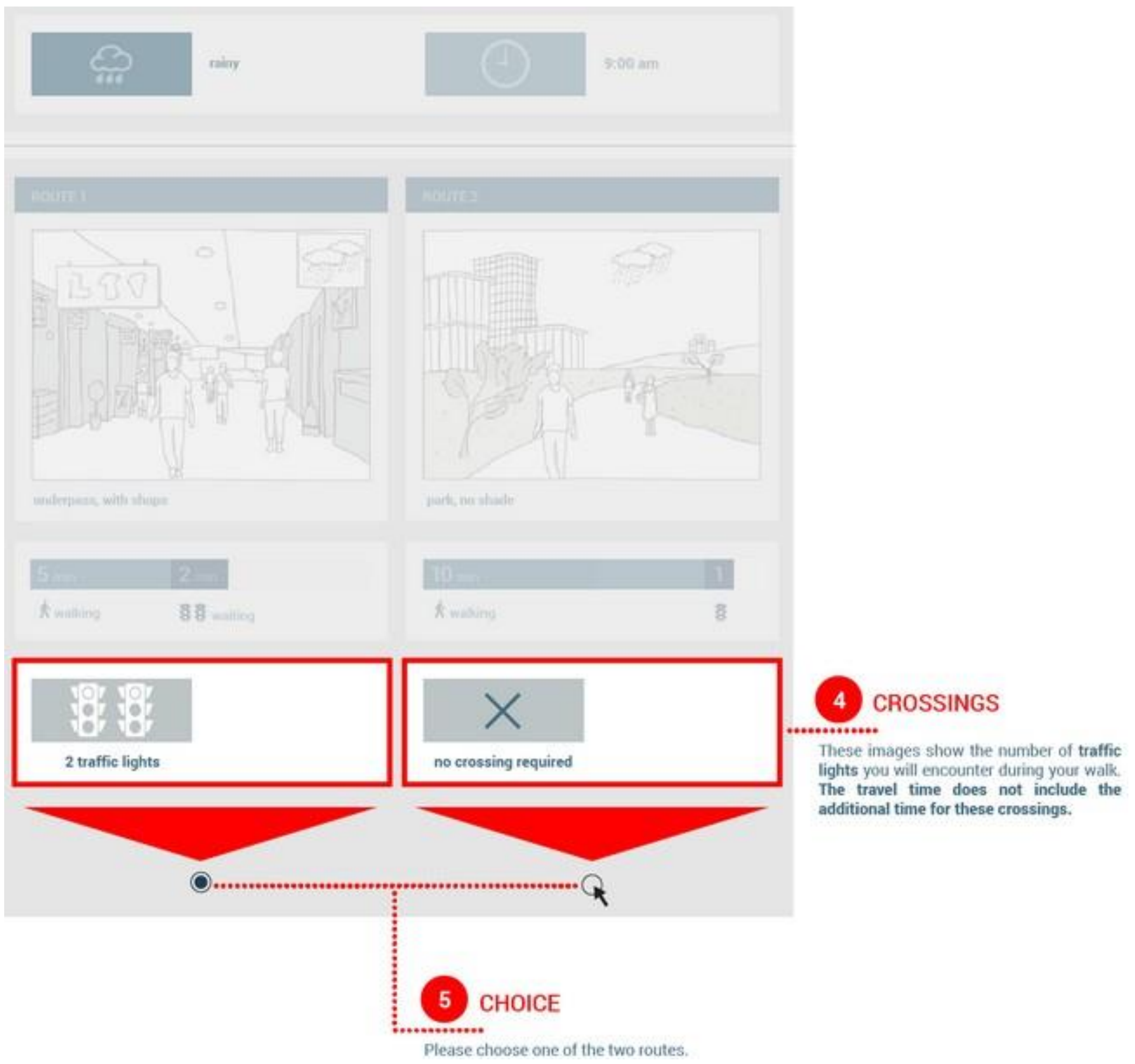
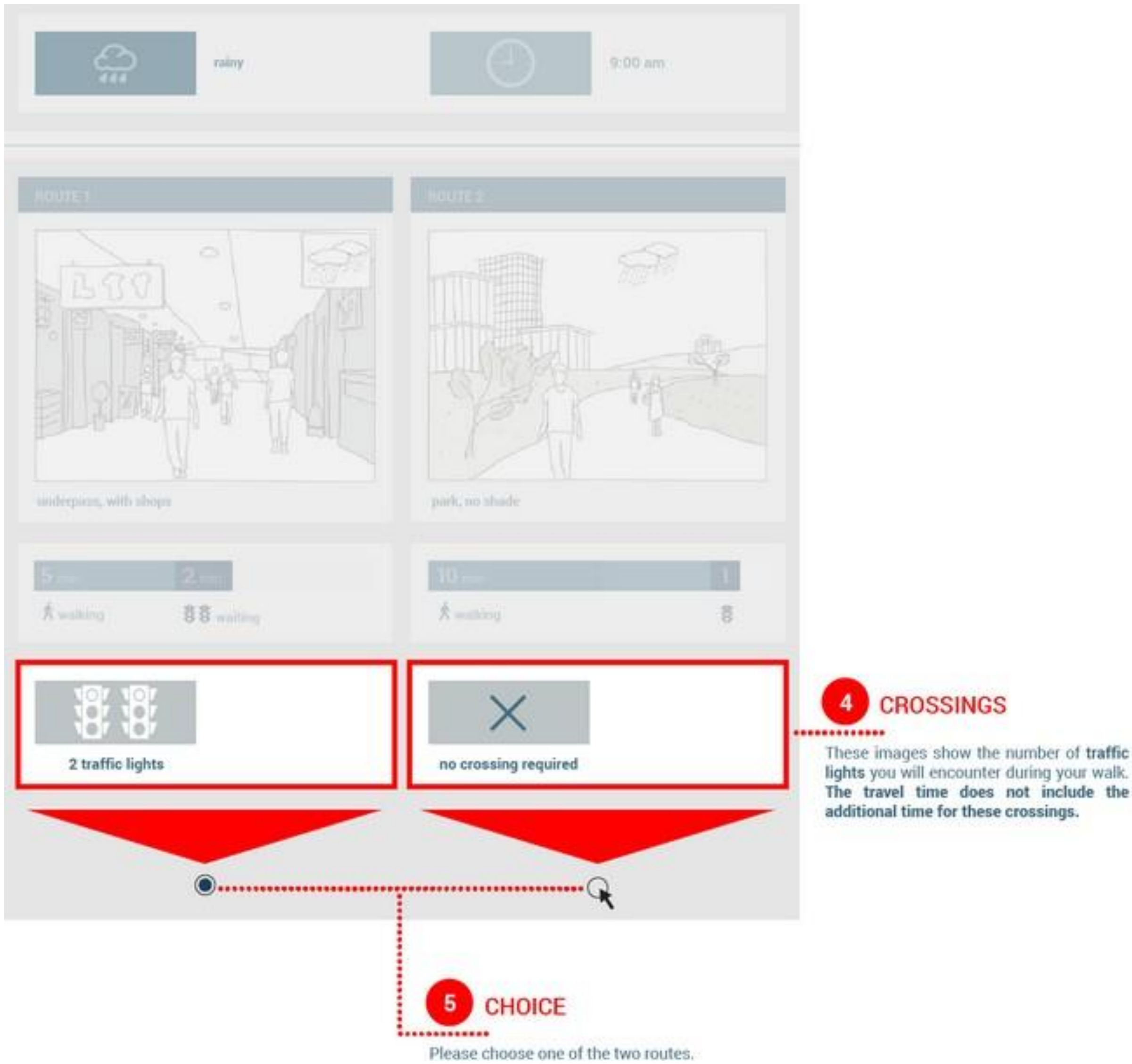


Figure 29 Explanation for the stated preference experiment (Part 5)

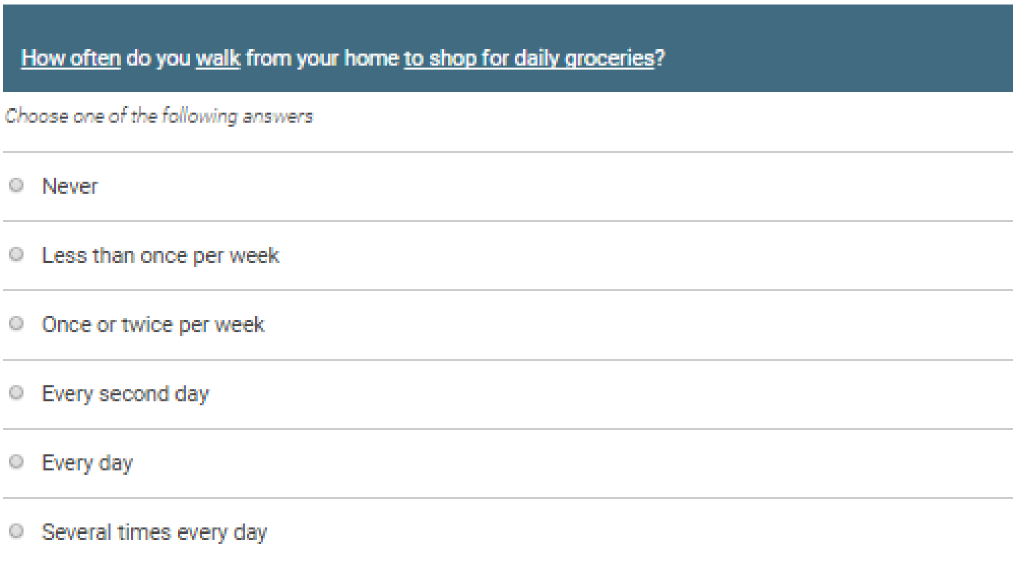


Dominant Alternative

Figure 30 Dominant alternative of the consistency test

Questions about Walking Behavior

Figure 31 Walking frequency for daily groceries

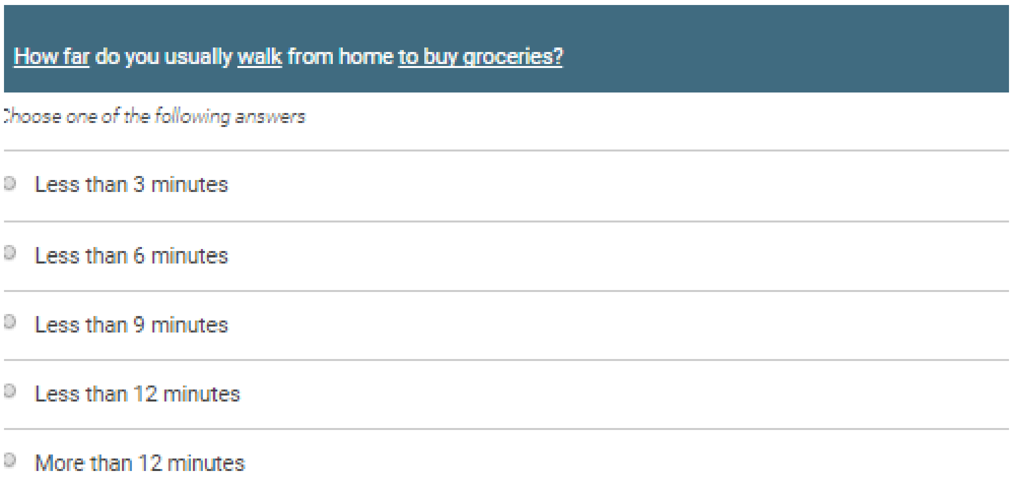


How often do you walk from your home to shop for daily groceries?

Choose one of the following answers

- Never
- Less than once per week
- Once or twice per week
- Every second day
- Every day
- Several times every day

Figure 32 Walking time for daily groceries



How far do you usually walk from home to buy groceries?

Choose one of the following answers

- Less than 3 minutes
- Less than 6 minutes
- Less than 9 minutes
- Less than 12 minutes
- More than 12 minutes

Figure 33 Walking frequency for eating out

How often do you walk from your home to eat out?

Choose one of the following answers

- Never
- Less than once per week
- Once or twice per week
- Every second day
- Every day
- Several times every day

Figure 34 Walking time for eating out

How far do you usually walk from home to eat out?

Choose one of the following answers

- Less than 3 minutes
- Less than 6 minutes
- Less than 9 minutes
- Less than 12 minutes
- More than 12 minutes

Figure 35 Walking frequency for leisure activities

How often do you walk from your home to leisure activities?

Choose one of the following answers

- Never
- Less than once per week
- Once or twice per week
- Every second day
- Every day
- Several times every day

Figure 36 Walking time for leisure activities

How far do you usually walk to leisure activities?

Choose one of the following answers

- Less than 3 minutes
- Less than 6 minutes
- Less than 9 minutes
- Less than 12 minutes
- More than 12 minutes

Respondent Location Questions

Figure 37 Location in Singapore

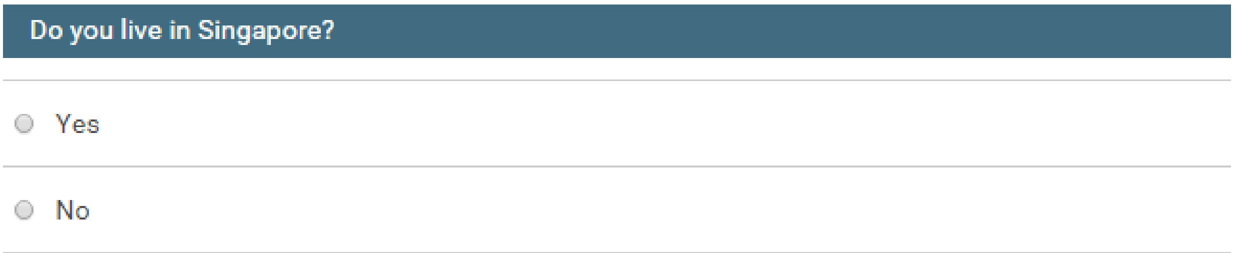


Figure 38 Country of residence (if not from Singapore)

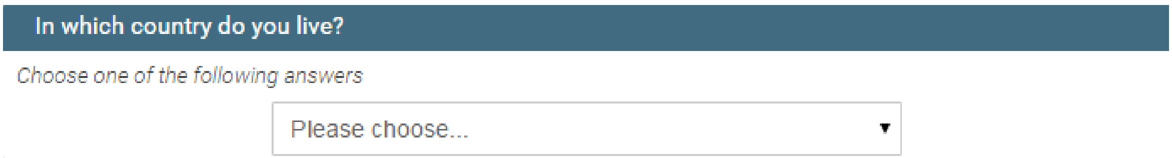


Figure 39 Location urbanity (only for MTurk respondents)

Figure 40 Municipality size (only for MTurk respondents)

What is the size of the locality / agglomeration where you live? (number of inhabitants)

Choose one of the following answers

- I don't live in a locality
- Less than 200
- 200 to 499
- 500 to 999
- 1,000 to 1,999
- 2,000 to 4,999
- 5,000 to 9,999
- 10,000 to 19,999
- 20,000 to 49,999
- 50,000 to 99,999
- 100,000 to 499,999
- 500,000 to 1 million
- Between 1 million and 5 millions
- Between 5 millions and 10 millions
- 10 millions or more

Socio-Demographic Questions

Figure 41 Respondent Gender

Figure 42 Respondent Age

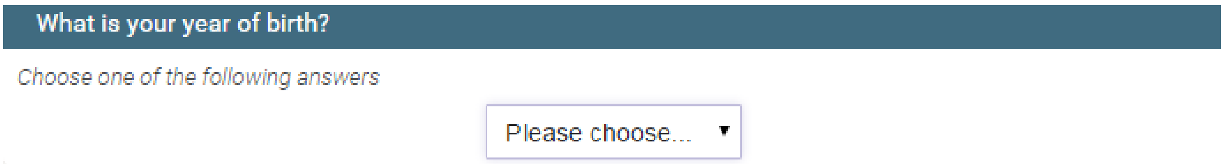


Figure 43 Residential Status of the respondent (only for Singaporean residents)

Figure 44 Respondent Ethnicity (only for Singaporean residents)

What is your ethnicity?

Choose one of the following answers

- Chinese
- Malay
- Indian
- Other

Figure 45 Respondent Nationality (only for Singaporean residents)

What is your nationality (in your passport)?

Choose one of the following answers

- Singaporean
- Chinese
- Malaysian
- Indian
- Indonesian
- Other

Figure 46 Travel frequency to CBD (only for Singaporean residents)

How often do you visit the downtown area of Singapore?

Choose one of the following answers

- Never
- Less than once per week
- Once or twice per week
- Every second day
- Every day
- Several times every day

This area includes the CBD, Orchard, Chinatown, Little India and Bugis.

Figure 47 Highest level of education

Which is the highest level of education you achieved?

Choose one of the following answers

- No education
- Primary
- Lower Secondary
- Secondary
- Post-Secondary (Non-Tertiary)
- University or Polytechnic
- Professional Qualification or Other Diploma
- Other

Figure 48 Present work situation

Which of these propositions best describes your present work situation?

Choose one of the following answers

- Full-time employed (more than 30 hours per week)
- Part-time employed (less than 30 hours per week)
- Student
- Retired
- Homemaker
- Mandatory Military Service
- Unemployed
- Permanently sick / disabled
- Other

Figure 49 Personal Income (in SGD)

What is your monthly income (in SGD)?

Choose one of the following answers

Please choose... ▾

Figure 50 Personal Income (in USD)

What is your monthly income (in US Dollars)?

Choose one of the following answers

Please choose...

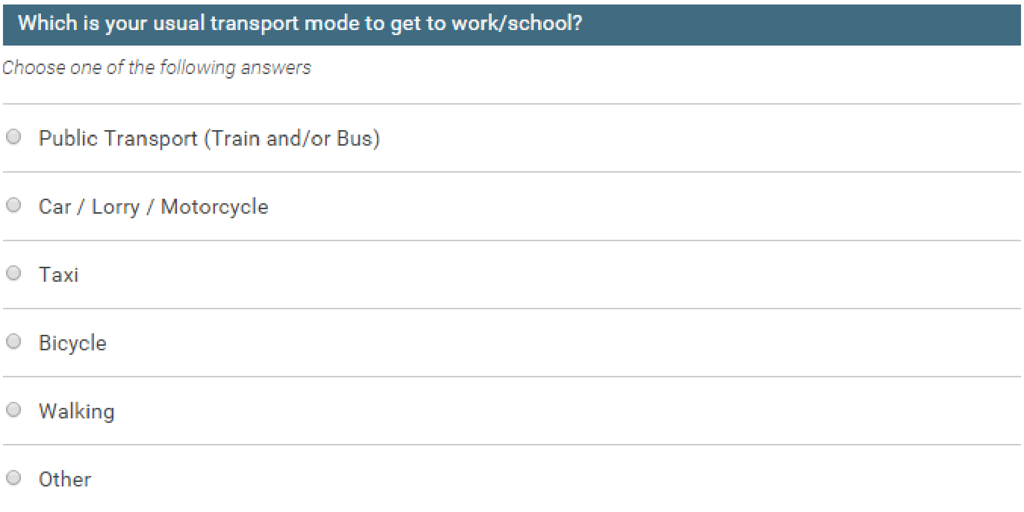
Figure 51 Usual transport mode to get to school/work

Which is your usual transport mode to get to work/school?

Choose one of the following answers

- Public Transport (Train and/or Bus)
- Car / Lorry / Motorcycle
- Taxi
- Bicycle
- Walking
- Other

Figure 52 Usual transport mode for leisure activities

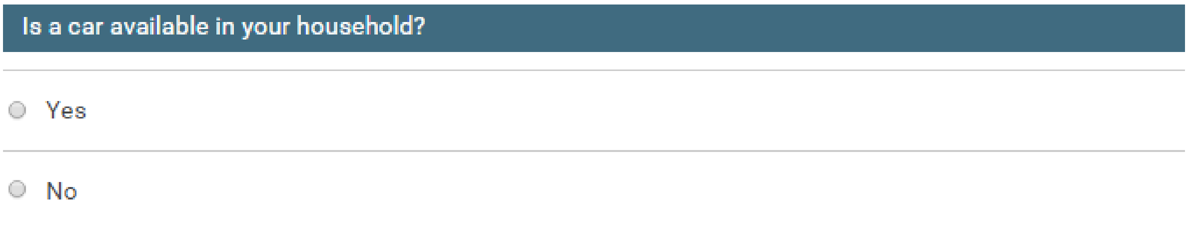


Which is your usual transport mode to get to work/school?

Choose one of the following answers

- Public Transport (Train and/or Bus)
- Car / Lorry / Motorcycle
- Taxi
- Bicycle
- Walking
- Other

Figure 53 Car Ownership in the respondent household



Is a car available in your household?

- Yes
- No

Figure 54 Main language of the household

Which language do you speak at home?

Choose one of the following answers

- English
- Mandarin / Chinese Dialect
- Malay
- Indian Language
- Other

Figure 55 English proficiency level of the respondent (only for MTurk respondents)

How do you rate your English level?

Choose one of the following answers

- First language or Bilingual
- Advanced level
- Intermediate level
- Basic Knowledge
- Other

A 2 Data Analysis

Response Rate Estimation

An ex-ante completion rate of the designed survey is estimated with the help of the regression models of Axhausen et al., respectively of Schmid and Axhausen (Axhausen et al. 2015; Schmid and Axhausen 2015). First, the response burden of the survey is calculated using the using the scoring system of Axhausen et al. (Axhausen et al. 2015). A minimal and a maximal burden are estimated, based on the branching logic of the survey design. The answers are presented in Table 47, Table 48 and Table 49.

Table 47 Estimation of the response burden of the commercial survey

Question number	Description	Points Min.	Points Max.
INTRO1	6 Lines Text	6	6
INTRO2	2 Lines Text	2	2
INTRO3	Map + 3 Lines Text	3	3
INTRO4	Map + 4 Lines Text	4	4
INTRO5	Map + 1 Line Text	3	3
INTRO6	Map + 3 Lines Text	3	3
INTRO7	Map + 2 Lines Text	3	3
SP1	1 Line + 1 SP 2 Alternatives + 7 variables	11	11
SP2	1 Line + 1 SP 2 Alternatives + 7 variables	11	11
SP3	1 Line + 1 SP 2 Alternatives + 7 variables	11	11
SP4	1 Line + 1 SP 2 Alternatives + 7 variables	11	11
SPPart2-INTRO	Map + 6 Lines Text	4	4
SP5	1 Line + 1 SP 2 Alternatives + 7 variables	11	11
SP6	1 Line + 1 SP 2 Alternatives + 7 variables	11	11
SP7	1 Line + 1 SP 2 Alternatives + 7 variables	11	11
SPcons	1 Line + 1 SP 2 Alternatives + 7 variables	11	11
SP8	1 Line + 1 SP 2 Alternatives + 7 variables	11	11
QWalking-INTRO	2 Lines Text	2	2
QW1	1 Line Text + Scale with more than 3 Notes	4	4
QW2	1 Line Text + Scale with more than 3 Notes	0	4
QW3	1 Line Text + Scale with more than 3 Notes	4	4
QW4	1 Line Text + Scale with more than 3 Notes	0	4
QW5	1 Line Text + Scale with more than 3 Notes	4	4
QW6	1 Line Text + Scale with more than 3 Notes	0	4
QLocation-INTRO	2 Lines Text	2	2
QL1	1 Line Text + Yes/NO	3	3
QL2	1 Line Text + Scale with more than 3 Notes	4	0
QL3	1 Line Text + Scale with more than 3 Notes	0	4
QL4	1 Line Text + Number Answer	0	3
QPers-INTRO	2 Lines Text	2	2
QP1	1 Line + YES/NO	3	3
QP2	1 Line Text + Scale with more than 3 Notes	4	4
QP3	1 Line Text + Scale with more 3 Notes	0	4
QP4	1 Line Text + Scale with more than 3 Notes	0	4
QP5	1 Line Text + Scale with more than 3 Notes	0	4
QP6	1 Line Text + Scale with more than 3 Notes	4	4
QP7	1 Line Text + Scale with more than 3 Notes	4	4
QP8	1 Line Text + Scale with more than 3 Notes	4	4
QP8	1 Line Text + Scale with more than 3 Notes	0	4
QP10	1 Line Text + Scale with more than 3 Notes	0	4
QP11	1 Line Text + Scale with more than 3 Notes	0	4
QP12	1 Line Text + YES/NO	3	3
QP13	1 Line Text + Scale with more than 3 Notes	4	4
QP14	1 Line Text + Scale with more than 3 Notes	4	4
QFinal	2 Lines Text	2	2
	Burden Total	**184**	**223**

Table 48 Estimation of the response burden of the convenience survey

Question number	Description	Points Min.	Points Max.
INTRO1	6 Lines Text	6	6
INTRO2	5 Lines Text	5	5
INTRO3	Map + 3 Lines Text	3	3
INTRO4	Map + 4 Lines Text	4	4
INTRO5	Map + 1 Line Text	3	3
INTRO6	Map + 5 Lines Text	5	5
SP1	1 Line + 1 SP 2 Alternatives + 7 variables	11	11
SP2	1 Line + 1 SP 2 Alternatives + 7 variables	11	11
SP3	1 Line + 1 SP 2 Alternatives + 7 variables	11	11
SP4	1 Line + 1 SP 2 Alternatives + 7 variables	11	11
SPPart2-INTRO	Map+6 Lines Text	4	4
SP5	1 Line + 1 SP 2 Alternatives + 7 variables	11	11
SP6	1 Line + 1 SP 2 Alternatives + 7 variables	11	11
SP7	1 Line + 1 SP 2 Alternatives + 7 variables	11	11
SPcons	1 Line + 1 SP 2 Alternatives + 7 variables	11	11
SP8	1 Line + 1 SP 2 Alternatives + 7 variables	11	11
QWalking-INTRO	2 Lines Text	2	2
QW1	1 Line Text + Scale with more than 3 Notes	4	4
QW2	1 Line Text + Scale with more than 3 Notes	0	4
QW3	1 Line Text + Scale with more than 3 Notes	4	4
QW4	1 Line Text + Scale with more than 3 Notes	0	4
QW5	1 Line Text + Scale with more than 3 Notes	4	4
QW6	1 Line Text + Scale with more than 3 Notes	0	4
QLocation-INTRO	2 Lines Text	2	2
QL1	1 Line Text + Yes/NO	3	3
QL2	1 Line Text + Scale with more than 3 Notes	4	0
QL3	1 Line Text + Scale with more than 3 Notes	0	4
QL4	1 Line Text + Number Answer	0	3
QPers-INTRO	2 Lines Text	2	2
QP0	1 Line Text + Scale with more 3 Notes	4	4
QP1	1 Line + YES/NO	3	3
QP2	1 Line Text + Scale with more than 3 Notes	4	4
QP3	1 Line Text + Scale with more 3 Notes	0	4
QP4	1 Line Text + Scale with more than 3 Notes	0	4
QP5	1 Line Text + Scale with more than 3 Notes	0	4
QP6	1 Line Text + Scale with more than 3 Notes	4	4
QP7	1 Line Text + Scale with more than 3 Notes	4	4
QP8	1 Line Text + Scale with more than 3 Notes	4	4
QP8	1 Line Text + Scale with more than 3 Notes	0	4
QP10	1 Line Text + Scale with more than 3 Notes	0	4
QP11	1 Line Text + Scale with more than 3 Notes	0	4
QP12	1 Line + YES/NO	3	3
QP13	1 Line Text + Scale with more than 3 Notes	4	4
QP14	1 Line Text + Scale with more than 3 Notes	4	4
QFinal	2 Lines Text	2	2
	Burden	**190**	**229**

Table 49 Estimation of the response burden of the MTurk survey

Question number	Description	Points Min.	Points Max.
INTRO1	6 Lines Text	6	6
INTRO2	5 Lines Text	5	5
INTRO3	(Map + 3 Lines Text)	0	3
INTRO4	(Map + 4 Lines Text)	0	4
INTRO5	(Map + 1 Line Text)	0	3
INTRO6	(Map + 5 Lines Text)	0	5
SP1	1 Line + 1 SP 2 Alternatives + 7 variables	11	11
SP2	1 Line + 1 SP 2 Alternatives + 7 variables	11	11
SP3	1 Line + 1 SP 2 Alternatives + 7 variables	11	11
SP4	1 Line + 1 SP 2 Alternatives + 7 variables	11	11
SPPart2-INTRO	(Map +) 6 (2) Lines Text	3	4
SP5	1 Line + 1 SP 2 Alternatives + 7 variables	11	11
SP6	1 Line + 1 SP 2 Alternatives + 7 variables	11	11
SP7	1 Line + 1 SP 2 Alternatives + 7 variables	11	11
SPcons	1 Line + 1 SP 2 Alternatives + 7 variables	11	11
SP8	1 Line + 1 SP 2 Alternatives + 7 variables	11	11
QWalking-INTRO	2 Lines Text	2	2
QW1	1 Line Text + Scale with more than 3 Notes	4	4
QW2	1 Line Text + Scale with more than 3 Notes	0	4
QW3	1 Line Text + Scale with more than 3 Notes	4	4
QW4	1 Line Text + Scale with more than 3 Notes	0	4
QW5	1 Line Text + Scale with more than 3 Notes	4	4
QW6	1 Line Text + Scale with more than 3 Notes	0	4
QPers-INTRO	2 Lines Text	2	2
QP1	1 Line + YES/NO	3	3
QP2	1 Line Text + Scale with more than 3 Notes	4	4
QP6	1 Line Text + Scale with more than 3 Notes	4	4
QP7	1 Line Text + Scale with more than 3 Notes	4	4
QP8	1 Line Text + Scale with more than 3 Notes	4	4
QP9	1 Line Text + Scale with more than 3 Notes	0	4
QP12	1 Line + YES/NO	3	3
QP14	1 Line Text + Scale with more than 3 Notes	4	4
QLocalization-INTRO	1 Line Text + Scale with more than 5 Notes	5	5
QL5	1 Line Text + Scale with more than 3 Notes	4	4
QL6	1 Line Text + Scale with more than 3 Notes	4	4
QFinal	2 Lines Text	2	2
	Burden	**170**	**202**

The completion rate is estimated using equation 10 from Schmid and Axhausen (Schmid and Axhausen 2015). Regression coefficients are presented in Table 50.

$$Completion_{estimate,average} = \beta_0 \cdot \exp(\beta_{burden} \cdot \frac{burden_{average}}{100}) \qquad (10)$$

$Completion_{estimate,average}$: Estimation of the average ex-ante completion rate [%]
β_0: Model's constant [-]
β_{burden}: Average burden coefficient [-]
$burden_{average}$: Average burden of the survey [points]

Table 50 Regression coefficients for predicting the ex-ante completion rate up in dependency of the presence of incentive and/or recruitment

Variable	**Samples' Type**		
	Recruit + Incentive	Recruit + No Incentive	No Recruit + No Incentive
β_0	93.297	75.256	38.462
β_{burden}	-0.047	-0.062	-0.116

Source: Schmid and Axhausen 2015

For the SSIl and MTurk surveys, the coefficient estimates of the model with recruitment and incentive have been considered. Indeed, it is assumed that the information provided by Amazon, respectively by the SSI might be similar to recruitment. In the case of the convenience sample, the situation is less clear. On the one hand, no recruitment occurs but on the other hand, the proximity of some respondents to the author might be related to a kind of recruiting effect. In this case, the two models recruiting and no incentive, respectively no recruiting and no incentive are estimated. The results are presented in Table 12 of section 5.1.

Completion Time

Table 51 Answering time per burden point for the whole survey

Sample	Answering Time Average [sec/point]	Median [sec/point]	SD [sec/point]
SSI	2.95	2.05	2.76
MTurk India	2.43	2.25	0.99
MTurk USA	2.04	1.86	0.79
Convenience	2.93	2.41	2.23

Table 52 Answering time per burden point for the stated preference questions

Sample	Answering Time Average [sec/point]	Median [sec/point]	SD [sec/point]
MTurk India	1.65	1.50	0.83
MTurk USA	1.59	1.36	0.91
Convenience	2.45	1.72	4.30

Table 53 Answering time per burden point for the introduction texts and multiple choice questions

Sample	Answering Time Average [sec/point]	Median [sec/point]	SD [sec/point]
MTurk India	3.44	3.20	1.47
MTurk USA	2.72	2.44	1.03
Convenience	3.32	2.79	1.76

Figure 56 Histogram of the answering time per burden point of the SSI sample for the whole survey

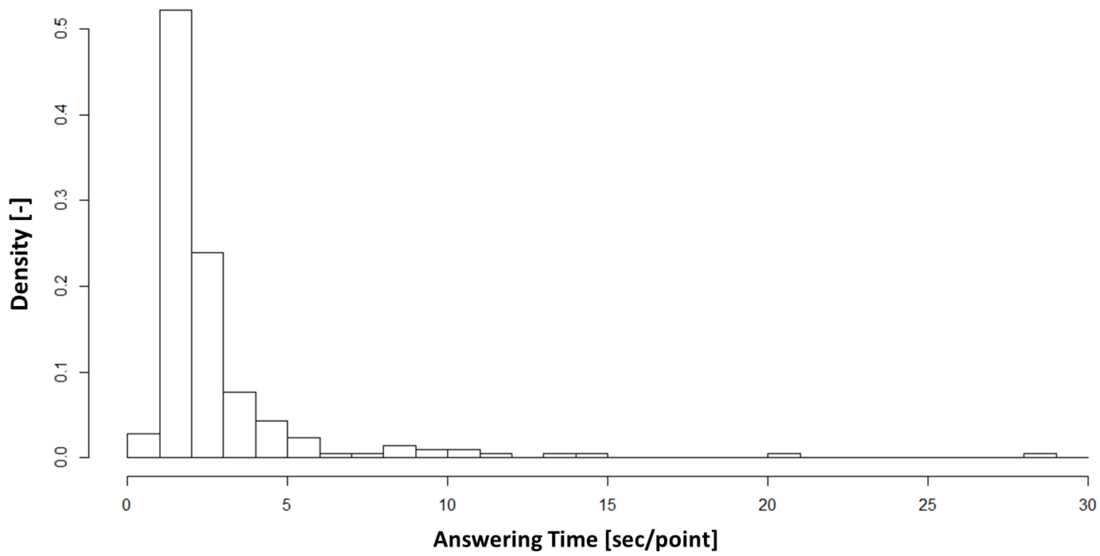


Figure 57 Histogram of the answering time per burden point of the Indian MTurk sample for the whole survey

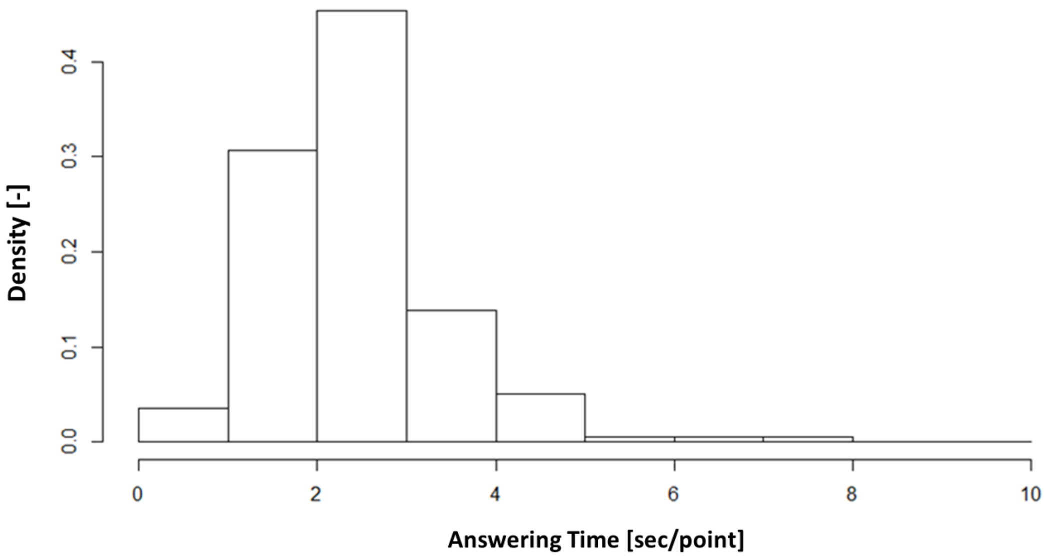


Figure 58 Histogram of the answering time per burden point of the American MTurk sample for the whole survey

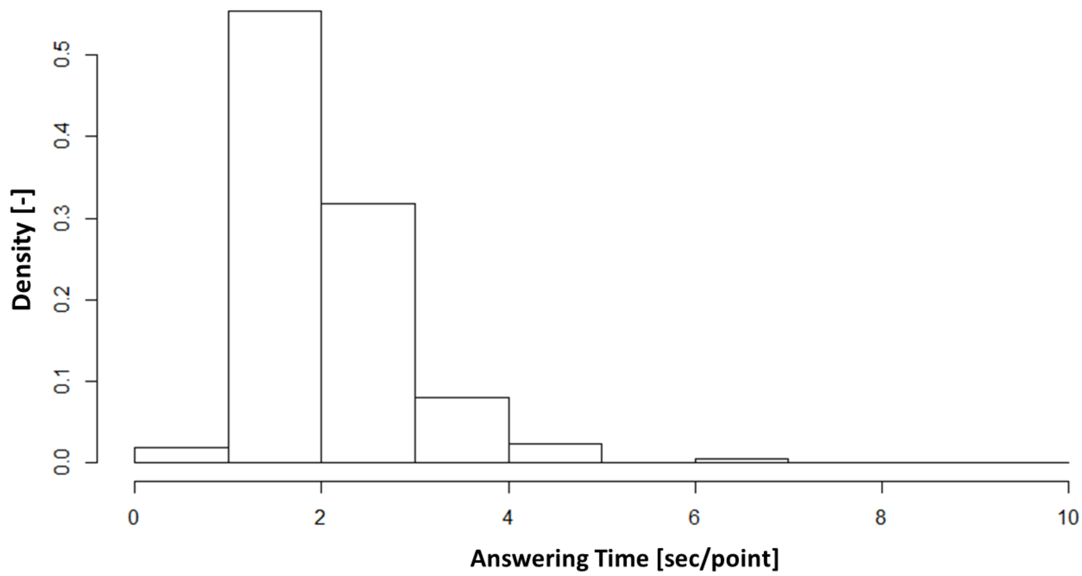


Figure 59 Histogram of the answering time per burden point of the convenience sample for the whole survey

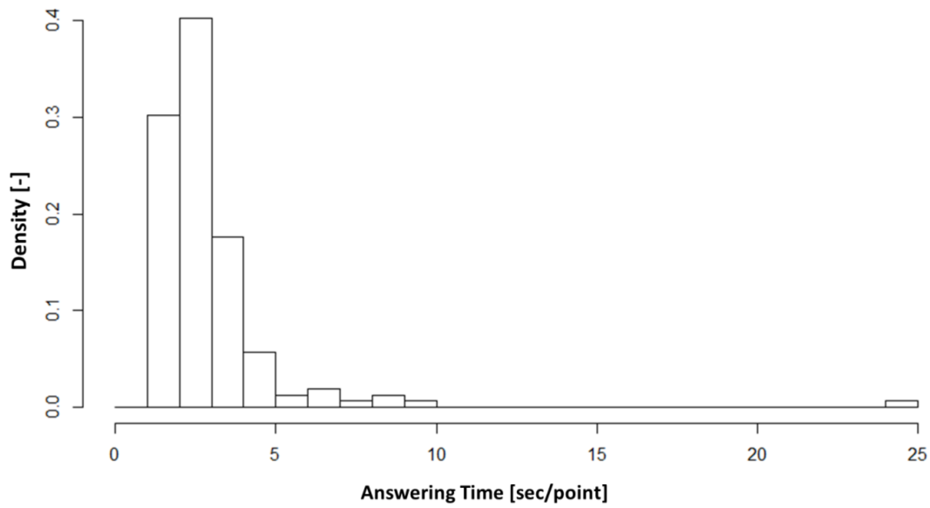


Figure 60 Histogram of the answering time per burden point of the Indian MTurk sample for the SP questions

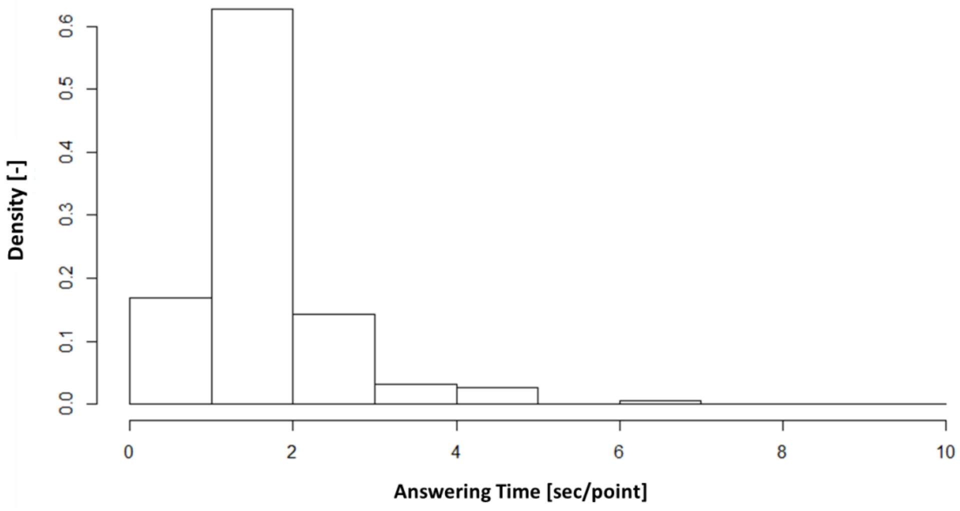


Figure 61 Histogram of the answering time per burden point of the American MTurk sample for the SP questions

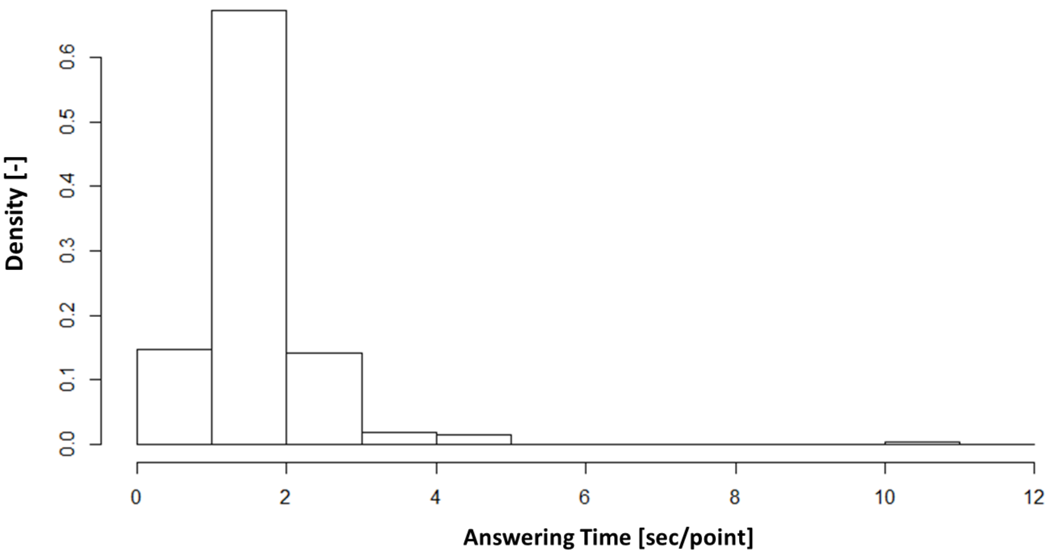


Figure 62 Histogram of the answering time per burden point of the convenience sample for the SP questions

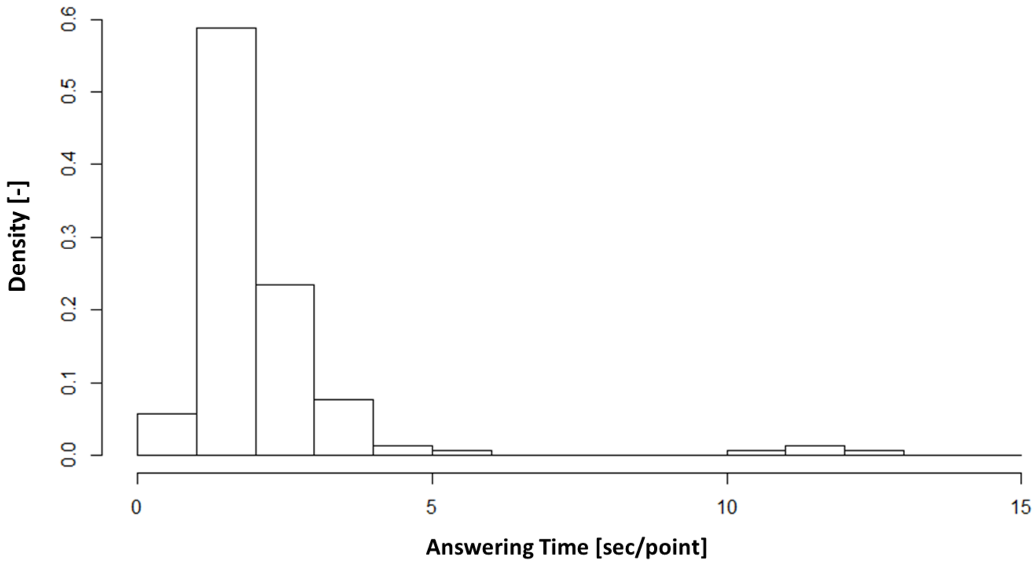


Figure 63 Histogram of the answering time per burden point of the Indian MTurk sample for the text introductions and personal questions

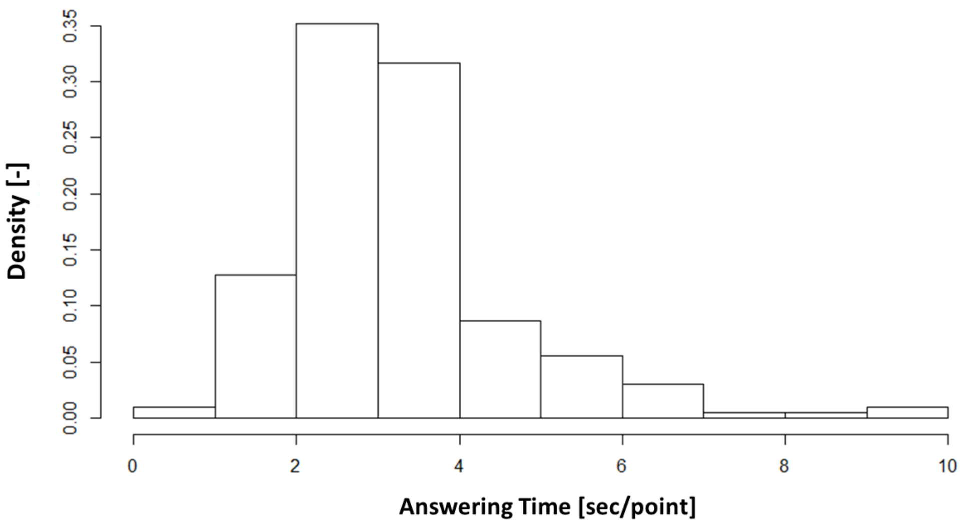


Figure 64 Histogram of the answering time per burden point of the American MTurk sample for the text introductions and personal questions

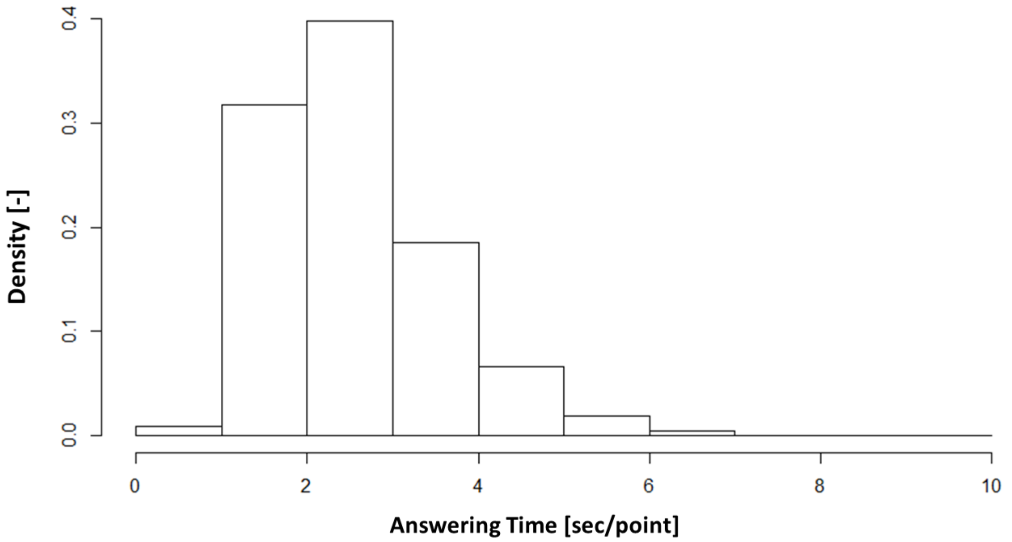
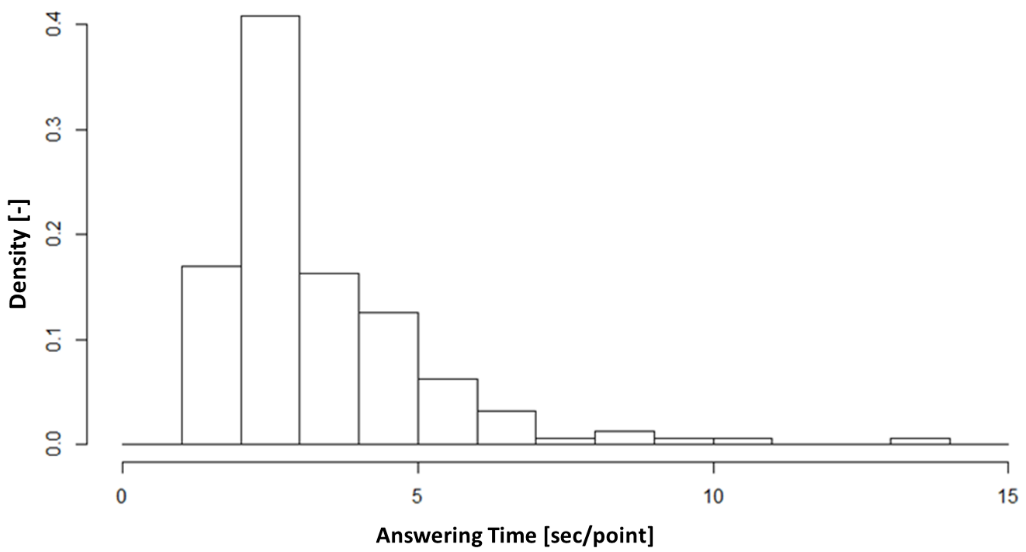


Figure 65 Histogram of the answering time per burden point of the convenience sample for the text introductions and personal questions



Walking Characteristics

Figure 66 Heatmap of respondent shares [in %] of the SSI sample according to walking time and frequency to buy daily groceries

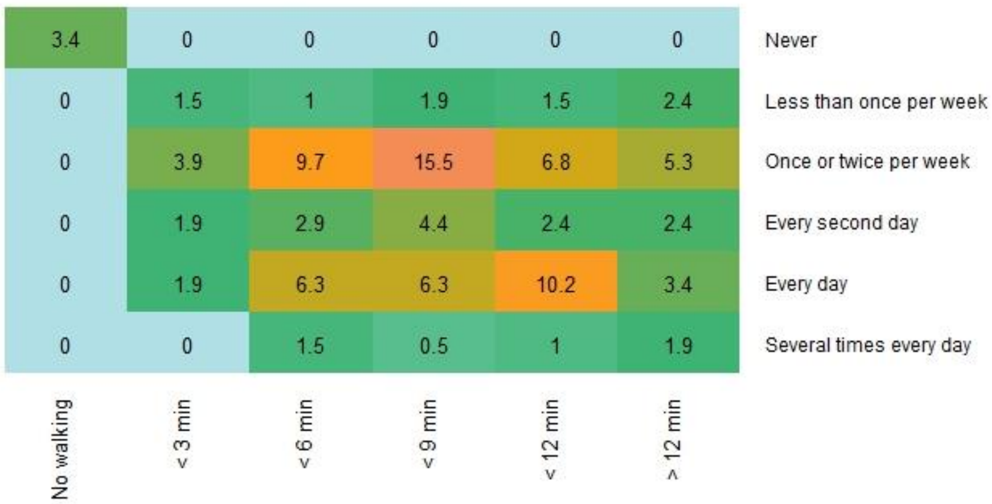


Figure 67 Heatmap of respondent shares [in %] of the MTurk sample according to walking time and frequency to buy daily groceries

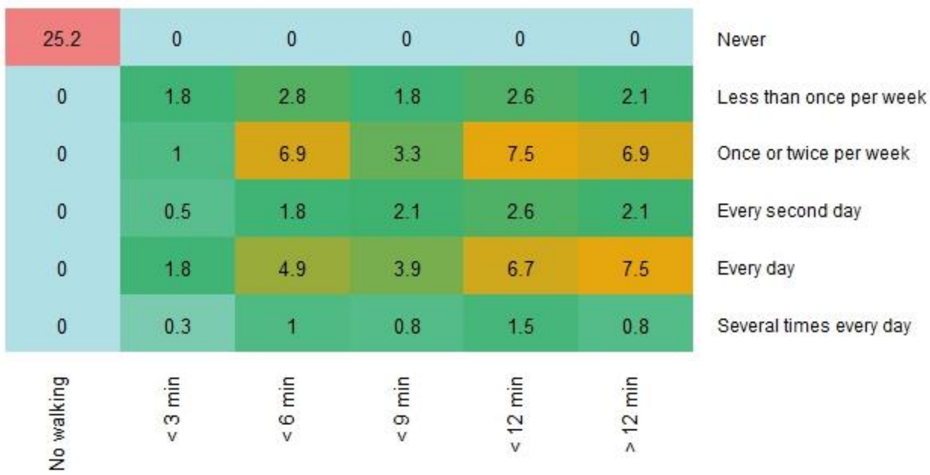


Figure 68 Heatmap of respondent shares [in %] of the convenience sample according to walking time and frequency to buy daily groceries

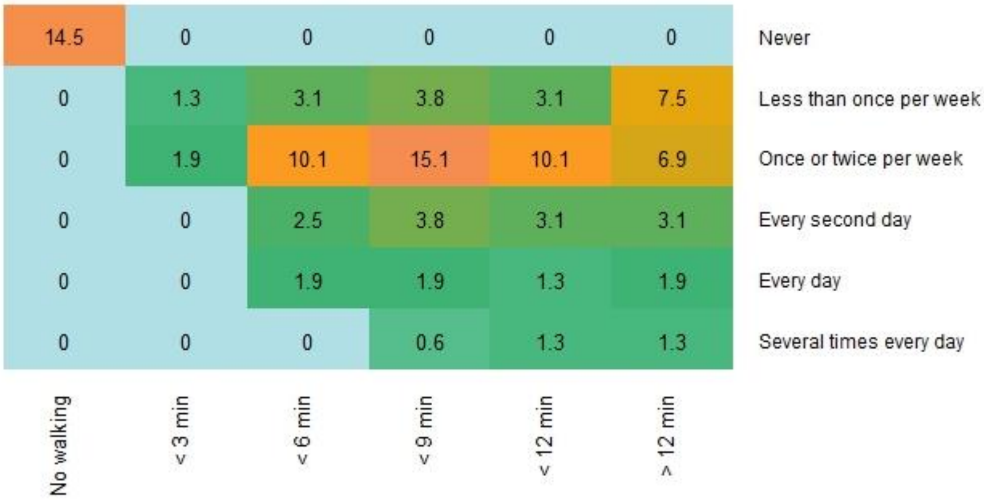


Figure 69 Heatmap of respondent shares [in %] of the SSI sample according to walking time and frequency to eat out

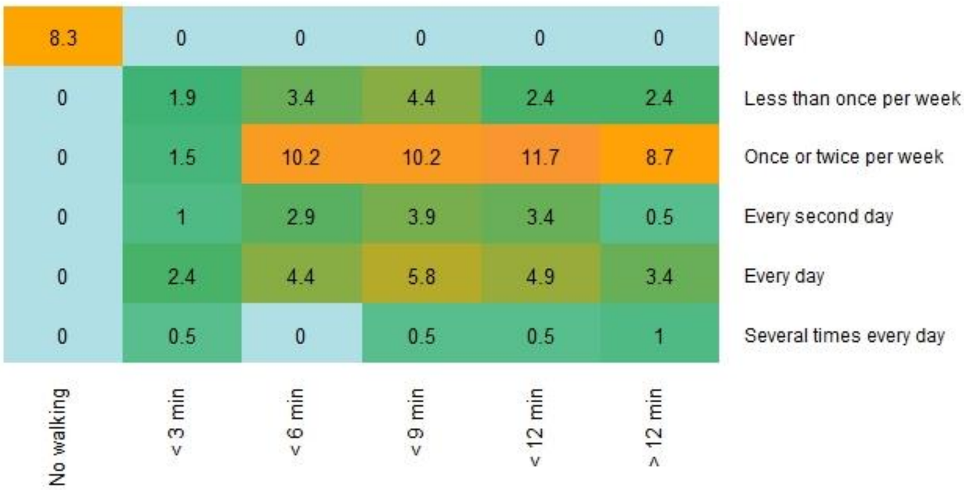


Figure 70 Heatmap of respondent shares [in %] of the MTurk sample according to walking time and frequency to eat out

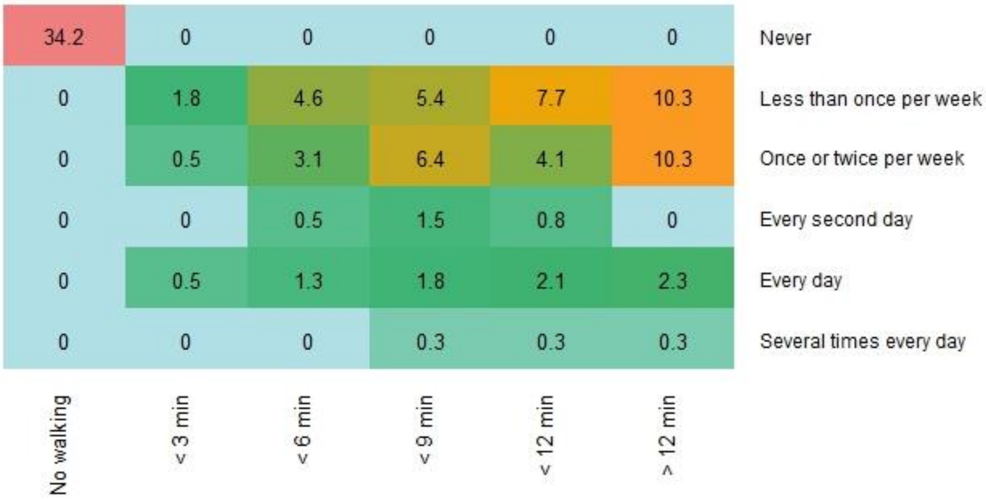


Figure 71 Heatmap of respondent shares [in %] of the convenience sample according to walking time and frequency to eat out

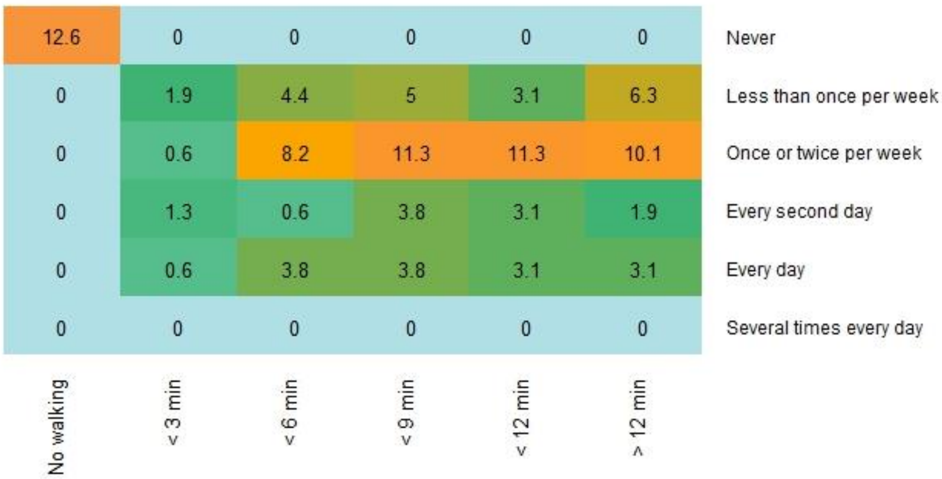


Figure 72 Heatmap of respondent shares [in %] of the SSI sample according to walking time and frequency to get to leisure activities

Figure 73 Heatmap of the respondent proportions [in %] of the MTurk sample according to walking time and frequency to get to leisure activities

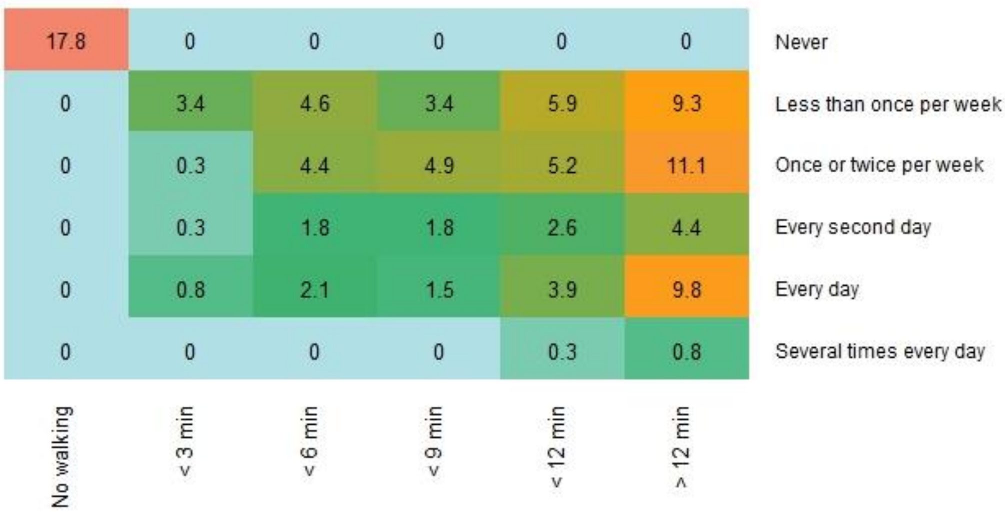
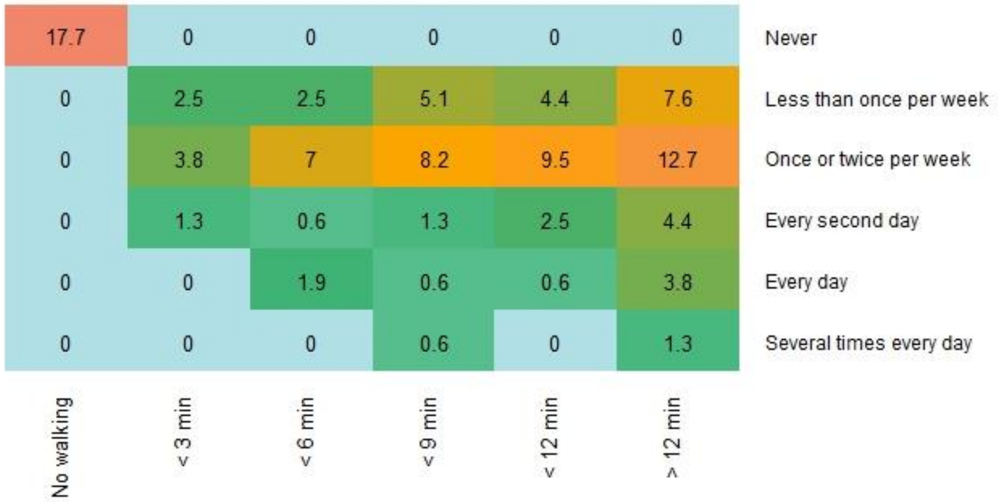
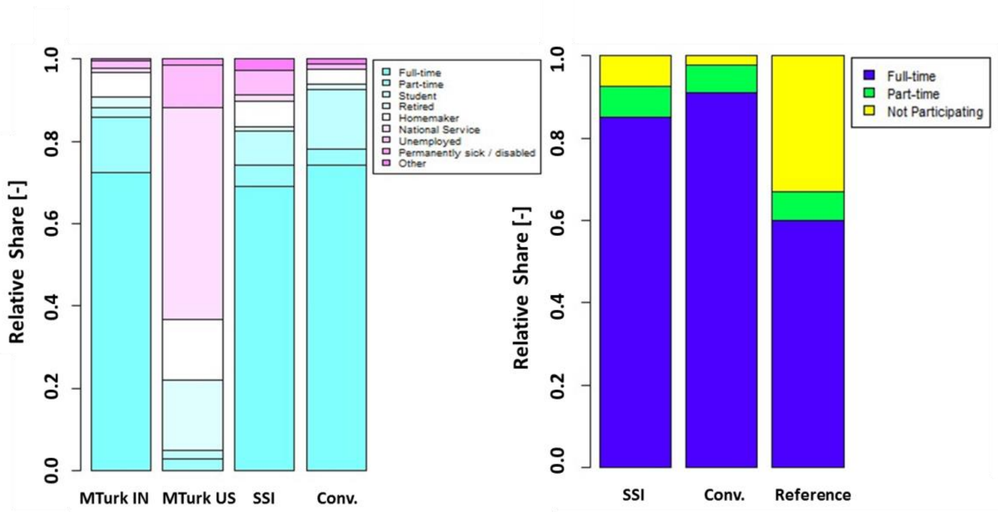


Figure 74 Heatmap of respondent shares [in %] of the convenience sample according to walking time and frequency to get to leisure activities



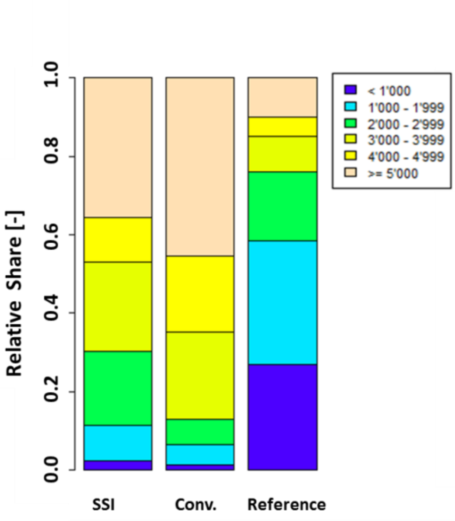
Socio-Demographic Characteristisc

Figure 75 Employment statistics of the whole population per sample type (left) and of the Singaporean population with reference value (right)



References' source: Singapore Ministry of Manpower (2014a)

Figure 76	Monthly wage [SGD] statistics for the Singaporean population with reference value



References' source: Singapore Ministry of Manpower (2014b)

Figure 77	Monthly wage [USD] statistics for the MTurk respondents per country of origin

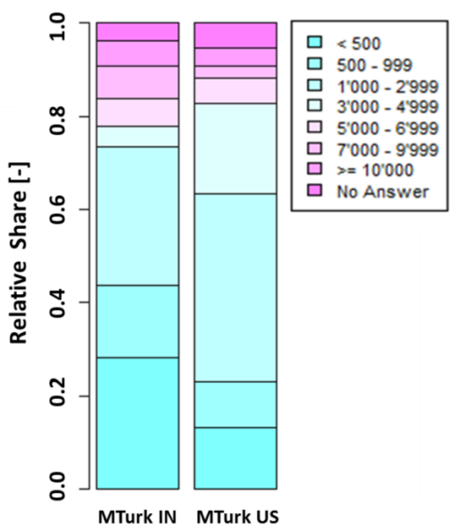


Figure 78 Transport mode statistics for commuting purpose per survey type

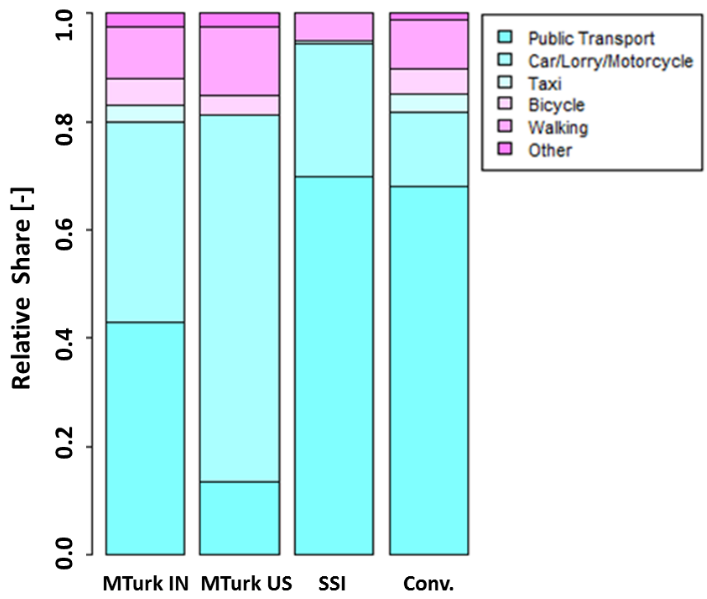
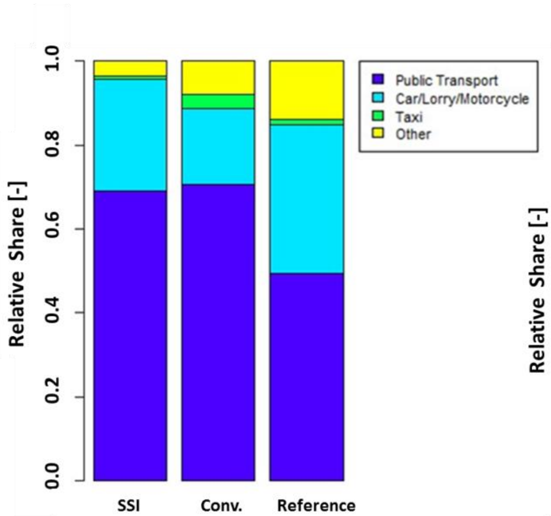
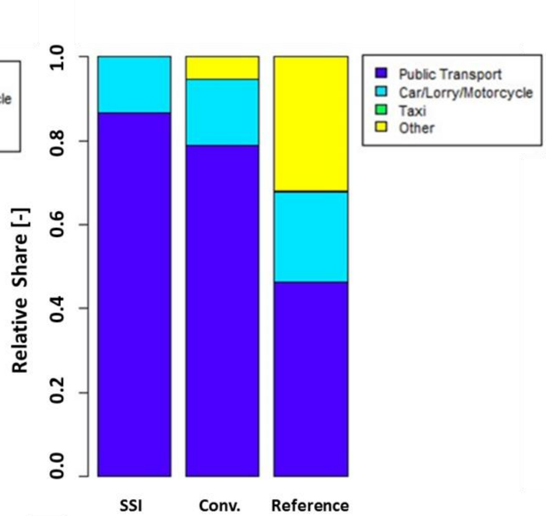


Figure 79 Transport mode statistics for commuting to work (left) respectively to school (right) for the Singaporean population with reference value





References' source: Singapore Department of Statistics (2010)

Figure 80 Transport mode statistics for leisure purpose per survey type

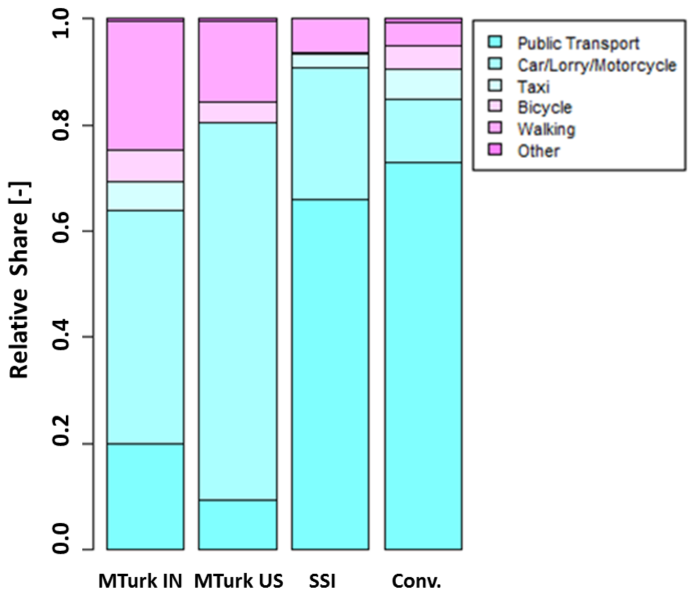


Figure 81 Car ownership statistics per survey type

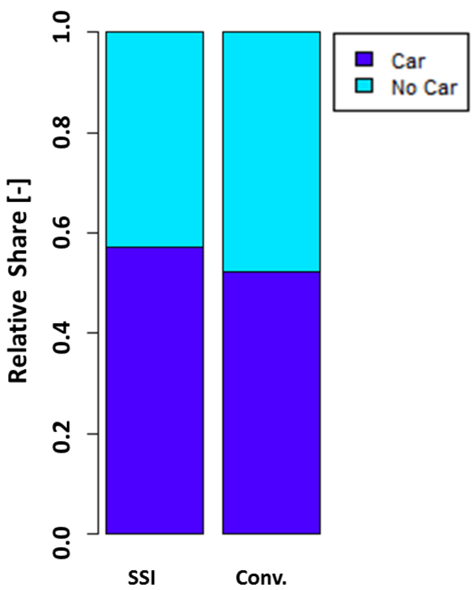
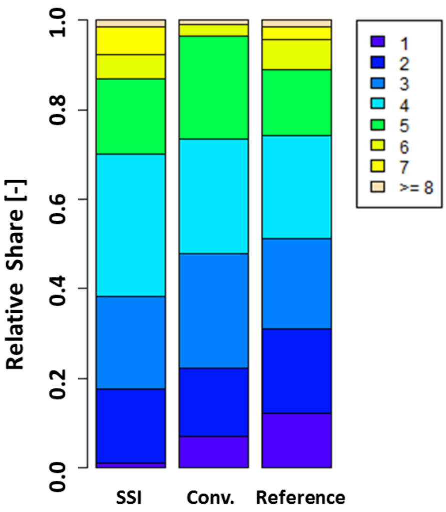
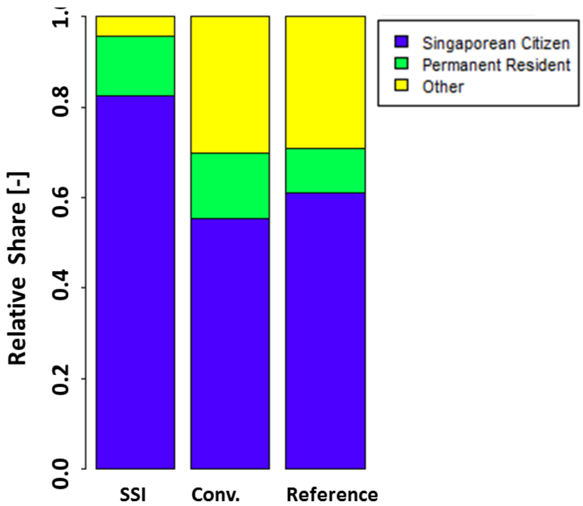


Figure 82 Household size statistics [persons] for the Singaporean population with reference value



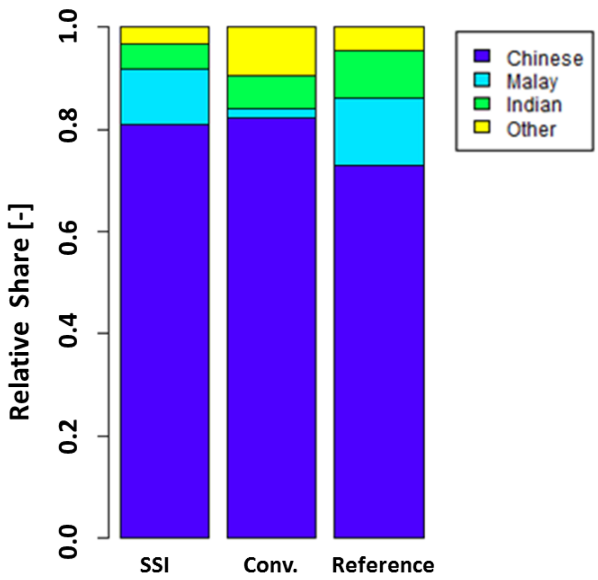
References source: Singapore Department of Statistics (2014)

Figure 83 Residential status statistics for the Singaporean population with reference value



Reference source: Singapore Department of Statistics (2014)

Figure 84 Ethnicity statistics for the Singaporean population with reference value



Reference source: Singapore Department of Statistics (2010)

Figure 85 Nationality statistics for the Singaporean population

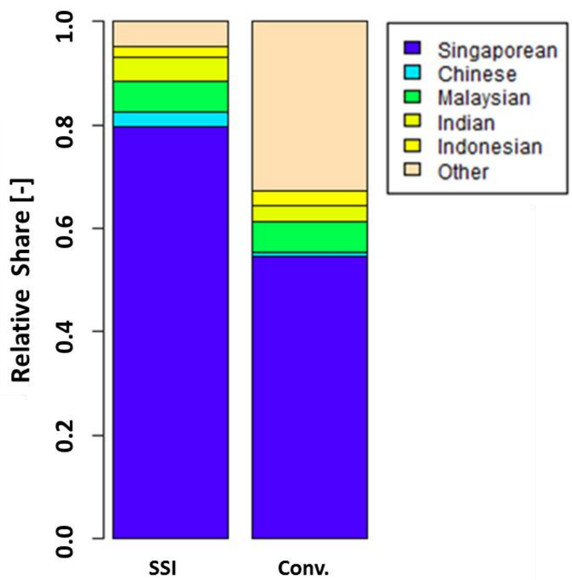


Figure 86 Trip frequency to CBD for the Singaporean population

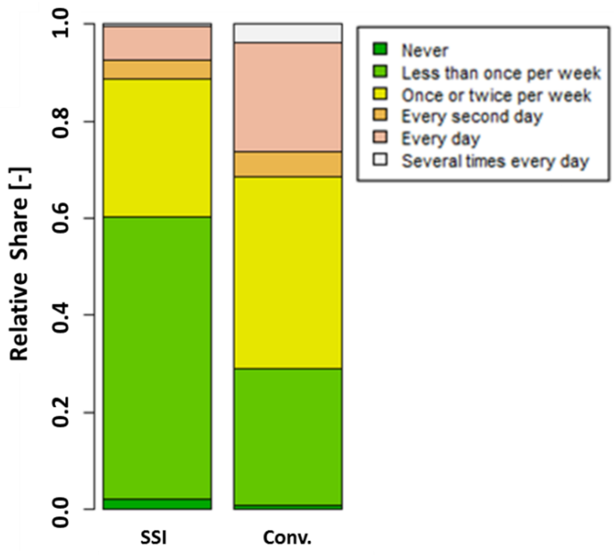
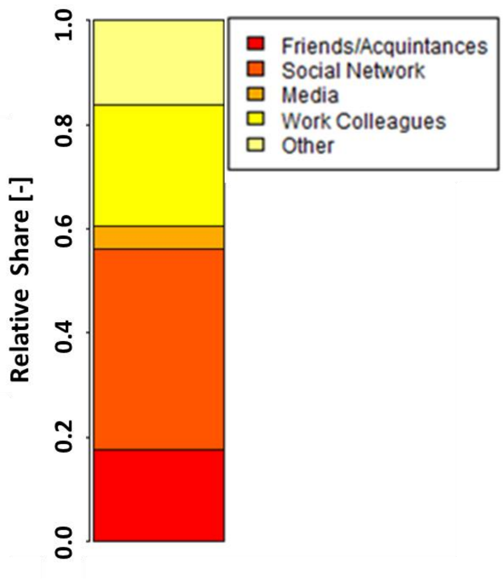


Figure 87 Source of the survey link for the convenience sample respondents



A 3 Discrete Choice Models

Utility Functions

Basic Model Survey Design Analysis (1)

```
asc_left +
b_time * (1 + b_time_text * survey_text) * walking_time +
b_major * type_major +
b_park * type_park +
b_inside * type_inside +
b_shops * shops +
b_greenery * greenery +
b_cover_sidewalk * cover_sidewalk +
b_1trafficlight * 1trafficlight +
b_2trafficlights * (1 + b_2trafficlights_text * survey_text) *
2trafficlights +
b_overpass_lift * overpass_lift +
b_overpass_stairs * overpass_stairs +
b_jaywalk_2lanes * jaywalk_2lanes +
b_jaywalk_4lanes * jaywalk_4lanes
```

Basic Model with All MTurk Data (2)

```
asc_left +
b_time * walking_time +
b_major * type_major +
b_park * type_park +
b_inside * type_inside +
b_shops * shops +
b_greenery * greenery +
b_cover_sidewalk * cover_sidewalk +
b_1trafficlight * 1trafficlight +
b_2trafficlights * 2trafficlights +
b_overpass_lift * overpass_lift +
b_overpass_stairs * overpass_stairs +
b_jaywalk_2lanes * jaywalk_2lanes +
b_jaywalk_4lanes * jaywalk_4lanes
```

Basic Model with MTurk Data Without Text Design (3)

```
asc_left +
b_time * walking_time +
b_major * type_major +
b_park * type_park +
b_inside * type_inside +
b_shops * shops +
b_greenery * greenery +
b_cover_sidewalk * cover_sidewalk +
b_1trafficlight * 1trafficlight +
b_2trafficlights * 2trafficlights +
b_overpass_lift * overpass_lift +
b_overpass_stairs * overpass_stairs +
b_jaywalk_2lanes * jaywalk_2lanes +
b_jaywalk_4lanes * jaywalk_4lanes
```

Basic Model Country of Origin (4)

```
asc_left +
b_time * walking_time +
b_major * type_major +
b_park * type_park +
b_inside * type_inside +
b_shops * shops +
b_greenery * greenery +
b_cover_sidewalk * cover_sidewalk +
b_1trafficlight * 1trafficlight +
b_2trafficlights * 2trafficlights +
b_overpass_lift * overpass_lift +
b_overpass_stairs * overpass_stairs +
b_jaywalk_2lanes * (1 + b_jaywalk_2lanes_india * india)
* jaywalk_2lanes +
b_jaywalk_4lanes * (1 + b_jaywalk_4lanes_india * india)
* jaywalk_4lanes
```

Basic Model with US MTurk Data only (5)

```
asc_left +
b_time * walking_time +
b_major * type_major +
b_park * type_park +
b_inside * type_inside +
b_shops * shops +
b_greenery * greenery +
b_cover_sidewalk * cover_sidewalk +
b_1trafficlight * 1trafficlight +
b_2trafficlights * 2trafficlights +
b_overpass_lift * overpass_lift +
b_overpass_stairs * overpass_stairs +
b_jaywalk_2lanes * jaywalk_2lanes +
b_jaywalk_4lanes * jaywalk_4lanes
```

Basic Model with Indian MTurk Data only (6)

```
asc_left +
b_time * walking_time +
b_major * type_major +
b_park * type_park +
b_inside * type_inside +
b_shops * shops +
b_greenery * greenery +
b_cover_sidewalk * cover_sidewalk +
b_1trafficlight * 1trafficlight +
b_2trafficlights * 2trafficlights +
b_overpass_lift * overpass_lift +
b_overpass_stairs * overpass_stairs +
b_jaywalk_2lanes * jaywalk_2lanes +
b_jaywalk_4lanes * jaywalk_4lanes
```

SSI Model with Walking Time Interaction (7)

```
asc_left +
b_time * (1 + b_major_time * type_major + b_park_time * type_park +
b_inside_time * (1 + b_inside_time_rain * rainy) * type_inside) * (1
+ b_cover_sidewalk_time * (1 + b_cover_sidewalk_time_rain * rainy) *
cover_sidewalk) * (1 + b_shops_time * shops) * (1 + b_tree_road_time
* tree_road) * walking_time +
b_1trafficlight * 1trafficlight +
b_2trafficlights * 2trafficlights +
b_overpass_lift * overpass_lift +
b_overpass_stairs * (1 + b_overpass_stairs_age50 * age50) * (1 +
b_overpass_stairs_female * female) * overpass_stairs +
b_jaywalk_2lanes * jaywalk_2lanes +
b_jaywalk_4lanes * jaywalk_4lanes
```

FCL Study's Model with Walking Time Interaction (8)

```
asc_left +
b_time * (1 + b_major_time * type_major + b_park_time * type_park +
b_inside_time * (1 + b_inside_time_rain * rainy) * type_inside) * (1
+ b_cover_sidewalk_time * (1 + b_cover_sidewalk_time_rain * rainy) *
cover_sidewalk) * (1 + b_shops_time * shops) * (1 + b_tree_road_time
* tree_road) * walking_time +
b_1trafficlight * 1trafficlight +
b_2trafficlights * 2trafficlights +
b_overpass_lift * overpass_lift +
b_overpass_stairs * (1 + b_overpass_stairs_age50 * age50) * (1 +
b_overpass_stairs_female * female) * overpass_stairs +
b_jaywalk_2lanes * jaywalk_2lanes +
b_jaywalk_4lanes * jaywalk_4lanes
```

SSI Model with Direct Effects (9)

```
asc_left +
b_time * walking_time +
b_major * type_major +
b_park * (1 + b_park_1pm * time_1pm) * type_park +
b_inside * (1 + b_inside_rain * rainy) * type_inside +
b_shops * (1 + b_shops_1pm * time_1pm) * shops +
b_tree_road * tree_road +
b_cover_sidewalk * (1 + b_cover_sidewalk_rain * rainy +
b_cover_sidewalk_sun * sunny) * cover_sidewalk +
b_1trafficlight * 1trafficlight +
b_2trafficlights * 2trafficlights +
b_overpass_lift * overpass_lift +
b_overpass_stairs * (1 + b_overpass_stairs_age50 * age50) * (1 +
b_overpass_stairs_female * female) * overpass_stairs +
b_jaywalk_2lanes * jaywalk_2lanes +
b_jaywalk_4lanes * jaywalk_4lanes
```

Convenience Model with Direct Effects (10)

```
asc_left +
b_time * walking_time +
b_major * type_major +
b_park * (1 + b_park_rain * rainy + b_park_sun * sunny) * type_park+
b_insideRain * type_insideRain +
b_shops * shops +
b_cover_sidewalk * (1 + b_cover_sidewalk_sun * sunny) * (1 +
b_cover_sidewalk_female * female) * cover_sidewalk +
b_1trafficlight * 1trafficlight +
b_2trafficlights * 2trafficlights +
b_overpass_stairs * (1 + b_overpass_stairs_age50 * age50) * (1 +
b_overpass_stairs_female * female) * overpass_stairs +
b_jaywalk_2lanes * jaywalk_2lanes +
b_jaywalk_4lanes * jaywalk_4lanes
```

MTurk Model with Direct Effects (11)

```
asc_left +
b_time * walking_time +
b_major * type_major +
b_park * (1 + b_park_rain * rainy + b_park_sun * sunny) * type_park+
b_inside * (1 + b_inside_sun * sunny) * type_inside +
b_cover_sidewalk * (1 + b_cover_sidewalk_sun * sunny) * cov-
er_sidewalk +
b_1trafficlight * 1trafficlight +
b_2trafficlights * 2trafficlights +
b_jaywalk_2lanes * (1 + b_jaywalk_2lanes_age50 * age50) * jay-
walk_2lanes +
b_jaywalk_4lanes * jaywalk_4lanes
```

Insignificant Attributes

Basic Model Survey Design Analysis (1)

Table 54 Insignificant attributes tested in the basic model of the survey design analysis

Attribute	Definition
b_1trafficlight_text	Interaction of b_1trafficlight if survey type is text only
b_cover_sidewalk_text	Interaction of b_cover_sidewalk if survey type is text only
b_greenery_text	Interaction of b_greenery if survey type is text only
b_inside_text	Interaction of b_inside if survey type is text only
b_jaywalk_2lanes_text	Interaction of b_jaywalk_2lanes if survey type is text only
b_jaywalk_4lanes_text	Interaction of b_jaywalk_4lanes if survey type is text only
b_major_text	Interaction of b_major if survey type is text only
b_overpass_lift_text	Interaction of b_overpass_lift if survey type is text only
b_overpass_stairs_text	Interaction of b_overpass_stairs if survey type is text only
b_park_text	Interaction of b_park if survey type is text only
b_shops_text	Interaction of b_shops if survey type is text only
b_1trafficlight_image	Interaction of b_1trafficlight if survey type is image only
b_2trafficlights_image	Interaction of b_2trafficlights if survey type is image only
b_cover_sidewalk_image	Interaction of b_cover_sidewalk if survey type is image only
b_greenery_image	Interaction of b_greenery if survey type is image only
b_inside_image	Interaction of b_inside if survey type is image only
b_jaywalk_2lanes_image	Interaction of b_jaywalk_2lanes if survey type is image only
b_jaywalk_4lanes_image	Interaction of b_jaywalk_4lanes if survey type is image only
b_major_image	Interaction of b_major if survey type is image only
b_overpass_lift_image	Interaction of b_overpass_lift if survey type is image only
b_overpass_stairs_image	Interaction of b_overpass_stairs if survey type is image only
b_park_image	Interaction of b_park if survey type is image only
b_shops_image	Interaction of b_shops if survey type is image only
b_time_image	Interaction of b_time if survey type is image only

Basic Model with Country of Origin (4)

Table 55 Insignificant attributes tested in the basic model of the analysis of the country of origin effects

Attribute	Definition
b_1trafficlight_india	Interaction of b_1trafficlight if respondent origin is India
b_2trafficlights_india	Interaction of b_2trafficlights if respondent origin is India
b_cover_sidewalk_india	Interaction of b_cover_sidewalk if respondent origin is India
b_greenery_india	Interaction of b_greenery if respondent origin is India
b_inside_india	Interaction of b_inside if respondent origin is India
b_major_india	Interaction of b_major if respondent origin is India
b_overpass_lift_india	Interaction of b_overpass_lift if respondent origin is India
b_overpass_stairs_india	Interaction of b_overpass_stairs if respondent origin is India
b_park_india	Interaction of b_park if if respondent origin is India
b_shops_india	Interaction of b_shops if respondent origin is India
b_time_india	Interaction of b_time if respondent origin is India

SSI Model with Walking Time Interaction (7)

Table 56 Insignificant attributes tested in SSI model with walking time interaction

Attribute	Definition
b_cover_sidewalk_time_1pm	Interaction of b_cover_sidewalk_time if daytime is 1 pm
b_cover_sidewalk_time_sun	Interaction of b_cover_sidewalk_time if sunny
b_inside_time_1pm	Interaction of b_inside_time if daytime is 1 pm
b_inside_time_age50	Interaction of b_inside_time if age greather or equal than 50 years
b_inside_time_sun	Interaction of b_inside_time if sunny
b_jaywalk_2lanes_age50	Interaction of b_jaywalk_2lanes if age greather or equal than 50 years
b_jaywalk_2lanes_female	Interaction of b_jaywalk_2lanes if female
b_jaywalk_4lanes_age50	Interaction of b_jaywalk_4lanes if age greather or equal than 50 years
b_jaywalk_4lanes_female	Interaction of b_jaywalk_4lanes if female
b_major_time_1pm	Interaction of b_major_time if daytime is 1 pm
b_major_time_rain	Interaction of b_major_time if rainy
b_major_time_sun	Interaction of b_major_time if sunny
b_overpass_stairs_1pm	Interaction of b_overpass_stairs if daytime is 1 pm
b_overpass_stairs_sunny	Interaction of b_overpass_stairs if sunny
b_park_time_1pm	Interaction of b_park_time if daytime is 7 pm
b_park_time_rain	Interaction of b_park_time if rainy
b_park_time_sun	Interaction of b_park_time if sunny
b_shops_time_7pm	Interaction of b_shops_time if daytime is 7 pm

SSI Model with Direct Effects (9)

Table 57	Insignificant attributes tested in the SSI model with direct effects

Attribute	Definition
b_inside_1pm	Interaction of b_inside if daytime is 1 pm
b_inside_age50	Interaction of b_inside if age greather or equal than 50 years
b_inside_sun	Interaction of b_inside if sunny
b_jaywalk_2lanes_age50	Interaction of b_jaywalk_2lanes if age greather or equal than 50 years
b_jaywalk_2lanes_female	Interaction of b_jaywalk_2lanes if female
b_jaywalk_4lanes_age50	Interaction of b_jaywalk_4lanes if age greather or equal than 50 years
b_jaywalk_4lanes_female	Interaction of b_jaywalk_4lanes if female
b_major_1pm	Interaction of b_major if daytime is 1 pm
b_major_rain	Interaction of b_major if rainy
b_major_sun	Interaction of b_major if sunny
b_overpass_stairs_1pm	Interaction of b_overpass_stairs if daytime is 1 pm
b_overpass_stairs_sunny	Interaction of b_overpass_stairs if sunny
b_park_rain	Interaction of b_park if rainy
b_park_sun	Interaction of b_park if sunny
b_shops_7pm	Interaction of b_shops is 7 pm

Convenience Model with Direct Effects (10)

Table 58	Insignificant attributes tested in the convenience model with direct effects

Attribute	Definition
b_cover_sidewalk_rain	Interaction of b_cover_sidewalk if rainy
b_inside_1pm	Interaction of b_inside if daytime is 1 pm
b_inside_age50	Interaction of b_inside if age greather or equal than 50 years
b_inside_sun	Interaction of b_inside if sunny
b_jaywalk_2lanes_age50	Interaction of b_jaywalk_2lanes if age greather or equal than 50 years
b_jaywalk_2lanes_female	Interaction of b_jaywalk_2lanes if female
b_jaywalk_4lanes_age50	Interaction of b_jaywalk_4lanes if age greather or equal than 50 years
b_jaywalk_4lanes_female	Interaction of b_jaywalk_4lanes if female
b_major_1pm	Interaction of b_major if daytime is 1 pm
b_major_rain	Interaction of b_major if rainy
b_major_sun	Interaction of b_major if sunny
b_overpass_stairs_1pm	Interaction of b_overpass_stairs if daytime is 1 pm
b_overpass_stairs_sunny	Interaction of b_overpass_stairs if sunny
b_park_1pm	Interaction of b_park of daytime is 1 pm
b_shops_7pm	Interaction of b_shops is 7 pm

MTurk Model with Direct Effects (11)

Table 59 Insignificant attributes tested in the MTurk model with direct effects

Attribute	Definition
b_cover_sidewalk_rain	Interaction of b_cover_sidewalk if rainy
b_inside_1pm	Interaction of b_inside if daytime is 1 pm
b_inside_age50	Interaction of b_inside if age greather or equal than 50 years
b_inside_rain	Interaction of b_inside if rainy
b_jaywalk_2lanes_age50	Interaction of b_jaywalk_2lanes if age greather or equal than 50 years
b_jaywalk_2lanes_female	Interaction of b_jaywalk_2lanes if female
b_jaywalk_4lanes_age50	Interaction of b_jaywalk_4lanes if age greather or equal than 50 years
b_jaywalk_4lanes_female	Interaction of b_jaywalk_4lanes if female
b_major_1pm	Interaction of b_major if daytime is 1 pm
b_major_rain	Interaction of b_major if rainy
b_major_sun	Interaction of b_major if sunny
b_overpass_lift	Overpass with lift
b_overpass_stairs	Overpass with stairs
b_park_1pm	Correction of b_park of daytime is 1 pm
b_shops	Shopping amenities
b_tree_road	Tree presence on road

Willingness to avoid

Table 60 Values of the computed willingness to avoid with statistics for the SSI model with walking time interaction

Willingness to avoid	Value	SE robust	p-value robust
1 trafficlight	1.22	0.35	0.00
2 trafficlights	2.55	0.39	0.00
Jaywalk 4 lanes	5.68	0.72	0.00
Overpass with lift	1.79	0.70	0.01
Overpass with stairs	4.50	0.89	0.00

Table 61 Values of the computed willingness to avoid with statistics for the FCL study's model with walking time interaction

Willingness to avoid	Value	SE robust	p-value robust
2 trafficlights	0.91	0.23	0.00
Jaywalk 2 lanes	1.41	0.53	0.01
Jaywalk 4 lanes	3.63	0.50	0.00
Overpass with lift	1.57	0.50	0.00
Overpass with stairs	2.86	0.43	0.00

Table 62 Values of the computed willingness to avoid with statistics for the SSI model with direct effects

Willingness to avoid	Value	SE robust	p-value robust
1 trafficlight	1.43	0.43	0.00
2 trafficlights	2.49	0.51	0.00
Covered sidewalk	-2.91	0.61	0.00
Jaywalk 4 lanes	6.87	0.88	0.00
Major road	0.77	0.32	0.02
Overpass with lift	2.06	0.85	0.02
Overpass with stairs	5.36	1.04	0.00
Park	-3.95	0.81	0.00
Shops	-1.70	0.40	0.00
Trees	-0.90	0.36	0.00
Inside building	-2.42	0.67	0.00

Table 63 Values of the computed willingness to avoid with statistics for the convenience model with direct effects

Willingness to avoid	Value	SE robust	p-value robust
1 trafficlight	6.81	-	-
2 trafficlights	9.85	-	-
Covered sidewalk	-13.89	-	-
Inside building if rainy	-9.88	-	-
Jaywalk 2 anes	9.17	-	-
Jaywalk 4 lanes	11.03	-	-
Overpass with stairs	24.76	-	-
Park	-33.98	-	-
Shops	-4.33	-	-

Table 64 Values of the computed willingness to avoid with statistics for the MTurk model with direct effects

Willingness to avoid	Value	SE robust	p-value robust
1 trafficlight	4.10	1.12	0.00
2 trafficlights	5.62	1.28	0.00
Covered sidewalk	-3.72	1.29	0.00
Inside building	-2.85	1.15	0.01
Jaywalk 2 lanes	5.58	1.85	0.00
Jaywalk 4 lanes	7.58	2.27	0.00
Park	-17.30	4.11	0.00