Investigating Determinants in the Acceptance of Automated Taxis: Evidence from Online Screen-based and Virtual Reality-based Stated Choice Experiments

SCHOOL OF ENGINEERING
NEWCASTLE UNIVERSITY

By
Hao Yin

A thesis submitted for the degree of
Doctor of Philosophy in Transport at
Newcastle University

July 2022

Declaration

This thesis is submitted to Newcastle University for the degree of Doctor of Philosophy. I declare that this thesis and the original work presented in it are my own and have been generated by me. The material presented has never been submitted to any other educational establishment for the purpose of obtaining a higher degree.

July 2022 Hao Yin

Abstract

The introduction of Automated Taxis (ATs) has the potential to bring economic, environmental, and social benefits to the current transport system. The success of ATs, as a nascent technology, crucially depends on whether potential users adopt and accept to use them and what determinants will affect their acceptance of ATs. Despite relative novelty of the topic, there is already a relatively vast literature on the factors influencing the choice of owning and/or using general Autonomous Vehicles (AVs), but only few papers have dealt with Automated Taxis (ATs). Due to the absence of an AT real market to obtain Revealed Preference (RP) data, Stated Choice (SC) methods are typically an appropriate way to elicit travellers' preferences. However, even though customised and careful designed, a major problem implicit in the SC experiments is the 'lack of realism', leading to the well-known hypothetical bias. This problem is particularly relevant and predominant when SC experiments are used to study innovative products, because, as amply discussed in the context of electric vehicles, respondents lack of experience with the product and often-insufficient knowledge has a significant impact in the preferences elicited with SC experiments.

In line with this discussion, the aim of this PhD research consists in providing empirical, theoretical and methodological evidence to contribute to the state of the art to understand, measure and quantify the determinants in the acceptance of ATs and it proposes a methodology to implement SC experiments in immersive virtual reality (VR) environments. Four datasets are collected using the same instrument but applied online and embedded on the VR. Online data were collected in the UK and China, allowing for a comparison of the impacts across nations. Other two sets of data were collected among respondents in Newcastle using the online survey for one set of data and the VR for another set of data. A comparison between these datasets allows testing the impact of VR on the preferences elicited. Finally the impact of living in a city where AT systems do exist was also explored using the data collected in China.

Many interesting results were found. For example UK respondents are willing to pay on average 5 times and 10 times as much as the WTPs of Chinese respondents to save one hour of travel time and are willing to pay on average between 2-11 times more than Chinese participants to save one hour of waiting time. On the other hand, no significant differences were found in the preference for in-vehicle features. Overall, results showed that the Chinese AT market is less elastic than the UK market to changes in level of services characteristics, in-vehicle futures and social conformity measures, with the exception of some population segments. Interestingly, the British AT market share is still significantly affected by the latent psychological construct Hedonic Motivations. As for the VR impact, perhaps the most interesting result, however, is that the attribute to measure AT adoption rate, which is a very problematic attribute in the online SC experiments (as proved in all researches conducted for other innovations), was significant when measured in the VR environment. Results need to be confirmed by further evidence, but are promising that VR might help achieve better results in particular to measure social conformity effects.

The PhD research provides an extensive discussion of the willingness to pay measure for all the attributes tested and a comparison among the four datasets collected as well as with the values reported in the international literature. It provides also methodological guidelines on how to build SC experiments embedded into immersive VR settings and how to empirically carry out the experiments. Methodology-related, policy-related implications, and limitations of the current study are also discussed.

Acknowledgment

Four years of PhD studying at Newcastle University are coming to an end. The years of life in Newcastle are full of ups and downs, but more harvest, growth and happiness. I would like to express my gratitude to all those who helped, accompanied and guided me during my PhD study over the past years.

First and foremost, I deeply appreciate my supervisor Prof. Elisabetta Cherchi and her help, inspiration and encouragement in the study and research throughout the years. A special thanks to her for providing me with constructive feedbacks and suggestions, and for being patient to correct thesis chapters, reports and papers in details. Her constructive advices and criticism indeed not just help me in PhD study but also guide me in future academic life.

I also gratefully acknowledge the help of my second supervisor Prof. Daniel Nettle. I wish to express my sincere and warm thanks for providing me with his helps and guidance in PhD research in general and in psychology-related knowledge in particular.

In addition, I do appreciate my schoolmates and friends in Newcastle University as well. Thank all of you for helping me overcome setbacks, and supporting and encouraging me during my depression period.

Particularly, I am thankful for my loving wife Tianle Ke for her support, patience and company during this journey of doctoral study. Without her encouragement and support during the special period of covid pandemic, I cannot smoothly complete my doctoral study.

At last, I would like to thank my lovely family for their financial and spiritual support throughout these years.

Table of Content

Abstract ii

Acknowledgment iii

List of Tables vii

List of Figures x

List of Abbreviations xi

Chapter 1 Introduction 1

1.1 Background 1
1.2 Research aim and objectives 5
1.3 Research Question 5
1.4 Thesis outline 6
1.5 Thesis main innovations and contributions 8
Chapter 2 Literature Review 11

2.1 Introduction 11
2.2 Factors influencing the choice to use AVs and taxi services 11
2.2.1 Service and vehicle characteristics 12
2.2.2 Individual-related characteristics 17
2.2.3 Attitudinal characteristics 20
2.2.4 Attitudinal characteristics in the primary appraisal stage 21
2.2.5 Cross-national comparison 29
2.3 Survey techniques to measure choice/use of AVs and taxi services 30
2.3.1 Service and vehicle characteristics 30
2.3.2 Psychological constructs 32
2.3.3 Other survey methods 33
2.4 Virtual Reality 34
2.4.1 General literature on virtual reality 34
2.4.2 Virtual Reality in Transport 35
2.5 Summary 37
Chapter 3 Data Collection Methodology 38

3.1 Introduction 38
3.1.1 Rationale for the data collection methodology 40
3.2 Context of reference 42
3.2.1 Selection of the context in Newcastle upon Tyne 42
3.3 Focus Group 45
3.3.1 FG survey methodology 46
3.3.2 FG main results 48
3.4 Pre-SC information 51
3.4.1 AT safety information 51
3.4.2 Personal safety, privacy and routing information 53
3.5 Stated Choice experiment 55
3.5.1 Attributes definition 56
3.5.2 Property of the design 62

3.6 Attitudinal questions 66
3.7 VR-based SC Experiment in Newcastle (UK) 67
3.7.1 Constructing virtual reality environment 67
3.7.2 Constructing the SC experiment embedded in the VR environment 71
3.8 Additional information 73
3.8.1 Section 1: Introduction to the survey 73
3.8.2 Part 4: Individual information 74
3.9 Sample recruitment 74
3.9.1 VR-based SC survey protocol 77
3.10 Summary 79
Chapter 4. Descriptive Data Analysis 81
4.1 Introduction 81
4.2 Data from the *Survey-online UK* and the *Survey-online China* 81
4.2.1 Socio-demographic characteristics 81
4.2.2 Trip characteristics 85
4.2.3 Knowledge levels of AVs and ATs 88
4.2.4 Psychological statements 90
4.2.5 Stated Choices 92
4.3 Data from the *Survey-online NCL* and the *Survey-VR NCL* 93
4.3.1 Socio-demographic characteristics 94
4.3.2 Trip characteristics 95
4.3.3 Knowledge levels of AVs or ATs 97
4.3.4 Stated choice and other VR questions 98
4.4 Data from the *Survey-online China – AT existence* 101
4.4.1 Socio-demographic characteristics 101
4.4.2 Trip characteristics 103
4.4.3 Knowledge of AVs and ATs 104
4.4.4 Psychological statements 104
4.4.5 Stated Choices 105
4.5 Summary 106
Chapter 5 Modelling Methodology 108
5.1 Introduction 108
5.2 Theoretical foundation of DCM 108
5.2.1 Random utility-based theory 108
5.2.2 Operationalisation of the random utility-based theory 109
5.3 Mixed Logit model 110
5.4 Hybrid Choice model 112
5.5 Joint estimation 114
5.6 Building DCMs 115
5.6.1 Utility specification in the Mixed Logit models 115
5.6.2 Utility specification Hybrid Choice Models 116
5.6.3 Model estimation 117
5.6.4 Model evaluation 118
5.6.5 Marginal Utilities and Willingness to pay 119
5.7 Summary 121
Chapter 6 Modelling Estimations & Discussions 122
6.1 Introduction 122

6.2 Survey-online UK and the Survey-online China 123
6.2.1 Impact of level of services attributes 127
6.2.2 Impact of in-vehicle features 128
6.2.3 Impact of normative conformity 129
6.2.4 Knowledge of AVs and ATs 130
6.2.5 Other effects 131
6.2.6 Latent psychological constructs 132
6.2.7 Willingness to pay 137
6.3 Survey-online NCL and the Survey-VR NCL.. 141
6.3.1 Impact of the level-of-services attributes... 144
6.3.2 Impact of in-vehicle features 146
6.3.3 Impact of normative conformity 146
6.3.4 Impact of knowledge of AVs and ATs... 147
6.3.5 Other effects 147
6.3.6 Willingness to pay 148
6.4 Survey-online China – Impact of AT existence 151
6.4.1 Impact of the level-of-services attributes... 153
6.4.2 Impact of in-vehicle features 153
6.4.3 Impact of normative conformity 154
6.4.4 Impact of knowledge of AVs and ATs... 154
6.4.5 Other effects 155
6.4.6 Latent psychological constructs 155
6.4.7 Willingness to pay 156
6.5 Policy Implications 157
6.5.1 Scenario N1: variations in the LOS attributes... 159
6.5.2 Scenario N2: variations in the in-vehicle features... 161
6.5.3 Scenario N3: improvements in the customer reviews 161
6.5.4 Scenario N4: improvements in the latent variables 162
6.6 Summary 164
Chapter 7 Conclusion 166
7.1 Main findings and contributions... 166
7.2 Methodological recommendations... 169
7.3 Policy recommendations... 170
7.4 Limitations and future research.. 172
Appendices 176
Appendix 1 – Focus Groups.. 176
Appendix 2 – Pilot Tests 180
Appendix 3 – VR Urban Layout... 190
Appendix 4 – Descriptive Data Analysis 194
Appendix 5 – Modelling Results 198
References 201

List of Tables

Table 2-1 Summary of service/vehicle characteristics studied to explain the choice of using AVs....14

Table 2-2 Summary of factors studied to explain the perceived taxi service quality........................15

Table 2-3 Summary of individual-related characteristics and travel habit studied to explain AV choices...19

Table 2-4 summary of latent constructs affecting behaviour intention to use AVs..................................22

Table 2-5 Summary of social conformity effects and measurement methods..................................25

Table 2-5 (Cont'd) Summary of social conformity effects and measurement methods....................27

Table 2-6 Summary of survey techniques used for studying service/vehicle characteristics..........31

Table 2-7 Summary of survey techniques used for psychological constructs.....................................33

Table 2-8 Summary of other survey techniques...34

Table 3-1 Pre-identified key factors related to the taxis, AVs and social conformity.........................45

Table 3-2 questions for defining car safety...53

Table 3-3 SC Attributes and attributes levels in the UK and in China...64

Table 3-4 summary of four surveys for data collection...77

Table 4-1 Socio-demographic characteristics- Survey-online UK & China 82

Table 4-2 Age distribution by employment status- Survey-online UK & China.................................83

Table 4-3 Income scales- Survey-online UK & China..84

Table 4-4 Trip characteristics- Survey-online UK &China...85

Table 4-5 Age distribution by trip segment- Survey-online UK & China..87

Table 4-6 Trip purpose distribution by employment status- Survey-online UK & China......................88

Table 4-7 Frequency distribution of taxi usage by trip segment- Survey-online UK & China.............88

Table 4-8 Knowledge levels of AVs and ATs- Survey-online UK & China...89

Table 4-9 Knowledge of AVs or ATs distribution by gender- Survey-online UK & China...................90

Table 4-10 Psychological statements- Survey-online UK & China...90

Table 4-11 Psychological statements by age- Survey-online UK & China..91

Table 4-12 Psychological statements by familiarity of automation levels- Survey-online UK & China 92

Table 4-13 Stated Choices by trip distance segment- Survey-online UK & China................................92

Table 4-14 Stated Choices classified by age and gender- Survey-online UK & China........................93
Table 4-15 Stated Choices Stated Choices classified if like driver carrying luggage and knowledge level on AVs and ATs- Survey-online UK & China 93
Table 4-16 Socio-demographic characteristics – Survey online NCL & Survey VR NCL.......................95
Table 4-17 Trip characteristics – Survey online NCL & Survey VR NCL..96
Table 4-18 Knowledge levels of AVs and ATs – Survey online NCL & Survey VR NCL...........................97
Table 4-19 Stated Choices – Survey online NCL & Survey VR NCL..98
Table 4-20 Stated Choices classified by age and gender – Survey online NCL & Survey VR NCL......99
Table 4-21 Stated Choices classified by if like driver carrying luggage and knowledge level on AVs and ATs – Survey online NCL & Survey VR NCL 99
Table 4-22 Socio-demographic characteristics – Survey online China...102
Table 4-23 Trip characteristics– Survey online China...103
Table 4-24 Knowledge of AVs and ATs – Survey online China 104
Table 4-25 Psychological statements – Survey online China...105
Table 4-26 Stated Choices – Survey online China...106
Table 4-27 Stated Choices classified by frequency of taxi usage – Survey online China.................106
Table 6-1 Systematic heterogeneity & Interactions effects- Survey-online UK & China 124
Table 6-2 Marginal utilities- Survey-online UK & China...126
Table 6-3 Latent psychological constructs- Survey-online UK & China..135
Table 6-3 (continued) Latent psychological constructs- Survey-online UK & China..........................136
Table 6-4 WTP mean values- Survey-online UK & China...137
Table 6-5 WTP t-test and confidence intervals- Survey-online UK & China.....................................138
Table 6-6 Willingness to pay for travel time – from the literature 140
Table 6-7 Minutes of travel time that equates WTP for in-vehicle characteristics............................140
Table 6-8 Minutes of travel time that equates WTP for good reviews...141
Table 6-9 model estimation results – Survey-online NCL and the Survey-VR NCL...........................143
Table 6-10 Marginal utilities – Survey-online NCL &survey-VR NCL...144
Table 6-11 WTP mean values and t-tests – Survey-online NCL & Survey-VR NCL...........................148
Table 6-12 Minutes of travel time that equates WTP for in-vehicle characteristics........................150

Table 6-13 AT-existence – Survey online China..152
Table 6-13 (continued) AT-existence – Survey online China...153
Table 6-14 Willingness to pay mean values, t-tests and confidence levels – AT-existence China...156
Table 6-15 Minutes of travel time that equates WTP for change destination and taxi with good reviews...157
Table 6-16 Mean values of attribute levels in the no-policy scenario- UK & China 158
Table 6-17 Aggregate choice probability in the base scenario for different segments.....................159
Table 6-18 Market share variations for changes in level of services...160
Table 6-19 Market share variations for changes in the in-vehicle characteristics............................161
Table 6-20 Market share variations for changes in the Customers Reviews.....................................161
Table 6-21 Market share variations for changes in the latent variables...162
Table 6-22 Marginal rates of substitution of the three latent variables with respect to travel time 163

List of Figures

Figure 3.1 Process of development of four surveys……………………………………………………………40
Figure 3.2 Location of some of the taxi ranks considered in Newcastle…………………………………44
Figure 3.3 Street view of Northumberland Road in Google Maps…………………………………………44
Figure 3.4 In-vehicle communication forms with the AT operator…………………………………………49
Figure 3.5 a) Car safety information in the UK ……………………………………………………………52
Figure 3.5 b) Car safety information in China (translated from Chinese)………………………………52
Figure 3.6 Privacy and routing information…………………………………………………………………54
Figure 3.7 a) How to use ATs…………………………………………………………………………………55
Figure 3.7 b) What to do within the ATs……………………………………………………………………55
Figure 3.8 Example of choice task presented…………………………………………………………………56
Figure 3.9 In vehicle attributes in the SC scenarios………………………………………………………59
Figure 3.10 Presentation formats of payment options ……………………………………………………61
Figure 3.11 Presentation formats of payment instructions………………………………………………62
Figure 3.12 Destination list in the SC experiment in Newcastle (UK)……………………………………63
Figure 3.13 The re-design of taxi rank area: drawing and VR environment built………………………68
Figure 3.14 Ticket Board in the immersive VR-based SC experiment……………………………………70
Figure 3.15 Information board in the immersive VR-based SC experiment……………………………71
Figure 3.16 Destination list presented to respondents in VR SC experiment…………………………72
Figure 3.17 Example of SC choice task in the VR environment…………………………………………72
Figure 3.18 Textual and pictorial description of reference context………………………………………74
Figure 4.1 Income distribution by trip segments- Survey-online UK……………………………………86
Figure 4.2 Income distribution by trip segments- Survey-online China…………………………………87
Figure 4.3 Previous VR experience Survey-VR NCL………………………………………………………100
Figure 4.4 Reality of VR experiment Survey-VR NCL……………………………………………………100
Figure 4.5 Respondents noticing ATs and NTs colours and passengers queuing Survey-VR NCL……101
Figure 5.1 Framework of hybrid choice model………………………………………………………………113
Figure 6.1 AT market share variations for different segments……………………………………………160

List of Abbreviations

APT	Automated Public Transport
ARTS	Automated Road Transport System
AS	Autonomous Shuttle
AT	Automated Taxi
AV	Autonomous Vehicle
BL	Binary Logit
CaaP	Autonomous vehicle-as-a-Product
CaaS	Autonomous vehicle-as-a-Service
CNY	Chinese Yuan
DCM	Discrete Choice Model
DfT	Department for Transport
DOI	Diffusion of Innovation
EE (PEoU)	Effort Expectancy (Perceived Ease of Use)
EV	Electric Vehicle
FC (PBC)	Facilitating Condition (Perceived Behavioural Control)
FG	Focus Group
GBP	British Pound Sterling
HCM	Hybrid Choice Model
HM	Hedonic Motivation
IN	Injunctive Norm
Knw	Knowledge
LV	Latent Variable
ML	Mixed Logit
MNL	Multinomial Logit
MRS	Marginal Rate of Substituion
MS	Market Share
MSL	Maximum Simulated Likelihood
MU	Marginal Utility
NCL	Newcastle upon Tyne
NT	Normal Taxi
PAV	Private Autonomous Vehicle
PE (PU)	Performance Expectancy (Perceived Usefulness)
PPP	Purchase Power Parity

PR	Perceived Risk
PS	Perceived Safety
PT	Public Transport
RP	Revealed Preference
SAV	Shared Autonomous Vehicle
SC	Stated Choice
SI (SN)	Social Influence (Subjective Norm)
SP	Stated Preference
T	Trust
TAM	Technology Acceptance Model
TPB	Theory Of Planned Behaviour
UTUAT	Unified Theory of Use and Acceptance of Technology
VR	Virtual Reality
WTP	Willingness to Pay

Publications and Conferences

Publications

Yin, H. and Cherchi, E. 2022. Willingness to pay for automated taxis: a stated choice experiment to measure the impact of in-vehicle features and customer reviews. *Transportation*, pp. 1-22 https://doi.org/10.1007/s11116-022-10319-3

Yin, H., Cherchi, E. and Nettle, D. 2023. Living in a city where Automated Taxis are operating: Does this affect consumers' preferences? *Transportation Research Record*, 0(0). https://doi.org/10.1177/03611981231155181

Yin, H. and Cherchi, E. 2023. Conducting Stated Choice Experiments within a Virtual Reality environment: an application to the choice of automated taxi. *Transportation Research Procedia*. (Accepted)

Yin, H. and Cherchi, E. 2023. Heterogeneity in the willingness to pay for Automated Taxis: A cross-national comparative study between the UK and China. *Transportation Behaviour and Society* (In preparation)

Conferences

Yin, H. and Cherchi, E. 2022. A stated choice experiment embedded in immersive virtual reality to estimate preference for fully automated taxis. 16th International Conference on Travel Behaviour Research Chile.

Yin, H and Cherchi, E. 2022. The impacts of in-vehicle features and social influence on the choice of autonomous taxis. 101st Seminar on Transportation Research Board, Washington DC, USA

Yin, H. and Cherchi, E. 2022. Conducting Stated Choice Experiments within a Virtual Reality environment: an application to the choice of automated taxi 12th International Conference on Transport Survey Methods. Travel Survey and Big Data: how to make the best of both worlds, Portugal.

Yin, H., Cherchi, E. and Nettle, D. 2023**.** Living in a city where Automated Taxis are operating: Does this affect consumers' preferences? 102nd Seminar on Transportation Research Board. Washington DC, USA.

Chapter 1 Introduction

1.1 Background

With the rapid development of driving technologies (wireless communications, sensors, vehicle-to-infrastructure and vehicle-to-vehicle recognition systems communication, mapping and navigational technology), Autonomous Vehicles (AVs, also known as self-driving vehicles, driverless vehicles, automated vehicles, or robotic vehicles) are leading to a revolutionary change of the current transport system (Fagnant and Kockelman, 2014; Fagnant and Kockelman, 2015; Greenblatt and Shaheen, 2015). In general, AVs means highly or fully automated vehicles, which require no manual steering (Level4 or Level5)[1] (SAE, 2018). The application of AVs may become a reality in the future, and as stressed by many researchers (e.g. Fagnant and Kockelman, 2014; Zhang *et al.*, 2015; Mena-Oreja *et al.*, 2018), in particular Automated Taxis (ATs) have the potential to become a market player in the transport system. The global AT market is expected to be equivalent to $1.03 billion in 2023, and is projected to reach $38.61 billion by 2030, of which $16.89 billion is estimated to account for passenger transportation (Abhay *et al.*, 2020). The AT market is a rapidly growing and evolving field, where a number of high-tech companies and IT giants over the world are developing AT services to create a new urban travel paradigm. In the United Arab Emirates, Txai began piloting an AT project, with a fleet of five vehicles with a safety officer on board to serve passengers at nine stops (Arabian Business, 2021). In the US, Waymo have launched the first commercial autonomous ride-hailing service (Waymo, 2020). In China, WeRide, Baidu Apollo, DiDi Chuxing and AutoX have commenced in 2020 the pilot AT service project open to the public in some metropolitan cities (Changsha Evening News, 2020; DiDi, 2020; WeRide, 2020; XINHUANET, 2020). In the UK, ServCity completed in 2023 the pilot testing of an AT service on the streets of London, although the service is not yet open and available to the public (Transport Research Laboratory, 2023).[2] Furthermore, automotive manufacturing giants like Ford, Cruise and Tesla are also promoting the commercialisation of AT services, which are expected to become among the most promising business models for AVs (Saeed *et al.*, 2020).

[1] According to classification of Society of Automotive Engineers (SAE), there are six levels in regards to vehicle automation, from Level 0 (no automation) to Level 5 (full automation). Level 0 represents no automation; Level 1, Level 2 and Level 3 represent low, partly and conditional automation respectively; While Level 4 and Level 5 represent high and full automation: The vehicle is capable of performing all driving functions under certain conditions (Level 4) or all conditions (Level 5).

[2] The majority of these services were not available when this dissertation started.

AT services are on-demand mobility services with the potentials to improve growing externalities of road traffic in modern cities (Nicolaides *et al.*, 2017; Walker and Marchau, 2017), substantially enhance environments, reducing requirement of parking space and increasing road capacity. Fagnant and Kockelman (2014) indicated that reductions of varying degrees for energy use and pollutant emission (e.g. carbon monoxide, volatile organic compounds and greenhouse gas etc.) are anticipated during the lift-cycle period if 3.5% of human trip would be served by ATs. Zhang *et al.* (2015) estimated a sharp reduction up to 90% for parking space at only 2% penetration rate of ATs. Mena-Oreja *et al.* (2018) observed the road capacity might increase by 9.39% to 39.21%, at 100% penetration rate of ATs.

Nevertheless, if ATs will be not adopted, as simulation studies assume they will be (all possible impacts are based on different assumptions on the levels of penetration rates), it will be less likely to have all the estimated benefits that ATs would bring to current transport systems (for example, according to Mena-Oreja *et al.* (2018), the impact of ATs on road capacity turns to be positive until the penetration rate reaches an appropriate threshold value). The potential benefits of the ATs on current transport systems crucially depend on how successfully ATs will be in attracting customers. A low level of public acceptance possibly becomes a significant obstacle to the commercialisation of the ATs and the realisation of the potential benefits. This involves several questions that have not received yet a clear answer in the literature.

Even though ATs have been tested and deployed in a relatively small scale around the world, as an emerging and highly innovative new mode of local passenger transport, are still in the trial stage and related taxi services are not available for the most public at the current stage. Consumers do not have much knowledge of ATs (and AVs in general) and have never (or very rarely in few trials) experienced them personally. As such, they have no preferences yet for this technology and this poses the question about the role of knowledge and experience (direct and indirect) in shaping consumers preferences and how to measure preferences for new products.

Stated Choice (SC) methods are the most established way to elicit users' preferences for products that are not available in the real market, or are available but with significantly different characteristics. Due to the absence of ATs' consumers to obtain revealed preference data, SC methods are an appropriate way to gauge AT travellers' preferences. Nonetheless, even if customised and careful designs, SC experiment are prone to hypothetical bias, due to the fact that hypothetical situations provide a stimulus in the consumers that is different from that in real contexts. Studies in neuroscience found that more functions are involved in a real choice setting and most of the stimuli used to study human behaviour are not likely to evoke all the functional human responses that are present in natural contexts (see discussion in Cherchi, 2020). The lack of realism is a predominant issue implicit in all SC experiments, which could result in direct or cross elasticity, willingness to pay (WTP) measures derived from these data could be somewhat problematic and unreliable (Cherchi and Hensher, 2015).

Mitigation techniques, such as certainty calibration[3] and cheap talk[4] (Fifer *et al.*, 2014) and in particular the use of picture (Sillano and de Dios Ortúzar, 2005; Hensher and Mulley, 2015; Sottile *et al.*, 2015), are typically employed to compensate hypothetical bias in SC experiments. However, even if images are used, SC experiments carried in the conventional text-based format, which cannot convey the stimulus of the real experience and the understanding of the innovative technologies (e.g. ATs) is still based on individual's mental image.

In this line, recent applications have also used Virtual Reality (VR) techniques, which have recently become more available and affordable for scientific research purposes. VR techniques offer close-to-realistic, immersive and interactive environments, which greatly improve the degree of realism in hypothetical scenarios, but also it can reduce the respondents' mental or cognitive burdens for processing complicated description expressed textually, which might considerably alleviate error variance caused by participants' poor mental imagination. VR techniques are an interesting and promising new area of research, but still at its infancy and significant work is still needed to understand potentiality, limits and whether the closer to reality VR environment allows for better estimation of consumers preferences.

A number of recent applications have used immersive VR technology with stated preference experiments (e.g. Farooq *et al.*, 2018; Sobhani and Farooq, 2018; Birenboim *et al.*, 2019; Arellana *et al.,* 2020) to account for attributes that are difficult to measure in a standard screen-based survey and to control for the characteristics of the environment. These studies are applied to pedestrian and cycling behaviour, which involve a continuous movement and this makes it difficult to assess internal validity, i.e. the impact of the VR experience in the elicited consumer

---

3 Following the SC choice tasks, respondents are asked how certain they are about their preferred option.

4 A text script is shown to respondents prior to completing an experiment. The script emphasises the importance of the respondent's answers.

preferences, compared to a traditional stated choice screen-based survey. Few papers have compared text-based and VR-based SC experiment (Patterson *et al.*, 2017; Arellana *et al*., 2020; Rossetti and Hurtubia, 2020; Mokas *et al.*, 2021) and results are not consistent. Some results suggest no significant differences between text-based and VR-based, some other found an impact on the WTP estimates, some others found that VR data allows for efficient estimation results than the traditional format. More evidence is needed and none of these works focuses on ATs.

Extensive research in psychology and social science has provided significant evidence that psychological and social factors play a key role in explaining acceptance of innovations and technologies and can hinder or facilitate their acceptance. There is also a relatively vast literature specifically on AVs (see the literature reviews by Becker and Axhausen, 2017; Gkartzonikas and Gkritza, 2019; Narayanan *et al*., 2020), but this has mostly focused on using and/or owning AVs in general. Less attention has been paid instead on the use of ATs. Given their innovative nature, it is likely that psychological and social factors will have an important role in the acceptance of ATs. In particular, according to the Diffusion of Innovation theory (Rogers, 1962), innovative products diffuse over time through a social system, starting from early adopters, i.e. people who behave differently from what they did previously, and are then followed by other people who imitate the early adopters. And research conducted in a similar innovation (electric vehicles) has shown that social influence represents a strong predictor of the adoption of this kind of innovations. At the same time, since ATs are completely controlled by intelligent systems trust in this autonomous technology is vital for their adoption, in particular in the early stage, among the early adopters. Additionally, while it has been found (Venkatesh *et al*., 2012) that using a technology is typically considered fun and pleasant (strong positive effect of Hedonic Motivation), other research (Kyriakidis *et al.* (2015); Nordhoff *et al.* (2019), argue that the increase in the automation will make using car (i.e. AVs) less enjoyable. This is probably based on the assumption that people like to talk with the driver, though it might not be true for all passengers. The role of hedonic motivation is then not clear, and hence interesting to be explored.

Last, but not least, another unexplored and interesting topic is whether individuals across different regions or countries have different preferences in ATs and their characteristics. Several papers on AVs have been applied to many different countries but results are often not comparable and hence no transferable. Some studies have focused on the differences in the determinants or barriers across countries and provided some descriptive statistic results (Schoettle and Sivak, 2014; Kyriakidis *et al.*, 2015).

1.2 Research aim and objectives

The aim of this PhD research consists in providing theoretical, methodological and empirical, evidence to contribute to the state of the art in understanding and measuring the determinants in the acceptance of automated taxis (ATs) versus normal taxis (NTs). Given the research gaps identified, this dissertation sets the following specific objectives:

1. to investigate the key determinants affecting the choice of ATs versus NTs, focusing in particular into the role of in-vehicle features, normative conformity and the knowledge of AVs and ATs, in addition to the level of services attributes;
2. to identify appropriate ways to present and measure level of services attributes, in-vehicle features, normative conformity within SC experiment;
3. to propose and develop a novel methodology to implement SC experiments into immersive Virtual Reality environments, in order to improve low realism implicit in traditional online screen-based SC experiments;
4. to investigate the role of key latent psychological constructs into the choice of AT, focusing in particular into the role of knowledge of AVs and ATs in explaining these psychological constructs;
5. to investigate cross-national heterogeneity in the preferences for ATs and for their characteristics between UK and China using online SC surveys;
6. to investigate cross-methodological heterogeneity in the preferences for ATs and their characteristics and the impact of the immersive virtual environment in the preferences elicited with SC survey compared to online screen-based SC surveys;
7. to investigate the cross-experience heterogeneity in preferences for ATs and their characteristics, i.e. heterogeneity between those living in a city whether ATs are operating and available for the public and those living in a city where ATs are not operating.

1.3 Research Question

Based on the research aims and objectives, the research carried out in this dissertation aims to address the following research questions:

Research question 1: What are the key factors that ATs ought to have to be competitive with normal taxis with taxi driver?

Research question 2: How can these factors be appropriately measured in a SC survey?

Research question 3: What are the methodological implications of building a traditional SC survey embedded into a VR environment?

Research question 4: To which extent does living the choice in an immersive VR environment affect consumers' preferences?

Research question 5: Are preferences for ATs measured with SC experiments homogeneous across nations?

Research question 6: Does living in cities where ATs are operating and available to public have an impact on consumers' preference for AT and its characteristics?

1.4 Thesis outline

To achieve the research aims and objectives set up, the remainder of this dissertation is organised as follows:

An extensive literature review is presented in **Chapter 2 Literature Review**. This chapter starts with a critical review of factors affecting the choices of AVs and taxi services, mainly involving three aspects: service and vehicle features, individual-related characteristics, attitudinal characteristics and social conformity. The heterogeneity in the preference across-countries is also discussed. The next two sections then focus on the discussion of the survey techniques used to elicit preference for AVs and the state of art of VR technology. The section concludes discussing the research gaps identified.

Chapter 3 Data Collection Methodology reports the detailed description of four surveys (an online survey collected in the entire UK, an online survey collected in China, an online survey only in Newcastle and a survey collected in Newcastle using VR) built as part of this PhD research. The detailed process of constructing these four surveys includes the implementation of focus groups, several pilot tests, construction of experiments (attributes, attribute levels and experimental design), selection of attitudinal constructs and definition of additional individual information. Particular attention has been put in defining methods to enhance the realism of the SC survey. This includes the layout of the SC experiment, the definition and design of the attributes as well as the pre-information provided. The most interesting contribution refer to the methodology to embed SC experiment into immersive VR environment. Challenges and potentialities are discussed, along with methodological implications. Sample recruitment and implementation of these four surveys are described in the last section.

Chapter 4 Descriptive Data Analysis presents a descriptive analysis of the data collected and an initial discussion of the phenomenon as depicted in the data. The following five aspects are analysed: socio-demographic characteristics (gender, age, education level, etc.), trip characteristics (frequency of taxi usage, frequency of talking with driver, etc.), knowledge levels of AVs and ATs (heard of AVs, familiarity of 5 levels of automation, etc.), latent psychological factors, and frequency of the stated choices. The two datasets collected in the UK and China using the online surveys are compared to explore differences in the sample characteristics as a basis for the analysis of cross-national heterogeneity. The two datasets collected in Newcastle using the online survey and the VR survey are compared as a basis for the analysis of cross-methodological heterogeneity. Finally, the dataset collected in China using the online survey is further analysed comparing the characteristics of the respondents living in cities where ATs are operating and those living in cities where ATs are not operating, as a basis for the analysis of the cross-experience heterogeneity.

After that, **Chapter 5 Modelling Methodology** outlines the modelling approaches used to analyse and quantify the impact of key determinants. It first reports a description of the theoretical foundation of the Discrete Choice Models (DCMs), including the definition of the typical steps to build a model: utility specification, model estimation and model evaluation. The models formulations presented are mainly the Mixed Logit model for panel correlation and the Hybrid Choice model (HCM) to account for latent psychological constructs. The problem of joint estimation with different datasets and the scale heterogeneity is also discussed. The chapter presents the exact specifications used in the model estimated, the results of which are discussed in **Chapter 6.**

Chapter 6 Modelling Estimations & Discussions presents the results of the models estimated and specifically discusses the impact of the level of service attributes, in-vehicle features. normative conformity, knowledge of AVs and ATs, other effects, and impact of latent psychological factors. The chapter is organised in three sections. The first section discusses the results for UK and China, and highlight the cross-national heterogeneity. The second section discusses the results specifically for Newcastle, and highlight the cross-methodological heterogeneity and the impact of the immersive VR experience in the preferences elicited. The third section instead focuses on the results for China and it discusses the potential cross-experience heterogeneity in the preferences due to living in cities where AT services are operating or not. Willingness of Pay (WTP) for the AT characteristics and the normative influence effects are calculated and discussed also in comparison with the values reported in the literature, providing references for formulating policies. The results of the application of some policy scenarios using the models estimated are discussed at the last section of this chapter.

Finally, the dissertation concludes in **Chapter 7 Conclusion**, where the main findings of the research are summarised, the methodological implication of this research are discussed and the policy implications highlighted. The chapter and the thesis concludes with a discussion of the limitations of this study and suggestions for future research.

1.5 Thesis main innovations and contributions

The main innovations and contributions of this dissertation can be summarised into the following four aspects: (1) Data collection; (2) Model estimation; (3) Results and (4) Policy implications.

(1) Data collection

The dissertation proposes a novel methodology to collect data using SC experiments embedded within an immersive VR environment. The overall objective of this new methodology is to reduce the problem of 'lack of realism' in traditional SC experiments. The contribution stands mainly in two aspects:

1. How to build a SC experiment embedded into a VR environment. Here the novel contribution consists in providing guidelines on technical aspects related to building VR experiments to study transport choices and in highlighting three major elements that contribute to lack of realism in the standard screen-based SC experiments: a) assuming that the trip destination does not change among scenarios; b) including the payment method simultaneously with other attributes such as travel time; c) presenting the choice tasks continuously without a break.
2. How to build a SC experiment to measure the impact of novel attributes such as in-vehicle features and social conformity. Here the novel contribution consists in enhancing the current discussion on a) the images to define attributes, in particular for the in-vehicle features and the customer reviews, b) the conscious selection of the attributes, without mixing attributes that are evaluated at different stages of the decision process, c) a more realistic setting of the decision process across scenarios, as well as the precise definition of these attributes and layout of SC survey.

(2) Model estimation

The dissertation makes use of mixed logit and hybrid choice models to estimate individual's preferences for automated versus normal taxis. The contribution stands in the use of these models jointly across multiple datasets to estimate:

1. Preference heterogeneity across nations
2. Impact of living in cities where ATs are operating
3. Impact of living the choices in a VR environment

In addition, the modelling methodology discusses the calculation of the WTPs confidence intervals in the case of the systematic heterogeneity, a key aspect that is often overlooked.

(3) Results

The most significant results of this dissertation can be summarised as follows:

1. The option to 'change the destination' and the option to 'chat with an operator' are significantly relevant in terms of using ATs but the significance of these in-vehicle features is different across segments of population.
2. Results confirmed the difficulty in capturing the effect of descriptive norms within a SC experiment online, while the other social conformity variable, customer reviews, could be measured correctly within the SC and confirmed to be significant in the choice of the type of taxi.
3. The cross-methodological comparison revealed that the survey methods can influence the estimated preferences. Results showed that, interestingly, the descriptive norm was found to be positive and significant at 95% in the VR experiment, which probably reflected the important role of realism level when capturing the effect of descriptive norms. Results also showed that choices stated in the VR experiment were less sensitive to price.
4. Significant heterogeneity between the UK and China, was observed after carefully controlling sample heterogeneity in the socio-demographic characteristics. Differences across nations were found not in the preferences for the level of services (travel cost, travel time and waiting time) but in the impact of two latent psychological constructs. Hedonic motivation and trust in the UK had a much higher on the choices of ATs than in China, probably reflecting the cultural differences in the adoption of ATs.
5. Regarding the impact of indirect experience, the most interesting findings were that

participants from AT cities (where ATs are in operations) care less about the attribute 'change the destination' but more about customer reviews compared to the participants from NT cities (where ATs are not in operations). The latent variable 'hedonic motivation' is only significant among respondents living in AT cities, while the latent variable 'trust' is only significant among respondents living in NT cities.

(4) Policy implications

The dissertation investigates the AT market share variation under different policy scenarios. In addition to the traditional level of services variables, the study also explores to what extent the enhancements in in-vehicle features, customer reviews, and three latent psychological variables contribute to the increased demand for the ATs. Here the contribution stands in identifying the relative impact of improvements in objective (level of services and in-vehicle features) and subjective (latent psychological) characteristics in the AT market share and the comparison between the UK and the Chinese market share.

Chapter 2 Literature Review

2.1 Introduction

This chapter provides a critical review of the literatures on three key aspects related to seven specific research objectives. In detail: **Section 2.2** discusses the factors affecting the choice of autonomous vehicles (AVs) and taxi services in general and automated taxis (ATs) in particular. It also looks at cross-national heterogeneity in the choice of AVs and use of taxi services. **Section 2.3** discusses the survey techniques used to investigate individuals' preferences for AVs, mostly Stated Choice (SC) experiments and **Section 2.4** discusses virtual reality (VR) technology used in behavioural research. Finally, **Section 2.5** concludes this chapter.

There are two main operational models for highly or fully automated vehicles (Mourad *et al.*, 2019). Autonomous vehicle-as-a Product (CaaP) means a fully AV that is available for private purchase (i.e. private owned AVs) (e.g. Daziano *et al.*, 2017; Potoglou *et al.*, 2020 etc.; Ding *et al.*, 2021). Autonomous vehicle-as-a-Service (CaaS) means an on-demand service deploying fully AVs (i.e. shared AVs (SAVs) or ATs) (e.g. Krueger *et al.*, 2016; Yap *et al.*, 2016 etc.). CaaP involves long-term mobility decisions while CaaS involves short-term mobility decisions. The following discussion focuses on the factors associated with usage of AV, not factors about privately owning an AV (e.g. purchase cost and parking cost, etc.), as those factors are not relevant to ATs.

2.2 Factors influencing the choice to use AVs and taxi services

The factors influencing the choice of AVs and taxi services have been extensively studied in the literature (See research reviews: Becker and Axhausen, 2017; Gkartzonikas and Gkritza, 2019 etc.). These studies found that the choice of AVs (including owning and/or using AVs) are generally influenced by a multiplicity of determinants that are generally classified into three groups: 1) service characteristics and/or vehicle characteristics (for taxi usage, it also includes characteristics of taxi ranks); 2) individual-related characteristics; and 3) attitudinal characteristics and social influence. However, as discussed in more detail below, most of this literature focuses on psychological aspects behind the intention and the benefits and barriers to adopt the technology. Most of the literature focuses also on private ownership of AVs, few studies have instead focused on using AVs, like in a taxi service.

2.2.1 Service and vehicle characteristics

Among the variables related to the service and/or vehicle characteristics affecting the choice of AVs, travel cost/operational cost, travel time, waiting time, access or egress time/distance are common and main factors affecting traveller's preferences in using AVs (Krueger *et al.*, 2016; Yap *et al.*, 2016; Bansal and Daziano, 2018). The majority of these attributes are typically significant also in the choice of conventional mode choices, i.e. they are not specific for AVs. However, level of automation is a specific factor affecting the acceptance for owning AVs (Daziano *et al.,* 2017; Potoglou *et al.,* 2020), and using AVs (Bansal and Daziano, 2018). Bansal and Daziano (2018) indicated that based on investigation of 303 New Yorkers, automation (i.e. no driver) is a significant factor affecting the use of ATs and an average of $3.2 per trip can compensate New Yorkers for using automation in a commuting trip. Furthermore, AV-related literature found that potential AV users are worried about safety issues. However, the majority of papers have analysed individual's perception of safety, not safety as an objective measure (Shabanpour *et al.* (2018) defined safety in choice task as two relative levels: lower and higher than current average vehicle). Other vehicle and trip-related characteristics have also been found to affect the purchase and/or use of AVs. These are: trip purpose (Kolarova *et al.*, 2019), emission rate (Shabanpour *et al.*, 2018), car model (shape) and car size (Potoglou *et al.*, 2020), driving range and exclusive lane (Shabanpour *et al.*, 2018) as well as level of reliability (i.e. level of congestion) (Stoiber *et al.*, 2019).

Among this literature, few papers have dealt with SAVs. Krueger *et al.* (2016) studied Australian preferences for dynamic ride-sharing (SAV). They used a SC experiment with three alternatives (two AVs with and without ride-sharing and the current public transit) and three attributes (travel time, travel cost and waiting time). Respondents were informed that SAV could be imagined as driverless taxi services. Bansal and Daziano (2018) studied willingness to share a ride with strangers using a SC experiment with three alternatives (two Uber modes with and without ride-sharing and the current travel mode), where a dummy variable was used to indicate whether the Uber was with or without driver. Yap *et al.* (2016) investigated the choice of the egress modes of train trips, where two of the available options are cybercars (driving yourself and automatic driving). Alternatives are described in terms of travel cost, waiting time, travel time and walking time to the destination, plus a dummy variable to indicate if the cybercar is a shared vehicle or not.

Tables 2-1 provides an overview of the service and vehicle characteristics most studied in the literature, to explain the choice of using AVs and/or taxi services.

There is very little research specifically on the choice of taxis. Taxis are generally described by some service attributes used for describing AVs such as waiting time, travel time, travel cost and access time/distance as well. Wong *et al.* (2020a) investigated 1410 taxi customers' selection between electric taxis and liquefied petrol taxis in Hongkong, with the two types of taxis described by 4 attributes: walk time, on-street wait time, off-street and taxi fares. Wong *et al.* (2020b) then investigated the willingness to travel between ordinary and accessible taxis among the elderly and paid specific attention to the attribute 'subsidized taxi fare'. Other attributes considered were walk time and wait time. Wong and Szeto (2022) considered the effect of taxi surcharge in congested areas on customers' decision to use taxis. They used the common service attributes: walk and wait time, travel time and travel fare to depict the taxi option with surcharge. Besides these choice behaviour studies, the studies on perceived taxi service quality influencing customer satisfaction can also reflect the importance of service and vehicle characteristics.

Locations	Sample size	Dependent variables	Service characteristics								Vehicle characteristics		Source
			Travel cost	Travel time	Waiting time	Congestion time	Subscription cost	Vehicle availability	Access/egress time or distance	Availability of Sharing	Emission Rate	Level of automation	
Netherlands	761	PT/bicycle/manual AV/ automatic AV	x	x	x				x	x			(Yap *et al.*, 2016)
Australia	435	Current mode/ SAV without ride-sharing/ SAV-with ridesharing	x	x	x								(Krueger *et al.*, 2016)
Israel/North America	721	Current Car/PAV/SAV		x			x						(Haboucha *et al.*, 2017)
USA	298	Current car/ AT without ridesharing/ AT with ridesharing	x	x	x				x		x	x	(Bansal and Daziano, 2018)
USA	721	Buying vehicle Yes/No	x								x		(Shabanpour *et al.*, 2018)
Switzerland	709	Short-term: PAV/AT/AS	x	x	x				x	x			(Stoiber *et al.*, 2019)
		Long-term: PAV/AT/AS	x		x		x		x	x			
Germany	511	Walk/Bike/PT/ Current car/AV/SAV	x	x	x				x				(Kolarova *et al.*, 2019)
Australia	512	Conventional car/AV/PT	x	x		x							(Krueger *et al.*, 2019)
Austin, USA	1021	Current car/AV/SAV	x		x		x						(Asmussen *et al.*, 2020)
6 countries	6033	Petrol car and 3 alternative fuelled cars	x									x	(Potoglou *et al.*, 2020)
Australia	1433	6 out of 10 alternatives (taxi, PT, current car etc.)	x		x							x	(Zhou *et al.*, 2020)
Korea	511	Human-driven vehicle/AV	x	x									(Lee *et al.*, 2021)
China	542	Carsharing/AV/SAV	x					x	x				(Tian *et al.*, 2021)
Germany	484	Walk/Bike/PT/ Current car/AV/SAV	x	x	x					x			(Kolarova and Cherchi, 2021)

Table 2-1 Summary of service/vehicle characteristics studied to explain the choice of using AVs

PT: Public Transport; AV: Autonomous vehicle; AT: Automated taxis; PAV: Private autonomous vehicle; SAV: Shared autonomous vehicle; AS: Autonomous Shuttle

Location	Sample	Taxi										Taxi rank					Taxi driver					Source:
		Travel time	Travel cost	Waiting time	Walking time	punctuality	safety	Cleanliness	Comfort (Crowd, noise)	Car Condition	Payment method	Accessibility	Locations	Safety	Cleanliness	Comfort (Crowd, noise)	Professionalism	Drive safely	Appearance	Behaviour (carry luggage)	Service attitude	
Doha, Qatar	180	X	X			X	X	X	X			X		X	X	X	X	X	X		X	(Shaaban and Kim, 2016)
India	105		X			X			X	X							X	X		X		(Chaudhary *et al.*, 2016)
Santander, Spain	215	X	X	X			X	X	X		X	X	X							X		(Alonso *et al.*, 2018)
Hongkong, China	1008	X		X	X	X			X			X				X	X	X			X	(Wong and Szeto, 2018)
Melbourne, Australia	439						X	X	X	X							X	X			X	(Rose and Hensher, 2018)
Iran	559	X	X	X			X	X	X	X							X	X		X	X	(Askari *et al.*, 2021)

Table 2-2 Summary of factors studied to explain the perceived taxi service quality

As shown in **Table 2-2**, in evaluating perceived taxi service quality except typical service characteristics, some specific attributes relevant to taxi vehicle (like safety, cleanliness, comfort, car condition, etc.), taxi rank (like safety, cleanliness, comfort, etc.) and taxi driver (like professionalism, greetings, possibility of carrying luggage etc.) might be considered as non-ignorable factors. Shaaban and Kim (2016) recruited 180 taxi users for assessing taxi service quality in Doha, Qatar and they found that some characteristics relevant to taxi vehicle and taxi rank, such as level of crowding, noise and cleanliness, etc. significantly affect the evaluation of taxi service quality. Moreover, taxi driver characteristics such as safe driving, greetings etc. also play pivotal roles. In India, based on the investigation of 105 participants, Chaudhary *et al.* (2016) found driver professionalism and convenience of booking are two key factors influencing customer satisfaction. In evaluating the taxi service quality of Santander, Spain, Alonso *et al.* (2018) analysed 215 taxi users' perceived quality for taxi mode and found the impacts of payment method and possibility of carrying luggage were essential. Wong and Szeto (2018) found besides longer waiting time for taxis, Hong Kong taxi users have prioritised requests in enhancing the problems relevant to hire refusal by taxi drivers, attitude of taxi driver and professionalism of taxi drivers. Differently, Rose and Hensher (2018) discovered that Melbourne citizens focus more on taxi services in these two areas: 'driving knowledge of route' and 'driver driving ability'. Finally, Askari *et al.* (2021) investigating four factors (driver behaviour, vehicle condition, service management, and total travel time) influencing customer satisfactions on fixed-route taxi service in Iran, found taxi drivers' behaviour is the most important determinant in passenger satisfaction regarding taxi services.

Differently from a SAV and a traditional taxi, the AT service ought to be designed to satisfy various taxi passenger requests (such as adjusting the heating or air conditioning, changing the destination and adding a stopover) which in a normal taxi are dealt with a direct communication between the passenger and the driver (Kim *et al.*, 2020). It is unknown if in-vehicle features are necessary from taxi passenger perspectives and there is little evidence on what the role of in-vehicle features will play in choosing an AT for potential AT passengers. In-vehicle features might highly influence the potential passenger's choice of ATs, and this knowledge is very valuable for AT manufacturers and operators for developing ATs. Nordhoff *et al.* (2020) investigated users' perception with respect to the possibility of manually steering an automated shuttle and of having a button inside the automated shuttle which they can press to stop it, while Paddeu *et al.* (2020) used a naturalistic experiment where respondents were asked to rate the impact of the direction of seat (backwards/forwards) on comfort and trust in a shared AS.

2.2.2 Individual-related characteristics

Individual-related characteristics include socio-demographic characteristics and current travel behavioural habit or experience. Among the socio-demographic characteristics, the literature found that gender, age, income, education level, employment status, presence of children and possession of driving license, all have been commonly studied in the choice of owning and/or using AVs (Casley *et al.*, 2013; Bansal *et al.*, 2016; Yap *et al.*, 2016; Zmud *et al.*, 2016; Haboucha *et al.*, 2017). The evidence for the effect of gender on the acceptance of AVs is inconclusive. Men, for example, were found to prefer both purchasing and using AVs more than women do (Casley *et al.*, 2013; Payre *et al.*, 2014; Kyriakidis *et al.*, 2015).This could be due to the fact that men are less worried about full automation and the potential safe issues. AVs in fact are considered safer among men (Casley *et al.*, 2013; Kyriakidis *et al.*, 2015). Differently, Greek females are more likely to use AVs (Panagiotopoulos and Dimitrakopoulos, 2018). It seems that in Greece there is less gender gap in the acceptance of AVs. Hohenberger *et al.* (2016) provide empirical evidence from Germany to explain gender difference in the acceptance of AVs might be because of its mediator effect of affective response (i.e. anxiety and pleasure). However, no gender difference was significantly affecting intention to use AVs by Madigan *et al.* (2017), and owning AVs by Shabanpour *et al.* (2018) and Potoglou *et al.* (2020).

Age was also found to have an inconsistent impact on the choice of AVs. Payre *et al.* (2014) claimed that the age has no significant effect on using and owning AVs. Furthermore, Madigan *et al.* (2017) reported that age has no moderating effect in explaining the behavioural intention to use automated public transport. On the contrary, Haboucha *et al.* (2017) reported that older people preferred owning and using regular vehicles, as they are less open to trying new technologies and new ideas. Hulse *et al.* (2018) indicated that young adults displayed a greater acceptance of AVs. Regarding the income, 347 respondents in Austin indicated that travellers living in urban areas, with higher-income, are more willing to pay to own AVs (Bansal *et al.*, 2016). Differently, the income, however, has no significant impact on intention to use automation technology (Bansal *et al.*, 2016). Haboucha *et al.* (2017) found that people with higher education level favour owning and using the SAVs and private AVs, and there is a higher probability for household with the more than one child to choose SAVs. In addition, some other socio-demographic variables were also considered in AV ownership-related literature, such as ethnicity, marital status (Daziano *et al.*, 2017), and presence of disability (Shabanpour *et al.*, 2018).

Besides socio-demographic variables, among current travel behaviour habits or experiences, the travel mode previously used has a significant effect on the choice of AVs, and accident experience and daily/annual travel distance also significantly affect AV purchase decisions. Krueger *et al.* (2016) indicated that the individuals who get accustomed to using private vehicles are less likely to use SAVs. This is probably because private-car-oriented individuals have exclusive travel habit and it could be difficult to change this habit. Multi-modal travellers are more likely to adopt SAVs since SAVs could enhance their current travel habit (i.e. multimodality) and this group of individuals may be more open for novel mobility. Shabanpour *et al.* (2018) and Bansal *et al.* (2016) demonstrated that the majority of people who have experienced an accident and with higher annual/daily travel distances are more likely to buy AVs in future. This could be that this group of people would like to increase driving safety by buying AVs and productively using the travel time in longer travelling. Finally, some other individual -related characteristics, such as home/work location (Shabanpour *et al.*, 2018) and car availability (Krueger *et al.*, 2016; Kolarova and Cherchi, 2021), experience with Advanced Driver Assistance System (ADAS) (Kolarova and Cherchi, 2021) may also affect the purchase of AVs and use of AVs, respectively.

Table 2-3 shows a brief summary of individual-related factors studied to explain the acceptance of AVs. Individual-related characteristics involves socio-demographic characteristics and current travel behavioural habit or experience.

Location	Sample	Context	Individual-related characteristics													Source
			Gender	Age	Education	Income	Work status	Children	Driving license	Knowledge of AVs	Past crash experience	Public transport pass	Car availability	Current transport mode	Trip Frequency	
France	421	Use	X	X												(Payre *et al.*, 2014)
USA	347	Use	X	X	X	X	X	X	X	X	X					(Bansal *et al.*, 2016)
Australia	435	Use	X	X		X		X					X	X		(Krueger *et al.*, 2016)
Netherland	761	Use	X	X	X	X			X			X	X	X	X	(Yap *et al.*, 2016)
Israel and North America	721	Use and ownership	X	X	X	X	X	X								(Haboucha *et al.*, 2017)
Greece	315	Use	X	X											X	(Madigan *et al.*, 2017)
Germany	384	Use	X	X			X									(Nordhoff *et al.*, 2018b)
Greece	483	Use	X	X												(Panagiotopoulos and Dimitrakopoulos, 2018)
USA	1013	Ownership	X	X	X	X					X				X	(Shabanpour *et al.*, 2018)
Germany	511	Use				X			X			X				(Kolarova *et al.*, 2019)
China	1355	Ownership	X	X	X	X	X		X	X						(Liu *et al.*, 2019a)
USA	1021	Use and ownership	X	X	X	X	X	X								(Asmussen *et al.*, 2020)
6 Countries	6033	Ownership	X	X	X	X		X		X						(Potoglou *et al.*, 2020)
Australia	1433	Use	X	X	X				X							(Zhou *et al.*, 2020)
Germany	484	Use	X	X	X	X			X			X	X			(Kolarova and Cherchi, 2021)

Table 2-3 Summary of individual-related characteristics and travel habit studied to explain AV choices

2.2.3 Attitudinal characteristics

Extensive research has been conducted to analyse the impact of psychological factors on the intention to purchase and/or use AVs (e.g. Choi and Ji, 2015; Hewitt *et al.*, 2019). Psychological factors generally can be categorised as two types.

Type I: these studies focus on individuals' concerns, perceptions and attitudes affecting the intention to and/or willingness to pay for using AVs by using a sort of naturalistic tests (using for example descriptive analyses). This type of studies generally focuses on the perceived risk and perceived benefits that AVs could bring to individuals. Among the aspects studied in this category (**Type I),** safety concern is probably one of the most influential. For example, Casley *et al.* (2013) in a US study indicated that 82.41% of the participants ranked 'perceived' safety (in general terms of trust to computer) as the most important factor related to AVs and safety was regarded as the most influential factors for owning AVs, followed by legality. Begg (2014) collected perceptions of over 3500 respondents from London regarding the improvement of safety in general (i.e. it does not refer to specific aspects of safety) of road users. Results indicated that 36% of respondents agreed and 24% strongly agreed that AVs can enhance safety for all road users. However, other psychological factors have been studied and found to be significant. These are liability concerns, privacy concerns, software hacking/misuse concerns, environmental concerns, pleasure of driving, and pro-AV attitude, etc. (Casley *et al.*, 2013; Begg, 2014; Howard and Dai, 2014; Schoettle and Sivak, 2014; Kyriakidis *et al.*, 2015; Bansal *et al.*, 2016; Zmud *et al.*, 2016; König and Neumayr, 2017; Kaur and Rampersad, 2018; Sanbonmatsu *et al.*, 2018).

Type II: these studies focus on respondents' attitudes or perceptions of AV-related technology by using psychometric indicators. Most of the papers in this area follow some (extended) psychological underlying theory such as Technology Acceptance Model (TAM), Theory of Planned behaviour (TPB) and Unified Theory of Use and Acceptance of Technology (UTUAT). Among the studies in this category (**Type II),** as shown in **Table 2-3**, performance expectancy (perceived usefulness), effort expectancy (perceived ease of use) and subjective norms are the constructs included in almost all the studies. Facilitating condition (perceived behaviour control), trust, attitude toward AVs, perceived safety, perceived risk, facilitating condition and hedonic motivation are also common determinants in the acceptance of AVs. These studies though do not focus on ATs Some other psychological factors are also considered to play key roles in the acceptance of AVs and AV-related technologies, e.g. driving-related personality traits and locus of control (Choi and Ji, 2015), self-efficacy and anxiety (Hewitt *et al.*, 2019), compatibility (Rahman *et al.,* 2019), innovativeness (Chen *et al.*, 2020) etc.

As mentioned in **Table 2-4**, a variety of psychological factors have been found to play crucial roles in intention to use AVs (see also the literature reviews by Becker and Axhausen, 2017; Gkartzonikas and Gkritza, 2019; Narayanan *et al.*, 2020). In contrast, only few papers have considered the effects of the latent psychological factors in the choice of AVs or ATs as mode of transport, jointly with the objective characteristics of the AVs. Haboucha *et al*. (2017) considered the impacts of five latent psychological constructs on the use and purchase of AVs: public transport attitude, pro-AV attitude, perceived enjoyment of driving, technology interest and environment concern. In the context of use of AVs as last-mile travel mode, Yap *et al.* (2016) incorporated three attitudinal factors: trust in AVs, attitudes towards service reliability and attitudes towards sustainability. Finally, Kolarova and Cherchi (2021) studied the impact of trust and travel experience on the travel time of riding autonomously.

2.2.4 Attitudinal characteristics in the primary appraisal stage

According to the cognition-motivation-emotion framework proposed by Lazarus (1991), individual's decision making process includes three main stages: evaluating importance (primary appraisal), analysing behavioural options (secondary appraisal) and outcome stage. At the primary appraisal stage, the individuals analyse the relevance and significance of using AVs for travelling, then at the secondary appraisal stage the individuals will perform a careful and systematic appraisal of cost-benefit analysis for travelling by AVs (Ribeiro *et al.*, 2022). The outcome stage is the actual choice to use or not AVs. Due to lack of actual experience with ATs, respondents might find it hard to carefully and systematically consider perceived risks or benefits brought from using AVs. The primary appraisal stage is of interest for this PhD study. According to Ribeiro *et al.* (2022) the three components at the primary appraisal stage in the acceptance of AVs are: hedonic motivation, trust and social influence. Empirical studies provided only limited evidence in explaining the impact of these constructs on the choice of AVs in general and even less on the choice of ATs' in particular, with the exception of Tussyadiah *et al.* (2017), that studied the impact of trust and negative attitude toward AVs on intention to use specifically ATs.

Location	Sample	Context	Behaviour theory	PE (PU)	EE (PEoU)	SI (SN)	FC (PBC)	Trust	Atti	PR	PS	HM	Knw	Source
South Korea	552	AV	Extended TAM	X	X			X		X				(Choi and Ji, 2015)
France and Switzerland	349	APT	UTUAT	X	X	X								(Madigan *et al.*, 2016)
Greece	315	APT	UTAUT2	X	X	X	X					X		(Madigan *et al.*, 2017)
USA	312	AT	N.A.					X	X					(Tussyadiah *et al.*, 2017)
Greece	483	AV	Extended TAM	X	X	X		X						(Panagiotopoulos and Dimitrakopoulos, 2018)
Germany	384	AS	UTAUT	X	X	X					X	X		(Nordhoff *et al.*, 2018b)
China	300	AV (level3)	Extended TAM	X	X			X			X			(Xu *et al.*, 2018)
USA	187	AV (at 6 autonomy scenarios)	Extended UTAUT	X	X	X	X		X		X			(Hewitt *et al.*, 2019)
China	906	AV and SAV	Extended TPB			X	X		X	X			X	(Jing *et al.*, 2019)
USA	173	AV (level5)	Extended TAM and TPB	X		X		X	X		X			(Rahman *et al.*, 2019)
China	216	AV (level 3)	Extended TAM	X	X			X	X	X				(Zhang *et al.*, 2019)
China	913	AS	Extended UTAUT	X	X	X	X			X				(Chen *et al.*, 2020)
South Africa	121	AV	Extended UTAUT	X	X	X		X				X	X	(Morrison and Van Belle, 2020)
Australia/France/Sweden	1563	AV (level 4)	TPB/UTAUT	X	X	X	X		X					(Kaye *et al.*, 2020)
Iran	338	AV	Extended UTAUT	X	X	X								(Farzin and Mamdoohi, 2021)

Table 2-4 summary of latent constructs affecting behaviour intention to use AVs

PE (PU): Performance expectancy (Perceived Usefulness); EE (PEoU): Effort expectancy (Perceived Ease of Use); SI (SN): Social influence (Subjective norms) ; FC (PBC) :Facilitating condition (Perceived Behavioural Control); PR: Perceived Risk; PS: Perceived Safety; Knw: Knowledge; APT: automated public transport; AS: automated shuttle. Autonomy levels mentioned in the table 2-4 are based on SAE (2018)

Hedonic Motivation

Hedonic Motivation (HM) refers to the fun or pleasant derived from using a technology, originated from Unified Theory of Acceptance and Use of Technology 2 (UTAUT2) (Venkatesh *et al*., 2012). HM has been found to be a strong predictor in intention to use or accept a technology in many domains (e.g. information system, Van der Heijden, 2004). The impact of HM in the general intention of using AV was found to be positive (Keszey, 2020). In the specific context of automated shuttle, HM was found to be the strongest determinant among five UTAUT constructs (PE, EE, SI,FC and HM), affecting intention to use an automated road transport systems (which is similar to SAE L4 automated shuttle system, with safety steward who can intervene automated shuttle) (Madigan *et al.*, 2017). Nordhoff *et al.* (2018b) also support this result that HM is strongly correlated with taking a ride with an automated shuttle and even more strongly among those who never experienced the automated shuttle before. However, there is no empirical evidence of the impact of HM in the specific context of ATs. As argued by Kyriakidis *et al.* (2015) and Nordhoff *et al.* (2019), with the automation level increases, the less enjoyable the respondents perceived riding when using AVs and the full automation (Level 5) was regarded as the least enjoyable automation level. However, this result refers to the comparison with a normal car, where the users enjoy driving. It is unclear whether the role of hedonic motivation still plays an essential role in the choice of fully ATs, given that in both AT and NT the customer does not drive.

Trust

Trust refers to *the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability*, which is a foundation to human-automation interaction (Lee and See, 2004). However, as highlighted in Kolarova and Cherchi (2021) the term "trust" in the context of AVs does not have a unique definition. It is often used to represent vehicle safety (Jardim *et al.*, 2013; Gkartzonikas and Gkritza, 2019), reliability of the system (Yap *et al*., 2016; De Looff *et al.,* 2018), but it has also been used to describe affective reactions, such as being nervous or being afraid of using an AV.

Trust plays indeed a vital role in various sectors, particularly for innovative products. Several empirical studies have confirmed the critical role of trust on the intention to use AVs in general (Choi and Ji, 2015; Panagiotopoulos and Dimitrakopoulos, 2018; Zhang *et al.*, 2019; Ribeiro *et al.*, 2022) and ATs in particular (Tussyadiah *et al.*, 2017) and also in the choice among modes, where AV is one alternative (Yap *et al.*, 2016; Kolarova and Cherchi, 2021). Hewitt *et al.* (2019)

found that public perceptions or attitudes on AVs vary with different automation levels. However, only little research did focus on the automation level and considered its potential impact in investigating the role of psychological constructs (e.g. Panagiotopoulos and Dimitrakopoulos, 2018; Xu *et al.*, 2018; Zhang *et al.*, 2019).

Additionally, the context where the AT is used might also impact public acceptance of ATs and the context for ATs is likely to be different from the one for AVs. Given this, the role of trust in accepting in fully ATs might differ from the role of trust acceptance of AVs in general. It is then worth investigating the role of trust and quantifying the impact of trust in choosing fully ATs.

Social conformity

Research in social science suggests that human beings tend to consciously or unconsciously change attitudes, beliefs and behaviours to fit to groups of reference (Crutchfield, 1955; Cialdini and Goldstein, 2004). Turner (1991) states that "social influence relates to the process whereby people agree or disagree about appropriate behaviour, form or maintain social norms and the social conditions that give rise to, and the effects of such norms", where social norms are "a generally accepted way of thinking, feeling or behaving that is endorsed and expected because it is perceived as the right and proper thing to do. It is a rule, value or standard shared by members of a social group that prescribes appropriate or desirable attitudes and conduct in matters relevant to the group". Social influence is linked to social norms and social conformity is a type of social influence to match the group's normative belief (Cialdini and Goldstein, 2004). Social conformity, has been extensively studied in the psychological and social literature (Cialdini and Goldstein, 2004; Cialdini, 2005; Cialdini, 2007; Goldstein and Cialdini, 2011).

The impact of social influence or social conformity has been studied also in the transport literature, but differently from HM and trust, it has been studied also within SC experiments, i.e. as an attribute to trade-off with other objective characteristics. **Table 2-5** provides a summary of the social conformity effects studied and the methods used to measure them. Measuring social impacts is particularly complex. In particular, social conformity is a complex phenomenon that is affected by many psychological and objective factors and can manifest itself in different ways. Measuring social conformity is challenging, even more so if the goal is to identify an objective measure of it rather than using the classical psychological constructs used in the psychological literature (see Cherchi, 2017 for a detailed discussion).

By nature, ATs are highly innovative transport systems, and it is known that social influence plays a critical role in explaining individual choices for innovations. As discussed previously, several papers have studied the impact of subjective norms on the acceptance of AV, as part of psychological constructs (e.g. Madigan *et al.*, 2017; Nordhoff *et al.*, 2018b; Panagiotopoulos and Dimitrakopoulos, 2018 etc.), within TPB or UTAUT and their extensions framework. Madigan *et al.* (2017) tested the impact of social influence on intention to use the Automated Road Transport Systems (ARTS). Social influence was measured by three generic psychological statements adapted from UTAUT. Nordhoff *et al.* (2018b) tested the impact of social influence in the context of an automated shuttle by asking respondents "whether they would like to have their friends or family or other important people to them adopt the automated shuttle before they themselves do", and "whether people who are important to them would like it if the respondent used an automated shuttle".

A couple of studies have attempted to measure the effect of social influence or social conformity within SC experiments. However, these studies are applied to clean energy vehicles, which is considered an innovative product, though not as disruptive as AVs. The majority of the studies focused on the effects of descriptive norms, i.e. social adoption of the innovation by others (Kuwano *et al.*, 2012; Rasouli and Timmermans, 2013a; Araghi *et al.*, 2014; Kormos *et al.*, 2015; Okushima, 2015). For example, Rasouli and Timmermans (2013a) studied the effects of different market shares of electric cars in friends, relatives, colleagues and larger peer group on acceptance of electric cars. The results showed that these elements in social networks increased the probability of purchasing the electric vehicles. Araghi *et al.* (2014) analysed the willingness to offset flight-related carbon emissions. They divided the sample in 3 random groups, each was presented with different collective offsetting rate (5%, 50% and 90%). They found that the utility slope of carbon offsetting increased when the collective offsetting rate was high, which reflected high social conformity effects.

Case	Type of social conformity effect	Measurement method	Source
Intention to use public transport	• Important people's approval of use of public transport	• Rating on psychometric indicators	(Bamberg *et al.*, 2007)
Commute satisfaction	• Comparison of travel mode, stress level and travel time	• Firstly, enquire reference groups whose commute is familiar to them • Secondly, enquire travel mode of reference group, then rate on 5-point Likert scale about stress level and travel time compared to reference groups	(Abou-Zeid and Ben-Akiva, 2011)
Mode choice to work	• Share of social reference groups (level of income) • Share of spatial reference groups (residential postal codes)	• Attributes in choice model	(Walker *et al.*, 2011)
Purchase of electric vehicles	• Share of EVs	• A 4-level attribute in SP design	(Kuwano *et al.*, 2012)
Purchase of electric cars	• Market share of EV distinguished by 4 reference groups (friends, colleagues, peers and larger family) • Positive/negative review provided	• 4-level attributes in SP design	(Rasouli and Timmermans, 2013a)
Purchase of electric cars	• Market share of EV distinguished by 4 reference groups (friends, colleagues, peers and larger family) • Positive/negative review provided	• 4-level attributes in SP design	(Kim *et al.*, 2014)
Willingness to offset flight-related carbon emissions	• Collective offsetting rates	• A 3-level attribute outside the SP design	(Araghi *et al.*, 2014)
Sustainable commuting behaviours	• Percentage of commuters who switched to sustainable transport mode	• Set 3 groups: control, low and high social norm conditions and record one-month reduction of self-reported private vehicle use	(Kormos *et al.*, 2015)
Sustainable transport mode choice and preference for clean energy vehicles	• Proportion of commuters in your community who selected a sustainable transport mode A close friend owning CEVs	SP regarding mode change in case of the given local mode share with bus Ask preference in case that a close friend already owns a CEV	(Okushima, 2015)

Table 2-5 Summary of social conformity effects and measurement methods

Case	Type of social conformity effect	Measurement method	Source
WTP for SAVs	• AV adoption rate by friends	• Ask the question under the situation how many friends adopt AVs, respondents will also adopt	(Bansal *et al.*, 2016)
Intention to use public transport	• Public transport usage frequency of people among 4 reference groups (family members, friends, colleagues, and people in my living community) • Important people's approval of use of public transport	• Rating on psychometric indicators	(Zhang *et al.*, 2016)
Purchase of Electric Vehicles (EVs)	• Friend's opinion about EVs • Number of EVs recently bought • Important people's approval of use of EVs • Effect of being watched	• Ask a friend 3 questions outside SP design • A 3-levels attribute in the SP design • Rate on 7-point Likert scale on 3 psychometric indicators • An image of a pair of eyes as an attribute in the SP design	(Cherchi, 2017)
Car-sharing decisions	• Car-sharing rate by family, friends and others considering social distance that measure the strength of relationship between individuals	• 4-level attributes in stated adaption experiment	(Kim *et al.*, 2017)
Acceptance of Automated Road Transport system	• Important people's approval of use of Automated Road Transport System	• Rating on psychometric indicators	(Madigan *et al.*, 2017)
Acceptance of Automated Shuttle	• Important people's approval of use of Automated Shuttle	• Rating on psychometric indicators	(Nordhoff *et al.*, 2018b)
Intention to use autonomous driving	• Be proud if people saw me using AVs • Important people's approval of use of AVs	• Rating on psychometric indicators	(Panagiotopoulos and Dimitrakopoulos, 2018)

Table 2-5 (Cont'd) Summary of social conformity effects and measurement methods

The majority of this literature focused only on one aspect of normative conformity, namely social adoption. This could be due to fact that social conformity is not directly associated with features of the alternatives and measuring other aspects of conformity as an objective attribute in SC experiments is difficult and might not be perceived as realistic (Cherchi, 2017). Moreover, the above literature fails to distinguish between informational and normative conformity, which indeed measure different effects (Toelch and Dolan, 2015). Informational conformity occurs when an individual is an ambiguous (i.e. unclear) situation or lacks knowledge and hence turns to the members of their group for guidance, while normative conformity occurs because of the desire to be liked and accepted. Cherchi (2017) studied both aspects. She measured normative conformity in terms of both descriptive and injunctive norms and extended the measure of normative conformity measuring adoption and other signalling. She found that all these conformity effects were highly significant in the choice between electric car versus internal combustion vehicles and these different types of conformity had different effects on the choices of electric cars.

Finally, it is important to mention that Rasouli and Timmermans (2013) tested different adoption rates for different groups (friends, relatives, colleagues and general peers) and also included an attribute to measure the impact of public reviews, which were defined as: only positive, mainly positive, mainly negative, only negative (4 levels). They found that adoption rate was significant only for few groups and few levels of penetration rate. Negative reviews were not significant, while positive reviews had a significant positive impact on the intention to purchase electric vehicles, with no significant difference though between "only positive" and "mainly positive" reviews. Cherchi (2017) found instead that negative information had a significant impact in reducing the probability of buying an EV. The impact was not symmetric (positive information was less significant). These results are confirmed also by studies outside the transport-related domain. Zhao *et al*. (2015) found a significant negative relationship between negative online reviews and hotel booking intentions, while the impact of positive reviews was not statistically significant. Evidence from marketing suggest that reviews are critical factors for customer decision making (Vermeulen and Seegers, 2009; Mudambi and Schuff, 2010; Liu and Park, 2015). Customer reviews are important cues to help consumers evaluate the quality of the products to reduce the level of perceived uncertainty before experiencing or purchasing a product (Ye *et al*., 2011). Zhao *et al*. (2015) found a significant negative relationship between negative online reviews and hotel booking intentions, while the impact of positive reviews was not statistically significant. Zhu and Huberman (2012) measured how often respondents' choices change due to others' recommendations. They found that "other people's opinions significantly sway people's own choices" and the influence is stronger when facing a moderate, as opposed to large, number of opposing opinions. Customer reviews (representing general public opinion or a form of word of mouth), are increasingly used in reality as a form of social influence, but in transport have rarely been studied.

2.2.5 Cross-national comparison

Preferences and attitudes for AVs and AV-related attributes or AV-related psychological factors might be different among countries. As discussed in the previous sections most of the existing literature focused only on one country. Some studies have analysed to what extent preferences and attitudes are heterogeneous across countries, but the majority of them focus only on descriptive comparison with respect to some psychological factors (e.g. attitudes or concerns) towards AVs. Schoettle and Sivak (2014) administered a survey on attitudes and concerns towards Level 3 and Level 4 AVs among 6 countries: China, India, Japan, Australia, USA and UK. They found that concerns about AVs varied across these 6 countries. UK and Australian residents reflected moderate responses to 11 potential risks (e.g. safety consequence of equipment/system failure) compared with the Chinese, Indian, Japanese and American counterparts. According to 5000 responses from 109 countries, given cross-national effects, Kyriakidis *et al*. (2015) found that participants are generally concerned about software hacking/misuse and legality, however, respondents from high-income countries are more concerned about the threat of data misusing or sharing, while respondents from low-income countries are more concerned about the basic safety needs. A cross-national survey based on 7755 respondents from 116 countries was conducted by Nordhoff *et al.* (2018a). Their results indicate that, due to differences in thrill-seeking personality, participants from low-income countries (probably because they suffer from more transport-related problems) generally have high acceptance scores on AVs, as AVs might be a solution to these problems (e.g. parking place problem).

An investigation of attitudes towards automated road transport systems was implemented across European countries by Alessandrini *et al.* (2014) who found no commonalities in attitudes towards AVs across these European countries. Haboucha *et al.* (2017) compared and analysed preference heterogeneity in the ownership and use of AVs between Israel and North America. They found that Israelis are more likely to accept AVs than North Americans are and observed various differences in acceptance of AVs between the two countries. For example, Israelis care more about marginal cost rather than capital cost compared with North American. Etzioni *et al.* (2020) investigated the acceptance of AVs among 6 countries (Slovenia, Cyprus, UK, Hungary, Iceland, Montenegro) and individuals from high-income countries like UK or Iceland showed large hesitations towards AV acceptance. Individual differences across different countries in the acceptance of AVs were also observed. For example, the elderly (older than 59 years old) preferred to choose conventional cars in the UK, while this effect was not found in Montenegro and Hungary. Potoglou *et al.* (2020) examined the purchase behaviour of AVs in 6 countries (Germany, India, Japan, Sweden, UK and US) and found significant heterogeneities within and across countries. For instance, Indian respondents in Class 1 (those who were self-identified as pro-environment, have university degree, etc.) preferred to choose medium and large size cars with unique design, while Swedish respondents in Class 1 have no preference differences in terms of choosing different car size and design. Finally, Polydoropoulou *et al.* (2021) specifically focused on gender differences in the choice of SAVs across 7 countries (Cyrus, Greece, Israel, Hungry, Finland, Iceland, United Kingdom). Heterogenous gender-wise effects of co-passengers was found to play an important role when using SAVs (i.e. sharing with strangers).

Understanding the national differences in acceptance of AVs is critical in order to implement suitable policies and offer insights for different market segmentations (Nordhoff *et al.*, 2018a). Research on AV is clearly biased toward developed countries. In terms of taxi service, the studies of passengers' perceptions of taxi service quality also have largely focused on urban transport in developed countries. Directly employing strategies or conclusions into urban transport system in developing countries may not be transferable (Askari *et al.*, 2021). Much more efforts should be paid to study AV acceptance in developing countries (like some countries in Asia and Africa) in order to have a more comprehensive understanding of AV acceptance and develop rational policies for different country-segmentations (Jing *et al.*, 2020).

2.3 Survey techniques to measure choice/use of AVs and taxi services

Different survey methods have been employed to investigate factors affecting users' choice of AVs and taxi services. The survey method used depends on research objectives. **Table 2-6, 2-7 and 2-8** provide an overview of the survey techniques used for different research objectives.

2.3.1 Service and vehicle characteristics

Papers aiming at studying the characteristics of the vehicles and that of the service, typically use SC methods. For example, focusing on energy efficiency and autonomous features, Daziano *et al.* (2017) conducted an online SC experiment involving 1260 respondents to investigate public WTP for three levels of self-driving vehicles. An online SC questionnaire including 1053 respondents across Netherlands was also employed by Yap *et al.* (2016) to explore the difference of commuters' preferences between existing transport modes and AVs as egress mode of train trips. Krueger *et al.* (2016) also conducted a SC experiment comprising 435 individuals to study respondents' differences in WTP for service attributes of SAVs. When studying the choice of traditional taxi services, focusing on the service attributes of taxis, several face-to-face SP surveys were administrated in Hong Kong to analyse respondents' preferences in the choice of traditional taxi service. (Wong *et al.*, 2020a; Wong *et al.*, 2020b; Wong and Szeto, 2022). Differently from single-alternative SP experiments, a ranked SP experiment was used by Asmussen *et al.* (2020) to elicit Austin preferences on AV adoption, in which AV option was characterised by three service attributes and a best-worst SP experiment was employed by Shabanpour *et al.* (2018) to examine the most attractive or least attractive AV features affecting adoption decision of Chicago residents.

SC experiments are commonly used to investigate users' choice of AVs as this is a product not yet available in the market. Lack of realism and the related hypothetical bias affects all SC experiments but it is more marked in the case of highly innovative products as respondents have no experience with them and could not have formed preference for the product (see a discussion in Cherchi and Hensher, 2015). In this case, lack of realism in SC can give invalid results.

Location	Sample	Survey Method	Sources
Netherland	1053	Online SP survey	(Yap *et al.*, 2016)
Australia	435	Online SP survey	(Krueger *et al.*, 2016)
Israel&North America	721	Online SP survey	(Haboucha *et al.*, 2017)
New York/US	1260	online SP survey	(Daziano *et al.*, 2017)
Chicago/US	1253	Online Best-Worst SP survey	(Shabanpour *et al.*, 2018)
New York/US	303	Online SP survey	(Bansal and Daziano, 2018)
Switzerland	709	Online SP survey	(Stoiber *et al.*, 2019)
Germany	511	Online SP survey	(Kolarova *et al.*, 2019)
6 countries	6033	Online SP survey	(Potoglou *et al.*, 2020)
Austin/USA	1021	Ranked SP survey	(Asmussen *et al.*, 2020)
Germany	484	Online SP survey	(Kolarova and Cherchi, 2021)
Taxi Hong Kong/China	1410	On-site SP survey	(Wong *et al.*, 2020a)
Taxi Hong Kong/China	580	On-site SP survey	(Wong *et al.*, 2020b)
Taxi Hong Kong/China	773	On-site SP survey	(Wong and Szeto, 2022)

Table 2-6 Summary of survey techniques used for studying service/vehicle characteristics

Some researchers specifically considered the problem of realism in SC experiment when investigating AV-related preferences. For example, when Bansal and Daziano (2018) examined intentions to use two types of low-emission ATs versus current travel mode, for enhancing the realism in the experiment, a SC experiment with attribute levels pivoted around realistic values that respondents provided in the reference alternative was presented to respondents. However, most of the respondents still lack of knowledge on ATs. Lack of knowledge of the alternatives will affect the ability of SP experiment to reflect the real respondent preference (Cherchi and Hensher, 2015). For this reason, another method has been used among the literature to mitigate the realism problem in SC experiments: introduction of AV-related and trip-related information by videos or images before SC experiments (e.g. Krueger *et al.*, 2016; Kolarova *et al.*, 2019). For instance, some scholars (e.g. Howard and Dai, 2014) have attempted to offset the problem by showing AV video to participants before conducting survey or using images/videos in text-based surveys. Although videos and images help in improving realism, the SC experiment is still present in the standard format, which is in itself a non-realistic way in which choices are made in real life. Kolarova *et al.* (2019) assessed the effect of AVs on value of travel time savings. Before SC choice task part, two short animated videos about how a trip might look like with an AV and a text description of SAVs were presented to participants. This may help in familiarising the innovation to the certain extent ('level of familiarity' indeed played a critical role in choice of AVs (see Bansal *et al.*, 2016; Asmussen *et al.*, 2020)). Images and videos related to AVs are rarely neutral and it might have a priming effect on individuals, which further positively or negatively affect the real individual preference. Hence building a relatively real choice environment is an important prerequisite for accurately and reliably investigating customer preferences on ATs.

2.3.2 Psychological constructs

Papers aiming at studying psychological constructs mainly focuses only on (adapted) psychometric indicators measured with Likert scales. For example, Choi and Ji (2015) conducted an online survey incorporating 552 drivers in South Korea to examine the importance of 9 psychological constructs on intention to use AVs and each psychological construct was measured by 3 psychometric indicators. Panagiotopoulos and Dimitrakopoulos (2018) recruited 483 Greeks to examine the role of 4 psychological constructs measured by 13 psychometric indicators on behavioural intention to use AVs. Rahman et al. (2019) explored the willingness to use self-driving vehicles using 5 psychological constructs from the perspectives of American elderly people as pedestrians and users. Madigan *et al.* (2017) used 17 refined psychometric indicators for measuring 5 psychological constructs to investigate their impacts on intention to use automated road transport system in Greece. The same survey technique also was conducted to examine the role of different psychological constructs in using AVs by Zhang *et al.* (2019) and Hewitt *et al.* (2019), etc.

Location	Sample	Survey method	Source
South Korea	552	Online questionnaire 30 indicators, 10 constructs	(Choi and Ji, 2015)
France and Switzerland	349	7 indicators, 4 constructs	(Madigan *et al.*, 2016)
Netherland	761	Post-SP online questionnaire，16 indicators, 3 constructs	(Yap *et al.*, 2016)
Greece	315	20 indicators, 6 constructs	(Madigan *et al.*, 2017)
Israel/ North America	721	Pre-SP online questionnaire, 24 indicators, 5 constructs	(Haboucha *et al.*, 2017)
Greece	483	14 indicators, 5 constructs	(Panagiotopoulos and Dimitrakopoulos, 2018)
China	300	17 indicators, 6 constructs	(Xu *et al.*, 2018)
USA	173	15 indicators for pedestrians, 5 constructs /19 indicators for users, 6 constructs	(Rahman *et al.*, 2019)
China	216	23 indicators, 7 constructs	(Zhang *et al.*, 2019)
USA	187	26 indicators, 9 constructs	(Hewitt *et al.*, 2019)
Germany	384	19 indicators, 6 constructs	(Jing *et al.*, 2019)
Germany	484	Post-SP online questionnaire，6 indicators, 2 constructs	(Kolarova and Cherchi, 2021)

Table 2-7 Summary of survey techniques used for psychological constructs

Only three papers, studied the effect of psychological constructs together with service/vehicle characteristics, influencing the choice of AVs. Before the SC experiments, Haboucha *et al.* (2017) used 24 indicators to measure the role of 5 psychosocial constructs for eliciting preferences on the choice among current car, PAVs and SAVs. For the last-mile travel mode choice after train trip Yap *et al.* (2016) used 16 out of 23 indicators (after SC experiment) to measure impacts of 3 psychological constructs on the choice among 5 alternatives, including cyberwar with/without full automation. After SC experiments, Kolarova and Cherchi (2021) employed 6 indicators to measure the effect of two psychological constructs on the choice of 5 alternatives, also including PAVs and SAVs.

2.3.3 Other survey methods

Papers aiming at studying public perceptions, attitudes, fears, concerns or perceived benefits towards AV technology mainly use interviews, focus groups or questionnaires, typically including ranking or rating questions measured with Likert scales. For example, 35 semi-structured interviews have been used by Kaan (2017) to explore what factors respondents will come up with and also examined the importance of identified factors in a sample of 35 young individuals. Silberg *et al.* (2013) used focus group to investigate opinions about self-driving vehicles in a sample of 32 participants. Qualitative questions have been used also by Howard and Dai (2014) to investigate general attractiveness of AV compared to existing transport modes and by Zmud *et al.* (2016) who interviewed 44 residents in Austin (Texas) to investigate how opinions and perception of AVs influence intention to use and to own AVs. Ranking question, rating questions and text questions have also been used by Casley *et al.* (2013), Schoettle and Sivak (2014) and Kyriakidis *et al.* (2015) to investigate public feeling, beliefs, concerns towards AV-related issues and WTP for automations. For example, according to a 63-question Internet-based survey, Kyriakidis *et al.* (2015) studied 5000 respondents' concerns and WTP towards AVs with different automation levels. Schoettle and Sivak (2014) investigated 6 countries' respondents' perceived benefits and perceived concerns when using level 3 and level 4 AVs and their willingness to pay for self-driving technology. These methods allow for statistical analyses about the associations between psychological factors and intentions or WTP for using and owning AVs, but they cannot effectively measure the extent to which the specific characteristics of the innovative products (i.e. the AVs) affect the intention to use the AVs and the willingness to pay for it. This piece of information is crucial to identify the specific characteristics the AVs need to have and in which combination, in order to be accepted by the population or segment of it.

Location	Sample	Survey Method	Sources
US cities	32	Focus group	(Silberg *et al.*, 2013)
Worcester, US	467	Online attitudinal survey	(Casley *et al.*, 2013)
Berkeley, US	107	Semi-structured interview with video about AVs	(Howard and Dai, 2014)
6 countries	1722	Online attitudinal survey	(Schoettle and Sivak, 2014)
109 countries	5000	Online attitudinal survey	(Kyriakidis *et al.*, 2015)
Austin, Texas	2167	Online attitudinal survey	(Bansal *et al.*, 2016)
Austin, Texas	556/44	Online attitudinal survey/semi structured Interview	(Zmud *et al.*, 2016)
Netherland	35	semi-structured Interview	(Kaan, 2017)

Table 2-8 Summary of other survey techniques

2.4 Virtual Reality

2.4.1 General literature on virtual reality

As the rapid advances and popularisation of the VR techniques, these have become available and affordable for scientific research purposes. Differently from the traditional or standard survey techniques, VR techniques can offer close-to-realistic, immersive and interactive environments, which greatly improve the degree of 'immersion' and 'presence' in hypothetical scenarios (Sanchez-Vives and Slater, 2005).

VR experiments represent a new area of research that promises to change radically the way surveys are conducted to measure preferences. According to the theory of 'ecological rationality' (Gigerenzer *et al*. 1999), the 'decision environment' plays a crucial role in decision-making process due to cognitive constraints. Fiore et al (2009) demonstrated VR can generate sufficiently natural and familiar field, providing 'field cues' or 'field hints' occurring in real world. Individuals would react or behave to VR environment as if they are in the real world driven by the experimental stimuli after immersing the virtual environment, which have crucial effects on conscious and volitional behaviours of individuals (Sanchez-Vives *et al*. 2005). In pursuit of accurately presenting or describing complex alternatives or stimuli in experiments, VR techniques are gradually used as the survey technique in various disciplines to explore individual's perceptions or behavioural responses: tourism (Tussyadiah *et al.*, 2018), marketing (Loureiro *et al.*, 2019), environnemental policy (Fiore *et al.*, 2009) and economics (Innocenti, 2017) etc. For instance, to bridge the gap between laboratory experiments and field experiments, Fiore *et al.* (2009) proposed a replicable lab experiment in the natural 'look and feel' field domain using VR technology for wildfire management policy-making.

Some studies have also started to use SP experiments employed in VR environment to explore and to assess relatively complex scenarios. A couple of papers have compared text-based and VR-based SC experiment. Mokas *et al.* (2021) found that the presentation format has an impact on the WTP estimates for environmental elements in the streets and that the VR reduces the randomness in making choices. Rossetti and Hurtubia (2020) studied the ecological validity of VR experiment, but the focus is on the perceptions of public spaces. Patterson *et al.* (2017) in the context of neighbourhood choices, but found no significant differences between text-based and VR-based SC experiment, though they use a non-immersive VR experiment.

2.4.2 Virtual Reality in Transport

A number of recent applications have also used VR technology in transport-related research, particularly, parking behaviour, pedestrian and cyclists behaviours. For gaining in-depth insights into driver parking behaviour, Ben-Elia *et al.* (2015) exploited the 'ParkGame' serious game platform, where cruising takes place in virtual environments that incorporate a realistic road network, and priced on-street parking places and parking lots. Using the same 'ParkGame' platform, Fulman *et al.* (2020) explored drivers' two instantaneous parking choices in the city centre: when to quit cruising and where to cruise, while Geva *et al.* (2022) investigate the dynamics of the parking search behaviour under unique on-street price distributions. In order to understand the impact of distracted pedestrians' crossing behaviours, Sobhani and Farooq (2018) examined the importance of smart LED light safety treatment under three road-crossing conditions. Lovreglio *et al.* (2016), Arellana *et al.* (2020) and Feng *et al.* (2022) studied the pedestrian behaviour in the case of evacuation. Arellana *et al.* (2020), in the context of pedestrian crossing and evacuation, found that results from discrete choice estimation were more efficient using VR data than the traditional format, though more effects were significant in the traditional than in the VR survey. Birenboim *et al.* (2019) and Bogacz *et al.* (2021) studied instead cycling behaviour, controlling for car flow and geometric characteristics. These studies make use of SP experiments to account for attributes that are difficult to measure in a standard screen-based survey and to control for the characteristics of the environment (i.e. flow speed etc.) where the respondents perform a continuous behaviour (walk or cycle). This makes it more difficult to assess internal validity, i.e. the impact of the VR experience in the elicited consumer preferences, compared to a traditional SC screen-based survey. Pedestrian and cycling behaviours in fact involve continuous movement. In some studies, for modelling purposes, the continuous behaviour has been converted into a choice, but, from a neurological point of view, motor actions (like walking and cycling) activate different circuitries in users' brain compared to choice-based actions. Motor actions in fact show a good overlap between brain activities during imagined and real movements (see the discussion in Cherchi, 2020) , while the overlap disappears in the case of choice behaviours.

Recently, studies on VR techniques in autonomous driving has also been advanced (see literature review by Riegler *et al.*, 2021). VR technology, providing a safe and controlled environment, has been used to study the autonomous driving and experience from 12 application areas (such as vehicle navigation, driving behaviour, safety, user interface design, etc.). For example, VR was used to examine the enhancement of human-machine interface for AVs from the perspectives of different road users. such as testing passengers' in-vehicles experiences- entertainment and games (Wang *et al.*, 2016), backseat productivity work (Li *et al.*, 2020), etc. Farooq *et al.* (2018) and Velasco *et al.* (2019) studied pedestrian behaviour when crossing a street where a flow including normal cars and AVs (in different proportions) is driving through, controlling for the speed of the cars, the geometric characteristics of the street and the weather condition. There was much less research focusing on SC or SP experiment to investigate the preferences or travel behaviour in the choice of AVs using VR tools (to my best knowledge, only Farooq *et al.* (2018) (pedestrian preferences when crossing a street in the presence of AVs) and Djavadian *et al.* (2020) (driver behavioural responses to connected autonomous vehicles (CAVs)). The use of VR mitigates the disadvantages existing in text-based and image/video-based SC experiments and respondents have more consistent preferences with the increase of understanding on hypothetical situations (Farooq *et al.*, 2018). It seems that data collected with VR experiments have the potential to significantly improve the estimates of users' preference and users' acceptance for innovations compared to the survey methods currently used. But, there is as yet no research on this area.

2.5 Summary

Based on above extensive literature review, current studies have made considerable efforts in understanding the determinants of AV acceptance, preference heterogeneity in AV, enhancing survey techniques in investigating individual preferences on AVs. Four main research gaps identified from literature review were summarised as below:

- Although there are various studies relevant to the investigation of determinants in using and/or owning AVs, there is no evidence on the preferences and consequent adoption behaviours for ATs and none has studied in particular the impact of in-vehicle features in the choice of ATs. In-vehicle features might highly influence passenger's choice of ATs.
- While several papers have studied the impact of psychological constructs in the choice of AVs, very few have studied them jointly with the objective characteristics of the vehicles. Among these few, none have addressed specifically the three critical constructs (HM, trust and SI) in the primary appraisal in the individual's decision making process.
- The existing research has recognised the preference heterogeneity across countries towards AV acceptance. But, there is still little evidence. It mostly refers to psychological constructs and developed countries. This knowledge is important and useful for developing the country-segmentation strategies.
- The vast majority of the studies that measured users preferences for AVs characteristics, used traditional SC experiments. These suffer from 'lack of realism' problem. The VR technology can provide a high level of realistic environment, but very little research exists to study the extent to which VR affects the measurement of consumers' preferences for innovation. No studies have used SC experiments embedded in VR environment. There is nearly no studies in existing literature for guiding on this task, particularly in relation to the choices of ATs.

Chapter 3 Data Collection Methodology

3.1 Introduction

This chapter describes the process followed to design and develop the experiment to collect the data. The core of the data collection methodology is a stated choice experiment built to identify what affects consumers' choice between a normal taxi (NT) and a fully automated taxi (AT). But the overall survey includes also psychological statements to measure injunctive norms, hedonic motivation and trust along with several socio-economic characteristics as well as other information about current use of taxis and familiarity with AVs and ATs. In order to achieve the specific objectives of this PhD research the overall questionnaire and in particular the stated choice experiment was built in a way that the data collected online in the UK and in China could be compared and the data collected in the UK online and with the Virtual Reality (VR) environment could also be compared. Building a survey that allows cross-national comparison, cross-methodological comparison and cross-experience comparison, along with eliciting the importance of AT in-vehicle features, social conformity and awareness of AV was challenging.

In total, four datasets were collected:

Dataset 1: collected among people living in Newcastle metropolitan areas using a screen-based online survey. It is called from now on '*survey-online NCL*'.

Dataset 2: collected among people living in Newcastle metropolitan areas using an immersive VR-based survey for the SC experiment. Other parts of the survey were collected still in the lab but using a screen-based survey (i.e. questions about a recent trip by taxi, socio-economic/travel characteristics, as well as attitudes and/or perceptions towards fully automated taxis). It is called from now on '*survey-VR NCL*'.

Dataset 3: collected among people living in any cities in the UK using an online screen-based survey. It is called from now on '*survey-online UK*'.

Dataset 4: collected among people living in selected cities in China using an online screen-based survey. It is called from now on '*survey-online China*'.

The rationale for selecting these datasets is to answer the three specific objectives defined in **Chapter 1**, which are the cross-methodological, cross-national and cross-experience comparison. For the cross-methodological comparison, the choice of Newcastle upon Tyne as location, was motivated by the fact that the VR-based experiment requires participants to come to the laboratory that was physically located in Newcastle. In this case the data were collected from the same city (Newcastle) but with different methodology. For the cross-experience comparison, the choice of China as location was motivated by the fact that was one of the very few countries where ATs were in commercial or in trial operations, allowing for a comparison of the preferences' between taxis users who are from AT cities (i.e. where ATs are available for the public) and from NT cities (where ATs are not available for the public). In this case the data were collected from cities of similar dimension within China. Given these two datasets, the cross-national comparison was of course performed between China and UK. In this case, differences in socio-economic characteristics were controlled for, in order to ensure the validity of the comparison.

The same questionnaires was used in all these four surveys (both in the UK and China, both online and VR-based), with minor adjustments, as described in details in the rest of the chapter. The questionnaire was organised in five major parts, in the following order:

Part 1: Introduction to the survey, screen out questions, and questions to customise the SC experiment.

Part 2: SC experiment.

Part 3: Socioeconomic and travel behaviour information.

Part 4: Psychological statements.

Part 5: Post survey. Only in the *survey-VR NCL*.

Figure 3.1 summarises the main steps followed to build the survey and in particular the SC experiment. The initial steps were conducted in the Newcastle upon type, UK: these include three focus groups and some initial pre-tests for the '*survey-online NCL'*. At the same time, the SC online experiment was thoroughly tested in the VR-based setting and modified to ensure that it would be realistic and work properly also embedded into the close-to-realistic VR environment (based on a specific taxi rank in Newcastle as it will be discussed later) *'survey-VR NCL'*. The final survey was then slightly adjusted to ensure compatibility with the general Chinese context and general UK's context used to collect also the data in China *('Survey-online China'*) and in the UK *('Survey-online UK'*).

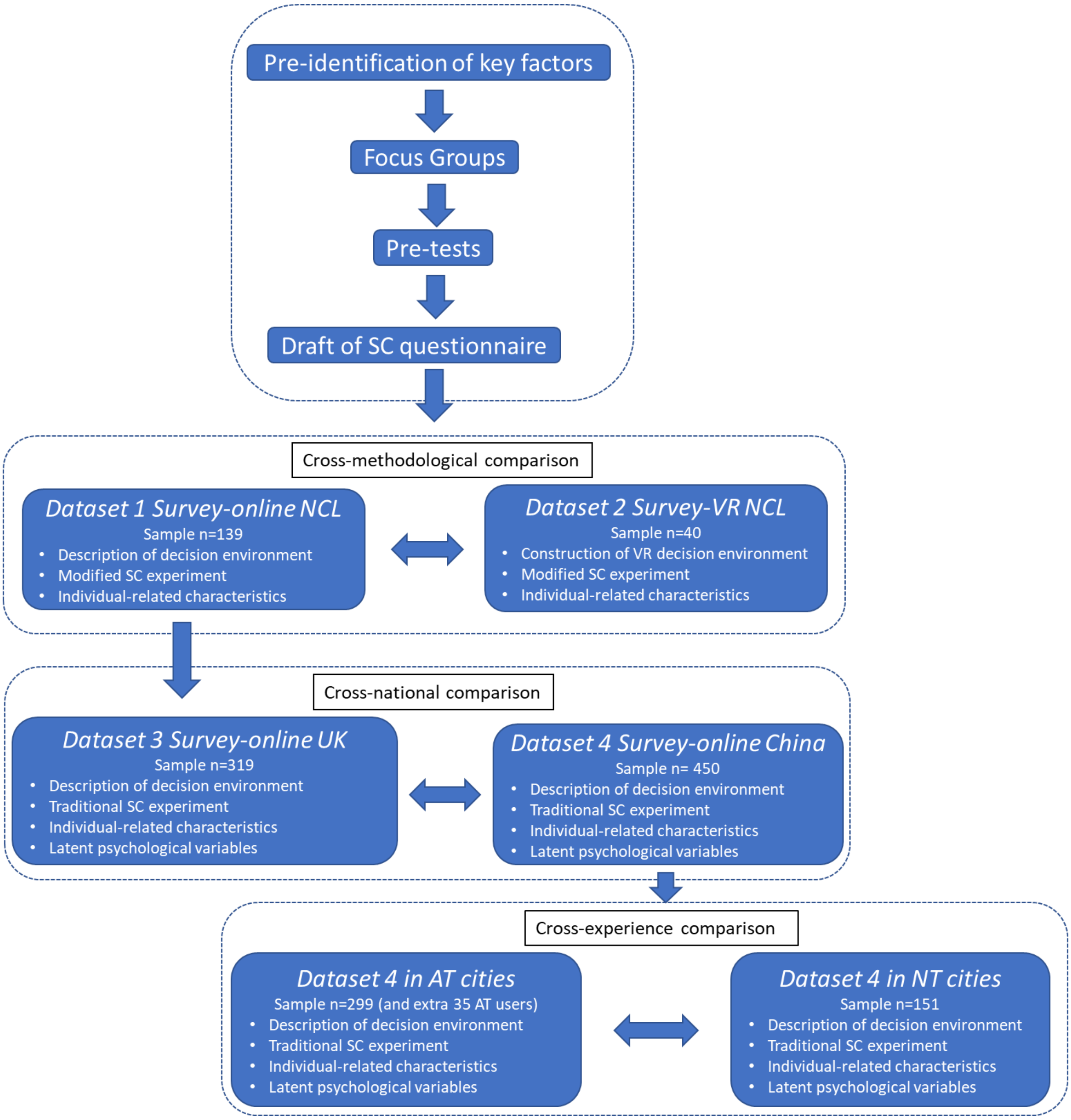


Figure 3.1 Process of development of four surveys

3.1.1 Rationale for the data collection methodology

As discussed in **Chapter 2**, several survey methods have been used to collect information about AVs and taxi services. The selection of the survey method to use depends on the research objectives. The methods can be qualitative, if the aim is to explore in-depth some aspects of the phenomenon or opinions, or quantitative, if the aim is to collect data to elicit users preferences for or users attitudes towards AV or AT.

The best approach (see a summary in Cherchi and Hensher, 2015) consists in using qualitative methods to inform on how to best collect (all or some of) the quantitative data. This is the methodological approach used in this dissertation.

Among the qualitative methods, the most common are by far focus groups and in-depth interviews to users or experts. Focus groups are used precisely to inform quantitative data collection, and this is why these are used also in this dissertation.

Among the quantitative methods, Revealed Preference (RP) and Stated Choices (SC)/Stated Preference (SP) surveys are the methods used to elicit consumers' preferences and willingness to pay for certain products or services. In the data collection conducted for this dissertation, SC experiments were selected for the following reasons:

1. RP surveys allow measuring real behaviours, which is a great advantage. However, AT services are still under the testing phase and a small-scale deployment of AT services around the world is not sufficient to recruit enough respondents for modelling purposes. In addition, users of this very innovative AT services are early adopters, and would not be representative of the preferences of the broader population, mostly represented by followers (Rogers, 1962).
2. RP surveys are limited to the existing alternatives and attributes, which some time are to measure and/or can lack enough variations in the trade-offs between alternatives (Louviere *et al.*, 2000). In addition, the dissertation entails an investigation of some novel attributes like in-vehicle features and social conformity attributes, which cannot be measured with RP surveys due to their non-existence in the current scenarios.

Nevertheless, as discussed in **Chapter 1**, there are also limitations or weaknesses associated to data collected with traditional SP experiments, mainly due to hypothetical bias (Haghani *et al.*, 2021a; Haghani *et al.*, 2021b). This motivates the innovative proposal of this dissertation to test the use of immersive VR technology to collect the data. Finally, as far as I am aware, the only method available to measure attitudes is by statements using a Likert scale.

The remaining of this chapter is organised as follows: **Section 3.2** describes the selection of the context of reference for the three surveys. **Section 3.3** describes the Focus Groups (FG) carried out to explore in depth the attributes identified from the literature review and the main results achieved. Based on these results some attributes were included as Pre-SC information (and discussed in **Section 3.4**), other attributes were instead included as part of the SC experimental design (and discussed in **Section 3.5**). **Section 3.6** describes the attitudinal questions defined to measure the latent psychological constructs identified as relevant for this research, while **Section 3.7** reports the work done to build the VR environment with the SC experiment embedded. Finally, **Section 3.8** briefly describes the additional information included in Part 1 1 and in Part 4 of the questionnaire, while **Section 3.9** describes the sample recruitment and **Section 3.10** summarises the chapter.

3.2 Context of reference

Identifying a realistic context of reference is a relevant task in any surveys and in particular in any stated choice experiment. In this research, the task for the UK experiment was complicated by the need to build the same (or a consistent/comparable) SC for the online-based SC experiment and the VR-based SC experiment. The SC experiment set up in this research consists of a choice between a conventional taxi (with driver) and a fully automated taxi (no driver and no steering wheel). To ensure realism, it is always recommended to ask people to describe a recent trip, in our case a trip made by taxi, and then customise the SC experiment around this trip. In the online SC experiment, any trip by taxis can be considered. Even if pictures or videos of ATs are presented to respondents, these do not have to refer to specific locations. Indeed images or videos often picture imaginary cities. In the case of the VR-based SC experiment, the definition of the precise context is instead critical for the experiment, because respondents will make their choice within the virtually real environment. The first step then consisted in identifying a realistic context where passengers have the possibility to see the taxis operating while they are making their choice. This was a first challenge and the only solution was to locate the experiment at a taxi rank. Despite the diffusion of car-hailing and private hire services, traditional taxi services are still extensively used in Newcastle upon Tyne, England and there are numerous taxi ranks in the city centre. In China, the experiment was run only online, the selection of the context of reference did not add additional problems compared to normal surveys.

3.2.1 Selection of the context in Newcastle upon Tyne

As mentioned before, a taxi rank in Newcastle was chosen as context of reference. This was due to the need to run a comparable experiment screen-based and VR-based. A thorough analysis of the taxi ranks located in Newcastle city centre was conducted. First, the following 6 possible taxi ranks were selected and compared:

1. In Northumberland Rd between Primark and Santander bank
2. At Newcastle Central Station, Neville Street (Busy traffic)
3. In Newgate Street
4. In John Dobson Street, near Newcastle City Library
5. Between Haymarket Metro Station and M&S
6. Percy Street near Tesco Express.

Pros and cons of each location were analysed and compared based on the following criteria:
Requirements for the specific SC experiment:

1. Being very familiar to many possible respondents.

Requirements for the VR environment:

2. Simple buildings
3. No trees, straight road
4. No long horizons
5. Walking distance compatible with the space available in the room where the experiment will take place
6. Enough space to allow taxis standing in 2 lines, one for the NTs and one for the ATs.

The taxi rank (N1) located in Northumberland Rd. was selected as the final VR scene location, as it fulfilled all the above criteria. Northumberland Road is located adjacent to a major pedestrian shopping street in Newcastle and it is very well-known by locals. The taxi rank located there is used by more than 1,000 people a week. **Figure 3.2** reports some pictures of some possible taxi ranks locations analysed (from left to right, location N2, N3 in the first row, N4, N5 in the second row), while **Figure 3.3** reports two pictures of Northumberland Road, which as we see fulfils almost all the above criteria. There is a long horizon, but this is far from the point where the respondents will be located, which allowed us to use a low definition for the buildings far in the horizon. The street is not very wide but there is enough space to allow taxis ranks and their movements. The blue sign in the pictures is the sign of the current taxi rank.

Figure 3.2 Location of some of the taxi ranks considered in Newcastle

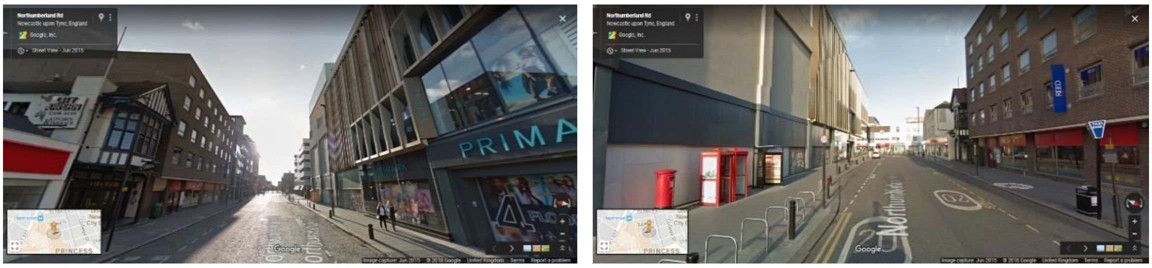


Figure 3.3 Street view of Northumberland Road in Google Maps

3.2.2 Selection of the general context in the UK and China

With respect to context of reference in survey-online UK, as mentioned previously, the traditional taxis are still commonly used at taxi rank, thus the context of reference for the survey in the UK was still a taxi rank, but not a specific one as in Newcastle but any taxi rank.

In survey-online China the experiment was run online, hence the context of reference did not present major problems. In China, app-hailing taxis and traditional taxis are two main taxi commercial modes. Despite the diffusion of app-hailing taxis (like DiDi), traditional taxis in China still dominate the taxi economy, accounting for 68.6% of total transactions according to analysis of China's taxi passenger volume and market size in 2020[5]. Traditional taxis in China are a type of cruising taxis, which means taxi users can hail a taxi both on the street (empty taxis can temporarily stop on the street to pick up passengers or taxi passengers can pre-book the taxis) and at taxi rank. Apart from major transit hubs (e.g. airport, train stations, etc.), nowadays the taxi rank are no longer a common place for Chinese taxi passengers to use

[5] See: http://www.huaon.com/channel/trend/739739.html [In Chinese]

traditional cruising taxi in their daily life. Therefore when defining the context of using taxis in China, the situation where a passenger hail a cruising taxi on the street was referred for designing the SC experiments.

3.3 Focus Group

The Focus Group (FG) is a qualitative survey consisting of interviews in which a selected small group of people participate in a planned discussion intended to elicit consumer perceptions about a particular topic or area of interest. Unlike direct questions or one-to-one interview, interactive group discussion allows to elicit more information by a process of sharing and comparing (Morgan and Krueger, 1998). This is why the FG allows going deeper in exploring particular aspects of the phenomenon and helps understanding the complexity of travel behaviours (Clifton and Handy, 2003; Thomas *et al.*, 2022). FGs were used in this research as a preliminary step to the SC experiment, to gather a better understanding of the attributes and their levels influencing passengers' choices between ATs and NTs and to identify the best way to present these factors.

Based on the literature review discussed in Chapter 2, **Table 3-1** summarises the list of different characteristics that have been found relevant in the choice of taxis, autonomous vehicles and social conformity. Since the survey was intended for the VR experiment, it was also considered if the variables were feasible and realistic for the VR environments. Not all the attributes identified from the literature review were relevant and/or feasible. In this dissertation, the focus is on the pre-trip choice, i.e. when travellers take the decision to take a taxi. Some factors, such as seat quality, level of noise, etc., which are not strongly related to characteristics of ATs, were filtered out. The characteristics of the taxi rank or the point where taking the taxi are important, but there is no reason to assume that these characteristics would be different for ATs and NTs, unless there is a specific interest for studying this aspect.

		Possible factors
Taxi	Vehicle	Taxi brand, Cleanliness, Car condition, Car model, Sitting position, Safety
	Driver	Welcoming, Appearance or dressing, Knowledge of direction, Communication during the trip
	Level of service	Waiting time, Travel time, Travel cost, Payment method
AVs		Safety, Vehicle control, Ease of use technology, Privacy
Social Conformity		Number and type of people in queue, Customer reviews, Information received, Who gives the information, How the information was provided

Table 3-1 Pre-identified key factors related to the taxis, AVs and social conformity

On the other hand, since in the VR part of the experiment (described in **Section 3.7)**, respondents can 'live' the environment and see how the new taxi system operates, it was important to discuss the exterior and interior features of the ATs, as wells as car conditions and car models and the sitting position (back or front seats). All factors related to the choice of buying an AV, including the different levels of automation, were not considered either in the analysis.

Finally, given the aim of this dissertation, the in-vehicle features represented also key factors to be tested. These refer to in-vehicle communication forms with the AT operator, and social interaction with the driver/AT operator (to communicate the destination, to get the price, to simply chat and to pay).

The characteristics identified in **Table 3-1** and the in-vehicle features were then analysed in detail in a series of focus groups[6], with the specific objectives of:

1. discussing the important objective factors which can influence the choice of ATs versus NTs. First the key attributes identified from the literature review, and then other objective factors raised by the participants;
2. identifying individual's opinions related to the potential use of ATs;
3. identifying how or in what possible ways information like safety, social conformity might affect participants' choices of taxis or innovative products (i.e. to find a more realistic way to present these information);
4. investigating the best way to present these factors according to respondent's understanding and experience.

3.3.1 FG survey methodology

Based on the above 4 specific objectives, the skeleton of the FG discussion was designed and

---

6 Focus Groups were run in collaboration with the team of the Veronica project, financed by ESRC, UK. Veronica project aims to develop and employ the virtual reality experiment to understand the public acceptance of fully autonomous vehicles. This PhD shared the work for the Focus Groups with the Veronica project, but the SC experiments built are different as well as the samples collected. In addition, the SC experiments used in this dissertation were built before those used in the Veronica project. More details about Veronica project can be found in: https://www.veronicaproject.org/.

organised in three parts (see **Appendix I**):

Part 1: experience and opinion in general about AVs (i.e. without driver) (include 3 questions).
Part 2: experience about AT and first general opinion about ATs (include 4 questions).
Part 3: what factors are considered relevant if you need to take an AT (include 13 questions).

The FG were run only in the UK for practical reasons. In the UK three FG of one-hour each have been run: two on 4th July 2018 and the third on 24th July 2018. The first 2 FG were organised at the Institute for Ageing, Newcastle University. Participants were recruited on a voluntary base among people who participate to the group activities of the Voice North Research Group. They did not receive any incentives. There were 6 attendees (3 males and 3 females) in the first focus group (FG1) and 7 attendees (2 males and 5 females) in the second (FG2). Basic socio-economic characteristics were asked at the end of the FG. Participants were mainly elderly, and with a high level of education. A member of the Institute of neuroscience moderated both FGs, using a set of slides as a guidance (see **Appendices 1**). The author primarily introduced the project background and aim, and then acted as observer, along with my supervisor, and took notes during the discussion. The FGs were recorded (only the voice), permission was asked to the participants, before starting the registration.

The third FG was instead organised after 3 weeks, and more importantly after analysing the results from the first two FG. Based on these results, the author made minor changes in the description of some questions but the major body of the slides stayed the same. The third FG was run at the University premises and it was moderated by a colleague expert in FG. My supervisor and the author acted as observers as before. Participants were recruited via a panel available at the Institute of Neuroscience. 7 people (4 males and 3 males) attended the third (FG3). One person had a last minute problem and could not participate. The participants in this third FG were mainly young, but apart from that, it was mostly a heterogeneous group in terms of age, status, education and other characteristics. As a reward, the recruited participants were informed that they would receive a gift cards at the end of FG. The FG was recorded in the ways of both audio and video. Before starting the registration, the consent of the participants was obtained.

The transcripts of FG discussions were analysed by an expert in qualitative surveys[7]. There was no significant difference among type of participants (i.e. among the 3 FGs) in terms of their opinion about AVs and all respondents showed quite good knowledge of what AVs are and the different level of automations

3.3.2 FG main results

The most relevant results from the FG are the following:

Level of Service

The FGs confirmed that all the level of service attributes listed in **Table 3-1** are relevant and these were all included and tested in the SC experiment. Payment method in particular attracted participants attention during FGs. It mainly involved two points: the form of payment (e.g. card, cash, etc.) that can be used for ATs (i.e. someone reported that *"…would [it] accept cash or card? If I knew where I was going, and they would accept a payment card, I would get that…"* others reported *"I normally pay in cash.")* and payment time (i.e. if AT fare should be paid in advance because someone worried *"Like if there is no one there to take the money off you, what's going to stop people from just jumping out not paying")*

Driver

Among the factors related to the *driver,* communication forms appeared to be the most important for the participants. The other attributes discussed (welcoming, appearance and knowledge of direction) were not considered key priorities. During the FGs participants were presented with the options in **Figure 3.4** and were asked to indicate which form of communication they prefer, and a discussion was then open on the reasons of their choice. The presence of a button was considered generally relevant, while, surprisingly participants expressed concerns about relying only on the app, as well as on the reliability of voice control (e.g. recognising different accents, etc.). Several participants asked why all three options could not be used, which seems to reflect some anxiety about being in a car without a driver. The discussion in the FG also confirmed that lack of human interaction, i.e. the possibility to interact with the driver was a relevant factor. In particular elderly members of the FG highlighted that

---

[7] This is in order to have a neutral point of view, as the researchers might unconsciously analyse the data searching for confirmation to their initial positions. The analysis prepared by the external expert was validated by the researchers based on the notes they took as observers during the FG.

in the AT they would have missed even just chatting with the driver during the trip, even if they have no specific request to the taxi driver *("I think one disadvantage of it, when you get a taxi, normally they say, "have you had a good day? ... You can talk to them.").* This is a relevant information for the development of the AT service and it was then selected to be tested in the SC experiment.

Figure 3.4 In-vehicle communication forms with the AT operator

The other in-vehicle feature selected to be tested is the possibility to change destination. This prompted from the FG open discussion as a participant said *"What if you were there and you suddenly took ill, or felt ill and ... you couldn't stop it and couldn't change the route (...) how quickly will you get any kind of response? A taxi driver would immediately.*" While changing the destination (even simply drop off a bit earlier or in a particular point close to the destination) is very common and easy to perform in a AT, the lack of driver can potentially limit flexibility during the trip. It is important to know if communication-related devices and a hidden driver or an operator are needed and to which extent the lack of this in-vehicle feature might affect the use of the AT services.

AVs & Taxi vehicle

Among the attributes related to the vehicle (to the taxi and the general AV), results confirmed that safety is indeed an important factor, though different participants were concerned by different aspects of safety (e.g. car safety, personal safety and privacy). The critical issue here is that information about safety can be understood or interpreted differently by different respondents and this is likely to have an impact on respondents' choices. It was discussed in depth whether to include safety as an attribute in the SC experiment. The major problem discussed in this case is that if safety is an attribute in the SC, its value (whatever this is, depending on how the safety is specified) will change among scenarios. This might be perceived as not realistic by the respondents because in each scenario they are asked to imagine that this is another day (that can be also the day after the previous scenario) where they have to take a taxi. It might not be realistic that safety will change over such short period of time (safety measures typically are aggregated yearly or monthly values). At the same time, safety information is extremely important for AV and needs to be provided to respondents. It was then decided to present them as pre-information, before respondents will start the SC experiment. The same information was presented to all participants and was not varied among tasks, not among participants. A thorough analysis was carried out about which information to provide, how to provide it and form which source should the information come from. This discussion is reported in the next section. Car conditions were valued relevant but not top of the list. Moreover, most of the car conditions are not specific features of the AT, hence less relevant for this study. The model of the AT was discussed extensively during the FG, because automated vehicles can be like normal cars, or having a distinctive model. Results showed a general preference for a normal size of AT and with the appearance of a normal car (for example, someone responded *"Well, I'm personally very fussy about my cars, and I certainly wouldn't have that one ('distinctive model') on the right there*"). This attribute was then not tested further in the SC experiment, where a normal car was used also for the AT. Interestingly, instead participants in the FGs showed a clear preference for an AT without steering wheel, mainly because it looked like what participants expected an AT to be *("I would choose A without steering wheel if it was going to be autonomous"*). As for the sitting position, there was not a clear preference (someone responded *"I always like to sit at the front"*. Others responded *"I'm just very used to sitting in the back of a taxi."*)

Other attributes, such as taxi brand, cleanliness, ease of using new technology, driver's welcoming were discussed but finally not included in the SC experiments for different reasons: cleanliness, driver's appearance or dressing are not easy to define objectively; taxi brand involves other aspects like loyalties, marketing that would require a specific dedicated research; ease of using new technology should instead be measured via psychological statements.

Social Conformity

Initially, the goal was to measure social conformity manipulating the number and characters of the passengers queuing to take the taxi. For this reason, the discussion in the focus group pointed at exploring which type of people respondents would notice in the queue and whether this would have an impact on their decision to take the taxi. The discussion was not conclusive, because no specific category appeared to be associated with taxi choice in both positive or negative way. Given this result, the author decided not to focus on the type of people but only on the number of people queueing, which is a measure of descriptive norms.

3.4 Pre-SC information

Based on the results from the FG, three types of information were provided before the actual SC experiment: (1) car safety information, i.e. safety of the AT circulating in the traffic, (2) general information about private safety, privacy and routing information and (3) specific information about how the AT operates and some optional and compulsory in-vehicle features once on board before choice tasks.

3.4.1 AT safety information

Significant effort has been devoted to define the *car safety* content and format. As mentioned before, it is important and necessary to provide respondents with safety information in order to give them a common background. At the same time information should be as objective as possible, to avoid affecting positively (or negatively) individual preferences, and as close as possible to how respondents would get the information in real life. Several options were tested asking respondents opinions about (1) media channel (i.e. who reported this information and where it was reported), (2) source of information (i.e. which institutions or organisations investigate and report taxi safety information) and (3) type of safety (e.g. accident rate, fatalities, injuries by type, total versus relative numbers). **Figure 3.5** a) shows the information provided to UK respondents. The same information was presented also to Chinese respondents, translated and adjusted for the Chinese context. **Figure 3.5** b) shows the safety information adjusted for the Chinese context.

In order to define the safety pre-information, two pre-tests were run among 22 friends and 31 taxi users in the UK, with a set of questions, aiming at clarifying the definition of car safety (see **Table 3-2**). The vast majority (60.9%) of respondents indicated the national-level media (e.g. BBC news) as the most likely channel to get safety information about ATs and the national-level government agency as the most trustworthy source of safety information (e.g. Department for Transport). To define the safety type, after reviewing the literature on willingness to pay to reduce accident (e.g. Rizzi and de Dios Ortúzar, 2003; Rizzi and Ortúzar, 2006) information from Vehicle Safety Report for current AV tests (e.g. Tesla Vehicle Safety Report https://www.tesla.com/VehicleSafetyReport) it was finally decided to present the information in terms of "accidents recorded per miles travelled by ATs compared to the NTs". According to Department for Transport (2015;2019), 5,359 taxi-related accident recorded per year and 3,396,750,600 miles travelled by taxis per year, that is one accident recorded every 633,840 miles travelled by NTs. The accident rate for ATs was assumed to be half of the accident rates for NTs. This value has to reflect the scenario of full adoption of AV but it also aims to make respondents feel safe to use ATs. To increase realism, we also carefully design the layout of safety information to be the same as the official BBC news layout[8].

Figure 3.5 a) Car safety information in the UK

Figure 3.5 b) Car safety information in China (translated from Chinese)

In the Chinese context, it was decided to present the safety information as news in the CCTV (China Central Television)[9] reporting results provided by the Ministry of Transport of People's Republic of China. As Chinese accident rate of ATs was not found, taxi accident rate was then computed taking as reference the UK's taxi accident data and travel mileage by taxis, adjusting it based on the Chinese accident rate for normal cars. The accident rate for ATs was still assumed to be half of the accident rates for ATs. Additionally, for choice task, the attributes and layout were kept consistent with the UK counterpart.

[8] See: https://www.bbc.co.uk/news

[9] See https://english.cctv.com/news/index.shtml?spm=C69523.PDoRdCIUTBov.EBfl1JN80NdJ.3

Q1 If you would like to receive road safety information, which of following road safety institutions or organisations would you like to get the information from?
National-level government agency or institutions (e.g Department for Transport)
Regional-level or local level government agency or institution (e.g. Newcastle City Council, Northumbria Police)
University research (e.g. Newcastle University)
Non-profit transport-related organisations (e.g. Chartered Institutions of Highways & Transportation (CIHT))
They are all the same
Other, please specify
None, I do not care
Q2 If you were to read about the road safety information from the media, which of the following media sources would you like to receive it from?
National-level media (e.g. BBC, Guardian)
Regional-level media (e.g. Heart North East, Northern Echo)
Local-level media (e.g. Newcastle Chronicle)
They are all the same
Other, please specify
None, I do not care
Q3 which of following expressions about taxi safety is easier to understand for you?
Number of fatalities by automated taxis per year (e.g. 5 fatalities per year)
Number of fatalities caused by the automated taxis per million km travelled (e.g. average 4.6 fatalities caused by million km travelled)
Number of fatalities per 100,000 registered automated taxis (e.g. average 2 fatalities per 100,000 registered automated taxis.)

Table 3-2 questions for defining car safety

3.4.2 Personal safety, privacy and routing information

An equally thorough analysis was conducted to check the personal safety, privacy and routing information. Pre-tests results confirmed the far majority of respondents indeed care about the personal safety (risk of personal crime, etc.), privacy (location data or personal data) and routing information. After several tests, it was then determined to define them as:

Personal safety: all automated taxis are equipped with a 24h security surveillance camera and a 'SOS' button.

Privacy: your trip location data or other data will not be recorded when using automated taxis.

Routing information: automated taxis provide an optimal route to your destination according to a computerised algorithm.

Figure 3.6 shows the information provided to UK respondents. The same information was presented also to Chinese respondents, without changes just translated into Chinese.

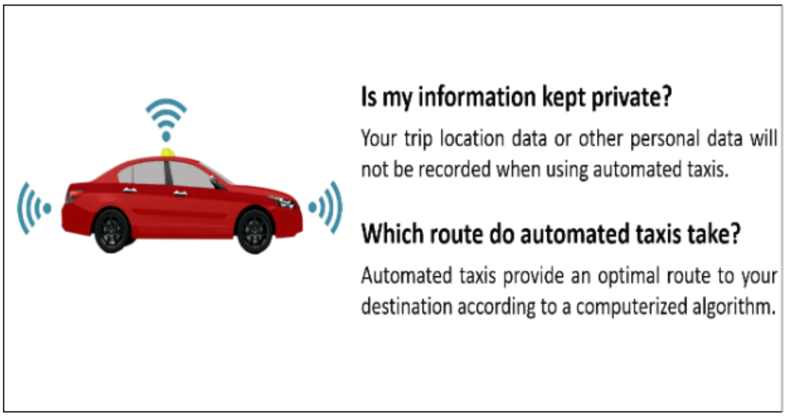


Figure 3.6 Privacy and routing information

3.4.3 Information about how the AT operates

In addition, during various pre-tests ran, it was found necessary to supplement the information about the 'in-vehicle features' and 'how to use an automated taxi'. It is consistent with our real-life experience that some instruction information should be provided when experiencing something new.

The majority of the SC experiments involving automated vehicles have opted for detailed descriptions about what it is possible to do within an automated vehicle with often images (also virtual reality images) or videos featuring how vehicles can be used (e.g. Krueger et al., 2016; Kolarova et al., 2019). Providing information is particularly important in case of automated vehicles, as these are relatively unknown to the majority of the population, who certainly have no experience with them. In this case, however, it was decided not to include this information not any image or video for two main reasons. First, the experiment features a choice only between taxis, then there is no difference in the type of activities that can be performed within a normal (i.e. with driver) and automated (without driver) taxi. Second, even though images and videos help familiarising with the innovation, they are rarely neutral and it is likely they have a priming effect on individual preferences. Finally, since this research involves also a comparison with a SC embedded in a VR environment, the choice was made not to include videos.

Figures 3.7a) and b) show the information provided to UK respondents. The same information was presented also to Chinese respondents, without changes just translated into Chinese.

For the Automated Taxi:

a) The AT will stop exactly at the fixed pick-up point
b) Open the BACK door as you do in a normal taxi and sit inside
c) Fasten seat belt, scan the QR code to confirm the order information and press 'start' button
d) The AT will drive you to fixed drop-off point near your destination selected

Figure 3.7 a) How to use ATs

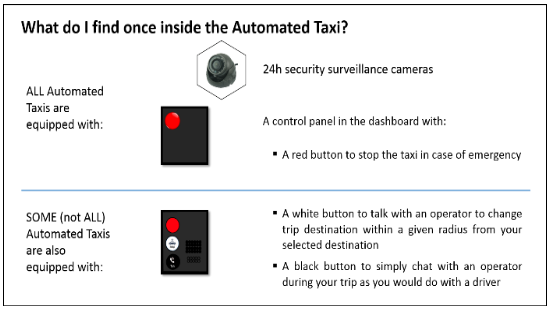


Figure 3.7 b) What to do within the ATs

3.5 Stated Choice experiment

The choice experiment set up for this research consists of a binary choice between an AT and a normal taxi. Based on the results of the FG, a first stated choice experiment was built using an orthogonal design, as priors for the coefficients were not available. A series of 4 pilots were run in the UK (with 45 friends and university staff in the first 2 pilots and 66 taxi users from a panel provider in the last 2 surveys) before getting to the final version. Other 2 pilots were run in China (with 99 taxi users from a panel provider) to test the adjustment of the UK survey to the Chinese context.

The final experiment includes 7 attributes, three attributes refer to standard level of service attributes (waiting time, travel time and fixed journey fare), the other 4 attributes have been specifically designed to test specific features available inside the AT (talking with an operator and changing the destination) and to measure the impact of social conformity (the number of customers in the last hour and the customer rating yesterday).

A significant effort has been devoted during the pilot tests in designing the layout of the tasks in order to present it in a way that looks as realistic as possible. In particular, worth noting that the layout of task is not designed as a traditional table style but the format is similar to the one used in reality in the ticket machines. The use of images, not just textual description, to present the in-vehicle features and customer reviews, is also inspired to current experience.

Figure 3.8 reports an example of the task presented. The same format and attributes were used for the Chinese experiments, translated into Chinese. The values of the attributes (the levels) differ in the UK and China. These will be discussed in **Section 3.5.2**.

Figure 3.8 Example of choice task presented

3.5.1 Attributes definition

This section provides a critical description of the attributes tested in the SC experiment, and the work done to define them. Due to the fact that the same SC experiment was implemented both online and embedded in the VR environment, the way attributes were defined and presented required extra attention. In particular, during the tests it was discovered that some of the attributes commonly used in the online SC were not realistic in the VR environment. The definition of the levels will be discussed in **Section 3.5.2**.

Level of service attributes

Fixed journey fare

Travel cost, namely taxi price, is the monetary cost that users need to pay for their trip by taxi. Taxi price is a key attribute influencing customer's choice, it is also critical to compute willingness to pay and user benefit, as such it must be included in the SC experiment. The definition of the taxi price in the design is in principle not complicated, because respondents are used to pay to use the taxi. However, the problem consists in that typically the payment in the taxi occurs at the end of the journey and it is based on the actual distance travelled and time spent on board. In the SC experiment instead respondents are presented with the taxi price before the trip and asked to make a choice based on this information, which in real trips can change at the end of the trip. This of course causes uncertainty in the monetary cost, which is not desirable, given the key role of this attribute and its coefficient as economic measure. From one side, it can be said that the attribute has to be intended as 'expected price", which can change and this is consistent with real booking. However, in this specific experiment, the problem was complicated by the fact that the survey will also aim to measure the preferred form of payment and if this is different between AT and NT. As it will be discussed later, in the AT the only feasible form of payment is ‘pay in advance’ (i.e. before the actual trip will take place. This is because taxi passenger can possibly jump off the taxi after arriving at the destination and do not make any payments). In this case, the taxi price cannot change. Hence, the taxi journey fare was explicitly defined as ‘fixed’ for both alternatives (NT and AT), i.e. the fare displayed in each scenario does not depend on the actual travel time experienced on-board, even if respondents decide to change destination as allowed in the experiment (change destination is another SC attribute discussed below). Pre-tests confirmed that this attribute had a significantly negative effect.

Travel time

Travel time denotes the time users spend travelling within the taxi, from the origin to the destination. This is also a key attribute for the choice of taxi and also not difficult to include in the SC experiment. In any booking, the travel time displayed represents an ‘expected travel time’ and this is also how it was defined in this experiment. Differently from the taxi price, it is not realistic to define a fixed travel time. Uncertainty is then implicitly associated to this attribute. However, this applies equally to AT and NT and no information was given to respondents that could suggest that AT could experience different (shorter or longer) travel times than NT. As reported in the pilot tests (**Appendix 2**) it was also discussed the possibility not to include travel time in the SC experiment, because of the above uncertainty. But, given the importance of travel time, it was then decided to include it and make it generic between AT and NT, without assigning any a priori lower value to AT travel time with respect to NT.

Waiting time

Waiting time denotes the time users spend waiting at the origin of their trip. This is a key attribute affecting taxi user’s choice and, in the AV literature, it is typically included in the SC experiments as the time subjects would spend waiting outside the AV (Krueger *et al.*, 2016; Yap *et al.*, 2016). The problem arises for the VR-based SC experiment because respondents can see people queuing and the taxis departing, so they might have their own judgment about how long they should wait. Three possible options were discussed: 1) ’expected waiting time’ 2) ‘waiting time as a function of number of people queuing and the number of taxis departing’; 3) ‘perceived waiting time’ Option 1 is the option commonly used in the online choice task. Option 2 is more realistic but each respondent might associate a different waiting time to the same queue length, or also to queue lengths within a given range of people. There is no possibility for the modeller to know which waiting time each respondent associates to the number of people presented in the choice tasks. Option 3 is probably the most realistic one, because users normally judge the waiting time according to their personal knowledge and experience. This option can reflect the difference between traditional SC and VR SC because respondents normally cannot experience 'waiting' when they are completing the standard SC choice task. However, respondents experience the waiting time only after they have booked the taxi (i.e. made the choice in my case). Instead of using the experience to define the waiting time, this effect could be used to make respondents "pay" the consequence of their choice, this option could dramatically increase the experiment time. Finally, it was decided to use option 1 to define 'waiting time' in SC experiment. Pre-tests showed that this attribute had an expected effect (i.e. significantly negative).

In-vehicle features

What you can do during the trip (1: Change the destination and 2. Chat with an operator)

These two attributes are specific for the AT option and refer to the possibility offered to the passenger while travelling on board the AT to change the destination selected before starting the trip and the possibility to chat with an operator as they would do with a normal taxi driver. As discussed previously, AT services ought to be designed to satisfy various taxi passenger requests which in a normal taxi are dealt with a direct communication between the passenger and driver (Kim *et al.*, 2020). Although the in-vehicle feature refers to something that passengers experiment within the vehicle, this is necessary information that passengers should know before they board the taxi. The definition of these attributes required several tests. While the above two attributes (change the destination and chat with an operator) were tested specifically within the SC experiment, other in-vehicle information (about how to operate the AT) needed also to be provided to respondents. In the initial tests, all in-vehicle features were presented together with detailed images and textual description, but this was found to be incompatible with the resolution allowed in the VR environment. Information was then simplified but still the majority of respondents during VR pre-tests reported that it was difficult to read so much information reported all together. Details of all the tests are reported in **Appendix 2**. It was then decided to divide the in-vehicle feature into two parts. The features that do not change across SC scenarios, were presented in the pre-information (as described in **Section 3.4, Figure 3.7**) while in the SC scenarios were included only the two in-vehicle features (change the destination and chat with an operator). Several layouts were tested before arriving to the final configuration reported in **Figure 3.9.** This was further tested, and none of respondent reported difficulty in understanding and reading this information in both VR and online survey.

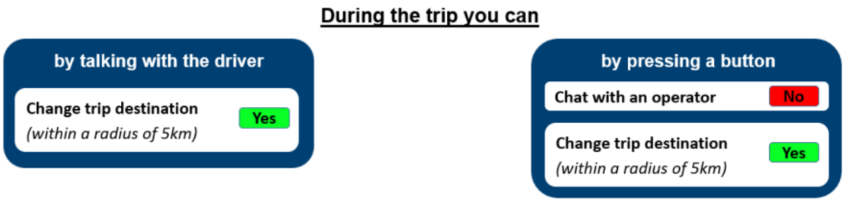


Figure 3.9 In-vehicle attributes in the SC scenarios

Social conformity

Number of customers in the last hour

As discussed in the FG, initially, social conformity was initially defined and tested as *number of people in the queue (i.e. in the queue to take the taxi)*. The reason was that the queue was mentioned in the FGs and in the VR it is possible for participants to see the number of people queuing ahead right when they are making the choice. Descriptive norms are expected to have a positive effect on respondents' choices. However, when tested, the effect of the attribute *number of people in the queue* turned out to be negative, which is different from what is expected according to the psychological theory of social conformity. However, in this case, it could be that the number of people ahead (in the queue) was perceived as correlated with waiting time (i.e. more people queuing, more time a customer needs to wait to get a taxi). The model was estimated with the interaction between these two attributes, but the interaction was not significant and the linear effect of the descriptive norm attribute remained negative. Following previous research, (e.g. Rasouli and Timmermans, 2013b; Cherchi, 2017), it was then decided to use the *number of customers in the last hour*. The choice to use the number of customers specifically *in the last hour* as a period of reference was based on the consideration that each SC scenario represents a different day when the respondent takes a taxi. Given the nature of the service, a short time reference (todays or shorter) was considered more realistic, because the number of customers change among scenarios, and each scenario represents another day where the respondent takes the taxi. 'Today' is relative period reference, not absolute period reference. This probably causes the situation where respondents perceived different time frames depending on when respondents participate in the survey. For instance, respondents who join the survey in the morning may perceive 'Today' shorter than those who join the survey in the evening, causing the difficulty in defining the realistic attribute values. Given these considerations, it is then more realistic "the last hour" was chosen as period of reference.

Customer rating yesterday

Customer reviews or ratings, representing general public opinion or a form of word of mouth, are increasingly used in reality as a form of social influence. It is common for example in ride-hailing taxi Apps like Uber, that passengers evaluate the drivers, but in transport studies the impact of customer reviews has rarely been studied. There is significant evidence suggesting that customer reviews are a critical factor for customer decision making. Indeed, this has become also very popular in all websites. However, how customer rating is actually presented is also critical. Cosley *et al*. (2003) investigated users' satisfaction, rating consistency, and recommendation accuracy when rating movies under three different scales: a binary scale, a ±3 scale with no zero, and a five-star rating scale with half-star increments. They found that users like the five-star scale best, and they found evidence suggesting that as scale granularity increases, recommendation accuracy increases. Sparling and Sen (2011) evaluated four types of rating scales and concluded that users prefer the five-star scale overall, although the thumbs scales come in as a relatively close second choice for product reviews. Chen (2017) comparing different rating systems, found that the five-star rating system allows cognitive fit (match between task, problem representation and individual problem-solving skills) which increases perceived information quality and decrease cognitive decision efforts. Based on this evidences and with the aim to increase realism, we decided to present the customers review using the 5 star system. We also specified that these reviews refer to "yesterday" customers, to make it realistic for respondents to see different customer rating in different scenarios, as each scenario represents another day. Pre-tests confirmed that this attribute had a significantly positive effect on the choice of taxis, which is consistent with what we expected. This reference period used for the customer review is based on consideration that it typically requires some time for the customers to write a review and for the provider to process them (though the stars can be counted automatically). This is why it was felt that 'reviews from previous hours' could be perceived as non-realistic by the respondents. The number of customers instead can be counted immediately, hence the definition 'number of customers in the last hour'. Additionally, using the same reference period for social adoption and customer review there is the risk that respondents might consider these two attributes correlated (intuitively, that the number of customers can depend on the average customer ratings). Different reference periods used for these two attributes can reasonably avoid this problem.

Other attributes

Payment method

Payment method refers to how respondents can pay for the cost of their trip by taxi, (cash, CC, App) and when they pay (in advance or at the end of the trip). Based on the results from the FGs, initially it was decided to test two attributes with two levels each. The first attribute was the *type of payment*, with levels "pay cash" and "pay with major credit/debit cards", as these were considered the most common methods used in practice. The second attribute was *when to pay*, with levels "pay before the trip" and 'pay to taxi driver' for the normal taxi and fixed to one level ("pay before the trip") for the AT. It is important to mention that paying to the driver is currently the most common option, as taxis have typically a taximeter. However, paying in advance for a taxi is not unrealistic; this is for example what happens with Uber or some pre-booking.

Initially both attributes were included in the SC experiment. However, while in the traditional (i.e. screen-based) SC experiment the inclusion of this attributes in the choice task sounds perfectly fine, in the VR environment, we noticed that it appeared unrealistic. When we choose a transport option in reality, we are only presented with the characteristics of the options, and after we make the choice the ticket machine, we are asked how we would like to pay. The realism of the SC experiment embedded in the VR experiment made this problem evident.

For this reason, after pre-tests, it was decided to remove this attribute from the SC experiment. In the "*survey-online NCL*" and "*survey-VR NCL*" the question was asked after each scenario. i.e. after the respondent has chosen the type of taxi, it appears the question "how do you want to pay?" with a list of options. A major effort was devoted to define how to present this information. A description of the tests carried out is reported in the **Appendix 2**. **Figures 3-10** shows the final layout used in both surveys ("*survey-online NCL*" and "*survey-online VR*"). After each choice scenario, respondents were asked to choose the Payment option (**Figures 3-10** shows the options if a normal taxi was chosen. For the AT, the two options to pay to the taxi driver were not displayed).

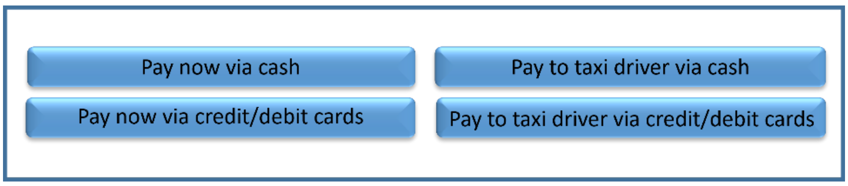


Figure 3.10 Presentation formats of payment options

In the immersive VR survey ("*survey-VR NCL*"), in order to complete the ticket purchase process, which is important to guarantee the realism, after the payment options, a page of instruction about how to pay was then presented, customised depending on the payment option selected. **Figures 3.11** shows the layout used.

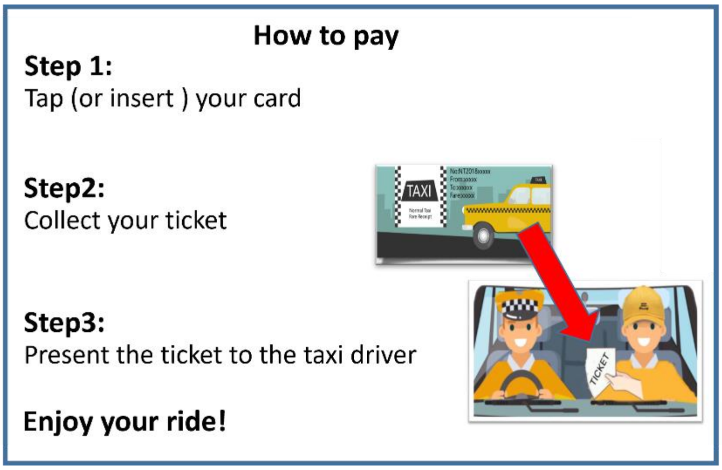


Figure 3.11 Presentation formats of payment instructions

In the "*survey-online UK"* and "*survey-online China*" the payment options were not added for two reasons. In China, paying the taxi fare by credit/debit cards is almost impossible. In the UK it was not sure if those form of payments were available everywhere in the UK, since this is not a key information for this research, and in order to keep the UK survey consistent (and more comparable) with the Chinese survey, it was decided not to include it.

3.5.2 Property of the design

A heterogeneous Bayesian D-error efficient design was built using Ngene (ChoiceMetrics, 2012) The design was customised based on the travel time of the last trip by taxi described by the respondent. Three segments were defined both in the UK and in China: 5 km trips (for short trips between 2.5km to 7.5km), 10km trips (for medium trips between 7.5km to 12.5km) and 15km trips (for long trips between 12.5km to 17.5km).

In Newcastle, there are 58 geographic district (and 7 non-geographic district) according to postcodes. Using google map, the average distance between the taxi rank and the central point of the different districts was computed according to road network distribution. The length distribution of the trip made by taxi from 2013-2017 in England confirmed that nearly 90% passengers take a taxi within the travel distance of 10 miles (i.e. less than nearly 15km) (DfT, 2018), which is consistent with the result from a small-scale investigation of the taxi drivers in

Northumberland Rd. Therefore, the list of destinations within 15 km covers the vast majority of taxi passengers' destinations in Newcastle. The destinations were then divided into three groups based on the average distance.

Post Code	District	Post Code	District	Post Code	District
NE1	Newcastle city centre	NE11	Dunston, Metro Centre, Team Valley, Kibblesworth	NE15	Lemington, Throckley, Newburn, Scotswood
NE2	Jesmond, Spital Tongues	NE12	Killingworth	NE20	Ponteland
NE3	Gosforth, Fawdon, Kingston Park, Great Park (East)	NE13	Airport, Wideopen, Dinnington, Great Park (West), Woolsington	NE21	Blaydon, Winlaton
NE4	Fenham, Westgate, Wingrove	NE16	Whickham, Sunniside, Burnopfield	NE23	Whitley Bay, Seaton Sluice
NE5	Westerhope, Newcastle West	NE27	Shiremoor, West Allotment, Backworth, Holystone, Murton Village	NE25	Monkseaton, New Hartley, Holywell, Seaton Delaval
NE6	Walker, Byker, Heaton	NE29	North Shields, Royal Quays, Billy Mill, New York	NE26	Cramlington, Seghill
NE7	High Heaton, Benton	NE31	Hebburn	NE30	Marden, Tynemouth, Cullercoats
NE8	Gateshead, Bensham	NE32	Jarrow, Fellgate, South Tyne Tunnel	NE33	Town Centre, Deans, High Shields
NE9	Low Fell, Springwell	NE35	Boldon Colliery	NE34	Harton, Horsley Hill, Marsden, Simonside, Brockley Whins
NE10	Felling, Leam Lane, Pelaw, Bill Quay	NE36	East Boldon, West Boldon	NE38	Town Centre, Oxclose, Fatfield, Harraton
NE28	Battle Hill, Willington, Wallsend, North Tyne Tunnel	NE37	Usworth, Sulgrave, Albany	NE39	Rowlands Gill, High Spen, Hamsterley Mill, Eastern Chopwell Wood

Figure 3.12 Destination list in the SC experiment in Newcastle (UK)

In the "*survey-online NCL"* and "*survey-VR NCL*", before each choice scenario, respondents were presented with a list of destinations and asked to select one specific destination for their trip that day (i.e. in that scenario). Differently from the standard practice, we allowed respondents to choose different destinations in each scenario, and hence the 6 scenarios presented to each respondent can belong to any of 3 segments (5km, 10km or 15km). This breaks the efficiency of the designs, but it gains significantly in realism, especially in the VR-based SC experiment. This aspect will be discussed further in **Section 3.7. Figure 3.12** shows the list of destinations presented in the UK experiment. The first two columns include all the destinations (post code and district) within an average distance from Northumberland Rd. taxi rank of approx. 2.5km-7.5km (short distance trip). The next two columns are the destinations within an average distance of approx. 7.5km-12.5 km (medium distance trip), and the last two columns include destinations within an average distance of approx. 12.5km- 17.5km (long distance trip).

In the "*survey-online UK"* and "*survey-online China*" we could not present a list of destinations (because it involves too many cities in the UK and China). Therefore, before the entire SC experiment, respondents were asked to describe a recent trip by taxi, including origin, destination, purpose and travel time. The SC experiment was then customised, based on the travel time declared. Based on the declared travel time reported, respondents were divided into same three group distances as discussed before, i.e. 5km, 10km and 15 km.

Weight factors were used to account for the proportion of the population in each segment. The weight factors used in the UK are: 75% for the 5 km segment, 20% for the 10 km segment, and 5% for the 15 km segment. The weight factors used in China are: 50% for the 5 km segment, 30% for the 10 km segment, and 20% for the 15 km segment. The weight factors were computed based on the real travel distance distribution of taxi trips in the UK[10] and on the distribution measured from pilot tests in China. A uniform distribution was used for all parameters to avoid extreme parameter values. 12 choice tasks in each segment were generated and randomly assigned into 2 blocks of 6 choice tasks each. To get the priors, fractional factorial orthogonal designs were built and tested, in the UK with a sample of 35 taxi users, and in China with a sample of 48 taxi users.

Attributes (*description)*		Attribute levels		
		Short (5km)	Medium (10km)	Long (15km)
Waiting Time [Minutes]	UK	1/6/11		
	China	1/7/13		
Travel Time [Minutes] (*in-vehicle travel time)*	UK	6/10/14	11/16/21	16/22/28
	China	6/11/16	12/18/24	18/25/32
Fixed Journey Fare *(for the trip)*	UK [GBP]	4/7/10	10/14/18	16/21/26
	China [CNY]	10/15/20	24/30/36	38/45/52
Number of Customers in the last hour	UK	17/80/143		
	China	51/240/429		
Change the destination (*possibility to change destination*)	UK & China	0:No 1:Yes		
Chat with an operator (AT) (*possibility to chat with an operator)*	UK & China	0: No 1:Yes		
Customer rating Yesterday	UK & China	0: Bad Reviews (2 Stars) 1: Good Reviews (4.5 Stars)		

Table 3-3 SC Attributes and attributes levels in the UK and in China

Table 3-3 reports the attributes and levels used in the final designs in the UK and in China. In the UK, the experimental design was the same for the screen-based and VR-based SC experiment. The levels of all the attributes were tested and adjusted during the various pre-tests, as discussed in **Appendix 2**. The levels of the attributes travel time and journey fare were customised based on the distance travelled by each respondent. Based on three distance intervals, the reference values of the travel time for the short, medium and long distance

[10] Department for transport 2019. Taxi and Private Hire Vehicle Statistics, England: 2019 assets.publishing.service.gov.uk

segments are based on the travel time computed on Google Map for UK and Baidu Map for China, for each of this distance. In addition, another consideration was to ensure that the lower level value of travel time was higher than the shortest time spent within OD due to speed limit in urban highway. Analogously, the full travel cost was computed based on the travel time reference defined for each interval and pivoting around this value. Within each distance interval, travel time and costs were defined as generic variables between AT and NT, with three levels each (+/- the reference value).

Three levels were also designed for the waiting time, ranging from almost no waiting time (1 minute) to 11 minutes in the UK (13 minutes in China), which is a quite high waiting time for both UK and China. To define the reasonable levels for the attribute *number of customers in the last hour*, it was measured that in Northumberland Rd. a taxi leaves at every nearly 5-6 minutes. Normally there are 5-6 taxis waiting there. This gives approximately 50-80 passengers per hour (estimated hourly passengers at 7 taxi ranks in Newcastle could be also used as reference to define this attribute levels (SouthTynesideCouncil, 2016))

The in-vehicle features (*change destination* and *talk with an operator*) were specified as dummy variables (two levels: the option was either available or not available). The possibility to change destination was given (and specified in the scenario) within a given radius, i.e. it was thereby controlled the radius of travel distance that respondents are allowed to change. The radius was set equal to the distance of reference travelled. For example, in a 10km trip, participants can change their destinations only within a 10 km radius from the origin. The same for the other distances 5 km and 15 km.

The levels defined for the attribute *customer reviews yesterday* were based on the study of Pang and Lee (2005). They proved that, within a rating scale of four or five stars and a separation of one star and a half, 100% of the users are capable of discerning the relative difference. Based on this evidences and with the aim to increase realism, it was decided to use 2 stars for bad reviews and 4 and a half stars for good reviews, i.e. with no extreme evaluations.

Finally, in the "*survey-online NCL"* and "*survey-VR NCL*" a short 'break' section was provided between scenarios. In the several preliminary tests ran with the VR environment, it was discovered that presenting the choice tasks in sequence or unceasingly, like it is always done in the online survey, was extremely unrealistic in the VR, and could jeopardise the level of realism for the entire experiment. It was decided therefore to add a break section between scenarios with the written "*your taxi journey went as planned"*. In the last break section instead a different message was included informing respondents that they had to wait at taxi rank 10 minutes longer than planned. This information was introduced to try to simulate the impact of "paying the consequences" of the choice made. It was introduced only before the last task in order to control the impact, i.e. that only the last task was affected by the message.

3.6 Attitudinal questions

As discussed in the literature review, **Chapter 2,** public perception or attitude towards ATs might vary with different automation levels. Before attitudinal questions, respondents were firstly informed that *A "fully automated taxi" refers to an automated taxi with an automation level of 5 (i.e. without safety steward and without steering wheel).* 10-item statements were then used in this study to measure the impact of injunctive norm, hedonic motivation and trust

Social conformity/social influence/subjective norm plays an important role in mode choices. The subjective norm or social influence is a common factor used in studying the acceptance of autonomous vehicle technology using UTAUT or TPB theory. The literature (e.g. Madigan *et al.,* 2017) rarely distinguished different types of social influence, except Cherchi (2017). Two types of social conformity (i.e. social adoption and customer reviews) have been objectively measured and mentioned in **Section 3.5.1.** Cherchi (2017) found after a couple of tests it would be difficult to objectively measure the impact of injunctive norms (injunctive norms refer to when the individual's behaviour is influenced by what other people think of her/his doing something). Therefore, in this part three statements adapted from Cherchi (2017) were allocated and presented in a random order to respondents and the second statement was reversely scaled to avoid the situation where respondents might respond the same or similar answers for similar questions. Hedonic motivation and trust statements were adapted from Venkatesh *et al*. (2012), as well as Choi and Ji (2015) and Liu *et al.* (2019b). For all these statements, a 7-point Likert response scale was used, ranging from 'Strongly Disagree' to 'Strongly Agree':

Three statements were adapted from Cherchi (2017) to measure the impact of injunctive norms (the second item was reversely scaled):

IN1:People who are important to me (friends, family) would approve of me using a fully automated taxi
IN2:People who are important to me (friends, family) would think that using a fully automated taxi is not appropriate
IN3: People who are important to me (friends, family) would think that more people should use fully automated taxis

Three statements were adapted from Venkatesh *et al*. (2012) to measure the impact of hedonic motivation (the third item was reversely scaled):

HM1: I believe using a fully automated taxi will be fun
HM2: I believe using a fully automated taxi will be pleasant
HM3: I believe using a fully automated taxi without driver will be boring

Four statements were adapted from Choi and Ji (2015) and Liu *et al.* (2019b) to measure the impact of trust (the first three items were reversely scaled):

T1: Overall, I do not trust fully automated taxis.
T2: I don't trust that fully automated taxis will be adequately supervised
T3: I don't trust that a computer can drive vehicle without assistance from the driver
T4: I trust I can relax while riding in a fully automated taxi without driver

3.7 VR-based SC Experiment in Newcastle (UK)

This section describes the methodology used to construct the VR environment for the specific context in Newcastle upon Tyne, UK, and some specific considerations related to SC experiment embedded into this VR environment.

3.7.1 Constructing virtual reality environment

As discussed in **Section 3.2.1**, when building a SC experiment embedded in a VR environment it is critical to carefully select the context of reference, unless a significant budget (both in terms of time and money) is available. Once the context is chosen, the next step consists in redesigning the urban structure to make the inclusion of ATs in the taxi rank realistic. Currently, normal taxis are operating in Northumberland Road and they stand in one line, along the pathways. The layout of the street was modified to accommodate both taxis, NT and AT. Several alternative designs were evaluated and reported in **Appendix 3**. **Figure 3.13** shows the geometrical design selected (on the left) and its implementation in the VR environment (on the right). As it can be seen:

- The taxi rank was organized with two separated lines of taxis, as this facilitates the experiment where respondents have to appreciate the characteristics of each taxi to make their choice.
- The black line in the drawing is a small island, where a *Ticket Board* is located as well as the space to queue for the taxis. The creation of this island has the main purpose to allow participants to walk from the ticket board to the queues, without the problem of

crossing the street in front of the cars, as this could add safety concerns that needed to be controlled in the experiment.

- People waiting for the taxi will be standing in the small island. This allows passengers to be equally distant from both NTs and ATs, and see the front of both type of taxis.
- The island needs to be wide enough to allow two clearly distinct lines of passengers. Careful considerations were required to ensure this was feasible given the real dimension of the area.
- The left side of the waiting area is for normal taxis and the right side for ATs. Ideally, the position should be randomised among the tasks presented to the respondents. However, changing the position of the cars in the VR environment is costly, and we decided not to implement this feature.

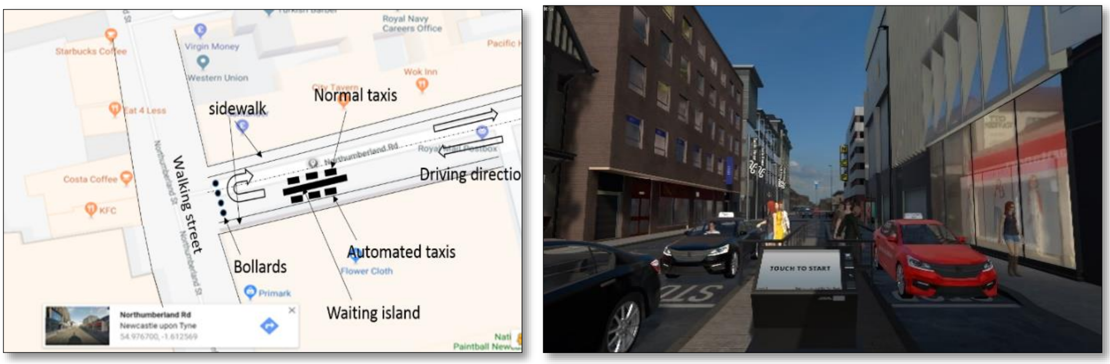


Figure 3.13 The re-design of taxi rank area: drawing and VR environment built

The VR platform was built using Unity 3D Engine for gaming and virtual reality[11], which allowed pushing the realism to very high standard. It features an extremely detailed reproduction of the buildings. Respondents using the VR have then the possibility to see how the taxi system (and in particular the AT) works and is used: passengers queue waiting for the taxi, the taxis (normal and the automated) arrive, pick up passengers and leave the taxi rank. Since respondents are immersed in this environment, they can really live in the environment: they can go close to the taxi, queue with the other passengers, cross the street, see the ATs passing by, closed to them.

[11] The VR environment was built by Animmersion, a leading supplier of digital visualization tools specialized in VR https://www.animmersion.co.uk/

A potential problem with realistic environments is that unintended information might affect the respondent choices since VR environments include richer contexts (Patterson *et al.*, 2017). This is also what happens in real-life contexts that seldom allow experimental control, i.e. to capture the effects of the stimuli. The VR environment built in this dissertation represents the real environment with high fidelity. With this, we ensured that the VR does not introduce additional distractions compared to the real environment. The only new element, compared to the real environment is the presence of an AT. For the ATs we used the same shape as that of a normal car. This was analysed in a series of focus groups (FGs) ran as preparatory analysis (as discussed in **Section 3.3**). Respondents showed a preference for cars of normal size and normal shape (i.e. similar to the current vehicles). This result guided the choice of cars that appears in **Figure 3.13.** For the normal taxi, we used a black standard car that reproduces the shape of the taxis that operate at the taxi ranks in Newcastle.

Another discussion that was held during the FG was whether AT passengers should sit in the front seat or in the back and whether AT should have a steering wheel (though without having the driver) or not. As for the sitting position, there was not a clear preference. We chose to show passengers sitting in the front, as this makes it even more evident the absence of driver. In principle, it is possible to show in the VR experiment, passengers sitting both in the front or the back seats, but it is more expensive. Interestingly, instead participants in the FGs showed a clear preference for an AT without steering wheel, mainly because it looked like what participants expected an AT to be.

Another non-trivial decision referred to how to present the information about the taxis (which is the core of the SC experiment) and how respondents would make their choice. An option considered was to present the information in a cell-phone, but this is not easy to implement within the VR environment due to the difficulty to guarantee sufficient quality of the images in such small dimensions. It was then decided to design a ticket board that simulates a touch-screen, which respondents can activate using the trigger in the VR handles. To enhance realism, the ticket board was carefully designed to resemble existing ticket boards, both in its external appearance and in the format of the contents presented. Several formats were tested, the final version (reported in **Figure 3.14)** includes a wide screen designed to present the choice scenarios (this characterises the taxis) and on its right the features to insert banknote and coins, to tap the bank cards and to receive ticket and change for improving realism in VR, which is consistent the real-life choices of purchasing the ticket.

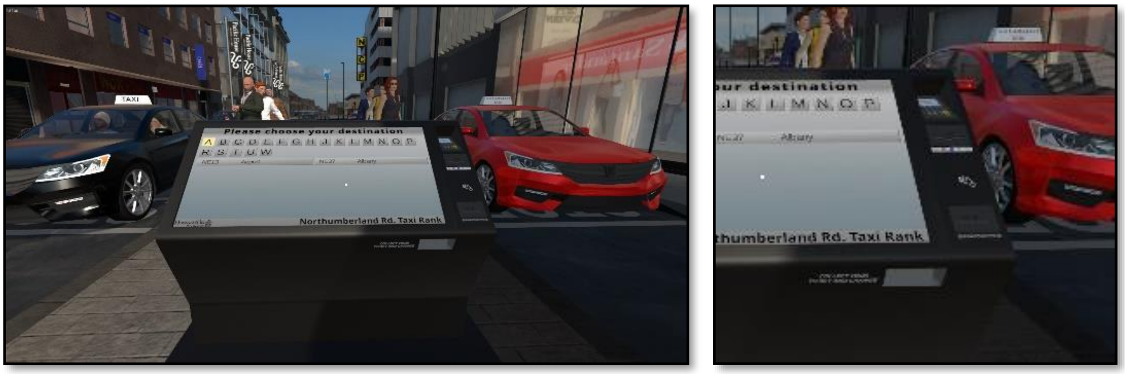


Figure 3.14 Ticket Board in the immersive VR-based SC experiment

The ticket board was positioned facing the taxis in order to give respondents the possibility to see the vehicles and the other passengers queuing while reading the information about the taxis available and making their choice.

Aspects that need to be carefully considered when designing the ticket board in the VR environment are:

- Size of text and images. Despite the very good quality of the VR environment, the resolution is not comparable with that of a computer or smart phone. This posed a constraint in the amount of information that could be presented in the ticket board. A significant amount of work was needed to identify how to present the information (i.e. the attributes of the SC experiment and the text to include) without altering the SC experiment designed.
- Colour of text and images. The vision with the VR is slightly different from the one we have in real world. For this reason, also a simple task like the colour of the text and that of the background required several tests, in order to identify the best combination in terms of ease to read.
- Height of ticket board. It has to be high tall enough to allow respondents to read easily the information provided (given the resolution of the text in the VR environment) but not too high, otherwise respondents are not able to see the taxis and the other passengers queuing, which is an important part of the immersive VR experiment. In the VR, the height of the board can be adjusted based on the height of each respondent, but this option is relatively costly. Therefore, we opted for a fixed height. The appropriate height was identified after several tests where feedbacks from respondents were collected and analysed.

As discussed in **Section 3.4.1**, since ATs do not exist in reality, it is necessary to provide respondents with background information (i.e. pre-information). The VR environment was constructed in such a way that respondents can see how the system operates and walk around the moving vehicles without the need to describe it. However, the same pre-information presented in the screen-based online survey was also presented in the VR-based survey, in an information board (**Figure 3.15**). As shown in the right part of **Figure 3.15**, at the top of the information board screen there is a menu, and respondents can select the information (i.e. safety, privacy, routing information and optional/compulsory in-vehicle features) to display using the trigger in the VR handle. Note, as respondents can observe how other passengers use ATs in VR, the information relevant to 'how to use ATs' was not included in the information board.

Analogously to the ticket board, also for the information board several tests were run to identify the optimal height as well as the appropriate size of the text and images.

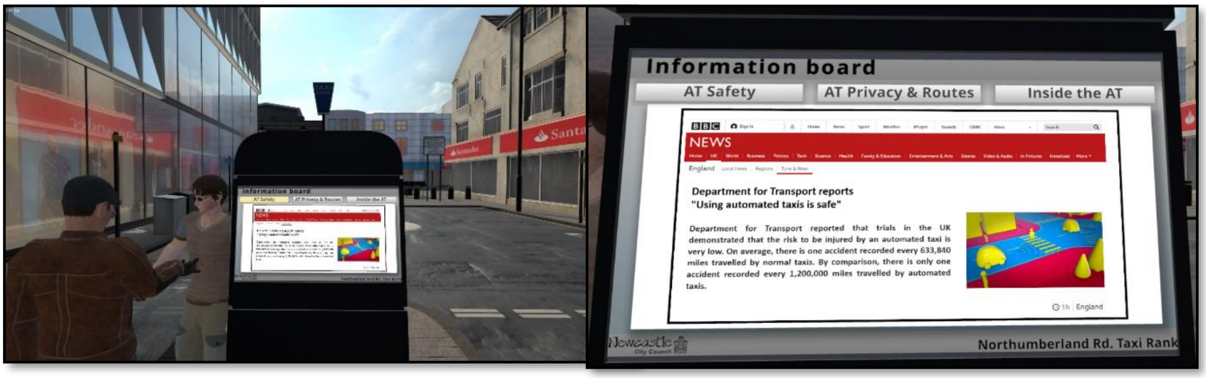


Figure 3.15 Information board in the immersive VR-based SC experiment

3.7.2 Constructing the SC experiment embedded in the VR environment

The process to build the VR-based SC experiment was described in **Section 3.5.** In this section it will discuss how the VR environment was built to implement the SC experiment. As discussed in **Section 3.5** before presenting the choice tasks, respondents were asked to choose the destination for their trip in order to customise the design based on the distance of their destination.

Figure 3.16 Destination list presented to respondents in VR SC experiment

a) destination list (e.g. with first letter H)

b) a customised choice task

c) payment method and instruction

d) 'break' section

Figure 3.17 Example of SC choice task in the VR environment

To make it realistic, the list of destinations in the VR environment was presented in the ticket board. The first information asked to respondents was to "select your destination". In standard SC experiments, this information is asked once at the beginning of the SC and before each scenarios respondents are asked to assume that they have to do a trip always with the same destination. Interestingly, this standard procedure, that sounds perfectly reasonable in the screen-based SC, it appeared extremely unrealistic in the VR environment. This is because each scenario in the VR starts at the ticket board, where the first information is “select your destination” and during the pre-tests respondents doubted why they need to choose same destination in each scenario. This motivated the decision to allow respondents to choose different destinations in each scenario.

Figure 3.16 shows how the list of the destination appeared in the ticket board in the VR environment. In order to make the experiment more realistic, respondents can query destination by the first letter in VR. **Figure 3.17** shows how the entire experiment appeared in the VR environment.

3.8 Additional information

This section completes the discussion of the survey structure, describing the additional information included in Section 1 (introduction) and in Section 4 (individual information).

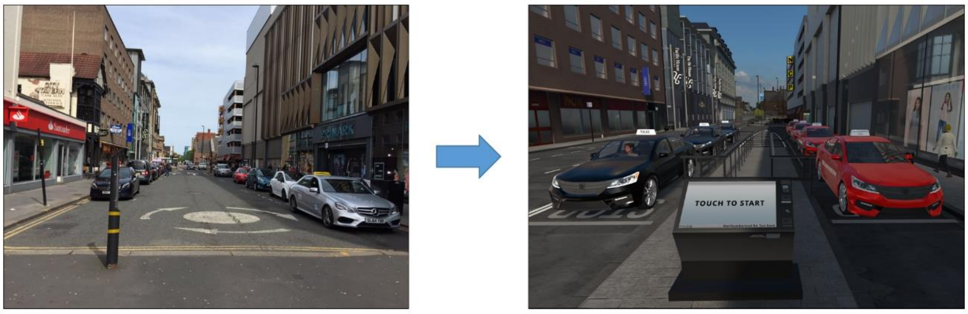
3.8.1 Section 1: Introduction to the survey

The first section of the survey included three types of information: Screen out, knowledge of automated vehicles and recent taxi trip information. The survey was intended for taxi users and the screen-out question was set up to include only respondents who have used a taxi at least once in the last year. Some questions were included to measure the level of familiarity with AVs in general and in particular with the AT system operating in both UK and China. In this occasion, respondents were also given a description of what is an automated vehicle and the different levels of automation. In terms of knowledge of automated vehicles, they were asked four specific questions: if they heard of AVs, how familiar they are with the 5 levels of automation, if they heard of ATs testing or operating in the UK/China and from whom/where they heard of ATs testing or operating in the UK/China. Finally, in the “survey-online UK” and “survey-online China” respondents were asked to describe the last trip made by taxi (used then to customise the SC experiments, while in the survey-online NCL and survey-VR NCL, the SC experiments were customised based on their selected destination at each scenario) this includes: origin and destination of the trip, purpose, what time and where they took a taxi, travel time, how they booked the trip, etc. Additionally, in survey-online NCL respondents were reminded that taxi rank at Northumberland Rd. has been redesigned for accommodating the ATs which is consistent with the situation mentioned in survey-VR NCL in **Section 3.7**, as shown in **Figure 3.18**.

Assume that you were in Northumberland Street (or in any of its branch roads) doing the following activity: *Leisure (e.g. shopping, pub,etc.)* ,
and then you go to take a taxi at the taxi rank in Northumberland Rd.

Assume that Northumberland Rd. has been redesigned as in the picture below.

There are 2 types of taxis you can choose from:
- **an automated taxi** (i.e. without taxi driver and without steering wheel)
- **a normal taxi** (i.e. like the taxi circulating now)



Real view of Northumberland Rd. **Redesigned view of Northumberland Rd.**

Figure 3.18 Textual and pictorial description of reference context

3.8.2 Part 4: Individual information

Questions related to socio-demographic characteristics and taxi riding behaviour characteristics were asked in all four surveys. This section was used to collect socio-demographic characteristics of respondents including gender, age, level of education, employment status, and personal monthly disposable income. Questions relevant to education level and personal monthly disposable income are different between China and UK survey. In terms of education level, A level and GCSE (general certificate of secondary education) were asked in the UK survey which are considered the same as education level high school and junior high school in China survey. In addition income categories asked are different due to the income difference between China and UK. Moreover, this section was also used to collect respondent taxi riding behaviour characteristics: frequency of using taxi, frequency of taking with taxi driver, whether they enjoy talking with a driver, whether they like driver to help carry luggage and whether they are capable to take a taxi without any help. Finally, only for "*survey-VR NCL",* some extra questions were supplemented in the survey to investigate their previous VR experience (i.e. frequency of using VR devices) and their feeling and experience about current VR SC experiments: whether they observed the colour of normal taxis and ATs, whether they observed the passenger queue and rating the reality level of this experiment.

3.9 Sample recruitment

Table 3-4 summarises the main characteristics of the four surveys ran. It is important to mention that due to Covid-19 the online surveys in the UK and China were run before the surveys (online and VR) in Newcastle, even though the research was developed first for the Newcastle surveys (both online and VR) and then adjusted for the two online national surveys.

The *survey-online UK* was administrated in selected main the UK cities, e.g. London, Leeds, Sheffield, etc to current users of normal taxis (i.e. with driver). Data were collected in December 2021. A pilot test with approximately 50 respondents was conducted before the final survey-online UK to check if respondents can fully understand the entire survey and avoid possible problems. The final sample was recruited using the panel provider Prolific Academics. A pre-survey was used to screen out those who have not used a normal taxi in the last year. 384 participants joined the pre-survey, 25 participants were screened out as they did not satisfy the requirements: 1) are more than 18 years old; 2) have used the normal taxi in the last year; 40 participants were excluded from the final UK sample as they did not reply to the whole questionnaire. After data cleaning, a final sample consisting of 319 respondents (19 from Newcastle and 300 from the rest cities of UK) was available for the analysis. Since each respondents provided 6 choices, a total sample of 1914 pseudo-observations was available for modelling purposes.

The *survey-online* China was administrated in selected main Chinese cities, e.g. Shanghai, Guangzhou, Changsha, Chongqing, Wuhan etc. to current users of normal taxis (i.e. with driver). In Changsha, the AT service called Apollo, was introduced by Baidu, a Chinese IT giant, in September 2019, with an initial fleet of 45 AT vehicles. Apollo was then fully open to the public and offered free trial AT rides on the open urban roads in April 2020. According to Changsha Evening News (2020), Apollo AT service covers 130 km^2 including residential, commercial and leisure areas, and industrial parks in Xiangjiang New District. The operation hours are from 09:30 to 16:20 from Monday to Sunday except holidays. One of the participants reported the ride of Apollo AT taxi was smoother than they expected. As of the end of December 2019, Apollo AT taxis had served more than ten thousand safe trips. In Shanghai, the Chinese ride-hailing giant, DiDi, unveiled the AT service for the public in June 2020, starting with a fleet of dozens AT vehicles. According to DiDi (2020) and XINHUANET (2020), users need to register online and they can hail a free AT within the 53.6-km designated roads in Shanghai's Jiading District after passing the ID check. Operation hours and safety-related incidents were not reported. After two months AutoX, a startup AT company, also launched the AT service for the public in Jiading District, Shanghai. There were no fee charges for the pilot AutoX AT service. According to AutoX (2020), the initial pilot AT service included a fleet of 100 vehicles, but specific operating hours and safety information are not available to obtain. In Guangzhou, the WeRide Robotaxi company, a start-up of autonomous driving, was the first to launch at the end of November 2019 the commercial operation of AT services for the public in China, with a fleet of over 100 vehicles. Instead of offering the free rides in trial operations, WeRide AT service charges the same tariff as the conventional taxis. According to WeRide passenger survey report (WeRide, 2020), AT service areas cover a territory of 144 km^2 and it operates from 8 am to 10pm every day of a week. There were 209 pick-up and drop-off spots on the urban roads in Huangpu and Guangzhou Development District, where 40 AT vehicles have been deployed in the operation area. As of November 2020, WeRide has served more than sixty thousand passengers in the first year's operation. Approximately 90% of riders expected a wider service area. Safety driving performance scored relatively high, and no accident-related incident was reported.

Data were collected between March and April 2021. The final sample was largely recruited using the panel provider SurveyEngine and in a small proportion self-recruited. The survey targeted only respondents 18 years or older, who have used a normal taxi in the last year and who have never used ATs before (during the experiment, AT services are available for the public in some Chinese cities). 633 participants joined the survey in China. 89 participants were screened-out as they did not satisfy the requirements: 1) are more than 18 years old; 2) have used the normal taxi in the last year; 59 participants were excluded as they did not reply to the whole questionnaire. The final valid sample consists of 485 respondents. Since in China there are ATs operating, the aim was to collect information for both those who used the AT and those who did not (ATs are available to the public during the period of survey administrated in some cities, China. i.e. Changsha, Shanghai and Guangzhou; AT service was commercially operating in Guangzhou and AT services were under trial operation in other two cities). Out of the 485 respondents, 35 are the respondents who tried the AT and 450 those who have not tried the AT themselves. The sample was initially selected randomly, and after it was tried to specifically reach out those who had tried an AT, but this is still a niche group, and it was very difficult to reach them out in a normal online survey. A total sample of 2,700 pseudo-observations was available for modelling purposes.

The *survey-online NCL* was conducted in January 2022 and targeted only residents in Newcastle upon Tyne or surrounding areas in northeast of England and satisfied the same requirements: to be 18 years or older and have used a normal taxi in Newcastle in the last year.

An online pre-survey was also applied in advance to screen out those who were ineligible to join the survey and those 16 respondents who have already answered the *survey-online UK*. After data cleaning, 28 participants were excluded as they did not reply to the whole questionnaire and 8 participants were removed due to lexicographic problems. The final sample consists of 139 valid responses and A total sample of 834 pseudo-observations was available for modelling purposes.

The *survey-VR UK* was conducted in the premises of Newcastle University in December 2021 and February 2022, and targeted also residents in Newcastle upon Tyne or surrounding areas in northeast of England and satisfied the same requirements as the online survey, to be 18 years or older and have used a normal taxi in Newcastle in the last year. An online pre-survey was also applied in advance to screen out those who were ineligible to join the survey and those recruited from Prolific Panel who have participated in *survey-online UK and survey-online NCL*. After data cleaning, 2 participants who never used taxi in the last year in Newcastle, the final sample consists of 40 valid responses. The sample was recruited from the panel provider Prolific Panel, and from advertisement at Newcastle University. A total sample of 240 pseudo-observations was available for modelling purposes.

Survey type	Sample size	Context	Time
Survey-online UK	319	UK - Several cities	December 2021
Survey-online China	450	China - Several cities	March - April 2021
Survey-online NCL	139	Newcastle	January 2022
Survey-VR NCL	40	Newcastle	December 2021 and February 2022

Table 3-4 summary of four surveys for data collection

3.9.1 VR-based SC survey protocol

For the VR-based SC experiment, a rigorous protocol was defined, which consists of the following four steps:

Step 1 Introduction

a. Participants were welcomed and asked to sign a consent form explaining the purpose of the study and some covid safety measures implemented to mitigate risk of infection during the experiment;
b. Participants were shown the basic components of the VR and given a brief explanation of the entire experiment.

Section 2 Tutorials

Respondents wore the VR headset and went through two short tutorials about VR:

c. A general tutorial, a video about how an immersive VR environment works;
d. A specific tutorial, respondents found themselves in the virtual Northumberland road and were asked to walk and move around to familiarise with the environment of the final experiment.

These tutorials were particularly important and aim to ensure that all respondents have enough capabilities to complete choice task in the VR environment and to avoid distractions during the experiment. During the several pilot tests, respondents, particularly those without the previous experience with VR, expressed high excitement at the first time when they found themselves in the virtual street of Northumberland Road and were clearly distracted by what they were seeing, including the products displayed in the windows, though these are the same they could see in reality. This tutorial to familiarise with the environment is then critical to avoid participants to be distracted by unintended information (Patterson *et al.*, 2017; Birenboim *et al.*, 2019)

Short break

Participants were asked to remove the headset and to have a short break (2 minutes-5 minutes). It is important to mention that, during the experiment and at the end of each tutorial, respondents were asked how they felt. The experiment would have been immediately stopped if respondents suffered from any of adverse symptoms (e.g. motion sickness the most likely, but it could be also dizziness or nausea, or sense of claustrophobia, etc.)

Section 3 Formal VR-based SC experiment

e. Respondents were asked to wear again the immersive VR headset to perform the final VR experiment, where respondents were informed that they have now to imagine that they are in Northumberland Road to take a taxi as they do in reality. No further instructions were provided at this point regarding the SC experiment in itself. Respondents were given only some information about how to use the controllers to make their choices into the VR experiment.

Section 4 Post survey

f. At the end of the VR experiment, respondents were asked to fill in a short online questionnaire, including questions mentioned at part 1, part 3 and part 4 in **Section 3.1** and some extra questions about their experience with the VR-based SC experiment (also discussed in section 4.3.4), such as:

- Can you tell ‘what was the colour of the autonomous taxi in the VR experiment?’
- Can you tell ‘what was the colour of the normal taxi in the VR experiment?’
- Did you notice the passengers’ queueing when you were doing VR experiment?
- Can you tell how many people approximately were queueing for the ATs and how many for the NTs in the last scenario?
- How realistic was the VR experiment?

These questions were supplemented to investigate respondents’ experience with VR and to examine if they observed the surrounding environment when selecting taxis. It is important to mention that during the VR experiment no respondent experienced sick motion. Only in the final experiment, one respondent suffered from claustrophobia and could not continue the experiment.

3.10 Summary

This chapter described the data collection methodology defined in this study in order to reach the research objectives. The core of the data collection methodology is the stated choice experiment built to identify what affects consumers’ preferences in the choice of ATs versus NTs. In particular, the key contribution of the SC methodology proposed consists in the study of the in-vehicle features, the study of the normative conformity and the implementation of the SC experiment in an immersive VR environment.

In-vehicle features refer to those services that are typically provided by the taxi driver. After extensive work, including several FGs and pilot tests, two in-vehicle feature attributes are identified and defined in this PhD research: the possibility of changing the destination and the possibility of chatting with an operator. Normative conformity refers to the impact of the choice and the reviews of other customers. After extensive work, two normative conformity attributes are defined in this PhD research: number of customers in the last hour and customer rating yesterday. Given the nature of the AT service, a short time reference (number of customer in the last hour) is proposed because it is more realistic: the number of customers change among scenarios, and each scenario represents another day where the respondent takes the taxi. For this customer rating, a 5 star system, with no extreme evaluations is proposed. Reviews refer to “yesterday” customers, to make it realistic for respondents to see different customer rates in different scenarios, as each scenario represents another day. The SC experiment included also three level of service attributes (waiting time, travel time and fixed journey fare).

In order to achieve the specific objectives of this PhD research the SC experiment was built in a way that it could be used to collect data in the UK, in China, online and with the VR environment. The chapter discussed the challenges and the methodological implication of building such a survey, with particular reference to the implementation of the SC experiment in the immersive VR environment. The most important and interesting issue is that some of the elements that are typically used in the online SC and are considered perfectly acceptable, look quite unrealistic when used in a VR-based environment. In particular, this chapter highlights three elements of complexity: in the definition of the choice context (in the VR has to be a specific real location), in the definition of the attributes (the VR makes it clear that not all attributes can be included and trade-off in the SC design), in the customisation (specifically, the VR makes it clear that is not realistic asking the destination of the trip only once at the beginning of the SC experiment and presenting the choice tasks continuously without break).

The overall questionnaire included also attitudinal statements (i.e. injunctive norm, hedonic motivation and trust), information about a recent taxi trip, information about the knowledge level of AVs and ATs and individual-related socio-demographic characteristics.

Finally, the sample recruitment methods for the four surveys were summarised in the last section. 319, 450, 139 and 40 valid responses were respectively collected for *survey-online UK*, *survey online China*, *survey-online NCL* and *survey-VR NCL*. Detailed discussion and comparison of these four datasets will be presented in the following **Chapter 4 Data descriptive analysis**.

Chapter 4. Descriptive Data Analysis

4.1 Introduction

This chapter reports the descriptive analyses of the data collected. In line with the objectives of this PhD thesis, the data are analysed looking at all the information collected (socio-demographics, trip characteristics, knowledge of AV and psychological statements), but also at a comparison between samples. An initial descriptive summary of the stated choices will also be discussed. In particular, **Section 4.2** describes and compares the samples gathered with the *Survey-online UK* and the *Survey-online China*. **Section 4.3** describes and compares the samples gathered with the *Survey-online NCL* and the *Survey-VR NCL.* **Section 4.4** explores instead in more detail the impact of living in cities where ATs are operating compared to cities where they are not in operation. This information is available only for the *Survey-online China*. Finally, **Section 4.5** summarises this chapter

4.2 Data from the *Survey-online UK* and the *Survey-online China*[12]

As mentioned in **Section 3.9**, the final valid sample collected in the UK consists of 319 respondents, while in China it consists of 485 respondents, of which 35 are the respondents who tried personally the AT and 450 those who did not try it. Since ATs are not available in the UK, nobody has tried them. The 35 respondents who tried AT in China will then not be included in the comparison between samples, but analysed separately in **Section 4.4**. The UK population at the end of 2021 counted 68.2 million people. The Chinese cities where the sample was collected (hereafter called 'main cities') in 2021 had a population of approximately 74.6 million people. None of the two samples can be considered representative of their respective populations, even though an attempt was made to try to match the population distribution at least in terms of gender and age.

4.2.1 Socio-demographic characteristics

Table 4-1 reports the distribution of the socio-demographic characteristics in the UK and in China, including only those who have not tried an AT. Based on the chi-squared tests, there are significant differences between the two samples from UK and China in all the characteristics analysed (all p-values are less than 0.05). In particular, in the Chinese sample the proportion of male and that of youngers (18-29 years) is higher than in the UK, while the proportion of elderly

[12] It is useful to remember that China in this dissertation means mostly the main cities of Shanghai, Guangzhou, Changsha, Chongqing, Wuhan, Guiyang, Xianning and Huangshi.

(60 years or older) is lower. However, we note that this reflects the distribution of the population respectively in the UK and in China. According to the last Census in England and Wales in 2011[13] the distribution was as follows: 51% female 49% males; 21% less than 30 years old, 17% 30-39 years old, 19% 40-49 years old, 15% 50-59 years old and 29% 60 years or older. According to the 6th Census of the Chinese population in 2011, the distribution in the main cities, was as follows: 51.3 % female and 48.7% male; 39.2% less than 30 years old, 16.5% 30-39 years old, 16.7% 40-49 years old, 12.9% 50-59 years old and 14.7% 60 years or older.

Socio-demographic characteristics			UK N (%)	China N (%)	χ^2test (p-value)
Total sample			**319**	**450**	
Gender	Female		164 (51.4)	200 (44.4)	4.023 (0.047)
	Male		152 (47.6)	249 (55.3)	
	Rather not to say		3 (0.9)	1 (0.2)	
Age	18-29		74 (23.2)	154 (34.2)	25.29 (0.000)
	30-39		60 (18.8)	83 (18.4)	
	40-49		49 (15.4)	71 (15.8)	
	50-59		46 (14.4)	76 (16.9)	
	60 or above		90 (28.2)	66 (14.7)	
Education level	Secondary school and below		56 (17.5)	32 (7.1)	24.22 (0.000)
	High school		71 (22.3)	90 (20.0)	
	Bachelor degree		140 (43.9)	231 (51.3)	
	Master degree		44 (13.8)	88 (19.6)	
	Doctorate degree		8(2.5)	9 (2.0)	
Current work status	Employed full-time (30+ hours per week)		126 (39.5)	295 (65.6)	137.31 (0.000)
	Employed part-time (<30 hours per week)		53 (16.6)	4 (0.9)	
	Self-employed		41 (12.9)	8 (1.8)	
	Jobless		9 (2.8)	8 (1.8)	
	Students		35 (11.0)	81 (18.0)	
	Retired		50(15.7)	54 (18.0)	
	Others		5 (1.6)	0 (0.0)	
Personal monthly income	Less than £500	<¥1000	71 (22.3)	38 (8.4)	N.A.
	£501-£1500	¥1001 - ¥ 3000	118 (37.0)	22 (4.9)	
	£1501-£ 2500	¥3001 - ¥5000	65 (20.4)	53 (11.8)	
	£2501-£ 3500	¥5001 - ¥10000	27 (8.5)	160 (35.6)	
	£3501-£ 4500	¥10001 - ¥20000	11 (3.4)	94 (20.9)	
	> £4500	>¥20000	4 (1.3)	33 (7.3)	
	I do not wish to disclose it		23 (7.2)	50 (11.1)	

Table 4-1 Socio-demographic characteristics- Survey-online UK & China

13 The results from the 2021 Census are not available for the public during the data analysis period. However according to the World Bank in 2020 the distribution of age and gender is similar to that in 2011.

The percentage of respondents with high education level in the UK sample (60.2% bachelor degree and above) is lower than that in the China sample (72.9%). The distribution of the education level for each of the two samples is also different from that of the respective populations. However, as already said, the aim was not to get representative samples. In the UK around 2% of the population aged between 21 and 65 holds a PhD and 6.2% a MSc. The chi-squared test between the sample and population for the education level in the UK allows rejecting the assumption of equal distribution at 99.99%. The distribution of education level in the main cities of China is 61.5% secondary school and below, 20% high school (and college degree) 18.5% Bachelor degree or above. The chi-squared test between the sample and population for the education level in China allows rejecting the assumption of equal distribution at 99%.

With respect to the employment status, the proportion of participants in the UK sample who are full-time employees or retired is much less than in China and the proportion of those who are part-time employed or self-employed much higher. In addition, as reported in **Table 4-2**, in the UK the proportion of participants over 30 who are not full-time employee is more than twice than in China. And conversely in China the proportion of participants below 60 and full-time employee is almost twice than in the UK. It is not possible to make a comparison with the distribution in the respective populations, because comparable values were not found.

Age	UK N (%)		China N (%)		χ^2 test (p-value)
	Full-time employee	Other status	Full-time employee	Other status	
18-29	25 (7.8)	49 (15.4)	69 (15.3)	85 (18.9)	91.713 (0.000)
30-59	86 (27.0)	69 (21.6)	208 (46.2)	22 (9.6)	
60+	15 (4.7)	75 (23.5)	18 (4.0)	48 (10.7)	
All	126 (39.5)	193 (60.5)	295 (65.6)	155 (34.4)	

Table 4-2 Age distribution by employment status- Survey-online UK & China

Table 4-3 reports a comparison between the income scales in the UK and China. At 30 May 2022, the conversion rate was 1 GBP = 8.43 CNY. But due to different income categories used in the UK and China survey, income level cannot be directly compared. The personal monthly income in the UK sample is skewed toward low income group ‘less than £500’ and ‘£501-£1,500’ accounting for 22.3% and 37.0%, respectively (the median monthly pay of the UK is

£2,076 in 2022). The Chinese per capita disposable income in 2020[14] was 3,820 GBP (32,200 CNY) on average nationwide and 5,200 GBP (43,800 CNY) on average for urban residents. This means an average per-capita monthly income of approximately 320 GBP nationwide and 430 GBP at urban level. In terms of income, the Chinese sample is concentrated in the middle and high income accounting for 35.6% for ¥5,001-¥10,000 [£593 - £1,186] and 20.9% for ¥10,001 - ¥20,000 [£1,186 - £2,373], respectively.

	UK categories	China categories	
	GBP (£)	CNY (¥)	Converted into GBP (£)
Personal monthly income	Less than £500	<¥1000	< £119
	£501-£1500	¥1001 - ¥ 3000	£119 - £356
	£1501-£ 2500	¥3001 - ¥5000	£356 - £593
	£2501-£ 3500	¥5001 - ¥10000	£593 - £1186
	£3501-£ 4500	¥10001 - ¥20000	£1186 - £2373
	More than £4500	>¥20000	>£2373

Table 4-3 Income scales- Survey-online UK & China

It is interesting to note that a common feature for both UK and China is that about 95% of the sample do not pay more than 7-8% of their monthly disposable income to travel by taxis. In fact a taxi in the UK costs approximately £15 for a trip of 20 minutes (travel time), with only 2.5% of sample spending 32% of disposable income in travelling by taxi and 96% of the sample spending less than 7% of their disposable income travelling by taxi. A taxi in China costs approximately £4.50 for a trip of 20 minutes (travel time), and the percentage of disposable income spent travelling by taxi in our sample is less than 2%. Only 1% of the sample spends on average 42% of their disposable income on taxi, 4% of spends 14% and 95% of the sample spends less than 8% of their disposable income travelling by taxi.

In comparing the income between the UK and China, it is important to bear in mind the difference in the purchase power between the two countries. The Purchase Power Parity (PPP) conversion factor for gross domestic products (GDP) in local currency unit per USD is 4.19 (CNY/USD) for China and 0.68 (GBP/USD) for the UK in 2021 (The World Bank, 2021). Therefore, PPP conversion factor for GDP between China and UK is 6.16 (GBP/CNY). The PPP approach reflects the differences in the prices of general products and measures the purchasing power of a currency by comparing the prices of a basket of goods and services.

14 Information from the National Bureau of Statistics of China – updated at 19/01/2021 - http://www.stats.gov.cn/english/PressRelease/202101/t20210119_1812523.html

4.2.2 Trip characteristics

Table 4-4 reports the characteristics (travel time and purpose) of the last trip made by taxi, frequency of usage of taxi, and information about interaction with the taxi driver (frequency of talking with the taxi driver, enjoying talking with the taxi driver, whether they like the taxi driver to help carrying luggage and whether they can take the taxi without any help).

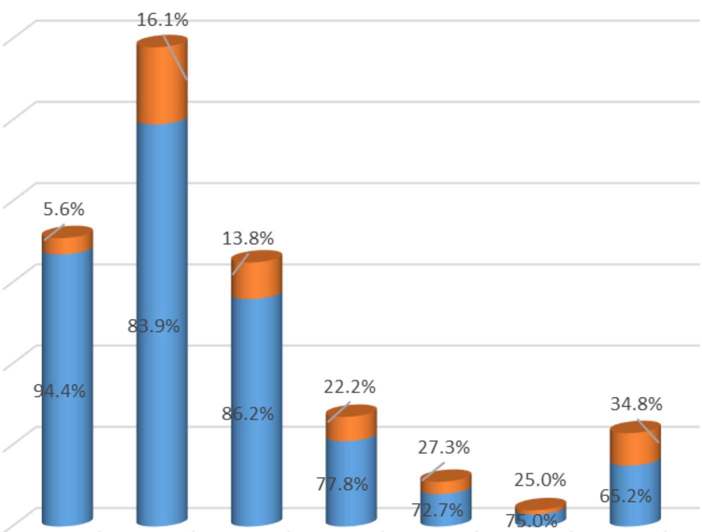
As it can be seen, the two samples (UK and China) differ in terms of distribution of the trip characteristics analysed (all p-values are less than 0.05), with two exceptions: in both the UK and China the vast majority of the respondents like the driver to help them carry the luggage and are able to take the taxi without any help.

Trip characteristics		UK N (%)	China N (%)	χ^2 test (p-value)
Travel time most recent trip by taxi	Short (around 10 min. or less)	177 (55.5)	118 (26.2)	76.12 (0.000)
	Medium (around 20 min.)	92 (28.8)	163 (36.2)	
	Long (around 30 min or more)	50 (15.7)	169 (37.6)	
Activities before taxi trip	Business	20 (6.3)	46 (10.2)	72.90 (0.000)
	Commuting (work or school)	15 (4.7)	97 (21.6)	
	Leisure (e.g. shopping, pub)	143 (44.8)	150 (33.3)	
	Visiting friends	44 (13.8)	43 (9.6)	
	Holiday	30 (9.4)	14 (3.1)	
	I was at home	47 (14.7)	90 (20.0)	
	Other, please specify	20 (6.3)	10 (2.2)	
Trip purpose most recent trip by taxi	Business	17 (5.3)	54 (12.0)	87.21 (0.000)
	Commuting (work or school)	14 (4.4)	98 (21.8)	
	Leisure (e.g. shopping, pub)	126 (39.5)	171 (38.0)	
	Visiting friends	38(11.9)	55 (12.2)	
	Holiday	22 (6.9)	13 (2.9)	
	Going home	80 (25.1)	42 (9.3)	
	Other, please specify	22 (6.9)	17 (3.8)	
Frequency of using taxis	At least once a week	31 (9.7)	193 (42.9)	190.50 (0.000)
	Less than once a week, at least once a month	81 (25.4)	164 (36.4)	
	Less than once a month, more than twice a year	127 (39.8)	82 (18.2)	
	At most twice a year	80 (25.1)	11 (2.4)	
Frequency of talking with the driver	Very infrequently	39 (12.2)	54 (12.0)	58.03 (0.000)
	Somewhat infrequently	34 (10.7)	63 (14.0)	
	Occasionally	121 (37.9)	255 (56.7)	
	Somewhat frequently	81 (25.4)	66 (14.7)	
	Very frequently	44 (13.8)	12 (2.7)	
Enjoying talking with taxi driver	Always	36 (11.3)	35 (7.8)	9.19 (0.010)
	Sometimes	249 (78.1)	335 (74.4)	
	Never	34 (10.7)	80 (17.8)	
Like the driver to help carry luggage	Yes	209 (65.5)	291 (64.7)	0.06 (0.807)
	No	110 (34.5)	159 (35.3)	
Take the taxi without any help	Yes	313 (98.1)	447 (99.3)	2.38 (0.174)
	No	6 (1.9)	3 (0.7)	
Sample size		319	450	

Table 4-4 Trip characteristics- Survey-online UK & China

Looking at the length of the trips by taxi, more than half of the respondents in the UK used taxis for a short trip, which is consistent with the distribution of trip length by taxi in England (DfT 2019[15]). In China, instead medium and long distances account each for 36-37% of the trips, which seems reasonable given the larger extension of the Chinese cities, though data on the trip length of trips by taxi in China are not available. In line with these comments, the chi-squared test confirms that UK and China samples are different at 95%with respect to trip length.

Figures 4.1 and 4.2 report the income distribution by the trip segments respectively for UK and China. As we can see, while in both samples those in the low income categories tend to use taxi more for short trips, in China the difference is much less evident than in the UK. In the UK, only around 10.0% of respondents with disposable monthly income less than £2500, take a taxi for long trips (>=30 minutes). By comparison, in China, more than 30.0% of the respondents with disposable monthly income are less than ¥5000, still take a taxi for long trips (>= 30 minutes). This is certainly due to the large dimension of the Chinese cities, where a trip longer than 30 minutes is probably not considered a long trip. In addition, as shown in **Table 4-5,** in the UK respondents over 60 takes short trips 3 times more than respondents in China. And young respondents below 30 years in China take long trip 3 times more than young UK respondents do.



	< £500	£501-£1500	£1501-£2500	£2501-£3500	£3501-£4500	> £4500	Not reported
Long trip	4	19	9	6	3	1	8
Shorter trip	67	99	56	21	8	3	15

Figure 4.1 Income distribution by trip segments- Survey-online UK

15 DfT (2019). Taxi and Private Hire Vehicle Statistics: England 2019 (publishing.service.gov.uk)

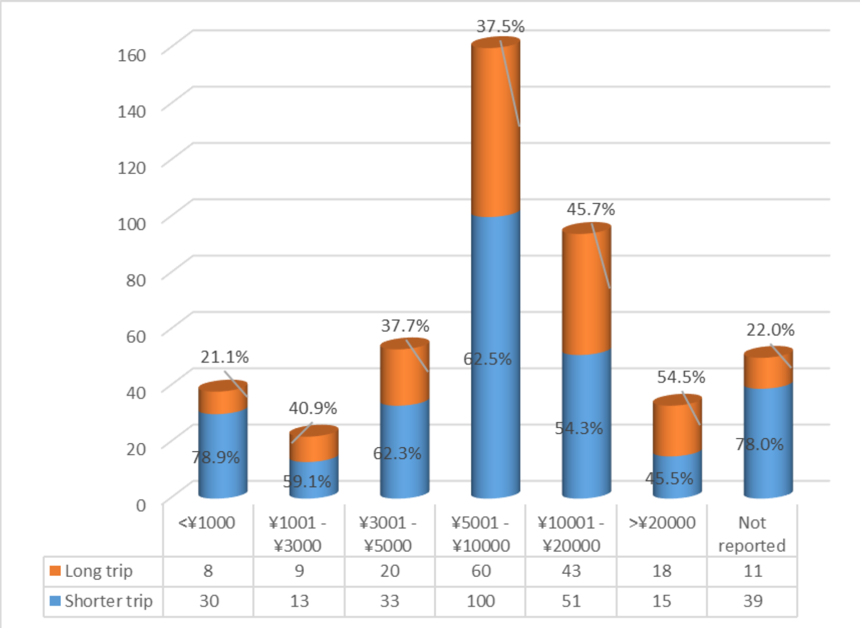


Figure 4.2 Income distribution by trip segments- Survey-online China

Age	UK N (%)			China N (%)			χ^2 test (p-value)
	Short	Medium	Long	Short	Medium	Long	
18-29	41 (12.9)	22 (6.9)	11 (3.4)	43 (9.6)	55 (12.2)	56 (12.4)	107.936 (0.000)
30-59	96 (30.1)	37 (11.6)	22 (6.9)	61 (13.6)	88 (19.6)	81 (18.0)	
60+	40 (12.5)	33 (10.3)	17 (5.3)	14 (3.1)	20 (4.4)	32 (7.1)	
All	177 (55.5)	92 (28.8)	50 (15.7)	118 (26.2)	163 (36.2)	169 (37.6)	

Table 4-5 Age distribution by trip segment- Survey-online UK & China

In both samples (the UK and China), most of the trips by taxi are made for leisure purposes (39.5%[16] in the UK and 38.0% in China, **Table 4-4**), but the second most common trip purpose in China is commuting while in the UK is going home. In this case, it was asked what was the activity performed before taking the taxi and, not surprisingly, in 74.0 % of the cases the purpose was leisure and visiting friends. Differences are also found in the other purposes and the chi-squared test confirms that UK and China samples are significantly different at 95%. It is interesting to note that in China, 26.2% of the respondents who are full-time employees use taxi for business or commuting, against 5.0% in the UK (**Table 4-6**).

[16] The percentage of leisure trips by taxi or PHV trips in England is 47%. This information is not available for the Chinese cities.

Trip purpose	UK N (%)		China N (%)		χ^2 test (p-value)
	Full-time employee	Other work status	Full-time employee	Other work status	
Business or commuting	16 (5.0)	15 (4.7)	118 (26.2)	34 (7.6)	61.482 (0.000)
Leisure	56 (17.6)	70 (21.9)	100 (22.2)	71 (15.8)	
Other purposes (holiday, going home etc.)	54 (16.9)	108 (33.9)	77 (17.1)	108 (11.1)	

Table 4-6 Trip purpose distribution by employment status- Survey-online UK & China

In terms of frequency of using taxis, it is evident (from **Table 4-7**) that Chinese respondents do use taxi much more frequently (79.3% use taxis at least once a month) than the UK respondents do (64.9% use taxis less than once a month). This is probably explained by the fact that in China the sample was gathered in main (very large) cities, while in the UK, everywhere (and taxi is used more in cities than in the countryside). But this difference might also root in differences in the economic and social system. In line with this argument, in China the majority of the trips are frequent and long (18.4%), while in the UK the majority of the trips are infrequent and short.

Frequency	UK N (%)			China N (%)			χ^2 test (p-value)
	Short trip	Medium trip	Long trip	Short trip	Medium trip	Long trip	
At least once a week	15 (4.7)	9 (2.8)	7 (2.2)	39 (8.7)	71 (15.8)	83 (18.4)	233.104 (0.000)
Less than once a week, at least once a month	49 (15.4)	21 (6.6)	11 (3.4)	48 (10.7)	62 (13.8)	54 (13.8)	
Less than once a month, more than twice a year	72 (22.6)	35 (11.0)	20 (6.3)	26 (5.8)	27 (6.0)	29 (6.4)	
At most twice a year	41 (12.9)	27 (8.5)	12 (3.8)	5 (1.1)	3 (0.7)	3 (0.7)	

Table 4-7 Frequency distribution of taxi usage by trip segment- Survey-online UK & China

Finally, from **Table 4-4**, we note that almost 40.0% of the UK respondents talk (somewhat or very) frequently with the taxi drivers during the trip, compared to only around 17.0% of the Chinese respondents. The chi-squared test confirms that the two samples are significantly different at 95% with respect to this characteristic. Interestingly, the far majority of the participants (more than 74%), both in the UK and in China, enjoy only occasionally to talk with the taxi driver, and the hypothesis that the two samples have the same distribution is rejected at 95% (p-value $0.010 < 0.05$). The proportion of Chinese respondents who never enjoy talking with the taxi driver is slightly higher.

4.2.3 Knowledge levels of AVs and ATs

Table 4-8 reports the characteristics of the samples with respect to the level of knowledge of AVs in general and ATs in particular. Four types of information were collected and are analysed here: whether respondents have heard of AVs, level of familiarity with the 5 levels of automation, whether they have heard of ATs testing/operating in the UK/China and from whom respondents heard about it.

Approximately 80% of both British and Chinese participants have heard of AVs in general before joining this survey. However, Chinese participants have more knowledge of the levels of automation than the UK participants do. The samples are significantly different at 95% with respect to knowledge of the 5 levels of automation. Among those who heard about AVs, in the UK only 37% of the respondents have also heard about the tests carried out, while in China, more than half of the participants (57.0%) have heard that there are AT operating. As shown in **Table 4-9**, there are gender difference regarding knowledge of AVs, in the UK, there are relatively more females who have not heard of AVs and are either not familiar or just slightly familiar with 5 levels of automation. While in China, this percentage is lower.

Knowledge levels of AVs and ATs		UK N (%)	China N (%)	χ^2 test (p-value)
Heard of AVs	Yes	255 (79.9)	362 (80.4)	0.03 (0.927)
	No	64 (20.1)	88 (19.6)	
Familiarity with the 5 levels of automation	Not at all familiar	102 (32.0)	87 (19.3)	53.84 (0.000)
	Slightly familiar	126 (39.5)	124 (27.6)	
	Moderately familiar	74 (23.2)	155 (34.4)	
	Very familiar	12 (3.8)	66 (14.7)	
	Extremely familiar	5 (1.6)	18 (4.0)	
Heard of ATs testing in the UK / operating in China	Yes	95 (29.8)	205 (45.6)	22.49 (0.000)
	No	160 (50.2)	157 (34.9)	
	Not heard of AVs at all	64 (20.1)	88 (19.6)	
From whom they heard about ATs testing in the UK / operating in China	From someone who participated in tests (UK) or used ATs (China)	23 (7.2)	49 (10.9)	0.003 (0.954)*
	From someone who heard about ATs or others	72 (22.6)	156 (34.7)	
	Not heard of ATs testing in the UK / operating in China	224 (70.2)	245 (54.4)	19.53 (0.000)

Table 4-8 Knowledge levels of AVs and ATs- Survey-online UK & China
(*) the chi-squared test is computed considering only the two groups who have heard of ATs testing in the UK/China

Among these who heard about ATs testing/operating, only in 24.0% of the cases (both in the UK and China) the information comes from someone who had direct experience, i.e. has participated directly in the trial (UK) or used the AT (China). In **Table 4-9**, among the females who heard of AVs, those who know about ATs testing/operating in their country only accounts for around 1/3 of the sample in the UK, 2/3 in China. This difference is not as pronounced for males. In all other cases, the information comes from someone who heard about it. Most of the respondents in the UK and in China heard of AVs and according to the chi-squared test (considering only those who have not heard of AT testing in the UK/China), no significant difference was observed between the two samples (p-value 0.954).

		UK N (%)		China N (%)		χ^2 test (p-value)
		Female	Male	Female	Male	
Heard of AVs	Yes	119 (37.7)	133 (42.1)	161 (35.9)	201 (44.8)	10.314 (0.016)
	No	45 (14.2)	19 (6.0)	39 (8.7)	48 (10.7)	
Familiarity with the 5 levels of automation	Not at all	73 (23.1)	29 (9.2)	44 (9.8)	42 (9.4)	70.345 (0.000)
	Slightly	67 (21.2)	57 (18.0)	57 (12.7)	67 (14.9)	
	Moderately	22 (7.0)	51 (16.1)	64 (14.3)	91 (20.3)	
	Very	2 (0.6)	10 (3.2)	27 (6.0)	39 (8.7)	
	Extremely	0 (0.0)	5 (1.6)	8 (1.8)	10 (2.2)	
Heard of ATs testing / operating	Yes	37 (11.7)	56 (17.7)	100 (22.3)	105 (23.4)	40.125 (0.000)
	No	82 (25.9)	77 (24.4)	61 (13.6)	96 (21.4)	
	Not heard of AVs	45 (14.2)	19 (6.0)	39 (8.7)	48 (10.7)	

Table 4-9 Knowledge of AVs or ATs distribution by gender- Survey-online UK & China

4.2.4 Psychological statements

Table 4-10 reports the average scores of the statements for each psychological latent construct. In general, we observe that the distribution of the rating scores is significantly different between the UK and China, with only two exceptions.

Constructs	Items	UK Mean	China Mean	t-tests	χ^2test （p-value）
Injunctive norms	IN1	4.14	4.90	-7.386	59.822 (0.000)
	IN2	4.33	4.12	1.838	18.203 (0.006)
	IN3	3.69	4.69	-9.912	105.410 (0.000)
Hedonic Motivation	HM1	4.24	5.25	-9.393	92.522 (0.000)
	HM2	4.34	5.13	-7.767	60.879 (0.000)
	HM3	4.73	4.58	1.305	14.756 (0.022)
Trust	T1	3.88	4.62	-5.792	53.392 (0.000)
	T2	3.80	4.40	-4.988	35.355 (0.000)
	T3	3.98	4.71	-5.864	46.978 (0.000)
	T4	3.79	4.84	-8.926	93.185 (0.000)

Table 4-10 Psychological statements- Survey-online UK & China

The direction of reversed statements IN2, HM3, T1, T2 and T3 was changed for comparison

For the injunctive norm, IN2 (*People who are important to me (friends, family) would think that using a fully automated taxi is not appropriate*), and the hedonic motivation, HM3 (*I believe using a fully automated taxi without driver will be boring*), the assumption of equal average scores between UK and China is rejected respectively at 94% and 81% (two-tail test). We also note that the average scores of all the trust statements in the China sample are higher than in the UK sample, indicating that Chinese respondents might have a higher trust in ATs.

Constructs	Items	UK Mean		China Mean		t-tests UK-Chjna	
		Age18-29	Age 30+	Age18-29	Age 30+	age 18-29	age 30+
Injunctive norms	IN1	4.41	4.06	4.45	5.13	-0.246	-9.059
	IN2	4.09	4.40	3.84	4.26	1.176	1.029
	IN3	3.76	3.67	4.09	5.01	-1.807	-11.389
Hedonic Motivation	HM1	4.66	4.12	5.03	5.36	-1.969	-9.606
	HM2	4.73	4.22	4.84	5.28	-0.652	-8.452
	HM3	4.72	4.73	4.47	4.64	1.142	0.700
Trust	T1	3.97	3.85	4.27	4.80	-1.370	-6.070
	T2	3.73	3.82	4.10	4.55	-1.840	-4.955
	T3	4.11	3.95	4.67	4.73	-2.636	-5.145
	T4	4.01	3.72	4.42	5.07	-1.789	-9.734
Sample size		74	245	154	296		

Table 4-11 Psychological statements by age- Survey-online UK & China

As shown in **Table 4-11**, age seems to have an impact on the average score of the statements, and the impact is stronger in China than in the UK. In China, people older than 30 years scored higher than those aged 30 or younger in all statements in all the psychological constructs. The mean values between age groups are also significantly different at 99%. This indicates that older people in China are more likely to conform to or to receive others' suggestions, they do perceive AT as enjoyable and they trust AT more than people 30 years old or younger do. Similar effect is observed in the UK, but the average scores are significantly different only in two statements: HM1 (*I believe using a fully automated taxi will be pleasant*) and T3 (*I don't trust that a computer can drive vehicle without assistance from the driver*). Moreover, in China respondents scored higher than in the UK in all statements.

Additionally, **Table 4-12** shows that in both UK and China, those who are (very or extremely) familiar with the 5 levels of automation have a high tendency to comply more with others' approval, enjoy ATs and trust them. As expected, the more familiar they are with AVs, the more they trust ATs. Even in this case, though, the effect is stronger among Chinese than among UK respondents. In the Chinese sample, the assumption of equal average score between those who are familiar and those who are not is rejected at 99% in all statements, except HM3 (*I believe using a fully automated taxi without driver will be boring*). In the UK sample, instead, it is rejected at less than 92% in all statements except IN3 (*People who are important to me (friends, family) would think that more people should use fully automated taxis*) and T1 (*Overall, I do not trust fully automated taxis*).

Constructs	Items	UK Mean		China Mean		t-tests UK-China	
		Familiar	Not Familiar	Familiar	Not Familiar	Familiar	Not Familiar
Injunctive norms	IN1	5.00	4.09	5.62	4.73	-1.788	-5.964
	IN2	5.12	4.28	4.64	3.99	0.991	2.429
	IN3	4.12	3.67	5.35	4.54	-3.494	-8.252
Hedonic Motivation	HM1	5.24	4.19	5.75	5.13	-1.500	-8.301
	HM2	5.65	4.27	5.83	4.97	-0.590	-6.568
	HM3	5.12	4.71	4.69	4.55	0.936	1.275
Trust	T1	4.06	3.87	5.19	4.49	-2.360	-4.613
	T2	5.00	3.74	5.27	4.20	-0.589	-3.798
	T3	4.82	3.94	5.23	4.59	-0.833	-5.084
	T4	4.88	3.73	5.48	4.70	-1.516	-7.771
Sample size		17	302	84	366		

Table 4-12 Psychological statements by familiarity of automation levels- Survey-online UK & China

4.2.5 Stated Choices

Finally, this section briefly analyses descriptively the choices made in the SC experiments. Since each individual responded 6 choice scenarios, the sample analysed in this section consists of 1,914 pseudo-observations in the UK and 2,700 pseudo-observations in China. **Table 4-13** presents the choice share by length of the trip by taxi (as chosen in the SC experiment). In line with the distance travelled in the last trip by taxi, discussed in **Section 4.2.1**, in China the long distance segment is chosen almost twice as much as in the UK and, interestingly, UK participants chose overall NT significantly more (64%) than Chinese participants (49%). This seems to be related to the fact that in the UK there are more short trips and AT is chosen more for short than for longer trips.

SC segments	UK N (%)		China N (%)		χ^2 test (p-value)
	NTs	ATs	NTs	ATs	
Short	689 (36.0)	373 (19.5)	363 (13.4)	345 (12.8)	537.877 (0.000)
Medium	338 (17.7)	214 (11.2)	504 (18.7)	474 (17.6)	
Long	203 (10.6)	97 (5.1)	451 (16.7)	563 (20.9)	
All	1230 (64.3)	684 (35.7)	1318 (48.8)	1382 (51.2)	

Table 4-13 Stated Choices by trip distance segment- Survey-online UK & China

Tables 4-14 reports the stated choice share by age and gender. Results show that respondents younger than 30 years old, in both UK and China, chose NT more than ATs. But more interestingly, we note that while in the UK also those older than 30 years still choose NT more than AT, in China is the opposite, clearly indicating an age effect on the choice of type of taxi. Interestingly, in our samples, instead the assumption that the distribution of age is the same among type of taxis can be rejected at 83%. We expected a stronger result from the chi-squared test, given the significance of male in the choice of AT, reported in the literature.

		UK N (%)		China N (%)		χ^2 test (p-value)
		NTs	ATs	NTs	ATs	
Age	18-29	266 (13.9)	178 (9.3)	502 (19.3)	404 (15.0)	203.388
	30+	964 (50.4)	506 (26.4)	798 (29.6)	978 (36.2)	(0.000)
Gender	Female*	660 (34.8)	324 (17.1)	572 (21.2)	628 (23.3)	135.963
	Male	553 (29.2)	359 (18.9)	740 (27.5)	754 (28.0)	(0.000)

Table 4-14 Stated Choices classified by age and gender- Survey-online UK & China

….* 3 people in the UK and 1 in China preferred not to declare their gender. Given the low numbers, these are not included in the table.

Table 4-15 presents the stated choices classified by whether respondents would like driver to help with luggage or not. As we can see, in China, there are no obvious difference in the choices between ATs and NTs. On the contrary. in the UK, those who like driver carry the luggage selected NT more than twice as much as AT. This effect will be explored in **Section 6** when the results of the demand models will be discussed.

		UK N (%)		China N (%)		χ^2 test (p-value)
		NTs	ATs	NTs	ATs	
Like driver to carry luggage	Yes	879 (45.9)	400 (20.9)	847 (31.4)	899 (33.3)	131.345
	No	351 (18.3)	284 (14.8)	471 (17.4)	483 (17.9)	(0.000)
Familiarity with the 5 levels of automation	Not at all	436 (22.8)	166 (8.7)	303 (11.2)	219 (8.1)	371.169 (0.000)
	Slightly	455 (23.8)	271 (14.2)	388 (14.4)	355 (13.1)	
	Moderately	295 (15.4)	194 (10.1)	410 (15.2)	520 (19.3)	
	Very	34 (1.8)	33 (1.7)	175 (6.5)	221 (8.2)	
	Extremely	10 (0.5)	20 (1.0)	42 (1.6)	66 (2.4)	
Heard of ATs	Yes	366 (19.1)	244 (12.7)	512 (19.0)	718 (26.6)	194.263 (0.000)
	No	605 (31.6)	320 (16.7)	495 (18.3)	447 (16.6)	
	Not heard of AVs	259 (13.5)	120 (6.3)	311 (11.5)	217 (8.0)	

Table 4-15 Stated Choices classified if like driver carrying luggage and knowledge level on AVs and ATs- Survey-online UK & China

4.3 Data from the *Survey-online NCL* and the *Survey-VR NCL*

As mentioned in **Section 3.9**, the final valid samples collected in Newcastle upon Tyne (NCL) consists of 139 valid responses (834 pseudo-individual SC) from the screen-based *Survey-online* NCL and 40 valid responses (240 pseudo-individual SC) from the *Survey-VR*. We note that this sample size is small but in line with other immersive VR experiments published so far. For example 48 respondents from University staff and students is the sample size used in Bogacz *et al*. (2021); 36 respondents is the sample size in Feng *et al*. (2022); 28 participants is the sample size in Zou *et al*. (2021). Nevertheless, the sample is small to get robust results for the psychological constructs. These will not be analysed.

Having the same respondent participating in both surveys (*Survey-online* and *Survey-VR*) would have been beneficial for the purpose of comparison. However, this requires controlling for the order effect (*VR-based SC* answers could be affected by *screen-based online* SC answers or vice versa), and this requires at least twice as many respondents. Different samples were then collected to avoid order effects. The population in the Newcastle metropolitan area at end of 2019 counted 0.3 million people. Neither samples can be considered representative, especially the VR sample, even though an attempt was made to match at least the age distribution of both samples with that of Newcastle, as discussed in the next section.

4.3.1 Socio-demographic characteristics

Table 4-16 reports the distribution of the socio-demographic characteristics for the two samples (online and VR) collected in Newcastle upon Tyne. As expected, given the sample size and the different recruitment methods (as discussed in **Section 3.9**), the two samples are significantly different in all the socio-demographic characteristics, except in terms of age (p-value 0.332) and income (p-value 0.152). Interestingly, the age distribution in both samples reflects that of Newcastle city[17] that is 33.5% 18-29 years old and 66.5% 30 years or older. On the other hand, in both samples people with higher education (master and PhD) seem to be overrepresented. At UK nation-level[18], only about 10% of the population holds a MSc or PhD degree. We do not have the distribution for Newcastle, but we note that in Newcastle, there are two large universities, hence the proportion of highly educated people can be higher than the national average. Nevertheless 60% of the sample with MSc or PhD in the *VR-based SC* sample is particularly high. This reflects the fact that, due to Covid pandemic, part of the participants was recruited at Newcastle University. In the *online screen-based* sample the proportion of highly educated people is 1/3 that in the *VR-based* sample (22.3%) but still twice the national value. Finally, the average monthly disposable income in Newcastle in 2022 is £802.00[19] and the median monthly wage was £2061.00 (after-tax wage). As mentioned before, a taxi in the UK costs approximately £15.00 for a trip of 20 minutes (travel time), in both VR and online sample 95% respondents spend less than 5% of their disposable income for travelling by taxi.

---

[17] 2021 04 NFNA City Profile.pdf (newcastle.gov.uk)

[18] Education level proportion at Newcastle-city level was not found.

[19] https://www.finder.com/uk/disposable-income-around-the-uk

Socio-demographic characteristics		Online SC N (%)	VR- SC N (%)	χ^2 test (p-value)
Gender	Female	79 (56.8)	15 (37.5)	4.66 (0.047)
	Male	60 (43.2)	25 (62.5)	
Age	Below 30 years old	40 (28.8)	15 (37.5)	1.11 (0.332)
	30 years old or above	99 (71.2)	25 (62.5)	
Education level	Bachelor degree or below	108 (77.7)	16 (40.0)	20.74 (0.000)
	Master or Doctorate degree	31 (22.3)	24 (60.0)	
Current work status	Full-time employees	76 (54.7)	14 (35.0)	24.21 (0.000)
	Students	15 (10.8)	18 (45.0)	
	Others	48 (34.5)	8(20.0)	
Personal monthly disposable income	Less than £500	38 (27.3)	7 (17.5)	5.29 (0.152)
	£501-£1500	61 (43.9)	20 (50.0)	
	£1501-£ 2500	22 (15.8)	11 (27.5)	
	£2501- more	12 (8.6)	1 (2.5)	
	I do not wish to disclose it	6 (4.3)	1 (2.5)	
Sample size		139	40	

Table 4-16 Socio-demographic characteristics – Survey online NCL & Survey VR NCL

The *VR-based SC* sample has also a significantly higher proportion of male (62.5%) and students (45.0%) compared to the *screen-based online*. This also reflects the fact that many participants came from the school of Engineering, where there proportion of male is a higher than that of female. On the other hand, the proportion of the employed in the online sample (54.7% full-time and 16.5% part-time) is 72.2%, which is close to the 70.0% employment rate in Newcastle, among the working-age population in 2019[20].

4.3.2 Trip characteristics

Table 4-17 reports the trip characteristics of the two samples. Differently from the survey collected in the UK (as discussed in **Section 4.2.2**) in these samples respondents are assumed to be (or found themselves in the VR experiment) in Northumberland Rd. and they were asked to report the activities performed before going to Northumberland Rd. to take the taxi. All other characteristics are the same as collected in the *survey-online UK*.

Interestingly, despite the socio-demographic characteristics are significantly different between the two samples, the trip characteristics are instead similar, with the exception of the distance of the destinations selected and the activity performed before taking the taxi in Northumberland Rd.. Looking at the trip distance (SC segment), more than half of the participants in the *screen-based online* survey selected a destinations in the 5km segment. This is consistent with the *Survey-online UK* and also consistent with the distribution of trip length by taxi in England.

[20] Statistics and intelligence | Newcastle City Council

Differently, participants in the VR experiment choose evenly 5km and 10km. Some studies in the literature (Meißner *et al.*, 2020) report that variety seeking behaviour increases in VR experiments. Variety seeking behaviour refers to respondents being more open to explore new products under new situations and environments (Kahn, 1995). It can be that in the VR experiment respondents have selected also less common destinations, i.e. where they go less frequently in the real life.

Trip characteristics		Online SC N (%)	VR SC N (%)	χ^2 test (p-value)
SC segment	5km destinations	461 (55.3)	91 (37.9)	30.20 (0.000)
	10km destinations	169 (20.3)	86 (35.8)	
	15km destinations	204 (24.5)	63 (26.3)	
Activities at Northumberland Rd	Commuting or Business	5 (3.6)	13 (32.5)	46.65 (0.000)
	Leisure (e.g. shopping, pub)	127 (91.4)	18 (45.0)	
	Others (visiting friends, holiday, at home, etc)	7 (5.0)	9 (22.5)	
Frequency of using taxis	At least once a week	12 (8.6)	2 (5.0)	1.07 (0.785)
	Less than once a week, at least once a month	50 (36.0)	13 (32.5)	
	Less than once a month, more than twice a year	62 (44.6)	21 (52.5)	
	At most twice a year	15 (10.8)	4 (10.0)	
Frequency of talking with the driver	Very infrequently	7 (5.0)	6 (15.0)	7.18 (0.127)
	Somewhat infrequently	17 (12.2)	8 (20.0)	
	Occasionally	47 (33.8)	9 (22.5)	
	Somewhat frequently	46 (33.1)	12 (30.0)	
	Very frequently	22 (15.8)	5 (12.5)	
Enjoying talking with taxi driver	Always	12 (8.6)	6 (15.0)	1.91 (0.385)
	Sometimes	114 (82.0)	29 (72.5)	
	Never	13 (9.4)	5 (12.5)	
Like the driver to help carry luggage	Yes	77 (55.4)	27 (67.5)	1.87 (0.205)
	No	62 (44.6)	13 (32.5)	
Take the taxi without any help	Yes	138 (99.3)	40 (100.0)	N.A
	No	1 (0.7)	0 (0.0)	

Table 4-17 Trip characteristics – Survey online NCL & Survey VR NCL

The distribution of the activity performed before taking the taxi is interesting and it seems to reflect once more the differences in the socio-demographic characteristics between the two samples. Northumberland Rd. is a side street of the most important shopping street in Newcastle. It is not surprising that in the *screen-based* sample in more than 91.0% of the cases the activities performed before taking the taxi is leisure. At the same time, part of the *VR-based* sample was recruited at the University (that is very closed to Northumberland Rd.), which might explain why in the *VR-based* sample only in 45.0% of the cases the activities performed before taking the taxi is leisure, while in 32.0% of the cases is commuting or business. With respect to the frequency of using taxis, most respondents in both surveys used taxis less than once a month, followed by at the least once a month. This result is consistent with the result in *survey online UK*, where more than half respondents used taxi at less than once a month.

In terms of interaction with taxi driver, in both samples, around 40.0% of the participants talk frequently with taxi drivers during the trip. However, the far majority of them only sometimes enjoy talking with a taxi driver and nearly all respondents are able to take a taxi without other helps. Finally, in both samples collected in Newcastle (*online-based* and *VR-based)*, the majority of the sample enjoyed that the taxi driver helps them to carry luggage or heavy bags. All these characteristics are also consistent with the UK sample (**Table 4-4**).

4.3.3 Knowledge levels of AVs or ATs

Table 4-18 reports the characteristics of the samples regarding knowledge of AVs and ATs. It is interesting to note that the two samples are not significantly different in all these characteristics (all p-values are > 0.05), except of the last one (from whom they have heard about ATs testing in the UK). The far majority of the respondents in both samples have heard about AVs, though the vast majority have no or little knowledge about them and have never heard about the testing carried out in the UK. Among the few (30% in the *VR-based* sample and 17% in the *online-based* sample) who have heard about the testing, in the majority of the case this is through a 'word-of-mouth'. This is in line with the results found in the *Survey-online UK* and *Survey-online China* (**Section 4.2.3**).

Knowledge levels of AVs and ATs		Online SC N (%)	VR- SC N (%)	χ^2 test (p-value)
Heard of AVs	Yes	107 (77.0)	32 (80.0)	0.16 (0.686)
	No	32 (23.0)	8 (20.0)	
Familiar with 5 levels of automation	Not at all familiar	46 (33.1)	8 (20.0)	6.86 (0.076)
	Slightly familiar	60 (43.2)	13 (32.5)	
	Moderately familiar	25 (18.0)	12 (30.0)	
	Very or extremely familiar	8 (5.7)	5 (12.5)	
Heard of ATs testing in the UK	Yes	24 (17.3)	12 (30.0)	3.14 (0.208)
	No	83 (59.7)	20 (50.0)	
	Not heard of AVs at all	32 (23.0)	8 (20)	
From whom they heard of ATs testing in the UK	From someone who participated in the test	10 (7.2)	2 (5.0)	2.25 (0.260)*
	From someone or others who heard about ATs	14 (10.0)	10 (25.0)	
	Not heard of ATs testing in the UK	115 (82.7)	28 (70.0)	6.02 (0.049)

Table 4-18 Knowledge levels of AVs and ATs – Survey online NCL & Survey VR NCL

* it means that the chi-squared test is computed only between the first two categories, without 'not heard of ATs testing in the UK'

Although not statistically significantly different, we note that in the *VR-based* sample the proportion of respondents who have heard of AV testing through 'word-of-mouth' is higher (25%) than in the *online-based* sample (10%). Analogously the proportion of those who are (moderately or very) familiar with AV (42.5%) is higher than in the *online-based* sample (23.7%). This difference, which as said is not statistically different (maybe due to the sample size), might be due to the higher proportion of participants with high education level recruited in VR experiment. In accordance with the literature review, they are more familiar with the innovative technology.

4.3.4 Stated choice and other VR questions

This section briefly analyses descriptively the choices made in the stated choice experiments and some post-questions related to VR experience (their previous VR experience and some questions about VR-SC experiments). With each individual answering 6 choice scenarios, the samples analysed in this section consist of 834 pseudo-observations for the *online-based NCL* survey and 240 pseudo-observations for the *VR-based NCL* survey. **Table 4-19** presents the choice share between NTs and ATs in these two samples. Interestingly, ATs is chosen slightly less in the *online-based* survey (41.5%) than in the *VR-based* survey, but this difference is not significant at 95%. This is more marked for short trips than longer trips. It is instead reversed for medium distance trips. This phenomenon was also observed in the UK online survey **in section 4.2.5.** It seems that Newcastle respondents (also British respondents) are less willing to embrace and use the innovative ATs. Though this result is not confirmed in the VR experiment. This effect will be discussed more in **Section 6** in the model estimation.

SC segments	Online NCL N (%)		VR NCL N (%)		χ^2 test (p-value)
	NTs	ATs	NTs	ATs	
Short	275 (33.0)	186 (22.3)	46 (19.2)	45 (18.8)	41.366 (0.000)
Medium	98 (11.8)	71 (8.5)	35 (14.6)	51 (21.3)	
Long	115 (13.8)	89 (10.7)	33 (13.8)	30 (12.5)	
All	488 (58.5)	346 (41.5)	114 (47.5)	126 (52.5)	

Table 4-19 Stated Choices – Survey online NCL & Survey VR NCL

Tables 4-20 reports the stated choice share by age and gender. Results show that in the *online-based* survey respondents younger than 30 years chose NTs more than ATs. while in *VR-based* survey the share is opposite, but less marked. Interestingly, we found that the assumption that the distribution of gender is the same between ATs and NTs can be rejected at 83%. We expected instead a higher percentage of ATs among male, given the significance of male in the choice of

AT, reported in the literature.

		Online NCL N (%)		VR NCL N (%)		χ^2 test (p-value)
		NTs	ATs	NTs	ATs	
Age	18-29	134 (16.1)	106 (12.7)	42 (17.5)	48 (20.0)	15.292
	30+	354 (42.4)	240 (28.8)	72 (30.0)	78 (32.5)	(0.002)
Gender	Female	300 (36.0)	174 (20.9)	41 (17.1)	49 (20.4)	36.271
	Male	188 (22.5)	172 (20.6)	73 (30.4)	77 (32.1)	(0.000)

Table 4-20 Stated Choices classified by age and gender – Survey online NCL & Survey VR NCL

Table 4-21 presents the stated choices classified by whether respondents would like driver to help with luggage, by familiarity with ATs and whether they heard about ATs. As we can see, the samples differ significantly in all these aspects, and in particular, among those who have heard about ATs, we note that twice as many respondents in the VR experiment have heard of AT testing in the UK, compared to the respondents in the online survey.

		Online N (%)		VR N (%)		χ^2 test (p-value)
		NTs	ATs	NTs	ATs	
Like driver to carry luggage	Yes	277 (33.2)	185 (22.2)	76 (31.7)	86 (35.8)	21.520
	No	211 (25.3)	161 (19.3)	38 (15.8)	40 (16.7)	(0.000)
Familiarity with the 5 levels of automation	Not at all	180 (21.6)	96 (11.5)	26 (10.8)	22 (9.2)	73.316 (0.000)
	Slightly	210 (25.2)	150 (18.0)	32 (13.3)	46 (19.2)	
	Moderately	78 (9.4)	72 (8.6)	34 (14.2)	38 (15.8)	
	Very	13 (1.6)	23 (2.8)	17 (7.1)	13 (5.4)	
	Extremely	7 (0.8)	5 (0.6)	5 (2.1)	7 (2.9)	
Heard of ATs testing in the UK	Yes	72 (8.6)	72 (8.6)	34 (14.2)	38 (15.8)	38.304 (0.000)
	No	284 (34.1)	214 (25.7)	64 (26.7)	56 (23.3)	
	Not heard of AVs	132 (15.8)	60 (7.2)	16 (6.7)	32 (13.3)	

Table 4-21 Stated Choices classified by if like driver carrying luggage and knowledge level on AVs and ATs – Survey online NCL & Survey VR NCL

As discussed in **Section 3.9**, after running the *survey-VR NCL*, respondents were asked to answer a short post survey to investigate their experience with the VR. As can be seen in **Figure 4.3**, the majority of the participants were not familiar with the VR settings, half of them had not tried a VR setting before and 25% had tried it only once.

The vast majority of the respondents found it very easy to perform the choice tasks in the VR environment and felt the VR was highly realistic (75%), though this does not necessarily mean that everybody felt as if they were making a choice in reality (**Figure 4.4)**. The question about the colour of the taxis was added to test the concern that there could be a colour effect in the experiment and this could affect the preferences elicited. Interestingly, results show that only 50% remember the colour of the normal taxi, while 67% remember that of the automated taxi

(**Figure 4.5**). This seems to suggest that there might not be a colour effect. Among the 33% of the respondents who did not notice it, 47.4% chose NT and 52.6% AT (78 times). By comparison, 67% of those who noticed it, 47.5% chose NT and 52.5% AT (162 times). According to chi-squared test, there was no significant difference (p-value= 1.000). Almost all participants noticed the people waiting in the queues, confirming that 1) the setting of the scene proposed was correct (i.e. the position of the ticket board allows a good view of the context, as described in **Section 3.7** *VR-based SC*); and 2) participants did look around at the context. This is one of the key features of the immersive experience.

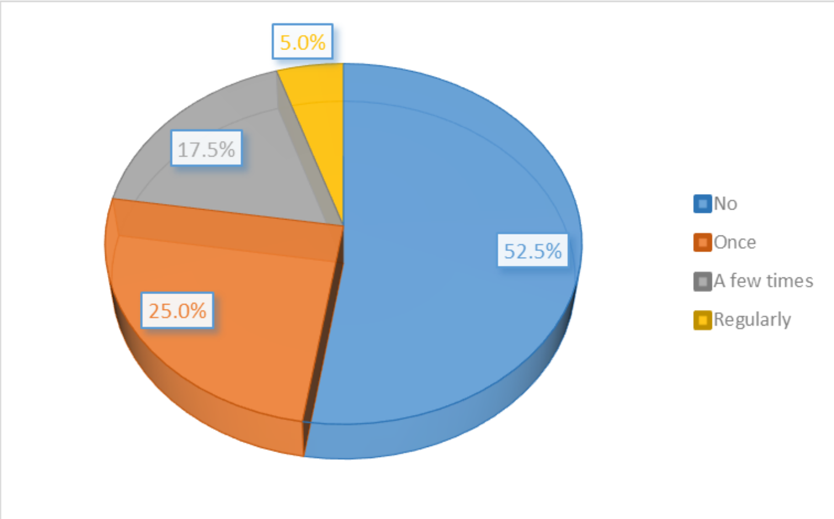


Figure 4.3 Previous VR experience-Survey VR NCL

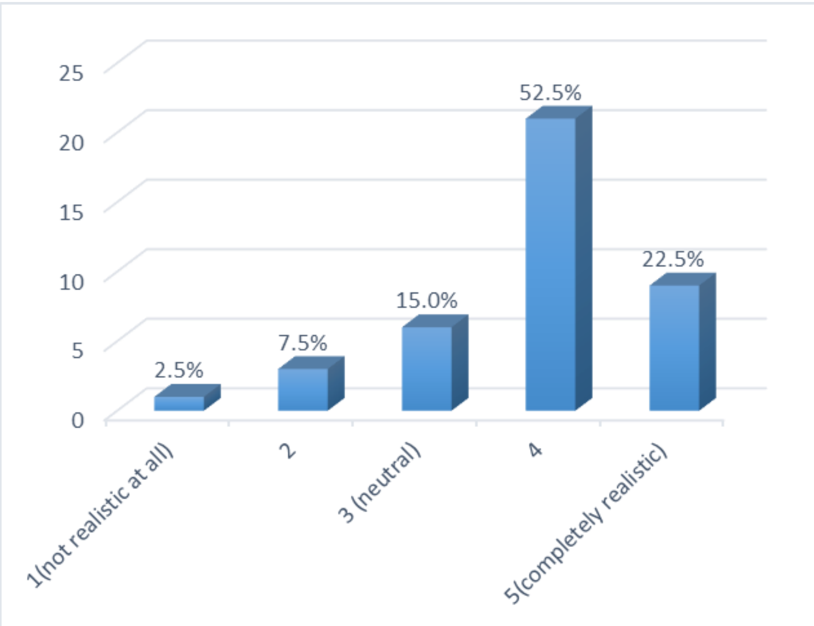


Figure 4.4 Reality of VR experiment-Survey-VR NCL

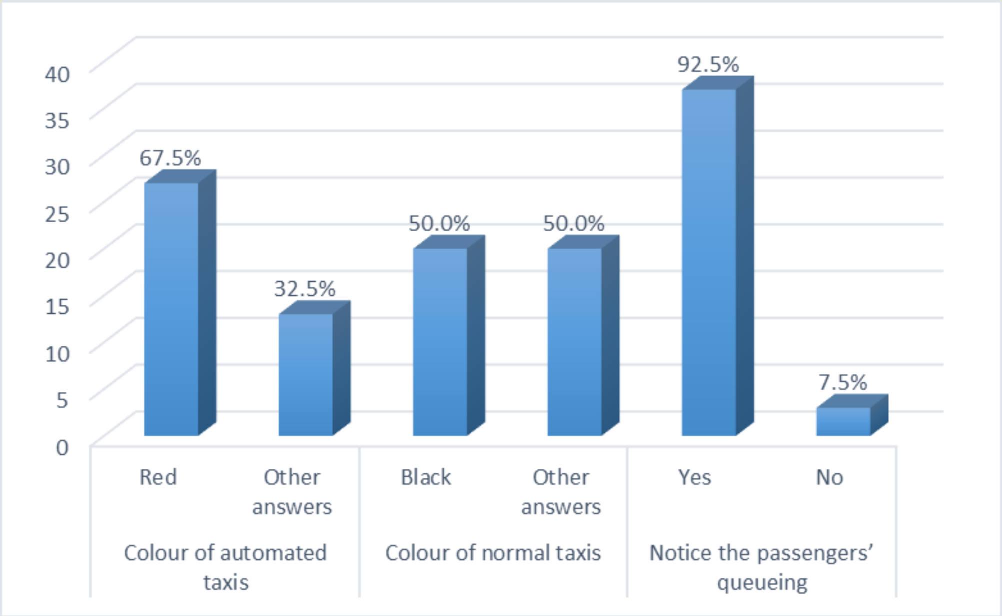


Figure 4.5 Respondents noticing ATs and NTs colours and passengers queuing-Survey-VR NCL

4.4 Data from the *Survey-online China – AT existence*

As discussed in **Section 3.1**, in China we collected 485 responses, of which 450 comes from respondents who have never used an AT (these were analysed in **Section 4.2**) and 35 from respondents who have tried an AT. Among the 450 respondents who never tried an AT, 299 live in cities where AT operates while 151 live in cities where AT are not operating. This section reports a comparison among these three subsamples: AT users (group A), NT users living in AT cities (defined as AT cities, group B), and NT users living in NT cities (defined as NT cities, group C).

4.4.1 Socio-demographic characteristics

Table 4-22 reports the distribution of the socio-demographic characteristics in these three subsamples. We note first that some categories are aggregate, because they have few respondents, and that the chi-square test is conducted only to check differences between the groups B and C (i.e. non-AT users), because the number of AT users is quite small. The chi-square test requires 25% of the expected cell frequency to be greater than 5. The income variable has some unreported values from respondents in NT cities, the comparison with AT cities was then not analysed. The distribution with all the categories is reported in **Appendix 4, Table 4-C**.

Socio-demographic characteristics		Group A AT users	Group B AT cities	Group C NT cities	χ^2 test B-C (p-value)
		N (%)	N (%)	N (%)	
Gender	Female	24 (68.6)	145 (48.5)	57 (37.1)	4.926
	Male	11 (31.4)	154 (51.5)	95 (62.9)	(0.028)
Age	Below 30	7 (20.0)	65 (21.7)	89 (58.9)	61.683
	30 or above	28 (80.0)	234 (78.3)	62 (41.1)	(0.000)
Education level	Below Bachelor degree	3 (8.6)	77 (25.8)	45 (29.8)	0.832
	Bachelor degree or above	32 (91.4)	222 (74.2)	106 (70.2)	(0.371)
Current work status	Employed full-time	30 (85.7)	203 (67.9)	92 (60.9)	2.156
	Non employed full-time	5 (14.3)	96 (32.1)	59 (39.1)	(0.143)
Sample size		35	299	151	

Table 4-22 Socio-demographic characteristics – Survey online China

As we can see, there are significant differences among these three subsamples. Among those who tried an AT, 2/3 are females, against half of those who never tried an AT even though they live in a city where AT is operating and 1/3 of those who never tried an AT and living in a city where AT is not operating. The difference is remarkable and interesting, and the chi-square test between none- AT users (groups B and C) confirms that there is a significant gender difference between these two groups (p-value 0.028< 0.05)

In terms of age, in the cities where AT is operating, 80% of the sample (with no difference between AT and non-AT users) is 30 years or older, against 41% of the sample in the cities where AT is not operating. This can be due to the characteristic of the cities, even though we tried to match at least age and gender distribution while sampling.

In terms of education level, as discussed in **Section 4.2.1** taxi in China is used more by people with university education. From **Table 4-22**, it seems that this effect is even more marked for AT users (more than 91% of the AT users have a university degree against around 70% of the non- AT users). This is in line with the research discussed in the literature review that found a correlation between level of education and AV acceptance (willingness to use and more positive attitude, (e.g. Haboucha *et al.*, 2017). Interestingly, instead the level of education is not significantly different among non-AT users whether they live in a city where AT is operating or not (p-value 0.371> 0.05).

Finally, there is a higher proportion of full-time employees among AT users, but it is not possible to say if this is significantly different from the non- AT users. It is only possible to say that the proportion of non-AT users full-time employees is not significant different whether they live in a city where AT is operating or not (p-value 0.143> 0.05).

4.4.2 Trip characteristics

Table 4-23 reports the characteristics of the last trip made by taxi as reported by the respondents, distinguished by AT and non- AT users and whether they live in a city where AT is operating

Trip characteristics		Group A AT users N (%)	Group B AT cities N (%)	Group C NT cities N (%)	χ^2 test B-C (p-value)
Self-reported travel time of a recent trip	Short (10 minutes or less)	7 (20.0)	51 (17.1)	67 (44.4)	43.839 (0.000)
	Medium (around 20 minutes)	20 (57.1)	112 (37.5)	51 (33.8)	
	Long (30 minutes or more)	8 (22.9)	136 (45.5)	33 (21.8)	
Activities before taxi trip	Commuting or Business	11 (31.4)	88 (29.4)	55 (36.4)	6.900 (0.032)
	Leisure (e.g. shopping, pub)	16 (45.7)	112 (37.5)	38 (25.2)	
	Others (visiting friends, holiday, at home, etc)	8 (22.9)	99 (33.1)	58 (38.4)	
Trip purpose	Commuting or Business	12 (34.2)	96 (32.1)	56 (37.1)	1.501 (0.472)
	Leisure (e.g. shopping, pub)	11 (31.4)	119 (39.8)	52 (34.4)	
	Others (visiting friends, holiday, at home, etc)	12 (34.3)	84 (28.1)	43 (28.5)	
Enjoying talking with taxi driver	Always	7 (20.0)	27 (9.0)	8 (5.3)	1.961 (0.375)
	Sometimes	25 (71.4)	220 (73.6)	115 (76.2)	
	Never	3 (8.6)	52 (17.4)	28 (18.5)	
Frequency of talking with the driver	Very infrequently	2 (5.7)	32 (10.7)	22 (14.6)	3.227 (0.521)
	Somewhat infrequently	2 (5.7)	43 (14.4)	20 (13.2)	
	Occasionally	14 (40.0)	168 (56.2)	87 (57.6)	
	Somewhat frequently	13 (37.1)	46 (15.4)	20 (13.2)	
	Very frequently	4 (11.4)	10 (3.3)	2 (1.3)	
Frequency of using taxis	At least once a week	26 (74.3)	156 (52.2)	37 (24.5)	44.845 (0.000)
	Less than once a week, at least once a month	8 (22.9)	104 (34.8)	60 (39.7)	
	Less than once a month, more than twice a year	1 (2.9)	36 (12.0)	46 (30.5)	
	At most twice a year	0 (0.0)	3 (1.0)	8 (5.3)	
Like the driver to help carry luggage	Yes	24 (68.6)	205 (68.6)	86 (57.0)	5.917 (0.016)
	No	11 (31.4)	94 (31.4)	65 (43.0)	
Take the taxi without any help	Yes	35 (100.0)	298 (99.7)	149 (98.7)	N.A.
	No	0 (0.0)	1 (0.3)	2 (1.3)	

Table 4-23 Trip characteristics– Survey online China

As before, due to low frequency, two categories (activity before the trip and trip purpose) have been aggregated. The distribution with all the categories is reported in **Appendix 4, Table 4-D.** As it can be seen, the two samples have different distribution for all the trip characteristics, except trip purpose, whether enjoying talking with driver and frequency of talking with driver. In particular, in cities where AT is operating, a higher proportion of respondents use taxi frequently (at least once a week) and make long trip (30 minutes or longer). Both effects are due probably to the size of the cities. Although all major cities were selected in China for recruitment, the cities where ATs are operating are the biggest.

4.4.3 Knowledge of AVs and ATs

Table 4-24 reports the characteristics of the samples with respect to the level of knowledge of AVs in general and ATs in particular, distinguished by AT and non -AT users and whether they live in a city where AT is operating or not.

As expected, respondents living in AT cities have much more knowledge of AVs and ATs that those living in NT cities, and this difference is significant at 99%. Very few respondents (14%) have not heard of AVs at all (against 30% of the respondents living in cities where ATs are not operating). However, surprisingly, almost 48% of those living in cities where AT do operate are not aware that the system is operating in their city.

Knowledge of AVs and ATs		Group A AT users N (%)	Group B AT cities N (%)	Group C NT cities N (%)	χ^2 test B-C (p-value)
Heard of AVs	Yes	35.(100.0)	257(86.0)	105 (69.5)	17.189 (0.000)
	No	0 (0.0)	42(14.0)	46 (30.5)	
Familiar with 5 levels of automation	Not at all familiar	0 (0.0)	37 (12.4)	50 (33.1)	49.853 (0.000)
	Slightly familiar	6 (17.1)	69 (23.1)	55 (36.4)	
	Moderately familiar	6 (17.1)	118 (39.5)	37 (24.5)	
	Very familiar or above	23 (65.7)	65 (25.1)	9 (5.9)	
Heard of ATs operating in China	Yes	35(100.0)	156 (52.2)	49 (32.5)	22.710 (0.000)
	No	0 (0.0)	101 (33.8)	56 (37.1)	
	Not heard of AVs at all	0 (0.0)	42 (14.0)	46 (30.5)	
From whom they heard of ATs operating in China	From whom have used ATs	24 (68.6)	39 (13.0)	10 (6.6)	16.089 (0.000) 0.432 (0.569)*
	From whom only heard about ATs or others	11 (31.4)	117 (39.1)	39 (25.8)	
	Not heard of ATs operating in China	0 (0.0)	143 (47.8)	102 (67.6)	

Table 4-24 Knowledge of AVs and ATs – Survey online China

* it means that the chi-squared test is computed only between the first two categories, without 'not heard of ATs testing in the UK'

4.4.4 Psychological statements

Table 4-25 reports the mean values of the scores of the statements for the psychological constructs, distinguished by AT and no AT users and whether they live in a city where AT is operating or not. The mean value of the rating scores is significantly different between AT cities and NT cities for half of the statements (IN1, IN3, HM1,HM2 and T4). We also note that the average scores of and hedonic motivation statements in AT cities are higher than those in NT cities, indicating that living in cities where ATs are operating makes probably respondents more prone to perceive enjoyable using ATs.

Constructs	Items	Group A AT users	Group B NT users in AT cities	Group A&B AT cities	Group C NT cities	T-test	χ^2 test A&B-C （p-value）
		Mean	Mean	Mean	Mean		
Injunctive norm	IN1	5.71	5.05	5.12	4.59	-3.974	22.206 (0.001)
	IN2	3.83	4.22	4.18	3.90	-1.753	11.657 (0.070)
	IN3	5.37	4.88	4.93	4.33	-4.563	33.498 (0.000)
Hedonic motivation	HM1	5.51	5.38	5.40	4.98	-3.514	31.430 (0.000)
	HM2	5.46	5.22	5.25	4.95	-2.522	25.009 (0.000)
	HM3	3.77	4.68	4.59	4.38	-1.366	15.009 (0.020)
Trust	T1	4.09	4.70	4.63	4.46	-1.103	14.289 (0.027)
	T2	3.74	4.42	4.35	4.36	0.104	17.744 (0.007)
	T3	4.14	4.69	4.63	4.75	0.723	22.394 (0.001)
	T4	5.29	5.05	5.08	4.43	-4.765	39.065 (0.000)

Table 4-25 Psychological statements – Survey online China
The direction of reversed statements IN2, HM3, T1, T2 and T3 were changed for comparison

4.4.5 Stated Choices

Table 4-26 presents the choice share between AT and NT, distinguished by AT and non-AT users and whether they live in a city where AT is operating or not. In this case, since each participant provided six choices, the sample is sufficient to perform a direct comparison (using the chi squared test) among all three groups. The samples analysed in this section consist of 210 pseudo-observations in group A (AT users), 1794 pseudo-observations in group B (respondents living in cities with AT) and 906 in group C (respondents living in cities without AT). It is interesting to note that no matter the trip distance travelled respondents from AT cities choose more ATs, while respondents from NT cities choose more NTs. Not surprisingly, this phenomenon is even more marked for AT users.

Finally, **Table 4-27** presents the stated choices classified by frequency of taxi usage. Interestingly, in AT cities, those who use taxi frequently, at least once a week, selected much more frequently ATs than NTs. This does not occur in NT cities. This effect will be explored in **Section 6** when results of the demand models will be discussed.

	Group A AT users N(%)		Group B AT cities N(%)		Group C NT cities N(%)		χ^2 test A-B (p-value)	χ^2 test B-C (p-value)
	NTs	ATs	NTs	ATs	NTs	ATs		
Short	16 (7.6)	26 (12.4)	147 (8.2)	159 (8.9)	216 (23.8)	186 (20.5)	80.743 (0.000)	283.270 (0.000)
Medium	27 (12.9)	93 (44.3)	324 (18.1)	348 (19.4)	180 (19.9)	126 (13.9)		
Long	20 (9.5)	28 (13.3)	341 (19.0)	475 (26.5)	110 (12.1)	88 (9.7)		
All	63 (30.0)	147 (70.0)	812 (45.3)	982 (54.7)	506 (55.8)	400 (44.2)	17.802 (0.000)	27.009 (0.000)

Table 4-26 Stated Choices – Survey online China

		Group A+B AT cities (AT & NT users) N(%)		Group C NT cities N(%)		χ^2 test (p-value)
		NTs	ATs	NTs	ATs	
Frequency of using taxis	At least once a week	385 (19.2)	707 (35.3)	116 (12.8)	106 (11.7)	243.303 (0.000)
	Less than once a week	490 (24.5)	422 (21.1)	390 (43.0)	294 (32.5)	

Table 4-27 Stated Choices classified by frequency of taxi usage – Survey online China

4.5 Summary

The objective of this chapter was to report the descriptive analyses of the samples that will be used for the modelling in **Chapter 6**. This chapter discussed the distribution of the sample characteristics and compared samples with the respective populations (though representativeness was not sought) and between them. The sample characteristics analysed include socio-demographics, familiarity with AV and ATs, choices made in the SC experiments, psychological statements and information about a current trip by taxi. In analysing these characteristics, the following comparisons were carried out: cross-national comparison between UK and China (*survey-online UK* and *survey-online China* were used), cross-methodological comparison between online and immersive VR surveys (*survey-online NCL* and *survey-VR NCL* were used), as well as, cross-experience comparison between AT existent cities and AT non-existent cities.

The descriptive analyses show that UK and China samples are different with respect to most of the characteristics. The UK sample has a lower proportion than the Chinese sample of male, younger, high educated, full-time employees or retired and low income people. Chinese respondents use taxi more frequently and for long trip and the majority of full time workers in China took a taxi for business or commuting purpose. Both UK and Chinese participants have heard of AVs, but Chinese participants have more knowledge of the levels of automation than the UK participants do. Interestingly, however, UK and Chinese respondents are very similar in terms of income proportion spent travelling by taxi, with 95% of each samples spending less than 8% of their disposable income. As expected, due to cultural differences, the samples differ also in the average scores of most of the psychological statements. Finally, in the UK sample respondents chose more NTs than ATs. The opposite in China, though the proportion varies depending on the length of the trip and the socio-economic characteristics.

The samples (online and VR) collected among those living in Newcastle also differ significantly in terms of socio-demographic characteristics, with a higher proportion of male, higher education level and students in the VR experiment participants, due to the recruitment process. But, income distribution is similar between the two samples as well as the trip characteristics (except destination selections) and the knowledge levels of AVs and ATs. Interestingly, ATs are chosen slightly less in the *online-based* survey than in the *VR-based* survey, in particular for short trips. Finally, it is worth mentioning that most respondents were not familiar with the VR settings before participating to the experiment, but the majority of respondents could easily complete the VR SC task and felt the VR experiments provided high level of realism.

Among the Chinese participants, the descriptive analysis showed that in the cities where AT is operating, the majority of the participants are young, highly educated, use taxis frequently and much more knowledge of AVs and ATs that those living in NT cities. Among those who tried an AT, 2/3 are female, against half of those who never tried an AT even though they live in a city where AT is operating and 1/3 of those who never tried an AT and living in a city where AT is not operating. Interestingly, respondents living in cities where ATs are operating are more prone to perceive enjoyable using ATs, and not surprisingly, chose ATs more frequently than respondents living in cities where ATs are not operating.

Chapter 5 Modelling Methodology

5.1 Introduction

This chapter outlines the modelling approaches used to analyse and quantify the impact of the key determinants in the adoption of ATs. **Section 5.2** reports a description of the theoretical foundation of the Discrete Choice Models (DCMs). **Section 5.3** describes the Mixed Logit (ML) model and **Section 5.4** the Hybrid Choice model (HCM) to account for latent psychological constructs. **Section 5.5** presents the formulation of joint models estimated with different datasets to account for scale heterogeneity. **Section 5.6** presents the exact specifications used in the estimated models (**Section 5.6.1 and Section 5.6.2**), the results of which are discussed in **Chapter 6**, the estimation process (**Section 5.6.3**), the tools used to evaluate the models (**Section 5.6.4**) and the process to compute marginal utilities and willingness to pay (**Section 5.6.5**).

5.2 Theoretical foundation of DCM

5.2.1 Random utility-based theory

The models that transport researchers currently use to describe how people choose among a discrete set of alternatives are based on the traditional economic assumptions of 'perfect rationality'. This implies preference rationality (i.e. consumers preferences are assumed to be primitive, consistent, and immutable), perception-rationality (i.e. consumers are assumed to behave as if they possess the formal tools to calculate the optimum) and process-rationality (i.e. the cognitive process is assumed to be simply based on preference maximization).

According to this theory, given a vector **A** of alternatives available in a given context, and a vector **X** of measurable attributes describing the characteristics of the individuals and that of the alternatives, individuals:

- are endowed with a particular set of attributes $\mathbf{X}' \in \mathbf{X}$.
- face a choice set $\mathbf{A}(q) \in \mathbf{A}$ and know all alternatives available in their choice set $\mathbf{A}(q)$.
- evaluate each alternative $j \in \mathbf{A}(q)$ based on its characteristics $\mathbf{X}'$.
- associate to each alternative a level of satisfaction, that is measured using an index defined 'utility'
- compare the alternatives based on the level of satisfaction perceived and always choose the alternative most attractive (i.e., the one that has the highest utility) subject to environmental constraints.

The formulation of the consumer choice models is based on the economic theory, but also recognises that modellers lack complete information about all the elements (**X**') considered by the individual in making a choice and that individual behaviours deviate from 'perfect rationality' (Tversky, 1972). The existence of '*features of taste template that were heterogeneous across individuals and unknown to the analyst, as well as unobserved aspects of experience and of information on the attributes of alternatives*' are interpreted as random factors (McFadden, 2000). This led to the formulation of the DCM based on the random utility theory, as reported in textbooks (Ben-Akiva and Lerman, 1985; Ortúzar and Willumsen, 2011; Train, 2009) and commonly used in research.

5.2.2 Operationalisation of the random utility-based theory

The theory described in the previous section is operationalised as follows. Let U_{jqt} be the utility that an individual q in a population Q ($q \in Q$) derives from each alternative j, within a set $\boldsymbol{A(q)}$ of alternatives available to the individual ($j \in \boldsymbol{A(q)}$, $A(q) = \{j_1, \ldots, j_A\}$) in the period (or scenario) t, with $t \in \boldsymbol{T}$, $T = \{t_1, \ldots, t_T\}$. The choice of alternative j by the individual q will be observed if the utility associated to the alternative j is higher that the utility associated to any other alternative i available to the individual:

$$U_{jqt}(X'_{jqt}) \geq U_{iqt}(X'_{iqt}) \quad \forall (i, j) \in A(q), i \neq j \qquad (5.1)$$

According to the random utility approach the utility is a random variable including measurable and known characteristics and at least one random component that captures everything that deviates from perfect rationality and/or all the relevant aspects of the phenomenon not explicitly known by the modeller. Therefore, the modeller is only able to observe a subset ($\mathbf{X} \subset \mathbf{X}'$) of the real vector of characteristics (Manski, 1977; Williams, 1977), while everything else is captured by a random component (ε'_{jqt}). A key assumption for the derivation of the DCM is that the random utility can be defined as the sum of the observable part (V_{jqt}), which is a function of known attributes $\mathbf{X}_{jqt}$, and the random unobservable part (ε_{jqt}):

$$U_{jqt} = V(\mathbf{X}_{jqt}) + \varepsilon'_{jqt} \qquad (5.2)$$

Since the utility is a random variable, it is not possible to know with certainty which alternative has the highest utility. It is only possible to compute the probability that a given alternative will be chosen:

$$P_{jqt} = \text{Prob}\left\{V(\mathbf{X}_{jqt}) + \varepsilon'_{jqt} \geq V(\mathbf{X}_{iqt}) + \varepsilon'_{iqt}, \forall \text{ j} \in A(q), \text{j} \neq \text{i}\right\} \tag{5.3}$$

Denote $f(\varepsilon') = f(\varepsilon'_1, ...\varepsilon'_N)$ the density function of the error term, the probability is then defined as:

$$\mathbf{P}_{jqt} = \int_{R_N} f(\varepsilon')d\varepsilon' \tag{5.4}$$

Where,

$$\text{R}_N = \begin{cases} \varepsilon'_{iqt} \leq (V_{jqt} - V_{iqt}) + \varepsilon'_{jqt} & \forall i \neq j, (i,j) \in A(q) \quad \forall t \in T \\ V_{jqt} + \varepsilon'_{jqt} \geq 0 \end{cases}$$

DCM can assume various forms depending on the assumptions on the distribution of the error terms. The simplest DCM, the Multinomial Logit (MNL) model is obtained assuming the error terms are independently and identically distributed (iid) Extreme Value type 1 (EV1). The MNL will be discussed in the next section, as part of the of the Mixed Logit model.

5.3 Mixed Logit model

The Mixed Logit (ML) model is one of the most powerful models currently available, because as demonstrated by McFadden and Train (2000) it can approximate any random utility model. The ML is characterised by an error term with at least two components: one distributed EV1 that gives the logit probability and a second one that can have any distribution.

Let write the random term in equation (5.2) as the sum of two components $\varepsilon'_{jqt} = \varepsilon_{jqt} + \mu_{jqt} z_{jqt}$, where $\varepsilon_{jqt}$is the component distributed iid EV1, $\mu_{jqt}$is the component (one or a vector) that can have any distribution with parameters (0, Ω) and Z_{jqt} is an attribute (or a vector of attributes) that might be known (i.e. the same as X_{jqt}) or unknown (and thus set equal to one). With this, the utility can be rewritten as:

$$U_{jqt} = V_{jqt}(X_{jqt}) + \underbrace{\mu_{jqt} z_{jqt}}_{\eta_{jqt}} + \varepsilon_{jqt} \tag{5.5}$$

Using equation (5.4), the mixed logit probabilities are the integrals over the density functions *f(ε)* and *f(η)*. Since ε and η are independent, the integral over *f(ε)* gives the typical MNL and the ML model is the integral of the MNL probability over density functions $f(\eta)$:

$$P_{jqt} = \int_{\eta} L_{jqt}(\eta) d\eta \tag{5.6}$$

Where L_{jqt} is the MNL probability conditional on the realisation of η:

$$L_{jqt}(\eta) = \frac{e^{V_{jqt}+\eta_{jqt}}}{\sum_{i \in A(q)} e^{V_{iqt}+\eta_{iqt}}} \tag{5.7}$$

The vector η of unobserved components can be decomposed to better capture the different aspects of individual random heterogeneity. The most typical components are the random parameter and the error components:

$$U_{jqt} = V_{jqt} + \underbrace{\sum_{t} \mu_{jqt} X_{jqt} + \sum_{m} \mu_{qm} z_{jm}}_{\eta_{jqt}} + \varepsilon_{jqt} \tag{5.8}$$

where μ_{jqt} are individual parameters, randomly distributed with zero mean, and X_{jqt} is a vector of level-of-service attributes. This product allows to account for random heterogeneity around the mean. μ_{qm} are also individual parameters, randomly distributed with zero mean and fixed over periods/scenarios while z_{jm} is an index that equals one if m appears in utility function j, and zero otherwise. This product allows to account for intra-individual correlation, i.e. correlation among different parameters in the same 'state', and correlation among alternatives.

The Random Parameters (RP) component of the ML involves the error term sharing the vector of attributes with the systematic component of utility, while the "pure" Error Components (EC) are completely unknown terms that account for response heterogeneities specific to each alternative or group of alternatives (Train, 2009).

In case of EC for intra-individual correlation, the ML probability is the probability that individual q will make a sequence of choices $j_t = \{j_1, \ldots, j_T\}$ and takes the following form:

$$P_{jqt} = \int_{\eta} \prod_{t=1}^{T} L_{jqt}(\eta) d\eta \tag{5.9}$$

An important issue in the ML model is the choice of the distribution to reproduce the heterogeneity underlying population preferences. The Normal distribution is the most used distribution but allows parameters to be either positive or negative, so it is difficult to understand whether the proportion of the population reproduced with a wrong sign is due to the data were wrongly coded or respondents answered untruthfully. Bounded distributions have been proposed, but they all have some problems: the lognormal for example allows avoiding wrong signs but has a long right-hand tail, returning high probabilities of yielding large portions of cumulative mass close to zero. The choice of distribution has an essential impact on the RP version of ML models, while it is not relevant in the EC version used in this PhD research.

A problem that is instead critical in the EC version of the ML is the theoretical identification. Walker (2000) reports an analysis of the three conditions that an EC ML model needs to satisfy to be theoretically identified. These are 'order conditions' (give the maximum number of coefficients that can be estimated), 'rank conditions' (give the actual number of coefficients that can be estimated), 'positive definiteness' (allows to verify if the chosen normalisation is valid). These conditions do not need to be applied in case the model includes only two alternatives. In this case the identification allows to include only one random term in one of the two alternatives.

5.4 Hybrid Choice model

The traditional discrete choice model has focused on the analysis of observable variables such as alternative-related attributes or socio-economic characteristics. Nonetheless, the impacts of subjective factors on choice process or disaggregate behaviours play a non-negligible role and have also been paid more attentions over two decades (e.g. Anable, 2005). The HCMs, allowing for the inclusion of latent behavioural or psychological constructs and overcoming the drawbacks of traditional discrete choice models, were first proposed by McFadden (1986) and Train *et al.* (1987) and were then popularised by the works of Walker (2001) and Ben-Akiva *et al.* (2002). HCMs has been extensively applied and adopted in transport context for capturing the impacts of psychological factors on several choice behaviours.

Figure 5.1 reports the framework of the HCMs (also called Integrated Discrete Choice and Latent Variable (ICLV)) models, integrating two components: discrete choice models (capable to account for latent variables as independent variables) and latent variables models. As shown in **Figure 5.1**, this specification can measure both direct and indirect effects of observed variables and latent variables on choice utilities, where the latent variables can be regarded as the explanatory variables in the utility specification of choice alternatives.

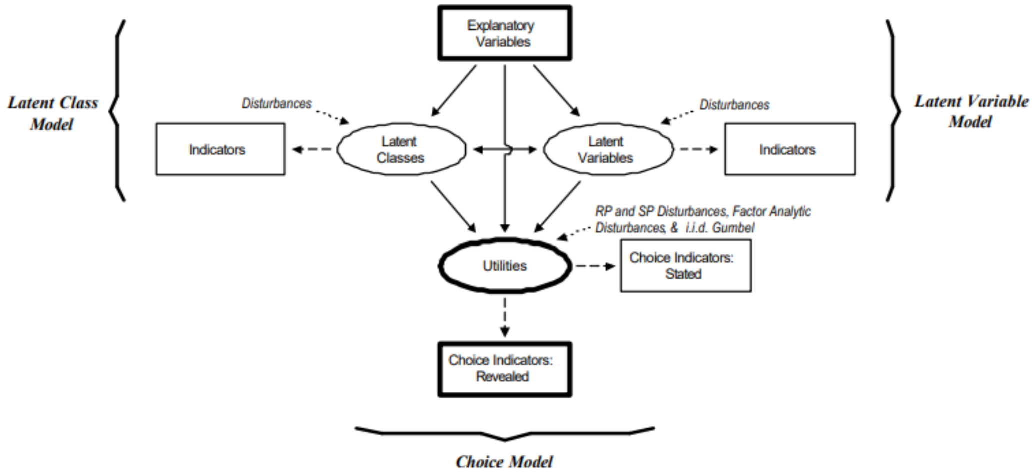


Figure 5.1 Framework of hybrid choice model (Walker, 2001; Ben-Akiva *et al.*, 2002)

The discrete choice part of the HCM is a typical DCM, which can be any form. But typically it is a ML. The latent variable model incorporates instead a set of structural equations (where the latent variables are explained by the individual or alternative characteristics) and a set of measurement equations (that explain the perception indicators) (Ortúzar and Willumsen, 2011).

The structural equation model is typically a linear regression:

$$LV_q = \alpha + \lambda X_q + \omega_q \tag{5.10}$$

Where, LV_q is the latent variable for individual q; α is the constant, X_q is a vector of observable variables (e.g. individual characteristics or other characteristics), λ is the associated coefficients and ω_q is a random term typically distributed Normal (0, σ_ω).

The measurement model takes typically the following expression:

$$I_{rq} = \delta_r + \theta_r LV_q + \nu_{rq} \qquad r = 1, \ldots, R \tag{5.11}$$

Where, I_{rq} is the r indicator for the latent variable, δ_r is a constant, and θ_r is the coefficient associated to the latent variable. ν_{rq} is the random terms distributed typically with Normal (0, σ_υ). But the indicators can be either discrete or continuous (Walker, 2001). Due to the identification issues, the δ_1 and θ_1 are generally normalised as 0 and 1 respectively.

The HCM probability is the integral of the DCM probability over the density function of ω. If the DCM is a ML, then it takes the form:

$$P_{jqt} = \int_{\omega}\int_{\eta} L_{jqt}(\eta,\omega)\, f_{LV_q}(\omega) \prod_{r} f_{Iq}(\mathrm{I}_{rq})\, \mathrm{d}(\eta)\, \mathrm{d}(\omega) \tag{5.12}$$

Where,

$$f_{LV_q}(\omega_q) = \frac{1}{\sigma_{\omega_q}} \phi\left(\frac{LV_q - (\alpha + \lambda X_q)}{\sigma_{\omega_q}} \right) \tag{5.13}$$

$$f_I(\mathrm{I}_{rq}) = \frac{1}{\sigma_{\nu_{rq}}} \phi\left(\frac{I_{rq} - (\delta_r + \theta_r LV_q)}{\sigma_{\nu_{rq}}} \right) \qquad r = 1,\ldots,R \tag{5.14}$$

5.5 Joint estimation

The joint estimation is a method used to enrich the estimation process. It is often difficult to obtained rich enough datasets containing all the information necessary for the study at hand. The joint estimation allows overcoming this problem, by pooling together different datasets. The joint estimation can be performed with any number of datasets, as long as their joint estimation is theoretically meaningful, and technically there is at least one attribute whose marginal utility is common among datasets. When a subset of the attributes included in the utility specification is common among the datasets, we talk of ‘partial enrichment’. When the entire set of attributes is common, we talk of ‘full enrichment’. Partial enrichment is much more flexible and nowadays the most used approach.

It is reasonable to believe that different segments of population or sources of data might have the same preferences for some attributes or alternatives but they do often have different scale. Whatever approach is used (full or partial), the scale of the data needs to be explicitly estimated. Let dataset A and dataset B be two datasets that we use to estimate jointly in a ML. The utility specification take the form:

$$U^A_{jqt} = \mu^A (V^A_{jqt} + \eta^A_{jq} + \varepsilon^A_{jqt})$$
$$U^B_{jqt} = \mu^B (V^B_{jqt} + \eta^B_{jq} + \varepsilon^B_{jqt}) \tag{5.15}$$

Where,

$$\varepsilon^k_{jqt} \sim EV1(0, \sigma^2_{\varepsilon^k}) \quad k = A, B \text{ , } \eta^k_{jq} \sim N(0, \sigma^2_{\eta^k}) \quad k = A, B$$

μ^A and μ^B are the scale parameters of the datasets, which are proportional to the inverse of the variance of the data. In the most general case, these are different and need to be estimated. Both scale parameters cannot be estimated. But as long as at least one coefficient is generic between the dataset, the relative variance of $\theta = \frac{\mu^A}{\mu^B}$ can be simultaneously estimated. This is equivalent to normalise one scale and estimate the other one. If $\theta = \frac{\mu^A}{\mu^B}$ is not significantly different from 1, the scales between two datasets can be considered as the same.

5.6 Building DCMs

5.6.1 Utility specification in the Mixed Logit models

The utility associated to each alternative can have any form and can include any type of attributes in any form. However, the most typical specification assumes linearity in the parameters to ease the estimation process, while there are no limits in terms of attributes. The utility specification used in this PhD dissertation to study the characteristics that affect the choice between ATs and NTs has the following general expression.

$$U^k_{jqt} = \lambda^k \left(ASC_j + \beta X + \beta^{SE}_j SE_q + \theta_j SE^{'}_q X^{'} + \eta_{jq} + \varepsilon_{jqt} \right) \tag{5.16}$$

With:

$$X = \left\{ LOS_{jqt}, IV_{jqt}, SC_{jqt}, Knw_q, TrCh_q \right\}$$
$$\beta = \left\{ \beta^{LOS}_j, \beta^{IV}_j, \beta^{SC}_j, \beta^{Knw}_j, \beta^{TrCh}_j \right\} \tag{5.17}$$

and

$$\beta X = \beta^{LOS}_j LOS_{jqt} + \beta^{IV}_j IV_{jqt} + \beta^{SC}_j SC_{jqt} + \beta^{Knw}_j Knw_q + \beta^{TrCh}_j TrCh_q \tag{5.18}$$

Where:

U_{jqt} is the utility that individual q assigns to alternative $j = \{normal\ taxi,\ automated\ taxi\}$ in choice task $t = \{1,\ 2,\ \ldots,\ 6\}$;

LOS is a vector including the level of services attributes (travel cost, travel time and waiting time);

IV is a vector including the in-vehicle features (change the destination, and chat with an operator);

SC is a vector including the normative conformity attributes (descriptive norms and customer reviews);

SE is a vector of socio-economic characteristics (e.g. age, gender, etc.);

Knw is a vector including the attributes related to the knowledge levels on AVs and ATs (e.g. familiarity with 5 levels of automation, etc.);

$TrCh$ is a vector including travel characteristics (e.g. frequency of taxi usage, etc.);

β is a vector of coefficients associated to all these characteristics, except the SE characteristics whose coefficient is β_j^{SE};

θ_j is the coefficient of the interactions between SE' and the vector X', that accounts for systematic heterogeneity in the X characteristics;

η_{jq} is the error term distributed Normal (0, σ_η), accounting for the correlations among choice tasks for same individual and , ε_{jqt} is the error term iid EV1;

ASC_j is the alternative specific constant for alternative j, takes value of 1 if j= AT, 0 otherwise;

$\lambda^k$is the scale parameter for the kth dataset. The joint estimation was performed with maximum two datasets, so k=1,2, and one scale was normalise ($\lambda^1 = 1$) and the other estimated.

5.6.2 Utility specification Hybrid Choice Models

The utility specification of the hybrid choice model is the same as for the ML models discussed in the previous section, with the addition of the latent variables. The utility takes the expression:

$$U_{jqt}^k = \lambda^k \left(ASC_j + \beta X + \beta_j^{SE} SE_q + \theta_j SE_q^{'} X^{'} + \beta_j^{LV} LV_q + \eta_{jq} + \varepsilon_{jqt} \right) \qquad (5.19)$$

Where, all coefficients and variables have the same definition described in the previous section and LV_q is a vector of latent variables that includes injunctive norm, hedonic motivation and trust, and β_j^{LV} is the vector of coefficients associated to the latent variable.

In addition, in the HCM, the structural equations of the latent variables (i.e. psychological constructs) is defined as:

$$LV_q^m = \alpha^m + \lambda^m Ind_q^m + \omega_q^m \qquad m = \{IN, HM, Trust\} \qquad (5.20)$$

Where, α^m is the constant for each LV; Ind_q^m is a vector of individual-related characteristics that can be different from the vector of SE, TrCh and Knw in equation (5.17), and λ^m is the vector of corresponding coefficients and ω_q^m is the normally distributed error term with mean zero and standard deviation of σ_ω^m.

The indicators measured in a 7-points Likert scale were assumed continuous.

5.6.3 Model estimation

In general, the maximum likelihood estimation is commonly performed for the DCM estimation, in which the logarithm of the product of probability is maximised by identifying estimated parameters (Ben-Akiva and Lerman, 1985). When the integral has a closed form (e.g. MNL or BL), the log-likelihood function given by the logarithm of the product of the probabilities P_{jqt} for individual q among sample size Q can be analytically calculated as:

$$LL = \sum_{q=1}^{Q} \ln(\mathrm{P}_{jqt}) \tag{5.21}$$

However, due to the assumption about the distribution of random components or flexibility of MLs, a closed form of probability function is not always suitable for MLs. The choice probability can be approximated through simulation, rather than exactly estimated (Train, 2009). Therefore, Maximum Simulated Likelihood (MSL) estimation as the most common simulation techniques was utilised for solving this problem (Hensher and Greene, 2003; Train, 2009). The Simulated Log-Likelihood (SLL) is derived as:

$$SLL = \frac{1}{R} \sum_{r=1}^{R} \prod_{t=1}^{T} L_{jqt}(\eta^r, \omega^r) f_{LV_q}(\omega^r) \prod_{r} f_{Iq}(\mathrm{I}_{rq}) \tag{5.22}$$

Where, R is the number of draws, (η^r, ω^r) are the parameter in the r^{th} draw for individual *q*. In order to get a good simulation of the true distribution, a large of number of draws is required. This relation between the number, the type of draws and the efficiency of the simulation has also been widely discussed (e.g. Bhat, 2003; Hess *et al.*, 2006).

The functional forms of HCMs are complex (e.g. complicated multidimensional integrals), making the estimating HCMs more complicated compared with MLs. Different maximisation methods can be applied (e.g. Method of Moments), but the Maximum Simulated Likelihood method is still commonly implemented due to its straightforward interpretation (Walker, 2001).

In the past the sequential estimation (i.e. a two-stage model: first estimating the latent variable model and then including the latent variables into the choice model estimation) was used. Nowadays, advances in computer performance, and availability of codes make the simultaneous estimation (it is a simultaneous estimation of traditional choice model together with latent variable models) the common approaches to estimate HCMs.

It is recommended to adopt simultaneous estimation as it theoretically solved the issues existing in sequential estimations. Although simultaneously estimating HCMs requires high computational power, which is the not issue at the current stage, the simultaneous simulated maximum likelihood approach was therefore employed in the following study (Ortúzar and Willumsen, 2011).

5.6.4 Model evaluation

The models estimated are evaluated using the typical statistical tests, and the microeconomic conditions. Among the statistical tests, the t-test is used to test if each single coefficient estimated is different from a value of reference, typically zero, and to compute the probability at which we can reject the assumption H0: β=0. Given S_{β_k} the standard deviation of the k^{th} estimated coefficient, the expression of the t-test is:

$$t = \frac{\beta_k^* - 0}{S_{\beta_k}} \tag{5.23}$$

The t-test is also used to test if there is a linear relationship between estimated parameters:

$$t = \frac{\beta_k^* - \beta_l^*}{\sqrt{S_k^2 + S_l^2 - 2S_{k,l}}} \tag{5.24}$$

Where, β_k^* and β_l^* are the two estimated coefficients, which are compared with each other; S_k^2 and S_l^2 are their respective standard deviations and $S_{k,j}$ is the their covariance. If the compared coefficients β_k^* and β_l^* are from two separate and independent models, then $S_{k,j}$=0.

The t-test for samples larger than 30 observations approximates a Normal distribution, and a test of hypothesis can be performed.

Another statistical test particularly useful is the *likelihood ratio test*. This allows to test if a model, which implies a set of linear restrictions (r) over a general function can be accepted. The expression is:

$$LR = -2\left\{l^*\left(\beta_{res}\right) - l^*\left(\beta\right)\right\} \sim \chi_r^2 \tag{5.26}$$

Where, $l^*(\beta_{res})$ is the restricted maximum values of log-likelihood function; $l^*(\beta)$ is the unrestricted maximum values of log-likelihood function. It is possible to test the null hypothesis that all coefficients except constant are equal to zero, in which $l^*(\beta_{res})$ is the log-likelihood of a model with only constant (market share model) and $l^*(\beta)$ the log-likelihood of the estimated model. The likelihood ratio test has a chi-squared distribution with degrees of freedom equal to the number of linear restrictions.

Finally, the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) are very useful tests in case the models are not restricted versions.

$$AIC = -2(LL(\beta) - K) \tag{5.27}$$

$$BIC = -2LL(\beta) + K \cdot \ln(Q) \tag{5.28}$$

Where, $LL(\beta)$ is the log-likelihood of estimated model, K is the number of estimated coefficients and Q is the number of observations. The lower the values of these criterion the better the model is. BIC is more restrictive because it penalises the model accounting also for the sample dimension.

5.6.5 Marginal Utilities and Willingness to pay

The microeconomic conditions establish the direction of the impact of the marginal utility of the various characteristics. Namely, the impact has to be positive or negative. If the utility is linear in the attributes, the estimated coefficients represent the marginal utility of the attributes associated to those coefficients. If the utility is not linear, the marginal utility is the partial derivative of the utility with respect to the attribute.

As an example, in the case of systematic heterogeneity between a LOS and a SE (as in Equation (5.16)), the marginal utility with respect to a characteristics *x* included in the vectors X and X', will be:

$$MU(X^{k}_{jqt,1}) = \frac{\partial U^{k}_{jqt}}{\partial X^{k}_{jqt,1}} = \lambda^{k}\left(\beta_{X^{k}_{jqt,1}} + \theta_{j} SE_{q}\right) \tag{5.29}$$

Since the marginal utility is a function of the estimated coefficients, it is a random variable that is normally distributed, like the coefficients estimated. As such, it is possible to compute the confidence interval and use hypothesis testing.

The willingness to pay is the amount of money individuals are willing to give up to save one unit in one characteristic. It is computed as the ratio between the marginal utility of the characteristics and the marginal utility of the cost at constant utility:

$$WTP(X^{k}_{jqt,1}) = \left. \frac{\partial U^{k}_{jqt}}{\partial X^{k}_{jqt,1}} \middle/ \frac{\partial U^{k}_{jqt}}{\partial C^{k}_{jqt,1}} \right|_{V} \tag{5.30}$$

Since the WTP is the ratio of marginal utilities, it is a function of the coefficients estimated and it is a random variable. However, the ratio of normally distributed random variables does not have a known distribution. The simulation then needs to be used to compute the t-test and confidence interval.

Monte Carlo simulation can be employed to calculate WTPs. Following Kolarova and Cherchi (2021), mean values and standard deviation for the WTP can be computed as below:

$$E\left(WTP_{\mathrm{ig}}\right) = \frac{1}{R}\sum_{r=1}^{R}\sum_{g=1}^{G}\left(WTP_{\mathrm{ig}}(r)\right) \tag{5.31}$$

$$S.D.\left(WTP_{\mathrm{ig}}\right) = \sqrt{\frac{1}{R-1}\sum_{r=1}^{R}\sum_{g=1}^{G}\left(WTP_{\mathrm{ig}}(r) - E\left(WTP_{\mathrm{ig}}\right)\right)^{2}} \tag{5.32}$$

Where, WTP_{ig} is the willingness to pay for attribute i for user group g and draw r. The t-tests of WTP_{ig} were then computed as the ratio between the mean and the standard deviation and the 95% confidence intervals were then computed.

5.7 Summary

This chapter presented the modelling methodology used in the PhD research. The core of the modelling methodology is based on the discrete choice models (DCMs). This chapter started with the theoretical foundations of discrete choice models, then described the utility specification and model estimation of mixed logit models, hybrid choice models, and in particular the DCMs jointly estimation with multiple datasets, which is the core methodology used in this dissertation. This chapter includes a discussion on the microeconomic derivation of the DCM and on the role of systematic heterogeneity in the individual's preferences for objective determinants (i.e. level of service characteristics, in-vehicle features and social conformity factors) and in the individual's preferences for specific alternatives. In the hybrid choice models, this heterogeneity can be indirect, through the impacts of the latent variables. Last not least, the chapter discusses the calculation of the WTPs confidence intervals in the case of the systematic heterogeneity, a key aspect that is often overlooked.

Chapter 6 Modelling Estimations & Discussions

6.1 Introduction

This chapter presents and discusses the results of the models estimated using the data collected in the four surveys described in **Chapter 3,** in order to understand and quantify what affects the choice of ATs versus NTs. This section is organised as follows:

Section 6.2 discusses the models estimated using the data collected with the *survey online-UK* and *survey online-China*¸ allowing for a cross-national comparison between UK and China on the preference for AT. In highlighting this comparison it discusses in detail the impact of the level-of-service attributes (**Section 6.2.1**), in-vehicle features (**Section 6.2.2**), normative conformity (**Section 6.2.3**), knowledge of AVs and ATs (**Section 6.2.4**), other effects (**Section 6.2.5**) and the impact of latent psychological constructs (**Section 6.2.6**), on the choice of ATs versus NTs. The section concludes with a discussion and a comparison of the willingness to pay (WTP) for these characteristics (**Section 6.2.7**).

Section 6.3 discusses the models estimated using the data collected with the *online-NCL* and *VR-NCL* allowing for a cross-methodology comparison between online and VR surveys on the preference for AT. As in the previous section, in highlighting this comparison it discusses in detail the impact of the level-of-service attributes (**Section 6.3.1**), in-vehicle features (**Section 6.3.2**), normative conformity (**Section 6.3.3**), and knowledge of AVs and ATs (**Section 6.3.4**), and other effects (**Section 6.3.5**) on the choice of ATs versus NTs. Given the relatively small size of the samples, in this case the impact of the latent psychological constructs is not estimated. WTP for all the characteristics analysed is discussed in **Section 6.3.6.**

Section 6.4 discusses the models estimated using the data collected with the *survey online-China* distinguishing between respondents living in cities where an AT service is operating (Guangzhou, Shanghai and Changsha) and respondents living in cities where an AT service is not operating (Chongqing, Guiyang, Wuhan etc.). This allows for a cross-experience comparison regarding the impact of AT existence on the preference for AT. It is an indirect experience, due only to living in a city where ATs are available to public. Using the best model estimated (as identified in **Section 6.2**) it discusses in detail the impact of AT existence on the preference for the level-of-service attributes (**Section 6.4.1**), in-vehicle features (**Section 6.4.2**), normative conformity (**Section 6.4.3**), knowledge of AVs and ATs (**Section 6.4.4**), and the impact of latent psychological constructs

(**Section 6.4.5**), on the choice of ATs versus NTs. The section concludes with a discussion and a comparison of the WTP for these characteristics (**Section 6.4.7**).

Section 6.5 discusses the results of the application of some policy scenarios using the models estimated. Four scenarios are tested and compared. A variations in the LOS attributes (**Section 6.5.1**), a variations in the in-vehicle features (**Section 6.5.2**), improvements in the customer reviews (**Section 6.5.3**) and improvements in the latent variables (**Section 6.5.4**).

Section 6.6 concludes this chapter with a summary of the main results.

6.2 Survey-online UK and the Survey-online China

Table 6-1 includes the results of mixed logit models with panel effects estimated using each single dataset separately. The first two models include only the main effects and are reported for comparison purposes, the last two models instead include also systematic heterogeneity and interaction effects. These latter models (ML2) only include the effects that were significant, however, all possible effects were tested (i.e. using all the socio-demographic data available in interaction with all the level of services variables, in-vehicle attributes and social conformity attributes). Particular attention was dedicated to those characteristics that allow accounting for differences between the two samples (UK and China) as discussed in **Section 4.2**. These are age, education level, employment status, etc. These were tested in the models to control for differences in the sample characteristics and their potential impact in the preferences heterogeneity between UK and China.

Before discussing each effect in detail, it is important to note that not all linear effects are significant. Looking at models ML1_UK and ML1_China in **Table 6-1**, we note that all level of service attributes tested are highly significant at more than 95% and all the marginal utilities (MUs) are negative as expected, in line with the microeconomic theory. Among the AT features and the social conformity variables, however, 'chat during the trip' in the UK dataset and the 'number of customers in the last hour' in both datasets are not significant at 95%. The most critical results however, is that the MU of the number of customers is negative, i.e. the effect is opposite to what is expected based on the psychological theory. In models ML2, which account for systematic heterogeneity, it was possible to identify for the Chinese sample, a specific category of respondents for whom the number of customers in the last hour has a correct positive effect. For the UK sample, however, the attribute was not significant for any category of respondents. This effect will be discussed in detail in the **Section 6.2.3**.

DISCRETE CHOICE MODEL	ML1_UK		ML1_China		ML2_UK		ML2_China	
	Value	Rob. t-test	Value	Rob. t-test	Value	Rob. t-test	Value	Rob. t-test
ASC (AT)	-1.34	-7.02	-0.209	-2.14	-0.118	-0.48	-0.672	-4.05
SIGMA (AT)	2.13	11.63	1.440	12.96	1.960	11.65	1.340	12.49
Level of services								
Travel cost [GBP][a]	-0.361	-9.65	-0.594	-8.67	-0.414	-10.60	-0.607	-8.34
*Long trip (30 minutes or more)					0.112	2.07	-0.266	-2.38
Travel time [minutes]	-0.116	-7.05	-0.036	-5.63	-0.126	-8.06	-0.035	-5.73
Waiting time [minutes]	-0.116	-6.46	-0.054	-6.07	-0.108	-5.53	-0.077	-9.48
* Age60_more					-	-	0.037	2.13
* Full-time employee					-0.050	-2.15	-	-
AT features								
Change the destination	0.240	2.99	0.334	5.35	0.283	3.51	0.200	3.27
Chat during the trip	0.314	1.91	0.530	5.59	0.330	2.09	0.555	5.56
Social Conformity								
Number of customers in the last hour/100	-0.063	-0.94	-0.035	-0.71	-	-	-	-
… * Heard of ATs from those who tested it					-	-	0.230	2.44
Good review yesterday	0.838	7.76	0.591	9.39	0.660	6.18	0.370	4.67
… * Age18_29					0.698	2.90	-	-
… * Long trip (30 minutes or more)					-	-	0.672	4.88
Systematic heterogeneity in AT alternative								
Age18_29					-	-	-0.485	-2.83
Frequently use taxis (at least once a week)					-	-	0.761	4.69
Like to carry luggage					-1.110	-3.94	-	-
Enjoy talking with driver					-1.880	-3.33	-	-
Heard of ATs operating in China or testing in the UK					-	-	0.654	4.05
Not familiar at all: 5 levels of automation					-0.983	-3.17	-	-
Summary of statistics								
Number of draws	500		500		500		500	
Log-Likelihood Market Share	-1212.160		-1770.972		-1212.160		-1770.972	
Maximum Log-likelihood	-959.248		-1610.846		-923.569		-1558.831	
Akaike Information Criterion (AIC)	1936.496		3239.692		1875.138		3147.662	
Bayesian Information Criterion (BIC)	1986.509		3292.802		1952.935		3236.177	
Number of individuals	319		450		319		450	
Number of observations	1914		2700		1914		2700	

Table 6-1 Systematic heterogeneity & Interactions effects- Survey-online UK & China

[a] In all models estimated with the data *online China*, the unit of travel cost was converted from CNY to GBP Conversion rate: 1 GBP = 8.43 CNY

Looking at the overall performance of the models, all models in **Table 6-1** are significantly different and superior to the market share model (based on the Likelihood Ratio test, the assumption that each model is equal to the market share model can be rejected at 99%).

To be able to compare the results from UK and China, scale differences need to be tested. Models estimated with the UK and the China data might have different scale parameters, i.e. differences in the error variances due to different context and maybe also different culture. To control for scale differences, a joint mixed logit model was estimated with both datasets, *survey online China* and *survey online UK*. The joint estimation was performed using the same utility specification used in model ML2_UK and ML2_China. The scale of the UK dataset was estimated, while the scale of the China dataset was set to one (no matter which scale is normalised, it has not an impact). The scale can be estimated only if at least one other coefficient is constrained to be the generic between datasets. Following Cherchi and Ortúzar (2006), we first computed the ratio of all equal coefficients estimated with each data separately to identify candidate coefficients to be constrained in the joint estimation. Since there was not a set of clear candidate coefficients to be constrained (i.e. with the same ratio), we tested all coefficients once at a time. The scale parameter in all these cases was never significantly different from one. Consistently the assumption that the scale is equal to one was always rejected only at less than 60%. **Table A5-A** in **Appendix 5** reports one of the UK-China joint estimation. As we can see the scale is not significantly different from 1 (H0 is rejected at 51% in a two-tailed test). China dataset and UK dataset have the same variance, it is then possible to compare directly the results reported in **Table 6-1**.

Table 6-2 reports the MUs of all the characteristics tested in ML2_UK and ML2_China for different categories of respondents. The table reports also the t-test between the MUs estimated in the UK and in China, and the sample size of each category. The t-test is computed assuming independence of the MU between UK and China. However, due to systematic heterogeneity, the standard deviation of the MU of each attribute for each specific category is computed accounting also for the correlation between the coefficients for different categories. In **Table 6-2** the MU are computed using equation (5.29) in **Chapter 5**, and the t-test for generic coefficients, using equation (5.24) still in **Chapter 5**.

The results from models ML2_UK and ML2_China in **Table 6-1** and their MU computed in **Table 6-2** are used in the next sections to discuss the specific effect of each characteristic on the choice of AT versus NT and the cross-countries difference.

	Marginal utilities			Sample size	
DISCRETE CHOICE MODEL	ML2_UK	ML2_ China	t-test	UK	China
Level of services					
Travel cost * Short trips (< 30 min) [GBP]	-0.414	-0.607	2.34	269 (84%)	281 (62%)
Travel cost * Long trip (>= 30 min) [GBP]	-0.302	-0.873	5.05	50 (16%)	169 (38%)
Travel time [minutes]	-0.126	-0.035	-5.42	319 (100%)	450 (100%)
Waiting time * Age < 60 * Full-time empl. [minutes]	-0.158	-0.077	-3.65	111 (35%)	277 (62%)
Waiting time * Age >= 60 * Full-time empl. [minutes]	-0.158	-0.040	-4.50	15 (5%)	18 (4%)
Waiting time * Age < 60 * other empl. status [minutes]	-0.108	-0.077	-1.47	118 (37%)	107 (24%)
Waiting time * Age >= 60 * other empl. status [minutes]	-0.108	-0.040	-2.68	75 (23%)	48 (11%)
AT features					
Change the destination	0.283	0.200	0.82	319 (100%)	450 (100%)
Chat during the trip	0.330	0.555	-1.20	319 (100%)	450 (100%)
Social Conformity					
Number of customers in the last hour/100 * Heard about ATs from those who used ATs	0.230		N.A.	23 (7%)	49 (11%)
Good review * Age <30 * short trip (< 30 min)	1.358	0.370	4.09	63 (20%)	98 (22%)
Good review * Age >=30 * short trip (< 30 min)	0.698	0.370	1.29	206 (65%)	183 (41%)
Good review * Age <30 * Long trip (>= 30 min)	1.358	1.042	1.25	11 (3%)	56 (12%)
Good review * Age >=30 * Long trip (>= 30 min)	0.698	1.042	-1.30	39 (12%)	113 (25%)
Systematic heterogeneity in AT alternative					
Age < 30	-0.118	-1.157	3.41	74 (23%)	154 (34%)
Frequently use taxis (at least once a week)	-0.118	0.089	-0.71	31 (10%)	193 (43%)
Like to carry luggage	-1.228	-0.672	-2.02	209 (65%)	291 (65%)
Enjoy talking with driver (always)	-1.998	-0.672	-2.17	36 (13%)	35 (8%)
Heard of ATs operating in China or testing in the UK	-0.118	-0.018	-0.34	160 (50%)	157 (35%)
Not familiar with 5 levels of automation at all	-1.101	-0.672	-1.12	102 (32%)	87 (19%)
All other categories	-0.118	-0.672	1.87	-	-

Table 6-2 Marginal utilities- Survey-online UK & China

6.2.1 Impact of level of services attributes

As mentioned before, all coefficients associated to the level of service attributes are significant at 99% in both UK and China and all MUs have the correct negative sign for all the categories in the samples. Results show that the MU of travel cost in both UK and China varies with the length of the trip. However, interestingly, why in the UK the MU of cost diminishes with the length of the trip, in China, the effect is the opposite. This effect could be linked to differences in income distribution between the two samples. As discussed in **Section 4.2.1, Chapter 4**, the percentage of monthly disposable income spent travelling by taxi is the same in the UK and China (less than 7%-8% for 95% of each sample). However, further analyses reported in **Section 4.2.5** show that in the UK most of the respondents in the lower income categories selected short trips, while in China it is less clear but the distribution is the opposite. This means that there can be a confounding effect between income and distance travelled. Following Jara-Díaz and Videla (1989), income effect[21] was tested including the squared of cost (that should be positive and significant to suggest presence of income effect) and also estimating different MU of cost for different income groups (MU decreasing as income increases would suggest income effect). None of these tests confirmed the presence of income effect, in both datasets.

As expected from the previous discussion, the MU of travel cost is significantly different between UK and China (t-test rejected at 99%) and Chinese respondents are much more sensitive to taxi fare than the UK respondents are. As discussed previously, this phenomenon does not seem to be due to income effect, but maybe to Chinese respondents using taxi more frequently and for longer trips (as discussed in **Section 4.2.5)**. This effect was tested in the model, but results were not stable, meaning that the interaction between frequency of the trip and cost was significant if this was the only interaction included, but the effect becomes non- significant when the interaction with the distance was included.

Waiting time is a very important variable in the choice of taxi. Results in **Table 6-2** show heterogeneity in the MU of waiting time as a function of age and type of occupation, and

---

[21] According to the economic theory (Hicks, 1956) the total impact of a change in the price is the sum of substitution effect (substitution between characteristics at constant utility) and income effect (variation in the utility due to a variation in the purchase power or real income). Typical utility specifications linear in the price only account for substitution effect. If income effect is present, a utility non-linear in the price, or better in income minor price needs to be specified.

heterogeneity between UK and China respondents. In the UK, full-time employees value a unit change in waiting time almost 50% more negatively than those with different occupational status. This is certainly due to the fact that full-time employees have fixed schedules and are more likely to be restricted on other activities outside work (as shown in **Table 4-6** only 13% of full time workers in the UK took a taxi for business or commuting purposes, against 40% in China). However, the impact seems to be related to the constraint imposed by being full time worker, not by the purpose of the trip, because the interaction between waiting time and the purpose of the trip was tested and it was not significant. Among the Chinese respondents, instead age is the most important factors. Chinese people who are older than 60 years care less about waiting time.

Overall, in the UK, all respondents have MU of waiting time more than twice as big as respondents in China, and of course the MU are significantly different at 99%, with the exception of the category of respondents younger than 60 years and not full-time employees. The MU of waiting time for this category is not significantly different between the UK and China.

Finally, no systematic heterogeneity was found for the MU of travel time, both in the UK and China. This result was not expected because in the literature, travel time is probably the level of service attribute with the strongest evidence of systematic heterogeneity. However, as discussed in **Section 2**, the few papers dealing with AT did not test heterogeneity in the preference for travel time, so a direct comparison is not possible. Results show however that the MU of travel time for the UK respondents is much higher than for the Chinese respondents (H0 rejected at 99%).

6.2.2 Impact of in-vehicle features

In-vehicle features represent two key variables in this research. Testing their impact in the choice of AT versus NT is one of the objectives of this PhD dissertation. The two in-vehicle features tested are both coded with dummy variables. The possibility to change destination and to chat with an operator takes value 1 if the feature is available in the AT, 0 otherwise. The MU of these features is expected to be positive. Having the possibility to change destination and to chat with an operator during the trip should increase the probability to choose an AT over a NT. As we can see, both features have a positive and highly significant effect in both datasets. The t-test allows rejecting the assumption that the coefficient are equal to zero at 97% for 'chat with an operator' in the UK and at more than 99% for all other in-vehicle features both in the UK and in China. The request to chat with an operator was mentioned in the FGs by some elderly participants and by those who said to enjoy talking and to frequently talk with a taxi driver. Based on this information, it was tested if the MU of the in-vehicle feature 'chat with an operator' was different for people who enjoy talking and those who frequently talk with the driver. However, none of these interaction effects was significant. Maybe, this is due to the fact that in both samples (UK and China) only few people reported to really enjoy talking with a driver and to frequently talk with a driver (see **Table 4-4, Chapter 4**). Other systematic heterogeneity effects were tested but none was found significant.

Looking at the difference between UK and China, interestingly, results in **Table 6-2** show that there is no significant difference in the MU of changing destination and chatting with an operator between UK and Chinese respondents. This result is interesting, though not expected, because for example chatting with the driver (or an operator in case of AT) is something that is likely to be related to cultural traits. In China the MU is a bit higher, but not significantly different from UK (according to the t-test, H0 is rejected at 77% in a two-tail test).

6.2.3 Impact of normative conformity

As discussed in the introduction of this chapter, the descriptive norm 'number of customers in the last hour' shows some problems. Its effect is not significant for the entire sample and it has also a counterintuitive negative sign in both samples. This is in line with the results from the studies on electric vehicles (as discussed in **Chapter 2**), where this attribute has always been problematic, when tested as an attribute within a SC experiment. Several utility specifications were then tested to identify if there was systematic heterogeneity in the preference for this descriptive norm among groups of respondents (i.e. socio-demographic characteristics, taxi trip characteristics and knowledge level of ATs or AVs). In the UK sample, none of these interaction effects was found significant and more importantly none of them had the expected positive sign (some of these tests are reported in **Table A5-2** in **Appendix 5**). It was then decided not to include this attribute in the UK specification. In the Chinese sample, on the other hand, among all the interactions tested, it was found that this descriptive norm was highly significant and with a positive effect for those who heard of ATs from those who have used it (accounting for 10.9% of total sample). This is a plausible result, as there is of course a link between the number of customers and those who used AT from whom the respondent heard about ATs.

Confirming the results from the marketing literature, good reviews measured by a high rating from yesterday's customers, have a significant positive impact in the choice of the type of taxi in both survey online in the UK and in China. Some systematic heterogeneity was found in the preference for good reviews, but interestingly, results are quite different between UK and China. As shown in **Table 6-2**, in the UK it seems that good reviews have a bigger impact than in China, with the exception of the group of respondents older than 30 years when the trip by taxi is longer than 30 minutes. In this case the impact among the Chinese respondents is much higher than among the British ones.

In China, those who use taxis for long trips are more sensitive to the impact of good reviews. This result makes sense because the longer the time spent within a taxi, the more respondents wish to be reassured about the overall quality of the service. Differently from the result of China survey, in the UK, the younger the customer, the greater the impact of good reviews. This seems to be in line with the literature on consumer reviews, that found the younger people tend to be more influenced by positive reviews and aggregate information, while older adults more affected by a single negative review (Von Helversen *et al.*, 2018).

6.2.4 Knowledge of AVs and ATs

As discussed extensively in the introduction and in the literature review (**Sectio**n **2.3.1**) knowledge of AVs or lack of it, is considered critical in estimating properly users' preferences for AVs. This is why, as discussed in **Section 3.8** information about AVs are included before all SC surveys, to allow respondents to get the knowledge they lack. In this dissertation, four types of information were collected to measure AV knowledge: heard of AV, familiarity with the 5 levels of automation, heard of AT operating/testing in the country, person from whom the respondent heard of AT. Interestingly, results show that knowledge of AV/AT does not affect the preference for the AT characteristics, but does affect significantly the specific preference for AT versus NT. As discussed in the next **Section 6.2.6**, knowledge of AV and AT have an indirect impact on the choice of AV through the latent psychological constructs.

We already discussed that respondents have an overall preference for NT over AT (the ASC-AT is negative and the systematic heterogeneity remains negative for all categories of users, as reported in **Table 6-2**). In addition, we note that complete lack of familiarity with the 5 levels of automation is a strong deterrent in the choice of AT, especially in the UK where the percentage of respondents in this category counts 32% of the sample, against 19% in China. On the other hand, having heard of AT operating in China has a strong positive impact in the choice of AT. This is another interesting result. Firstly, we note that what affects the choice of AT is not having generic knowledge of AV but having specific knowledge about AT operating in China. In terms of systematic heterogeneity in the preference for ATs, it only matters that the person has heard about ATs operating in China, it does not matter from who they heard about them. This phenomenon was not observed in the UK sample probably because in the UK, differently from China, there are no ATs operating, there are only tests with minibuses. It is not surprising that having an AT system operating in the country has a stronger impact than having generic tests.

6.2.5 Other effects

As reported in **Table 6-1** and **Table 6-2,** other systematic heterogeneities in the preference for ATs were found, and these effects are different in the UK and in China. In **Table 6-2** the MU for these effects are computed as the sum between the ASC for AT and the coefficient of each systematic heterogeneity estimated. The ASC_AT gives a measure of the preference for the AT alternative compared to the NT alternative, everything else being equal. This preference is average for the entire sample. The socio-demographic characteristics added to the AT alternative gives a measure of the preference for AT compared to NT by that specific segment of the sample. The net preference is then the sum of the ASC and the coefficients of the socio-demographics. Results show that surprisingly both in the UK and in China, there is a preference, everything else equal, for NT with respect to AT. This finding is also consistent with the findings in some AV-related literature (e.g. Krueger *et al.*, 2016; Haboucha *et al.*, 2017, etc.). This can also be due to inertia effect. Respondents normally have a tendency to select the existing travel mode. To note that the alternative specific constant (ASC) in the AT alternative is negative and significant in this dissertation as well as in this previous literature. While there is an exception of those in China who use frequently taxi (at least once a week). This result is less intuitive and might be due to the fact that they had not good experiences using NTs (e.g. unnecessary detours is a common problem in China) which increases the willingness to change to AT. At the same time, this result could also be due to curiosity about the AT services. AT services are still not common in China, and none of the respondents in the sample used to estimate these models has the experience with ATs. In addition, young respondents below 30 years old are less likely to choose ATs. This result is inconsistent with other findings in the literature (e.g. Haboucha *et al.*, 2017), as young people are generally more likely to embrace new products compared with older counterpart. A possible explanation could be that this variable indirectly affects the choice of ATs through the latent psychological constructs and will be discussed in the **Section 6.2.6**).

In line with expectation, in the UK, respondents who like the taxi driver to help with their luggage or heavy bags and who enjoy talking with the taxi driver have an even lower (compared to the mean value represented by the ASC) preference for AT versus NT, everything else equal. It is plausible due to the fact that those respondents probably get accustomed to communicating and interacting with a taxi driver when using NTs and probably dislike the fully ATs without any human interactions.

Finally, it is worth noting that the ASC for AT in the UK model is not significantly different from zero, indicating that the variables specified reproduce well the difference between AT and NT.

6.2.6 Latent psychological constructs

Hybrid choice models are estimated to study the impact of the three latent psychological constructs, as discussed in **Chapter 2.** These are injunctive norm, hedonic motivation and trust, in addition to the above objective effects. The results of the best hybrid choice models are reported in **Table 6-3.**

First, comparing the results of the discrete choice model part of the HCM (in **Table 6-3**) with the results of the ML2 models (in **Table 6-1**), we note that all the coefficients, except those measuring systematic heterogeneity in AT, have the same values. The ratio between the HCM and the ML coefficients is around 1.05, which is an indication of the scale difference between the two model specifications. Differences are instead found in the preferences specific for ATs. This is expected because the latent psychological constructs are summed to the AT utility, and several of the attributes that have a significant effect directly in the preference for AT are also significant in explaining the latent psychological construct. In this line, we note also that the ASC-AT is highly significant in the HCM. In the remaining of this section, the discussion will focus then only on the effect of the latent psychological constructs.

As expected, all three latent psychological constructs have a positive impact in the choice of AT versus NT and are also significant at 99%. Interestingly, the injunctive norm has a similar impact in the choice of ATs in both survey-online UK and survey-online China, while hedonic motivation and trust in the UK have an impact roughly twice and four times than in China, respectively.

First we note that a couple of effects that in the ML models seemed to have a direct and significant in the preference for AT, in the HCM it is clear that their impact is indeed indirect through the latent psychological constructs. Respondents younger than 30 years old in China prefer NT over AT, but the real effect is that they are less affected by injunctive norms and they seems also to trust less AT than respondents older than 30 years. Analogously, in the UK lack of familiarity with the 5 levels of automation has a negative impact on the preference for AT, but the true effect seems to be that familiarity increases the probability of being positively affected by injunctive norms, trust and also hedonic motives. These effects are not significant anymore when directly summed to the utility of AT, but are highly significant in the structural model of the latent psychological constructs.

Some other effects instead have both a direct impact in the utility of AT and an indirect impact through the latent psychological constructs. Interestingly this is the case for all the variables related to knowledge of AV and AT, as well as for the frequency of usage of taxi, which interestingly is an indirect measure of habit.

In particular, we found that in China being familiar with the 5 levels of automation, having heard of AT (by those who either have tried or not an AT) and using frequently taxi (at least once a week) is positively related with injunctive norms (more than 99%). These results suggest that the higher the level of familiarity with taxis and the level of knowledge of AV and AT the more people tend to comply with the approval of using AT from those who are important for them. It seems knowledge reinforces compliance.

In the Chinese sample, we found that young aged below 30 years are less likely to be affected by what other people think is right to do, probably because young people are more informed and more assertive regarding using innovative products and less prone to be influenced by others. An important point to note is that the variable age younger than 30 years old has a negative direct impact in the choice of ATs when included in the AT alternative in the ML. This is a counter-intuitive result, because young people are typically more likely to accept innovative modes However, in our case, the direct impact of younger than 30 years old in the ML2_China is spurious and it becomes not significant in the HCM where the attribute is included also in the injunctive norm (the direct effect was removed from the estimation in the HCM in **Table 6-3**). In our data, the correct impact of being younger than 30 years old is an indirect effect via the injunctive norm, which is revealed correctly in the HCM.

Familiarity with the 5 levels of automation and having heard of AT by those who have tried an AT has a positive impact also on the hedonic motivation both in the UK and in China. Maybe more knowledge is associated also with more positive opinion of AT. This can be explained partly by the fact that if people seek information there might be already a positive attitude toward AV, and the knowledge tends to support and reinforce the positive opinion. Partly by the fact that the majority of the information currently available are positive in the light of boosting the development and acceptance of the AV. Those who frequently use taxis also tend to have more positive hedonic motivation, though there is a difference in terms of frequency: in China is once a week, in the UK once a month. This difference might simply be due to the distribution in the samples (the majority of the sample in China use taxi once a week, while in the UK between once a month and twice a year. Finally, in the UK, differently from China, the variable young (age below 30 years old) was also found to have a significantly positive effect on hedonic motivation. This is intuitive because young people are curious and prone to try something new, and they probably associate using new AT with something fun to do.

Finally, trust is positively linked to familiarity with level of automation both in the UK and in China. This is expected and an intuitive result. This is, however, the only common effect between UK and China. In the UK, being male and frequent users of taxi, is associated with higher level of trust in AT. This is plausible because men are generally more open to innovative technologies. This is also in line with several results in the literature (Schoettle and Sivak, 2014; Kyriakidis *et al.*, 2015; Schoettle and Sivak, 2015; Sweet and Laidlaw, 2020). The fact that users who frequently use normal taxis have higher trust in AT could be explained by a higher level of confidence in taxis in general that is probably extended to any type of taxis, normal and automated.

In China instead, trust is related only to age. In particular, young people (less than 30 years old) trust AT less, which is a counter-intuitive result, as young people tend to be more prone to or less scared by technology and innovation. Hence, they are expected to trust more. However, trust in AT implies probably a more reasoned evaluation, based on information, while young people might be more prone to innovation in general not necessarily because they trust it.

	HCM_ UK		HCM _China	
DISCRETE CHOICE MODEL	Value	Rob. t-test	Value	Rob. t-test
ASC (AT)	-8.920	-7.98	-5.750	-10.76
SIGMA (AT)	0.955	7.25	0.913	8.04
Level of services				
Travel cost	-0.397	-10.66	-0.585	-8.32
*Long trip (30 minutes or more)	0.106	1.94	-0.245	-2.29
Travel time	-0.120	-8.08	-0.033	-5.60
Waiting time	-0.104	-5.62	-0.074	-9.44
… * Age60_more	-	-	0.036	2.17
… * Full-time employee	-0.045	-1.96	-	-
AT features				
Change the destination	0.274	3.52	0.196	3.32
Chat during the trip (AT)	0.335	2.18	0.520	5.39
Social Conformity				
Number of customers in the last hour/100	-	-	-	-
… * Heard about ATs from those who used ATs	-	-	0.212	2.23
Good review yesterday	0.663	6.26	0.361	4.71
… * Age18_29	0.581	2.61	-	-
… * Long trip (30 minutes or more)	-	-	0.644	4.83
Systematic heterogeneity in AT alternative				
Age18_29	-	-		
Frequently use taxis (at least once a week)	-	-	0.278	1.87
Like to carry luggage	-0.421	-1.90	-	-
Enjoy talking with driver	-1.430	-3.62	-	-
Heard of ATs operating in China or testing in the UK	-	-	0.313	2.04
Not familiar with 5 levels of automation at all	-	-	-	-
Latent psychological constructs				
IN: Injunctive norms (AT)	0.344	2.92	0.406	3.11
HM: Hedonic Motivation (AT)	0.727	4.58	0.449	3.37
T: Trust (AT)	0.887	2.70	0.200	2.31
Summary of statistics				
Number of draws	500		500	
Log-Likelihood Market Share	-9200.30		-13434.80	
Maximum Log-likelihood	-6291.34		-8905.78	
Akaike Information Criterion (AIC)	12690.68		17923.56	
Bayesian Information Criterion (BIC)	12990.75		18254.02	
Number of individuals	319		450	
Number of observations	1914		2700	

Table 6-3 Latent psychological constructs- Survey-online UK & China

LATENT VARIABLE MODEL	IN _UK		HM _UK		T _UK		IN _China		HM _China		T _China	
	Value	Rob. t-test	Value	Rob. t-test	Value	Rob. t-test	Value	Rob. t-test	Value	Rob. t-test	Value	Rob. t-test
Structural model												
Constant	4.070	42.45	3.710	21.97	3.610	23.57	4.610	36.26	4.860	56.62	4.74	41.13
Standard deviation of error term	0.235	3.78	0.417	7.12	-0.314	-1.55	0.057	0.78	0.043	0.69	0.352	7.25
Male	-	-	-	-	0.321	2.49	-	-	-	-	-	-
Age18_29	-	-	0.398	1.90	-	-	-0.542	-3.86			-0.478	-2.60
Frequent use of taxis (at least once a week)							0.464	3.18	0.447	4.14	-	-
Frequent use of taxis (at least once a month)	-	-	0.616	2.87	0.313	2.55	-	-	-	-	-	-
Familiar with 5 levels of automation	0.773	2.71	1.160	5.80	0.371	2.4	0.367	1.60	0.488	3.46	-	-
Heard of ATs testing or operating	-	-	-	-	-	-	-	-			-	-
Heard of ATs from those who used ATs	-	-	0.732	2.44	-	--	0.760	4.11	0.383	2.56	-	-
Heard of ATs from those who not used ATs	-	-	-	-	-	-	0.366	2.37	-	-	-	-
Measurement model												
Constant in indicator N2	1.300	4.01	0.583	2.88	-3.760	-2.27	1.630	4.41	0.191	0.50	0.021	0.07
Constant in indicator N3	-0.071	-0.24	2.620	8.35	-4.450	-2.51	-0.001	0.00	1.830	4.45	0.415	1.57
Constant in indicator N4					-4.740	-2.63					2.570	8.21
Coeff in indicator N2	0.732	9.83	0.886	19.3	1.950	4.65	0.507	6.34	0.941	13.71	0.948	14.64
Coeff in indicator N3	0.909	12.74	0.497	7.64	2.170	4.85	0.958	18.49	0.523	6.42	0.930	17.88
Coeff in indicator N4					2.200	4.82					0.493	8.05
Standard deviation indicator N1	-0.354	-3.44	-0.231	-2.37	0.569	13.65	-0.188	-2.75	-0.387	-5.29	-0.137	-1.73
Standard deviation indicator N2	0.204	3.33	-0.233	-2.62	-0.034	-0.41	0.424	12.78	-0.259	-3.93	-0.015	-0.20
Standard deviation indicator N3	-0.170	-2.08	0.344	8.19	-0.030	-0.35	-0.168	-3.07	0.375	10.06	0.005	0.08
Standard deviation indicator N4					0.019	0.26					0.244	6.70

Table 6-3 (continued) Latent psychological constructs- Survey-online UK & China

6.2.7 Willingness to pay

Tables 6-4 reports the mean value of the WTP estimated for all the attributes tested for both dataset and both the ML and the HCM. As expected, the WTP in the ML and HCM are almost identical. We will comment only on the WTP estimated from the ML model results. **Tables 6-5** also reports t-tests and confidence intervals only for the ML models, these are computed using Monte Carlo simulations with 5,000 draws from a multivariate truncated Normal distribution.

	ML		HCM	
	UK	China	UK	China
Short or medium trip (< 30 minutes)				
Level of services [GBP/hour]				
Travel time	18.26	3.46	18.14	3.43
Waiting time * Age < 60 * Full-time employee	22.86	7.57	22.50	7.60
Waiting time * Age >= 60 * Full-time employee	22.86	3.93	22.50	3.88
Waiting time * Age < 60 * other empl. status	15.65	7.57	15.72	7.60
Waiting time * Age >= 60 * other empl. status	15.65	3.93	15.72	3.88
In-vehicle features [GBP/unit]				
Change the destination	0.68	0.33	0.69	0.34
Chat during the trip (AT)	0.80	0.91	0.84	0.89
Social conformity [GBP/unit]				
Number of customers in the last hour/100 … * Heard about ATs from those who used ATs		0.38	-	0.36
Good review * Age <30	3.28	0.61	3.13	0.62
Good review * Age >=30	1.59	0.61	1.67	0.62
Long trip (>= 30 minutes)				
Level of services [GBP/hour]				
Travel time	25.03	2.41	24.74	2.41
Waiting time * Age < 60 * Full-time employee	31.33	5.26	30.70	5.36
Waiting time * Age >= 60 * Full-time employee	31.33	2.74	30.70	2.73
Waiting time * Age < 60 * other empl. status	21.46	5.26	21.44	5.36
Waiting time * Age >= 60 * other empl. status	21.46	2.74	21.44	2.73
In-vehicle features [GBP/unit]				
Change the destination	0.94	0.23	0.94	0.24
Chat during the trip (AT)	1.09	0.64	1.15	0.63
Social conformity [GBP/unit]				
Number of customers in the last hour/100 … * Heard about ATs from those who used ATs	-	0.26	-	0.26
Good review * Age <30	4.50	1.19	4.27	1.21
Good review * Age >=30	2.19	1.19	2.28	1.21

Table 6-4 WTP mean values- Survey-online UK & China

	ML model			
	UK		China	
	t-test	95% Confidence interval	t-test	95% Confidence interval
Short or medium trip (< 30 minutes)				
Level of services [GBP/hour]				
Travel time	6.53	[12.90,23.97]	4.67	[2.04, 4.98]
Waiting time * Age < 60 * Full-time employee	4.72	[13.46,32.58]	6.00	[5.21,10.26]
Waiting time * Age >= 60 * Full-time employee	4.72	[13.46,32.58]	2.05	[0.18,8.01]
Waiting time * Age < 60 * other empl. status	4.94	[9.50,22.01]	6.00	[5.21,10.26]
Waiting time * Age >= 60 * other empl. status	4.94	[9.50,22.01]	2.05	[0.18,8.01]
In-vehicle features [GBP/unit]				
Change the destination	3.37	[0.29,1.09]	2.96	[0.11,0.56]
Chat during the trip (AT)	2.05	[0.03,1.58]	4.48	[0.52,1.34]
Social conformity [GBP/unit]				
Number of customers in the last hour/100 * Heard about ATs from those who used ATs	-	-	1.81	[-0.03,0.77]
Good review * Age <30	4.65	[1.92,4.71]	3.98	[0.31,0.92]
Good review * Age >=30	5.36	[1.02,2.20]	3.98	[0.31,0.92]
Long trip (>= 30 minutes)				
Level of services [GBP/hour]				
Travel time	3.25	[10.50,42.28]	4.15	[1.30,3.63]
Waiting time * Age < 60 * Full-time employee	2.98	[11.35,54.96]	5.09	[3.33,7.51]
Waiting time * Age >= 60 * Full-time employee	2.98	[11.35,54.96]	2.00	[0.05,5.68]
Waiting time * Age < 60 * other empl. status	3.02	[7.98,37.41]	5.09	[3.33,7.51]
Waiting time * Age >= 60 * other empl. status	3.02	[7.98,37.41]	2.00	[0.05,5.68]
In-vehicle features [GBP/unit]				
Change the destination	2.53	[0.22,1.76]	2.81	[0.07,0.40]
Chat during the trip (AT)	1.79	[-0.11,2.42]	4.11	[0.34,0.96]
Social conformity [GBP/unit]				
Number of customers in the last hour/100 * Heard about ATs from those who used ATs	-	-	2.20	[0.03,0.51]
Good review * Age <30	2.91	[1.55,7.97]	4.40	[0.68,1.77]
Good review * Age >=30	3.06	[0.83,3.80]	4.40	[0.68,1.77]

Table 6-5 WTP t-test and confidence intervals- Survey-online UK & China

We note that all WTPs are highly significant at 95% (since the WTP is always positive, one tail test is used) and with a narrow confidence interval. We also note that, even if the MU estimated are significantly different between categories, some of the WTP confidence intervals overlap, indicating that there is a probability that respondents of different groups do have indeed the same WTP. This occurs for almost all the categories tested. Because of the different MU of travel cost between UK and China, the WTP for almost all the characteristics are also very different. In the UK, the WTP for all characteristics for short trips is higher than for long trips.

In China, it is the opposite. This reflects the discussion on the impact of income as a function of the trip duration, in **Section 6.2.1.**

Level of services attributes

In terms of level of service attributes, first we mention that all WTPs are measured in GBP/hour. We notice that UK respondents are willing to pay to save on hour of travel time more than 5 times (£18.26) the Chinese respondents are (£3.46) for shorter trips (<30 minutes) and more than 10 times (£25.03 vs. £2.41) for long trips (>= 30 minutes). In order to understand these differences, **Table 6-6** reports a summary of some of the WTPs for travel time reported in the literature, converted in GBP. As we can see, mean estimated values range from approximately £4.70 to £28.21. The WTP for the UK calculated in our sample is consistent with the values reported in the literature for the UK (Etzioni *et al.*, 2020; Polydoropoulou *et al.*, 2021). The WTP computed in our Chinese sample is smaller than the values found in the literature, but we note also that none of the studies refers to Chinese context. It is interesting to note that WTP for saving travel time in AVs varied considerably among studies. This is certainly due to different contexts and different definition of AVs (some study focuses on AT, other on SAV etc.). Another reason can also be the designs, studies use different alternatives, different levels and these affect the WTP estimated (Cherchi and Hensher, 2015). Finally, differences in the WTP are also due probably to the fact that AVs are new alternatives, unknown by the respondents and this might carry some bias in the estimation.

Analogously, results show that UK participants are willing to pay between 2 and 11 times more than Chinese participants to save one hour of waiting time, depending on the distance travelled and respondents socio-demographic characteristics (age and employment status). In the literature, waiting time is typically valued 2.5 times more than travel time. Our results for UK are in line with this general literature. The specific AV literature does not discuss much this point. Kolarova and Cherchi (2021) found the MU of waiting time to be almost twice the MU of travel time for SAV, in the ML model. But they do not discuss WTP for waiting time. Similarly, Bansal and Daziano (2018) also found that New Yorkers value the WTP for out of vehicle time (wait time and walk time) is almost twice that of in-vehicle travel time. Overall, for both travel time and waiting time, UK respondents are willing to pay more to save one hour compared to Chinese respondents. This is because, as discussed previously, UK respondents have lower MU of cost and higher MU of travel time and waiting time.

Source	Travel mode	Location	Category	WTP TT [GBP/hour]	Confidence interval
Yap *et al.* (2016)	AV	Netherland	Egress trip after train	10.16	-
Bansal and Daziano (2018)	AT	New York	All attributes	17.41	-
			Selected attributes	8.35	-
Correia *et al.* (2019)	AV	Netherland	Office-interior	4.70	-
			Leisure interior	6.98	-
Etzioni *et al.* (2020)	AV and SAV	UK	Regular trip	28.21	-
Kolarova and Cherchi (2021)	PAV	Germany	Current commuting trip	6.28	[-0.24, 12.81]
	SAV			5.32	[-0.89, 11.38]
Polydoropoulou *et al.* (2021)	SAV with strangers	UK	Regular journey	25.45	-
	AV			18.39	-

Table 6-6 Willingness to pay for travel time – from the literature
(Conversion rate: 1 GBP = 1.17 Euros; 1 GBP =1.39 USD)

In-vehicle features

In terms of in-vehicle features, interestingly UK and Chinese respondents have the same WTP for the option to talk with an operator in the AT, but not for the option to change destination: UK respondents' WTP is twice the Chinese one. However, more than this similarity/difference, what is interesting is the trade-off between the WTP for in-vehicle features and the WTP for the level of services attributes. Following Cherchi (2017), **Table 6-7** reports the amount of travel time such as the WTP for this amount equates the WTP for the in-vehicle features. The value is computed as the ratio between WTP for the in-vehicle feature and the WTP to save one minute of travel time. Results show that the amount of money UK respondents are willing to pay to have the option to change destination or to chat with an operator is the same that they are willing to pay to save 2-3 minutes of travel time. In China instead, respondents WTP for the option to chat with an operator is equivalent to the WTP to save almost 16 minutes of travel time.

	Change destination [Minutes]		Talk with an operator [Minutes]	
	UK	China	UK	China
shorter trip (<30 minute)	2.23	5.72	2.63	15.78
Long trip (>= 30minutes)	2.25	5.73	2.61	15.93

Table 6-7 Minutes of travel time that equates WTP for in-vehicle characteristics

Social conformity attributes

Table 6-8 reports the amount of travel time such as the WTP for the social conformity attributes equates the WTP for good reviews. We can see that UK respondents under 30 years old value a good review the same as saving around 11 minutes, no matter the length of the trip. And this value is same for Chinese respondents short trips, no matter the age. On the other hand UK respondents, older than 30 years, value a good review the equivalent of only 5 minutes saved in travel time.

		Good reviews [Minutes]	
		UK	China
Shorter trip (<30 minutes)	Age<30	10.78	10.58
Shorter trip (<30 minutes)	Age>=30	5.22	10.58
Long trip (>=30 minutes)	Age<30	10.79	29.63
Long trip (>=30 minutes)	Age>=30	5.25	29.63

Table 6-8 Minutes of travel time that equates WTP for good reviews

6.3 Survey-online NCL and the Survey-VR NCL

Table 6-9 includes the results of mixed logit models with panel effects estimated using the datasets from the *survey-online NCL* and the *survey-VR NCL* separately. The first two models include only the main effects and are reported for comparison purpose, the last two models instead include also systematic heterogeneity and interaction effects (only the effects that are significant). As discussed for the comparison between UK and China in **Section 6.2**, all possible effects were tested (i.e. using all the socio-demographic data available in interaction with all the level of services variables, in-vehicle attributes and social conformity attributes), with particular attention to those characteristics that allow accounting for differences between the two samples (online and VR) as discussed in **Section 4.2**. These are education level, employment status, etc. and they were tested in the models to control for differences in the sample characteristics and their potential impact in the preferences heterogeneity between survey online and with the VR.

Before discussing each effect in detail, we note that all linear effects (ML1_online NCL and ML1_VR NCL) for the level of service attributes are highly significant, at more than 95% and all the MUs are negative as expected, in line with the microeconomic theory. However, none of the AT features is significant at 95% and the option to change destination has also a counterintuitive negative sign. Among the social conformity measures, good reviews is highly significant in both datasets collected online and with the VR, confirming the importance of this piece of information. The most interesting result, however, is the MU of the number of customers. In the data collected online (i.e. in the same way as for the overall UK, as discussed in **Section 6.2**) the MU is negative, i.e. the effect is opposite to what the psychological theory expects, though not significantly different from zero. On the other hand instead, in the data collected with the VR, the MU is correct, it is positive and also significantly different from zero at more than 95%.

In models ML2, accounting for systematic heterogeneity, it was possible to identify specific categories of respondents for whom the in-vehicle features and number of customers in the last hour (for the online survey) has a correct positive effect. These effects will be discussed in detail in the **Section 6.3.2** and **6.3.3**. Looking at the overall performance of the models, all models in **Table 6-9** are significantly different and superior to the market share model (based on the Likelihood Ratio test, the assumption that each model is equal to the market share model can be rejected at 99%).

Analogously to the discussion for the comparison between UK and China, also in this case, scale differences need to be tested. Models estimated with the data collected online and the data collected with the VR might have different scale parameter, i.e. differences in the error variances due to different method to elicit preferences. To control for scale differences, a joint mixed logit model was estimated with both datasets, from *survey online NCL* and *survey VR NCL*. The joint estimation was performed using the same utility specification used in model ML2_online and ML2_VR. The scale of the VR dataset was estimated, while the scale of the online dataset was set to one (which scale is normalised has not an impact). Also in this case there was not a set of clear candidate coefficients to be constrained (i.e. with the same ratio), various coefficients once at a time were tested. The scale parameter in all these cases was never not significantly different from one. **Table 5-B** in **Appendix 5** reports one of the online/VR NCL joint estimations. As we can see the scale is not significantly different from 1 (H0 is rejected at 52% in a two-tailed test). Results were then compared directly as reported in **Table 6.9**.

Table 6-10 reports the MUs of all the characteristics tested in ML2_online NCL and ML2_VR NCL for different categories of respondents. The table reports also the t-test between the MUs estimated online and with the VR, and the sample size of each category. The t-test is computed assuming independence of the MU between online and VR. However, due to systematic heterogeneity, the standard deviation of the MU of each attribute for each specific category is computed accounting also for the correlation between the coefficients for different categories.

	ML1_ online NCL		ML1_ VR NCL		ML2_online NCL		ML2_VR NCL	
DISCRETE CHOICE MODEL	Value	Rob. t-test	Value	Rob. t-test	Value	Rob. t-test	Value	Rob. t-test
ASC (AT)	-0.778	-3.51	0.064	0.25	-1.090	-3.94	0.063	0.31
SIGMA (AT)	1.810	8.00	0.587	2.15	1.820	7.86	-0.572	-2.10
Level of services								
Travel cost	-0.416	-8.60	-0.267	-3.48	-0.467	-9.99	-0.264	-3.61
Travel time	-0.104	-4.59	-0.143	-3.77	-0.100	-4.35	-0.138	-3.75
* Master degree and PhD					-0.094	-2.97	-	-
Waiting time	-0.141	-5.50	-0.119	-2.92	-0.127	-5.00	-0.115	-3.04
* Frequent use of taxi (at least once month)					-0.074	-2.29	-	-
AT features								
Change the destination	0.143	1.09	-0.080	-0.51	-	-	-	-
... * Male					0.529	2.78	-	-
... * Age18_29							0.557	2.67
Chat during the trip	0.358	1.59	0.332	1.1	-	-	-	-
... * Age18_29					0.589	1.75		
...* Infrequent talking with driver							0.844	2.34
Social Conformity								
Number of customers in the last hour/100	-0.126	-0.97	0.318	1.99	-	-	0.320	1.95
...* Frequent use of taxi (at least once a week)					0.633	2.09	-	-
Good review yesterday	0.899	5.60	0.555	2.53	0.877	5.86	0.639	2.97
Systematic heterogeneity in AT alternative								
Male					0.915	2.48		
Summary of statistics								
Number of draws	500		500		500		500	
Log-Likelihood Market Share	-546.757		-168.628		-546.757		-168.628	
Maximum Log-likelihood	-419.323		-143.941		-405.140		-140.334	
Akaike Information Criterion	856.646		305.9		834.280		298.668	
Bayesian Information Criterion	899.182		337.2		890.995		329.993	
Number of individuals	139		40		139		40	
Number of observations	834		240		834		240	

Table 6-9 model estimation results – Survey-online NCL and the Survey-VR NCL

In **Table 6-10** the MU are computed using equation (5.29) in **Chapter 5**, and the t-test for generic coefficients, using equation (5.24) still in **Chapter 5**.

The results from models ML2_online and ML2_VR in **Table 6-9** and their MU computed in **Table 6-10** are used in the next sections to discuss the specific effect of each characteristic on the choice of AT versus NT and the cross-methodological difference.

	Marginal utilities			Sample size	
DISCRETE CHOICE MODEL	ML2_ online	ML2_VR	t-test	Online NCL	VR NCL
Level of services					
Travel cost [GBP]	-0.467	-0.264	-2.34	139 (100%)	40 (100%)
Travel time * Master and PhD [minutes]	-0.194	-0.138	-1.17	31 (22%)	24 (60%)
Travel time * up to Bachelor [minutes]	-0.100	-0.138	0.88	108 (78%)	16 (40%)
Waiting time * Frequent use of taxi (once a month) [minutes]	-0.201	-0.115	-1.78	77 (55%)	25 (62%)
Waiting time * Not frequent use of taxi (less than once a month) [minutes]	-0.127	-0.115	-0.26	62 (45%)	15 (38%)
AT features					
Change the destination * male * Age <30	0.529	0.557	-0.10	15 (11%)	7 (18%)
Change the destination * female * Age <30	-	0.557	-	25 (18%)	8 (20%)
Change the destination * male * Age >=30	0.529	-	-	45 (32%)	18 (45%)
Change the destination * female * Age >=30	-	--	-	54 (39%)	7 (18%)
Chat during the trip * Age <30 * Infreq. Talk to driver	0.589	0.844	-0.52	14 (10%)	7 (18%)
Chat during the trip * Age >=30* Infreq. Talk to driver	-	0.844	-	10 (7%)	7 (18%)
Chat during the trip * Age <30 * Freq. talk to driver	0.589	-	-	26 (19%)	8 (20%)
Chat during the trip * Age >=30* Freq. talk to driver	-	-	-	89 (64%)	18 (45%)
Social Conformity					
N. Customers (..) * Frequent use of taxi (once a week)	0.633	0.320	0.91	12 (9%)	2(5%)
N. Customers (..) * Infrequent use of taxi (less than …)		0.320	-	127 (91%)	38 (95%)
Good review yesterday	0.877	0.639	0.91	139 (100%)	40 (100%)
Systematic heterogeneity in AT alternative					
Male	-0.175	0.063	-0.57	60 (43%)	25 (62%)
Female	-1.090	0.063	-3.44	79 (57%)	15 (38%)

Table 6-10 Marginal utilities – Survey-online NCL &survey-VR NCL

6.3.1 Impact of the level-of-services attributes

As mentioned before, all coefficients associated to the level of service attributes in both datasets (online and VR) are significant at 99% and all MUs have the correct negative sign for all the categories in the sample. Given the sample size, in particular for the sample-VR, the amount of systematic heterogeneity tested can be limited. However, systematic heterogeneities in the preference for travel cost were tested, but found not significant at 95%. In particular, the interaction between income and travel cost was not significant. In **Table 6-10**, the t-test of the MU of travel cost between the sample-online and sample-VR shows that these MU are significantly different at 99% level and the respondents who filled in the survey online are twice as much sensitive to the effect of travel cost than the respondents who participated in the immersive VR experiment.

In the sample-online the proportion of participants in the lowest level of income is higher than in the sample-VR but the distribution of income between the two samples is not statistically different (as discussed in **Section 4.3.1**). According to Meißner *et al.* (2020) and Fang *et al*. (2021), customers in high immersive VR are generally less price-sensitive. Their argument is that customer excitements triggered by enriched contents presented in high immersive VR may lower price-sensitivity. In this papers they refer to 'situational' price sensitivities, i.e. how customers respond to the price differences or changes in different contexts (Wakefield and Inman, 2003). This phenomenon (i.e. low price-sensitivity) probably can be explained by the increased psychological ownership (i.e. feeling of possession and psychological bond) in VR. For non-owners or buyers, this perceived ownership can be increased by merely touching the object (Peck and Shu, 2009). In line with this, despite the respondents do not actually own NT or AT, there could be a psychological bond due to the experience lived interacting with the taxis, experiencing first-hand the re-designed taxi rank at Northumberland Rd., buying the tickets in a touch screen ticket board. While this can be an explanation, these results need to be verified, because Meißner *et al.* (2020) and Fang *et al*. (2021) studies referred to food purchase. According to the psychological literature, food purchase triggers more brain activities than other purchases because eating is an essential need for human beings. Nevertheless, these results are relevant as represent the first evidence of different sensitivity for price and other characteristics measured with survey-online and survey-VR.

Waiting time is a very important variable in the choice of taxi and where we expected the immersive experience might highlight differences in the preferences between the sample-online and sample-VR. Results in **Table 6-10** confirm our expectation, with the MU computed from the survey-online almost twice the MU computed with the survey-VR for those who use taxi frequently (more than once a month), which account for more than 65% of both samples. It makes sense that those who use taxi frequently have a higher (in absolute terms) MU, though it is interesting that this effect is not captured in the immersive VR experiment, probably due to the small sample size. One possible explanation could be that in the VR experiment respondents could look at the number of passengers queuing and maybe had a feeling that the waiting time was not too long. This problem was discussed in the initial SC designs (see **Section 3.5**). On the other hand respondents in the survey-VR have much (1.5 times) higher MU for travel time than respondents in the survey-online and higher MU (1.2 times) than the MU for waiting time.

This result is again not expected, as waiting time is typically valued 2.5 more than travel time; however there are two possible considerations. The 2.5 ratio between MU of waiting time and travel time is typically computed for public transport, and for taxi it could be different. As far as we are aware, there are no studies specifically for taxi. Secondly, again, the area where the taxi rank is located is a pleasant area, close to a high street; it could be that waiting time could not be perceived as too unpleasant.

6.3.2 Impact of in-vehicle features

Results in **Table 6-9** show that in-vehicle features are not significant in both samples. This result was not expected, in particular giving the results from the sample UK discussed in **Section 6.2.2.** One reason for lack of significance could be the small sample size. But it can also be due to the different context, in the datasets collected in NCL the context is specifically a taxi rank, in the dataset collected in the UK the context is not restricted to a taxi location. Nevertheless, we found that in-vehicle features are highly significant for some categories, though these categories are different between sample-online and sample-VR and it is not straightforward to identify the reasons. It is interesting to note that those who do not frequently talk with drivers do value positively the option of chatting with an operator on board of the ATs. This might be because chatting with an operator is quite different from chatting with a real person. Similar to the Chinese datasets, those who enjoy talking with taxi driver and some elderly people requested this 'chat' option in FGs, do not have a preference to have this request within the ATs, either. It is interesting also to note that this effect appears only in the survey-VR not in the survey-online, where instead young (< 30 years old) seems to prefer the option to chat with an operator. Analogously, the option of changing destination is valued positively by male in the survey-online and by young respondents in the survey-VR. This result is not clear to justify, and it is likely to be due to the sample size, which does not allow to get robust results for these attributes. We would not try then to elaborate more on that.

6.3.3 Impact of normative conformity

Perhaps the most interesting result in this part of the research is the effect of the descriptive norm 'number of customers in the last hour' that in the immersive VR experiment is positive and significant at 95%, while in the survey-online it is significant only for those who use taxi frequently, that account for 65% of the sample. As discussed in the literature review, descriptive norms do have a strong impact in consumers choices, but it seems to be difficult to measure this effect within SC experiments. It is then very interesting that this attribute is significant when measured in the immersive VR experiment, as it probably reflects the importance of realism in the ability to capture the impact of the descriptive norm. On the other hand, the effects of good reviews is confirmed to have a strong impact on the choice whatever is the instrument used to collect the data, with no significant differences in both datasets at 95% (in **Table 6-10**, the t-test between MU of good reviews in the survey- online and the survey-VR is 0.91, H0 rejected at 64%).

6.3.4 Impact of knowledge of AVs and ATs

Systematic heterogeneity in the preference for AT alternative and for all the attributes as a function of the knowledge of AVs and ATs was tested in both samples, but none was found significant. In both samples (survey-online NCL and survey-VR NCL) the proportion of respondents who has heard of AV and AT testing, and has knowledge of the levels of automation is similar to that registered in the sample UK where the impact was significant (see **Section 6.2.4**). Except for the sample dimension or maybe the context, we could not find other reason why the impact of knowledge does not have an impact.

6.3.5 Other effects

Finally, we note that in the online survey males show a stronger preference for ATs, which is consistent with other finding in the literature, as men are more open to the new technology (see discussion in the **Section 6.2.5**). The effect is not confirmed in the VR experiment. Probably this is due to the fact that in the VR experiment the effect cannot be disentangled because there is a high proportion of male and high educated participants (though these two attributes are not significantly related), which are more likely to be the savvy-tech person, thus they are more likely to accept the use of automated taxis.

Another effect tested was the impact of the final message presented before the last SC scenario informing respondents that in their last trip they had to wait 10 minutes longer than what reported in the ticket board when they bought the ticket. This message was intended to give respondents a feedback on the consequence of their trip. Interesting results show that this message did not seem to have an impact on the preference for waiting time, not in the survey-online, not in the survey-VR.

Finally, we note also that the ASC is not significantly different from zero in the VR-based experiment, while it is highly significant in the online experiment. Interestingly, it seems that living the choice in the immersive VR environment reduce the ‘label’ effect, i.e. the effect associated to the specific alternative besides its characteristics.

6.3.6 Willingness to pay

Tables 6-11 reports the mean value of the WTP estimated for all the attributes tested for both dataset using the ML results reported in **Table 6-9** or equivalently in **Table 6-10**. **Tables 6-11** also reports t-tests and confidence intervals only for the ML models. These are computed using Monte Carlo simulations with 5,000 draws from a multivariate truncated Normal distribution. We note that all WTPs are significantly different from zero at 95% (two-tailed t-test) in the online survey, but not in the VR. This is due to the sample size, because the standard deviation of the coefficients estimated is inversely related to the sample size (Bliemer and Rose, 2009). This affects also the confidence interval that is quite large in both samples.

	Online NCL			VR NCL		
	Mean value	t-test	Confidence interval	Mean value	t-test	Confidence interval
Level of services [GBP/Hour]						
Travel time * MSc and PhD	24.91	4.44	[14.10,36.38]	34.72	1.99	[0.59,68.84]
Travel time * up to Bachelor	12.85	4.00	[6.63,19.37]	34.72	1.99	[0.59,68.84]
Waiting time * Frequent use of taxi (once a month)	25.76	4.38	[14.42,37.74]	29.02	1.84	[-1.85,59.90]
Waiting time * Not frequent use of taxi (less… a month)	16.32	4.41	[9.12,23.74]	29.02	1.84	[-1.85,59.90]
AT features [GBP/unit]						
Change the destination * male * Age <30	1.13	2.67	[0.30,1.97]	2.34	1.80	[-0.20, 4.88]
Change the destination * female * Age <30				2.34	1.80	[-0.20, 4.88]
Change the destination * male * Age >=30	1.13	2.67	[0.30,1.97]			
Change the destination * female * Age >=30	-	-	-	-	-	-
Chat during the trip * Age <30 * Infreq. Talk to driver	1.26	1.75	[-0.15,2.71]	3.55	1.58	[-0.86,7.96]
Chat during the trip * Age >=30* Infreq. Talk to driver				3.55	1.58	[-0.86,7.96]
Chat during the trip * Age <30 * Freq. talk to driver	1.26	1.75	[-0.15,2.71]			
Chat during the trip * Age >=30* Freq. talk to driver	-			-		
Social Conformity [GBP/unit]						
N. Customers (..) * Frequent use of taxi (once a month)	1.36	2.05	[0.03,2.72]	1.21	1.43	[-0.50, 3.17]
N. Customers (..) * Infrequent use of taxi (less than …)				1.21	1.43	[-0.50, 3.17]
Good review yesterday	1.88	4.91	[1.14, 2.66]	2.42	1.75	[-0.32, 5.72]

Table 6-11 WTP mean values and t-tests – Survey-online NCL & Survey-VR NCL

Level of services attributes

In terms of level of service attributes, we notice that among the respondents living in Newcastle, those with low level of education are willing to pay £12.85 to save an hour of travel time, which is almost half the amount that people with higher level of education are willing to pay (£24.91). This result might be related with the distance travelled.

From **Table 6-4** we see that UK respondents are WTP £18.26 for short trips and £25.03 for medium/long trips. From **Table 4-17** in **Chapter 4** we notice that 55% of the respondents online selected a destination within 5 km, while in the sample-VR, 62% of the respondents selected medium-long distances. It seems then that there might be some confounding effect between level of education and distance travelled, and it is not possible to distinguish which effect is playing a role, though the interaction with distance was tested also in these data (online NCL and VR NCL) but it was not found significant. Interestingly, in the sample who used the VR experiment, respondents' WTP for travel time is even higher (£31.36). This might occur because the sample is formed mainly by highly educated people and they mostly selected medium/long distance trip.

Regarding waiting time, in the online survey Newcastle residents who infrequently use taxis (at least once a month) are willing to pay £16.32 to save one hour of waiting time, while those who frequently use taxis in Newcastle are willing to £25.76 to save one hour of waiting time. It is plausible that the more frequently they use taxi, the less time they are willing to waste waiting for taxis. This value for those who frequently used taxis is close to the WTP for saving one hour of waiting time in VR survey (£26.14).

In-vehicle features

Table 6-12 reports the amount of travel time such as the WTP for this amount equates the WTP for the in-vehicle features. The value is computed as the ratio between WTP for the in-vehicle feature and the WTP to save one minute of travel time. Results show that in the survey-online the amount of money male respondents are willing to pay to have the option to change destination is equivalent to the amount that are willing to pay to save approximately 5.3 minutes of travel time for those with low education level and 2.7 minutes of travel time for those with high education level. In the survey-VR instead, to have the option of change the destination (£2.11) young participants are willing to pay an amount equivalent to saving 4 minutes of travel time.

Regarding the possibility to change destination during the trip and to chat with an operator, participants in the survey-online are willing to pay much less than participants in the survey-VR. However, none of these WTPs in the survey-VR is significantly different from zero at 95% (even in a one-tail tests, the assumption can be rejected at 94%), and the confidence intervals are quite wide. This makes it difficult to compare these results. A bigger sample is needed to be able to reach conclusions on the impact of the in-vehicle features.

For completeness, in comparison with the analyses reported for the UK and Chinese datasets, **Table 6-12** reports the minutes of travel time that equates mean WTP for in-vehicle features. In line with the discussion above, in the survey-VR the trade-off with travel time is twice than in the survey-online

	Change destination [minutes]		Talk with an operator [minutes]	
	Online	VR	Online	VR
Among master and PhD degree				
Male * Age <30	2.72	4.04		
Female * Age <30	-	4.04		
Male * Age >=30	2.72	-		
Female * Age >=30	-	-		
Age <30 * Infreq. Talk to driver			3.03	6.13
Age >=30* Infreq. Talk to driver			-	6.13
Age <30 * Freq. talk to driver			3.03	-
Age >=30* Freq. talk to driver			-	-
Up to bachelor degree				
Male * Age <30	5.28	4.04		
Female * Age <30	-	4.04		
Male * Age >=30	5.28	-		
Female * Age >=30	-	-		
Age <30 * Infreq. Talk to driver			5.88	6.13
Age >=30* Infreq. Talk to driver			-	6.13
Age <30 * Freq. talk to driver			5.88	-
Age >=30* Freq. talk to driver			-	-

Table 6-12 Minutes of travel time that equates WTP for in-vehicle characteristics

Social conformity attributes

Results show that participants in the survey-online NCL are willing to pay on average £1.88 to use a taxi that got a good customer rating. While this value is slightly higher for respondents in the VR survey, who are willing to pay on average £2.42, which is equivalent to the amount of saving 4.6 minutes of travel time and 5.6 minutes of waiting time.

6.4 Survey-online China – Impact of AT existence

As discussed in **Chapter 3,** the sample collected in China included 369 respondents who live in the cities of Guangzhou, Changsha and Shanghai where AT services were in operation (this group of respondents was defined in **Chapter 4** as *AT cities*) and 151 respondents who live in cities where ATs do not operate (defined in **Chapter 4** as *NT cities*). The models discussed in this section aims to identify the impact of living in a city where AT is operating on the preference for AT. The Chinese sample includes also 35 respondents who have used AT in China. It is very interesting to test if those who actually used the AT have different preference, however, given the small sample, we could not find significant effects, and it is not clear if this is due to the lack of effect, or the sample size. The remaining of this section will focus on the impact of AT existence.

Table 6-13 reports the best model results (only the effects that are significant at more than 80%). It only presents those coefficients that are significantly different between AT cities and NT cities at more than 80%. The models are a mixed logit model with panel effects estimated using the best model specification reported for the online China dataset (model ML2-China in **Table 6-1**) and a HCM estimated using the best model specification reported for the online China dataset (model HCM-China in **Table 6-3**). These model specifications were then extended to examine the impact of AT-existence in preference on AT and its corresponding characteristics. The coefficients, which were not significantly different between these two types of cities, were defined as generic.

All possible interaction effects with AT-existence were tested. But, several MU were not significantly different between AT-cities and NT-cities (based on the t-test for generic coefficients) and then coefficients were constrained to be generic between these two categories. The model with specific coefficients of all attributes is included in **Appendix 5, Table 5-C**.

Looking at the models in **Table 6-13**, we note that the ML and the HCM returns the same estimates with the exception of the ASC and the category of respondents who heard of AT operating in China. This makes sense because the LVs in the HCM are summed into the utility of AT, which of course directly affects the ASC and any category that has a direct impact on the preference for one alternative over the other. Apart from that, all other results are identical. Comments will then be reported only for the HCM.

	ML		HCM	
DISCRETE CHOICE MODEL	Value	Rob. t-test	Value	Rob. t-test
ASC (AT)	-0.677	-4.27	-5.520	-12.00
SIGMA (AT) *AT existence	1.130	9.94	0.917	7.83
SIGMA (AT) *AT non-existence	1.910	8.84	1.360	6.99
Level of services				
Travel cost [GBP][a]	-0.588	-8.00	-0.590	-8.09
*Long trip (30 minutes or more)	-0.257	-2.37	-0.251	-2.33
Travel time [minutes]	-0.031	-5.20	-0.031	-5.21
Waiting time [minutes]	-0.075	-9.39	-0.075	-9.36
* Age60_more	0.027	1.70	0.028	1.75
AT features				
Change the destination				
*AT existence	0.156	2.23	0.158	2.25
*AT non-existence	0.371	3.46	0.352	3.46
Chat during the trip	0.517	5.20	0.518	5.25
Social Conformity				
Number of customers in the last hour/100				
… * Heard of ATs from those who used it	0.239	3.05	0.242	3.00
Good review yesterday				
… * AT existence	0.525	5.41	0.530	5.39
… * AT non-existence	0.188	1.83	0.181	1.81
… * Long trip (30 minutes or more)	0.547	3.93	0.549	3.91
Systematic heterogeneity in AT alternative				
Age18_29	-0.446	-2.67	-	-
Frequently use taxis (at least once a week)				
… * AT existence	0.916	5.66	0.562	3.46
… * AT non-existence	-	-	-	-
Heard of ATs operating in China	0.679	4.40	0.259	1.86
Latent psychological constructs				
IN: Injunctive norms (AT)			0.490	5.09
HM: Hedonic Motivation (AT)				
… * AT existence			0.495	5.31
T: Trust (AT)				
… * AT non-existence			0.564	5.55
Summary of statistics				
Number of draws	500		500	
Log-Likelihood Market Share	-1895.77		-14451.35	
Maximum Log-likelihood	-1654.33		-9580.61	
Akaike Information Criterion (AIC)	3344.662		19279.223	
Bayesian Information Criterion (BIC)	3452.228		19631.801	
Number of individuals	485		485	
Number of observations	2910		2910	

Table 6-13 AT-existence – Survey online China

[a] the unit of travel cost was converted from CNY to GBP Conversion rate: 1 GBP = 8.43 CNY

LATENT VARIABLE MODEL	IN _China		HM _China		T _China	
	Value	Rob. t-test	Value	Rob. t-test	Value	Rob. t-test
Structural model						
Constant	4.690	44.13	4.930	64.15	4.640	45.37
Standard deviation of error term	0.021	0.32	-0.111	-1.64	0.396	9.12
Male						
Age18_29	-0.597	-4.57			-0.406	-2.39
Frequent use of taxis (at least once a week)	0.374	3.18	0.345	3.35		
Familiar with 5 levels of automation	0.395	2.47	0.538	4.12		
Heard of ATs from those who used ATs	0.645	4.19	0.236	2.04		
Heard of ATs from those who not used ATs	0.261	1.81				
Measurement model						
Constant in indicator N2	1.750	4.85	-0.708	-1.50	0.329	1.31
Constant in indicator N3	-0.054	-0.19	1.980	4.32	0.562	2.42
Constant in indicator N4					3.060	11.00
Coeff. in indicator N2	0.473	6.13	1.110	12.90	0.879	16.85
Coeff. in indicator N3	0.968	17.57	0.482	5.36	0.897	19.94
Coeff. in indicator N4					0.396	7.25
Standard deviation indicator N1	-0.200	-3.07	-0.208	-3.01	-0.242	-3.11
Standard deviation indicator N2	0.441	14.59	-0.371	-4.72	0.049	0.76
Standard deviation indicator N3	-0.198	-3.39	0.406	12.48	0.007	0.11
Standard deviation indicator N4					0.257	7.92

Table 6-13 (continued) AT-existence – Survey online China

6.4.1 Impact of the level-of-services attributes

As expected, the existence AT services in the city where respondents live do not have any impact on the MU of travel cost, travel time and waiting time. The full specifications with the MU of all the level-of-services attributes specific for AT cities and NT cities is reported in **Appendix 5**, **Table 5-C,** where it can be appreciated that the hypothesis that the MU of AT cities is equal to the MU of NT cities can be rejected at less than 60% in a two-tailed test for all three level of service attributes.

6.4.2 Impact of in-vehicle features

Among the in-vehicle features, interestingly, respondents have the same MU for the option to chat with an operator whether they live in an AT city or NT city (the H0 is rejected at 28% in a two-tailed test). This can be due to the fact that the AT operating in China still have a safety driver inside, i.e. there is a physical person in the car, although it is not actually driving. On the other hand, respondents living in NT cities, have much lower (around half) MU for the option ‘change the destination’ than those living in AT cities (the MU is different at 88% in a two-tailed test, t-test = 1.56).

In terms of familiarity with AV or ATs (see **Table 4-24**), no matter generic AVs knowledge and specific ATs knowledge, the proportion of respondents from AT cities who are familiar with AVs or heard of AVs or ATs operating in China are significantly more than that of respondents from NT cities, indicating average knowledge level regarding AVs and ATs is higher among respondents from AT cities compared to NT counterpart. Therefore, they would be more confident in using AT and whether ATs are equipped with this feature have less impact on them. Another reason might be higher frequency of usage of taxis, as shown in **Table 4-23** trip characteristics, compared to respondents from NT cities, the vast majority of respondents from AT cities used taxis at least once a month (of which, there are over half using taxis more than once a week). This probably indicates that they might be also familiar with traditional taxis and maybe less probability to change destination selection during the trip.

6.4.3 Impact of normative conformity

Despite the fact that the descriptive norm 'number of customer in the last hour' was not significant among the entire sample collected in China, it was expected that this variable could have been significant at least for the respondents from AT cities, as they have a higher knowledge of AVs and ATs. Therefore, AT-existence was also tested in interaction with this descriptive norm. Unexpectedly, the descriptive norm was not significant in either samples: not AT cities nor the sample of NT cities.

Interestingly, the AT-existence was found to have a significant impact on the MU of 'good review yesterday'. The assumption that the MU of good review between AT cities and NT cities is equal is rejected at 99% (t-test= -2.56), and the MU of good review yesterday in AT cities is about 2.5 times higher than that in NT cities. This result makes sense because respondents from AT cities have of course a higher possibility to access and use ATs, as an alternative travel mode. Even if they have not used the ATs personally, for them the customer reviews provide a stronger hint or cue for assessing and comparing the quality of these two types of taxi services, because they know that this review comes from real customers in their city.

6.4.4 Impact of knowledge of AVs and ATs

Finally, results in **Table 6-13** show that the impact of having heard of AT operating on the preference for AT is not different whether respondents live in AT cities or NT cities (H0 is rejected at 23%). This result is not in line with the assumption that living in a city where ATs are operating provides stronger cue, as discussed for the reviews. However, the difference is that a review implies an evaluation (positive or negative) while having only heard of it from others is a ‘neutral’ information. Other three variables measuring knowledge of AVs and ATs (heard of AVs, familiarity with 5 levels of AVs and from whom they heard about ATs) were not significant in the entire Chinese sample, they were also tested to examine if only significant among either the sample of AT cities or the sample of NT cities. Nevertheless, none of them was found to be significant at 95%. Additionally, the direct impact of AT-existence in the preference for AT was also tested but not found to be significant at 95%, either. The t-test allowed rejecting the assumption that the coefficient are equal to zero at only 23%.

6.4.5 Other effects

As shown in the **Table 6-2**, we found that in China those who used taxis at least once a week have a higher preference for ATs rather than NTs. However, after including the impact of AT-existence, results show that only those respondents from AT cities who frequently use taxis have a significantly (99%) positive preference for ATs. This effect is instead insignificant among those from NT cities (H0 assumption that this coefficient is equal to zero is rejected at 55%, robust t-test = 0.75). This is in line with the results from the descriptive analyses in **Table 4-27** that reports that approximately 65% of those who used taxis at least once a week in AT cities choose ATs while less than 30% who live in NT cities choose ATs.

6.4.6 Latent psychological constructs

Finally, regarding the latent psychological constructs, we note that injunctive norms are highly significant for respondents living in both AT and NT cities, but not significantly different between them (the t-test for generic injunctive norms coefficients between AT cities and NT cities is -0.11, i.e. H0 assumption rejected at 9%). On the other hand, the other two latent constructs are significantly different from zero only for one category: hedonic motivation is significant only for respondents living in AT cities, while trust is significant only for those living in NT cities. It is plausible that those who live in NT cities do not have the possibility to see the ATs operating, and then perceived trust as important. On the contrary, those living in AT cities, have probably seen AT circulating, though they have not used them in person, and could directly see that AT can be trusted. This is consistent with the result presented in **Table 4-25**, where the average ratings in the three statements of hedonic motivation, are higher for those living in AT cities than those living in NT cities. This might be attributable to higher knowledge level of AVs and ATs among respondents from AT cities (See **Table 4-24**) and in line with the modelling result in **Table 6-13**, the impact of hedonic motivation is positively correlated with knowledge levels of AVs and ATs

6.4.7 Willingness to pay

Table 6-14 reports the mean value of the willingness to pay (WTP) estimated for all the attributes estimated with the HCM reported in **Table 6-13** also reports t-tests and confidence intervals. As before, these are computed using Monte Carlo simulations with 5,000 draws from a multivariate truncated Normal distribution. We note that all WTPs are highly significant (t-test >1.96) and with a narrow confidence interval, with some exceptions of 'number of customers in the last hour *used ATs' and 'good review yesterday' in NT cities.

	Mean value		T-test		95% Confidence level	
	AT cities	NT cities	AT cities	NT cities	AT cities	NT cities
Short or medium trip (< 30 minutes)						
Level of services [GBP/hour]						
Travel time	3.31		4.24		[1.78, 4.84]	
Waiting time * Age < 60	7.77		5.89		[5.18, 10.35]	
Waiting time * Age >= 60	5.00		2.50		[1.08, 8.93]	
In-vehicle features [GBP/unit]						
Change the destination	0.27	0.64	2.10	3.12	[0.02,0.52]	[0.24,1.04]
Chat during the trip (AT)	0.64		3.20		[0.25, 1.04]	
Social conformity [GBP/unit]						
Number of customers in the last hour/100 * Heard about ATs from those who used ATs	0.41		2.84		[0.13, 0.70]	
Good review	0.90	0.32	4.32	1.75	[0.49, 1.30]	[-0.04, 0.69]
Long trip (>= 30 minutes)						
Level of services [GBP/hour]						
Travel time	2.34		3.85		[1.15, 3.54]	
Waiting time * Age < 60	5.49		4.97		[3.32, 7.65]	
Waiting time * Age >= 60	3.54		2.43		[0.69, 6.38]	
In-vehicle features [GBP/unit]						
Change the destination	0.19	0.45	2.09	3.00	[0.01, 0.37]	[0.16, 0.75]
Chat during the trip (AT)	0.62		3.85		[0.31, 0.94]	
Social conformity [GBP/unit]						
Number of customers in the last hour/100 * Heard about ATs from those who used ATs	0.29		2.70		[0.08, 0.50]	
Good review	1.31	0.91	4.31	3.48	[0.72, 1.91]	[0.40, 1.42]

Table 6-14 Willingness to pay mean values, t-tests and confidence levels – AT-existence China

Looking at the WTPs that are different between AT cities and NT cities, we note that for 'change the destination', the respondents from NT cities are willing to pay more than twice to have this option in the taxi compared to respondents from AT cities, which is equivalent (see **Table 6-15)** to save about 11.5 minutes of travel time. In terms of 'good review yesterday', we can see that for short trip, respondents from AT cities are willing to pay almost 3 times (£0.90/unit) more than respondents in NT cities. This is equivalent to the amount of money they are WTP to save about 16 minutes of travel time. For long trip, respondents from AT cities are willing to pay £1.31 to use taxis with good reviews, this value is equivalent to save 33-34 minutes of travel time, which is about 44% higher than the amount that respondents from NT cities are willing to pay to use taxis with good customer reviews.

	Change destination [minutes]			Good review yesterday [minutes]		
	AT cities	NT cities	China	AT cities	NT cities	China
Shorter trip (<30 minute)	4.89	11.60	5.72	16.31	5.80	10.58
Long trip (>= 30minutes)	4.87	11.54	5.73	33.59	23.33	29.63

Table 6-15 Minutes of travel time that equates WTP for change destination and taxi with good reviews

6.5 Policy Implications

This section will discuss the results of the application of some policy scenarios using the models estimated. Since models are estimated with SC data, the models do not reproduce the real market share, but the hypothetical one implicit in the SC scenarios. Models estimated with SC data cannot be used in the forecast, as the alternative specific constants (ASCs) and the scale of the model need to be calibrated to reproduce the current market share (Cherchi and Ortúzar, 2006). However, this calibration is not possible or very difficult to do in the case of innovative products, because a current market share does not exist or it is too small to be used as a base reference to test policies, as results will be biased downwards (See a discussion in Jensen *et al.*, 2017).

The analyses reported in this section provide then a sensitivity test to discuss the applicability and forecasting capabilities of the model developed, but cannot be used to discuss the actual prediction of the ATs. We assume the estimated market share (i.e. using the ASC estimated in the model) as the reference market share and we will discuss the impact of different policies with respect to this reference market share. Dataset 3 (Survey-online UK) and Dataset 4 (Survey-online China) will be used for this analysis because these datasets are much bigger than Dataset 1 and Dataset 2 and model results more robust. The HCMs reported in Table 6-3 are used for testing different policy scenarios, as these are the best models estimated for the UK and China, respectively.

The base scenario is a "no policy" scenario and it corresponds to the actual values used in the SC experiments for both the UK and China. These values are summarised in **Table 6-16**. For the dummy variables the values represent the proportion of the sample that had available an AT equipped with the option 'chat with an operator' or the option "change the destination" or had available a NT and an AT rated with good reviews. For the continuous variables, the values represent the average travel time in the sample, the average waiting time and the average travel cost.

No policy scenario (values in the sample)	UK		China	
	NT	AT	NT	AT
Change the destination (%)	50	50	50	50
Chat with an operator (%)	n.a	50	n.a.	50
Rated with good reviews (%)	50	50	50	50
Number of customers (average)	80.00	80.43	240.00	240.90
Travel time [minutes] (average)	13.61	13.64	18.79	18.84
Travel cost [£][a] (average)	11.21	11.21	3.76	3.77
Waiting time [minutes] (average)	5.98	5.99	7.07	7.00

Table 6-16 Mean values of attribute levels in the no-policy scenario- UK & China
n.a. not applicable
[a] the unit of travel cost was converted from CNY to GBP Conversion rate: 1 GBP = 8.43 CNY

The models are applied for segments of the population based on the socio-economic characteristics included in the utility specification directly and indirectly through the latent variables. In using the model in prediction, panel effect was not included, as this is only needed for estimation purposes, while the random parts of each latent variables need to be taken into account. For each segment, the predicted conditional probabilities were integrated over the latent variable distribution using Monte Carlo simulation.

Table 6-17 reports the simulated MS of the reference scenario for the entire samples and for the groups (or segments) relevant in the models estimated. 12 segments were identified based on the SE characteristics that were significant in HCM models (i.e. gender, age, and employment status). Groups 2, 5 and 9 account for more than half sample in the UK, while Groups 5, 7 and 11 account for more than half sample in China. Under the no-policy scenario, the market shares of AT were predicted to be 31.2% in the UK and 49.8% in China. We remember, as reported in **Chapter 4**, that the stated choice shares of ATs in the UK and China are 35.7% and 51.2% respectively. The difference in the stated market share among groups has been discussed already in the previous chapters.

Group	Respondents' Characteristics			Sample Distribution %	
	Gender	Age	Employment status	UK	China
Gr1	Female	Age18_29	Not Fulltime employee	8.9	7.1
Gr2	Female	Age30_59	Not Fulltime employee	18.0	3.3
Gr3	Female	Age60+	Not Fulltime employee	2.8	6.9
Gr4	Female	Age18_29	Fulltime employee	4.4	7.1
Gr5	Female	Age30_59	Fulltime employee	17.4	18.5
Gr6	Female	Age60+	Fulltime employee	0.3	1.6
Gr7	Male	Age18_29	Not Fulltime employee	6.3	11.8
Gr8	Male	Age30_59	Not Fulltime employee	3.5	1.6
Gr9	Male	Age60+	Not Fulltime employee	20.9	3.8
Gr10	Male	Age18_29	Fulltime employee	3.2	8.0
Gr11	Male	Age30_59	Fulltime employee	9.8	27.8
Gr12	Male	Age60+	Fulltime employee	4.4	2.4
Whole sample				100.0	100.0

Table 6-17 Aggregate choice probability in the base scenario for different segments.

The following four scenarios are tested and compared with the reference scenario:

Scenario N1: variations in the LOS attributes

Scenario N2: variations in the in-vehicle features

Scenario N3: variations in the social conformity attributes

Scenario N4: improvements in the latent constructs

6.5.1 Scenario N1: variations in the LOS attributes

The first scenario tested implies a 20% decrease in the LOS characteristics of the ATs:

A 20% decrease of travel cost; this would mimic the policy measures that ATs are subsidised by the government or AT fares receive tax reduction etc.;

A 20% decrease of travel time; this policy measure can be achieved through the construction of the exclusive automated vehicle lanes or enhancement of AT travel speeds etc.; and

A 20% decrease of waiting time; this policy measure can be achieved by introducing more AT vehicles etc.

Table 6-18 reports the AT market shares variations after applying Scenario N1 and **Figure 6.1** visualises the effect of Scenario N1 on AT market shares. A reduction in AT travel cost is the policy that has the highest impact in the AT market share. However, the impact shows significant heterogeneity among groups, with female, age 60+ and fully employed (group 6) showing the lowest impact in the UK but the highest in China and vice versa male, age 18-29 and no full time employee (group 7) showing the lowest impact in China and male, over 60 fulltime employed (group 12) the highest impact in the UK. In general, with few exceptions (group 2, 4 and 6), the impact of a change in travel cost is always lower in the Chinese than the UK market.

Group	Respondents' characteristics			AT market share variations (%)					
				UK			China		
				20% decreases in			20% decreases in		
	Gender	Age	Employment status	TC	TT	WT	TC	TT	WT
Gr1	Female	Age18_29	Not Fulltime employee	10.6	4.1	1.6	10.3	2.4	1.9
Gr2	Female	Age30_59	Not Fulltime employee	10.1	3.8	1.6	11.0	2.5	2.0
Gr3	Female	Age60+	Not Fulltime employee	12.8	4.9	1.5	10.5	2.5	1.0
Gr4	Female	Age18_29	Fulltime employee	10.0	3.7	2.1	11.6	2.6	1.9
Gr5	Female	Age30_59	Fulltime employee	10.4	4.0	2.3	9.1	2.3	1.9
Gr6	Female	Age60+	Fulltime employee	7.4	3.1	2.9	13.0	2.7	0.9
Gr7	Male	Age18_29	Not Fulltime employee	12.8	4.9	1.9	8.5	2.1	1.9
Gr8	Male	Age30_59	Not Fulltime employee	10.1	4.0	1.6	10.8	2.5	2.1
Gr9	Male	Age60+	Not Fulltime employee	12.4	4.7	1.7	10.3	2.4	1.0
Gr10	Male	Age18_29	Fulltime employee	12.2	4.6	2.7	10.5	2.4	1.9
Gr11	Male	Age30_59	Fulltime employee	12.0	4.5	2.4	9.7	2.3	1.9
Gr12	Male	Age60+	Fulltime employee	13.1	4.7	2.3	11.2	2.6	1.0
Whole sample				11.3	4.3	1.9	9.9	2.4	1.8

Table 6-18 Market share variations for changes in level of services attributes

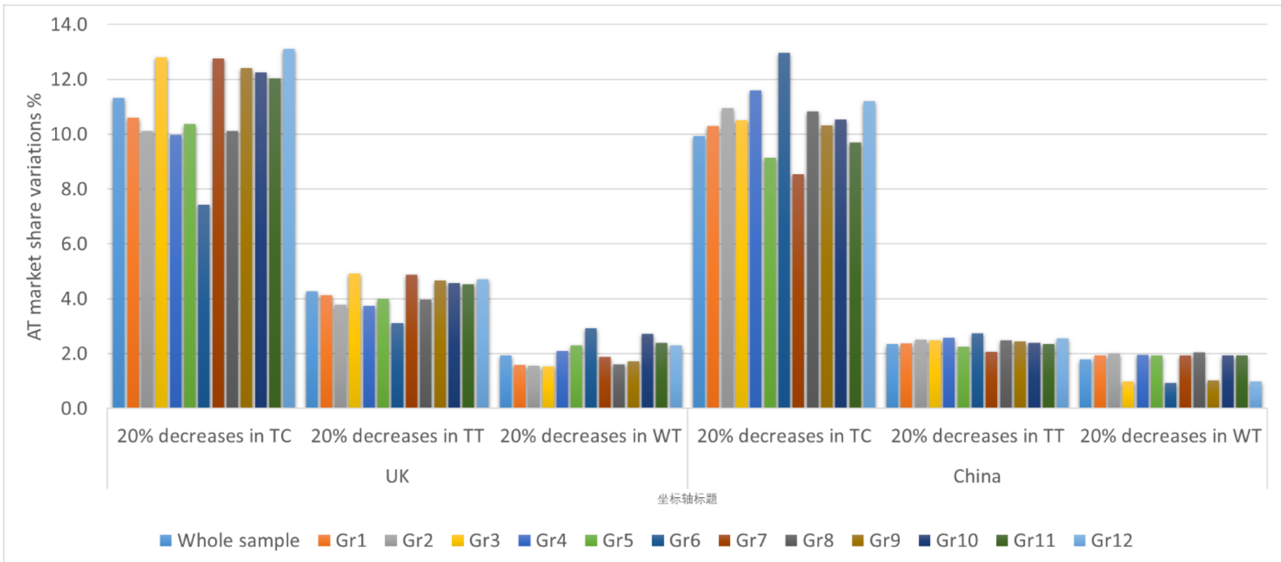


Figure 6.1 AT market share variations for different segments

The impact in the AT market share of a change in travel time and waiting time is much less pronounced and much more homogeneous across groups than the impact of a change in travel cost. We also note that the impact of a change in travel time is always lower in the Chinese than the UK market. The impact of a change in waiting time is also lower in China than in the UK with the exception of groups 1, 2 and 8.

6.5.2 Scenario N2: variations in the in-vehicle features

The second scenario tested assumes the following enhancement of in-vehicle features:

All ATs equipped with the option 'chat with an operator';

All ATs equipped with the option 'change the destination';

This measure can be implemented by AT companies improving in-vehicle functions of all vehicles

Policy	AT market share variation (%)	
	UK	China
In-vehicle characteristics		
All ATs with the option 'Chat during the trip' (from 50% to 100%)	+ 2.2	+4.7
All ATs with the option 'Change the destination' (from 50% to 100%)	+1.7	+1.9

Table 6-19 Market share variations for changes in the in-vehicle characteristics

As shown in **Table 6-19**, when all ATs are equipped with this feature, the AT market shares in the UK and China increase but by a small proportion. The increase in China is twice as bigger as in the UK, bust still below 5%. The impact of equipping all taxis (instead of only 50% of the fleet) with the option 'change the destination' has a positive but small impact in the demand for ATs

6.5.3 Scenario N3: improvements in the customer reviews

All ATs rated with good reviews; this measure can be implemented by AT companies or operators ensuring a good travel experiences for the passengers, e.g. maintaining vehicles clean and tidy, ensuring taxi passengers feel comfortable during the trip, etc.

As shown in **Table 6-20** the implementation of this measure has a similar impact in the UK and in China, around 5% and 6% respectively.

Policy	AT market share variation (%)	
	UK	China
Social conformity		
1. All ATs rated with Good Reviews (from 50% to 100%)	+4.9	+5.8

Table 6-20 Market share variations for changes in the Customers Reviews

6.5.4 Scenario N4: improvements in the latent variables

As discussed in Chorus and Kroesen (2014), the main challenge associated with hybrid choice models is the applicability in policy-making using latent variables results. The latent variables are generally defined as a function of individual-related characteristics and measured via statements. Testing a variation in the value of the latent variables is difficult to justify and controversial because the latent variables do not have a clear unit.

Since LVs are function of SE-related characteristics, it is possible to test trend scenarios, assuming a change in these characteristics in the population. **Table 6-21** show the market share variations under different trend scenarios. We did not test the scenario where the entire population is female or male, for obvious reasons. Each scenario assumes only a change in one characteristic at a time. As we can see, in the majority of the scenarios, the impact of the LV in the percentage of AT adoption is very modest, and it is in general lower in China than in the UK. It is interesting to note that the characteristics that has the highest impact are "having heard of ATs from those who used ATs" and "being familiar with the 5 levels of automation" in the UK, as they dramatically affect choice probability on ATs (+6.56% and 10.81% variation respectively) via Hedonic Motivations.

Trend scenarios	N	AT market share variation (%)					
		Injunctive norms		Hedonic Motivation		Trust	
		UK	China	UK	China	UK	China
Age18_29	1	-	-2.73	2.88	-	-	-1.19
Age30+	2	-	1.42	-0.85	-	-	0.61
Frequent use of taxis (at least once a week)	3	-	2.06	-	2.19	-	-
Infrequent use of taxis (less than once a week)	4	-	-1.52	-	-1.62	-	-
Frequent use of taxis (at least once a month)	5	-	-	3.75	-	2.29	-
Infrequent use of taxis (less than once a month)	6	-	-	-2.04	-	-1.28	-
Familiar with 5 levels of automations	7	3.26	2.30	10.81	3.38	4.06	-
Not Familiar with 5 levels of automation	8	-0.20	-0.52	-0.63	-0.76	-0.25	-
Heard of ATs from those who used ATs	9	-	5.18	6.56	2.89	-	-
Not heard of ATs from those who used ATs	10	-	-0.64	-0.48	-0.35	-	-
Heard of ATs from those who not used ATs	11	-	1.98	-	-	-	-
Not heard of ATs from those who not used ATs	12	-	-0.82	-	-	-	-

Table 6-21 Market share variations for changes in the latent variables

Another method to analyse the impact of a change in the LV consists in computing the marginal rate of substitution (MRS) between variables (one being the LV), defined as the ratio between marginal utilities of the two variables. Following Vij and Walker (2016), given the utility function (all terms are the same as discussed in equation (5.19), The utility is simplified to fit the purpose of this discussion):

$$U_{jqt} = ASC_j + \beta_{TT}TT + \beta_{LV}^{m}LV^{m} + \varepsilon_{jqt} \quad (6.1)$$

we know that a unit change in the individual's preference for a latent variable *m* has the same effect as a change in travel time of AT $\beta_{LV}^{m}/\beta_{TT}$ min; from measurement model:

$$I_{rq}^{m} = \delta_r^{m} + \theta_r^{m}LV_q + \nu_{rq} \quad (6.2)$$

We know that a unit change in individual attitude towards latent variable *m* transforms the outcome to the measurement indicator r by θ_r^{m}.

When it involves latent variables, the MRS take then the form:

$$MRS = \beta_{LV}^{m}/(\theta_r^{m} * \beta_{TT}) \quad (6.3)$$

Table 6-22 reports marginal rates of substitution of the three latent variables with respect to travel time. Using the indicator T4: *I trust I can relax while riding in a fully automated taxi without driver* as an example, if we assume that individual q1 rated one-point higher score than individual q2 on the indicator T4, then results say that, *ceteris paribus*, individual q1 is willing to spend more 3.36 min more to choose AT than individual q2. The MRS values depend also on the marginal utility of travel time that in China is much higher (in absolute terms) than in the UK.

Indicators	MRS Values	
	UK	China
Injunctive Norms		
IN1:People who are important to me (friends, family) would approve of me using a fully automated taxi	-2.87	-12.30
IN2:People who are important to me (friends, family) would think that using a fully automated taxi is not appropriate (R)	-3.92	-24.27
IN3: People who are important to me (friends, family) would think that more people should use fully automated taxis	-3.15	-12.84
Hedonic Motivations		
HM1: I believe using a fully automated taxi will be fun	-6.06	-13.61
HM2: I believe using a fully automated taxi will be pleasant	-6.84	-14.46
HM3: I believe using a fully automated taxi without driver will be boring (R)	-12.19	-26.02
Trust		
T1: Overall, I do not trust fully automated taxis (R)	-7.39	-6.06
T2: I don't trust that fully automated taxis will be adequately supervised (R)	-3.79	-6.39
T3: I don't trust that a computer can drive vehicle without assistance from the driver (R)	-3.41	-6.52
T4: I trust I can relax while riding in a fully automated taxi without driver	-3.36	-12.29

Table 6-22 Marginal rates of substitution of the three latent variables with respect to travel time.

6.6 Summary

In this chapter, the four datasets collected were used to quantify and discuss the impact of the level of service attributes, in-vehicle features, normative conformity, knowledge of AVs and ATs, other effects, and the impact of latent psychological factors on the choices of ATs versus NTs. In doing that, it also discusses cross-national heterogeneity, cross-methodological heterogeneity and cross-experience heterogeneity.

The most important results can be summarised as follows:

- The level of service attributes are significant in all the datasets analysed. UK respondents are less sensitive than Chinese respondents to the travel cost, while they care more about travel time and waiting time (with one exception of waiting time among respondents who are younger than 60 and have other employment status). There seems to be significant cross-national heterogeneity. Cross-methodological heterogeneity was captured in terms of travel cost. Participants joining in survey VR NCL are less price-sensitive. No significant difference in the preferences for travel time was observed among Newcastle participants from the online and the VR survey, with few exceptions. Those who frequently used taxi (at least once a month) in survey-online NCL are more sensitive to waiting time than those in survey-VR NCL. No cross-experience heterogeneity was found regarding level of service variables
- In terms of two in-vehicle variables tested, the option to 'change the destination' and the option to 'chat with an operator', these are highly significant and have a positive similar effects in both UK and China. While in the sample collected in Newcastle, these variables are significant only for some specific categories (male, young and those who do not like to talk with the taxi driver). On the other hand, among the Chinese respondents, those living in NT cities have much lower (around half) MU for the option 'change the destination' than those living in AT cities at 88%.
- Regarding the (objective) normative conformity, the impact of descriptive norms measured by the number of people in the last hour was problematic (significant only in some categories of respondents) in all samples, except in the dataset collected with the immersive VR SC experiment. On the other hand, good reviews from customers was significant and with a positive effect in all samples, i.e. in the UK, in China, in Newcastle with both methods used to elicit preference, and in China whether respondents live in AT or NT cities.

- In terms of knowledge of AVs and ATs, among four types of AV and AT knowledge information measured, it was found that UK individuals who completely lack familiarity with the 5 levels of automation are reluctant to choose ATs. In the sample from China instead, the systematic heterogeneity in preference for ATs only matters among people who have heard of ATs operating in China. But significant heterogeneity was not found among China samples from AT cities and NT cities. Nevertheless, none of these four types of knowledge affected Newcastle residents' preferences for ATs in both *survey-online NCL* and *survey-VR NCL*.
- The three latent psychological constructs tested, injunctive norms, hedonic motivation and trust all have a significantly positive impact on the choice of AT versus NT at 99% in both UK and China sample. The impacts of hedonic motivation and trust in the UK were higher than in China. However, Hedonic motivation is significant only for respondents living in AT cities, while trust is significant only for those living in NT cities, these reflecting the cross-national and cross-experience heterogeneity.
- The chapter has explored in detail the WTPs for all AT characteristics. Results show that British and Chinese respondents have significant different WTP for most of the characteristics. For example, UK respondents are willing to pay on average more on travel time and waiting time. However, respondents from AT cities and NT cities only differ in WTPs to have the option to change the destination and take the taxi with good reviews. Interestingly, differences are also find in the WTP computed using data from Newcastle depending on whether preferences are elicited with an online survey or with the immersive VR environment. However, a bigger sample is needed to confirm these differences.
- Last but not least, AT market share variations were explored from four scenarios. Results indicate that in both the UK and China, a reduction in travel cost has the strongest impact (increase) in the AT market shares but there is substantial heterogeneity among different segments. The impact of a reduction in travel time and waiting time is relatively modest and more homogeneous. The impact of an improvement in in-vehicle features and customer reviews is also relatively modest, while interestingly the biggest increase in the British AT market share is given by the word of mouth ("having heard of ATs from those who used ATs") and "being familiar with the 5 levels of automation", as they dramatically affect choice probability on ATs (6.56% and +10.81% variation respectively) via their impact on the latent psychological construct Hedonic Motivations.

Chapter 7 Conclusion

The success of ATs crucially depends on whether the public adopt and accept ATs. From a research point of view, this prompts several important research questions (RQ), as listed in **Chapter 1**, which refer to identify the determinants of AT adoption (RQ1), to the methodology to measure these factors, within SC experiments (RQ2) and within immersive VR environment (RQ3 and RQ4), to the transferability of the results across nations (RQ5) and the impact of living in a city where the innovation is actually implemented and available to public (RQ6).

In order to answer these questions, the methodology set up consisted of: 1) performing a thorough analysis of the literature and preliminary identification of the key characteristics affecting ATs; 2) running several FGs and pilot tests to get a deeper understanding of those characteristics and highlight new elements; 3) building a stated choice experiment to measure these characteristics along with other important socio-economic, contextual and latent psychological factors affecting adoption of ATs; 4) setting up a novel methodology to implement SC experiments embedded into an immersive VR environment; 4) collecting four sets of data in the UK, China, online and in the VR environment; 5) estimating advanced discrete choice models to measure consumers preferences and willingness to pay for ATs and their characteristics.

Based on the specific research objectives, the main findings and contributions of this PhD thesis are summarised in **Section 7.1**. Some useful insights and recommendations from a methodological and policy point of view are then elaborated in **Section 7.2** and **Section 7.3** respectively. Finally, **Section 7.4** discusses limitations of the current study and some future research suggestions.

7.1 Main findings and contributions

This section summarises and highlights the main findings from this research that allows answering the research questions:

Research questions 1: What are the key factors that ATs ought to have to be competitive with normal taxis with taxi driver?

Results from this research show that the level of services attributes are significant determinants also when it comes to innovative ATs. People still do care about price, travel time and in particular waiting time. This finding is consistent with many previous studies in terms of using AVs or shared AVs (e.g. Krueger *et al.*, 2016; Yap *et al.*, 2016; Kolarova *et al.*, 2019; Kolarova and Cherchi, 2021). This research has contributed to the current state of the research by investigating also the features that are instead specific to ATs, i.e. the in-vehicle characteristics that in an AT cannot be satisfied by the taxi driver. It identified in particular two in-vehicle features that have been found to be relevant: talking with an operator and changing destinations. The significance of these in-vehicle features is different across segments of the population but represents important features to be considered. Another key determinants identified in this research is the customer's knowledge of AVs in general and ATs in particular, while it is less relevant from whom the information is received. The research also contributes to shedding light on the impact of normative conformity on the choice of ATs. We know that conformity plays a critical role in consumers choice (Cialdini and Goldstein, 2004), but, as discussed in the literature review, this has been proved to be difficult to quantify in a way comparable with the level of services attributes. Results from this research confirm this difficulty for the adoption rates, with an exception that will be discussed later, while the customers reviews confirmed the marketing literature like e.g. Chen (2017), Vermeulen and Seegers (2009), and it is always highly significant in this study. Last but not least, the research highlights the importance of three latent constructs in the choice of ATs, namely injunctive norms (which complement the normative conformity measure), trust in innovation, and hedonic motivation (which was proved to be relevant for automation even if the use of AT does not imply giving up driving, because in the alternative NT, users do not drive either).

Research question 2: How can these factors be appropriately measured in a SC survey?
Several results and contributions have been provided. First of all, it sheds light on how to measure in a SC experiment standard level of service attributes. For example, it is suggested that price should be kept fixed and independent on the actual travel time, to avoid uncertainty in the evaluation of the options. Attributes like the type of payments should not be included in the SC tasks, as in real life this choice happens typically after respondents have selected the option to use. Another interesting contribution refers to the selection of the destinations that in normal SC experiment is asked at the beginning of the experiment and kept fixed through all the SC tasks. The research proposes instead to allow respondents to define the destination prior to each choice task. This reduces the efficiency of the SC design, but greatly improves the realism. The use of images in the definition of these attributes has also been explored and proved to enhance realism. The definition of the safety information was also tested thoroughly and FG

and pilot tests suggest that it is better to provide this information before the actual SC, again it proved to help realism. It is critical to mention that most of these results benefitted from the work done with the immersive VR environment embedding into the SC task. This point will be discussed in the next RQs. Another interesting contribution is represented by the design of the consumers rating, where borrowing from the marketing literature and the real life experience, a star system was tested and proved to be particularly significant.

Research question 3: What are the methodological implications of building a traditional SC survey embedded into a VR environment?
The research conducted as part of this thesis show that building a SC embedded into a VR environment is challenging. The contribution of this thesis in relation to this research question consists of the definition of a detailed process proposed for the construction of the SC experiment within an immersive VR environment in the domain of innovative transport choices (i.e. automated taxis versus normal taxis) along with a rigorous protocol for data collection. The research carried out in this thesis provides the first (at our best knowledge) guideline on how to build a SC experiment embedded into a VR environment, and it will be hopefully a reference for future research in the field.

In terms of methodological implications, we found that the following assumptions seem to jeopardise the level of realism in VR experiments:
1) assuming that the trip destination does not change among scenarios,
2) the inclusion of the payment method simultaneously with other attributes such as travel time,
3) presenting the choice tasks continuously without a break.

Research question 4: To which extent does living the choice in an immersive VR environment affect consumers' preferences?
This research question is the most challenging. Since we do not know the ground truth it is not possible to say if the VR environment allows getting 'better' measures of the consumer preferences. Previous studies have not yet tested what we did and it was only confirmed that VR tool provided more robust results. However, the results from this research show that the VR experiment exposed the limit (in terms of realism) of assumptions typically made in the screen-based SC and accepted without questioning, as discussed in the RQ3. Results also show significant differences in the preferences for some attributes, in particular those attributes that rely more on visual information. The most interesting results, however, is that the attribute

measure adoption rate of AT, which is a very problematic attribute (as proved in all researches conducted for other innovation), was significant when measured in the VR environment. Results need to be confirmed by further evidences, but are promising that VR might help getting better results in particular to measure social conformity effects.

Research question 5: Are preferences for ATs measured with SC experiments homogeneous across nations?
This is another challenging question, because to be able to compare samples, these should be perfectly comparable, and this is difficult to achieve. Differently from previous studies (such as Etzioni *et al.*, 2020; Potoglou *et al.*, 2020 etc.), differences in the sample characteristics are carefully controlled, and the general result is that there is significant heterogeneity across nations, at least between UK and China and results cannot be easily transferred. Bigger samples would be probably needed to control better for differences, but this the first comparison across nations with respect to objective characteristics as measured in a SC experiment. Hopefully, these results will serve as basis for more research in this line

Research question 6: Does living in cities where ATs are operating and available to the public have an impact on consumers' preference for AT and its characteristics?
This another interesting research question that is linked to an important area of research, that is the impact that indirect experience has on the adoption of innovations. Indirect experience refers to the experience made by others and communicated to the respondents either via word of mouth, or official information channels or other forms. In this research we extended this analysis by considering the impact of simply living in a city where AT are operating, even if nobody give necessary information to the participants. But just living in the city could represent a cue to respondents. Results confirm that there is an impact, though not as pronounced as we would originally expect. This is the first attempt to test the role of indirect experience on the choice of AT.

7.2 Methodological recommendations

Based on the results from the research carried out in this PhD, the most important methodological recommendation refers to the need to improve realism in the SC experiment. The research carried out with the VR setting allows us to discover limitations in the current way of building and conducting SC experiments that are normally accepted in the current practice. These limitations have been already described, and includes a better use of images, a conscious selection of the attributes, without mixing attributes that are evaluated at different stages of the decision process, a more realistic setting of the decision process across scenarios, allowing for identifying for example different destinations or different trips. What is important to highlight from a methodological point of view is that these recommendations do not apply only if the SC is carried out within the VR setting, but also in traditional online experiments.

Another recommendation refers to the definition of the attributes. Besides attribute level values, due to the difference in the context of AVs, the definition of these attributes and layout of SC survey should be also raised attention. It is essential to define the specific usage context and then determine the specific definition of attributes under this specific context. It is beneficial to examine the confounding effects between the defined attributes by pre-tests. The layout and format of SC experiments can be possibly presented in the way as respondents make a choice in their daily life.

Due to the fact that the price of VR program is still much higher than the traditional online method, the specific reference context i.e. "decision environment" where the SC experiments are embedded is recommended to choose a simple reference context to maximise the ratio between quality and cost of VR experiment. Some new elements (such as shape and interior of AT in this study) were introduced into the VR environment for incorporating new alternative should be paid attention. Due to limited resolution of VR environment, it is required to test the size and colour of texts and images presented and the position where the information would be presented.

On a more empirical level, in terms of implementing VR-based SC, a minimum 4mx4m room is recommended to run VR-based SC experiment for avoiding the problems such as sudden vibration, instantaneous movement from one point to another. Inclusion of auditory isolation within VR program can isolate the respondents from being affected by the noise from the real world (e.g. some voices from lab room). Finally, a rigorous protocol is required to be identified for conducting VR-based SC experiment.

7.3 Policy recommendations

The results of this study allows to provide also some evidence-based policy recommendations for policy makers, AT manufacturers and relevant stakeholders.

In-vehicle features

Our results suggest that more attention should be given to the design of direct communication within the automated vehicles. AT manufacturers should then consider equipping a direct communication with an operator and providing an option to change the destination for passengers when designing and developing automated taxis, as these in-vehicle features are key to attract the demand to ATs. The high impact of the request to communicate with an operator confirms the importance to ensure some "human" connection inside the automated taxis, which is in line with the broader concerns that technology eradicates the human innate tendency to seek connection with others. In terms of type of in-vehicle communication equipment, the recommendation would be to put an interactive 'screen', or even a 'simple button' to open the communication, rather than setting a 'phone app'. Despite the diffusion of 'phone apps', respondents seem not to trust it when it comes to communicate with an operator inside the AT. If possible, it is recommended to install more than one form of communication with an operator, as there seems to be some anxiety about being in a car without a driver. In line with this result, it is also highly recommended to install CCTV cameras in all vehicles.

Another interesting recommendation for AT manufacturer refer to the car models. Almost all AV advertisements show fancy cars, very different from the current models, probably also to highlight the potentiality (e.g. activities that can be performed while riding). However, still from the focus groups we found that respondents have a preference for normal models, like the cars they are used to use every day.

For the taxi operators, based on the results of the Focus Groups, we found that of course, the condition of the vehicle (cleanliness, age, model or brand) is important for an automated as for a normal taxi, but for an AT these are not top priorities for potential customers.

Social conformity

In line with the marketing literature (Cosley *et al.*, 2003; Pang and Lee, 2005; Sparling and Sen, 2011; Chen, 2017), we found that reviews from other customers have a strong impact on the choice of ATs. The use of a 5 star system also proved to be an effective way to report the consumers reviews. The 5 star system is by far the most common format, something consumers are very familiar with, and from a methodological point of view it confirms the importance of using realistic SC scenarios. From a practical point of view, the suggestion for AT operators, is first to pay special attention to maintaining a good reputation among customers and then to use customer ratings to advertise the system, as this confirms to be particularly relevant to boost the demand, as vastly demonstrated in online shopping or hotel booking.

Knowledge of AVs and ATs and latent psychological factors

Our results also highlight the importance of the knowledge about AT in the adoption of this innovation. This effect has been explored for other innovation such as EV (Jensen *et al.*, 2013), but not yet for automated taxis. Our results show that it is not the generic knowledge about autonomous vehicles that matters but the specific knowledge about automated taxis operating in China. This stresses the importance of tailoring the information provided, for example in a marketing campaign. We also note the impact of injunctive norms, hedonic motivation and trust in the adoption of ATs, which depends among other factors also on the experience of the person from whom the information about AT is obtained. This suggests that a word-of-mouth marketing campaign, such as organising activities that encourage interactions among consumers, would be an effective approach for attracting potential AT users. Moreover, the content and form of the information delivered to the public, in accordance with the test for conveying safety information, also matters. This implies that core mainstream media and transport sectors plays a non-negligible role when delivering key information of ATs to the public. It is then advised to pay particular attention to the type of information provided, using objective information preferably from national-level government agencies, as these are valued as the most trustable source of safety information, and to make use of core media channels to inform the public and increase their knowledge on ATs.

7.4 Limitations and future research

This research has provided interesting findings in investigating the critical determinants of preferences for ATs and AT characteristics and in setting up a methodology to measure them. The major limitation is the dimension of the samples. While the online samples have a good size that allowed robust results to be obtained, more data seem to be needed to measure some critical attributes, such as some in-vehicle features, and adoption rates. This is needed in particular to try to elicit preferences for groups of respondents, as it seems that there is quite significant heterogeneity depending on respondents' characteristics and experience with technology. The comparison between nations also would benefit from bigger samples, in order to have a better control of the sample characteristics.

A bigger sample is definitely needed from the immersive VR experiment, but this is extremely difficult to obtain. It requires significant resources in terms of time and money that often are not available for PhD research.

This PhD has provided a basis for future research on the field of AT and in particular using VR environment. The results found have challenged some of the typical assumptions made in the SC experiments and suggested changes in how to build and run these experiments. More research is needed to confirm these findings. Major areas for future research are in the definition and measure of the social conformity attributes, in particular the normative conformity. Some studies suggested that the facial expression is relevant in conforming to other behaviours (Pasupathi, 1999; Procházka *et al.*, 2016; Chen and Wyer Jr, 2020). This is an area where VR could be potentially very useful, but this would stretch the use of the VR to the limits. Another area of research is getting respondents to pay the consequence of their choice (Herriges *et al.*, 2010; Moser *et al.*, 2014; Krčál *et al.*, 2019).

In terms of further VR-based SC experiments, suggestions for future research are as follows:

- Propose an approach to effectively reduce the VR experiment time for efficiently collecting data. Due to the rigorous protocol proposed and the nature of VR experiments, the total experiment period is much longer than in the traditional online SC survey. Data collection efficiency in the VR environment is lower than using online SC survey. It would be worthy exploring an effective approach to increase efficiency in collecting valid data using VR experiments.
- Testing 'visual' attributes relevant to the choice of ATs. One of the benefits of the VR experiment is the ability to represent complex visualisations and dynamic variables. Some visual attributes such as weather conditions defined as the fixed variable (i.e. always good weather) in this study can be explored in further studies.
- Examining the role of in-trip experience. Due to time and cost limitations, in this study, special effort was put in defining the pre-trip experience. However, the role of in-trip experience is also important in the choice of ATs. A further study can investigate and explore preferences for ATs after a full experience (including both pre-trip and in-trip experience in VR).

In terms of further attributes relevant for ATs, suggestions for future research are as follows:

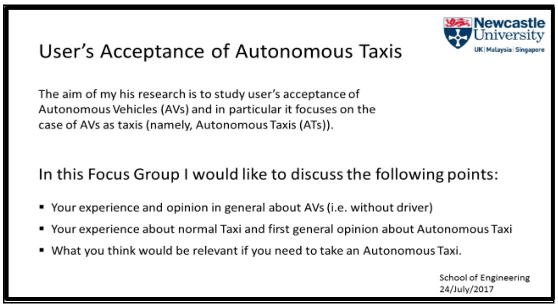
- Investigating the impacts of other latent psychological factors; as mentioned in **Chapter 2**, the underlying psychological theories, TAM, TPB and UTAUT, have been extensively used to investigate the impacts of a large number of psychological constructs on the intention to use AVs. It is also essential and meaningful to build a more comprehensive model accounting for all psychological factors based on these underlying psychological theories and measure these effects jointly with microeconomic effects. Nevertheless, as the mentioned by Thorhauge *et al.* (2019), this was proved to be challenging and required a huge amount of data;
- Investigating the impacts of pre-defined information e.g. safety information. Although in this study some pre-defined information was provided before the formal SC experiments, but all these information was kept same for all respondents. It would be also interesting to understand the role of this information in choice of ATs. One option could be to include the accident rates as attribute in the SC experiment across different scenarios like an attribute defined in the SC experiment. However, this could be problematic in terms of realism, because respondents might not believe safety levels can change in a short time as they answered six choice tasks in sequence. A better options consists in keeping it as pre-information but assigning different groups of people to different safety levels. This would however requires recruiting more participants for the modelling.
- Investigating the impact of payment method. To ensure high level of realism, the variable 'payment method' was included after respondents have chosen the type of taxi, which is consistent with real-life situations. To investigate the impact of this variable, an advanced choice model is required to develop.
- Investigating the impact of other social conformity in the choice of ATs. Different types of passengers queuing for taxis can be also used for measuring impact of other social conformity, as respondents might perceive they belong to different groups. Initial pre-tests were run to test the impact of this type of social conformity, however, defining the different type of people in the context of using taxis is challenging. Nevertheless, it is still an interesting direction to explore.
- Investigating the revealed preferences of the current ATs users, i.e. those who have already used ATs. This was attempted in this dissertation, but the number of AT users recruited was too small to examine their preferences. A possible method for recruiting AT users would be to contact participants at pick-up or drop-off points of AT services.

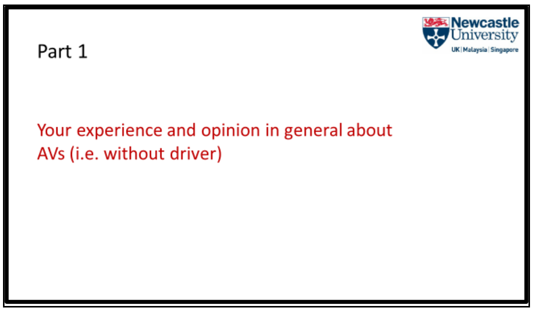
Given the increasing popularity of AT services in China, the group of AT users might become large enough, it is certainly interesting to interview them about their preferences or attitudes towards current AT services using RP survey or to investigate their preferences for further upgraded AT services using SP survey.

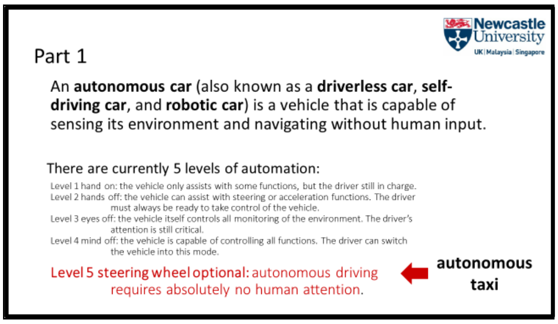
Appendices

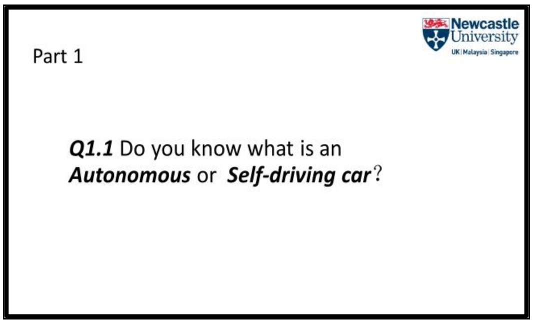
Appendix 1 – Focus Groups

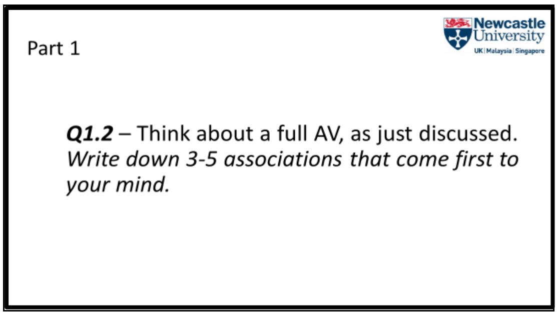
Slides prepared to guide the FG discussion[22].

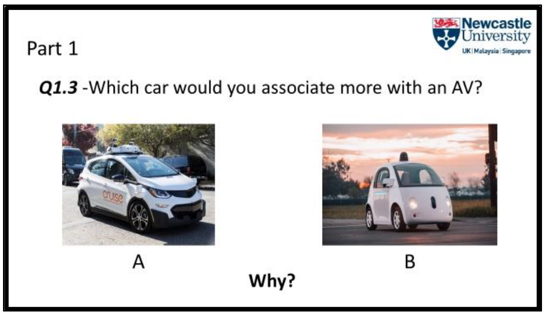












[22] Focus Groups were run in collaboration with the team of the Veronica project, funded by ESRC, UK. https://www.ncl.ac.uk/engineering/research/research-case-studies/future-mobility/veronica/

Part 2

Your experience about normal Taxi and first general opinion about AT

Part 2

Q2.1 -Think about when you take a (normal) taxi. Imagine you are at a taxi rank in Newcastle, *please write down 3-5 associations that come first to your mind.*

Part 2

Q2.2 - Now, imagine the same situation (i.e. you are at the taxi rank in Newcastle) but you have the option to take a taxi that is an AV.

If you have to chose between a normal taxi and an autonomous taxi, what would you look for?

Take a piece of paper and jot down 3-5 things that are important to you If you have to make this choice.

Part 2

Q2.3 – These are 2 possible interior of ATs

If you have the choice between these 2 AT (A and B), which one would choose?

A

B

Part 2

Q2.4 – Thinking about taking an AT

Where would you prefer to sit?

In the front sits (left or right?)

A1

A2

In the back sits

B

Why?

Part 3

What you think would be relevant if you need to take an Autonomous Taxi

Part 3

Q3.1 - Imagine you are at a taxi rank in Newcastle, and you can choose between a normal Taxi and an AT.

If you had to pick only one factor that is most important to you, what would it be? *You can pick something that you mentioned or something that was said by others.*

1. Brand of the taxi
2. Cleanliness
3. Car condition
4. Car model
5. Safety
6. Control of vehicle
7. Ease of using new technology
8. Privacy
9. Travel cost
10. Payment method
11. Waiting time
12. Travel time

Part 3

Let's analyse some of these factors in more detail.

Part 3

Ease of using new technology

Q3.2: How do you feel using technology?

Do you have a smart phone?

Have you every used the smart phone for one of the following activities?

- to search information online (google or other search engine)
- To check location of places in google map
- To book a taxi online

Part 3

Ease of using new technology

Q3.3: Think about taking a taxi

Typically we tell the taxi driver where we wish to go …

How would you feel if you can only use technology?

Part 3

Ease of using new technology

Q3.4: Think about taking a taxi

How typically do you get the information about the cost of your trip by taxi?

- Ask the taxi driver before taking the taxi
- Just pay what the taxi driver tell me when I arrive
- Check online when I book
- Other ….

Part 3

Payment method

Q3.5: Since the AT has no driver, you cannot pay to the taxi driver.

Which of the following payment methods would you prefer to use?

A) Phone app

B) Credit/debit card

C) Cash

Part 3

Waiting time

Q3.6: Thinking about the time you need to wait to take a taxi.

Which of the following information are more reliable for you?

- The waiting time displayed in the information panel
- The number of people queuing
- How fast the queue move
- Your previous experience
- The previous experience of friends/family/others you trust

Part 3

Communicating while on board

Q3.7: Since the AT has no driver, you have nobody in the car to talk to in case of need.

Which of the following forms would you prefer to communicate with the AT operator in case of need?

A) Phone app

B) Screen or button

C) Voice control

Part 3

Let's now talk about Safety of AV.

Part 3

Q3.8 *Write down 3-5 associations that come first to your mind when you think about safety of ATs.*

Part 3

Q3.9 Imagine you are at a taxi rank, ready to take an autonomous taxi.

Among the following safety issues, which one would you concern you the most? - *Pick only one* -

- Possibility of crashes with other cars
- The severity of crash with other cars
- Possibility of crashes with pedestrians
- The severity of crash with pedestrians
- Personal safety: others can take advantage of you being alone inside the car
- Other factors
- Nothing, no concerns

Part 3

Q3.10 Imagine you are at a taxi rank, ready to take an autonomous taxi.

How would like to receive information about safety?

- From the sign board before you take the taxi, along with the cost info etc.
- From your phone before you take the taxi
- Listening the experience of other customers (you do not know them) who have already used ATs
- Listening the experience from friends/family/others you trust who have already used ATs
- I only trust my own experience
- Other ...

Part 3

Q3.11 Imagine you are at a taxi rank, and there is a queue of people waiting.

Write down 3-5 associations that come first to your mind when you think about the people queuing for a taxi.

Part 3

Q3.12 Imagine you are at a taxi rank, queuing to take an autonomous taxi.

Which of the following people would you notice in the queue?

- People in the same range of age as you
- Family with young children
- Professionals, clearly going to work
- Youngers, clearly highly prone to technology
- None of them
- Other??

Part 3

Q3.13 Imagine you are at a taxi rank, queuing to take an autonomous taxi.

If you have noticed any of these people in the queue ...

- People in the same range of age as you
- Family with young children
- Professionals, clearly going to work
- Youngers, clearly highly prone to technology
- None of them
- Other??

Would this makes you more/less willing to take an AT or indifferent?

Part 3

1) Do you know VR experiment?

2) How would you feel in participating in an experiment where you are asked to use VR?

Appendix 2 – Pilot Tests

The pilot tests are an iterative process consisting in testing the overall survey and in particular the experimental design built, estimating discrete choice, evaluating the statistical and microeconomic significance of the results, adjusting the experimental design and testing it again in a new pilot. In order to finalise the survey instrument several pilot tests were run to test the content and structure of survey, the pre-information provided before the SC experiments (including safety, privacy, etc.), the attributes in the SC tasks (definition and levels) and the format or layout of the tasks.

Pilot Test N1 (UK)

For the first pilot survey, a standard SC questionnaire was designed in the platform 'Google Forms'. The questionnaire was kept initially quite simple. Other than the SC experiment, only a few socio-economic characteristics were asked at the beginning of the questionnaire. The SC experiment included 3 attributes at 3 levels (waiting time, travel cost, number of people ahead) and 2 attributes at 2 levels (payment method, when to pay). Since priors were not available, an orthogonal fractional design with only main effects was built that returned 16 choice tasks, divided in 2 blocks of 8 tasks each. The pilot test N1 ran between 29 Nov. 2018 and 1 Dec. 2019 mainly among Newcastle University students and staff. The sample was composed of 12 males and 11 females and the majority (17) were British. They did not receive incentives. 184 pseudo-observations were collected. According to initial MNL estimation result, the variable 'number of people ahead' had a negative sign and the variable 'when to pay' was not significant, which is not in line with the expected results. These two variables were then re-defined and tested in the next pilot.

Pilot Test N2 (UK)

Pilot Test N2 aimed to correct the problematic attributes reported in the Pilot Test N1, but also to test a more advanced structure of the survey. It was greatly improved in the following parts:
Section 1: Minor changes were made in the format of the introduction to the overall survey. Information regarding a recent trip experience by taxi was added at the beginning of questionnaire and was more detailed compared to first online pilot. The few about socio-economic characteristics were moved at the end of the survey.

Section 2: A major work was carried out to improve the SC design and in particular to increase realism, which is key for the quality of the estimation results. A better introduction with images to the SC was added and the format of SC experiment was greatly changed. The destination list was introduced, divided into three segments (5km, 10km and 15km) which covers most of the destinations by from Newcastle city centre. In addition, the layout of the choice task was greatly changed, in particular images were added in the description of the two taxis. An example of SC choice task is shown in **Figure 4.6**. The variable 'you will be the Xth passenger since the last 7 days' was used to replace 'the number of people ahead in queue', which were not significant according to results of the Pilot Test N1.

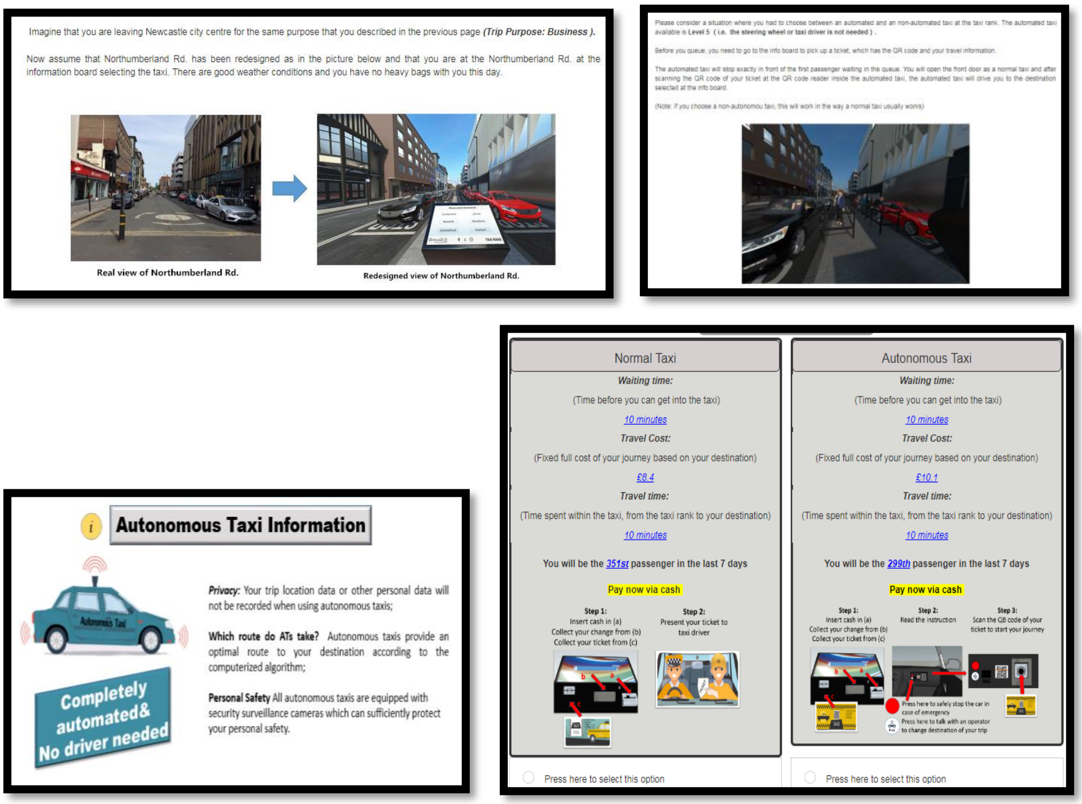


Figure 2.A an example of choice task in pilot test N2

Section 3: Specific questions were set up aiming to check the clarity of the format and content of each single part of the choice task. In addition, a set of questions were defined to investigate respondents' opinions, perception and level of understanding of vehicle safety, personal safety and privacy issues. This information is key for ATs, and more importantly if respondents have different information or perception about the level of vehicle safety, personal safety and privacy issues, this might affect the preference for ATs estimated in the model. These effects should be then controlled. This is the reason why a significant research effort was devoted to identify and test the best way to convey this information.

The survey in the second pilot was built using SurveyEngine. The Pilot Test N2 was run among friends and friends of friends, who live or lived in Newcastle and have used taxis before. They were invited to complete the survey by sending anonymous links. They did not receive incentives. Respondent answered 8 choice tasks each, therefore, 176 pseudo-observations were collected. The main objective of this second pilot was to test the format and how to appropriately present safety information and privacy information. Mixed logit models with panel effects were estimated. Results showed that all the attributes were significant at 95%, except the 'number of people in the last 7 days' and 'when to pay'. Previous studies confirmed that the first attribute (measuring normative conformity) it is difficult to test and often has low significance.

The other aim of the Pilot Test N2 pilot was to determine how to present the information about taxi safety, personal safety and privacy in the SC. **Table 2-A** reports the aspects tested. Respondents were asked 'How concerned are you about the following safety issues if they occur?' and the answers ranged from 'not concerned at all' to 'extremely concerned'.

Q1 The taxis you are travelling in has an accident involves:
Only taxi passengers, only other car users;
Only vulnerable road users (i.e. pedestrian, pedal cyclist and motorcyclist);
Q2 If the taxi you are travelling in has an accident that involved
Only fatalities (e.g. that caused death less than 30 days after the accident)
Both fatalities and seriously injured (e.g. for which a person in detained in hospital as an 'in-patient')
Q3 If you would like to receive road safety information, which of following road safety institutions or organisations would you like to get the information from?
National-level government agency or institutions (e.g Department for Transport)
Regional-level or local level government agency or institution (e.g. Newcastle Citiy Council, Northumbria Police)
University research (e.g. Newcastle University)
Non-profit transport-related organisations (Chartered Institutions of Highways & Transportation (CIHT))
They are all the same
Other, please specify
None, I do not care

Table 2-A Testing questions of taxi safety, personal safety and privacy issues

Q4 If you were to read about the road safety information from the media, which of the following sources would you like to receive it from?
National-level media (e.g. BBC, Guardian)
Regional-level media (e.g. Heart North East, Northern Echo)
Local-level media (e.g. Newcastle Chronicle)
They are all the same
Other, please specify
None, I do not care
Q5 Imagine you are at a taxi rank, ready to take an automated taxi
How would you like to receive the safety information?
From the sign board you take the taxi , along with the cost info etc.
Listening the experience of other customers (you do not know them) who have already used automated taxis
Listening the experience from friends/family/others you trust who have already used automated taxis
I only trust my own experience
Other, please specify
Q6 which of following expressions about taxi safety is easier to understand for you?
Number of fatalities by automated taxis per year (e.g. 5 fatalities per year)
Number of fatalities caused by the automated taxis per million km travelled (e.g. average 4.6 fatalities caused by million km travelled)
Number of fatalities per 100,000 registered automated taxis (e.g. average 2 fatalities per 100,000 registered automated taxis.)
Q7 In terms of personal safety, if you think of yourself taking an automated taxi, would you worry about your personal safety because there is no driver in the taxi?
Yes
No
If yes, what type of personal safety do you care about?
Risk of personal crime (e.g. being physically assaulted/molested)
Risk of property crime (e.g. suffering a robbery)
Other, please specify
Q8 Are you concerned about privacy issues (e.g. travel-location data misuse and crash data misuse) if you used the automated taxis?
Yes
No

Table 2-A (Cont'd) Testing questions of taxi safety, personal safety and privacy issues

The following considerations were drawn from the test:

- Regarding how to measure taxi safety (Q1 and Q2), people are more concerned about those within the taxi than others outside. Among the 'others' respondents are more concerned about both fatalities and seriously injured, as expected. In the next pilot, this question needs to be corrected, considering the low probability to have a fatality, and the higher chance to be injured.
- Regarding the way to present taxi safety in the SC experiment (Q3, Q4 and Q5), 56% of the respondents would like to receive road safety information from Regional-level or

local-level government agency or institution (e.g. Newcastle City Council, Northumbria Police), followed by National-level government agency or institution (e.g. Department for Transport) and University research (e.g. Newcastle University) both summing 39% of preferences. The far majority of the respondents (61%) would like to read these information from national-level media (e.g. BBC, Guardian), followed by the local-level media (e.g. Newcastle Chronicle) (26%). There is instead no clear preference about where to receive this information: From a 'sign board before you take the taxi' and 'Listening the experience from friends/family/others you trust who have already used autonomous taxis' got both 35% preferences; 'Listening the experience of other customers' collected 22% preferences. Since the same experiment will be carried out also with the VR, it could be easy to present this safety info in the sign board, rather than simulate the friends/family who respondents trust.

- Regarding the clarity of the safety expressions (Q6), it was asked to indicate which expression was easier to understand and respondents split almost equally among all options provided. However, this does not tell us if the options not chosen were difficult to understand, only that these were less easy. It might be better to ask to indicate which expressions are NOT clear or they DO not understand.
- Regarding personal safety (Q7 and Q8), almost all respondents (83%) are worried about their personal safety since there is no driver when they take the taxi. Among them, 74% really worry about the risk of personal crime (e.g. being physically assaulted/molested) and 58% worry about privacy issue Therefore, it is necessary to include clear information about personal safety into the SC experiment. However, since it is not the main objective of this PhD thesis, and it is less realistic to present a SC task that includes the 'privacy' attribute, privacy information was given as 'pre-information' before presenting the choice tasks.

Pilot Test N3 (UK)

The third online pilot questionnaire was performed among real taxi customers who have used taxi at least once in the last year. Based on the results form Pilot N2, the following three sections were changed:

Section 1: The format of the introduction to the overall survey and the structure of the questions regarding a recent trip experience by taxi were slightly modified.

Section 2: The introduction to the SC using images and the format of the SC experiment were improved. The layout of the choice tasks was greatly changed. Images were simplified to make it easy for respondents to understand the information contained. The attribute 'customers in past week at this taxi rank' was used to replace the 'You will be the _th passenger in the last 7 days'. And 'What you can do when travelling' was used to replace the 'Use of AT' in Pilot Test N2. This variable refers to the safety/communication options available inside the AT. Although this refers to something that happens within the vehicle while travelling, it is also a piece of necessary information that passengers should have before they board the taxi. In-vehicle features might highly influence the potential passenger's choice of ATs, and this knowledge is very valuable for AT manufacturers and operators for developing ATs. Finally using the coefficients estimated in the second pilot as prior parameters, an efficient design was built and tested in the Pilot Test N3.

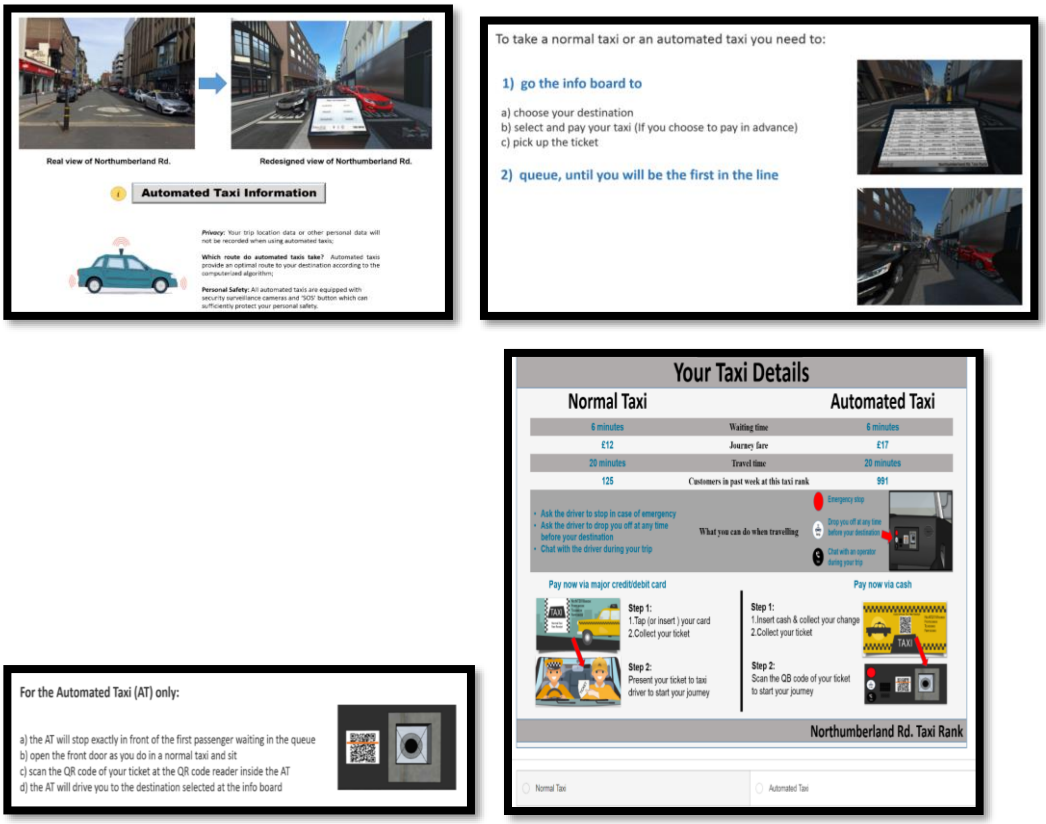


Figure 2.B an example of standard SC choice task in Pilot Test N3

Section 3: Several questions related to respondents' opinions, perceptions, level of understanding of vehicle safety, personal safety and privacy issues were retained, because some aspects about safety were still not clarified and the sample was different from the one used in the Pilot N2 and more heterogeneous. Some questions about travel purpose, travel frequency by taxi and respondent's income were also added.

The online pilot N3 was tested with a sample of 31 real taxi users, mixed logit models with panel effect were estimated using all 279 observations (each respondent answered 9 choice tasks). Results showed that 'travel cost' and 'waiting time' had the expected sign, and are significant at 95%. Four of the six attributes tested (i.e. what you can do when travelling, descriptive norms, payment method and when to pay) were not significant and in the model without interaction effects, the attributes 'Paying by credit card', 'customer in past week at this taxi rank', 'what you can do when travelling' had a counter intuitive sign.

The other aim of this pilot N3 was to determine how to present the information about taxi safety, which was not solved in the pilot N2. The results about personal safety and privacy from the pilot N3 are similar to those from pilot test N2, i.e. respondents do care about these two factors. Therefore, the summary is only related to the taxi safety.

- The far majority of the respondents thought that the 'fatality' is the least acceptable casualty severity for them and that accidents involving vulnerable road users (e.g. cyclists or pedestrians) were the least acceptable. Results also suggest that providing "mild" information about safety would be better and plausible. It was then decided to use the information 'only X minor casualties, not required hospitalisation in the last year or X accidents recorded' before the SC experiment. This information would not change across respondents.
- Regarding the way to present the taxi safety, the majority of the respondents would like to receive road safety information from national level government agency, followed by regional-level government agency (in the pilot test N2 most of the respondents preferred regional-level government agency. This problem might be caused by the convenience sample use (e.g. friends)). In pilot N3, national-level government agency was obviously chose by most of the respondents. It would be better to use national-level government agency like 'Department for Transport' as the source of road safety information. Similarly to pilot N2, the far majority of respondents chose the national-level media (e.g. BBC or Guardian) to report safety information about automated vehicles.
- Regarding the definition of safety, the far majority of the respondents reported that it would be easy way for them to understand taxi safety if expressed in terms of 'number of fatalities caused by automated taxis per year (e.g. 5 fatalities per year)'. We reviewed the literature on the willingness to pay to reduce accidents (e.g Rizzi and Ortúzar 2003,

2006; González *et al.* 2018) and the information from Vehicle Safety Report for current AV tests (e.g. Tesla Vehicle Safety Report https://www.tesla.com/VehicleSafetyReport), it was finally determined to present the safety information as 'accidents recorded per x miles travelled by ATs compared to the NTs.'

In summary, in terms the vast majority of respondents indicated the national-level media as the most likely channel to get safety information about ATs and the national-level government agency as the most trustable source of safety information. In terms of safety measure, it was finally decided to present the safety information in terms of "accidents recorded per x miles or kilometres travelled by ATs compared to the NTs.

Pilot Test N4 (UK)

The main aim of the Pilot Test N4 was to test the attributes and the format of the choice tasks. Compared to Pilot Test N3, he following major changes were implemented:

- For the pre-information, **Figure 2-C a)** shows the necessary safety information, and privacy and routing information. The layout of taxi safety was adjusted like the BBC news, which improve the realism of this information. The accident rate for ATs reported was assumed to be half the accidents in normal taxis. This is just an assumed value. While **Figure 2-C b)** shows the information presented about the specific features that respondent would find inside the AT, as well as the information about 'how to use ATs'. In order to further simplify the attribute 'what you can do', button features were moved into the pre-information, so that respondents do not need to read this information again in each choice task. It should be noted that the information reported in **Figure 2-C b)** is meant to inform respondents in advance about some features that are presented in all automated taxis (24h security surveillance camera and a 'SOS' button, which respondents in the FG and pilot tests considered necessary) plus some optional features that will be present in some of the AT, not all (this reflects what respondents will find in the different SC scenarios)..
- In the choice tasks, travel time was changed as 3-level attribute, instead of a fixed information provided in the choice task and two new dummy variables 'type of vehicle (petrol or electric)' and 'customer rating in the last week' were introduced. A discussion about these variables is reported in the main body of the PhD thesis. The payment method was removed from the experimental design and asked as a separate information after respondents made their choice of taxi.

The online pilot N4 was tested with a sample of 35 taxi users. Each respondent was presented with 9 choice tasks. ML models with panel effect were estimated using all 315 observations. The attribute 'Reviews' as well as the level of services attributes (travel cost, travel time and waiting time) were all significant at 95%. The attributes i.e. 'propulsion type of vehicles', 'chat during the trip_AT', 'change the destination' and 'number of today's customer' were not significant at 95% and without interaction effects all these attributes showed a counter intuitive sign.

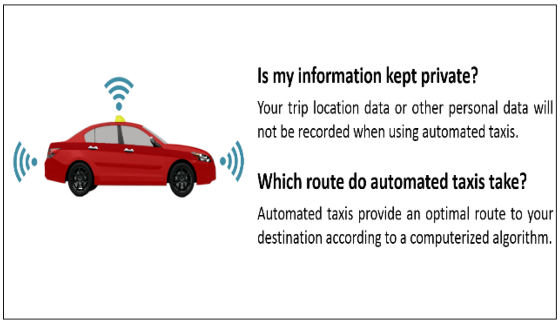


Figure 2-C a) Safety information Privacy and route information

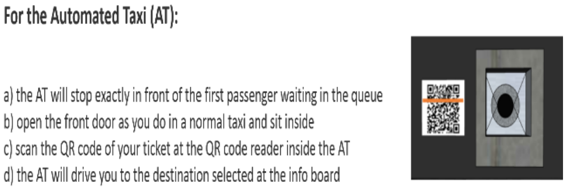


Figure 2-C b) What to do within the ATs How to use ATs

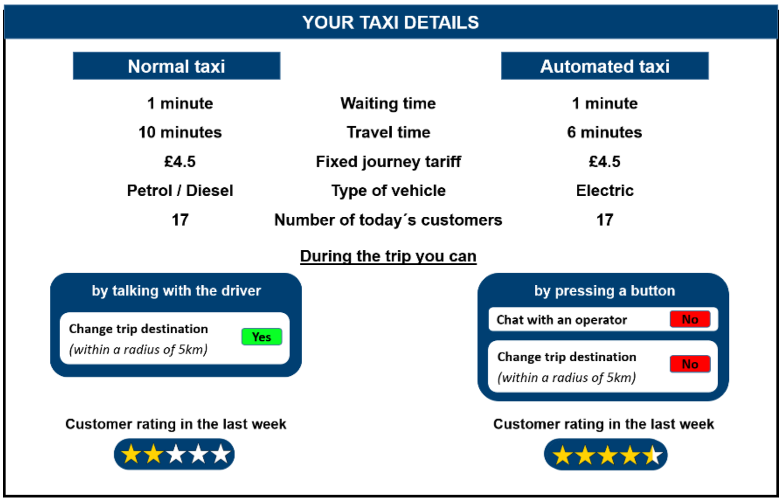


Figure 2-C c) an example of standard SC choice task in Pilot Test N4

Pilot Test N5 (China)

The Pilot N5 was used to test the survey in the Chinese context, after adjusting the following relevant information to the Chinese context. These adjustments are described in the main body of the PhD thesis.

For the SC experiment, an orthogonal design was initially used, as coefficients were not available for the Chinese context. The online pilot N5 was tested with a sample of 48 real taxi users, mixed logit models accounting for panel effect were estimated using 432 observations (each respondent answered 9 choice tasks). Results showed that travel cost', 'waiting time' and 'change the destination' were significant at 95%. The other four attributes tested (i.e. 'chat during the trip_AT', 'review' and 'number of today's customer' and 'travel time') were not significant at 95% but had the right sign.

Pilot Test N6 (China)

Differently from pilot N5, in this pilot survey, an heterogeneous efficient design was used. This final pilot test was used to check if the content or structure of survey was clear and understandable after translating into Chinese. Results showed that all coefficients had the right sign. 'travel cost', 'waiting time' and 'good review' were significant at 95%. The other four attributes tested were not significant at 95%. None of the respondents reported that some parts was unclear or difficult to understand.

Appendix 3 – VR Urban Layout

This appendix describes the work done to modify the urban layout of Northumberland Road, to make the inclusion of automated taxis in the taxi rank realistic and functional to the purpose of running a SC experiment embedded into a VR environment. **Figure 3.A** represents the location choice in Google map, while **Figures 3.B** show the street view, from Google map.

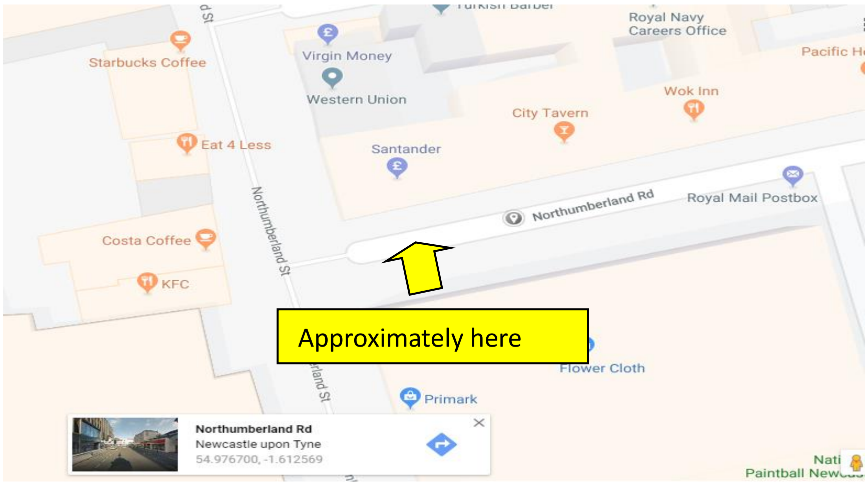


Figure 3.A Location of research area (the taxi rank at Northumberland Rd.)

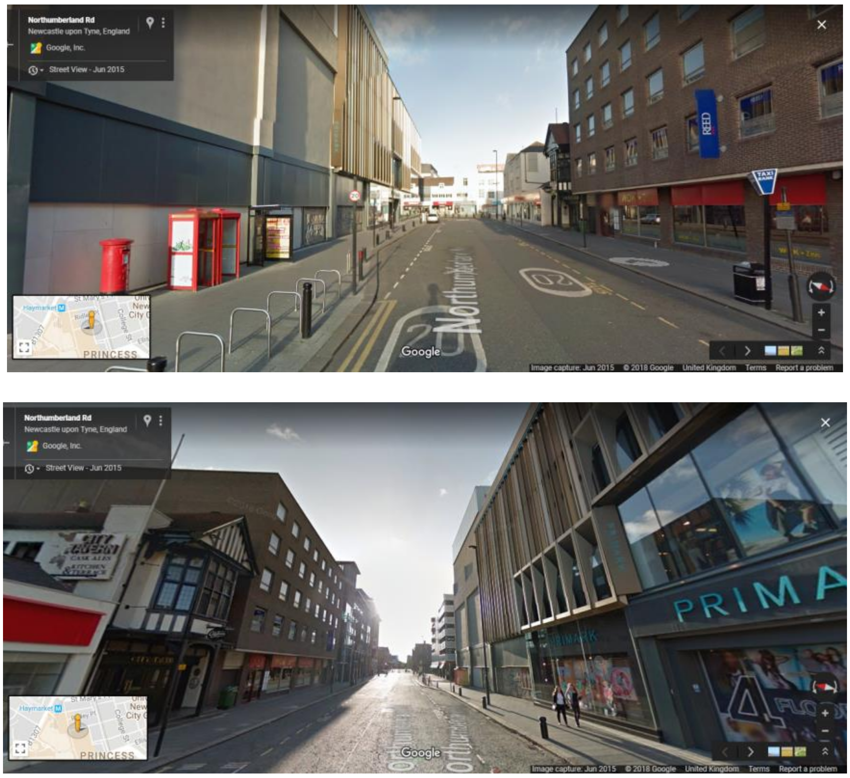


Figure 3.B Street view of Northumberland Rd in Google Maps

The taxi rank in Northumberland Rd. is currently located at one side of the street and taxis queue there. The street has 2 lanes, one for each direction, as normal cars also are allowed to drive in mainly to drop on – drop off passengers. In the SC experiment respondents have the choice of 2 types of taxi (normal and autonomous), the urban layout needs then to be modified to accommodate both options.

The re-design of the area was performed considering also the experimental design and the attributes that needed to be tested, as described in **Chapter 4** of the thesis. Keeping this in mind, several options were analysed. All options included 2 separated lines of taxis, as this facilitates the experiment where respondents have to clearly distinguish the characteristics of each taxi line and choose one of the 2 taxis. This unfortunately requires some more street space, and does not allow to have 2 lanes for the movement of the cars. With AVs, being these without drivers, it was important to avoid the problem of crossing the street in front of the AV, as this could have added safety concerns that should be controlled in the experiment. **Figures 3.C** show some of the options analysed.

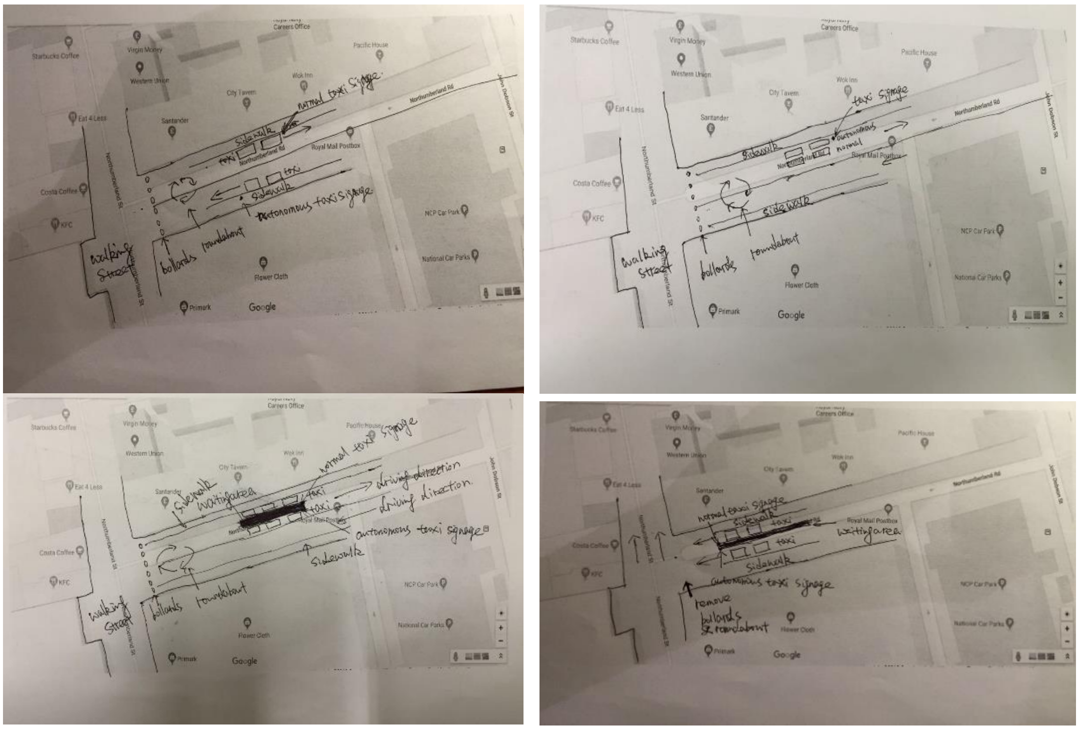


Figure 3.C Taxi rank design options

All the options were sketched and discussed with Animmersion[23]. Pros and cons were carefully evaluated and finally, it was decided to locate the taxi rank as sketched in **Figure 3.D**.

[23] Animmersion is a leading supplier of digital visualization tools specialized in VR https://www.animmersion.co.uk/

As it can be seen:

1. The waiting area of taxi users is set on the right vehicle lane.
2. Autonomous taxi and normal taxi carry passengers are located on the sides of the waiting area. The left side of the waiting area is for normal taxi and the right side for autonomous taxi. Ideally, this needs to be randomized among the tasks presented to the respondents, however, changing the position of the cars in the VR environment has a cost, and we might decide not to implement this feature.
3. The black line in the drawing is a kind of small island. The island needs to be wide enough to allow two lines of passengers, in order to be easy to identify those who chose AV and are waiting for it, and those who chose the normal taxi and are waiting for it.
4. Information about the taxis (costs, time etc,) will be located on the small island, where there is the arrow to signage.
5. People waiting for the taxi will be standing in the small island. This option allows passengers to be equally distant from both regular and AV taxis, so they can see the front of both type of taxis.
6. Respondents can start the experiment standing on the right sidewalk or close to the roundabout and be asked to go and take the taxi (we can observe if for example they choose to cross in front of the normal taxi or in front of the autonomous taxi)

The approximate width and length of Northumberland Rd (**Figure 3.D**) is 15 meters and 140 meters respectively according to Google measurement. To allow enough space for the above design, the path walks were slightly narrowed.

Figure 3.D shows the same design as it was implemented in the VR environment. Several checks were performed before we arrived to this final version **Figure 3.E**. The detailed checks were mainly related to ensure that the scene was clear and all the elements that the respondents will need to evaluate in the experiment were clearly distinguishable. In particular, I checked the signboard (i.e. direction, distance away from taxi queue and people queue, height, number and type), the waiting island (i.e. width of island), the taxi (i.e. colour and model) and the position of the people in the queue. For example, initially the waiting island was too narrow and people were standing too close one to another. Respondents could not easily see how many and what type people were queuing. Therefore, I suggest to enlarge the waiting island a bit to allow people in the queue to be seen and to give the space for respondents.

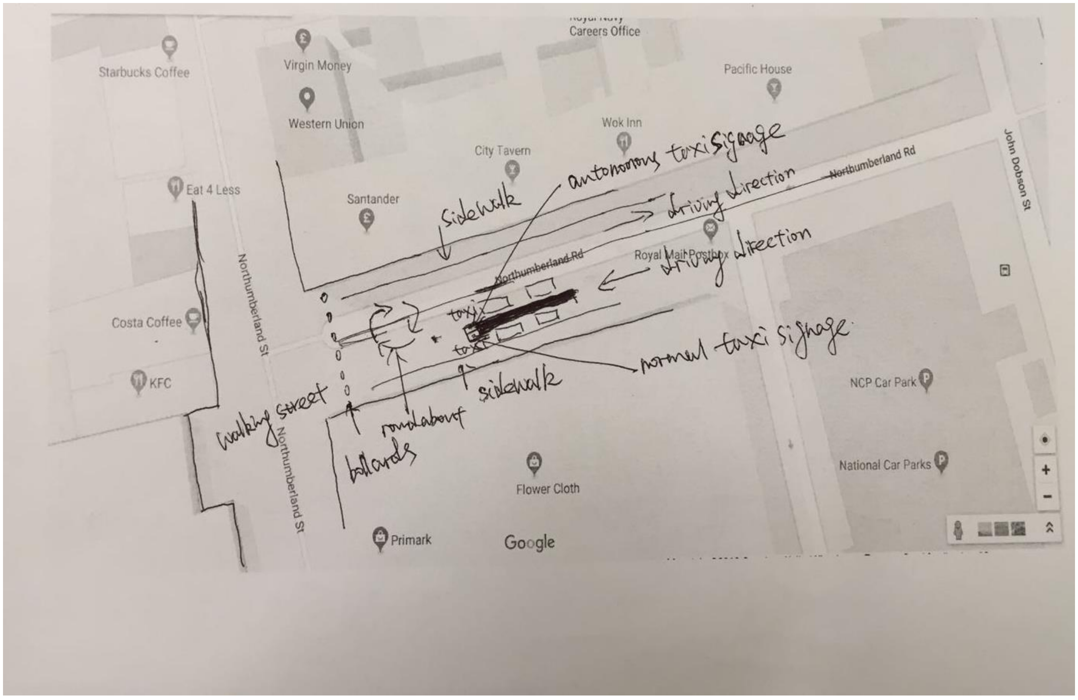


Figure 3.D Taxi rank design

Figure 3.E Final VR scene design

Appendix 4 – Descriptive Data Analysis

Socio-demographic characteristics		VR- SC N (%)	Online SC N (%)
Total sample		40	139
Gender	Female	15 (37.5)	79 (56.8)
	Male	25 (62.5)	60 (43.2)
Age	Below 30	15 (37.5)	40 (28.8)
	30-39	13 (32.5)	35 (25.2)
	40-49	5 (12.5)	31 (22.3)
	50-59	4 (10)	20 (14.4)
	60 or above	3 (7.5)	13 (9.4)
Education level	Less than GCSE	0 (0.0)	2 (1.4)
	GCSE or equivalent	2 (5.0)	13 (9.4)
	A level or equivalent	4 (10.0)	45 (32.4)
	Bachelor degree	10 (25.0)	48(34.5)
	Master degree	17 (42.5)	28 (20.1)
	Doctorate degree	7 (17.5)	3 (2.2)
Current work status	Employed full-time	14 (35.0)	76 (54.7)
	Employed part-time	4 (10.0)	23 (16.5)
	Self-employed	1 (2.5)	9 (6.5)
	Jobless	0 (0.0)	2 (1.4)
	Students	18 (45.0)	15 (10.8)
	Retired	3 (7.5)	7 (5.0)
	Others	0(0.0)	7 (5.0)
Personal monthly disposable income	Less than £500	7 (17.5)	38 (27.3)
	£501-£1500	20 (50.0)	61 (43.9)
	£1501-£ 2500	11 (27.5)	22 (15.8)
	£2501-£ 3500	0 (0.0)	9 (6.5)
	£3501-£ 4500	1 (2.5)	2 (1.4)
	More than £4500	0 (0.0)	1 (0.7)
	I do not wish to disclose this information	1 (2.5)	6 (4.3)

Table 4-A Socio-demographic characteristics Survey-online and the Survey-VR NCL-Full categories

Trip/travel behaviour characteristics		VR- SC N (%)	Online SC N (%)
Total sample		40	139
Travel time most recent trip by taxi	5km destinations	91 (37.9)	461 (55.3)
	10km destinations	86 (35.8)	169 (20.3)
	15km destinations	63 (26.3)	204 (24.5)
Activities at Northumberland Rd	Business	7 (17.5)	2 (1.4)
	Commuting (e.g. work or school)	6 (15.0)	3 (2.2)
	Leisure (e.g. shopping, pub)	18 (45.0)	127 (91.4)
	Visiting friends	2 (5.0)	6 (4.3)
	Holiday	1 (2.5)	0 (0.0)
	I was at home	5 (12.5)	0 (0.0)
	Other, please specify	1 (2.5)	1 (0.7)
Frequency of using taxis	At least once a week	2 (5.0)	12 (8.6)
	Less than once a week, at least once a month	13 (32.5)	50 (36.0)
	Less than once a month, more than twice a year	21 (52.5)	62 (44.6)
	At most twice a year	4 (10.0)	15 (10.8)
Frequency of talking with the driver	Very infrequently	6 (15.0)	7 (5.0)
	Somewhat infrequently	8 (20.0)	17 (12.2)
	Occasionally	9 (22.5)	47 (33.8)
	Somewhat frequently	12 (30.0)	46 (33.1)
	Very frequently	5 (12.5)	22 (15.8)
Enjoying talking with taxi driver	Always	6 (15.0)	12 (8.6)
	Sometimes	29 (72.5)	114 (82.0)
	Never	5 (12.5)	13 (9.4)
Like the driver to help carry luggage	Yes	27 (67.5)	77 (55.4)
	No	13 (32.5)	62 (44.6)
Take the taxi without any help	Yes	40 (100.0)	138 (99.3)
	No	0 (0.0)	1 (0.7)

Table 4-B Trip/travel behaviour characteristics - Survey-online and the Survey-VR NCL-Full categories

Socio-demographic characteristics		A (AT users)	B (AT cities)	C (NT cities)
Total sample		N=35 (%)	N=299 (%)	N=151 (%)
Gender	Female	24 (68.6)	145 (48.5)	56 (37.1)
	Male	11 (31.4)	154 (51.5)	95 (62.9)
Age	Below 30	7 (20.0)	65 (21.7)	89 (58.9)
	30-39	6 (17.1)	65 (21.7)	18 (11.9)
	40-49	5 (14.3)	51 (17.1)	20 (13.2)
	50-59	9 (25.7)	50 (16.7)	19 (12.6)
	60 or above	8 (22.9)	61 (20.4)	5 (3.3)
Education level	Secondary school and below	0 (0.0)	5 (1.7)	4 (2.6)
	High school	1 (2.9)	17 (5.7)	6 (4.0)
	College degree	2 (5.7)	55 (18.4)	35 (23.2)
	Bachelor degree	28 (80.0)	192 (64.2)	39 (25.8)
	Master degree	3 (8.6)	26 (8.7)	62 (41.1)
	PhD degree	1 (2.9)	4 (1.3)	5 (3.3)
Current work status	Employed full-time	30 (85.7)	203 (67.9)	92 (60.9)
	Employed part-time	0 (0.0)	4 (1.3)	0 (0.0)
	Self-employed	2 (5.7)	7 (2.3)	1 (0.7)
	Jobless	1 (2.9)	7 (2.3)	1 (0.7)
	Students	1 (2.9)	28 (9.4)	53 (35.1)
	Retired	1 (2.9)	50 (16.7)	4 (2.6)
	Others	0 (0.0)	0 (0.0)	0 (0.0)
Personal monthly disposable income	<¥1000	1 (2.9)	17 (5.7)	21 (13.9)
	¥1001 - ¥ 3000	2 (5.7)	4 (1.3)	18 (11.9)
	¥3001 - ¥5000	1 (2.9)	27 (9.0)	26 (17.2)
	¥5001 - ¥10000	7 (20.0)	131 (43.8)	29 (19.2)
	¥10001 - ¥20000	17 (48.6)	85 (28.4)	9 (6.0)
	>¥20000	7 (20.0)	31 (10.4)	2 (1.3)
	I do not wish to disclose this information	0 (0.0)	4 (1.3)	46 (30.5)

Table 4-C Socio-demographic characteristics- China AT existence Full categories

Trip characteristics		A (AT users)	B (AT cities)	C (NT cities)
Total sample		N=35 (%)	N=299 (%)	N=151 (%)
Self-reported travel time of a recent trip	Short (10 minutes or less)	7 (20.0)	51 (17.1)	67 (44.4)
	Medium (around 20 minutes)	20 (57.1)	112 (37.5)	51 (33.8)
	Long (30 minutes or more)	8 (22.9)	136 (45.5)	33 (21.8)
Activities before taxi trip	Business	1 (2.9)	30 (10.0)	16 (10.6)
	Commuting (e.g. work or school)	10 (28.6)	58 (19.4)	39 (25.8)
	Leisure (e.g. shopping, pub)	16 (45.7)	112 (37.5)	38 (25.2)
	Visiting friends	6 (17.1)	32 (10.7)	11 (7.3)
	Holiday	1 (2.9)	7 (2.3)	7 (4.6)
	I was at home	1 (2.9)	54 (18.1)	36 (23.8)
	Other, please specify	0 (0.0)	6 (2.0)	4 (2.6)
Trip purpose	Business	2 (5.7)	39 (13.0)	15(9.9)
	Commuting (e.g. work or school)	10 (28.6)	57 (19.1)	41 (27.2)
	Leisure (e.g. shopping, pub)	11 (31.4)	119 (39.8)	52 (34.4)
	Visiting friends	7 (20.0)	40 (13.4)	15 (9.9)
	Holiday	0 (0.0)	6 (2.0)	7 (4.6)
	Going home	1 (2.9)	28 (9.4)	14 (9.3)
	Other, please specify	4 (11.4)	10 (3.3)	7 (4.6)
Enjoying talking with taxi driver	Always	7 (20.0)	27 (9.0)	8 (5.3)
	Sometimes	25 (71.4)	220 (73.6)	115 (76.2)
	Never	3 (8.6)	52 (17.4)	28 (18.5)
Frequency of talking with the driver	Very infrequently	2 (5.7)	32 (10.7)	22 (14.6)
	Somewhat infrequently	2 (5.7)	43 (14.4)	20 (13.2)
	Occasionally	14 (40.0)	168 (56.2)	87 (57.6)
	Somewhat frequently	13 (37.1)	46 (15.4)	20 (13.2)
	Very frequently	4 (11.4)	10 (3.3)	2 (1.3)
Frequency of using taxis	At least once a week	26 (74.3)	156 (52.2)	37 (24.5)
	Less than once a week, at least once a month	8 (22.9)	104 (34.8)	60 (39.7)
	Less than once a month, more than twice a year	1 (2.9)	36 (12.0)	46 (30.5)
	At most twice a year	0 (0.0)	3 (1.0)	8 (5.3)
Like driver to help carry luggage	Yes	35 (100.0)	298 (99.7)	149 (98.7)
	No	0 (0.0)	1 (0.3)	2 (1.3)
Take the taxi without any help	Yes	35 (100.0)	298 (99.7)	149 (98.7)
	No	0 (0.0)	1 (0.3)	2 (1.3)

Table 4-D Trip characteristics- China AT existence Full categories

Appendix 5 – Modelling Results

	Joint ML online China& survey online UK			
	online UK		online China	
DISCRETE CHOICE MODEL	**Value**	**Rob. t-test**	**Value**	**Rob. t-test**
ASC (AT)	-0.084	-0.47	-0.676	-4.07
SIGMA (AT)	1.380	2.37	1.340	12.51
Scale parameter applied on UK dataset [t-test against 1]	1.410	2.39[0.69]	-	-
Level of services				
Travel cost	-0.292	-2.38	-0.607	-8.34
*Long trip (30 minutes or more)	0.080	1.72	-0.266	-2.38
Travel time	-0.089	-2.27	-0.035	-5.72
Waiting time	-0.076	-2.19	-0.077	-9.48
* Age60_more	-	-	0.037	2.13
* Full-time employee	-0.035	-1.64	-	-
AT features				
Change the destination **(generic)**	0.200	3.27	-	-
Chat during the trip	0.233	1.55	0.556	5.57
Social Conformity				
Number of customers in the last hour/100	-	-	-	-
… * Heard of ATs from those who tested it	-	-	0.230	2.44
Good review yesterday	0.466	2.22	0.370	4.67
… * Age18_29	0.493	1.89	-	-
… * Long trip (30 minutes or more)	-	-	0.672	4.88
Systematic heterogeneity in AT alternative				
Age18_29	-	-	-0.487	-2.84
Frequently use taxis (at least once a week)	-	-	0.762	4.70
Like to carry luggage	-0.783	-2.10	-	-
Enjoy talking with driver	-1.330	-2.04	-	-
Heard of ATs operating in China or testing in the UK	-	-	0.656	4.06
Not familiar at all: 5 levels of automation	-0.696	-1.92	-	-
Summary of statistics				
Number of draws	500			
Log-Likelihood Market Share	-2983.34			
Maximum Log-likelihood	-2482.410			
Akaike Information Criterion (AIC)	5022.820			
Bayesian Information Criterion (BIC)	5209.489			
Number of individuals	769			
Number of observations	4614			

Table 5-A joint mixed logit model *survey online China& survey online UK*

	Joint ML VR NCL & online NCL			
	VR NCL		**online NCL**	
DISCRETE CHOICE MODEL	Value	Rob. t-test	Value	Rob. t-test
ASC (AT)	-0.026	-0.17	-1.090	-3.93
SIGMA (AT)	0.368	1.58	1.820	7.86
Scale factor applied on VR dataset [t-test against 1]	1.350	2.77 [0.72]	-	-
Level of services				
Travel cost	-0.187	-3.18	-0.467	-9.98
Travel time **(Generic)**	-0.100	-4.35	-	-
* Master degree and PhD	-	-	-0.094	-2.97
Waiting time	-0.081	-2.91	-0.127	-5.00
* Frequent use of taxi (at least once month)	-0.074	-2.29	-	-
AT features				
Change the destination	-	-	-	-
… * Male	-	-	0.530	2.79
… * Age18_29	0.415	1.8	-	-
Chat during the trip	-	-	-	-
… * Age18_29	-	-	0.589	1.75
…* Infrequent talking with driver	0.688	2.15	-	-
Social Conformity				
Number of customers in the last hour/100	0.241	1.83	-	-
…* Frequent use of taxi (at least once a week)	-	-	0.633	2.09
Good review yesterday	0.504	1.99	0.877	5.86
Systematic heterogeneity in AT alternative				
Male	-	-	0.914	2.48
Summary of statistics				
Number of draws	500			
Log-Likelihood Market Share	-715.358			
Maximum Log-likelihood	-544.232			
Akaike Information Criterion (AIC)	1130.464			
Bayesian Information Criterion (BIC)	1235.026			
Number of individuals	179			
Number of observations	1074			

Table 5-B joint mixed logit model *survey VR NCL& survey online NCL*

	ML3_ AT-existence _China						
	Generic		NT cities		AT cities		Rob t-test between coeffs
DISCRETE CHOICE MODEL	Value	Rob. t-test	Value	Rob. t-test	Value	Rob. t-test	
ASC (AT)	-0.681	-4.05					
SIGMA (AT)	1.370	13.26					
Level of services							
Travel cost [GBP][a]			-0.528	-3.89	-0.544	-5.33	0.10
*Long trip (30 minutes or more)	-0.273	-2.39					
Travel time [minutes]			-0.037	-3.51	-0.027	-3.34	-0.72
Waiting time [minutes]			-0.079	-5.35	-0.063	-5.2	-0.84
* Age60_more	0.021	1.26					
AT features							
Change the destination			0.350	3.41	0.154	2.05	1.56
Chat during the trip			0.556	3.49	0.486	3.95	0.36
Social conformity							
Number of customers in the last hour/100			-0.066	-0.89	-0.022	-0.32	-0.46
… * Heard of ATs from those who tested it	0.240	2.79					
Good review yesterday			0.207	2.04	0.540	5.24	-2.54
… * Long trip (30 minutes or more)	0.530	3.74					
Systematic heterogeneity in AT alternative							
Age18_29			-0.425	-1.49	-0.473	-2.33	0.15
Frequently use taxis (at least once a week)			0.240	0.75	0.966	5.51	-0.29
Heard of ATs operating in China or testing in the UK			0.600	1.81	0.704	4.09	-0.29
Summary of statistics							
Number of draws	500						
Log-Likelihood Market Share	-1895.72						
Maximum Log-likelihood	-1659.49						
Akaike Information Criterion (AIC)	3370.985						
Bayesian Information Criterion (BIC)	3526.358						
Number of individuals	485						
Number of observations	2910						

Table 5-C ML3_AT-existence_China

References

Abou-Zeid, M. and Ben-Akiva, M. (2011) 'The effect of social comparisons on commute well-being', *Transportation Research Part A: Policy and Practice*, 45(4), pp. 345-361.

Alessandrini, A., Alfonsi, R., Delle Site, P. and Stam, D. (2014) 'Users' preferences towards automated road public transport: results from European surveys', *Transportation Research Procedia*, 3, pp. 139-144.

Alonso, B., Barreda, R., dell'Olio, L. and Ibeas, A. (2018) 'Modelling user perception of taxi service quality', *Transport Policy*, 63, pp. 157-164.

Anable, J. (2005) ''Complacent Car Addicts' or 'Aspiring Environmentalists'? Identifying travel behaviour segments using attitude theory', *Transport Policy*, 12(1), pp. 65-78.

Araghi, Y., Kroesen, M., Molin, E. and van Wee, B. (2014) 'Do social norms regarding carbon offsetting affect individual preferences towards this policy? Results from a stated choice experiment', *Transportation Research Part D: Transport and Environment*, 26, pp. 42-46.

Arellana, J., Garzón, L., Estrada, J. and Cantillo, V. (2020) 'On the use of virtual immersive reality for discrete choice experiments to modelling pedestrian behaviour', *Journal of choice modelling*, 37, p. 100251.

Askari, S., Peiravian, F., Tilahun, N. and Yousefi Baseri, M. (2021) 'Determinants of users' perceived taxi service quality in the context of a developing country', *Transportation Letters*, 13(2), pp. 125-137.

Asmussen, K. E., Mondal, A. and Bhat, C. R. (2020) 'A socio-technical model of autonomous vehicle adoption using ranked choice stated preference data', *Transportation Research Part C: Emerging Technologies*, 121, p. 102835.

Bamberg, S., Hunecke, M. and Blöbaum, A. (2007) 'Social context, personal norms and the use of public transportation: Two field studies', *Journal of Environmental Psychology*, 27(3), pp. 190-203.

Bansal, P. and Daziano, R. A. (2018) 'Influence of choice experiment designs on eliciting preferences for autonomous vehicles', *Transportation Research Procedia*, 32, pp. 474-481.

Bansal, P., Kockelman, K. M. and Singh, A. (2016) 'Assessing public opinions of and interest in new vehicle technologies: An Austin perspective', *Transportation Research Part C: Emerging Technologies*, 67, pp. 1-14.

Becker, F. and Axhausen, K. W. (2017) 'Literature review on surveys investigating the acceptance of automated vehicles', *Transportation*, 44(6), pp. 1293-1306.

Begg, D. (2014) 'A 2050 vision for London: what are the implications of driverless transport?'.

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

Ben-Akiva, M., McFadden, D., Train, K., Walker, J., Bhat, C., Bierlaire, M., Bolduc, D., Boersch-Supan, A., Brownstone, D. and Bunch, D. S. (2002) 'Hybrid choice models: Progress and challenges', *Marketing Letters*, 13(3), pp. 163-175.

Ben-Elia, E., Levy, N., Benenson, I. and Ashkenazi, S. (2015) 'Serious game-based study of urban parking dynamics', *NECTAR International Conference*. Ann Arbor.

Bhat, C. R. (2000) 'Incorporating observed and unobserved heterogeneity in urban work travel mode choice modeling', *Transportation science*, 34(2), pp. 228-238.

Bhat, C. R. (2003) 'Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences', *Transportation Research Part B: Methodological,* 37(9), pp. 837-855.

Bhat, C. R., Dubey, S. K. and Nagel, K. (2015) 'Introducing non-normality of latent psychological constructs in choice modeling with an application to bicyclist route choice', *Transportation Research Part B: Methodological*, 78, pp. 341-363.

Bierlaire, M. (2003). 'BIOGEME: A free package for the estimation of discrete choice models', *Proceedings of the 3rd Swiss transport research conference*. Ascona, Switzerland.

Bierlaire, M. and Fetiarison, M. (2009). 'Estimation of discrete choice models: extending BIOGEME', *9th Swiss Transport Research Conference*. Monte Veritá, Switzerland.

Birenboim, A., Dijst, M., Ettema, D., de Kruijf, J., de Leeuw, G. and Dogterom, N. (2019) 'The utilization of immersive virtual environments for the investigation of environmental preferences', *Landscape and Urban Planning*, 189, pp. 129-138.

Bliemer, M. C. and Rose, J. M. (2009) 'Efficiency and sample size requirements for stated choice experiments', *88th Transportation Research Board*. Washington DC, USA.

Bogacz, M., Hess, S., Calastri, C., Choudhury, C. F., Mushtaq, F., Awais, M., Nazemi, M., van Eggermond, M. A. and Erath, A. (2021) 'Modelling risk perception using a dynamic hybrid choice model and brain-imaging data: an application to virtual reality cycling', *Transportation research part C: emerging technologies*, 133, p. 103435.

Bollen, K. A. (1989) *Structural equations with latent variables*. John Wiley & Sons.

Casley, S. V., Jardim, A. S. and Quartulli, A. (2013) 'A study of public acceptance of autonomous cars', *Worcester Polytechnic Institute, Bachelor Thesis*.

Chaudhary, S., Tyagi, V. and Rao, P. H. (2016) 'Factors affecting customer satisfaction in the taxi service market in India', *Journal of Entrepreneurship & Management*, 5 (3), pp. 46-53.

Chen, C.-W. (2017) 'Five-star or thumbs-up? The influence of rating system types on users' perceptions of information quality, cognitive effort, enjoyment and continuance intention', *Internet Research*.

Chen, J., Li, R., Gan, M., Fu, Z. and Yuan, F. (2020) 'Public acceptance of driverless buses in China: An empirical analysis based on an extended UTAUT model', *Discrete Dynamics in Nature and Society*, 2020.

Chen, Y. and Wyer Jr, R. S. (2020) 'The effects of endorsers' facial expressions on status perceptions and purchase intentions', *International Journal of Research in Marketing*, 37(2), pp. 371-385.

Cherchi, E. (2017) 'A stated choice experiment to measure the effect of informational and normative conformity in the preference for electric vehicles', *Transportation Research Part A: Policy and Practice*, 100, pp. 88-104.

Cherchi, E. (2020) 'Our IATBR: 45 years contributing to travel behavior research', *in Mapping the Travel Behavior Genome*. Elsevier, pp. 17-28.

Cherchi, E. and Hensher, D. A. (2015) 'Workshop synthesis: Stated preference surveys and experimental design, an audit of the journey so far and future research perspectives', *Transportation Research Procedia*, 11, pp. 154-164.

Cherchi, E. and Ortúzar, J. d. D. (2006) 'On fitting mode specific constants in the presence of new options in RP/SP models', *Transportation Research Part A: Policy and Practice*, 40(1), pp. 1-18.

Chintagunta, P. K., Jain, D. C. and Vilcassim, N. J. (1991) 'Investigating heterogeneity in brand preferences in logit models for panel data', *Journal of Marketing Research*, 28(4), pp. 417-428.

Choi, J. K. and Ji, Y. G. (2015) 'Investigating the importance of trust on adopting an autonomous vehicle', International Journal of Human-Computer Interaction, 31(10), pp. 692-702.

ChoiceMetrics, N. (2012) '1.2 User Manual & Reference Guide, Australia'.

Chorus, C. G. and Kroesen, M. (2014) 'On the (im-) possibility of deriving transport policy implications from hybrid choice models', *Transport Policy*, 36, pp. 217-222.

Cialdini, R. B. (2005) 'Basic social influence is underestimated', *Psychological inquiry*, 16(4), pp. 158-161.

Cialdini, R. B. (2007) 'Descriptive social norms as underappreciated sources of social control', *Psychometrika*, 72(2), pp. 263-268.

Cialdini, R. B. and Goldstein, N. J. (2004) 'Social influence: Compliance and conformity', *Annu. Rev. Psychol*., 55, pp. 591-621.

Clifton, K. J. and Handy, S. L. (2003) Qualitative methods in travel behaviour research. Emerald Group Publishing Limited.

Correia, G. H. d. A., Looff, E., van Cranenburgh, S., Snelder, M. and van Arem, B. (2019) 'On the impact of vehicle automation on the value of travel time while performing work and leisure activities in a car: Theoretical insights and results from a stated preference survey', *Transportation Research Part A: Policy and Practice*, 119, pp. 359-382.

Cosley, D., Lam, S. K., Albert, I., Konstan, J. A. and Riedl, J. (2003) 'Is seeing believing? How recommender system interfaces affect users' opinions', *Proceedings of the SIGCHI conference on Human factors in computing systems*. pp. 585-592

Crutchfield, R. S. (1955) 'Conformity and character', *American psychologist*, 10(5), p. 191.

Daziano, R. A., Sarrias, M. and Leard, B. (2017) 'Are consumers willing to pay to let cars drive for them? Analyzing response to autonomous vehicles', *Transportation Research Part C: Emerging Technologies*, 78, pp. 150-164.

de Looff, E., Correia, G.H.D.A., van Cranenburg, S., Snelder, M., van Arem, B., 2018. 'Potential changes in value of travel time as a result of vehicle automation: a case- study in the Netherlands'. *The 97th Annual Meeting of the Transportation Research Board*. Washington, D.C., USA .

DfT (2015) Road accident data 2014. Available at: https://discover.ukdataservice.ac.uk/catalogue/?sn=7752&type=Data%20catalogue. (Accessed: 6 May).

DfT (2018) Taxi and Private Hire Vehicle Statistics, England: 2018. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/751202/taxi-and-phv-england-2018.pdf (Accessed: 18 January).

Ding, Y., Li, R., Wang, X. and Schmid, J. (2021) 'Heterogeneity of autonomous vehicle adoption behavior due to peer effects and prior-AV knowledge', *Transportation*, pp. 1-24.

Djavadian, S., Farooq, B., Vasquez, R. and Yip, G. (2020) 'Virtual immersive reality based analysis of behavioural responses in connected and autonomous vehicle environment', *in Mapping the Travel Behavior Genome*. Elsevier, pp. 543-559.

Domencich, T. A. and McFadden, D. (1975) Urban travel demand-a behavioral analysis.

Etzioni, S., Hamadneh, J., Elvarsson, A. B., Esztergár-Kiss, D., Djukanovic, M., Neophytou, S. N., Sodnik, J., Polydoropoulou, A., Tsouros, I. and Pronello, C. (2020) 'Modeling cross-national differences in automated vehicle acceptance', *Sustainability*, 12(22), p. 9765.

Fagnant, D. J. and Kockelman, K. (2015) 'Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations', Transportation *Research Part A: Policy and Practice*, 77, pp. 167-181.

Fagnant, D. J. and Kockelman, K. M. (2014) 'The travel and environmental implications of shared autonomous vehicles, using agent-based model scenarios', *Transportation Research Part C: Emerging Technologies*, 40, pp. 1-13.

Farooq, B., Cherchi, E. and Sobhani, A. (2018) 'Virtual immersive reality for stated preference travel behavior experiments: A case study of autonomous vehicles on urban roads', *Transportation research record*, 2672(50), pp. 35-45.

Farzin, I. and Mamdoohi, A. (2021) 'Acceptance of Autonomous Vehicles using a Combination of UTAUT and DOI', *Amirkabir Journal of Civil Engineering*, 53(8), pp. 15-15.

Feng, Y., Duives, D. C. and Hoogendoorn, S. P. (2022) 'Development and evaluation of a VR research tool to study wayfinding behaviour in a multi-story building', *Safety science*, 147, p. 105573.

Fifer, S., Rose, J. and Greaves, S. (2014) 'Hypothetical bias in Stated Choice Experiments: Is it a problem? And if so, how do we deal with it?', *Transportation research part A: policy and practice*, 61, pp. 164-177.

Fiore, S. M., Harrison, G. W., Hughes, C. E. and Rutström, E. E. (2009) 'Virtual experiments and environmental policy', *Journal of Environmental Economics and Management*, 57(1), pp. 65-86.

Fulman, N., Benenson, I. and Ben-Elia, E. (2020) 'Modeling parking search behavior in the city center: A game-based approach', *Transportation Research Part C: Emerging Technologies*, 120,

p. 102800.

Geva, S., Fulman, N. and Ben-Elia, E. (2022) 'Getting the prices right: Drivers' cruising choices in a serious parking game', *Transportation Research Part A: Policy and Practice*, 165, pp. 54-75.

Gkartzonikas, C. and Gkritza, K. (2019) 'What have we learned? A review of stated preference and choice studies on autonomous vehicles', *Transportation Research Part C: Emerging Technologies*, 98, pp. 323-337.

Goldstein, N. J. and Cialdini, R. B. (2011) 'Using social norms as a lever of social influence', in The science of social influence. *Psychology Press*, pp. 167-191.

Greenblatt, J. B. and Shaheen, S. (2015) 'Automated vehicles, on-demand mobility, and environmental impacts', *Current sustainable/renewable energy reports*, 2(3), pp. 74-81.

Haboucha, C. J., Ishaq, R. and Shiftan, Y. (2017) 'User preferences regarding autonomous vehicles', *Transportation Research Part C: Emerging Technologies*, 78, pp. 37-49.

Haghani, M., Bliemer, M. C., Rose, J. M., Oppewal, H. and Lancsar, E. (2021a) 'Hypothetical bias in stated choice experiments: Part I. Macro-scale analysis of literature and integrative synthesis of empirical evidence from applied economics, experimental psychology and neuroimaging', *Journal of choice modelling*, 41, p. 100309.

Haghani, M., Bliemer, M. C., Rose, J. M., Oppewal, H. and Lancsar, E. (2021b) 'Hypothetical bias in stated choice experiments: Part II. Conceptualisation of external validity, sources and explanations of bias and effectiveness of mitigation methods', *Journal of choice modelling*, 41, p. 100322.

Hensher, D. A. and Greene, W. H. (2003) 'The mixed logit model: the state of practice', *Transportation*, 30(2), pp. 133-176.

Hensher, D. A. and Mulley, C. (2015) 'Modal image: candidate drivers of preference differences for BRT and LRT', *Transportation*, 42(1), pp. 7-23.

Herriges, J., Kling, C., Liu, C.-C. and Tobias, J. (2010) 'What are the consequences of consequentiality?', *Journal of environmental economics and management*, 59(1), pp. 67-81.

Hess, S. and Rose, J. M. (2009) 'Should reference alternatives in pivot design SC surveys be treated differently?', *Environmental and Resource Economics*, 42(3), pp. 297-317.

Hess, S., Train, K. E. and Polak, J. W. (2006) 'On the use of a Modified Latin Hypercube Sampling (MLHS) method in the estimation of a Mixed Logit model for vehicle choice', *Transportation Research Part B: Methodological,* 40(2), pp. 147-163.

Hewitt, C., Politis, I., Amanatidis, T. and Sarkar, A. (2019) 'Assessing public perception of self-driving cars: The autonomous vehicle acceptance model', *Proceedings of the 24th international conference on intelligent user interfaces*. pp. 518-527.

Hicks, J. R. (1956) A Revision of Demand Theory. Oxford: Clarendon Press.

Hohenberger, C., Spörrle, M. and Welpe, I. M. (2016) 'How and why do men and women differ in their willingness to use automated cars? The influence of emotions across different age groups', *Transportation Research Part A: Policy and Practice*, 94, pp. 374-385.

Howard, D. and Dai, D. (2014) 'Public perceptions of self-driving cars: The case of Berkeley, California', *Transportation research board 93rd annual meeting*. pp. 1-16.

Hulse, L. M., Xie, H. and Galea, E. R. (2018) 'Perceptions of autonomous vehicles: Relationships with road users, risk, gender and age', *Safety Science*, 102, pp. 1-13.

Innocenti, A. (2017) 'Virtual reality experiments in economics', *Journal of behavioral and experimental economics*, 69, pp. 71-77.

Jara-Díaz, S. R. and Videla, J. (1989) 'Detection of income effect in mode choice: theory and application', *Transportation Research Part B: Methodological*, 23(6), pp. 393-400.

Jardim, A.S., Quartulli, A.M., Casley, S.V., 2013. A Study of Public Acceptance of Autonomous Cars. Available at: https://digitalcommons.wpi.edu/iqp-all/2596 (Accessed: 15 January).

Jensen, A. F., Cherchi, E. and Mabit, S. L. (2013) 'On the stability of preferences and attitudes before and after experiencing an electric vehicle', *Transportation Research Part D: Transport and Environment*, 25, pp. 24-32.

Jensen, A. F., Cherchi, E., Mabit, S. L. and Ortúzar, J. d. D. (2017) 'Predicting the potential market for electric vehicles', *Transportation Science*, 51(2), pp. 427-440.

Jing, P., Huang, H., Ran, B., Zhan, F. and Shi, Y. (2019) 'Exploring the factors affecting mode choice Intention of autonomous vehicle based on an extended theory of planned behavior—A case study in China', *Sustainability*, 11(4), p. 1155.

Jing, P., Xu, G., Chen, Y., Shi, Y. and Zhan, F. (2020) 'The determinants behind the acceptance of autonomous vehicles: A systematic review', *Sustainability*, 12(5), p. 1719.

Kaan, J. (2017) 'User acceptance of autonomous vehicles: Factors & implications'.

Kahn, B. E. (1995) 'Consumer variety-seeking among goods and services: An integrative review', *Journal of retailing and consumer services*, 2(3), pp. 139-148.

Kamargianni, M. and Polydoropoulou, A. (2013) 'Hybrid choice model to investigate effects of teenagers' attitudes toward walking and cycling on mode choice behavior', *Transportation research record*, 2382(1), pp. 151-161.

Kaur, K. and Rampersad, G. (2018) 'Trust in driverless cars: Investigating key factors influencing the adoption of driverless cars', *Journal of Engineering and Technology Management*, 48, pp. 87-96.

Kaye, S.-A., Lewis, I., Forward, S. and Delhomme, P. (2020) 'A priori acceptance of highly automated cars in Australia, France, and Sweden: A theoretically-informed investigation guided by the TPB and UTAUT', *Accident Analysis & Prevention*, 137, p. 105441.

Keszey, T. (2020) 'Behavioural intention to use autonomous vehicles: Systematic review and empirical extension', *Transportation Research Part C: Emerging Technologies*, 119, p. 102732.

Kim, J., Rasouli, S. and Timmermans, H. (2014) 'Expanding scope of hybrid choice models allowing for mixture of social influences and latent attitudes: Application to intended purchase of electric cars', *Transportation research part A: policy and practice, 69*, pp. 71-85.

Kim, J., Rasouli, S. and Timmermans, H. J. P. (2017) 'Investigating heterogeneity in social influence by social distance in car-sharing decisions under uncertainty: A regret-minimizing hybrid choice model framework based on sequential stated adaptation experiments', *Transportation Research Part C: Emerging Technologies*, 85, pp. 47-63.

Kim, S., Chang, J. J. E., Park, H. H., Song, S. U., Cha, C. B., Kim, J. W. and Kang, N. (2020) 'Autonomous taxi service design and user experience', *International Journal of Human–Computer Interaction*, 36(5), pp. 429-448.

Kolarova, V. and Cherchi, E. (2021) 'Impact of trust and travel experiences on the value of travel time savings for autonomous driving', *Transportation Research Part C: Emerging Technologies*, 131, p. 103354.

Kolarova, V., Steck, F. and Bahamonde-Birke, F. J. (2019) 'Assessing the effect of autonomous driving on value of travel time savings: A comparison between current and future preferences', *Transportation Research Part A: Policy and Practice*, 129, pp. 155-169.

König, M. and Neumayr, L. (2017) 'Users' resistance towards radical innovations: The case of the self-driving car', *Transportation Research Part F: Traffic Psychology and Behaviour*, 44, pp. 42-52.

Kormos, C., Gifford, R. and Brown, E. (2015) 'The influence of descriptive social norm information on sustainable transportation behavior: A field experiment', *Environment and Behavior*, 47(5), pp. 479-501.

Krčál, O., Peer, S., Staněk, R. and Karlinova, B. (2019) 'Real consequences matter: Why hypothetical biases in the valuation of time persist even in controlled lab experiments', *Economics of Transportation*, 20, p. 100138.

Krueger, R., Rashidi, T. H. and Dixit, V. V. (2019) 'Autonomous driving and residential location preferences: Evidence from a stated choice survey', *Transportation Research Part C: Emerging Technologies*, 108, pp. 255-268.

Krueger, R., Rashidi, T. H. and Rose, J. M. (2016) 'Preferences for shared autonomous vehicles', *Transportation Research Part C: Emerging Technologies*, 69, pp. 343-355.

Kuwano, M., Tsukai, M. and Matsubara, T. (2012) 'Analysis on promoting factors of electric vehicles with social conformity', *13th International Conference on Travel Behaviour Research*. pp. 15-20

Kyriakidis, M., Happee, R. and de Winter, J. C. (2015) 'Public opinion on automated driving: Results of an international questionnaire among 5000 respondents', *Transportation research part F: traffic psychology and behaviour*, 32, pp. 127-140.

Lazarus, R. (1991) '(1991a). Cognition and motivation in emotion. American *Psychologist*, 46, 352-367'.

Lee, J. D. and See, K. A. (2004) 'Trust in automation: Designing for appropriate reliance', *Human factors*, 46(1), pp. 50-80.

Lee, J., Lee, E., Yun, J., Chung, J.-H. and Kim, J. (2021) 'Latent heterogeneity in autonomous driving preferences and in-vehicle activities by travel distance', *Journal of transport geography*, 94, p. 103089.

Li, J., George, C., Ngao, A., Holländer, K., Mayer, S. and Butz, A. (2020) 'An Exploration of

Users' Thoughts on Rear-Seat Productivity in Virtual Reality', *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. pp. 92-95.

Liu, P., Guo, Q., Ren, F., Wang, L. and Xu, Z. (2019a) 'Willingness to pay for self-driving vehicles: Influences of demographic and psychological factors', *Transportation Research Part C: Emerging Technologies*, 100, pp. 306-317.

Liu, P., Yang, R. and Xu, Z. (2019b) 'Public acceptance of fully automated driving: Effects of social trust and risk/benefit perceptions', *Risk Analysis*, 39(2), pp. 326-341.

Liu, Z. and Park, S. (2015) 'What makes a useful online review? Implication for travel product websites', *Tourism management*, 47, pp. 140-151.

Loureiro, S. M. C., Guerreiro, J., Eloy, S., Langaro, D. and Panchapakesan, P. (2019) 'Understanding the use of Virtual Reality in Marketing: A text mining-based review', *Journal of Business Research*, 100, pp. 514-530.

Louviere, J. J., Hensher, D. A. and Swait, J. D. (2000) Stated choice methods: analysis and applications. Cambridge university press.

Lovreglio, R., Fonzone, A. and Dell'Olio, L. (2016) 'A mixed logit model for predicting exit choice during building evacuations', *Transportation Research Part A: Policy and Practice*, 92, pp. 59-75.

Madigan, R., Louw, T., Dziennus, M., Graindorge, T., Ortega, E., Graindorge, M. and Merat, N. (2016) 'Acceptance of automated road transport systems (ARTS): an adaptation of the UTAUT model', *Transportation Research Procedia*, 14, pp. 2217-2226.

Madigan, R., Louw, T., Wilbrink, M., Schieben, A. and Merat, N. (2017) 'What influences the decision to use automated public transport? Using UTAUT to understand public acceptance of automated road transport systems', *Transportation research part F: traffic psychology and behaviour*, 50, pp. 55-64.

Manski, C. F. (1977) 'The structure of random utility models', *Theory and decision*, 8(3), p. 229.

McFadden, D. (1974) 'The measurement of urban travel demand', *Journal of public economics*, 3(4), pp. 303-328.

McFadden, D. (1986) 'The choice theory approach to market research', *Marketing science*, 5(4), pp. 275-297.

McFadden, D. (2000) 'Disaggregate behavioral travel demand's RUM side', *Travel behaviour research*, pp. 17-63.

McFadden, D. and Train, K. (2000) 'Mixed MNL models for discrete response', *Journal of applied Econometrics*, 15(5), pp. 447-470.

Meißner, M., Pfeiffer, J., Peukert, C., Dietrich, H. and Pfeiffer, T. (2020) 'How virtual reality affects consumer choice', *Journal of Business Research*, 117, pp. 219-231.

Mena-Oreja, J., Gozalvez, J. and Sepulcre, M. (2018) 'Effect of the configuration of platooning maneuvers on the traffic flow under mixed traffic scenarios', 2018 IEEE Vehicular Networking Conference (VNC). IEEE. pp. 1-4.

Mokas, I., Lizin, S., Brijs, T., Witters, N. and Malina, R. (2021) 'Can immersive virtual reality increase respondents' certainty in discrete choice experiments? A comparison with traditional presentation formats', *Journal of Environmental Economics and Management*, 109, p. 102509.

Morgan, D. L. and Krueger, R. A. (1998) 'The focus group guidebook'. Sage.

Morrison, G. and Van Belle, J.-P. (2020) 'Customer Intentions Towards Autonomous Vehicles in South Africa: An Extended UTAUT Model', 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence). IEEE. pp 525-531

Moser, R., Raffaelli, R. and Notaro, S. (2014) 'Testing hypothetical bias with a real choice experiment using respondents' own money', *European Review of Agricultural Economics*, 41(1), pp. 25-46.

Mourad, A., Puchinger, J. and Chu, C. (2019) 'Owning or sharing autonomous vehicles: comparing different ownership and usage scenarios', *European Transport Research Review*, 11(1), pp. 1-20.

Mudambi, S. M. and Schuff, D. (2010) 'Research note: What makes a helpful online review? A study of customer reviews on Amazon. com', *MIS quarterly*, pp. 185-200.

Narayanan, S., Chaniotakis, E. and Antoniou, C. (2020) 'Shared autonomous vehicle services: A comprehensive review', *Transportation Research Part C: Emerging Technologies*, 111, pp. 255-293.

Nicolaides, D., Cebon, D. and Miles, J. (2017) 'An autonomous taxi service for sustainable urban transportation', 2017 Smart City Symposium Prague (SCSP). IEEE. pp. 1-6

Nordhoff, S., De Winter, J., Kyriakidis, M., Van Arem, B. and Happee, R. (2018a) 'Acceptance of driverless vehicles: Results from a large cross-national questionnaire study', *Journal of Advanced Transportation*, 2018.

Nordhoff, S., de Winter, J., Madigan, R., Merat, N., van Arem, B. and Happee, R. (2018b) 'User acceptance of automated shuttles in Berlin-Schöneberg: A questionnaire study', *Transportation Research Part F: Traffic Psychology and Behaviour*, 58, pp. 843-854.

Nordhoff, S., Kyriakidis, M., Van Arem, B. and Happee, R. (2019) 'A multi-level model on automated vehicle acceptance (MAVA): A review-based study', Theoretical Issues in *Ergonomics Science*, 20(6), pp. 682-710.

Nordhoff, S., Stapel, J., van Arem, B. and Happee, R. (2020) 'Passenger opinions of the perceived safety and interaction with automated shuttles: A test ride study with 'hidden' safety steward', *Transportation Research Part A: Policy and Practice*, 138, pp. 508-524.

Okushima, M. (2015) 'Simulating social influences on sustainable mobility shifts for heterogeneous agents', *Transportation*, 42(5), pp. 827-855.

Ortúzar, J. de D. and Willumsen, L. G. (2011) Modelling transport. John wiley & sons.

Paddeu, D., Parkhurst, G. and Shergold, I. (2020) 'Passenger comfort and trust on first-time use of a shared autonomous shuttle vehicle', *Transportation Research Part C: Emerging Technologies*, 115, p. 102604.

Panagiotopoulos, I. and Dimitrakopoulos, G. (2018) 'An empirical investigation on consumers' intentions towards autonomous driving', *Transportation research part C: emerging technologies*, 95, pp. 773-784.

Pang, B. and Lee, L. (2005) 'Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales', *arXiv preprint cs/0506075*.

Pasupathi, M. (1999) 'Age differences in response to conformity pressure for emotional and nonemotional material', *Psychology and aging*, 14(1), p. 170.

Patterson, Z., Darbani, J. M., Rezaei, A., Zacharias, J. and Yazdizadeh, A. (2017) 'Comparing text-only and virtual reality discrete choice experiments of neighbourhood choice', *Landscape and Urban Planning*, 157, pp. 63-74.

Payre, W., Cestac, J. and Delhomme, P. (2014) 'Intention to use a fully automated car: Attitudes and a priori acceptability', *Transportation research part F: traffic psychology and behaviour*, 27, pp. 252-263.

Peck, J. and Shu, S. B. (2009) 'The effect of mere touch on perceived ownership', *Journal of consumer Research*, 36(3), pp. 434-447.

Polydoropoulou, A., Tsouros, I., Thomopoulos, N., Pronello, C., Elvarsson, A., Sigþórsson, H., Dadashzadeh, N., Stojmenova, K., Sodnik, J. and Neophytou, S. (2021) 'Who is willing to share their AV? Insights about gender differences among seven countries', *Sustainability*, 13(9), p. 4769.

Potoglou, D., Whittle, C., Tsouros, I. and Whitmarsh, L. (2020) 'Consumer intentions for alternative fuelled and autonomous vehicles: A segmentation analysis across six countries', *Transportation Research Part D: Transport and Environment*, 79, p. 102243.

Procházka, J., Leugnerová, M., Docekalova, S., Hlavackova, A., Morhacova, V., Nesporova, K., Jezek, S. and Vaculík, M. (2016) 'Do men conform more than women in the recognition and labeling of emotions?', *Studia Psychologica*, 58(4), p. 251.

Rahman, M. M., Deb, S., Strawderman, L., Burch, R. and Smith, B. (2019) 'How the older population perceives self-driving vehicles', *Transportation research part F: traffic psychology and behaviour*, 65, pp. 242-257.

Rasouli, S. and Timmermans, H. (2013a) 'The effect of social adoption on the intention to purchase electric cars: A stated choice approach' Proceedings 92nd Annual Meeting of the Transportation Research Board, January 13-17, Washington DC, USA (On-line: 18 pp).

Rasouli, S. and Timmermans, H. (2013b) 'Incorporating mechanisms of social adoption in design and analysis of stated-choice experiments: illustration and application to choice of electric cars', *Transportation research record*, 2344(1), pp. 10-19.

Raveau, S., Álvarez-Daziano, R., Yáñez, M. F., Bolduc, D. and Ortúzar, J. de D. (2010) 'Sequential and simultaneous estimation of hybrid discrete choice models: Some new findings', *Transportation Research Record*, 2156(1), pp. 131-139.

Ribeiro, M. A., Gursoy, D. and Chi, O. H. (2022) 'Customer acceptance of autonomous vehicles in travel and tourism', *Journal of Travel Research*, 61(3), pp. 620-636.

Riegler, A., Riener, A. and Holzmann, C. (2021) 'A systematic review of virtual reality applications for automated driving: 2009–2020', *Frontiers in human dynamics*, p. 48.

Rizzi, L. I. and Ortúzar, J. de D. (2003) 'Stated preference in the valuation of interurban road safety', *Accident Analysis & Prevention*, 35(1), pp. 9-22.

Rizzi, L. I. and Ortúzar, J. de D. (2006) 'Estimating the willingness-to-pay for road safety improvements', *Transport Reviews*, 26(4), pp. 471-485.

Rogers, E. (1962) 'Diffusion of Innovations. New Work: The Free Press'.

Rose, J. M. and Hensher, D. A. (2018) 'User satisfaction with taxi and limousine services in the Melbourne metropolitan area', *Journal of Transport Geography*, 70, pp. 234-245.

Rossetti, T. and Hurtubia, R. (2020) 'An assessment of the ecological validity of immersive videos in stated preference surveys', *Journal of choice modelling*, 34, p. 100198.

SAE (2018) Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems. Available at: https://www.sae.org/standards/content/j3016_201806 (Accessed: 6 May).

Sanbonmatsu, D. M., Strayer, D. L., Yu, Z., Biondi, F. and Cooper, J. M. (2018) 'Cognitive underpinnings of beliefs and confidence in beliefs about fully automated vehicles', *Transportation Research Part F: Traffic Psychology and Behaviour*, 55, pp. 114-122.

Sanchez-Vives, M. V. and Slater, M. (2005) 'From presence to consciousness through virtual reality', *Nature Reviews Neuroscience*, 6(4), pp. 332-339.

Schoettle, B. and Sivak, M. (2014) 'Public opinion about self-driving vehicles in China, India, Japan, the US, the UK, and Australia'. *University of Michigan, Ann Arbor, Transportation Research Institute*

Schoettle, B. and Sivak, M. (2015) 'Motorists' preferences for different levels of vehicle automation'. *University of Michigan, Ann Arbor, Transportation Research Institute*

Shaaban, K. and Kim, I. (2016) 'Assessment of the taxi service in Doha', *Transportation Research Part A: Policy and Practice,* 88, pp. 223-235.

Shabanpour, R., Golshani, N., Shamshiripour, A. and Mohammadian, A. K. (2018) 'Eliciting preferences for adoption of fully automated vehicles using best-worst analysis', *Transportation research part C: emerging technologies*, 93, pp. 463-478.

Silberg, G., Manassa, M., Everhart, K., Subramanian, D., Corley, M., Fraser, H. and Sinha, V. (2013) 'Self-driving cars: Are we ready', *Kpmg Llp*, 1, pp. 1-36.

Sillano, M. and Ortúzar, J. de D. (2005) 'Willingness-to-pay estimation with mixed logit models: some new evidence', *Environment and Planning A*, 37(3), pp. 525-550.

Sobhani, A. and Farooq, B. (2018) 'Impact of smartphone distraction on pedestrians' crossing behaviour: An application of head-mounted immersive virtual reality', *Transportation research part F: traffic psychology and behaviour*, 58, pp. 228-241.

Sottile, E., Cherchi, E. and Meloni, I. (2015) 'Measuring soft measures within a stated preference survey: the effect of pollution and traffic stress on mode choice', *Transportation Research Procedia*, 11, pp. 434-451.

South Tyneside Council (2016). South Tyneside taxi survey. Available at:

https://www.southtyneside.gov.uk/media/32852/Taxi-Survey-Final-Report/pdf/Taxi_Survey_Final_Report_230317 (Accessed: 10 January).

Sparling, E. I. and Sen, S. (2011) 'Rating: how difficult is it?', *Proceedings of the fifth ACM conference on Recommender systems*. pp. 149-156

Stoiber, T., Schubert, I., Hoerler, R. and Burger, P. (2019) 'Will consumers prefer shared and pooled-use autonomous vehicles? A stated choice experiment with Swiss households', *Transportation Research Part D: Transport and Environment*, 71, pp. 265-282.

Sweet, M. N. and Laidlaw, K. (2020) 'No longer in the driver’s seat: How do affective motivations impact consumer interest in automated vehicles?', *Transportation*, 47(5), pp. 2601-2634.

The World Bank (2021) World Development Indicators, PPP Conversion Factor, GDP (LCU Per International $). Available at: https://data.worldbank.org/indicator/PA.NUS.PPP (Accessed: 14 April).

Thomas, J. R., Martin, P., Etnier, J. and Silverman, S. J. (2022) Research methods in physical activity. *Human kinetics*.

Thorhauge, M., Cherchi, E., Walker, J. L. and Rich, J. (2019) 'The role of intention as mediator between latent effects and behavior: application of a hybrid choice model to study departure time choices', *Transportation*, 46(4), pp. 1421-1445.

Thorhauge, M., Haustein, S. and Cherchi, E. (2016) 'Accounting for the Theory of Planned Behaviour in departure time choice', *Transportation Research Part F: Traffic Psychology and Behaviour*, 38, pp. 94-105.

Thorhauge, M., Swait, J. and Cherchi, E. (2020) 'The habit-driven life: Accounting for inertia in departure time choices for commuting trips', *Transportation Research Part A: Policy and Practice*, 133, pp. 272-289.

Tian, Z., Feng, T., Timmermans, H. J. and Yao, B. (2021) 'Using autonomous vehicles or shared cars? Results of a stated choice experiment', *Transportation Research Part C: Emerging Technologies*, 128, p. 103117.

Toelch, U. and Dolan, R. J. (2015) 'Informational and normative influences in conformity from a neurocomputational perspective', *Trends in cognitive sciences*, 19(10), pp. 579-589.

Train, K. E. (2009) Discrete choice methods with simulation. Cambridge university press.

Train, K. E., McFadden, D. L. and Goett, A. A. (1987) 'Consumer attitudes and voluntary rate schedules for public utilities', *The review of economics and statistics*, pp. 383-391.

Turner, J. C. (1991) Social influence. Thomson Brooks/Cole Publishing Co.

Tussyadiah, I. P., Wang, D., Jung, T. H. and Tom Dieck, M. C. (2018) 'Virtual reality, presence, and attitude change: Empirical evidence from tourism', *Tourism management*, 66, pp. 140-154.

Tussyadiah, I. P., Zach, F. J. and Wang, J. (2017) 'Attitudes toward autonomous on demand mobility system: The case of self-driving taxi', *in Information and communication technologies in tourism 2017*. Springer, pp. 755-766.

Tversky, A. (1972) 'Elimination by aspects: A theory of choice', *Psychological review*, 79(4), p. 281.

Van der Heijden, H. (2004) 'User acceptance of hedonic information systems', *MIS quarterly*, pp. 695-704.

Velasco, J. P. N., Farah, H., van Arem, B. and Hagenzieker, M. P. (2019) 'Studying pedestrians' crossing behavior when interacting with automated vehicles using virtual reality', *Transportation research part F: traffic psychology and behaviour*, 66, pp. 1-14.

Venkatesh, V., Thong, J. Y. L. and Xu, X. (2012) 'Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology', *MIS Quarterly*, 36(1), pp. 157-178.

Vermeulen, I. E. and Seegers, D. (2009) 'Tried and tested: The impact of online hotel reviews on consumer consideration', *Tourism management*, 30(1), pp. 123-127.

Vij, A. and Walker, J. L. (2016) 'How, when and why integrated choice and latent variable models are latently useful', *Transportation Research Part B: Methodological*, 90, pp. 192-217.

Von Helversen, B., Abramczuk, K., Kopeć, W. and Nielek, R. (2018) 'Influence of consumer reviews on online purchasing decisions in older and younger adults', *Decision Support Systems*, 113, pp. 1-10.

Wadud, Z. (2017) 'Fully automated vehicles: A cost of ownership analysis to inform early adoption', *Transportation Research Part A: Policy and Practice*, 101, pp. 163-176.

Wakefield, K. L. and Inman, J. J. (2003) 'Situational price sensitivity: the role of consumption occasion, social context and income', *Journal of Retailing*, 79(4), pp. 199-212.

Walker, J. L. (2001) Extended discrete choice models: integrated framework, flexible error structures, and latent variables. Massachusetts Institute of Technology.

Walker, J. L., Ehlers, E., Banerjee, I. and Dugundji, E. R. (2011) 'Correcting for endogeneity in behavioral choice models with social influence variables', *Transportation Research Part A: Policy and Practice*, 45(4), pp. 362-374.

Walker, W. E. and Marchau, V. A. (2017) 'Dynamic adaptive policymaking for the sustainable city: The case of automated taxis', *International Journal of Transportation Science and Technology*, 6(1), pp. 1-12.

Wang, S., Charissis, V., Campbell, J., Chan, W., Moore, D. and Harrison, D. (2016) 'An investigation into the use of virtual reality technology for passenger infotainment in a vehicular environment', *2016 International Conference on Advanced Materials for Science and Engineering (ICAMSE). IEEE*. pp.404-407

Wong, R. and Szeto, W. (2018) 'An alternative methodology for evaluating the service quality of urban taxis', *Transport Policy*, 69, pp. 132-140.

Wong, R. C. P. and Szeto, W. Y. (2022) 'The effects of peak hour and congested area taxi surcharges on customers' travel decisions: Empirical evidence and policy implications', *Transport Policy*, 121, pp. 78-89.

Wong, R., Szeto, W. and Yang, W. (2020a) 'Customers' selections between premium electric taxis and liquefied petroleum gas taxis', *Transportation Research Part D: Transport and Environment*, 78, p. 102172.

Wong, R., Yang, L., Szeto, W., Li, Y. and Wong, S. (2020b) 'The effects of accessible taxi service and taxi fare subsidy scheme on the elderly's willingness-to-travel', *Transport Policy*, 97, pp. 129-136.

Xu, Z., Zhang, K., Min, H., Wang, Z., Zhao, X. and Liu, P. (2018) 'What drives people to accept automated vehicles? Findings from a field experiment', *Transportation research part C: emerging technologies*, 95, pp. 320-334.

Yap, M. D., Correia, G. and van Arem, B. (2016) 'Preferences of travellers for using automated vehicles as last mile public transport of multimodal train trips', *Transportation Research Part A: Policy and Practice*, 94, pp. 1-16.

Ye, Q., Law, R., Gu, B. and Chen, W. (2011) 'The influence of user-generated content on traveler behavior: An empirical investigation on the effects of e-word-of-mouth to hotel online bookings', *Computers in Human behavior*, 27(2), pp. 634-639.

Yin, H. and Cherchi, E. (2022) 'The impacts of in-vehicle features and social influence on the choice of autonomous taxis.', *101th Seminar on Transportation Research Board*. Washington DC, USA.

Zhang, D., Schmöcker, J.-D., Fujii, S. and Yang, X. (2016) 'Social norms and public transport usage: empirical study from Shanghai', *Transportation*, 43(5), pp. 869-888.

Zhang, T., Tao, D., Qu, X., Zhang, X., Lin, R. and Zhang, W. (2019) 'The roles of initial trust and perceived risk in public's acceptance of automated vehicles', *Transportation research part C: emerging technologies*, 98, pp. 207-220.

Zhang, W., Guhathakurta, S., Fang, J. and Zhang, G. (2015) 'Exploring the impact of shared autonomous vehicles on urban parking demand: An agent-based simulation approach', *Sustainable Cities and Society*, 19, pp. 34-45.

Zhao, X., Wang, L., Guo, X. and Law, R. (2015) 'The influence of online reviews to online hotel booking intentions', *International Journal of Contemporary Hospitality Management*, 27(6), pp. 1343-1364.

Zhou, F., Zheng, Z., Whitehead, J., Washington, S., Perrons, R. K. and Page, L. (2020) 'Preference heterogeneity in mode choice for car-sharing and shared automated vehicles', *Transportation Research Part A: Policy and Practice*, 132, pp. 633-650.

Zhu, H. and Huberman, B. A. (2012) 'To Switch or Not To Switch: Understanding Social Influence in Online Choices', *American Behavioral Scientist*, 58(10).

Zmud, J., Sener, I. N. and Wagner, J. (2016) 'Consumer acceptance and travel behavior: impacts of automated vehicles'. *Texas A&M Transportation Institute*.