Transportation Research Part C 163 (2024) 104628

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

Preferences for automated taxis. A comparison between immersive virtual reality and screen-based stated choice experiments☆

Hao Yin, Elisabetta Cherchi *

School of Engineering, Newcastle University, Stephenson Building, Newcastle upon Tyne NE1 7RU, UK

ARTICLE INFO

Keywords:
Immersive virtual reality
Stated choice experiment
Internal validity
Fully automated taxis

ABSTRACT

Automated Taxi (AT) services, as a promising business model for Autonomous Vehicles (AVs), have paved the way for novel mobility options. Understanding users' preferences for this innovative alternative is crucial for its success but proves to be challenging, as the lack of a real market mostly requires setting hypothetical situations to elicit respondents' preferences. Recent applications have seen an increasing use of Virtual Reality (VR), as a way to control the context, or to provide visual representation of some attributes. Research is however still in its infancy, in particular in the transport context, and results often are not comparable or show opposite effects. In this research, we aim to contribute to this limited research by studying the impact of the immersive VR environment in the preferences for AT elicited with Stated Choice (SC) experiments. Differently from previous studies, our experiment is built to ensure that the SC experiment in the VR-based environment is perfectly comparable with the standard SC screen-based, for internal validity. A control group is also used to test the order effect in the presentation of the two surveys, and to control for the carryover effect. Using data collected from a quite large sample, compared to the existing studies, joint mixed logit models are estimated allowing to assess the impact of the immersive VR environment on the choice bias, heterogeneity in the preferences for specific attributes, as well as panel effect, and order effect. Our results show that the immersive VR-based experience has no impact on the preferences for level-of-service attributes, travel time and cost, unless respondents can, to some extent, experience them, like waiting time in our study. On the other hand, the hints and cues provided by the immersive VR environment seem to affect the evaluation of the social aspects, descriptive norms and customers' reviews. This highlights a significantly high heterogeneity in the preferences, not revealed in the screen-based SC experiment. Finally, differently from previous literature, our results show no difference in the choice bias between VR- and screen-based environment, but only if differences in the preferences for specific attributes between the two environments are properly taken into account.

1. Introduction

A relatively vast literature has emphasised the potential of Automated Taxi (AT) services, as a promising business model for Autonomous Vehicles (AVs) and its prospective economic, environmental and social advantages (Fagnant and Kockelman, 2014;

☆ This article belongs to the Virtual Special Issue on "special issue full title".

* Corresponding author.
E-mail addresses: Hao.Yin@newcastle.ac.uk (H. Yin), Elisabetta.Cherchi@newcastle.ac.uk (E. Cherchi).

https://doi.org/10.1016/j.trc.2024.104628
Received 30 September 2023; Received in revised form 18 April 2024; Accepted 20 April 2024
Available online 9 May 2024

Lokhandwala and Cai, 2018; Lu et al., 2018; Zhang et al., 2015). The extent of benefits that ATs can bring to existing transport systems significantly relies on how successfully ATs can be in attracting customers to use this service. Understanding user preferences for this innovative travel mode is then a crucial step for their success. Traditional online Stated Choice (SC) experiments are commonly used to investigate users' preferences for AVs, as this is a product not yet available in the market (see the review in Gkartzonikas and Gkritza, 2019). Only few papers have used SC experiments to study preferences for AT-like systems. Krueger et al. (2016) elicited Willingness to Pay (WTP) measures for service attributes for shared AVs, defined as driverless taxi services. Yap et al. (2016) studied preferences for egress mode choices after train trips, where one alternative was cybercar (with automatic driving). Bansal and Daziano (2018) investigated the willingness to share a ride with strangers, where a dummy variable "automation" was used to denote if the Uber ride was with driver or driverless, while Yin and Cherchi (2022) still using an online SC experiment, studied AT preferences in China, with a focus on the WTP for in-vehicle features and social influence. Interestingly the estimated WTPs from these studies are very different, ranging for instance for travel time from £3.08/hr to £17.41/hr (and up to £20.00/hr if we consider AV, not strictly AT-like services). This is related to different contexts and different designs of the SC experiments, but it also raises a question about the accuracy of eliciting preferences for ATs and their characteristics when using conventional online SC surveys. SC experiments suffer from hypothetical bias (people's behaviours in reality deviate from what they state in the experiment) but have their strength in the quality of the estimated marginal utilities for the attributes included in the design (Louviere et al., 2000). This benefit of the SC is however now under scrutiny, with the increasing use of SC to test highly innovative products that respondents have no experience with or even knowledge of.

It is known that preferences elicited with SC experiments are strongly influenced by the SC framing. Pictures and videos have in fact been used with the aim to increase the accuracy of the preferences elicited (e.g., Iglesias et al., 2013). Their use has increased recently in the AV study (e.g., Howard and Dai, 2014; Kolarova et al., 2018) as a way to give participants some knowledge about the new products before presenting them with the SC experiment. In this line, recent applications have also seen an increasing use of Virtual Reality (VR), as a way to control the context, i.e., to improve the preliminary information provided about the (non-) market goods respondents were going to evaluate (Bateman et al., 2009; Fiore et al., 2009; Matthews et al., 2017; Patterson et al., 2017). Studies have shown that the use of immersive VR (instead of videos) can generate a sufficiently natural and familiar field, able to provide 'field cues' or 'field hints' that occur in the real world (Fiore et al., 2009). It allows the sensation of immersion in the activities prompting individuals to act as if they were in the real world (Animesh et al., 2011; Faiola et al., 2013; Nah et al., 2011; Sanchez-Vives and Slater, 2005). Based on these studies, the expectation is that immersive VR experiments should elicit preferences more accurately than standard screen-based surveys. However, very little is known in this area.

Recent works have used immersive VR technology to study continuous behaviours (also defined as motor actions, opposite to choice-based actions). Farooq et al. (2018) and Velasco et al. (2019) studied pedestrian crossing behaviour controlling for flow composition (different proportions of normal cars and AVs), speed of the cars, geometric characteristics of the street and weather condition. Arellana et al. (2020) studied the pedestrian behaviour as a function of the traffic flow and other geometric characteristics, but without AVs. Sobhani and Farooq (2018) studied the impact of distracted behaviours, Lovreglio et al. (2016) and Arellana et al. (2020) also studied the pedestrian behaviour in the case of evacuation while Feng et al. (2022) dealt with wayfinding behaviour in a multi-story building. Birenboim et al. (2019) and Bogacz et al. (2021) studied instead the cycling behaviour, still controlling for car flow and geometric characteristics. These studies use stated preference experiments to control for the characteristics of the environment (i.e., flow speed etc.) where the respondents walk or bike. From a neurological point of view, motor actions (like cycling or walking) activate different circuitries in the human brain compared to choice-based actions and show an overlap between brain activities during imagined and real movements (see discussion in Cherchi, 2020). This does not appear during choice-based actions. The use of VR for choice-based actions allows revealing effects that do not appear in VR applications for motor actions.

Few papers have compared immerive VR-based and screen-based SC experiments in the context of choice-based actions. Patterson et al. (2017) studied neighbourhood choices. Their SC experiment included two attributes that describe the building environment. Although the experimental design was the same in the VR and the text-based SC, they differed in the presentation of these characteristics. These were described textually in the screen-based SC (i.e., "20ft space between buildings"), while presented only visually in the VR environment where respondents evaluated that distance themselves based on the environment presented. They used a non-immersive VR experiment and a between-subject experiment (184 respondents for each treatment), though they controlled the distribution of the two samples in terms of gender, age, income, and current residential status. Differently from previous works in similar contexts, they found no significant differences between text-based and VR-based SC experiment, notably, in the evaluation of the visual attributes. They found though that in the model estimated with VR data, it was possible to estimate some segmentations, suggesting that the VR platform was better able to focus respondents' attention to process the text as well as visual attributes. They found an overall better model fit (adjusted rho-square) with VR data than with text-only data. They estimated separate models and did not control for scale differences.

Mokas et al. (2021) used a SC experiment to study the value of urban greenery (i.e., trees, bioretention planters) using three different presentation formats (text only, video and immersive VR). The same SC experimental design was used but, as in Patterson et al. (2017), the presentation of the three urban greenery attributes was conveyed through written description in the text only SC, while they were presented visually in the other two formats. They used a between-subject experiment, where a total of 180 respondents was randomly assigned to one of the three treatments (60 participants for each treatment). They estimated a model pooling all data, allowing to test for scale differences across experiment types, and estimated a full random parameter model but they did not test for differences in the preference for specific attributes. In line with previous research in this field, they found statistically significant differences in the scales, implying lower error variance (less randomness in making choices) in the video and the VR subsets compared to the text sample. At the same time, they found significant random heterogeneity in almost all the coefficients estimated, including price, which may account for differences in the preferences among experiment types, but this has not been examined.

Finally, Rossetti and Hurtubia (2020) studied the ecological validity of VR experiments (i.e., whether the results can be generalized to real-life settings), but the focus was on the qualitative assessment of aspects of an urban environment. They used three groups with a mix technique of between and within-subject. They recruited 251 participants in the streets for the assessment of a real environment, 192 participants in the university campus for the immersive VR experiment, and of these 443 people plus extra new participants, 153 participated in the second experiment with images. It is anyway interesting to report that they found that immersive videos are ecologically valid for perception of safety and security, as this depends more on visual cues to judge it, not for the quality of the built environment and the attractiveness of a certain location. It is also interesting to note that in their immersive VR it seems that participants did not move in the environment but responded to the experiment while seated.

The literature review shows that: 1) no studies in the VR field have focused on choices related to AVs, even less to ATs; 2) papers dealing with choice-based actions, do not deal with AVs and use the VR mainly to study visual attributes, otherwise described in words in the online surveys and 3) research so far showed opposite or non-conclusive results on key questions related to the impact of VR in eliciting preferences. In this paper, we aim to contribute to this limited literature by studying the impact of the immersive VR experience in the preferences for AT and its characteristics, controlling for internal validity. We use a choice context between normal and automated taxis, because this is the setting where differences between imagined and real movements occur and can then be detected with VR applications[1], differently from motor actions contexts. Following previous literature, because of the higher involvement and thus attention in the VR environment compared to online, we expect the choice error to be lower in the VR-based than in the screen-based SC experiment, however and differently from the previous literature we expect this difference to depend on the model specification and to be smaller if systematic heterogeneity is properly captured (H1). Since context and experience play a key role in shaping individual preferences, we expect preferences to be different whether elicited with the VR-based or screen-based SC experiment (H2). However, and more important, we do expect the impact of the VR to be stronger on those attributes more difficult to measure (such as those capturing for example social influence) compared to Level of Service (LOS) attributes (times and costs) (H3), unless participants can experience the LOS attributes as part of the VR experiment (H4). Last not least, another important hypothesis is that we expect the VR-based SC experiment to be able to capture this heterogeneity better than the screen-based SC experiment, because of the hints and clues provided by the realistic (though virtual) environment that provide different and more stimuli to our brain (H5).

To test these hypotheses, a SC experiment was built and implemented both screen-based (online) and within an immersive VR environment. The SC experiment consists of a series of binary choices between a normal taxi with the driver (NT) and a fully automated taxi without driver and without steering wheel (AT). Differently from the existing literature on choice-based actions 1) all attributes in the stated choice experiment are presented in the same way in the VR-based and in the screen-based surveys, 2) the VR environment does not change across stated choice scenarios, and it is used to test the "living effect", i.e., to make respondents feel that they are living the choice. The advantage of this setting is that the SC experiment in the VR-based environment is perfectly comparable with the standard SC screen-based, allowing to disentangle the impact of the immersive VR experience in the elicited consumer's preferences compared to the screen-based standard SC experiment (internal validity). In addition, in our experiment 3) the same respondent replied both surveys (screen-based and VR-based), which allows a full control of the effects tested[2] and a control group was used to test the order effect in the presentation of the survey (i.e., VR first, online after) and the carryover effect implicit in the within-subject samples. The sample used (208 participants) is one of the largest in the field for non-university within-subject immersive VR experiment.

The paper is organised as follows: Section 2 describes the survey methodology, namely the key aspects of the immersive VR environment and the SC experiment built. Section 3 describes the data collection process and the main characteristics of the samples. Section 4 reports the modelling framework and Section 5 includes a discussion of the modelling results. Section 6 summarises the key conclusions from this research and the comparison between VR-based and screen-based surveys.

2. Survey design and data collection methodology

The core of the methodology set up in this research consists of a SC experiment built to elicit preferences for fully ATs without driver and without steering wheel versus NTs with driver. The same SC experiment was used to collect data online (screen-based) and within an immersive VR environment (VR-based). In the immersive VR experiment, respondents found themselves into a street where there is a taxi rank with two types of taxis (with and without driver) picking up customers, and a ticket board (as it is the case in reality) where to select and pay for the taxi that the respondents wish to use. Depending on the destination selected, the ticket board presents two options: one for the normal and one for automated taxi. This is a task of the SC experiment, which is exactly the same (also in terms of presentation) as that used in the standard screen-based SC. The SC experiments in the immersive VR environment and in the screen-based are identical and this allows us to study the impact of the immersive VR experience in the elicited preferences without confounding effects. While the SC experiment is the same in the screen-based and the VR-based experiments, respondents did not

[1] "Innovative methods that make stimuli more lifelike (e.g., virtual reality, or bidirectional social interactions) are likely to make evidence from even hypothetical choices a better guess about what mechanisms are involved in real choice" (Camerer and Mobbs, 2017).

[2] After carefully balancing the pros and the cons, we decided to use the within-subject samples, as it requires smaller samples than between-subject samples to get the same statistical power and does not require controlling for differences in individual characteristics between samples that can affect internal validity.

encounter exactly the same choice situations because the choice tasks presented depend on the destination selected by the respondents during the experiment, and the destination could actually vary between the screen-based and the VR-based experiment, within each experiment and across different choice tasks.

2.1. Context and virtual reality environment

The context for the experiment is Northumberland Road, a well-known street in the city centre of Newcastle upon Tyne, in the northeast of England, adjacent to the pedestrian high street, the focal point for business, especially shopping, well-known by the locals. Great effort was put in building the immersive VR environment with high fidelity. The objective was to ensure that the environment was highly realistic to include all the field-cues presented in reality, but without additional unintended information that could affect respondents' choices. The only new element, compared to the real environment was the presence of ATs that had, however, the same shape of normal cars. We used the black standard car for NTs, as this is the colour of the taxis that operate at the taxi ranks in Newcastle. We used the red colour for ATs. A possible colour effect in the choice was carefully discussed and tested. It was found that 63% of the participants did not notice the colour of the cars[3].

Fig. 1 shows a picture of Northumberland Road where the taxi rank of our experiment is located. The left figure is a picture of the real street, while the right figure is its reproduction in the virtual reality environment. In the VR, the street was slightly re-designed in order to accommodate the two types of taxis and make it possible for respondents to clearly see and appreciate the characteristics of each type of taxi and how the taxis operate when making their choices. These images in Fig. 1 were also used in the screen-based survey to present respondents with the choice context, accompanied by a typical text explaining the context of the experiment.

In the immersive VR environment respondents make the choice of the taxi using a *Ticket Board* that simulates a touch-screen board, which is activated 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. The ticket board is illustrated in the left part of Fig. 2. It includes a wide screen where the characteristics of the taxis available (the SC choice tasks) were presented and on its right, it reports the features to insert banknote and coins, to tap the credit cards and to collect ticket or money change, which is consistent with the real ticket boards. A registered voice was also included to accompany all information presented in the ticket board where respondents need to take an action (such as "select your destination", or "select your payment method"). The exact same information, without the voice, was presented also in the screen-based survey.

The immersive VR environment was built in a way that respondents could see how the system operates and walk around the moving vehicles. However, still some preliminary information was needed (1) about AV safety, privacy and routing information and (2) how the AT operates once on board. This information was included in an *Information Board* and presented to all respondents allowing to normalise the initial conditions. The right part of Fig. 2 shows an example of the information board in the VR experiment. The same information was presented, with exactly the same graphics, also in the screen-based survey. More details are reported in Yin and Cherchi (2022).

2.2. Stated choice experiment

The SC experiment consists of a series of binary choices between a NT and an AT. The attributes included in the SC experiment were defined based on the literature and a series of focus groups. The final experiment includes six attributes. Three level of service attributes (waiting time, travel time and fixed journey fare, with three levels each), one attribute to measure the impact of fuel type (with two levels: electric or petrol/diesel) and two attributes to measure the impact of social conformity (number of customers with three levels and customer rating with two levels). The adoption rate measures if individual behaviour is affected by what other people do (descriptive norm). Customer rating yesterday, represents the general public opinion or a form of word-of-mouth, which has been widely studied in marketing but rarely in transport.

A heterogeneous Bayesian D-efficient design was generated in Ngene (ChoiceMetrics, 2012). Priors were taken from models estimated in several pilot tests based on orthogonal designs. Three SC experiments were optimised based on travel distances of 5 km, 10 km, and 15 km to customise the designs with different levels for travel cost and travel time. 16 scenarios were generated for each design and randomly divided in two blocks. Each respondent was presented with eight scenarios, which could belong to any of the three designs (5 km, 10 km or 15 km) depending on the destination selected before each scenario. Fig. 3 shows an example of the SC experiment presented in the immersive VR-based experiment. The same SC experiment with the same format was also used in the screen-based version. Table 1 shows the attributes and attribute levels included in the SC experiment.

The process followed to build the SC experiment is not different from the standard process typically followed when building a SC. However, some adjustments were required to ensure the SC experiment was perceived as realistic in the immersive VR. This is because some approximations typically used and considered perfectly acceptable in the screen-based SC looked unrealistic when used in a VR-based environment. In particular, for the purpose of this paper, the following aspects are relevant:

1) Definition of the choice context

[3] The small test carried out does not completely rule out that the colour "red" might exert unconsciously attraction, but it gives confidence that there is not a systematic bias. At the same time, if the colour exerts an unconscious attraction, this should affect both the screen-based and the VR-based experiment, because the taxis have the same colour in both experiments.

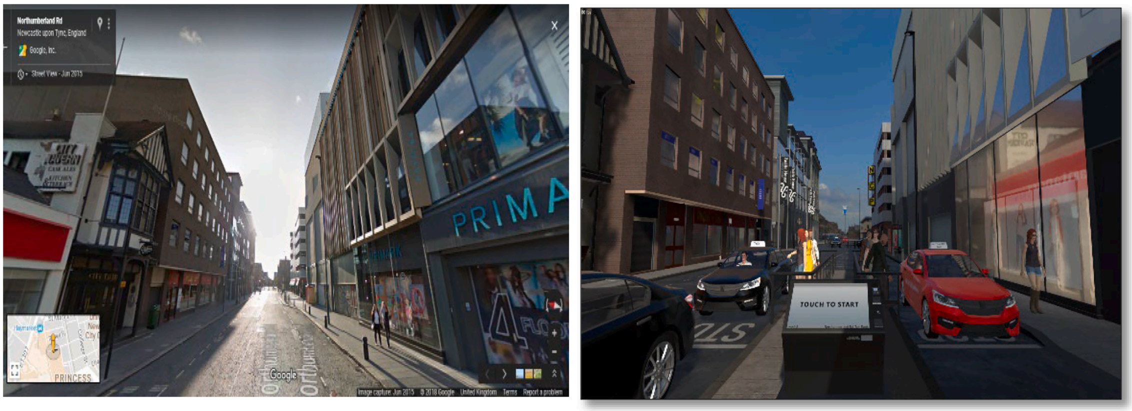

Fig. 1. The taxi rank area: real and virtual environment.

Fig. 2. Ticket and information board in the immersive VR-based SC experiment.

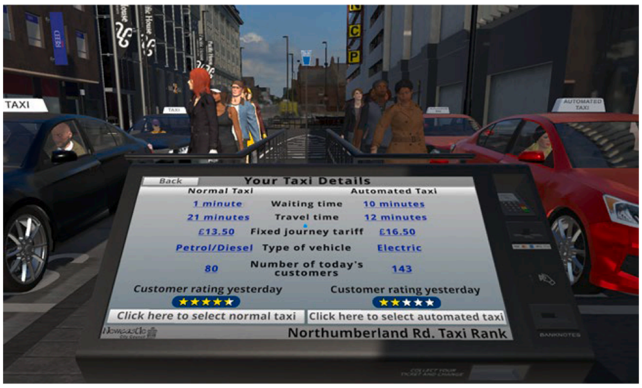

Fig. 3. Example of choice task in the VR-SC experiment (10 km segment).

It is typically recommended to customise the SC experiment based on a recent trip made by each respondent; in our case it would be a trip made by taxi. In a screen-based SC experiment, there are no specific constraints on the trips by taxi that can be considered, in terms of origin and destination, and booking methods (taxi rank, online, or by phone). Even if pictures or videos of automated taxis were presented to respondents, these do not have to refer to specific locations. Indeed, images or videos often refer to imaginary cities. In the case of VR-based SC experiments, however, a precise context needs to be defined, because respondents will make their choice within the virtually real environment. Not all contexts are feasible because the VR requires the experiment to take place in a specific physical location, where passengers have the possibility to see the taxis operating while they are making their choice. This is why we chose to locate the experiment at a taxi rank. Traditional taxi services are still extensively used in Newcastle upon Tyne and there are numerous taxi ranks in the city centre. The taxi rank located in Northumberland Road (the location chosen for the experiment) is used

Table 1
Attributes and attributes levels in the SC designs.

Attributes	Attribute levels		
	Short (5 km)	Medium (10 km)	Long (15 km)
Waiting time [Minutes]	1/6/10		
In-vehicle travel time [Minutes]	6/10/14	12/16/21	18/22/27
Fixed journey fare [GBP]	5.5/7/9	11.5/13.5/16.5	18/20.5/24
Fuel type for the AT	Petrol-Diesel / Electric		
Number of today's customers	17/80/143		
Customer rating yesterday	Bad Reviews (2 Stars) / Good Reviews (4.5 Stars)		

by more than 1,000 people per week.

2) Experimental design and customisation

In standard SC experiments, when customisation is used, respondents are typically instructed to assume that they have to make a trip with the same destination of the trip described as the most recent, and then all the scenarios presented refer to a trip with the same destination. Interestingly, this standard procedure that sounds perfectly reasonable in the screen-based SC experiment, it appeared extremely unrealistic in the VR environment. This is because in the VR environment respondents "live" the choice process, any assumption that they have to make the same trip felt unrealistic. In the several preliminary tests ran, it was discovered that presenting the choice tasks in sequence, like it is always done in the screen-based survey, was extremely unrealistic in the VR. Differently from the standard practice, respondents were then presented with a list of destinations at beginning of each scenario and were asked to select their destination before each scenario. Additionally, a short 'break' section was provided between scenarios. For consistency, the same procedure was adopted also in the screen-based SC survey.

2.3. Sample recruitment and overall survey

The VR-based SC experiment and its screen-based counterpart were administrated in Newcastle upon Tyne, UK, between 2022 and 2023. All participants were locals and knew Northumberland Road where the experiments took place. All participants did both experiments (within-subject comparison). 156 participants responded the screen-based first and after 1–2 weeks came to the lab and participated in the VR-based SC experiment. A total of 1248 observations were collected. To control for order effect, a fresh control group sample of 52 respondents (416 pseudo-observations) was collected and did the experiment in the reversed order (first the VR-based and then after a couple of months the screen-based). The control group had a much longer interval between experiments because of the higher engagement of the VR experiment, meaning that it is more likely that participants have a stronger recollection of it

Table 2
Socio-economic characteristics.

Socio-demographic characteristics		**Main sample** (N = 156)	**Control group** (N = 52)
	Female	52.6	58.0
Gender	Male	46.8	42.0
	Rather not to say	0.6	0.0
	18–29	19.9	16.0
	30–39	22.4	30.0
Age	40–49	22.4	16.0
	50–59	20.5	18.0
	60 or above	14.7	20.0
	Secondary school and below	11.5	20.0
	High school or equivalent	27.6	38.0
Education level	College degree	26.9	32.0
	Bachelor's degree	24.4	10.0
	Master/Doctorate degree	9.6	0.0
	Employed	76.9	80.0
Current work status	Self-employed	7.7	4.0
	Retired	8.3	8.0
	Others	7.1	8.0
	Less than £500	12.8	12.0
	£501-£1500	40.4	48.0
	£1501-£2500	21.8	22.0
Personal monthly disposable income	£2501-£3500	9.0	8.0
	£3501-£4500	2.6	4.0
	> £4500	0.0	0.0
	I do not wish to disclose it	13.5	6.0

compared to the online experiment.

A rigorous protocol was followed in the VR-based SC experiment (both in the main experiment and in the control group), which includes two tutorials, one general about how an immersive VR environment works and one specific to familiarise with the virtual Northumberland Road. These tutorials are particularly important as aimed to ensure that all respondents had enough capabilities to complete the choice tasks in the VR environment, and to remove what in the VR terms is called "wow effect", i.e., the high excitement that participants express the first time they found themselves in a virtual environment. A detailed description of the methodology followed to build the immersive VR environment and the VR-based SC experiment for the specific context of reference can be found in Yin and Cherchi (2024). The screen-based online SC experiment lasted around 15 minutes, while the VR-based lab session lasted between 40 and 60 minutes. This latter time includes the actual experiment, the tutorial, the breaks in between sessions to allow respondents to rest if needed and the various formalities, such as privacy information and consent forms. It is worth mentioning that none of the respondents complained about the length of the VR experiment, as it seems that the subjective perception of time speeds up when respondents are highly engaged in the VR experiment.

The overall survey consisted of the following parts: general questions about familiarity with AVs, information about a last trip performed by normal taxi, a SC experiment, a set of socioeconomic information and a set of statements to measure latent psychological constructs. All respondents interviewed were residents in the northeast of England and satisfied the requirements to be 18 years or older and have used a normal taxi in Newcastle in the last three years. Participants were recruited via a panel and received £45 each for their participation.

3. Samples characteristics

Tables 2 and 3 report a summary of the key characteristics of the sample collected. Our sample approximates the gender distribution of the Newcastle population but slightly underrepresents young people (20% in the main sample and 16% in the control group, against 25% in the Newcastle population). This difference can also be due to the recruitment screening questions. It can be that young people use taxis less often in real life, but there are no data to verify this information.

Although the size of our sample is one of the largest[4] among the studies using immersive VR, we did not aim for a representative sample. The characteristics of the control group present some differences with respect to the main sample and this will be controlled for in the modelling. In particular, respondents in the control group are mostly on their thirties or above 60 and have a lower level of education.

As discussed previously, travelling by taxi is a popular option in Newcastle, around 40% of the sample use the taxi at least once a month and around 50% of the sample talk with the taxi driver frequently and enjoy the interaction with the taxi driver, i.e., talking and/or getting help with luggage. As showed in Table 3, the majority of the sample have heard of AV (less in the control group) and are not relatively familiar with the 5 levels of automation. 1/3 of the sample had the experience with VR, though, upon investigation, only with video games. All participants commented that they had never seen anything like what presented in our experiment.

4. Modelling framework

A joint mixed logit choice model was used to estimate the impact of the immersive VR in eliciting preferences for ATs, versus a screen-based survey. The model specified allows in particular to estimate the following key features: 1) different bias in the choice between the immersive VR and the screen- based; 2) different bias in the choice depending on the order effect between the two treatments (main survey and cotrol group); 3) different preferences for the taxi characteristics, between the immersive VR and the screen based, beyond scale differences; 4) intra-respondent correlation and 5) systematic heterogeneity and non-linearity in the preferences for the taxi characteristics.

The model has the following expression:

$$\begin{aligned}
U_{jqt}^{VR.MS} &= \left(V_{jqt}^{VR.MS} + \sigma_j^{VR.MS}\nu_{jq}^{MS} + \varepsilon_{jqt}^{VR.MS}\right) \\
U_{jqt}^{Onl.MS} &= \lambda^{Onl.MS}\left(V_{jqt}^{Onl.MS} + \sigma_j^{Onl.MS}\nu_{jq}^{MS} + \varepsilon_{jqt}^{Onl.MS}\right) \\
U_{jqt}^{VR.CG} &= \lambda^{CG}\left[\left(V_{jqt}^{VR.CG} + \sigma_j^{VR.CG}\nu_{jq}^{CG} + \varepsilon_{jqt}^{VR.CG}\right)\right] \\
U_{jqt}^{Onl.CG} &= \lambda^{CG}\left[\lambda^{Onl.CG}\left(V_{jqt}^{Onl.CG} + \sigma_j^{Onl.CG}\nu_{jq}^{CG} + \varepsilon_{jqt}^{Onl.CG}\right)\right]
\end{aligned}$$

Where, U_{jqt} is the utility that individual q assigns to alternative j = [NT, AT] in choice task t = [1, 2, …, 8]. The superscripts *VR* and *Onl* refer to the survey environment, *VR* means that the utility is for the alternatives evaluated in the immersive VR-based survey, *Onl* for the screen-based (Online) survey. The superscripts *MS* and *CG* refer to the instrument to test order effect, *MS* stands for *main sample*, it means that the order in which participants replied to the SC is screen-based first and immersive VR-based second; *CG* stands for *control group*, and it means that the order in which participants replied to the SC is immersive VR-based first and screen-based second.

Three scales are estimated. Within each sample $\lambda^{Onl.MS}$ and $\lambda^{Onl.CG}$ are the scales of the screen-based data, normalised by the scale of

[4] Complementing the discussion in the literature review, for example, Bogacz et al. (2021) used 48 respondents; Feng et al. (2022) 36 respondents and Zou et al. (2021) 28 participants.

Table 3
Travel characteristics and Knowledge of AVs and VR.

Trip characteristics		**Main sample** (N = 156)	**Control group** (N = 52)
Travel time most recent trip by taxi	Short (around 10 min. or less)	46.2	38.0
	Medium (around 20 min.)	39.7	42.0
	Long (around 30 min or more)	14.1	20.0
Frequency of using taxis	At least once a week	10.3	16.0
	Less than once a week, at least once a month	42.3	32.0
	Less than once a month, more than twice a year	39.7	38.0
	At most twice a year	7.7	14.0
Frequency of talking with the driver	Very infrequently	8.3	2.0
	Somewhat infrequently	8.3	6.0
	Occasionally	30.1	32.0
	Somewhat frequently	36.5	34.0
	Very frequently	16.7	26.0
Enjoying talking with taxi driver	Always	16.0	18.0
	Sometimes	79.5	82.0
	Never	4.5	0.0
Like the driver to help carry luggage	Yes	64.1	70.0
	No	35.9	30.0
Take the taxi without any help	Yes	98.7	100.0
	No	1.3	0.0
Where taxis?	At the taxi rank in Northumberland Road	48.1	38.0
	At other taxi ranks in Newcastle but not in Northumberland Road	41.0	46.0
	Via taxi app or pre-book by phone	10.9	16.0
Knowledge of AV or VR			
Heard of AVs	Yes	69.2	54.0
	No	30.8	46.0
Familiarity with the 5 levels of automation	Not at all familiar	39.7	46.0
	Slightly familiar	35.3	30.0
	Moderately familiar	19.9	20.0
	Very familiar	4.5	4.0
	Extremely familiar	0.6	0.0
VR experience	Yes	35.9	36.0
	No	58.3	62.0
	Missing	5.8	2.0

the immersive VR-based data. $\lambda^{\cdot CG}$ is the scale of the control group compared to the main survey data.

$\sigma_j\sigma$ is the standard deviation of the error component terms that account for intra-individual correlation (or panel effect in the SC experiments). These are alternative-specific because within each environment and treatment, respondents have only two alternatives to choose from. We allowed this variance to be different between environments (*VR* and *Onl*) and between treatments (*MS* and *CG*), letting the results to inform if there are significant differences or not, as discussed in the results section. ν_{jq} is a random value from the standard normal distribution, that is the same between environments because in each treatment the same individual q participated in both the VR- and the screen-based experiments.

ε_{jqt} is instead the white noise, distributed EV1 independently and identically across alternatives and observations.

V_{jqt} is the portion of the overall utility U_{jqt} that depends on observable characteristics and allows accounting for different preferences for the taxi characteristics, between the immersive VR and the screen-based, beyond scale differences. It accounts also for systematic heterogeneity and non-linearity in the preference for the taxi characteristics. For simplicity of notation, let S be an indicator for the combination of survey environments (*VR, Onl*) and treatments (*MS, CG*). The utility takes the following form:

$$V_{jqt}^{S} = ASC_j^{S} + \beta_{j,X}^{S} f(X_{jqt}) + \beta_{j,SE}^{S} SE_{\mathrm{q}} + \beta_{j,X.SE}^{S} X_{jqt} SE_{\mathrm{q}}$$

Where:

ASC_j^S is the alternative specific constant for the alternative j; it takes value 1 if j = AT, 0 otherwise, and it varies among survey environments and treatments;
X_{jqt}^S is a vector containing all attributes used to define the alternatives in the SC experiment;
SE_q^S is a vector of socio-economic characteristics;
β^S are vectors of coefficients that measure the marginal effect of SC attributes, socio-economic characteristics, and their interactions. These are tested specific for survey environments and treatments.

Table 4
Models estimation results.

		Treatment			
		Estimated values		Rob. *t*-test	
	Environment	*Main survey*	*Control group*	*Main survey*	*Control group*
Level of Service					
Travel cost x income[1] ≥ £2,500	*Screen & VR*	−0.535		−8.45	
x income < £2,500	*Screen & VR*	−0.840		−8.19	
squared x income < £2,500	*Screen*	0.009	0.012	3.38	3.28
squared x income < £2,500	*VR*		0.019		6.12
Travel time	*Screen & VR*	−0.089		−5.97	
x more than very familiar with AVs	*Screen*	0.067		2.40	
x using taxis less than once a week, at least once a month	VR	−0.057		−2.04	
Waiting time	*Screen*	−0.124	−0.143	−6.54	−4.73
	VR	−0.162	−0.118	−10.06	−4.21
x work (purpose at destination)	*Screen & VR*	−0.149		−2.25	
EV **feature**					
Electric vehicle (AT)	*Screen*	0.299	0.926	1.55	2.60
	VR	0.503	–	2.92	–
Social conformity					
Good review	*Screen*	0.397	–	3.20	–
	VR		–		–
x bachelor max education	Screen & VR	0.583	−0.695	2.85	−2.69
x high school max education	VR	0.534	–	2.38	–
x slightly familiar with AVs	VR	−0.383	−0.815	−2.13	−2.66
x talk to taxi driver very frequently	VR	−0.296	–	−1.44	–
x usually take taxi at other taxi ranks in Newcastle but not in Northumberland Road	*Screen*	−0.515	–	−2.33	–
x using taxis less than once a month, more than twice a year	VR	0.412		2.38	
Number of today's customers/10	*Screen*	0.467	−0.426	2.21	−1.16
x no experience with VR	Screen	−0.850	–	−2.76	–
x take taxi at any taxi ranks	*Screen & VR*	−0.573		−3.20	
Alternative-specific effects					
ASC (AT)	*Screen*	−0.338	−1.100	−1.22	−1.94
	VR	−0.350	0.574	−1.63	2.60
Female (NT)	*Screen & VR*	0.944		3.90	
Age between 18 and 34 years (NT)	Screen	−0.581	–	−1.97	–
Not at all or slightly familiar with AVs (NT)	Screen	−0.844		−2.71	
Scale effects					
Scale screen-based data		0.920	1.390	7.02	3.72
(VR-based normalised)				[-0.61]	[1.04]
Scale control group			0.887		6.24
(main survey normalised)					[-0.79]
Panel effects					
SIGMA (AT)	*Screen*	1.790	2.000	6.86	3.87
	VR	1.500	0.702	6.53	2.60
Summary of Statistics					
Number of draws	500				
Maximum log-likelihood	−1532.40				
Log-likelihood for the Market Share	−2243.84				
Akaike Information Criterion	3150.80				
Bayesian Information Criterion	3413.53				
Number of individuals		156	52		
Number of pseudo-observations		2496	832		

(1) Monthly net income, after taxation, mortgage, living and other necessary expenses, excluding transport expenses.

For each treatment, the individual likelihood is given by the integral of the conditional probability to choose the sequence of alternatives $\mathbf{j} = \{j_1, \ldots j_8\}$ over the distribution of the random term ν_{jq}:

$$L_q(q \in MS) = \int_{\nu^{MS}} \left(\prod_{t=1}^{T} \frac{\exp\left(V_{jqt}^{VR.MS} + \sigma_j^{VR.MS}\nu_j^{MS}\right)}{\sum_{i \in A_q} \exp\left(V_{iqt}^{VR.MS} + \sigma_i^{VR.MS}\nu_i^{MS}\right)} \cdot \frac{\exp\lambda^{Onl.MS}\left(V_{jqt}^{Onl.MS} + \sigma_j^{Onl.MS}\nu_j^{MS}\right)}{\sum_{i \in A_q} \exp\lambda^{Onl.MS}\left(V_{iqt}^{Onl.MS} + \sigma_i^{Onl.MS}\nu_i^{MS}\right)} f\left(\nu^{MS}\right) d\nu^{MS} \right)$$

$$L_q(q \in CG) = \int_{\nu^{CG}} \left(\prod_{t=1}^{T} \frac{\exp\lambda^{CG}\left(V_{jqt}^{VR.CG} + \sigma_j^{VR.CG}\nu_j^{CG}\right)}{\sum_{i \in A_q} \exp\lambda^{CG}\left(V_{iqt}^{VR.CG} + \sigma_i^{VR.CG}\nu_i^{CG}\right)} \cdot \frac{\exp\lambda^{Onl.CG}\lambda^{CG}\left(V_{jqt}^{Onl.CG} + \sigma_j^{Onl.CG}\nu_j^{CG}\right)}{\sum_{i \in A_q} \exp\lambda^{Onl.CG}\lambda^{CG}\left(V_{iqt}^{Onl.CG} + \sigma_i^{Onl.CG}\nu_i^{CG}\right)} f\left(\nu^{CG}\right) d\nu^{CG} \right)$$

The unconditional log-likelihood for the entire sample is given by:

$$LL = \ln\left(\prod_{q \in MS} L_q(q \in MS) \prod_{q \in CG} L_q(q \in CG) \right)$$

The joint models were estimated by maximum simulated likelihood estimation, using PythonBiogeme (Bierlaire, 2016).

5. Modelling results

Table 4 reports the results of the joint mixed logit model estimated using a final sample of 3328 pseudo-observations (156 individuals in the main survey plus 52 in the control group replied each 8 scenarios in the screen-based and 8 in the VR-based survey). A standard process, as described in Cherchi and Ortúzar (2002), was used to identify the joint model reported in Table 4. Separate models using only VR-based and only screen-based data were first estimated to identify the set of coefficients candidate to be generic between experiment types. Simple models without interactions were estimated first and then systematic heterogeneity and no-linearities tested. In the joint estimation, t-tests for generic coefficient were used to decide which coefficients were constrained to be generic between experiment types.

All the variables measured and listed in Tables 2 and 3, where tested in the model. Table 4 reports only the effects that were significant at less than 25% significance level. To help the comparison of the results between the screen-based and the VR-based SC experiment, as well as between the main survey and the control group, the first two columns of Table 4 report the estimated mean values from the joint model and the last two columns their t-tests. When two values are reported, one in the first and one in the second column, it means that the coefficient is specific between the main survey and the control group. Otherwise, a unique value is reported for both surveys. The environment (screen or VR) refers to whether the value of the coefficient reported in the table is estimated specific (two values are reported) or generic (only one value is reported) between the screen-based and the VR-based samples.

Several interesting results can be drawn from the model estimated. These are discussed in detail hereafter.

5.1. Unobserved heterogeneity

We start with the "white" choice error, which is a measure of the error associated to the choice. Because of the higher involvement and thus attention in the VR environment compared to online, previous studies (e.g., Mokas et al., 2021) have found that the VR-based experiment elicits choices with less uncertainty than the screen-based SC experiment. The "white" choice error is inversely related to the model scale, and we observe in Table 4 that none of the three scales estimated is significantly different from one, indicating that in our study the choice error in the immersive VR is not different from that in the screen-based SC experiment. The reason could be due to the fact that in our SC experiment all attributes are presented exactly in the same way in both environments (VR and screen-based). Different choice bias reported in the literature could be related to different ways of presenting some attributes (text versus visual), rather than to the impact of the immersive VR environment. Another reason could be that differences in the choice bias can be due to differences in the preference for specific attributes, as in our first hypothesis (H1). We tested a specification where we constrained some of the coefficients to be the same between the screen-based and the immersive VR environment. We found that in this case the choice error in the screen-based was higher than in the immersive VR-based data and indeed significantly different between them (t-test = 2.12).

Regarding the order effect, results show that there is no difference in scale between the main survey and the control group, which is reassuring that potential differences are accounted for with measurable variables.

Looking at another unobserved error, results seem to suggest that the immersive VR reduces random choice heterogeneity, but mainly when respondents see the experiment first in the immersive VR-based environment. Results show that the panel correlation is not significantly different between the immersive VR and the screen-based SC experiment in the main survey (t-test = 1.10), while it is significantly different in the control group (t-test = 2.59), with random choice heterogeneity across individuals higher in the screen-based than in the immersive VR-based survey.

5.2. Level of Services

Since our SC experiment in the VR-based environment is perfectly comparable with the standard SC screen-based, it is possible to disentangle the impact of the immersive VR experience in the elicited consumer's preferences for Level of Service (LOS) attributes compared to the screen-based standard SC experiments (internal validity). In line with previous studies (e.g., Patterson et al., 2017), our results confirm that the immersive VR environment does not have an impact on individual preferences for LOS attributes. Our hypothesis (H2), that the immersive nature of the VR environment would have an impact on the preferences elicited on all attributes, is rejected because the marginal utility of the LOS attributes is not significantly different between the two environments (online and VR), except for waiting time and some socio-economic groups. In particular, for the majority of the sample the marginal utility of travel time is the same in both environments (VR and screen-based), with the exception of those who are very or extremely familiar with AVs, who value a change in travel time significantly less (in absolute terms, it is −0.022 versus −0.089) in the screen-based than in the VR-based environment, maybe because not living the choice they might have erroneously been misled by the knowledge that general AVs allowed a more efficient use of travel time. Those who use taxi moderately (less than once a week at most once a month) instead value a marginal variation in travel time almost doubled when elicited within the VR environment. The initial expectation was that frequent taxi users should be able to evaluate the disutility of travel time also in the screen-based surveys, because they are familiar with taxis. However, it seems that the opposite is happening. It could be that, because of familiar with taxis, they underestimate the disutility of travel time in the screen-based survey. There are no previous studies on this specific aspect for comparison. An external validity test is needed to confirm this effect. However, to put the results into context, for a short trip of 7.00 GBP, we found WTP for in vehicle travel time ranging from 7.50 GBP/hour for low-income people, not familiar with AT and frequent taxi users, to 17.50 GBP/hour for high-income people, familiar with AT. This is a bit higher, but still within the range reported in previous works in the field, as mentioned in the Introduction section.

The marginal utility of travel cost is not different between VR-based and screen-based either, but it shows a significant no-linear effect. In particular, the marginal utility is higher for low income people and in this group it also decreases with travel cost, suggesting the present of income effect (Jara-Díaz and Videla, 1989). This effect is the same between environments and treatments, though the impact of the squared cost term is doubled in the control group, which probably reflects the slightly higher proportion of respondents in the low-income groups (82% in the control group, 75% in the main survey).

Finally, as expected, even if respondents in the immersive VR do not actually wait to take the taxi (it would have significantly increased the time of the overall experiment in the lab) among all the LOS attributes, waiting time is perceived more negatively in the immersive VR environment and it is significantly different from the preference elicited in the screen-based environment (t-test = 2.04, the hypothesis that the marginal utility of waiting time is the same in the immersive VR and online is rejected at more than 99%). This result confirms our hypothesis (H4) that the VR environment has an impact only on the attributes that the participants can experience. In the immersive VR environment, respondents can see other customers waiting for their taxi, and waiting time would happen in the same location where the choice is taking place. The immersive VR environment allows participants to visually and spatially appreciate the implications of waiting, making the experience more salient and impactful on their decision-making, unlike the screen-based environment, where waiting time is not real and removed from a tangible context. However, interestingly, the impact of the experience with a particular aspect of the alternative (waiting time in our case) seems to be affected by the order in which participants see the experiment (order effect). Our results show that the marginal utility for waiting time is lower (in absolute values) the first-time participants reply to the survey, whatever environment comes first.

5.3. Social conformity and EV feature effects

Interestingly, when we look at the other attributes studied in the SC experiment, namely social conformity attributes and EV feature, results are much less "regular" and show significant heterogeneity among groups and significant differences between immersive VR-based and screen-based SC as well as significant differences between treatments. As we can see in Table 4, the marginal utility of the attribute EV, as well as almost all the marginal utilities of the social conformity attributes are significantly different between the VR-based and the screen-based environments. This confirms our hypothesis (H3) that the impact of the VR environment is stronger on those attributes more difficult to measure. This can be either attributes whose evaluation does not imply awareness, but unintentional impact exerted by action of others (as the social conformity attributes in our study) or visual attributes (as tested in Patterson et al., 2017; Mokas et al., 2021; Rossetti and Hurtubia, 2020).

Among the social conformity effects, the reviews from previous customers are no doubt the most significant. Interestingly, it shows a great deal of heterogeneity among groups and this heterogeneity is elicited almost only in the immersive VR-based survey, in line with our hypothesis (H5). In particular, those with a high school or bachelor education value good reviews almost twice as much as the rest of the sample. Similar effect for those who use taxi occasionally, who rely more on the reviews to compensate the little personal experience with the taxis. In both cases the hypothesis of homogeneity among these groups is rejected at more than 99%). Those who talk frequently to the taxi driver have a positive preference for good reviews, but smaller preference than those who do not talk to the driver. This is plausible, as they probably rely more on getting information from the taxi driver rather than general customers do. As expected, those who are slightly familiar with AVs have a positive marginal utility for good reviews (significantly different from the remaining sample at 99%), though closed to zero. As we said, the majority of these effects are elicited only in the immersive VR environment, confirming our hypothesis (H3) that the immersive VR environment provides more hints that seem to affect in particular the evaluation of these social aspects. Patterson et al. (2017) also found a couple of segmentations could be estimated only in the model estimated with VR data. Summarising, this seems to suggest that the higher realistic level of stimulus provided by the immersive VR generated much more diverse responses among participants and also makes it more difficult for the participants to focus on the social information provided. This effect is much more marked in the control group when the immersive VR was the first experiment performed.

We found a similar effect on the other social conformity attribute. The number of today's customers who have used the taxi is significant and has a positive correct sign only in the screen-based survey and only when this is the first experiment seen by the participants (i.e., in the main survey), but not for the entire sample. In the screen-based survey is significant and negative for those who have not experienced VR before and those who used to take taxis at any taxi ranks in Newcastle, while it is not significant at all in the immersive VR experiment except for those who used to take taxis at any taxi ranks (with a negative impact on the utility). While there is strong evidence in the literature that people tend to conform to what other people do or say, there is also evidence (Allen, 1965; Stricker et al., 1970; Willis, 1965) that some people do also show anti-conformity behaviours, i.e., do the opposite of what other do or what they are expected to do. The negative impact of the number of customers might indicate the presence of anti-conformity behaviour. For example, since the VR is an innovative tool, those who have not tried VR before coming to the lab, might be less pro-innovation and less pro-conformity, which could explain the negative sign for the descriptive norm attribute. However, more research in this area is needed to confirm this hypothesis.

Finally, the result for the attribute EV (presented as a dummy variable, and specific only for AT) seems to suggest a strong order effect. It seems that when respondents see the experiment for the first time, they do not value this attribute much, maybe they focus more on other attributes. When they repeat the experiment, instead, the attribute becomes highly significant, even if the second experiment is repeated after weeks, i.e., it is likely that they do not have the survey fresh in their mind. The attribute is in fact correctly positive, indicating that being the AT electric increases the probability to be chosen, however, the attribute is significant only in the second survey replied, whatever environment comes second.

5.4. Other effects

Results show more systematic heterogeneity in the preference for specific alternatives in the screen-based than in the immersive VR-based SC. We found that females are more likely to choose NT than AT, in line with the AV literature where females are typically classified as less tech-savvy and hence less prone to use AVs than males. All participants were informed upfront that the experiment included both an online survey and an immersive VR experiment and we might expect that females who accepted to participate in our experiment, that includes the VR, are probably more tech-savvy and more pro-innovation than the average females. Nevertheless they do still prefer NT over AT. The result about young participants, being more likely to choose AT, is also in line with the literature (Krueger et al., 2016) but the impact is not significant in the immersive VR experiment. The result about familiarity seems also reasonable, for opposite reasons. Those who are not at all or slightly familiar with AVs might have a higher probability to try AT, due to curiosity, than those who are familiar with AVs. Finally, we also tested if 'experience with VR' had an impact on the preference for specific alternatives. We found that having experience with VR increases the preferences for AT, but this effect is not significant (the null hypothesis can be rejected at 70%), with no difference between the main experiment and the control group.

6. Conclusions

This paper presents the results of a study carried out to shed light on the current discussion whether and to which extent preferences elicited with the immersive VR differ from those elicited with standard screen-based experiments. In particular, the research focuses on preferences elicited with SC experiments in the context of automated versus normal taxis. SC experiments are extensively used in transport field, and in particular to study innovative transport systems that are not yet available in the market. However, the lack of knowledge and experience for these new products adds another layer of concerns to the existing discussion about the accuracy of the preferences measured with SC experiments. In this context, the immersive VR technology offers a unique opportunity to allow respondents to make decisions in a realistic environment where the product is not available on the market.

Differently from previous literature, we built our experiment in order to ensure that the SC experiment in the VR-based environment was perfectly comparable with the one used in the standard screen-based. This allows internal validity. A sample of 156 participants replied both surveys (the screen-based first and the immersive VR-based after 1–2 weeks), and another sample of 52 participants was also used as control group to control for the order effect in the presentation of the two experiments. The control group and the ordering effect do not have direct implications for practice but are key in the evaluation of the results that have meaning in practical implications. Several interesting results were found.

Firstly, if the SC experiment used is exactly the same (with the attributes presented in the same way) in the immersive VR-based and in the screen-based environment, there is no difference in the choice bias between the two environments. This result might seem intuitive, however, what is interesting is that, even if the SC experiments are perfectly the same, the experience with the immersive VR environment does affect the individual preferences for the specific attributes presented in the SC experiment. Failing to model this difference might erroneously show the effect as choice bias instead.

Another interesting result is that among the attributes tested in the SC experiment, the immersive VR-based experience does not have an impact (compared to the screen-based) on the preferences for travel time and travel cost. It seems to have some effect in the preference for waiting time. Given the setting of our immersive VR experiment, where respondents can see other customers waiting for their taxi and take decisions in the same location where they would also wait for the taxi, this result seems to suggest that the immersive VR has an impact on the attributes that the respondents can live or do experience more closely. Maybe if the experiment would have taken place within the taxi, experiencing in-vehicle travel time, we would have observed an impact there. This stems from the fact that preferences typically derive from or tend to be shaped by our knowledge and experience. Allowing respondents to test the trip on board the taxi could be an interesting extension of the research. However, it must be noticed that in our study, respondents choose between two taxis, one normal and one automated, there is no difference then on how travel time can be used, compared to the studies discussed in the literature review that compared the WTP for travel time between a normal car (where the participant drives) and an automated car.

The most striking impact of the immersive VR was observed in the evaluation of the "non-economic" attributes, and in particular in the preference for the customer reviews and the number of customers who have chosen the taxi. The literature reports that the VR allows respondents to focus more on the decision. Our results suggest that the realism of the immersive VR experience induces a shift in the focus of the participants more on the LOS attributes than on the social information or other vehicle characteristics provided. In particular, for a norm to exert influence, it requires respondents' attention to be focused on the norms. The hints and cues provided by the immersive VR seems to drive respondents' attention away from these social attributes. This effect is much more marked when the immersive VR was the first experiment performed, i.e., when participants have not seen the experiment before. In addition, the immersive VR environment seems to be able to elicit much more heterogeneity in the preferences among respondents than the screen-based experiment do. This is an important result that has of course practical consequences if these results are used to inform policies makers or manufacturers.

While results are always context-specific, and more evidence is needed, we believe this research allows some cautious and general conclusions. It seems that using an immersive VR environment has indeed an impact on the preferences elicited with SC experiment. This impact is evident in three aspects: 1) on the attributes that the respondents is able to experience (also partially) within the VR environment; 2) on those attributes more difficult to measure, either because of visual attributes (such as urban environment) or because of attributes whose evaluation does not implies awareness, but unintentional impact exerted by action of others (such as social conformity attributes); 3) on the ability to elicit more systematic heterogeneity in the individual preferences for the alternatives evaluated and their characteristics. The extent to which these aspects can be elicited with the VR environment depends on the specific context and on the level of details measured in the data collection process and used to estimate the models. Our results seem to suggest that it is not possible to generalise the impact on the choice bias as this depends on the model specification. It is however recommended to carefully test differences in the preferences elicited with the screen-based and the VR-based data, as differences that are not explicitly estimated affect the choice bias.

The use of the control group allowed us to test if the impact of the VR environment could be affected by the fact that respondents had already seen the SC experiment before. Enough time was left between the two experiments to leave us confident that respondents do not remember the SC experiment (at least not in detail). However, it seems that participants who perform the experiment first in the immersive VR experiment are subject to more stimuli and tend to focus on fewer attributes and mostly times and costs. However, the sample used in our control group is smaller than that used in the main survey, and we did not run a randomised control trial. This can affect the results. Additional work is needed to confirm this finding from the control group. We also acknowledge that a full randomized control (FRC) trial design is probably needed to verify internal validity. However, it is worth nothing that FRC trials are not common in transport field (Merlin et al., 2022). Indeed, FRC trail design cannot be applied to any context and it is not without criticism, one of this is the cost and time needed to run them, which is a critical limitation in VR studies.

Since the immersive VR environment gives the perception of acting in a real world, the expectation is that preferences elicited with the immersive VR environment should be better (closer to reality) than the ones elicited with screen-based experiments. A real and consolidated market is needed to test the external validity of the results. But we share the feeling of Matthews et al. (2017) that immersive "virtual environments should seriously be considered", based on our results, to study products not available on the market and even more to study the impact of social conformity or social influence in its various aspects.

A final note, in our experiment we put much emphasis in building a highly realistic environment. However, as discussed in Cummings and Bailenson (2016) it is important to stress that while the level of details and the resolution of the images are relevant components in providing the sense of "being there", other aspects of the immersion, like the easiness of taking actions within the environment, the visual extension of the view, are even more important to allow participants to perceive themselves as being located in that space.

CRediT authorship contribution statement

Hao Yin: Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing. **Elisabetta Cherchi:** Funding acquisition, Supervision, Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This research is part of the Veronica project (https://www.veronicaproject.org/), financed by the Economic and Social Research Council (ESRC) (grant number: ES/S006885/1), UK.

References

Allen, V.L., 1965. Situational factors in conformity. Adv. Exp. Soc. Psychol. Elsevier 133–175.
Animesh, A., Pinsonneault, A., Yang, S.-B., Oh, W., 2011. An odyssey into virtual worlds: exploring the impacts of technological and spatial environments on intention to purchase virtual products. MIS Q. 789–810.
Arellana, J., Garzón, L., Estrada, J., Cantillo, V., 2020. On the use of virtual immersive reality for discrete choice experiments to modelling pedestrian behaviour. J. Choice Modelling 37, 100251.
Bansal, P., Daziano, R.A., 2018. Influence of choice experiment designs on eliciting preferences for autonomous vehicles. Transp. Res. Procedia 32, 474–481.
Bateman, I.J., Day, B.H., Jones, A.P., Jude, S., 2009. Reducing gain–loss asymmetry: a virtual reality choice experiment valuing land use change. J. Environ. Econ. Manag. 58 (1), 106–118.
Bierlaire, M., 2016. PythonBiogeme: A short introduction (Report TRANSP-OR 160706, Series on Biogeme). Transport and Mobility Laboratory, School of Architecture, Civil and Environmental Engineering, Lausanne, Switzerland.
Birenboim, A., Dijst, M., Ettema, D., de Kruijf, J., de Leeuw, G., Dogterom, N., 2019. The utilization of immersive virtual environments for the investigation of environmental preferences. Landsc. Urban Plan. 189, 129–138.
Bogacz, M., Hess, S., Calastri, C., Choudhury, C.F., Mushtaq, F., Awais, M., Nazemi, M., van Eggermond, M.A., Erath, A., 2021. Modelling risk perception using a dynamic hybrid choice model and brain-imaging data: an application to virtual reality cycling. Transp. Res. Part C: Emerg. Technol. 133, 103435.
Camerer, C., Mobbs, D., 2017. Differences in behavior and brain activity during hypothetical and real choices. Trends Cogn. Sci. 21 (1), 46–56.
Cherchi, E., Ortúzar, J. de D., 2002. Mixed RP/SP models incorporating interaction effects. Transportation 29, 371–395.
Cherchi, E. 2020. Our IATBR: 45 years of contribution to transport behaviour research. In Mapping the Travel Behavior Genome. K. Goulias (ed.) Elsevier, Chapter 1, 17-28.
ChoiceMetrics, 2012. Ngene 1.2 User Manual & Reference Guide, Australia.
Cummings, J.J., Bailenson, J.N., 2016. How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. Media Psychol. 19, 2272–2309.
Fagnant, D.J., Kockelman, K.M., 2014. The travel and environmental implications of shared autonomous vehicles, using agent-based model scenarios. Transp. Res. Part C: Emerg. Technol. 40, 1–13.
Faiola, A., Newlon, C., Pfaff, M., Smyslova, O., 2013. Correlating the effects of flow and telepresence in virtual worlds: Enhancing our understanding of user behavior in game-based learning. Comput. Hum. Behav. 29 (3), 1113–1121.
Farooq, B., Cherchi, E., Sobhani, A., 2018. Virtual immersive reality for stated preference travel behavior experiments: A case study of autonomous vehicles on urban roads. Transp. Res. Rec. 2672 (50), 35–45.
Feng, Y., Duives, D.C., Hoogendoorn, S.P., 2022. Development and evaluation of a VR research tool to study wayfinding behaviour in a multi-story building. Saf. Sci. 147, 105573.
Fiore, S.M., Harrison, G.W., Hughes, C.E., Rutström, E.E., 2009. Virtual experiments and environmental policy. J. Environ. Econ. Manag. 57 (1), 65–86.
Gkartzonikas, C., Gkritza, K., 2019. What have we learned? A review of stated preference and choice studies on autonomous vehicles. Transp. Res. Part C: Emerg. Technol. 98, 323–337.
Howard, D., Dai, D., 2014. Public perceptions of self-driving cars: The case of Berkeley, California, Transportation research board 93rd annual meeting, pp. 1-16.
Iglesias, P., Greene, M., de Dios Ortúzar, J., 2013. On the perception of safety in low income neighbourhoods: using digital images in a stated choice experiment. Edward Elgar Publishing, Choice Modelling.
Jara-Díaz, S.R., Videla, J., 1989. Detection of income effect in mode choice: theory and application. Transp. Res. B Methodol. 23 (6), 393–400.
Kolarova, V., Steck, F., Cyganski, R., Trommer, S., 2018. Estimation of the value of time for automated driving using revealed and stated preference methods. Transp. Res. Procedia 31, 35–46.
Krueger, R., Rashidi, T.H., Rose, J.M., 2016. Preferences for shared autonomous vehicles. Transp. Res. Part C: Emerg. Technol. 69, 343–355.
Lokhandwala, M., Cai, H., 2018. Dynamic ride sharing using traditional taxis and shared autonomous taxis: A case study of NYC. Transp. Res. Part C: Emerg. Technol. 97, 45–60.
Louviere, J.J., Hensher, D.A., Swait, J.D., 2000. Stated choice methods: analysis and applications. Cambridge University Press.
Lovreglio, R., Fonzone, A., Dell'Olio, L., 2016. A mixed logit model for predicting exit choice during building evacuations. Transp. Res. A Policy Pract. 92, 59–75.
Lu, M., Taiebat, M., Xu, M., Hsu, S.-C., 2018. Multiagent spatial simulation of autonomous taxis for urban commute: Travel economics and environmental impacts. J. Urban Plann. Dev. 144 (4), 04018033.
Matthews, Y., Scarpa, R., Marsh, D., 2017. Using virtual environments to improve the realism of choice experiments: A case study about coastal erosion management. J. Environ. Econ. Manag. 81, 193–208.
Merlin, L.A., Freeman, K., Renne, J., Hoermann, S., 2022. Clustered randomized controlled trial protocol of a Mobility-as-a-Service app for College campuses. Transp. Res. Interdisciplinary Perspect. 14, 100572.
Mokas, I., Lizin, S., Brijs, T., Witters, N., Malina, R., 2021. Can immersive virtual reality increase respondents' certainty in discrete choice experiments? A comparison with traditional presentation formats. J. Environ. Econ. Manag. 109, 102509.
Nah, F.-F.-H., Eschenbrenner, B., DeWester, D., 2011. Enhancing brand equity through flow and telepresence: A comparison of 2D and 3D virtual worlds. MIS Q. 731–747.
Patterson, Z., Darbani, J.M., Rezaei, A., Zacharias, J., Yazdizadeh, A., 2017. Comparing text-only and virtual reality discrete choice experiments of neighbourhood choice. Landsc. Urban Plan. 157, 63–74.
Rossetti, T., Hurtubia, R., 2020. An assessment of the ecological validity of immersive videos in stated preference surveys. Journal of Choice Modelling 34, 100198.
Sanchez-Vives, M.V., Slater, M., 2005. From presence to consciousness through virtual reality. Nat. Rev. Neurosci. 6 (4), 332–339.
Sobhani, A., Farooq, B., 2018. Impact of smartphone distraction on pedestrians' crossing behaviour: An application of head-mounted immersive virtual reality. Transport. Res. F: Traffic Psychol. Behav. 58, 228–241.
Stricker, L.J., Messick, S., Jackson, D.N., 1970. Conformity, anticonformity, and independence: Their dimensionality and generality. J. Pers. Soc. Psychol. 16 (3), 494.
Velasco, J.P.N., Farah, H., van Arem, B., Hagenzieker, M.P., 2019. Studying pedestrians' crossing behavior when interacting with automated vehicles using virtual reality. Transport. Res. F: Traffic Psychol. Behav. 66, 1–14.
Willis, R.H., 1965. Conformity, independence, and anticonformity. Hum. Relat. 18 (4), 373–388.
Yap, M.D., Correia, G., Van Arem, B., 2016. Preferences of travellers for using automated vehicles as last mile public transport of multimodal train trips. Transp. Res. A Policy Pract. 94, 1–16.
Yin, H., 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 1–22.
Yin, H., Cherchi, E., 2024. Conducting stated choice experiments within an immersive virtual reality environment: an application to the choice of automated taxi. Transp. Res. Procedia 76, 686–691.
Zhang, W., Guhathakurta, S., Fang, J., Zhang, G., 2015. Exploring the impact of shared autonomous vehicles on urban parking demand: An agent-based simulation approach. Sustain. Cities Soc. 19, 34–45.
Zou, X., O'Hern, S., Ens, B., Coxon, S., Mater, P., Chow, R., Neylan, M., Vu, H.L., 2021. On-road virtual reality autonomous vehicle (VRAV) simulator: An empirical study on user experience. Transp. Res. Part C: Emerg. Technol. 126, 103090.