A stated choice experiment to estimate preference for fully automated taxis: comparison between immersive virtual reality and online surveys

Hao Yin*[1], Elisabetta Cherchi[2]

[1] Research Assistant, School of Engineering, Newcastle University, UK

[2] Professor, School of Engineering, Newcastle University, UK

SHORT SUMMARY

In this paper, we aim to study the role of the immersive Virtual Reality (VR) experience in the preferences elicited with standard Stated Choice (SC) experiments embedded into a VR environment. For this purpose, a SC was built and implemented both online and within a VR environment and respondents were asked to reply to both surveys. The SC experiment consists of a binary choice between a normal taxi with the driver (NT) and a fully automated taxi (AT). The context is a well-known street in the city centre of Newcastle upon Tyne (UK). Hybrid choice models were estimated and results compared. Results suggest that VR experience indeed has significant effects on some attributes examined (waiting time and good reviews) and on the role of the latent variables in the choice of AT. Trust is significant only online, while injunctive norms and perceived safety only in the VR environment.

Keywords: Immersive Virtual Reality, Stated Choice Experiment, Fully Automated Taxis, Social Conformity, Internal validity

1. INTRODUCTION

Stated Choice (SC) experiments are commonly used to investigate users' acceptance of Autonomous Vehicles (AV), as this is a product not yet available in the market. 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 a preference for the product (see a discussion in Cherchi and Hensher, 2015). Pictures and videos have been increasingly used to provide more realistic information about the new products and, in particular in the case of AV to show respondents how the system could work (e.g., Haward and Dai, 2014; Kolarova et al., 2018). In this line, recent applications have also used Virtual Reality (VR) environment, mostly with the aim to control for the framing effect, i.e., to improve the preliminary information about the non-markets good respondents were going to evaluate (Bateman et al., 2009; Fiore et al., 2009; Phillips and Marsh 2015; Patterson et al., 2017).

VR experiments represent a new area of research that promises to change fundamentally the way in which consumers' preferences for innovations are measured. Studies have shown that people can develop realistic spatial knowledge in the VR environment that is like actual physical environments (O'Neill 1992; Ruddle et al., 1997; Tlauka and Wilson 1996). VR can generate a sufficiently natural and familiar field, able to provide 'field cues' or 'field hints' that occurr in the real world (Fiore et al., 2009). VR allows the sensation of immersion in the activities on the screen and with the virtual elements (Animesh et al. 2011; Faiola et al. 2012; Nah et al., 2011), prompting individuals to act as if they were in the real world (Sanchez-Vives et al. 2005). Based on this

theory the expectation is that VR experiments should perform better in eliciting individual preferences, i.e., should provide more realistic results than a standard online survey. However, very little is known in this area.

Some authors (Farooq et al., 2018; Arellana et al., 2020; Bogacz, et al., 2021; Feng et al., 2022) have used VR technology with stated preference experiments applied to pedestrian or cycling experiments, which involves continuous movement, not a choice among discrete alternatives. 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 cycling or walking) activate different circuitries in our brain compared to choice-based actions and show a better overlap between brain activities during imagined and real movements (see discussion in Cherchi, 2020). Moreover, the stated preference experiment in these studies is used to control the elements of the VR environment where respondents perform a continuous behaviour. 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 stated choice online survey. 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 is on the qualitative assessment of aspects of an urban environment.

In this paper, we aim to study the impact of the immersive VR experience in the preferences elicited with standard SC experiments embedded into a VR environment. For this purpose, a SC was built and implemented both online and within a VR environment and respondents were asked to reply both surveys. The SC experiment consists of a binary choice between a normal taxi with the driver (NT) and a fully automated taxi without driver and without steering wheel (NT). A set of attitudinal questions were also included in the survey to investigate if the VR environment has a mediating role on the impact of Trust, Injunctive Norm and Perceived Safety in the choice of the taxi type. Hybrid choice models were estimated with both surveys allowing assessing the impact of the VR experiment in the elicited consumer preferences.

2. METHODOLOGY

The core of the methodology set up in this research consists of a SC experiment built to elicit preferences for automated versus normal (i.e., with driver) taxis. The same SC experiment is used to collect information online and within a VR environment. Differently from the existing literature in the field, respondents go through a standard stated choice experiment while moving in the VR environment and "living" the choice experience. In our immersive VR experiment, respondents found themselves into the street where there is a taxi rank with the two types of taxis picking up customers and a ticket board (as it is the case in reality) where to select and pay for the taxi respondents wish to use. The advantage of this setting is that the SC experiment in the VR environment is perfectly comparable with the standard SC online, allowing to better disentangle the impact of the immersive VR experience in the elicited preferences.

The SC experiment built 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 2 levels: electric or gasoline) and two attributes to measure the impact of social conformity (number of customers who have used AT or NT, with three levels and customer rating with two levels). A heterogeneous Bayesian 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 three travel distances of 5 km, 10 km and 15 km (the only differences are the attribute level value of 'travel cost' and 'travel

time' among 3 SC designs). 16 choice scenarios were generated and randomly divided in 2 blocks. Each respondent was presented with 8 scenarios.

The above process is not different from a standard screen-based SC. However, building comparable SC experiments, when implemented in VR and online, poses some challenges. The most important and interesting issue is that some of the elements that are typically used in an online SC and are considered perfectly acceptable, look unrealistic when used in a VR-based environment. For example, we had to set the context at a taxi rank to allow respondents to make the choice in the virtual street. Traditional taxi services are still extensively used in Newcastle and there are numerous taxi ranks in the city centre. Some attributes appeared to be unrealistic when included in the SC in the VR environment. For example, the form of payment is a standard attribute often included in online SC experiments, but interestingly when used in the SC experiment within the VR environment it appeared clearly unrealistic. This is because when we choose a transport option in reality, we are only presented with the characteristics of the options, and only after we make the choice the ticket machine asks us how we would like to pay. The realism of the SC experiment embedded in the VR experiment made this problem evident. After several tests, the attribute was removed from the experimental design, and the question "how do you want to pay?" included after each SC scenarios (both in the online and the VR-bases survey). Finally, in standard SC experiments, respondents are typically presented with the destination of their most recent trip once at the beginning of the SC and before each scenarios 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, 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 do the same trip felt awkward. 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 the 3 designs (5km, 10km or 15km).

The VR-based SC experiment and its online counterpart were administrated in Newcastle in 2022. The final sample consists of 156 valid responses (1248 pseudo-individuals). These 156 respondents answered first the online SC survey, which according to the standard practice included: general questions about familiarity with automated vehicles, information about a last trip performed by normal taxi, a stated choice experiment, a set of socioeconomic information and nine statements to measure three latent psychological constructs: injunctive norm, perceived safety and trust. Approximately one week after, respondents were invited to the lab to perform the same SC experiment but this time embedded into the VR environment. All respondents interviewed are residents in the northeast of England and satisfy the requirements to be 18 years or older and have used a normal taxi in Newcastle in the last year. Table 1 reported the key characteristics of the sample. Our sample approximates the gender distribution of the Newcastle population but underrepresents young people (20% against 25% in the Newcastle population).

Table 1. Sample Characteristics

Socio-demographic characteristics		**%**
Gender	Female	52.6
	Male	46.8
	Rather not to say	0.6
Age	Younger than 30 years old	19.9
	30 years old or older	80.1
Education level	Bachelor degree or below	66.0
	Master or Doctorate degree	34.0
Current work status	Employed full-time	61.5
	Others	38.5
Personal monthly disposable income	Less than £500	12.8
	£501-£1500	40.4
	£1501-£2500	21.8
	£2501- £3500	9.0
	£3501-£ 4500	2.6
	More than £4500	0.0
	I do not wish to disclose it	13.5
Travel characteristics		**%**
Frequency of using taxis	More than once a week	10.3
	Between Once a month & Once a week	42.3
	Between Twice a year & Once a month	39.7
	Less than twice a year	7.7
Frequency of talking with driver	Very infrequently	8.3
	Somewhat infrequently	8.3
	Occasionally	30.1
	Somewhat frequently	36.5
	Very frequently	16.7
Knowledge levels of AVs and ATs		**%**
Heard of AVs	Yes	69.2
	No	30.8
Familiar with 5 levels of automation	Not at all familiar	39.7
	Slightly familiar	35.3
	Moderately familiar	19.9
	Very familiar	4.5
	Extremely familiar	0.6

3. MODELS ESTIMATED AND RESULTS

Hybrid choice models (HCM) were used to elicit user preferences. The discrete choice component of the HCM is a mixed logit (ML) model that allows estimating the trade-off between the attributes included in the SC experiment, controlling for panel effects (intra-individual correlation). The latent variable component of the HCM allows for estimating the impact of three latent psychological variables.

Model results are reported in Table 2. We note first that in both datasets, all the level of service attributes (travel time, waiting time and travel cost) have the expected sign and are significant at more than 99%. The same for the 'good reviews", and to some extent for the type of fuel ('EV').

While strong differences are found in the significance of the latent variables. We also note that the HCM estimated with the VR data has overall much better overall fit that the model estimated with online data (lower BIC and AIC).

Table 2. Models estimation results

	HCM Online		HCM VR	
	Estimated Value	**Rob. t-test**	**Estimated Value**	**Rob. t-test**
ASC(AT)	-7.730	-6.42	-5.800	-7.18
SIGMA (AT)	0.931	4.91	1.290	7.40
Level of Services				
Travel cost	-0.535	-9.88	-0.629	-12.47
Travel time	-0.083	-4.68	-0.127	-6.89
Waiting time	-0.130	-7.14	-0.189	-10.71
AT Vehicle Type				
EV	0.331	1.90	0.554	3.22
Social Conformity				
Good review	0.384	3.46	0.754	7.01
Systematic heterogeneity in alternatives				
Bachelor degree or above(AT)	0.614	2.38	-	-
Frequently talking with driver (AT)	-0.385	-1.57	-	-
Male (AT)	0.948	3.50	0.966	3.36
Latent variable				
Injunctive norm (AT)	0.336	1.18	0.485	2.41
Perceived Safety (AT)	0.397	1.50	0.805	4.05
Trust (AT)	1.400	4.07	0.339	1.32
Summary of Statistics				
Number of draws	500		500	
Maximum Log-likelihood	-2220.789		-2212.280	
Akaike Information Criterion	4537.578		4512.569	
Bayesian Information Criterion	4783.784		4738.258	
Number of individuals	156		156	
Number of observations	1248		1248	

Before looking at the differences between the two models, we note that a joint hybrid choice model was estimated to control for possible scale heterogeneity between online and VR-based SC data. The scale between online and VR-based SC data was not significantly different from one. We then compare the results using the models estimated separately. While the scale is not different, results show that the estimated preferences are significantly different between the online and the VR-based dataset for several attributes. The estimated marginal utility of travel cost, travel time, and fuel type is the same in the online and in the VR survey, while the estimated marginal utility of waiting time and good reviews are significantly higher in the VR than online. Willingness to pay (WTP) for saving waiting time is 1.2 times higher in the VR (18.0 £/hour) than in the online survey (14.6 £/hour) and the WTP for good reviews is more than 1.5 times higher in the VR (£ 1.2) than in the online survey (£ 0.72). This result is interesting and expected. Both results can be related to the higher realism provided by the VR environment. In the VR respondents can see other customers queueing, can move in the space where they are going to wait for the taxi, this could prompt a more realistic evaluation of the waiting time. If this argument is correct, then we should conclude that online surveys do underestimate the WTP for waiting time. If the choice context is more realistic, customer reviews provide respondents with a stronger hint or cue to

evaluate and compare the quality of these two types of taxi services, because respondents might have the feeling that they are really going to take the taxi.

Another interesting finding is that the VR environment indeed affects the role of the latent psychological constructs in the choice of automated taxis. Interestingly, trust significantly affects the preference for ATs only in the online survey, which makes sense because participants have not seen these innovative taxis operating, then trust in the automation is more important than in the VR experiment where they can see ATs operating around them. In line with that, when respondents see ATs really on the road, then the perception of safety becomes more important. Finally, injunctive norms are significant only in the VR environment, and this also makes sense, because the realism of the VR might give a feeling of being seen by others, which is related to the social norms.

4. CONCLUSIONS

This paper discussed the impact of the immersive VR experience in the preferences elicited with SC experiments. The same SC experiment was implemented in an online survey and within a VR environment and respondents were asked to reply to both surveys. Hybrid Choice models were estimated and preferences estimated in the two environments compared. Results suggest that the immersive VR environment does have an impact on the preferences elicited with SC experiments. Notably, it seems that the immersive experience has a strong impact on the preference for waiting time and customer reviews, both attributes are related to the experience that respondents can live in the immersive VR. Results also show that the role of the latent psychological factors tested (trust, injunctive norms, and perceived safety) is different (opposite) in the VR and in online. Before knowing or experiencing how the AT service works, trust plays a critical role. After or during the (virtual) experience, perceived safety becomes relevant instead of trust, and injunctive norms also become significant. Even if respondents replied to the two surveys at approximately one-week length one from the other, results could still be affected by order effect. At the same time, since the same respondents replied to both surveys, we are sure that the results are not affected by differences in socio-economic characteristics.

ACKNOWLEDGMENTS

This research was carried out as part of the Veronica project, funded by ESRC, UK.

REFERENCES

Animesh, A., Pinsonneault, A., Yang, S.B., and Oh., W. 2011. An Odyssey into Virtual Worlds: Exploring the Impacts of Technological and Spatial Environments on Intention to Purchase Virtual Products. *MIS Quarterly: Management Information Systems,* Vol. 35, No. 3, pp.789-810.

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*, Vol. 37, 100251.

Bateman, I.J., Day, B.H., Jones, A.P. and Jude, S., 2009. Reducing gain–loss asymmetry: a virtual reality choice experiment valuing land use change. *Journal of environmental economics and management*, Vol. 58, No. 1, pp.106-118.

Bogacz, M., Hess, S., Calastri, C., Choudhury, C.F., Mushtaq, F., Awais, M., Awais, M. Nazemi, M., van Eggermond, M.A.B. 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*, Vol. 133, 103435.

Cherchi, E. (2020) Our IATBR: 45 years of contribution to transport behaviour research. In *Mapping the Travel Behavior Genome*. K. Goulias (ed.) Elsevier, Chapter Vol. 1, 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*, Vol. 11, pp.154-164.

ChoiceMetrics, N., 2012. 1.2 *User Manual & Reference Guide*, Australia.

Faiola, A., Newlon, C., Pfaff, M. and Smyslova, O. 2012. Correlating the Effects of Flow and Telepresence in Virtual World: Enhancing Our Understanding of User Behavior in Game-Based Learning. *Computer in Human Behavior* Vol. 29, pp. 1113–1121.

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*, Vol. 2672, No. 50, pp.35-45.

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*, Vol. 147, 105573.

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*, Vol. 57, No. 1, pp.65-86.

Howard, D., Dai, D., 2014. Public perceptions of self-driving cars: the case of Berkeley, California. In: *93rd Annual Meeting of the Transportation Research Board*. Washington, D.C.

Kolarova, V., Steck, F., Cyganski, R. and Trommer, S., 2018. Estimation of the value of time for automated driving using revealed and stated preference methods. *Transportation research procedia*, Vol. 31, pp.35-46.

O'Neill, M.J. (1992). Effects of Familiarity and Plan Complexity on Wayfinding in Simu-lated Buildings. *Journal of Environmental Psychology* Vol. 12, No. 4, pp. 319-327.

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*, Vol. 157, pp.63-74.

Phillips, Y. and Marsh, D., 2015. Virtual Reality and Scope Sensitivity in a Choice Experiment About Coastal Erosion. 59th Conference Australian Agricultural and Resource Economics Society, February 10-13, 2015, Rotorua, New Zealand.

Rossetti, T. and Hurtubia, R., 2020. An assessment of the ecological validity of immersive videos in stated preference surveys. *Journal of choice modelling*, Vol. 34, 100198.

Ruddle, R.A., Payne, S.J. and Jones, D.M. 1997. Navigating Buildings in Virtual Environ-ments: Experimental Investigations Using Extended Navigational Experience. *Journal of Experimental Psychology: Applied*, Vol. 3, No. 2, pp.143-59.

Sanchez-Vives, M.V. and Slater, M., 2005. From presence to consciousness through virtu-al reality. *Nature Reviews Neuroscience*, Vol. 6, No. 4, pp.332-339.

Tlauka, M. and Wilson. P.N. 1996. Orientation-Free Representations from Navigation through a Computer-Simulated Environment. *Environment and Behaviour*, Vol. 28, No. 5, pp.647-664.