Connecting Choice Contexts: Evidence on Preference and Scale Differences from a Field Experiment

Yu Wang[1], Thoomas O. Hancock[1], Lucia Cebreiros Penzone[1], Charisma Choudhury[*1]

[1] Institute for Transport Studies, University of Leeds, Leeds, UK

*Corresponding author

SHORT SUMMARY

This study examines how travellers' preferences differ across connected choice contexts, including stated preference (SP), app-based travel planning, and real-world revealed preference (RP). To do so, we designed the Leeds-EXPERIENCE field experiment, in which 100 participants completed a pre-journey SP survey, an initial travel plan, a four-leg real-world journey under binding time and budget constraints, and a post-journey SP survey. A joint modelling framework was developed to analyse the four linked datasets. Preliminary results show clear shifts between planned and realised choices, indicating that information-based plans do not fully reflect real-world decisions. The joint estimates also suggest significant scale differences, with RP choices exhibiting lower scale than SP and planning data. The study highlights the value of connected multi-context data for improving behavioural realism and model transferability.

Keywords: field experiment; scale heterogeneity, stated preference; revealed preference; travel planning

1. INTRODUCTION

Understanding how travellers form and express preferences has long relied on evidence from multiple data sources, such as stated preference (SP) surveys, APP-based travel plan choices, and observed real-world behaviour (Hensher, 2012). Each approach provides useful but partial insights, reflecting different levels of realism, control over the choice context and attributes, information availability, restrictions on travellers' choice sets and decision processes. When used in isolation, however, each data source has limitations (Hensher, 2010; Danaf et al., 2019; Wang et al., 2025). For example, stated preference data are prone to hypothetical and reporting bias. These shortcomings limit the ability of single-source studies to fully represent travellers' preference structures and behavioural trade-offs.

To overcome these limitations, growing attention has been given to combining multiple data sources within joint modelling frameworks (Ben-Akiva & Morikawa, 1990). Such approaches, however, raise further challenges. In particular, systematic scale differences and context-dependent preference variations may arise from differences in decision making environment (Hensher et al., 1998). In addition, empirical applications are often constrained by the lack of connected datasets that observe the same individuals making comparable choices across multiple contexts (Wang et al., 2025). As a result, key questions about the nature and implications of preference and scale heterogeneity across decision environments remain open, restricting the development of realistic and transferable transport models.

Against this background, we designed and implemented the Leeds-EXPERIENCE field experiment to generate tightly connected travel choice data from the same individuals across multiple controlled and comparable decision contexts, including stated preference tasks, APP-based

travel plans, and real-world journey choices. The experiment aligns alternatives and attribute definitions across contexts while embedding real behavioural constraints, allowing like-for-like comparison of decisions. Building on this structure, we develop integrated models that jointly analyse the SP, planning, and real-world RP datasets, enabling a systematic assessment of preference stability, scale differences across contexts, and the transferability of behavioural parameters.

The three main research questions (RQ) investigated in this study are presented below:
(1) RQ1 (Method-related): How to design the experiment to generate connected multi-context choice data? Moreover, how to construct the model to analyse the preference and scale heterogeneity?
(2) RQ2 (Behaviour-related): To what extent do travellers' preferences and attribute sensitivities differ across connected choice contexts?
(3) RQ3 (Policy-related): How do the differences influence the reliability and transferability of model results for policy analysis and forecasting?

2. METHODOLOGY

A summary of the experiment procedure is presented here. For details of the experimental setup, please see Wang et al. (2026).

Experiment design and data

An incentive-compatible travel experiment named Leeds-EXPERIENCE (Leeds fiEld eXperiment on Preferences, Experience, and Real-world In-context EN-route ChoicE), was designed and conducted in Leeds, United Kingdom.

Participants were required to complete a multi-stage journey through four destinations, labelled A→B→C→D→A (as shown in Figure 1). Participants were told that they must finish the entire journey within a set time limit (105 min). At the start, each participant received a travel budget (£25) that they could spend on fares. They were free to choose among bus, train (only available for the second and third leg trips), and Uber for each trip. It should be noted that participants were not allowed to walk or cycle the entire trip. Any unspent budget remained theirs as a reward. Thus, the choices have real consequences. Moreover, to discourage participants from consistently choosing the cheapest (and simultaneously the slowest) travel option in order to maximise their reward, we introduced a lateness penalty mechanism. Participants who failed to complete all trips within the given time limit were penalized £0.2 per minute of delay which was deducted from their reward. In addition, participants received a fixed participation fee. Therefore, there was a clear incentive to balance time and cost. In designing the experiment, we expected that this incentive structure would generate increasing pressure as participants progressed through the later trips, particularly by Trips 3 and 4, as both time and monetary budgets became more constrained. This design was intended to expose participants to different levels of time and monetary pressure across decision points, as would typically be experienced in the real world. This is expected to increase behavioural realism of the experiment.

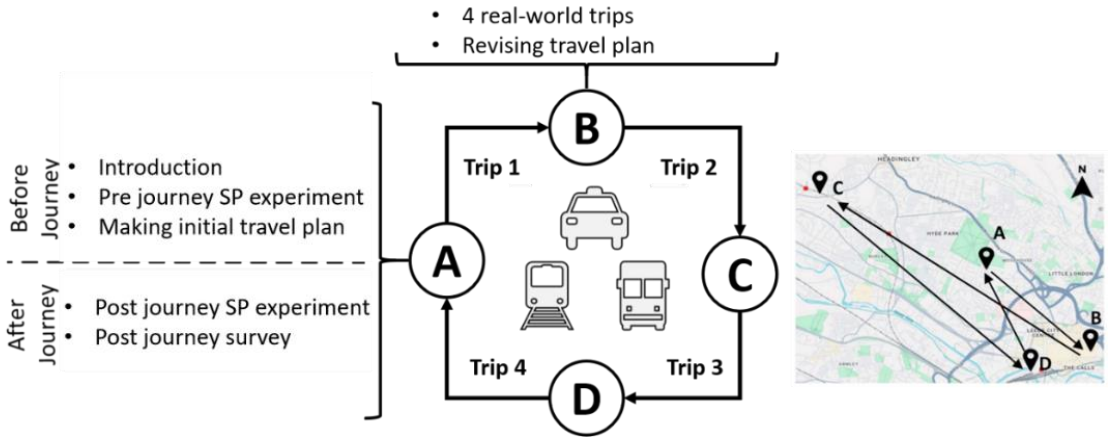


Figure 1 Experiment flow

Note:(A) Institute for Transport Studies, UoL; (B) The Markets Post Office, Leeds City Centre; (C) Station diner, Headingley; (D) The Brew Society, Leeds City Centre.

In the first stage, participants (travellers) completed a pre-journey stated preference experiment (SP1, 5 tasks) to measure their initial travel preferences, in which they were asked to choose among Uber, bus, and train in hypothetical scenarios. They then developed an initial travel plan (Plan, 4 tasks) for four real-world trips based on information from mobile applications. After making this plan, participants undertook a four-trip real-world journey (Real journey, 4 tasks) through four destinations (A→B→C→D→A) under binding time and budget constraints. At the beginning of the first trip and upon arrival at each subsequent destination, participants were required to record the travel time, cost, and waiting and walking time of the available modes for the next trip, ensuring that complete information on the choice set was collected. At these decision points, they were also allowed to revise their initial travel plan before making the next actual choice.

Upon reaching each destination (B, C, and D), participants also completed a brief questionnaire about their experience during the preceding trip (e.g., A→B, B→C, C→D). They reported their subjective mood, overall satisfaction, and specific experiential factors, such as perceived crowding and comfort, together with any immediate impressions. They were then asked to state their preferred mode in a similar future travel context, allowing the observation of potential preference updating. This procedure continued until participants returned to location A, where they completed a follow-up post-journey stated preference experiment (SP2, 5 tasks) to assess changes in preferences. Finally, participants completed a background questionnaire collecting demographic information, travel attitudes, and daily mode choice behaviour.

Before finalising the experimental design, a series of pre-tests were conducted to refine route selection, calibrate time and money budgets, determine the optimal balance between delay penalties and incentives, and assess procedural feasibility. Moreover, when designing the attribute levels, we referred to the actual data from the pilot of the real-world journey to ensure that the attribute values in the SP choice experiment were comparable in scale to those observed in the real-world journeys.

Models

A total of 102 participants registered, and 100 completed the experiment, yielding 1800 choice observations (Plan: 4 each, SP1: 5 each, RP from Real journey: 4 each, SP2: 5 each).

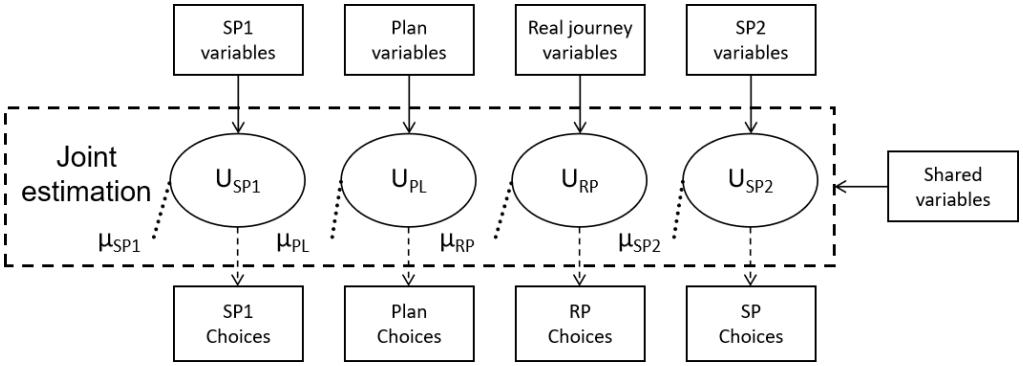


Figure 2 Model structure

For modelling, a joint framework has been developed to analyse the four choice datasets, as shown in Figure 2. Considering SP choice experiments both before and after the real-world journey, an aim was to explore whether actual travel experiences would lead to changes in stated preferences. Joint estimation assumes that different data sources capture comparable attributes and that individuals' perceptions remain broadly consistent across choice contexts, allowing common behavioural parameters to be identified. A key issue in joint modelling concerns scale differences in unobserved error variances across datasets (Ben-Akiva & Morikawa, 1990; Hensher, 2012). To account for this, separate scale parameters (μ) are incorporated for each choice context (Hensher et al., 1998).

3. RESULTS AND DISCUSSION

The modelling work in this study is ongoing. Nevertheless, the preliminary results already provide useful evidence on the behavioural relevance of the connected experimental design and on the presence of differences across choice contexts.

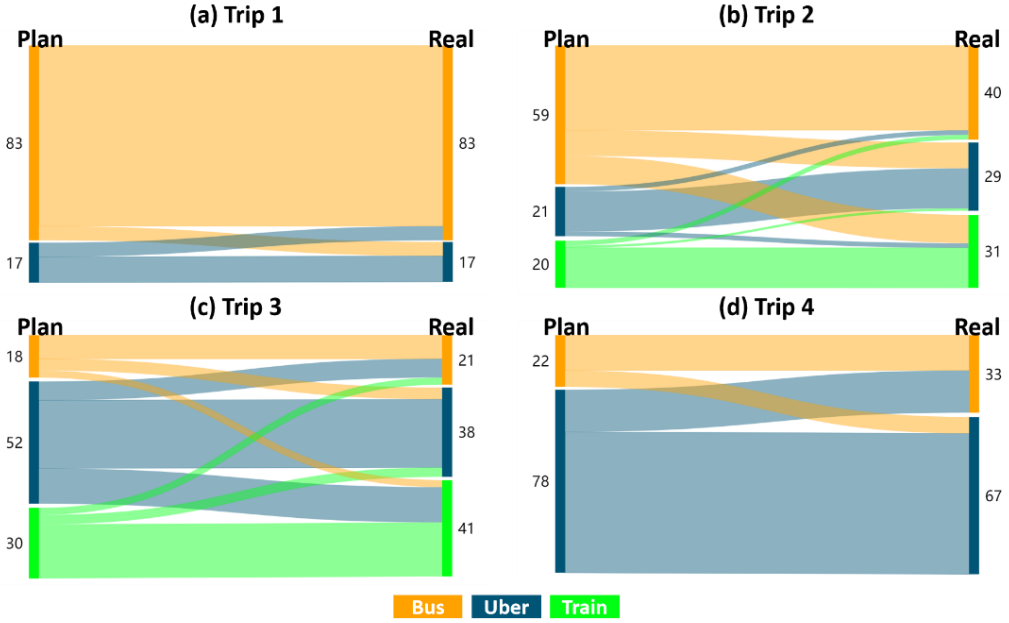


Figure 3 Mode shifts between plan and real choices

First, Figure 3 shows clear shifts between participants' initial planned choices and the modes ultimately chosen in the real-world journeys. This pattern is important because it indicates that choices made from pre-trip information alone cannot be assumed to be behaviourally equivalent to choices made at the point of travel. Although the alternatives and attribute definitions were

aligned across contexts, travellers still revised their decisions once they faced the actual trip environment. This finding supports the central premise of the experiment: even when the same individuals are observed and the choice set is controlled, decision outcomes remain context-dependent. In other words, the experiment is able to capture within-person variation in behaviour across hypothetical, information-based, and real travel settings.

Table 1 Scale differences

Parameters	Estimate	Rob.t.rat.(0)	Parameters	Estimate	Rob.t.rat.(1)
ASC_Uber	0.000	-	μ_SP1	1.000(fixed)	-
ASC_bus	1.425	3.81	μ_PL	0.914	-1.12
ASC_train	0.614	2.17	μ_RP	0.631	-4.31
beta_cost	-0.474	-7.02	μ_SP2	0.925	-0.80
beta_time	-0.1795	-10.77	LL(start)		-1815.32
beta_waiting+walking time	-0.205	-12.30	LL(final)		-1055.17
			Adj.Rho-squared		0.4143

The preliminary joint model provides evidence of systematic scale differences across the four datasets. Table 1 reports negative and statistically significant coefficients for cost, in-vehicle time, and waiting+walking time, which is consistent with behavioural expectations and supports the internal validity of the data. In substantive terms, travellers dislike paying more, travelling longer, and spending additional time waiting or accessing the service. Among the time-related attributes, the coefficient for waiting+walking time is slightly larger in absolute value than that for in-vehicle time, suggesting that out-of-vehicle components may be perceived as particularly burdensome in this setting. This pattern is consistent with the wider transport literature, where access and waiting time are often found to generate stronger disutility than in-vehicle time.

More importantly, the scale estimates indicate that the real-world RP data are associated with a lower scale than the SP and planning data. With μ_SP1 fixed to 1, the estimates for μ_PL and μ_SP2 are close to unity and not statistically different from the reference scale, whereas μ_RP is substantially lower and statistically significant. This implies that, relative to the SP1 data, the RP choices contain greater unexplained variability, or equivalently, that choices made in the real journey environment are less deterministic. Such a result is behaviourally plausible. Real-world choices are made in a setting where travellers face imperfect control, evolving situational conditions, and multiple sources of distraction or uncertainty that are difficult to fully capture using observed attributes alone. Even though the experiment was designed to keep alternatives comparable across contexts, the RP stage inevitably contains richer contextual influences than either the SP tasks or the app-based planning stage.

The results also suggest that the planning data occupy an intermediate behavioural position between SP and RP. Planning choices are more realistic than SP responses because they are based on real-time app information and linked to an actual journey that participants are about to undertake. At the same time, they are still formed before travellers experience the trip itself and before all contextual frictions fully materialise. This helps explain why the estimated scale for the planning data is closer to the SP datasets than to the RP dataset. Conceptually, this intermediate position is important, because it shows that "information-based realism" does not fully replicate "experienced realism". Thus, app-based planning data may reduce some of the abstraction associated with conventional SP, but they still cannot be treated as equivalent to revealed choices made in situ.

Another preliminary finding is the absence of a statistically meaningful scale difference between SP1 and SP2. This suggests that, at least at our case and within the current model specification, the real-world journey experience did not produce a major shift in the overall response consistency of the stated preference tasks administered before and after the experiment. In other words, exposure to the journey did not substantially alter how deterministically participants responded to the subsequent SP scenarios. This does not necessarily imply that experience has no effect. Rather, it may indicate that such effects are more subtle, heterogeneous across individuals,

or expressed through changes in specific taste parameters rather than through an overall scale shift. These issues require further investigation in richer model specifications. However, the current result is already informative, because it suggests that the most pronounced contrast in the present data is between experienced real-world choice and the other more controlled choice contexts, rather than between pre- and post-journey stated preferences.

Taken together, these preliminary results provide encouraging evidence for all three research questions. For RQ1, the experiment successfully generated connected multi-context data that reveal within-person behavioural changes across comparable decision settings. For RQ2, the early modelling results indicate that preferences are expressed differently across contexts and that the RP context is characterised by a distinct scale. For RQ3, the results point to potentially important implications for policy analysis, because model estimates obtained from one context may not transfer directly to another without accounting for scale and behavioural differences. In practical terms, policies evaluated using only SP or pre-trip planning data may overstate the stability and predictability of travellers' choices relative to real-world implementation conditions. Further modelling work will examine these implications more explicitly, including comparisons of derived indicators such as values of time and simulation-based policy forecasts across datasets.

4. CONCLUSIONS

This paper presents a connected field-experimental framework that makes it possible to observe the same travellers across multiple choice contexts, from hypothetical stated choices to app-informed plans and realised real-world decisions. The preliminary results show that these contexts are not behaviourally interchangeable: travellers revise their choices between planning and execution, and the joint model reveals a lower scale for real-world RP choices than for the SP and planning datasets. These findings highlight the value of collecting tightly linked multi-context data when the aim is to understand how preferences are expressed under different levels of realism. More broadly, the study contributes a practical basis for improving the behavioural credibility and transferability of transport choice models.

The next stage of this research will focus on extending the joint modelling framework to identify more precisely how preferences vary across contexts and how such differences affect model-based policy analysis. In particular, richer specifications will be estimated to test whether experience-related effects emerge through specific taste parameters rather than through scale alone, and whether these effects vary across individuals or trip stages. We will also compare behavioural indicators derived from different datasets, such as values of time and scenario-based predictions, to examine the extent to which conclusions drawn from SP or planning data remain valid when confronted with real-world choice behaviour.

ACKNOWLEDGEMENTS

The research is supported by UKRI Future Leader Fellowship [MR/T020423/1 and MR/Y034384/1].

REFERENCES

Ben-Akiva, M., & Morikawa, T. (1990). Estimation of switching models from revealed preferences and stated intentions. Transportation Research Part A: General, 24(6), 485-495.

Danaf, M., Atasoy, B., De Azevedo, C. L., Ding-Mastera, J., Abou-Zeid, M., Cox, N., ... & Ben-Akiva, M. (2019). Context-aware stated preferences with smartphone-based travel surveys. Journal of choice modelling, 31, 35-50.

Hensher, D. A. (2010). Hypothetical bias, choice experiments and willingness to pay. transportation research part B: methodological, 44(6), 735-752.
Hensher, D. A. (2012). Accounting for scale heterogeneity within and between pooled data sources. Transportation Research Part A: Policy and Practice, 46(3), 480-486.
Hensher, D., Louviere, J., & Swait, J. (1998). Combining sources of preference data. Journal of Econometrics, 89(1-2), 197-221.
Wang, Y., Cebreiros Penzone, L., Hancock, T. O., & Choudhury, C. F. (2026). Leeds-EXPERIENCE: a Leeds fiEld eXperiment on Preferences, Experience, and Real-world In-context EN-route ChoicE (Working paper). Institute for Transport Studies, University of Leeds.
Wang, Y., Hancock, T. O., Wang, Y., & Choudhury, C. (2025). Modelling residential relocation behaviour combining passive revealed preference data and stated preference survey data. Transport Policy, 103789.