Stated choices and simulated experiences

Differences in the value of travel time and reliability

Fayyaz, Muhammad; Bliemer, Michiel C.J.; Beck, Matthew J.; Hess, Stephane; van Lint, J. W.C.

DOI
10.1016/j.trc.2021.103145

Publication date
2021

Document Version
Final published version

Published in
Transportation Research Part C: Emerging Technologies

Transportation Research Part C 128 (2021) 103145

Contents lists available at ScienceDirect

Transportation Research Part C

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

Stated choices and simulated experiences: Differences in the value of travel time and reliability

Muhammad Fayyaz [a], Michiel C.J. Bliemer [a,*], Matthew J. Beck [a], Stephane Hess [b], J. W.C. van Lint [c]

[a] Institute of Transport and Logistics Studies, University of Sydney, Australia
[b] Choice Modelling Centre, Institute for Transport Studies, University of Leeds, United Kingdom
[c] Department of Transport and Planning, Delft University of Technology, the Netherlands

ARTICLE INFO

Keywords:
Hypothetical bias
Route choice behaviour
Stated choice experiments
Incentive compatible driving simulator experiments
Value of travel time
Value of travel time reliability

ABSTRACT

Surveys with stated choice experiments (SCE) are widely used to derive values of time and reliability for transport project appraisal purposes. However, such methods ask respondents to make hypothetical choices, which in turn could create a bias between choices made in the experiment compared to those in an environment where the choices have consequence. In this paper, borrowing principles of experimental economics, we introduce an incentive compatible driving simulator experiment, where participants are required to experience the travel time of their chosen route and actually pay any toll costs associated with the choice of a tolled road. In a first for the literature, we use a within respondent design to compare both the value of travel time savings (VTT) and value of travel time reliability (VOR) across a typical SCE and an environment with simulated consequence. Given the importance of VTT and VOR to transport decision making and the difficulty in estimating VOR using revealed preference data, our results are noteworthy and emphasise that more research on this topic is imperative. We provide suggestions on how the results herein may be used in future studies, to potentially reduce hypothetical bias that may be exhibited in SCE.

1. Introduction

1.1. The value of time and reliability

Route choice has been a topic of study for many decades in order to better understand the importance of various route attributes and to forecast behaviour in networks for transport planning and management purposes. Key in route choice analysis is the ability to derive the value of travel time (VTT) and the value of travel time reliability (VOR). VTT is the monetary value drivers assign to travel time changes, and VOR is the monetary value assigned to a change in travel time variability (unreliability). For the past five decades, the VTT has been considered an important value in transport policies and transport projects appraisal (Abrantes & Wardman, 2011). VTT serves two purposes, namely (i) as an input variable in cost-benefit analysis (CBA) of transport infrastructure projects, and (ii) as an explanatory variable in transport forecasting models (Shires & de Jong, 2009). More recently (for the last two decades), VOR has also received considerable attention in the CBA of transport projects and policies (de Jong & Bliemer, 2015).

* Corresponding author.
E-mail address: michiel.bliemer@sydney.edu.au (M.C.J. Bliemer).

https://doi.org/10.1016/j.trc.2021.103145
Received 1 September 2020; Received in revised form 15 March 2021; Accepted 9 April 2021
Available online 2 May 2021
0968-090X/

Given the importance of these inputs to fundamental transport decisions, there has similarly been much research into data types used to examine route choice decisions. Broadly speaking, the two types of data are stated preference (SP) and revealed preference (RP). SP data typically are collected with a stated choice experiment (SCE), in which respondents are asked to make choices in a series of hypothetical choice tasks. In contrast, RP data consists of route choices observed in the field, for example where drivers are tracked using mobile phones or other Global Positioning System (GPS) devices, remote sensing, or using driver-reported route information in interviews and questionnaires. VTT and VOR are often estimated using SP rather than RP data[1]. For example, in the UK, a large number of SCEs were designed for estimating VTT and VOR for different transport modes and trip purposes (Hess et al., 2017), and similar SP data collections have been conducted in many other countries.

While it is typically assumed that behaviour captured by SCEs reflects real-world behaviour, the hypothetical nature of SCEs may lead to biased results (Beck et al., 2016; Fifer et al., 2014). One of the potential reasons for hypothetical bias in typical SCEs is that "participants may not experience strong incentives to expend the cognitive efforts needed to provide researchers with an accurate answer" (Ding et al., 2005, p. 68). A SCE would therefore be incentive compatible only if it provides an incentive for participants to nudge them to truthfully reveal their preference towards an attribute. Experiments designed specifically to be incentive compatible are widely conducted in experimental economics and are often referred to as an economic experiment. Conversely, the tendency to simplify SCE settings, a practice that prevails despite extensive criticism (cf. Hess et al. 2020a), may lead to respondents enhancing the information in an unobserved manner to make up for the lack of experienced stimuli.

1.2. Contribution of this research

Given that differences may exist between values calculated on stated preference versus revealed preference data, research that seeks to examine this discrepancy is imperative given the importance of VTT and VOR to transport decision making. This research is, however, somewhat lacking. Following an extensive literature review on the comparison of stated preference (SP) and revealed preference (RP) data in the context of transportation, there are only two papers in the field that make the comparison between the two types of data with respect to VTT *and* VOR.

Brownstone & Small (2005) and Small et al. (2005) both make use of the same dataset wherein self-reported RP choices with regards to use of a tolled express lane on two highways in the US. By assuming that drivers know the distribution of travel times across days (which is a strong assumption), for any given time of day, they can measure VTT and VOR as preferences about this distribution. As with many studies that seek to examine VTT and VOR in a revealed preference context, Brownstone and Small (2005) acknowledge difficulties in data collection resulting in only one of the routes being able to satisfactorily identify coefficients of unreliability, with these results being sensitive to specification. In both studies, the SP survey asked people to choose among situations in which they trade off total travel time, the fraction of travel time in congested conditions, and trip cost between two otherwise identical routes. The SP tasks were generic across respondents and not pivoted around their actual experiences, though effort was made to segments split travellers into nine bands of total travel time. With regards to the data, there are 55 respondents from whom both RP and SP data was collected.

Brownstone and Small (2005) found differences in VTT estimated from both experimental treatments, and conclude that the examination of VOR is promising, but more research be done in studying this value. Small et al. (2005) use the same data and through a different modelling approach find that scale difference between SP and RP models was significant and that VTT and VOR were substantially underestimated based on SP data when compared to RP. A key limitation of both these studies, however, is that the SP and the RP data cannot be directly compared because the alternatives and levels shown are different across both data sources. As acknowledged by the authors, these differences can lead to travel time misperceptions or inconsistent behaviours.

To accommodate for this, Krčál et al. (2019), citing our original conference presentation[2] with preliminary findings of the research discussed herein, present an examination of VTT where the choice faced in the SP and RP setting were identical to allow for a more direct comparison. Note that the authors refer to RP results in their study, but the outcome was still experienced in a laboratory setting. While it is common for this type of experiment to be referred to as RP in experimental economics given that the monetary incentive is real, we argue that such data should be considered SP because the consequence is unlike that which would be experienced in a real world setting, despite an associated consequence with respect to time. Nonetheless, they find that VTT in the SP experiment is significantly lower than that revealed by the RP component of the experiment. Krčál et al. (2019) do not investigate VOR and instead of making all choices consequential, they randomly select only one choice task out of 80 to be consequential.

To better understand hypothetical bias, we need to understand the sources of hypothetical bias in a stated choice experiment. Is bias caused by the absence of consequences? Or because the environment is not realistic? Or perhaps because people provide socially desirable responses that do not reflect their actual behaviour? The presence of hypothetical bias varies across disciplines, for example in health economics it has been found that stated choice experiments have little to no hypothetical bias, while there is ample evidence in the environmental economics and consumer economics literature that the absence of consequences is an important source of hypothetical bias (Haghani et al., 2021). Haghani et al. (2021) distinguishes five classes of choice data, ranging from Class I (least

[1] Some exceptions exist (e.g., Carrion & Levinson, 2013; Fezzi et al., 2014; Prato et al., 2014), though historically RP data is rarely used to estimate VTT and VOR because many contexts cannot be (easily) examined in a real-world driving study (e.g., certain roads may not yet exist, certain toll levels may not yet exist). It is also very challenging to reconstruct objective or perceived route travel time distributions on road networks. More recently, there has been a renewed interest in using RP data for VTT research (e.g., Varela et al., 2018).

[2] Early results of our study were presented at the 15th International Conference on Travel Behaviour Research in July 2018 in Santa Barbara, USA.

realistic) data collected via typical non-consequential choice experiments to Class V (most realistic) data obtained via naturalistic choice observations. Class II refers to data from (partially) consequential choice experiments, Class III is associated with quasi-revealed or lab-in-the-field choice experiments, and Class IV refers to self-reported (or agent-aware) choice observations in the field. Existing studies on hypothetical bias in choice experiments have compared Class I data with data in a higher (more realistic, less hypothetical) class as benchmark, Table 1 summaries these studies based on the review of 57 articles across four applied economics domains in Haghani et al. (2021). This illustrates that there exist four studies in transport that have compared data from a typical hypothetical choice experiment with true choice observations (Class V) to demonstrate that hypothetical bias exists (Ghosh, 2001; Brownstone et al., 2000; Brownstone and Small, 2005; Small et al., 2005). However, it should be noted that these four studies used the same data set, illustrating that availability of Class V data is rare.

Our study contributes to the transport literature by investigating whether the absence of consequences in stated choice experiments are a source of hypothetical bias, i.e., we make a comparison with Class II data. Currently, only two such studies have been conducted in transport, namely Hultkrantz and Savsin (2017) and Krčál et al. (2019), who both identified significant bias in VTT due to consequences. However, they used a between-respondent design such that differences may be due to sample differences across the two data collection types. Our study adopts a within-respondent design to provide further evidence that the absence of consequences in a stated choice experiment is a source of hypothetical bias, not only in the context of VTT but also with respect to VOR.

Thus, the motivation for our study is to contribute to this very small body of research by examining drivers' route choice behaviour in both a stated preference experiment and an economic driving simulator experiment (DSE) that requires respondents to experience the travel times, costs, and travel time unreliability consequences of their choices in a more realistic setting. In particular, our experimental treatment ensures that the same respondents complete identical choice tasks in both experimental treatments, and in addition we are able to estimate VOR in not only the SP context, but also in the simulated RP experiment. To the best of our knowledge, this is the first study of its kind and this paper makes an additional contribution to the existing literature by being the first to use an economic DSE to estimate VTT and VOR measures in a route choice context.

1.3. *Outline of the paper*

The remainder of the paper is organised as follows. To help position the identified contribution of this research, we first provide a literature review of data collection techniques in a route choice context and their features. This is followed by an overview of the two experiments that are used to collect data, which is then followed by a discussion of the relatively novel experimental design that underpinned the values embedded in the stated choice experiments. We then provide information about the sample and some preliminary analysis, before looking at the results of comprehensive model estimation. Finally, we conclude with a discussion of the limitations of the current study, the over-arching outcomes of this research for the literature, and suggest future directions for research based on this analysis.

2. Literature review

Given the vast amount of literature on the collection of choice data, we limit this literature review to data collection techniques in the transport domain, with particular reference to route choice, VTT and VOR. We discuss these data collection methods in the light of the trade-off between (external) validity (do the data reflect real behaviour) versus the degree of experimental control in collecting the data, illustrated in Fig. 1, in which field data (RP) and experimental data (SP) are placed along both these axes. With respect to field data, the analyst has little control over the environment in which the data is collected, but the data exhibits low hypothetical bias. In contrast, while experimental data suffers from higher hypothetical bias, it can be collected with a high level of experimental control.

2.1. *Stated and revealed preference data*

RP data includes drivers' route choices in a real-world setting either by self-reported questionnaires and interviews (drivers complete a questionnaire regarding their past route choices), and/or GPS experiments or remote sensing (drivers are observed in real world traffic), making RP data either a subjective self-reported measurement or an objective observed measurement (Carrion & Levinson, 2012). There are many recent studies which examine route choice in an RP context (e.g., Djukic et al., 2016; Papinski et al., 2009; Ramos et al., 2012; Vacca et al., 2019; van Essen et al., 2019). One such study shows how mobile phone location data can be used to estimate meaningful VTT measures (Bwambale et al. 2019a). One common limitation with many RP studies is the difficultly in being able to generate robust VOR estimates. Indeed, there exists only a very limited number of studies that have estimated VOR using objective travel time measurement. Carrion & Levinson (2013) estimated VOR using travel time data collected from GPS devices. Prato

Table 1
Empirical studies into hypothetical bias of non-consequential choice experiments.

Applied economics domain	*Class II*	*Class III*	*Class IV*	*Class V*	*Total*
Transport economics	2	5	3	4	14
Environmental/resource economics	12	–	3	–	15
Consumer economics	13	2	–	3	18
Health economics	–	–	10	–	10

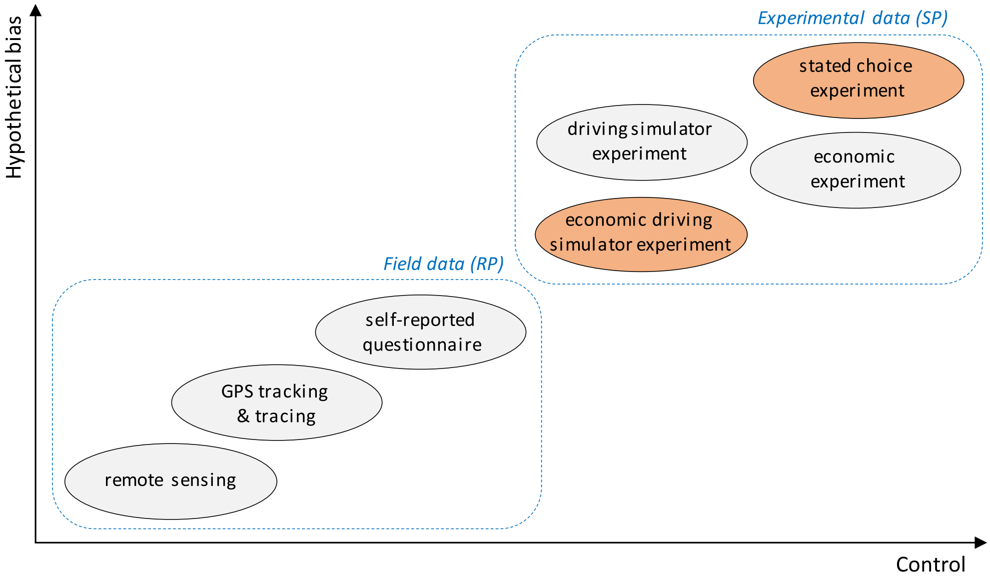


Fig. 1. Route choice data collection techniques.

et al. (2014) estimated VOR using GPS data collected in Denmark and recommend exploiting GPS data as technology is getting cheaper and the use thereof could result in "real" large scale models. In work related to travel time variability, Bwambale et al. (2019b) show how mobile phone data can be used to model departure time choices.

Given the limitation of RP data in some contexts (e.g., certain roads may not yet exist, certain toll levels may not yet exist, being able to construct the travel time distribution experienced or perceived by respondents, and typically the cost of data acquisition), SP methods are heavily used to examine travel behaviour and calculate both VTT and VOR (Hensher, 2001; Senna, 1994; Small, 1999). Participants of these experiments typically provide responses via an online survey, a pen-and-paper survey, or to a personal interviewer that brings a laptop or tablet to the participants' home or workplace. SCEs can also be conducted in a computer laboratory which allows analysts to show more complex choice tasks and could potentially also collect eye-tracking data, or other extra behavioural/ processing data. A trade-off with the high experimental control afforded by SP (due to all attributes and alternatives shown to respondents being controlled by the analyst) is the potential for hypothetical bias; participants tend to deviate from their stated responses when faced with the same situations in a real-life setting (see Broadbend, 2012; Fifer et al., 2014; Foster & Burrows, 2017; Loomis, 2011; Penn & Hu, 2018, 2019).

2.2. Comparisons between stated and revealed preference

Given the potential for differing results between RP and SP data, there have been a number of studies that examine the extent of hypothetical bias in transport economics. Brownstone et al. (2000) examine hypothetical bias in the context of alternatively fuelled vehicles, Haghani & Sarvi (2018, 2019) compare stated preference choices to those made in a simulated evacuation scenario, and the following authors examine differences between SP and RP data in terms of VTT and willingness to pay: Beck et al. (2016), Brownstone et al. (2000), Brownstone et al. (2005), Fifer et al. (2014), Ghosh (2001), Hultkrantz & Savsin (2018), Li et al., (2020), Nielsen (2004), and Peer et al. (2014). As discussed previously, there exist only two studies that examine both VTT and VOR in SP and RP contexts (Brownstone and Small 2005, Small et al. 2005), and one that addresses some of the limitations in previous SP and RP comparisons in the context of VTT (Krčál et al., 2019). In the majority of these studies, a significant bias between SP and RP contexts was observed though the direction of the bias was not consistent.

2.3. Bridging the stated and revealed gap

One technique to collect data that provides the analyst with some experimental control in a quasi-realistic environment, is driving simulator experiments (DSE). Driving simulators have been used extensively in road safety research (e.g. Donmez et al., 2007; Young et al., 2014), and increasingly also to study other aspects of traffic operations and driving behaviour, such as hysteresis (e.g. Saifuzzaman et al., 2017); drivers' responses to advanced traveller information systems (e.g., Bonsall, 2004; Koutsopoulos et al., 1994); and interactions with connected and / or automated vehicles (e.g. Ali et al., 2020a; Sharma et al., 2019), to name just a few examples. Bonsall (2004) argued that route choices in driving simulators provide more reliable data compared to data collected via typical surveys, the main advantage is that there is no need to inform a driver about an attribute of a route since the driver can experience it in the simulator.

In the past two decades, driving simulators have been increasingly used to investigate a range of driving and travel behaviour, including route choice behaviour. Bonsall et al. (1997) compared drivers' route choices in a driving simulator to their real-life decisions and found the decisions identical. Hess et al. (2020b) compare drivers' lane changing behaviour in an SCE and DSE and conclude that there are similarities as well as differences and suggests combining both data types. Despite the fact that DSEs allow participants to experience route attributes in a reasonably realistic virtual environment, they may still suffer from hypothetical bias. In real-life, driving is a goal-oriented activity where driving is for a reason (Levinson et al., 2004), which is why route choices made by a participant in a simulation may not reveal the participant's true preferences towards an attribute (e.g. travel time). For instance, arriving too late at a destination in a simulation is not the same as arriving too late for a meeting in real life.

This leads to another key difference between stated and revealed preference: the idea of consequence. Real world choices have real world consequences, whereas choices made in hypothetical situations are non-binding. Linked to this notion, early work in experimental economics examined conditions for experiments to be incentive compatible – broadly that a participant can achieve the best outcome to themselves by acting according to their true preferences. Smith (1976) outlined three conditions for an experiment to be incentive compatible; monotonicity, dominance and salience. Economic experiments in transportation are rare, outside of the study by Krčál et al. (2019), the only other being Dixit et al. (2015) who find that such experiments provide a high degree of experimental control, leading to internal validity and incentive compatibility.

Again, the unique contribution of this research is that we bring together the experimental control afforded by stated preference techniques, with the ability to create a quasi-real world environment via the driving simulators, to construct an experiment that is incentive compatible by incorporating approaches used in economic experiments. In doing so we are able to make direct comparisons between VTT *and* VOR estimates as a result of the type of experimental treatment only. How we constructed this novel experiment is outlined in the next section.

3. Overview of experiments

Our study consists of both an economic DSE and a typical SCE using a within-subject design, where the aim is to compare route choice behaviour of general population respondents in the two data collection techniques, and where the choice tasks are identical with the exception of simulated experiences and monetary incentives. Participants completed the study in two phases:

(i) A typical SCE consisting of five hypothetical choice tasks via an online survey.
(ii) The *same* five choice tasks where participants were made to experience their choices in the driving simulator (DSE) in the laboratory.

To control for order effects, we divided participants into two groups. In one group, participants completed the SCE before the DSE (order 1), while participants in the other group first completed the DSE (order 2). *Participants were not told that both experiments contain the same choice tasks*.

In making the choice, participants are asked to imagine travelling from home to work by car with two possible route alternatives, a stylised representation is shown in Fig. 2. The motorway alternative has a speed limit of 90 km/h, no traffic lights, a reliable travel time of 6 min across all choice tasks (i.e. exhibits no travel time variation), but the toll cost varies between $1 and $3[3]. The urban road alternative has a speed limit of 50 km/h and whilst having no toll, a driver will encounter four traffic lights that make the travel time unreliable, where it varies between 4 and 12 min according to a given probability distribution that changes across choice tasks[4]. Given that all participants in this study are from Sydney, the speed limit of 90 km/h on the motorway and 50 km/h on the urban road are based on existing tolled motorways and urban roads in Sydney (which varies between 80 and 100 km/h, and between 40 and 60 km/h, respectively).

3.1. Stated choice experiment

Presenting travel time unreliability[5] to participants in a SCE is not a straightforward exercise as the presentation formats vary considerably across empirical studies. For instance, Black & Towriss (1993) (cited in Tseng et al., 2009) suggested that participants can

[3] Australian dollars, where 1 Australian dollar is 0.79 US dollar or 0.64 euro (price level 28 Feb 2018, the approximate time period around which the experiment was conducted).

[4] While the travel times and toll costs for each choice task may not be representative of an average trip in Sydney, at an aggregate level the time and cost trade-off is appropriate. For instance, combining the travel time of all five choice tasks (in the simulator/stated choice survey) is comparable to the travel time of an average trip in Sydney where a driver spends between 25 and 30 min (Best Case, when the motorway is selected or a driver is lucky enough to get a (combination of) 4 min or 6 min travel time each time on the urban road) and 60 min (Worst Case) driving and spends up to $9 on tolls. We selected our toll levels based on distance-based charging used on some toll roads in Sydney, e.g., drivers usually pay around A $2 for a drive of 6 min on the M7 Motorway. The repetition of paying "small" tolls adds up to a total toll cost that is comparable to toll roads of similar proposed travel time savings over the length of the experiment.

[5] There exist two cases of travel time unreliability, one under risk (where a driver knows the likely travel time and their probability distribution on a route or at least able to predict, e.g., recurrent peak hour congestion) and one under uncertainty (where drivers do not know the travel time distribution and cannot accurately predict their travel times on a route, e.g., traffic accident). In this paper, we consider travel time outcomes under risk and not under uncertainty. That is, the travel time distribution on each route is clearly known to the respondents in each choice task, as is common in stated choice surveys that aim to estimate VTT and VOR.

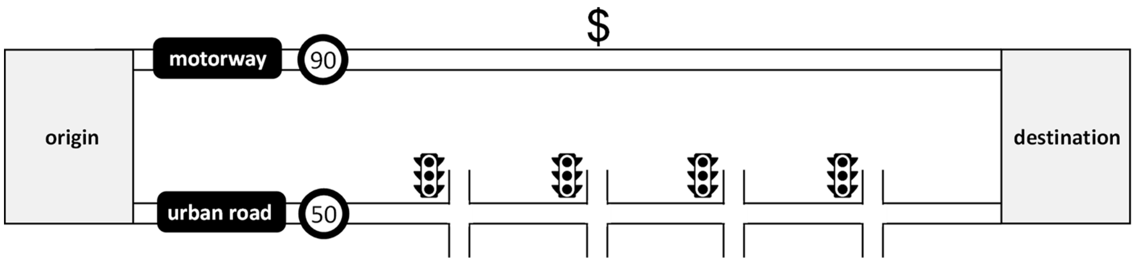


Fig. 2. Stylised representation of dual route network.

interpret a 5-point travel time distribution (in contrast to a 10-point) well. Tseng et al. (2009) conducted face-to-face interviews with 30 participants wherein they presented eight different formats taken from empirical studies in the literature on travel time unreliability. Based on responses to several indicators (e.g., clarity of reliability presentation, how easy it is to make a choice between two alternatives etc.) they recommended using a verbal description instead of a graph showing a probability distribution.

For this study, we designed the format of SCE choice tasks by considering the recommendations of Tseng et al. (2009). An example of the SCE choice task presented to participants is shown in Fig. 3. We explained to the participants that only the travel times of the urban route and the toll cost of the motorway vary over choice tasks, indicated in red. Respondents completed the choice tasks in the Qualtrics online survey instrument.

3.2. *Driving simulator experiment*

The driving simulators used in this study are based at the Travel Choice Simulation Laboratory at The University of Sydney (TRACSLab@USyd), as shown in Fig. 4. The lab consists of five driving simulators which allow the analyst control over the driving environment (this includes the road system, traffic signals and controls, as well as computer-simulated cars as background traffic) and facilitate the capture of all decisions made by a maximum of five human drivers at the same time. The driving simulators are detached and transformed Holden Commodores. Each simulator is comprised of functional pedals, steering wheel, automatic gearbox (neutral, drive, reverse, etc.), indicator levers, dashboard, radio/CD player and automatic/motorized seat adjustment controls. To run a simulation on the driving simulators and to control traffic, the SimCreator software (developed by Realtime Technologies Inc.) is used. For the purposes of this experiment, driving in the simulator was restricted to 12 min on any one route in order to avoid simulator sickness, with breaks of at least three minutes between consecutive driving episodes.

The motorway route was designed to have no intersections or buildings, thereby providing an experience of driving on a highway, as shown in Fig. 5(a). On the other hand, the urban road alternative replicates characteristics of a usual urban road which consists of a number of buildings, intersections, and traffic lights, see Fig. 5(b). Both routes also have speed signs and other road signs. Before choices are made in the simulator, each participant is asked to conduct a test drive on both the motorway and the urban road.

The choices presented in the DSE are *exactly the same* as the SCE except for the fact that in the DSE, the participant is faced with simulated experiences of the route travel time, travel cost, and travel time unreliability after having made a route choice decision.

The benefit of conducting driving simulator experiments (over a more naturalistic experiment) is that the researcher has a high degree of control over experienced travel times. Travel times in the simulator are controlled by adjusting the traffic light settings (longer red intervals means longer delays) and programming computer-controlled vehicles—lead vehicle, tail vehicle, and right-side vehicle(s). The lead and right-side vehicles ensure that a participant does not start speeding (above the speed limit) and overtaking, respectively; and the tail vehicle gives a perception of the general traffic to the participant in the rear-view mirror of the simulator. The scenario is designed in such a way that when the participant starts driving, the computer-controlled vehicles start driving too (a sensor is used to achieve this).

We did not opt for a long driving segment because (i) our aim of capturing route choice was well captured in this road segment, (ii) we need to consider multiple drives to reflect multiple choice tasks in a typical stated choice experiment, and (iii) a long drive in the simulation environment may induce fatigue earlier than the actual driving and increase the workload, which can compromise data quality. Therefore, we kept the maximum driving time to 12 mins for a scenario (with regular breaks between scenarios), allowing us to conduct five choice tasks while ensuring that the workload of the participants is reasonable and comparable to that in many existing studies using a driving simulator (Dell'Orco & Marinelli, 2017; Ali et al., 2020b,c). While the individual runs are shorter than what might be typical in a road safety experiment, the total driving time of each participant in the simulator (plus potential time savings and cost trade-offs) is representative of a typical trip within Sydney.

3.2.1. *Experiencing cost*

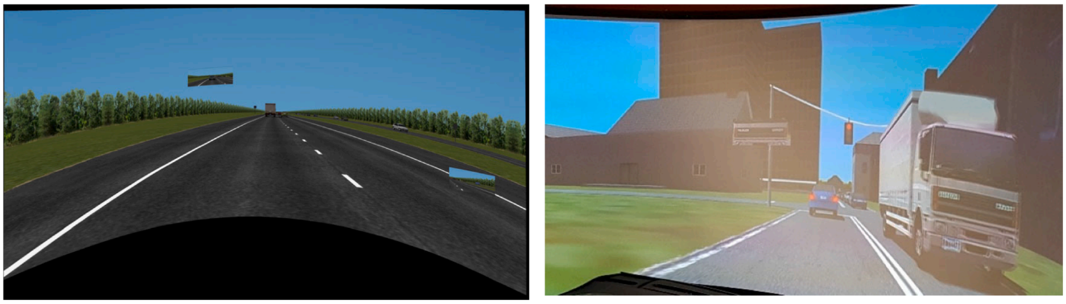
Each participant receives an initial endowment of $60 and are told during recruitment that their participation reward will vary on

Motorway	Urban road
Speed limit of **90 km/h**, **no traffic lights**.	Speed limit of **50 km/h**, **four traffic lights**.
The travel time is **6 minutes** every day. 6 6 6 6 6	The travel time varies. You will experience one of the following **travel times (in minutes)** with equal probability: 8 10 10 10 12
Toll cost: $ 2.00	Toll cost: **$ 0.00**

Fig. 3. Example of one of the SCE choice tasks presented to participants.

(a) Overview of the TRACSLab@USyd **(b) Cockpit of the simulator**

Fig. 4. Driving simulators at TRACSLab@USyd.



(a) Motorway **(b) Urban Road**

Fig. 5. Routes in driving simulator.

the choices made during the DSE[6]. To make the payment of the motorway toll consequential, respondents were shown a jar filled with one-dollar coins as per Fig. 6(a), and each time the motorway was chosen, respondents observed two dollars being removed from the reward jar placed into the "toll revenue" jar. After the experiment, the participant retains any remaining reward (paid out via a gift card). There is no cost associated with the urban road, so the reward jar is untouched if this alternative is chosen.

3.2.2. Experiencing travel time and unreliability

Irrespective of the alternative chosen, a participant must experience travel time by driving the chosen route. However, what is more novel in this experiment is the experience of travel time unreliability. Suppose again that the participant is faced with the example choice task shown in Fig. 3. The participant is asked to choose their preferred route, fully aware of the travel time distribution of the urban route (as well as other route characteristics). If the participant chooses the urban route, then we mimic unreliability by taking a random draw from the known travel time distribution. This is achieved by showing the participant five cards with travel times that replicate the travel time distribution (which the participant can verify). Fig. 6 (b) shows the five cards used for the travel time distribution shown in the example choice task in Fig. 3. Then the cards were shuffled face down and the participant was asked to pick a card (without seeing the travel time on the card). A scenario consistent with the randomly selected travel time was then loaded into the driving simulator for the participant to experience. It should be noted that:

- After the driving task the chosen card was revealed so that the participant could verify the experienced travel time. It is important to complete this verification process to maintain trust between the respondent and the analyst, and in the integrity of the experimental procedure.
- Respondents were clearly informed that the DSE would take a maximum of 90 min, however during the pre-DSE briefing they were told it would be possible to leave the experiment in a significantly shorter time period depending on the choices made[7].

4. Experimental design

We define four levels for toll cost, namely \$1, \$2, or \$3 for the motorway alternative and \$0 for the urban road alternative. Regarding the travel time distributions, the motorway alternative has a fixed travel time of 6 min (which equates to a mean travel time of 6 and a variance of 0) while for the urban road alternative, we consider combinations of five travel times, where each travel time can be 4, 6, 8, 10, or 12 min. These attribute levels were chosen for several reasons, but an important consideration is that each participant was limited to 90 min in the simulator. If a respondent was unlucky enough to generate 5 choice tasks each with a travel time of 12 min, this would require 60 min of driving time, with a further 15 min of enforced breaks (to avoid simulator sickness). With other associated time requirements of the experiment, this upper value of 12 min per task was the maximum that could reasonably be expected to be completed within the timeframe given. With regards to the remaining mix of times and costs, with these attribute levels, if a respondent was to select only the motorway option in each task, the expected cost across the 5 choice tasks would be \$10, with an approximate time saving of 30 min. This would equate to a value of time of approximately \$20 per hour, consistent with the VTT used by New South Wales Government.

In determining the correct design approach, it is imperative to already anticipate the specification of the econometric model, a point we turn to now. The two route alternatives are described by three attributes, namely travel (toll) cost, mean travel time, and the standard deviation of travel time (describing travel time unreliability). Adopting random utility theory in order to describe route choice, we define the utility of route i for respondent n in choice task t, denoted by U_{nti}, as

$$U_{nti} = V_{nti} + \varepsilon_{nti}. \tag{1}$$

where V_{nti} is the systematic utility and ε_{int} is a randomly distributed error term. We consider two routes, and assume that each route is described by a given travel time distribution, denoted by $\boldsymbol{T}_{nti}$, and given toll costs, C_{nti}. The systematic route utility is assumed to depend linearly on the average travel time, standard deviation of travel time, and toll cost. Furthermore, we assume an alternative-specific constant for the motorway alternative, using a dummy coded variable M_i that equals 1 for the motorway route and 0 for the urban road. Therefore, systematic route utilities are defined as:

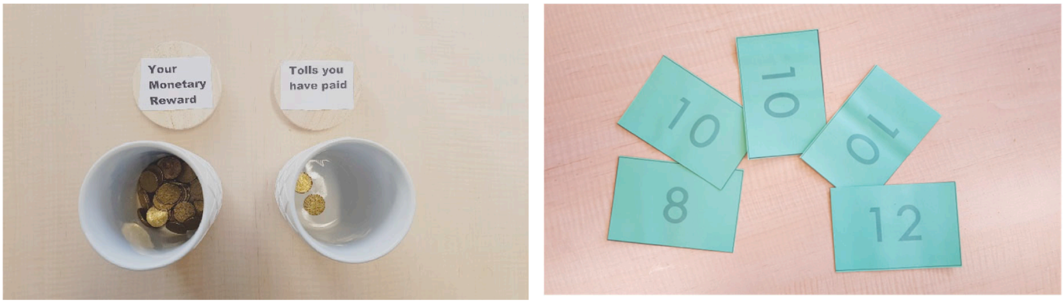
$$V_{nti} = \delta_M M_i + \beta_T E(\boldsymbol{T}_{nti}) + \beta_S \sqrt{var(\boldsymbol{T}_{nti})} + \beta_C C_{nti}. \tag{2}$$

where δ_M is the alternative-specific constant for motorway and $\boldsymbol{\beta} = (\beta_T, \beta_S, \beta_C)$ is a vector of marginal utility coefficients.

Eqn. (2) represents a mean–variance model that is often used to account for travel time unreliability. In the basic model, we assume that the error terms are independently and identically extreme value type I distributed, which means that route choice probabilities, P_{nti}, can be computed using a multinomial logit (MNL) model. That is:

[6] Participants were told the reward would vary between \$40 and \$60. Furthermore, a respondent received a fee of \$15 if they were not able to complete the study, e.g. due to motion sickness, or if they did not adhere to the rules regarding speeding and driving off-road in the driving simulator.

[7] That is, they can leave early if they finish driving the scenarios earlier if they select Motorway and pay the toll cost or if they get lucky on the Urban route by drawing a 4 or 6-minute travel time card each time.



(a) Experiencing cost **(b) Experiencing travel time unreliability**

Fig. 6. Simulated experiences.

$$P_{nti} = \frac{exp(V_{nti})}{\sum_{i'} exp(V_{nti'})}. \tag{3}$$

There exist 3 possible profiles for the motorway alternative, given the fixed travel time and the three possible toll levels. On the other hand, with five travel times shown for the urban road alternative in each scenario, drawn from five possible values (4, 6, 8, 10 or 12 min), there exist $5^5 = 3{,}125$ possible profiles for the urban road alternative. However, many of these profiles are essentially identical, as for example the travel time combination {4,6,6,10,12} represents the same distribution as {6,12,10,4,6} with the same mean and variance.

Considering only unique travel time distributions, which we represent in increasing travel time order to participants, there are 126 profiles left for the urban road alternative. Further, we only consider travel time distributions with a unique mean and variance, e.g. {6,6,10,10,10} has the same mean and variance as {6,8,8,8,12}, which removes 44 profiles. This means that in total, 246 different choice tasks exist, which are combinations of motorway and urban road profiles. Choice tasks where the urban road strictly or weakly dominates the motorway have been removed. For example, a choice task where the urban road has travel time distribution {4,4,4,4,4} is removed since the urban road strictly dominates the motorway (as it is cheaper, faster, and has the same reliability). If the urban road has travel time distribution {4,4,6,6,6} then it does not strictly dominate the motorway according to utility function (2) because while the urban road is cheaper and faster on average, it is less reliable. However, we can argue that a distribution of {4,4,6,6,6} is likely preferred over {6,6,6,6,6} despite travel time being less reliable, therefore we removed such choices tasks with a weakly dominant alternative from the candidate set. The final candidate set consists of 67 choice tasks.

Not all 67 choice tasks provide the same level of information for estimating the parameters in Eqn. (1). Given the limited number of participants in driving simulator experiments, we selected a subset of 15 choice tasks that provide high (Fisher) information for estimating the parameters in our model in Eqn. (1). This was achieved by generating a D-efficient experimental design through the modified Federov algorithm in Ngene (ChoiceMetrics, 2018) using our candidate set with 67 choice tasks. This design minimises the standard errors of the estimated model parameters and thereby minimises the required sample size (Bliemer et al., 2008; Rose & Bliemer, 2009). For the travel time and travel cost attributes, we assumed parameter priors based on estimates reported by Bliemer et al. (2017), who conducted a stated choice survey regarding route choice in Australia. For the standard deviation of travel time attribute, we assumed a prior consistent with a reliability ratio[8] of 0.85 (which is the average of the range 0.2 to 1.5 as reported in De Jong & Bliemer, 2015). The mean and standard deviation of travel time were computed for each of the 67 travel time distributions in the candidate set and once the experimental design was generated, they were converted back to travel time distributions consisting of five travel times.

Given that each participant could only spend a maximum of 90 min in the lab as stated in our ethics approval, each respondent was asked to complete only five choice tasks. Therefore, we blocked the design into three blocks of five choice tasks each where we aimed to have some degree of attribute level balance within each block (e.g. in each block the participant will face each of the toll levels at least once). The final experimental design is presented in Table 2.

The design of the experiment started in late 2016 and completed by mid-2017. It was followed by implementing, testing, and conducting pilot studies in the second half of 2017.

5. Sample and preliminary results

5.1. Participants

While many DSEs are conducted with students only, we opted to sample from the general population in order to get more variation in our sample (in particular with respect to age and income). We are not necessarily concerned with obtaining a representative sample

[8] Reliability ratio (RR) is defined as VOR divided by VTT (or marginal rate of substitution between travel time and standard deviation of travel time). Assuming a RR of 0.85, the resulting prior for standard deviation of travel time is equal to −0.122.

Table 2
Choice tasks and blocks.

Block	*Motorway*		*Urban Road*	
	Travel times (min)	Toll ($)	Travel times (min)	Toll ($)
I	6, 6, 6, 6, 6*	2	4, 4 ,10, 12, 12**	0
	6, 6, 6, 6, 6	1	8, 10, 10, 10, 12	0
	6, 6, 6, 6, 6	1	4, 6, 12, 12, 12	0
	6, 6, 6, 6, 6	2	4, 6, 6, 8, 10	0
	6, 6, 6, 6, 6	3	8, 10, 12, 12, 12	0
II	6, 6, 6, 6, 6	1	4, 6, 8, 10, 12	0
	6, 6, 6, 6, 6	1	8, 10, 12, 12, 12	0
	6, 6, 6, 6, 6	1	4, 4, 6, 8, 8	0
	6, 6, 6, 6, 6	2	4, 4, 4, 6, 12	0
	6, 6, 6, 6, 6	3	4, 6, 10, 12, 12	0
III	6, 6, 6, 6, 6	1	4, 6, 6, 8, 8	0
	6, 6, 6, 6, 6	2	8, 10, 10, 10, 12	0
	6, 6, 6, 6, 6	1	4, 6, 10, 12, 12	0
	6, 6, 6, 6, 6	2	8, 10, 12, 12, 12	0
	6, 6, 6, 6, 6	3	4, 6, 12, 12, 12	0

* 6 min every day
** 4 min two days per week, 10 min one day per week, and 12 min two days per week.

since we are not using the estimated VTT and VOR for appraisal purposes, but rather are mainly interested in differences in outcomes between the in SCE and DSE. Unlike SCEs, it is not easy to recruit non-student participants for the DSEs (despite a reward of up to A$60) since they typically require a larger time commitment and require the respondent to physically attend the lab.

We used two approaches to recruit participants, using advertisements and contacting participants from previous experiments conducted at TRACSLab. In advertisements, we adopted a number of different tools including a University of Sydney (USYD) volunteer database for research studies, posting study details via USYD Business School official Facebook and Twitter pages and the TRACSLab website, flyers at the University (including coffee shops), and free local classified ads (via www.gumtree.com.au).

During the recruitment process, participants complete a screening questionnaire which included supplementary questions designed to elicit information on their driving experience (i.e., how long they have been driving), age, sex, income, level of education, and occupation. The participants had to fulfil the following requirements:

a) Hold a driving license;
b) Drive at least 10 min, two or more days per week;
c) Be at least 18 years old;
d) Not suffer from motion sickness, vertigo, vestibular migraines or epilepsy;
e) Not suffer from any other medical conditions that impact the ability to drive.

Table 3
Socio-demographics of the participants.

Characteristics	Category	N = 74	
		Sample (%)	Greater Sydney (%)
Gender	Male	44.6	49.3
	Female	55.4	50.7
Age (in years)	18–29	28.4	26.9
	30–39	35.1	23.9
	40–65	36.5	49.1
Annual Personal Income (A$)	49,999 or less	29.7	50.4
	50,000–74,999	33.8	16.6
	75,000–99,999	17.6	10.5
	100,000 or more	16.1	13.4
	Not answered	2.7	9.2
Commuting to work (days per week)	≥ 5	44.6	NA
	4	21.6	NA
	≤ 3	33.8	NA
Occupation	Employed full-time	46.0	67.9
	Employed part-time	54.0	32.1
Education	Year 11 or less	0.0	17.2
	High School	13.5	35.5
	Associate degree (or Trade diploma)	14.9	15.2
	University degree or higher	70.2	22.5
	Not answered	1.3	9.6

The recruitment process was repeated three times in total, with waves in September 2017 to December 2017; February 2018; and finally November 2018. Although in these three attempts, we were able to recruit more than 300 participants, only 76 participants actually turned up at the TRACSLab. With two participants experiencing motion sickness, we retained a final sample of 74 participants. Each participant completed 5 choice tasks in both SCE and DSE, resulting in a total of 740 choice observations. Note that a sample size of 74 participants is high compared to most other driving simulator studies.

5.2. Descriptive statistics

The socio-demographic characteristics of the 74 participants that completed the experiment are presented in Table 3. Comparing our sample to the population of Greater Sydney (ABS, 2016), we do see an over-representation of respondents who are more highly educated (university degree or higher), and aged between 30 and 39 years. Respondents in the sample also tend to have relatively higher incomes.

Regarding the order of experiments, 43 participants completed the two experiments in Order 1 (SCE first) and 31 participants in Order 2 (DSE first). Fig. 7(a) shows the choice share of the motorway and urban road alternatives by experiment type. With respect to the experiment type, the Motorway is selected less often in the DSE (26%) compared to the SCE (36%). Therefore, these frequencies suggest that there may be differences in behaviour depending on the type of experiment.

Fig. 7 (b) further displays the choice shares depending on the order in which participants complete the experiments. We see that in Order 1 (SCE first), the motorway alternative is selected 34% in SCE compared to 22% in DSE. On the other hand, in Order 2 (DSE first), the motorway alternative is chosen 39% of the time in the SCE compared to 33% in the DSE. These differences in frequencies suggest that there exists an order effect, which is confirmed by a Chi-Square test ($\chi^2(1, N = 74) = 5.66, p < .05$). We account for ordering effects in our econometric analysis in Section 5. Note that these figures also reveal that the choices made in the SCE and the DSE differ to a greater extent in Order 1 than they do in Order 2.

6. Model Estimation and results

Given the novel experimental structure and exploratory nature of the work, we adopted a structured and incremental approach

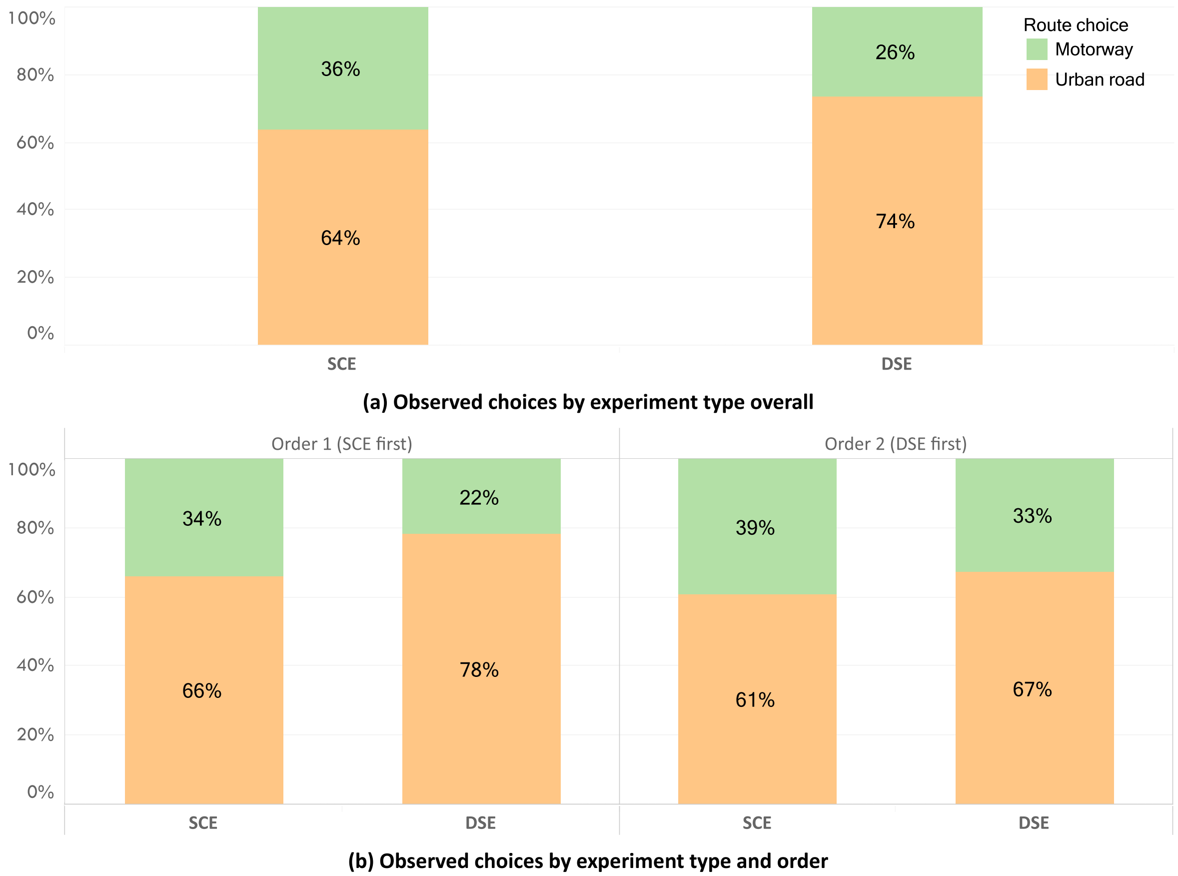


Fig. 7. Observed route choices in SCE and DSE.

toward estimating models starting with a basic route choice multinomial logit (MNL) then heteroscedastic logit (HL) models to account for scale differences across the two experiment types (SCE and DSE experiments) and also order effects, and finally a latent class (LC) model in order to account for preference heterogeneity across all attributes and the panel nature of the data (given that we have multiple observations from a single respondent). All models were implemented and estimated in R using Apollo (Hess & Palma, 2019). All models are estimated using pooled data (by combining SCE and DSE data). We excluded two participants from the dataset as they did not provide their income data, leading to a final sample of 72 respondents.

6.1. Multinomial logit model

In MNL 1, as discussed in Section 3.4, we estimate δ_M and $\boldsymbol{\beta}$. In MNL2, we investigate whether preferences for the motorway alternative are different across the two types of experiment, by estimating an additional shift parameter δ_{MD} for the motorway alternative in the DSE (i.e. when $M_i = 1$) which is used when task t for person n is a simulator task (i.e. when $D_{nt} = 1$):

$$V_{nti} = \delta_M M_i + \delta_{MD} M_i D_{nt} + \beta_T E(\boldsymbol{T}_{nti}) + \beta_S \sqrt{var(\boldsymbol{T}_{nti})} + \beta_C C_{nti}. \tag{4}$$

We next tested the inclusion of several sociodemographic variables where we found that only income had a statistically significant influence on route choice. In order to directly estimate the income elasticity λ_I we interact toll cost with a scaled income factor I_n for respondent n, where I_n is defined as income divided by average income. This gives us the following utility function for MNL 3:

$$V_{nti} = \delta_M M_i + \delta_{MD} M_i D_{nt} + \beta_T E(\boldsymbol{T}_{nti}) + \beta_S \sqrt{var(\boldsymbol{T}_{nti})} + \beta_C C_{nti} I_n^{\lambda_I}. \tag{5}$$

Table 4 presents parameter estimates for the three MNL models. In all models, we observe that the estimated parameters for the average travel time, toll cost, and travel time unreliability attributes have a negative sign, which is expected, and are statistically significant ($p < 0.05$). In all cases, the parameter of the motorway dummy is negative, indicating that the (tolled) motorway option is less attractive than the (untolled but unreliable) urban road option (ceteris paribus). Note that there was a relatively low negative correlation between the motorway dummy and toll cost (-0.22).

In MNL 2 and MNL 3, we see that the additional DSE shift for the motorway constant is negative and statistically significant, implying that motorway is disliked more in DSE than SCE. A possible explanation is that participants are more averse to the tolled motorway (irrespective of the toll level) when they have to pay actual toll costs in the DSE. We investigate this further in the latent class model in Section 6.3.

In MNL 3, the estimate of λ_I shows a negative and statistically significant income elasticity towards toll cost, meaning that the sensitivity towards toll cost decreases with an increase in income. The relatively low value for the λ_I may be explained by the fact that we have mostly high income participants in our sample. The improvements from MNL 1 to MNL 2 (the loglikelihood ratio (LR) test statistic is equal to 9.72 while $\chi^2_{1,\,95\,\%} = 3.84$) and then MNL 2 to MNL 3 (the LR test statistic is equal to 11.01 while $\chi^2_{1,\,95\,\%} = 3.84$) are statistically significant. MNL 3 will serve as the starting point for estimating heteroscedastic models in the next section where we take possible scale differences into account.

6.2. Heteroscedastic logit model

Our data originates from two different sources (SCE and DSE), and there may be different error variances (i.e. difference in scale) for each experiment type that need to be accounted for. In heteroscedastic logit (HL) models, we relax the assumption that the error variance is constant within the data. Furthermore, we also allow for difference in error variances based on order effects, i.e. which treatment was used first. This thus results in four scale terms in total.

Table 4
Estimation results for MNL models.

	MNL 1	*MNL 2*	*MNL 3*
Route attributes [a]			
Motorway(δ_M)	−1.631 (-4.73)	−1.381 (-3.77)	−1.376 (-3.75)
Motorway × DSE(δ_{MD})	—	−0.543 (-2.49)	−0.551 (-2.49)
Avg. travel time(β_T)	−0.512 (-7.30)	−0.519 (-7.29)	−0.528 (-7.42)
St. dev. of travel time(β_S)	−0.274 (-4.28)	−0.278 (-4.29)	−0.278 (-4.24)
Toll cost(β_C)	−0.799 (-7.46)	−0.812 (-7.44)	−0.780 (-6.66)
Income elasticity(λ_I)	—	—	−0.322 (-2.09)
Model fit			
LL (0)	−499.07	−499.07	−499.07
LL (final)	−397.33	−392.467	−386.963
Adj. Rho sq.	0.196	0.204	0.213
No. of parameters	4	5	6
BIC	820.97	817.83	813.40

[a] Robust t-ratio values against zero are in brackets.

We introduce a dummy variable O_n that equals 1 if respondent n faces the two experiments in Order 1 (SCE first), and zero in case of Order 2 (DSE first). Applying different scale parameters for each type of experiment depending on the order leads to the following formulation of the systematic route utilities:

$$V_{nti} = \overline{\mu}_{nt}\left(\delta_M M_i + \delta_{MD} M_i D_{nt} + \beta_T E(\boldsymbol{T}_{nti}) + \beta_S \sqrt{var(\boldsymbol{T}_{nti})} + \beta_C C_{nti} I_n^{\lambda_I}\right). \tag{6}$$

where we define

$$\overline{\mu}_{nt} = \begin{cases} \mu_{D1}, & \text{if } D_{nt} = 1 \text{ and } O_n = 1 \text{ (scale parameter for DSE, when SCE is taken first)}, \\ \mu_{S1}, & \text{if } D_{nt} = 0 \text{ and } O_n = 1 \text{ (scale parameter for SCE, when SCE is taken first)}, \\ \mu_{D2}, & \text{if } D_{nt} = 1 \text{ and } O_n = 0 \text{ (scale parameter for DSE, when DSE is taken first)}, \\ \mu_{S2}, & \text{if } D_{nt} = 0 \text{ and } O_n = 0 \text{ (scale parameter for SCE, when DSE is taken first)}. \end{cases} \tag{7}$$

In this model, which we refer to as HL1, we normalise $\mu_{D1} = 1$ and estimate three scale parameters, $\boldsymbol{\mu} = (\mu_{S1},\ \mu_{D2},\ \mu_{S2})$. Results are presented in Table 5.

In HL 1 we observe that all estimates for $\boldsymbol{\beta}$ remain statistically significant (and do so for all subsequent HL models). Examining the scale terms we note first that μ_{S1} is not statistically different from the base $\mu_{D1} = 1$. Second, scale parameter μ_{D2} is significantly different from the base of $\mu_{D1} = 1$, while μ_{D2} is not significantly different from scale parameter μ_{S2} (confirmed by testing whether $\mu_{D2} - \mu_{S2} = 0$, where this test incorporated the covariance between the estimates to ensure that the t-ratio on the difference reflects the maximum likelihood estimate properties of the original parameters; cf. Daly et al., 2012) and scale parameter μ_{S2} is significantly different from 1 at the 10% level. Our findings thus suggest that scale differences exist by order but not by experiment type (SCE vs DSE).

Given the above results, a second model (HL 2) was estimated that only includes scale differences due to experiment order, such that the systematic route utilities simplify to:

$$V_{nti} = \mu_2^{1-O_n}\left(\delta_M M_i + \delta_{MD} M_i D_{nt} + \beta_T E(\boldsymbol{T}_{nti}) + \beta_S \sqrt{var(\boldsymbol{T}_{nti})} + \beta_C C_{nti} I_n^{\beta_I}\right), \tag{8}$$

where we estimate scale parameter μ_2 corresponding to Order 2 (DSE first, $O_n = 0$) while scale is equal to 1 for Order 2 ($O_n = 1$). Parameter estimates shown in Table 5 suggest a lower scale parameter, and hence more error variance, for data collected using Order 2. One *possible* explanation for this result is that those respondents who complete the DSE first may exhibit more variety seeking behaviour and try out both routes in the simulator, e.g. participants are likely to sample the Motorway a number of times, or they are still learning about the nature of the experiment. Then, when completing the SCE two weeks later, they may make similarly stochastic choices in the context of the absence of consequence to the choices made.

On the other hand, another *possible* explanation is that, in Order 1 (SCE first), participants may be observed to make relatively more consistent choices because all information is presented in the SCE to participants and no travel time, travel time unreliability, or toll costs are experienced. When confronted with the DSE that now includes experience/experimental consequence, respondents are more incentivised to reveal their preferences towards these attributes and/or have learnt the nature of the experiment from the SCE and thus are less prone to variety seeking in the DSE.

While identifying one unique explanation for the result is not possible, a clear order effect can be observed via the impact on scale. Overall, we prefer HL 2 over HL 1 because model HL 2 is more parsimonious than HL 1 (the LR test statistic is equal to 0.71 whereas $\chi^2_{2,\,95\,\%} = 5.99$).

Building on model HL 2, we further explored preference heterogeneity towards route attributes in the two experiment types by estimating multipliers in the case of DSE, which leads to the following utility functions for model HL 3:

$$V_{nti} = \mu_2^{1-O_n}\left(\delta_M M_i + \delta_{MD} M_i D_{nt} + \kappa_T^{D_{nt}} \beta_T E(\boldsymbol{T}_{nti}) + \kappa_S^{D_{nt}} \beta_S \sqrt{var(\boldsymbol{T}_{nti})} + \kappa_C^{D_{nt}} \beta_C C_{nti} I_n^{\beta_I}\right). \tag{9}$$

where we estimate preference parameters $\boldsymbol{\beta}$, scale parameter μ_2 and attribute-specific multipliers $\boldsymbol{\kappa} = (\kappa_T, \kappa_S, \kappa_C)$, which only apply in case of the driving simulator experiment where $D_{nt} = 1$. Table 5 presents parameters for HL 3, where we clearly observe that none of the multipliers are statistically different from 1. Initially, we expected that participants would be more sensitive to time and cost attributes in the DSE because they had to experience them. However, from HL 3 we cannot draw this conclusion. HL 3 also does not improve model fit over HL 2 (the LR test statistic is equal to 1.02 whereas $\chi^2_{3,\,95\,\%} = 7.81$).

6.3. Latent Class model

To further explore potential differences in preferences, we estimate a latent class (LC) model with two classes, $q \in \{1,2\}$, with the following class-specific systematic utility functions:

$$V_{nti|q} = \mu_2^{1-O_n}\left(\delta_M M_i + \delta_{MDq} M_i D_{nt} + \kappa_T^{D_{nt}} \beta_{Tq} E(\boldsymbol{T}_{nti}) + \kappa_S^{D_{nt}} \beta_{Sq} \sqrt{var(\boldsymbol{T}_{nti})} + \kappa_{Cq}^{D_{nt}} \beta_C C_{nti} I_n^{\lambda_I}\right). \tag{10}$$

and for the class assignment model, we simply use utility function $V_q = \alpha_q$, where we normalise $\alpha_2 = 0$ and estimate only a constant for Class 1. We tried including sociodemographic variables in the class assignment utility function, but none were found to be statistically significant. We estimate class-specific preference parameters δ_{MDq} and $\boldsymbol{\beta}_q$ while we keep δ_M and λ_I generic across both classes since making them class-specific did not significantly improve the model fit. Also, attribute-specific multipliers $\boldsymbol{\kappa}$ are not considered class-specific to ensure that the model parameters are identifiable. We tried increasing the number of latent classes but a model with

Table 5
Estimation results for HL models.

	HL 1	*HL 2*	*HL 3*
Route attributes [a]			
Motorway(δ_M)	−1.977 (-3.34)	−1.784 (-4.14)	−1.578 (-3.27)
Motorway × DSE(δ_{MD})	−0.707 (-2.18)	−0.703 (-2.49)	−1.372 (-1.77)
Avg. travel time(β_T)	−0.721 (-5.56)	−0.654 (-9.84)	−0.606 (-6.77)
St. dev. of travel time(β_S)	−0.370 (-3.99)	−0.348 (-4.54)	−0.357 (-3.17)
Toll cost(β_C)	−1.010 (-6.07)	−0.937 (-7.54)	−0.990 (-4.78)
Income elasticity(λ_I)	−0.338 (-2.23)	−0.340 (-2.28)	−0.332 (-2.15)
Multipliers [b] *for DSE*			
Avg. travel time(κ_T)	—	—	1.244 (0.83)
St. dev. of travel time(κ_S)	—	—	1.045 (0.08)
Toll cost(κ_C)	—	—	0.910 (-0.28)
Scale [b]			
Order 1 DSE(μ_{D1})	1.000	—	—
Order 1 SCE(μ_{S1})	0.847 (-0.69)	—	—
Order 2 DSE(μ_{D2})	0.523 (-2.94)	—	—
Order 2 SCE(μ_{S2})	0.582 (-1.87)	—	—
Order 1(μ_1)	—	1.000	1.000
Order 2(μ_2)	—	0.590 (-2.68)	0.568 (-2.81)
Model fit			
LL (0)	−499.07	−499.07	−499.07
LL (final)	−381.79	−382.14	−381.629
Adj. Rho sq.	0.217	0.220	0.215
No. of parameters	9	7	10
BIC	822.79	810.34	829.05

[a] Robust *t*-ratio values against zero are in brackets; [b] Robust *t*-ratio values against one are in brackets

two latent classes was preferred based on the Bayesian Information Criterion (BIC), and consideration was also given to the application of the model and the practicability of the results (Beck et al., 2013).

Table 6 presents parameter estimates for the LC model where we accounted for the panel nature of the data[9] (i.e., we observe multiple choice observations from the same participant). We see a substantial improvement in model fit compared to the MNL and HL models (the LR test statistic for HL 2 vs. LC model is equal to 73.10 whereas $\chi^2_{8,\,95\,\%} = 15.51$). Based on $\alpha_1 = 0.135$, the shares of the two classes are 53% and 47% for Class 1 and Class 2, respectively, where these are not significantly different from an equal split.

Before discussing the two classes, we note that the order effect in terms of scale remains significant in this model. Further, we first discuss attribute-specific multipliers $\boldsymbol{\kappa}$. All multipliers are less than 1 suggesting that the impact of these attributes is reduced in the DSE relative to the SCE, though κ_T and κ_S are both not significantly different from 1 indicating that statistically these attributes carry the same preference weight in both experiments. However, with the case of κ_C, the multiplier is significantly less than 1, indicating that the impact of the toll cost attribute itself is less in the DSE than in the SCE. While this result may seem counterintuitive, the outcome needs to be considered in parallel to the shift in the alternative-specific constant for the motorway (δ_{MDq}).

Looking at the class-specific parameters, we see two different patterns. The parameters for Class 1 seem to indicate that respondents belonging to this latent class have a large aversion towards the tolled motorway in the DSE as expressed by the large negative value for δ_{MD1}, indicating that participants seem to prefer to avoid the motorway more in the DSE simply because it is a toll road, irrespective of the toll level (the estimated parameter for toll cost in the DSE is $\kappa_C\beta_{C1} = -0.524$). We also observe a relatively high aversion to travel time variability.

In contrast, in Class 2, respondents do not seem to have such a strong aversion to the motorway in the DSE (δ_{MD2} is not significant), nor to travel time variability (β_{S2} is not significant), rather seemingly trading mostly between travel time and cost. Toll levels matter to Class 2 respondents, given that the toll cost sensitivity is expressed by $\kappa_C\beta_{C2} = -1.482$, such that these respondents mainly switch to the urban road when the toll level on the motorway is high.

Overall, these results indicate that there are two classes of preference structures that exhibit differences in how the attributes of the alternatives are evaluated, and the significant cost multiplier μ_C shows that the evaluation of the cost attribute, and thus ultimately the trade-off between cost and time, differs in the case of the DSE as compared to the SCE. In other words, the absence or presence of consequences in a choice experiment seems to affect choice behaviour, in particular with respect to cost sensitivity. We explore the

[9] We also considered panel effects in each experiment (SCE and DSE) separately, where we found that only sensitivities towards toll costs were different in both experiments.

Table 6
Estimation results for the LC model.

	Class 1		*Class 2*
Route attributes [a]			
Motorway(δ_M)		−0.639 (-1.02)	
Motorway × DSE(δ_{MDq})	−2.530 (-2.81)		−1.500 (-1.68)
Avg. travel time(β_{Tq})	−0.926 (-6.34)		−1.013 (-3.77)
St. dev. Of travel time(β_{Sq})	−0.678 (-3.10)		−0.306 (-1.45)
Toll cost(β_{Cq})	−1.652 (-4.43)		−4.675 (-3.89)
Income elasticity(λ_I)		−0.237 (-1.66)	
Multipliers [b] *for DSE*			
Avg. travel time(κ_T)		0.757 (-1.57)	
St. dev. Of travel time(κ_S)		0.610 (-1.28)	
Toll cost(κ_C)		0.317 (-6.65)	
Scale [b]			
Order 1(μ_1)		1.000	
Order 2(μ_2)		0.568 (-2.81)	
Class allocation model [a]			
Class assignment factor(α_q)	0.135 (0.38)		0.000
Model fit			
LL (0)		−499.07	
LL (final)		−345.591	
Adj. Rho sq.		0.276	
No. of parameters		15	
BIC		789.87	

[a] Robust *t*-ratio values against zero are in brackets; [b] Robust *t*-ratio values against one are in brackets

implications of this result in the next section.

6.4. Comparison of VTT and VOR

In Table 7 we summarise the VTT values (β_T/β_C) and VOR values (β_S/β_C), assuming average income, i.e. $I_n = 1$. We also report the reliability ratios (RR = VOR/VTT). In a recent study based in Sydney, Douglas & Jones (2018) discuss VOR using the reliability ratio for which they found an average value equal to 0.37. We observe that the VTT values in this experiment are relatively high compared to other VTT values estimated using population samples in Australia, in particular, A$17.72/hr reported by TfNSW (2019), A$17.39/hr estimated by Hensher (2019) and A$7.33/hr estimated by Bliemer et al. (2017). In Table 8 we present sample and population average VTT and VOR in the LC model based on income representations in Table 2, which shows that the population average VTT and VOR values are indeed lower.

Our high VTT value may be explained by two factors. First, most of our participants have a relatively high income (see Table 3, for descriptive statistics). Second, since the motorway in our study has a non-zero toll (whereas the Urban Road has zero tolls), there are some confounding effects between the motorway constant, δ_M, and the toll cost attribute (with related parameter β_C) in capturing the departure from a zero cost.

For the above reasons, the VTT and VOR values found in this study are not directly comparable to other values found in the literature. However, we remind the reader that the objective of this paper is not to generate VTT and VOR values that apply to the population, rather we are seeking to compare route choice behaviour in the two data collection techniques, and where the choice tasks are identical with the exception of simulated experiences and monetary incentives being enforced in the DSE.

Fig. 8 visualises VTT and VOR results from the LC model. The confidence intervals for the error bars are calculated via the Delta method which uses the robust variance–covariance matrix resulting from model estimation. It can also be seen that the VTT and VOR

Table 7
Travel time-saving values (in A$/hr).

	MNL 1	*MNL 2*	*MNL 3*	*HL 1*	*HL 2*	*HL 3*		*LC*			
						SCE	*DSE*	*SCE*		*DSE*	
								Class 1	*Class 2*	*Class 1*	*Class 2*
VTT	38.41	38.35	40.59	42.85	41.88	36.69	43.10	33.63	13.00	44.88	17.35
VOR	20.57	20.54	21.34	21.99	22.29	21.62	23.14	24.64	3.93	30.13	4.81
RR	0.54	0.54	0.53	0.51	0.53	0.59	0.54	0.73	0.30	0.67	0.28

Table 8
Sample and population average VTT and VOR values for the LC model (in A$/hr).

		SCE		DSE	
		Class 1	Class 2	Class 1	Class 2
VTT	Sample	31.40	12.14	41.90	16.20
	Greater Sydney	29.56	11.43	39.44	15.25
VOR	Sample	23.01	3.67	28.13	4.49
	Greater Sydney	21.66	3.46	26.48	4.23

are statistically different across the classes, but not between SCE and DSE (within the same class). While we cannot conclude that VTT and VOR are different for SCE and DSE, we did observe significant differences in preferences towards toll road and toll level in the DSE. Therefore, failure to observe differences in VTT and VOR across experiment types may be the result of sample size limitations in this study.

7. Discussion and conclusions

This paper investigates drivers' route choice preferences in a typical hypothetical SCE and a DSE that exhibits a greater degree of incentive compatibility, namely via time and cost consequences experienced by participants, building on what is an otherwise sparse body of research. Our experimental treatment ensures that the same respondents complete identical choice tasks in both experimental treatments, being the first study of its kind to use a DSE to estimate VTT and VOR measures in a route choice context. This is particularly novel in the context of VOR, where the ability to reconstruct objective or perceived route travel time distributions on road networks is close to impossible in a revealed preference format. Importantly, our analysis reveals that when controlling for all else, there exist differences in the preference structures exhibited in a hypothetical SCE compared to a DSE with simulated consequences. As such, our research has found evidence in the transport domain that the absence of consequences in a typical SCE contributes to hypothetical bias.

Before turning to the concluding remarks there are limitations to this study that should be noted. While the DSE simulates time and cost consequences for respondents, we could not include any outcome-related consequences such as being early/late for work. Also, it is possible that some people obtained a positive utility from driving in the simulator due to 'game' feeling (although all people were observed to be generally tired at the end of the DSE). Moreover, risk-seeking drivers may choose the urban road alternative because of the gambling aspect (although this may equally be true in real life route choices). As previously stated, the sample size is relatively small and not representative, therefore VTT and VOR values reported in this paper should only be used for model comparison and not for appraisal purposes. Lastly, for future research related to the experimental design of such experiments, we suggest adding a toll level of zero to the motorway to clearly disentangle the relationship between toll cost and travel time.

With regards to the results presented in this paper, initial analysis on the choice shares revealed that the motorway alternative was significantly less attractive in the DSE compared to the SCE, particularly when the SCE was completed first (Order 1). Subsequent modelling revealed that the motorway alternative was indeed less attractive, as evidenced by the alternative specific constant which was significant across all estimated models; and the motorway alternative was found to be particularly less attractive in the DSE. Subsequent mathematical modelling revealed significant differences in the error variance of choices, with the consistently lower scale parameter for Order 2 indicating that choices exhibit greater error variance when the DSE is completed before the SCE.

Analysis also revealed preference heterogeneity among respondents; with one preference class broadly exhibiting strong aversion to the motorway alternative and travel time variability, while the other preference class showed dislike towards high toll cost. When examining any impact that the type of experiment might have on preferences, we found that the impact of cost was different in the DSE than in the SCE, in that the participants avoid the motorway more in the DSE. Regarding preferences towards travel time and travel time unreliability, we could not find significant differences between the two types of experiments. Therefore, the type of experiment (with or without simulated experiences) seems to affect responses to monetary attributes differently to time attributes. Despite these differences, we could not reject the hypothesis that VTT and VOR are same in the typical SCE and the DSE, which is likely due to our limited sample size.

The similarities and differences between the typical SCE and incentive compatible DSE from this research may have implications for the design of route choice studies. Our study has some insights particularly when there is a tolled road, given that cost has a different impact in the DSE than in a typical SCE. Many decisions regarding new toll road projects are based on analysis from SP surveys, given the significant investments are made in constructing toll roads there is a need for robust inputs into this decision-making process. Assuming that the DSE is closer to the true preferences of a respondent through greater incentive compatibility, we recommend the adoption of incentive compatible DSE to allow for the simulation of not only the new tolled route, but also to gather robust valuations of time and reliability that could be used to calibrate SCE data that may exhibit a higher degree of hypothetical bias. There is potential for future research to examine whether the DSE experiments are better able to recover what would be revealed preferences. For route choice studies that include only travel time and travel time unreliability, a SCE may suffice.

Future research may also seek to provide a robust explanation for why the experiment order produces a significant difference in error variances estimated on the data, and in turn what this might mean for determining which process may produce more robust estimates of VTT and VOR, given concerns of hypothetical bias in stated choice methods. Our analysis revealed that choices made in the SCE differ more to those in the DSE when done first (Order 1) and the subsequent modelling revealed that the Order 1 choices were

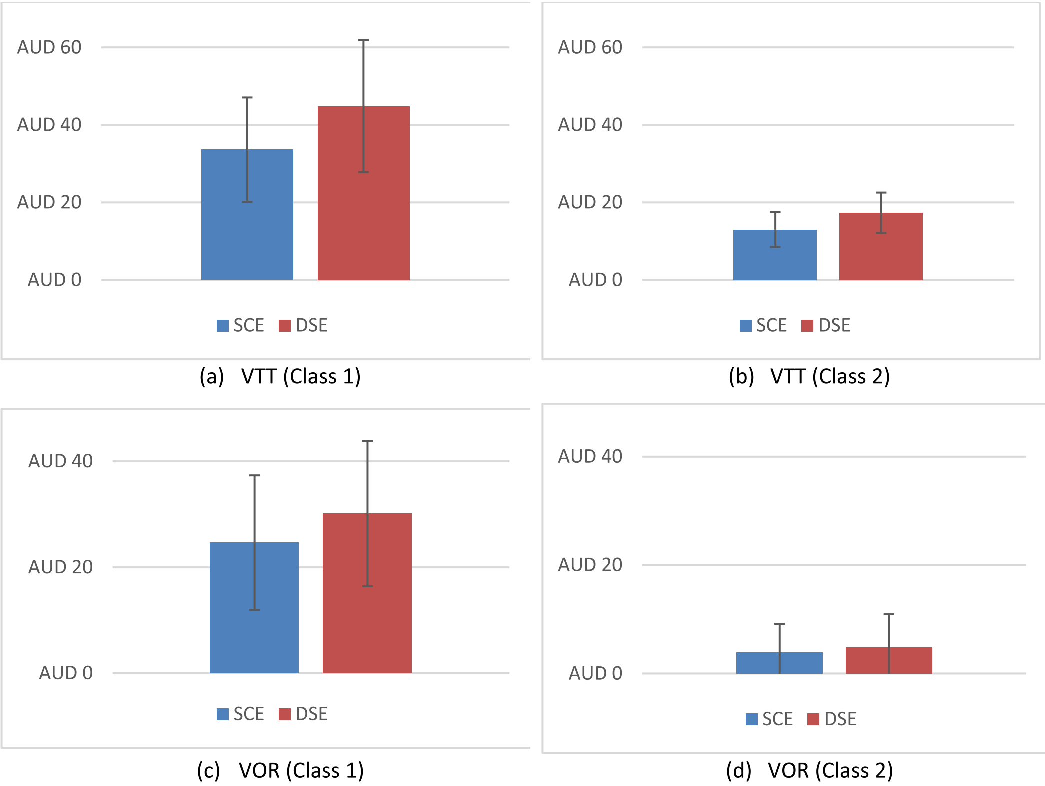


Fig. 8. VTT and VOR (per hour) across the two classes in the LC model.

also more deterministic, which is often the case in stated preference data. On the other hand, while the Order 2 choices produced a lower scale term and thus less deterministic choices, the difference in the choices observed across the two experimental treatments were not as great. We believe that it is likely the case that the choices in the DSE are more representative of the underlying preferences towards travel time and travel time variability (and thus exhibit a potentially more realistic level of variability). It could thus be argued that in completing the DSE first, respondents learn their preferences more fully in this simulated environment and these preference structures are recalled afterwards in the SCE. In may be the case that more robust values can be estimated cost effectively, by simply giving respondents a limited number of trials in a DSE environment, followed by a larger number of SCE tasks to elicit valuations.

This type of approach is not uncommon. For example, Hess et al. 2020b combined SCE and DSE data for efficiently estimating lane choice models, and a similar approach is also suggested by Bhat & Castelar (2002) who find insufficient variation in RP data to estimate a significant cost coefficient, arguing instead for a joint RP-SP approach. Indeed, the authors argue that their evidence points toward using SP experiments as the main data source for analysis and supplementing with small samples of RP data for anchoring with actual market activity. This research takes the first steps in examining differences produced by hypothetical choices and those with simulated consequence in the context of valuing travel time and reliability. However, given the potential implications of this research discussed in this section, ongoing research is strongly encouraged to explore the potential of driving simulators and/or simulated consequence as a way of bridging the gap between stated and revealed preference.

CRediT authorship contribution statement

Muhammad Fayyaz: Conceptualization, Methodology, Formal analysis, Writing - original draft, Writing - review & editing. **Michiel C.J. Bliemer:** Conceptualization, Methodology, Funding acquisition, Formal analysis, Writing - original draft, Writing - review & editing. **Matthew J. Beck:** Conceptualization, Methodology, Formal analysis, Writing - original draft, Writing - review & editing. **Stephane Hess:** Conceptualization, Methodology, Funding acquisition, Formal analysis, Writing - review & editing. **J.W.C. (Hans) Lint:** Conceptualization, Methodology, Writing - review & editing.

Acknowledgements

This research is funded by Australian Research Council grants DP150103299 and DP180103718. Stephane Hess also acknowledges support for his inputs by the European Research Council through the consolidator grant 615596-DECISIONS.

References

Abrantes, P.A.L., Wardman, M.R., 2011. Meta-analysis of UK values of travel time: An update. Transportation Research Part A 45, 1–17. https://doi.org/10.1016/j.tra.2010.08.003.

ABS. (2016). Australian Bureau of Statistics, General Community Profile, Greater Sydney, 2016, data cube: Excel spreadsheet, viewed 10 June 2019. https://quickstats.censusdata.abs.gov.au/census_services/getproduct/census/2016/communityprofile/1GSYD?opendocument.

Ali, Y., Zheng, Z., Mazharul Haque, M., Yildirimoglu, M., Washington, S., 2020a. Understanding the discretionary lane-changing behaviour in the connected environment. Accident Analysis & Prevention 137 (January), 105463. https://doi.org/10.1016/j.aap.2020.105463.

Ali, Y., Bliemer, M.C., Zheng, Z., Haque, M.M., 2020b. Comparing the usefulness of real-time driving aids in a connected environment during mandatory and discretionary lane-changing manoeuvres. Transportation Research Part C: Emerging Technologies 121, 102871.

Ali, Y., Sharma, A., Haque, M.M., Zheng, Z., Saifuzzaman, M., 2020c. The impact of the connected environment on driving behavior and safety: A driving simulator study. Accident Analysis & Prevention 144, 105643.

Beck, M.J., Fifer, S., Rose, J.M., 2016. Can you ever be certain? Reducing hypothetical bias in stated choice experiments via respondent reported choice certainty. Transportation Research Part B: Methodological 89, 149–167. https://doi.org/10.1016/j.trb.2016.04.004.

Beck, M.J., Rose, J.M., Hensher, D.A., 2013. Environmental attitudes and emissions charging: An example of policy implications for vehicle choice. Transportation Research Part A: Policy and Practice 50, 171–182. https://doi.org/10.1016/j.tra.2013.01.015.

Bhat, C.R., Castelar, S., 2002. A unified mixed logit framework for modeling revealed and stated preferences: formulation and application to congestion pricing analysis in the San Francisco Bay area. Transportation Research Part B: Methodological 36 (7), 593–616. https://doi.org/10.1016/S0191-2615(01)00020-0.

Black, I., Towriss, J., 1993. Demand effects of travel time reliability. Center for Logistics and Transportation, Cranfield Institute of Technology.

Bliemer, M.C.J., Rose, J.M., Chorus, C.G., 2017. Detecting dominance in stated choice data and accounting for dominance-based scale differences in logit models. Transportation Research Part B: Methodological 102, 83–104. https://doi.org/10.1016/j.trb.2017.05.005.

Bliemer, M.C.J., Rose, J.M., Hess, S., 2008. Approximation of Bayesian efficiency in experimental choice designs. Journal of Choice Modelling 1 (1), 98–126. https://doi.org/10.1016/S1755-5345(13)70024-1.

Bonsall, P., Firmin, P., Anderson, M., Palmer, I., Balmforth, P., 1997. Validating the results of a route choice simulator. Transportation Research Part C: Emerging Technologies 5 (6), 371–387. https://doi.org/10.1016/S0968-090X(98)00003-5.

Bonsall, P.W., 2004. Route Choice Simulators. In: Human Behaviour and Traffic Networks. Springer, Berlin Heidelberg, pp. 167–191. https://doi.org/10.1007/978-3-662-07809-9_7.

Broadbend, C. (2012). Hypothetical Bias, Consequentiality and Choice Experiments. In Intergovernmental Panel on Climate Change (Ed.), Economics Bulletin, AccessEcon (Vol. 32, Issue 3). Cambridge University Press. https://doi.org/10.1017/CBO9781107415324.004.

Brownstone, D., Bunch, D.S., Train, K., 2000. Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles. Transportation Research Part B: Methodological 34 (5), 315–338. https://doi.org/10.1016/S0191-2615(99)00031-4.

Brownstone, D., Small, K.A., 2005. Valuing time and reliability: Assessing the evidence from road pricing demonstrations. *Transportation Research Part A: Policy and Practice*, *39*(4 SPEC. ISSS.), 279–293. https://doi.org/10.1016/j.tra.2004.11.001.

Bwambale, A., Choudhury, C.F., Hess, S., 2019a. Modelling long-distance route choice using mobile phone call detail record data: a case study of Senegal. Transportmetrica A: Transport Science 15 (2), 1543–1568. https://doi.org/10.1080/23249935.2019.1611970.

Bwambale, A., Choudhury, C.F., Hess, S., 2019b. Modelling departure time choice using mobile phone data. Transportation Research Part A: Policy and Practice 130 (October), 424–439. https://doi.org/10.1016/j.tra.2019.09.054.

Carrion, C., Levinson, D., 2012. Value of travel time reliability: A review of current evidence. Transportation Research Part A: Policy and Practice 46 (4), 720–741. https://doi.org/10.1016/j.tra.2012.01.003.

Carrion, C., Levinson, D., 2013. Valuation of travel time reliability from a GPS-based experimental design. Transportation Research Part C: Emerging Technologies 35, 305–323. https://doi.org/10.1016/j.trc.2012.10.010.

ChoiceMetrics. (2018). Ngene 1.2 User Manual & Reference Guide, Australia.

Daly, A., Hess, S., de Jong, G., 2012. Calculating errors for measures derived from choice modelling estimates. Transportation Research Part B: Methodological 46 (2), 333–341. https://doi.org/10.1016/j.trb.2011.10.008.

de Jong, G.C., Bliemer, M.C.J., 2015. On including travel time reliability of road traffic in appraisal. Transportation Research Part A: Policy and Practice 73, 80–95. https://doi.org/10.1016/j.tra.2015.01.006.

Dell'Orco, M., Marinelli, M., 2017. Modeling the dynamic effect of information on drivers' choice behavior in the context of an Advanced Traveler Information System. Transportation Research Part C: Emerging Technologies 85, 168–183. https://doi.org/10.1016/j.trc.2017.09.019.

Ding, M., Grewal, R., Liechty, J., 2005. Incentive-Aligned Conjoint Analysis. Journal of Marketing Research 42 (1), 67–82. https://doi.org/10.1509/jmkr.42.1.67.56890.

Dixit, V.V., Harb, R.C., Martínez-Correa, J., Rutström, E.E., 2015. Measuring risk aversion to guide transportation policy: Contexts, incentives, and respondents. Transportation Research Part A: Policy and Practice 80, 15–34. https://doi.org/10.1016/j.tra.2015.07.002.

Djukic, T., Wilmink, I., Jonkers, E., Snelder, M., & van Arem, B. (2016). Exploratory Analysis of Traveller's Compliance with Smartphone Personal Route Advice: In a Field Trial Amsterdam.

Donmez, B., Boyle, L.N., Lee, J.D., 2007. Safety implications of providing real-time feedback to distracted drivers. Accident Analysis & Prevention 39 (3), 581–590. https://doi.org/10.1016/j.aap.2006.10.003.

Douglas, N., Jones, M., 2018. Estimating the Value of Private Travel Time for Nsw. *Australasian Transport Research.* Forum November, 1–31.

Fezzi, C., Bateman, I.J., Ferrini, S., 2014. Using revealed preferences to estimate the value of travel time to recreation sites. Journal of Environmental Economics and Management 67 (1), 58–70. https://doi.org/10.1016/j.jeem.2013.10.003.

Fifer, S., Rose, J., Greaves, S., 2014. Hypothetical bias in Stated Choice Experiments: Is it a problem? And if so, how do we deal with it? Transportation Research Part A: Policy and Practice 61, 164–177. https://doi.org/10.1016/j.tra.2013.12.010.

Foster, H., & Burrows, J. (2017). Hypothetical bias: a new meta-analysis. In Contingent Valuation of Environmental Goods (pp. 270–291). Edward Elgar Publishing. https://doi.org/10.4337/9781786434692.00016.

Ghosh, A., 2001. Valuing time and reliability: Commuters' mode choice from a real time congestion pricing experiment. PhD Thesis. University of California at Irvine.

Haghani, M., M.C.J. Bliemer, J.M. Rose, Harmen Oppewal, and E. Lancsar (2021) Hypothetical bias in stated choice experiments: integrative synthesis of empirical evidence and conceptualisation of external validity. Pre-print, arXiv:2102.02940.

Haghani, M., Sarvi, M., 2018. Hypothetical bias and decision-rule effect in modelling discrete directional choices. Transportation Research Part A: Policy and Practice 116 (March), 361–388. https://doi.org/10.1016/j.tra.2018.06.012.

Haghani, M., & Sarvi, M. (2019). Laboratory experimentation and simulation of discrete direction choices: Investigating hypothetical bias, decision-rule effect and external validity based on aggregate prediction measures. Transportation Research Part A: Policy and Practice, 130(April 2018), 134–157. https://doi.org/10.1016/j.tra.2019.09.040.

Hensher, D., 2019. Using the Average Wage Rate to Assess the Merit of Value of Travel Time Savings: A Concern and Clarification. Transport Findings 1–6. https://doi.org/10.32866/5772.

Hensher, D.A., 2001. The valuation of commuter travel time savings for car drivers: Evaluating alternative model specifications. Transportation 28 (2), 101–118. https://doi.org/10.1023/A:1010302117979.

Hess, S., Daly, A., Börjesson, M., 2020. A critical appraisal of the use of simple time-money trade-offs for appraisal value of travel time measures. Transportation 47 (3), 1541–1570. https://doi.org/10.1007/s11116-020-10097-w.

Hess, S., Choudhury, C. F., Bliemer, M. C. J., & Hibberd, D. (2020b). Modelling lane changing behaviour in approaches to roadworks: Contrasting and combining driving simulator data with stated choice data. Transportation Research Part C: Emerging Technologies, 112(May 2019), 282–294. https://doi.org/10.1016/j.trc.2019.12.003.
Hess, S., Daly, A., Dekker, T., Cabral, M.O., Batley, R., 2017. A framework for capturing heterogeneity, heteroskedasticity, non-linearity, reference dependence and design artefacts in value of time research. Transportation Research Part B: Methodological 96, 126–149. https://doi.org/10.1016/j.trb.2016.11.002.
Hess, S., Palma, D., 2019. Apollo: a flexible, powerful and customisable freeware package for choice model estimation and application version 0.0.6 User manual. Journal of Choice Modelling. https://doi.org/10.1016/j.jocm.2019.100170.
Hultkrantz, L., Savsin, S., 2018. Is 'referencing' a remedy to hypothetical bias in value of time elicitation? Evidence from economic experiments. *Transportation* 45 (6), 1827–1847. https://doi.org/10.1007/s11116-017-9803-1.
Koutsopoulos, H.N., Lotan, T., Yang, Q., 1994. A driving simulator and its application for modeling route choice in the presence of information. Transportation Research Part C 2 (2), 91–107. https://doi.org/10.1016/0968-090X(94)90002-7.
Krčál, O., Peer, S., Staněk, R., Karlínová, B., 2019. Real consequences matter: Why hypothetical biases in the valuation of time persist even in controlled lab experiments. Economics of Transportation 20 (February), 100138. https://doi.org/10.1016/j.ecotra.2019.100138.
Levinson, D., Harder, K., Bloomfield, J., Winiarczyk, K., 2004. Weighting waiting - Evaluating perception of in-vehicle travel time under moving and stopped conditions. Transportation Research Record: Journal of the Transportation Research Board 1898, 61–68. https://doi.org/10.3141/1898-08.
Li, Z., Hensher, D.A., Ho, C., 2020. An empirical investigation of values of travel time savings from stated preference data and revealed preference data. Transportation Letters 12 (3), 166–171. https://doi.org/10.1080/19427867.2018.1546806.
Loomis, J., 2011. What's to know about hypothetical bias in stated preference valuation studies? Journal of Economic Surveys 25 (2), 363–370. https://doi.org/10.1111/j.1467-6419.2010.00675.x.
Nielsen, O.A., 2004. Behavioral Responses to Road Pricing Schemes: Description of the Danish AKTA Experiment. Journal of Intelligent Transportation Systems 8 (4), 233–251. https://doi.org/10.1080/15472450490495579.
Papinski, D., Scott, D.M., Doherty, S.T., 2009. Exploring the route choice decision-making process: A comparison of planned and observed routes obtained using person-based GPS. Transportation Research Part F: Traffic Psychology and Behaviour 12 (4), 347–358. https://doi.org/10.1016/j.trf.2009.04.001.
Peer, S., Knockaert, J., Koster, P., Verhoef, E.T., 2014. Over-reporting vs. overreacting: Commuters' perceptions of travel times. Transportation Research Part A: Policy and Practice 69, 476–494. https://doi.org/10.1016/j.tra.2014.07.005.
Penn, J., Hu, W., 2018. Understanding Hypothetical Bias: An Enhanced Meta-Analysis. American Journal of Agricultural Economics 100 (4), 1186–1206. https://doi.org/10.1093/ajae/aay021.
Penn, J., Hu, W., 2019. Cheap talk efficacy under potential and actual Hypothetical Bias: A meta-analysis. Journal of Environmental Economics and Management 96, 22–35. https://doi.org/10.1016/j.jeem.2019.02.005.
Prato, C., Rasmussen, T., Nielsen, O., 2014. Estimating Value of Congestion and of Reliability from Observation of Route Choice Behavior of Car Drivers. Transportation Research Record: Journal of the Transportation Research Board 2412, 20–27. https://doi.org/10.3141/2412-03.
Ramos, G.M., Frejinger, E., Hoogendoorn, S., 2012. A revealed preference study on route choices in a congested network with real-time information. 13th International Conference on Travel Behaviour Research.
Rose, J.M., Bliemer, M.C.J., 2009. Constructing Efficient Stated Choice Experimental Designs. Transport Reviews 29 (5), 587–617. https://doi.org/10.1080/01441640902827623.
Saifuzzaman, M., Zheng, Z., Haque, M.M., Washington, S., 2017. Understanding the mechanism of traffic hysteresis and traffic oscillations through the change in task difficulty level. Transportation Research Part B: Methodological 105, 523–538. https://doi.org/10.1016/j.trb.2017.09.023.
Senna, L., 1994. The influence of travel time variability on the value of time. Transportation 21 (2), 203–228. https://doi.org/10.1007/BF01098793.
Sharma, A., Zheng, Z., Bhaskar, A., Haque, M.M., 2019. Modelling car-following behaviour of connected vehicles with a focus on driver compliance. Transportation Research Part B: Methodological 126, 256–279. https://doi.org/10.1016/j.trb.2019.06.008.
Shires, J.D., de Jong, G.C., 2009. An international meta-analysis of values of travel time savings. Evaluation and Program Planning 32 (4), 315–325. https://doi.org/10.1016/j.evalprogplan.2009.06.010.
Small, K. A. (1999). Valuation of travel-time savings and predictability in congested conditions for highway user-cost estimation. In NCHRP Report (Issue 431). National Academy Press.
Small, K.A., Winston, C., Yan, J., 2005. Uncovering the distribution of motorists' preference for travel time and reliability. Econometrica 73 (4), 1367–1382. http://www.jstor.org/discover/10.2307/3598825?sid=21105516803263&uid=3737664&uid=2&uid=4.
Smith, V.L., 1976. Experimental Economics: Induced Value Theory. American Economic Review 66 (2), 274–279.
TfNSW. (2019). Transport for NSW Economic Parameter Values (Issue September).
Tseng, Y.-Y., Verhoef, E., de Jong, G., Kouwenhoven, M., van der Hoorn, T., 2009. A pilot study into the perception of unreliability of travel times using in-depth interviews. Journal of Choice Modelling 2 (1), 8–28. https://doi.org/10.1016/S1755-5345(13)70002-2.
Vacca, A., Prato, C.G., Meloni, I., 2019. Should I stay or should I go? Investigating route switching behavior from revealed preferences data. Transportation 46 (1), 75–93. https://doi.org/10.1007/s11116-017-9788-9.
van Essen, M., Thomas, T., Chorus, C., van Berkum, E., 2019. The effect of travel time information on day-to-day route choice behaviour: evidence from a real-world experiment. Transportmetrica B: Transport Dynamics 7 (1), 1719–1742. https://doi.org/10.1080/21680566.2019.1699198.
Varela, J.M.L., Börjesson, M., Daly, A., 2018. Quantifying errors in travel time and cost by latent variables. Transportation Research Part B: Methodological 117, 520–541. https://doi.org/10.1016/j.trb.2018.09.010.
Young, W., Sobhani, A., Lenné, M.G., Sarvi, M., 2014. Simulation of safety: A review of the state of the art in road safety simulation modelling. Accident Analysis & Prevention 66, 89–103. https://doi.org/10.1016/j.aap.2014.01.008.