Article:

Hess, S orcid.org/0000-0002-3650-2518, Daly, A and Börjesson, M (2020) A critical appraisal of the use of simple time-money trade-offs for appraisal value of travel time measures. Transportation, 47. pp. 1541-1570. ISSN 0049-4488

https://doi.org/10.1007/s11116-020-10097-w

A critical appraisal of the use of simple time-money trade-offs for appraisal value of travel time measures

Stephane Hess
Institute for Transport Studies & Choice Modelling Centre
University of Leeds
s.hess@its.leeds.ac.uk

Andrew Daly
Institute for Transport Studies & Choice Modelling Centre
University of Leeds
andrew@alogit.com

Maria Börjesson
VTI Swedish National Road and Transport Research Institute and KTH Royal Institute of Technology
maria.borjesson@vti.se

Abstract

Stated choice surveys have established themselves as the preferred approach for value of travel time elicitation with the help of choice models. However, major differences exist in the approach used across regions and contexts. In Europe (particularly Northern Europe), value of travel time is often estimated in large national studies, which continue to rely extensively on simple time-money trade-offs. On the other hand, studies in Australia and South America in particular tend to have a more local focus and follow the notion that more complex setups are preferable. The European studies however are also those where the results are actually used in cost-benefit analysis (CBA) and data from European studies have formed a testbed for many advanced model specifications. The present paper aims to provide a critical appraisal of the use of simple time-money trade-offs, drawing from our experience in recent European studies. We highlight a number of issues, in terms of differences in valuations across formats as well as a lack of clarity on how respondents actually interpret travel time in these simple time-money trade-offs.

Keywords: value of time; time-money trade-offs; stated choice

1. Introduction

Monetary valuations of travel time (VTT) form one of the main inputs into the appraisal of new transport infrastructure and policy schemes, and their robust estimation is therefore of considerable societal, financial and environmental importance. The need to derive robust estimates has also motivated important developments in the theory and practice of modelling choices and willingness to pay, developments that extend beyond the context of value of travel time to more general valuation (De Borger and Fosgerau, 2008; Fosgerau, 2006, 2007; Rouwendal, et al., 2010).

While early studies (tracing back to Beesley, 1965 and Daly and Zachary, 1975) were based on real-world observation of travellers choosing between modes, the estimation of VTT has for the last three decades generally been performed via the analysis of Stated Choice (SC) data. While revealed preference (RP) data, including in the form of new *big* data sources, is gradually making a return in VTT work (e.g. Brownstone et al., 2005; Bwambale et al., 2019), such data is similarly beset by difficulties (e.g. see Varela et al., 2018), and especially large scale national VTT studies are likely to keep a strong focus on hypothetical data for a while yet.

At least in Northern Europe, and the UK in particular, where national value of time studies are popular, the analysis of simple time-cost trade-offs has been prioritised. These countries (UK, Denmark, The Netherlands, Sweden and Norway) account between them for a very large share of national VTT studies, and are also the key countries where VTT estimates are used extensively in cost-benefit analysis (CBA). However, in other parts of the world, more complicated trading exercises have been preferred in national VTT studies, notably in Switzerland (Axhausen et al., 2008) and Germany (Ehreke et al., 2015). The same applies to large numbers of more regional or even local VTT studies in Australia and South America. While an important argument in favour of simple trade-offs in applied work has been that other approaches are too complex or complicated, the counter-argument in academic work has been that what really matters is relevance (cf. Hensher, 2006). With the continuing reliance on simple trade-offs in some national studies, the issue therefore arises as to which approach should be adopted, especially if different results are obtained with different methods.

Although there is a geographic divide in terms of the type of approaches used, we argue that the credibility of the simple money-time trade-offs is relevant for a world-wide audience also because these experiments have been used as a testbed for developing many advanced modelling approaches, especially in terms of heterogeneity, such as in the work of Fosgerau (2006) and Fosgerau & Bierlaire (2007). Such experiments have also been used extensively in understanding that the data must support a sufficient range of the distribution to be able to estimate the mean of the VTT distribution and the sensitivity of the result with respect to distributional assumptions (Fosgerau, 2007; Börjesson et al., 2012). With simple time-money trade-offs, this is possible. However, if the experiments themselves are not (behaviourally) reliable, then this casts some doubt on this benefit.

This paper discusses the theoretical and behavioural implications as well as empirical evidence and finds that there are indications that simple time-cost trade-offs give significantly different results and this could lead to the conclusion that they are less satisfactory than more complex exercises in obtaining the required values. It is worth noting that our discussion specifically excludes the scenario where there is a choice between a free road and a faster toll road (or managed express lane). Such scenarios have been studied extensively in the United States and present travellers with a real time-money trade-off (see e.g. Brownstone and Small, 2005, Hess et al., 2011, Hossan et al., 2016, Lam and Small, 2005, and Small et al., 2005). However, they are not commonly used for national studies, which seek to establish a reliable long-term measure of VTT for appraisal rather than a journey specific one. In addition, valuations from toll road studies are often very corridor specific, are restricted to specific countries and can be beset by issues of strategic bias.

Before proceeding with our paper, it is worth briefly mentioning the history of these simple time-money trade-offs and provide some reasons for the resulting inertia in terms of methods. It was the 1987 national study of the value of travel time in the UK (MVA et al., 1987) that marked the acceptance of stated preference (SP), including SC, for use in important transport policy work. The argument in the 1987 study was that SC appeared to give acceptable error margins for VTT estimates with sample sizes that were feasible for the budgets available, while error margins from the revealed preference data available at that time were excessive. This study proved influential for example for the 1990 study for The Netherlands (Hague Consulting Group, 1990) and the following study for the UK (Accent and Hague Consulting Group, 1996). Perhaps unfortunately, a standard methodology was thus established and it then proved difficult for the civil services commissioning these important studies to deviate from the standard. For example, the subsequent Netherlands study (Hague Consulting Group, 1998) and the Swedish (Algers et al., 1995) and Norwegian (Ramjerdi et al., 2010) studies adopted very similar methods. In each case a simple binary time-money trade-off, sometimes presented as a route choice, formed a key component of the survey and analysis. Moreover, once repeated studies had been undertaken in a particular country, interest focussed on the change in VTT from the previous study, so that it was again very attractive for civil services to be able to compare results from a consistent methodology.

The above studies generally (though not always) also included scenarios other than simple time-money trade-offs, for example looking at the valuation of reliability or the valuation of travel time in different conditions of travel. However, the key appraisal values have been based on simple time-money trade-offs and results from other games (e.g. valuation of reliability) have come from separate (rather than joint) models and have been used simply as multipliers of the VTT measures from the time-money trade-offs, thus implicitly making the assumption that the base VTT is the same across formats.

As mentioned earlier, the use of simple time-money trade-offs in European national VTT studies differs substantially from the approach used elsewhere. Indeed, notably in Australian work (see e.g. Hensher & Rose, 2007), the argument has been put forward that it is preferable to present respondents with a single type of scenario (rather than different sets of choices across games) where this includes all attributes that may be relevant in the choice, and to allow respondents to focus on what really matters to them (see e.g. Hensher, 2006). Including all attributes together in a single experiment also has the statistical benefit that the parameters for all attributes are estimated on the basis of the full data.

Simple time-money trade-offs have often been the topic of heated discussions at conferences, with criticism notably from Australia and South America, but continue to be used in some national studies in Europe. In this paper, we revisit this issue on the basis of our own personal experience, going back to the 1970s for Daly and the early 2000s for Hess & Börjesson, having contributed to a number of national studies that have relied on simple time-money trade-offs, as well as to more complicated valuation studies. This gives us a strong position to re-evaluate the reliability of this approach without simply attacking the work done by others. We revisit the pros and cons of the different approaches, and present new empirical evidence. In doing so, we also address an issue not discussed in the literature relying on simple time-money trade-offs, namely that of consistency of results across different formats of experiments (e.g. time-money vs time-money-reliability) presented to the same respondent. If differences arise across separate survey formats in the valuation for the same journey component, for example the VTT, then the question arises which of these should be used. Furthermore, if valuations for different journey characteristics (e.g. value of time vs value of reliability) are obtained from different survey components, and if any common valuations (typically the VTT) are not consistent across components, then this raises doubts about the wisdom of combining results from different components. Our discussions focus solely on the setup of the choice tasks in terms of the attributes included (i.e. simply time-money vs more attributes) rather than the number of alternatives, which is a different discussion in itself.

The remainder of this paper is organised as follows. Section 2 presents some a priori considerations, while Section 3 explores the issues in more detail in the context of the recent British study. Section 4 presents additional evidence from other studies, before we summarise our findings in Section 5.

2. A priori considerations

We discuss the characteristics of simple time-money trade-offs under five headings, namely a) considerations of micro-economic theory, b) survey design considerations, c) behavioural considerations, d) modelling implications, and e) interpretation.

2.1. Considerations of micro-economic theory

The value of travel time depends on three factors: the resource value of time (or opportunity value, the utility that could be attained if the travel time was used for some other activity), the direct utility of travel time (compared to some reference activity), and the marginal utility of money (DeSerpa, 1971; Jara-Díaz & Guevara, 2003; Jara-Díaz, 2007), with $\text{VTT} = \frac{\text{resource value of time} - \text{direct utility of travel time}}{\text{marginal utility of money}}$. The direct utility of travel time is influenced by factors such as the comfort and productivity or pleasure gained on the trip. The resource value of time should increase the less available time the traveller has, and hence should vary depending on socio-economic status. The marginal utility of money is clearly related to income.

The resource value of time depends on the traveller's time-use pattern and scheduling function, which may vary between trip contexts. A higher resource value of time is probably the main reason for higher value of time for commuting than other trip purposes. To capture the resource value of time in stated choice scenarios, the experiments are generally related to a reference trip that the respondent has recently undertaken. Respondents are asked to respond to the choice scenario in the context of this reference trip and its related time-use pattern. We cannot disentangle the resource value of time from the direct utility of time. However, in all estimations, we model the variations in resource value of time across respondents by controlling for socio-economic variables such as trip purpose, travel time, family situation, employment, household size and obligations, and income, to name just a few. There are studies explicitly taking the daily time scheduling adaptation process into account for variations in the resource value of time between and within people (Habib et al. 2013, Weis et al. 2013, Schmid, 2017).

Although the respondents are instructed to have the context of a specific trip in mind when making their choices, it is possible that the hypothetical setting implies that they pay less attention to the impacts of their choices on the time availability for other activities. Aside from investigating the possibilities of revealed preference data, an interesting area for future work would be to test the differences in VTT that would arise from surveys looking not at a single trip but explicitly at the scheduling for an entire day. Even then, respondents may pay too little attention to the resource time of other days.

Assuming that the respondents follow the instructions and make the choices in the context of the reference trip, the VTT estimates produced from simple time-money trade-offs reflect the direct utility of travel time in the reference trip, including modal effects, but also reflect the resource value of time in this trip.[1] The differences in direct utility of time across modes and journey circumstances (e.g. purpose) can therefore only be explored by using an experiment in which the trip context and/or mode varies for a given traveller and reference trip (for which the resource value of time stays constant). Changing the mode of course leads to a requirement for a hypothetical leap of faith by the respondent which may impact data quality. Modal effects can clearly also be captured by looking at a mode choice experiment as opposed to separate route choice experiments with different modes but the dominance of specific modes may then impact on the data quality.

Assuming a representative sample, the overall results can then be used as values of time (including the direct value of time and the resource value of time components) in current conditions. However, their use would prevent us from studying changes in the VTT as a result of changes in the direct utility that would arise from changes in travel conditions (e.g. a change in highway congestion or public transport crowding), as these conditions do not change across choice tasks or alternatives in simple time-money trade-offs. For this purpose, data is needed from experiments that explicitly incorporate a quality-of-travel dimension.

In this section, we have stayed within the bounds of micro-economic theory which is what the appraisal frameworks are based on. However, there are of course many behavioural considerations that would imply departures from the pure microeconomic model. This includes numerous topics from behavioural economics, including for example the presence of different mental accounts (cf. Thaler, 1999), both for different types of costs and for money spent on journeys of different length and purpose. It would violate the microeconomic framework to let the sensitivity to different types of costs vary for a single individual. It would however not violate the microeconomic framework to let the sensitivity to travel time and different travel time components vary for an individual. The sensitivity to travel time is related to the wage rate (for travellers that can choose working hours freely), but also to the time constraints which vary across trips, and the direct utility of travel time which may also vary across trips (depending on the comfort etc). Hence, the departures from the micro-economic framework can be reduced by attributing the different values of time not to different cost sensitivities, but

[1] For instance, the differences in the VTT across modes arise from differences in the direct utility of time in the reference trip, due to differences in comfort and productivity of the travel time. However, the modal differences in the VTT arise also from self-selection, i.e. travellers with high resource value of time will tend to choose faster but more expensive modes, or those in which the travel time can be used for other purposes.

to different time constraints and differences in the direct utility of time. Another key issue is that of reference dependence. Demand functions of classical microeconomics, which describe consumer behaviour over a period and for a range of prices, are incompatible with reference dependence. However, reference dependence is established in many simple money - time trade off SP studies (De Borger and Fosgerau 2008; Hultkrantz and Mortazavi 2001; Börjesson and Eliasson 2014; Bates and Whelan 2001; Hjorth and Fosgerau 2011; Cantillo et al., 2006; Daly et al. 2014) and this is the context we are describing[2]. This could reflect real short-term preferences, because loss aversion and reference dependence are well established real preferences. However, they only exist in the short-term in relation to a well-defined reference point. In stated choice experiments based around a recently made trip, to increase realism, the reference point is strongly emphasised and the respondents are explicitly asked to reschedule a trip of a day with an already set schedule This is very different even from the context of a day some months or weeks ahead, for which scheduling is usually more flexible and the reference point less well-defined. Such reference-dependence would not reflect the population's long-term stable preferences, since the reference points and even the population change over time. Welfare economic analysis relies on long-term stable preference in the aggregate population. That is, even if preferences change as people grow old, and in fact the population itself changes, the preferences need to be stable at the aggregate level and not subject to short-term reference points.

2.2. Survey design considerations

In terms of survey design, an initial argument for simple time-money trade-offs might have been a greater ability to ensure that the choices presented were all meaningful (i.e. avoiding complete dominance, where one option is both cheaper and faster, or quasi-dominance, where for example one option is marginally faster but very much more expensive) and also covered the full spectrum of trade-offs (in terms of sign and size effects, i.e. looking at both increases and decreases in time and cost, and also changes of different amounts). While this argument would indeed have had some validity in the 1980s, experimental design techniques have evolved dramatically since, with algorithms and software that enable analysts to construct informative choice tasks with multiple attributes leading to data with good statistical properties.

There remain advantages and disadvantages for either a simple or a complex approach at the design stage. Simple time-money trade-offs can give greater control over the specific value of time trade-offs presented (as they are the only attributes). On the other hand, designs with more attributes and alternatives will lead to richer datasets (given the larger number of possible combinations) and also have more flexibility in terms of dealing with dominance. Indeed, they can present trade-offs where one alternative is slower and more expensive but has advantages for some other characteristic. It is important to note that such scenarios will still contribute to an analyst's ability to understand trade-offs between time and cost. Indeed, the majority of model implementations do not work in trade-offs between cost and time, but in differences in overall utility. Even if an alternative is slower and more expensive, the size of the actual differences will matter.

[2] See Hess, Daly & Batley (2018) for recent discussions on consistency with utility maximisation.

2.3. Behavioural considerations

An initial key argument for simple time-money trade-offs has been that of reduced respondent burden, which was a common concern in the early days of SC surveys. In choice modelling, the issue has been investigated across different areas of application for example by Hensher (2004, 2006).

There is also much discussion in the experimental and psychology literature about response burden. Fischhoff (2013) concludes that dealing with many attributes is cognitively and analytically cumbersome. Hence, presenting complex choices at random induces a risk that the respondent only focuses on a single aspect or spend less effort on some aspects to reduce the cognitive demands (Lichtenstein et al., 1978). In psychology, Jacoby et al. (1974) claimed to have found the existence of "information overload” in experiments with too many attributes. Staelin and Payne (1976) however do not find a declining accuracy in responses as the number of attributes increases. Malhotra (1982) finds that decision accuracy decreased when the number of attributes increased from 5 to 15 and even further to 25 (thus very far from two attributes). Bell at al. (1977) argue that two-attribute games might be useful, since the cognitive limitations of decision makers otherwise often force them to disregard information, even if it is relevant. Moreover, Keller and Staelin (1987) find that more information quantity does reduce decision effectiveness - but more information quality increases decision effectiveness (enriching the choice context). Wilson and Scholler (1991) advise against including all possible attributes in all decisions, because this forces subjects to think too much about their reasons for different choices, biasing the outcome. On the other hand, inspired by Hammond and Adelman (1976), Lichtenstein et al. (1978) reason that reducing the number of attributes destroys the respondents’ intuition. They argue that respondents’ cognitive processes work best in the context of past experiences and therefore are context dependent. Hence, attempts to evaluate a complex choice on the basis of a limited number of the attributes is likely to produce spurious outcomes. In summary, the exact number of attributes that is optimal is not clear from this literature. However, it is in most cases probably more than two. The lessons from such work are arguably too generic for the present paper, not least as the choices in a transport setting are often very familiar to respondents, which already decreases the a priori burden. The debate in academia about very complex surveys is ongoing, but here our focus is on VTT and the contrast between very simple and somewhat more complex games, and therefore this issue is not of central concern in the present paper.

The concerns about response burden in value of time surveys goes back to a time when most surveys were conducted on paper or via telephone survey. However, important improvements have been made and the move to computer or web-based surveys allows a customised survey environment where for instance explicit times can be calculated rather than showing “10 minutes more” etc[3]. Web-based surveys are also likely to reduce response error compared to telephone surveys (cf. Börjesson & Algers, 2011).

[3] This is of course different from saying that web-based surveys themselves have lower burden than paper-based surveys, and the work on respondent burden (Axhausen et al., 2015) suggests this is indeed not the case.

Choice modellers have engaged extensively with the issue of survey complexity in wide ranging empirical tests. Much of this work is influenced by the study of Hensher (2004) who showed variations in WTP measures as a function of design dimensionality. In related work, Caussade et al. (2005) show that the variance of the error term in choice models seems to indeed increase with the number of attributes presented – this however simply means that there is more scope for other factors to influence choices, including heterogeneity in sensitivities for more attributes and/or differences in interpretation, and does not necessarily lead to biased estimates or different choice behaviour. More recent evidence (Chintakayala et al., 2010) suggests that increases in error only occur with more than 6 attributes and in fact suggests that overly simplistic scenarios may lead to behaviour that is more difficult to model (i.e. more randomness). Other work has shown how analysts can allow for the fact that some respondents may focus on a subset of the attributes only, i.e. filtering out information not relevant to them (e.g. Hensher, 2008). In this context, Hess (2014) also suggests little or no difference in willingness-to-pay (WTP) measures when additional attributes are included in the SC scenarios, which could point towards simpler surveys being acceptable.

However, the insights from these studies for the present paper are somewhat limited – Caussade et al (2015) used different respondents for different treatments, while in Hess (2014), the full set of attributes was used for a respondent prior to focussing on a subset only, so directionality effects may exist. In typical European VTT studies, the simple experiment tends to be given first, and the directionality may thus be different.

Strongly related to the point about respondent burden is the argument that in studies primarily (or solely) interested in estimating the VTT, the inclusion of other attributes in the experimental setup may simply act as a distraction, for both the modeller and the respondent. We return to the modelling implications later on, and for now focus solely on the behavioural effect. The argument that a simple time-money trade-off leads to an '*unpolluted*' measure of the VTT can quite reasonably be undermined by a counter-argument that this setup in fact leaves more space for interpretation open to the respondent, and hence an increase in noise and potentially also bias. If respondents simply "struggle" to make the choices in the absence of other important information, then this will lead to increased error variance in the models. More seriously, respondents may enhance the information, for example inferring that a faster journey is less reliable (e.g. a faster route might be more popular and thus prone to accidents or congestion) or that a cheaper train is less comfortable. Such unobservable (to the analyst) effects may lead to bias in the estimates that are of interest, a point we return to in Section 2.4[4].

This point about missing information has directly led to the view by e.g. Hensher (2006) that what really matters is relevance. In other words, analysts should see possible increases in respondent burden, and/or the possibility of a respondent ignoring some of the presented attributes, as a lesser evil than the possibility of a respondent *enhancing* the presented information in an unobservable manner. As always, there are two sides to this argument and we should acknowledge that this may also be study specific.

[4] Although less studied in the literature, it is conceivable that there is also an impact on the statistical accuracy of estimates, i.e. standard errors.

The focus on just time and money may also have two additional implications. Firstly, a key motivation for using hypothetical choice settings rather than transfer price (i.e. contingent valuation) approaches is a recognition that the latter may lead respondents to purposefully bias their answers up or down depending on their vested interests. A simple time-money framework moves away from such direct elicitation, but is arguably still so transparent as to make it clear to respondents that they are being asked (albeit not directly) to give monetary valuations of travel time. Secondly, in real life contexts, respondents pay for improvements along a number of different dimensions (e.g. time, reliability, safety). To focus on just one in isolation can be imagined to either overstate the value (all the money can be spent on just travel time) or understate the value (an increase in cost is more visible in a two-attribute context).

Research in choice modelling has in recent years also focussed extensively on a number of behavioural effects, in particular non-linearity in preferences, and asymmetric preference formation around reference points (e.g. Hess et al., 2008; de Borger & Fosgerau, 2008), anchoring, lexicographic choice, inconsistent behaviour and non-trading (e.g. Hess et al., 2010) and heteroskedasticity (see discussions in Hess et al., 2017). These effects have been studied extensively in both simple and more complex trade-offs, but the question remains whether or not they are more prevalent in one or the other. This would lead to the conclusion that these effects are at least to some extent influenced by the experimental setup.

Non-trading and lexicographic choices especially have received extensive attention in stated preference work. To put this work into context, it should be noted that any real-world non-trading will be due to the fact that a given traveller is observed over a set of trips that present only limited incentive or reason to change behaviour. If there were indeed travellers who would never change their behaviour, independent of the incentive provided to them, then analysing their behaviour using random utility models or computing value of time measures for them would be meaningless as their VTT would be either zero or close to infinity. Neither scenario is in our view very likely. In stated choice surveys, it is of course not immediately clear whether a respondent does not trade because of a lack of engagement with the survey or because the incentives presented are too small. The former would be a cause for concern about survey validity while the latter is to be expected but may impact on the estimation of distributions of VTT. It is in this context that the work of Börjesson et al. (2012) is highly relevant. They showed that in the Swedish study, extending the ranges of the trade-offs presented to respondents led to reductions in the rates of apparent non-trading, providing some reassurance that respondents do indeed engage with the format. On the other hand, presenting respondents with excessively wide ranges just to encourage trading may lead to choice scenarios that are so far removed from a real world setting as to have other detrimental effects. Care is thus needed – do we want to extend ranges just to ensure respondents are trading or do we want to present only realistic trade-offs which might mean non-trading is reasonable?

Anchoring and apparent inconsistent behaviour present two additional issues. We readily acknowledge that real world behaviour may well be anchored to past experiences. Similarly, behaviour may change across days in such a way as to appear “inconsistent” (e.g. accepting to pay for a faster journey on one day that implies a VTT of at least v_1 while on another day refusing to pay in a trade-off that implies a VTT of not more than v_2, where $v_1 > v_2$). However, stated preference surveys are "instantaneous panels" where a given respondent is told to make a sequence of independent choices, all relating to the same journey. There is in that case little behavioural justification for finding evidence of anchoring or inconsistency other than seeing these as survey artefacts.

Our final point, and one that has motivated the most criticism of simple time-money trade-offs, is that of consistency with real world choice settings, often under the broad term of "realism". In the introduction to this paper, we made the point that it is rare for a traveller to face a real-world choice between options that are distinguished only by one being faster but more expensive and the other being slower but cheaper. While time-money trading may arise in a public transport context, it will often (though not always) involve a mode choice (e.g. expensive but fast train vs cheap but slow bus) or different sub-modes (e.g. high speed vs conventional rail), thus bringing in implicit issues of comfort, reliability and status as well as the simple time and money differences. SP studies have looked at both between and within mode choices for public transport. The core interest in many appraisal studies is on car, with the obvious real world context where a faster journey costs more arising in the case of toll roads. However, as already mentioned in the introduction, not only are toll roads rare or non-existent in many countries, but they are also too area specific to be suitable for national value of time studies, and are potentially subject to strategic bias. They have as a result largely been ignored in the experiments leading to national VTT measures. Studies have instead attempted to address the realism issue by not talking about route choice but a choice between *different hypothetical contexts*. Whether a respondent can be expected to make such a leap of faith is open to question.

The effects of the issues mentioned above are difficult to establish in empirical evidence given the latent nature of the behavioural processes that drive the results but also the fact that the data in different settings typically comes from different respondents. Our paper thus specifically focusses on comparisons where the same respondents provide data in more than one type of stated choice setting.

2.4. Modelling implications

The modelling implications of the number of attributes in an SC experiment fall into two broad categories, namely the role of presented and omitted attributes, and the detection and modelling of behavioural phenomena in the model.

The standard approach for estimating value of time measures is based in random utility theory. In general terms, the utility function for alternative *i* as faced by individual *n* in choice situation *t* can be written as:

$$V_{int} = \sum_{k=1}^{K} f_k(x_{kint}, \beta_{kn}) + \varepsilon_{int} \quad (1)$$

In this specification, $x_{k_{int}}$ is known to the analyst, while β_{kn} is unknown, but estimated from the data. If some of the attributes in $x_{int} =< x_{1int}, \dots, x_{kint} >$ do not influence behaviour, the associated parameters will become close to zero. Testing whether the inclusion of additional attributes leads to excessive burden is an empirical issue, testing for example whether $var(\epsilon_{int}) = h(K)$.

A completely different rationale applies to omitted attributes. Let $o_{int} = < o_{1int}, \dots, o_{Lint} >$ be a set of omitted attributes, which, for the decision maker, have a set of influences on behaviour represented through λ_n. We can then write:

$$\varepsilon_{int} \sim \sum_{l=1}^{L} g_l(o_{lint}, \lambda_{ln}) \tag{2}$$

Unlike in Equation (1), both o_{int} and λ_n in Equation (2) are unknown to the analyst, where this includes not just the values but also the existence of individual elements thereof.

A key assumption in random utility models in this specification is that, aside from independence, the error terms, ε_{int}, are distributed identically across alternatives and across choice scenarios. Respondents in surveys are commonly instructed to assume that any omitted characteristics are the same across all the presented alternatives. If this instruction is adhered to, then the above requirement would be satisfied. If however, respondents make assumptions about omitted attributes on the basis of the presented attributes, then we end up in a situation where:

$$var(\epsilon_{int}) \neq var(\epsilon_{jnt}), \forall(i,j) \tag{3}$$

and/or:

$$cov(V_{jnt}, \epsilon_{jnt}) \neq 0, \forall j. \tag{4}$$

Either of these issues can arise if the respondent attempts to infer additional information about a utility on the basis of the presented attributes, such as a faster journey being less reliable, for example. This creates correlation between observed and unobserved attributes, leading to endogeneity and potential bias in the estimates of β. It also creates heteroskedasticity across alternatives. While both issues can potentially be addressed empirically, they would necessitate additional information about how the respondents interpret the scenarios, lead to a requirement for a more complex modelling framework, and rely on further assumptions by the analyst about the processes taking place in ϵ_{jnt}.

Since the original development of choice models, there has been an interest in capturing the full richness of behavioural phenomena, be it heterogeneity across individual decision makers or the detection or modelling of specific behavioural traits. With both of these, the presence of just two attributes (and two alternatives) of course greatly simplifies the task faced by an analyst, and this can be seen as a key reason for the use of time-money trade-offs in a number of methodological studies in choice modelling, notably in work looking at ever more flexible treatments of inter-respondent heterogeneity (see e.g. Fosgerau, 2006, 2007; Fosgerau & Bierlaire, 2007).

With simple binary time-money trade-offs, a number of behavioural phenomena can also be detected directly by *observing* the data. Examples of this include apparent lexicographic, non-trading or inconsistent behaviour. The direct *inspection* of such issues becomes significantly more complicated in the case of more than two attributes (or more than two alternatives), or in the presence of non-linearity and asymmetry, and it is generally no longer possible to assign a single label to the observed behaviour. With simple binary time-money trade-offs, it also becomes possible to graphically inspect the data by plotting indifference curves (Fosgerau 2007, Börjesson et al. 2012) and this can then highlight the presence of heteroskedasticity and non-linearity/asymmetry in preferences. The ability to detect such phenomena directly in the data is of course an advantage that facilitates analysis and can help an analyst choose an appropriate model structure, but this should in no way mean that we no longer care about other possible disadvantages.

When it comes to the actual modelling of the behavioural phenomena, the presence of more than two attributes clearly complicates matters. When attempting a flexible specification of heterogeneity across decision makers with more complex datasets, there are clear increases in computational complexity and data requirements. In addition, the number of parameters to be estimated rises rapidly when seeking to capture correlation between individual randomly distributed sensitivities. With a simple time-money trade-off, an analyst is also able to determine a boundary VTT in each choice task, and this can be used not only to study anchoring effects but can also be beneficial in understanding random heterogeneity (c.f. Börjesson et al., 2012). However, once again, the study of these phenomena is also possible in the case of more than two attributes, as we highlight in the empirical section below. On the other hand, restricting the data to just time-money trade-offs (and in particular in a binary setting) also greatly limits the richness of the resulting data, given that there is less scope for variability and a lower dimensionality of the data. This can raise questions about the ability to reliably estimate models with distributed parameters.

2.5. Interpretation

The first issue arising in the interpretation of results from simple time-money trade-offs is the question of what type of time is valued. There is a risk that if the travel conditions are not stated explicitly for each alternative, it is not clear what conditions of travel (e.g. congestion, crowding) the respondent has in mind when valuing the time, and this would in turn influence the direct utility of the value of time.

In the interest of accurately capturing the resource value of time component of the VTT by making the respondents' time and money budget concrete, surveys often ask respondents to make choices as if they were in the context of a *current* or *recent* reference journey. The usual assumption is then that the VTT component reflecting the direct value of time coming out of time-money trade-offs reflects conditions similar or equal to those experienced on that reference journey. However, because the travel conditions are not explicitly stated it is far from certain whether this is the case. For example, a respondent may consider a difference in time between presented car options as necessarily involving a difference in congestion. If, as we hypothesised earlier, it is the case that respondents may *enhance* the information presented to them, by for example imagining that a faster journey is less reliable, then this is again likely to lead to issues with interpretation when focussing solely on changes in travel time and cost. This would then create the kinds of issues discussed in Section 2.4 and potentially lead to biased estimates of the VTT. The absence of such biases is an assumption that has in our opinion not been questioned sufficiently thus far.

It is on the other hand also possible that explicitly describing some of the travel conditions in the experiment gives rise to a focus effect related to the stated travel conditions, when in a real life setting, the travel conditions may not be as clear to a traveller a priori when making a journey decision. Empirical testing may not lead to conclusive findings, but we make some attempts later on in this paper.

The argument can be extended further to the cost attribute if one is willing to entertain the notion that respondents have different sensitivities to different cost components. This is problematic in itself since different sensitivities to different cost components are not micro-economically consistent. However, such results are supported by much empirical evidence from surveys with multiple cost attributes such as parking, toll and fuel costs for cars, for example (see for instance Vrtic et al., 2007).

In the earlier section on behavioural effects, we have already alluded to the question whether phenomena such as reference dependence and non-linearity (commonly referred to as sign and size effects) are real world effects or survey artefacts. Namely, if we observe in empirical work based on stated choice data that a respondent reacts more strongly to increases in time than to decreases (compared to a reference point) and that the sensitivity to a larger change is smaller per unit of time than the sensitivity to a smaller change, then would the same hold for real world behaviour by that respondent?

Reference dependence is of course clearly also a real-world phenomenon. Our question is simply whether what is picked up from data on hypothetical choices is the same as what happens in reality. A particular concern is that sign and size effects operate primarily in the short term, manifesting themselves only while there exists a well-defined reference point. In particular, value of time experiments implicitly require respondents to consider how to reschedule a reference trip to accommodate a travel time change in the short run, which is likely to increase reference dependence (Börjesson and Fosgerau, 2015). In the long run, e.g. the time horizon of a transport investment, there is however no stable reference point. Of course, in almost all cases, studies are required to produce a uniform VTT for appraisal that is independent of sign and size effects (cf. Daly et al., 2014), because in the long run there is no reference point (clearly implying that no WTP-WTA gap exists), among other important reasons.[5] Moreover, long-term stability of preferences is a fundamental assumption in welfare economics. This means that findings of substantial sign and size effects are problematic. An important question is thus again whether sign and size effects are stronger or weaker in simpler or more complex trade-off contexts, a point we address in our comparisons later on.

3. New evidence from the 2014/2015 GB national VTT study

The recently completed national VTT study for Great Britain provides a useful testbed for the points discussed above. The study presented people using car, bus, rail or other public transport

[5] A relevant question is what "long-term" means in different policy contexts. In the context of infrastructure investment, the time horizon of many decades ahead is so long term that one must assume that the reference point of the travellers has shifted and indeed the population of travellers has changed substantially. However, also for policies that can be implemented in shorter time such as congestion pricing evidence indicates that the reference point still shifts considerably. Evidence from the introduction of the Stockholm congestion charges show that even 6 months or a year is enough to substantially shift the reference point. The Stockholm system was abolished 7 months after the introduction, but the traffic levels did not return to their initial level suggesting a shift of reference points (Börjesson et al., 2010). Moreover, Peers and Börjesson (2018), comparing SP survey questions regarding trips the same day and SP survey questions regarding trips months ahead, show that the reference point may shift even over a couple of months.

(e.g. tram or London underground) with different types of SC games. Full details of the experimental setup and the modelling work conducted are given in Arup et al. (2015), with detailed discussions also in Hess et al. (2017).

For each of 11 purpose-mode combinations, multiple binary SC experiments involved different unlabelled trade-offs described by: time/money (SP1), time/money/reliability (SP2), and time/money/quality (SP3), where quality implied variation in either congestion or crowding. The SP2 and SP3 scenarios thus involved additional attributes compared to SP1, where the presentation varied across modes and between SP2 and SP3, as discussed in detail in Arup et al. (2015). To provide additional background, Figure 1 to Figure 4 present example choice scenarios from this study for car and rail.

Respondents received all three games, with 5 choices per game; this decision was made with a view to enabling us to make comparisons across games; the modelling used the data from all three games simultaneously. If data from different experiments comes from different people, it is likely that at least part of the variation stems from heterogeneity in preferences across people rather than across experimental settings. For the same reason, the GB study as far as possible made use of consistent modelling approaches across the different experiments.

As we mentioned in the introduction, a key argument behind the continued reliance on simple time-money trade-offs in many European studies has been that of continuity and comparability. This already hints at some underlying concerns that the values obtained with different experimental setups may vary (see also Widlert, 1994), and it is thus very surprising that only limited effort has gone into studying how the valuations obtained from the same respondents differ across contexts, something that is very different from studying differences in valuations across studies using different setups for different respondents. Our work described below addresses this issue.

In what follows, we focus on the key segment of commuters. Due to space considerations, we can only give a brief account of the actual work undertaken, and the focus here is on some key results. For in-depth discussion of the methods and model specifications, see Hess et al. (2017).

	Option A	Option B
One way fuel cost	£33.30	£35.00
One way travel time by car	4 hours 23 minutes	3 hours 30 minutes

	Option A	Option B

Figure 1: Example SP1 task for car travellers from 2014/2015 GB study

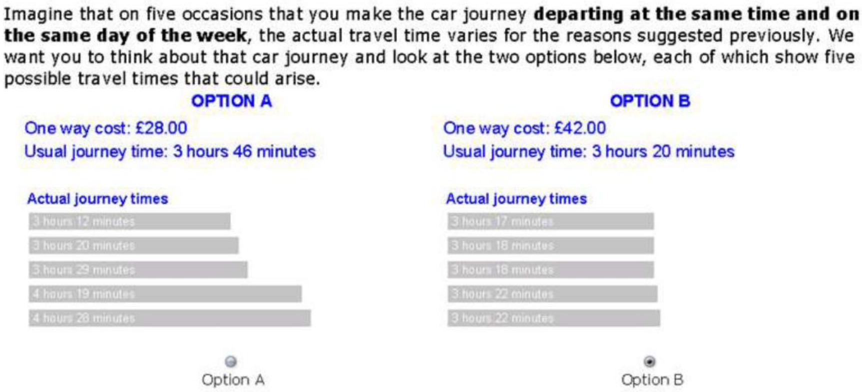

Figure 2: Example SP2 task for car travellers from 2014/2015 GB study

	Option A	Option B
One way travel cost	£37.60	£42.00
Traffic conditions	1 hour 45 minutes in heavy traffic 11 minutes in light traffic 2 hours 53 minutes in free flowing traffic	2 hours 11 minutes in heavy traffic 2 hours 36 minutes in light traffic 57 minutes in free flowing traffic
	Option A	Option B

Figure 3: Example SP3 task for car travellers from 2014/2015 GB study

	Option A	Option B
One way travel time	3 hours 54 minutes	3 hours 18 minutes
One way travel fare	£18.00	£24.00
Crowding level when you boarded	 Seated, 100% of seats occupied, eight people stood around each door	 Standing, 100% of seats occupied, one person stood around door
	Option A	Option B

Figure 4: Example SP3 task for rail travellers from 2014/2015 GB study

3.1. Apparent non-trading behaviour

We first look at the rates of apparent time and cost non-traders across games, i.e. respondents who always choose the fastest (time) or cheapest (cost) option across the five choice tasks in a given experiment. We use the term ‘apparent’ as it is not clear whether a respondent really does not trade (and hence does not engage with the choice task) or whether their preferences are such that the incentive to move away from either the cheapest or fastest are not strong enough in the scenarios presented to them. Similarly, a respondent might in fact use a heuristic to never accept a cost above a certain amount or a journey taking longer than a given amount of time, and the scenarios presented might thus impose the choice. Whatever the reason for such invariant choices, this type of behaviour can have substantial impact on model estimation in the case of random coefficients models, with extreme values being used to accommodate this behaviour, leading to long tails for the estimated distributions.

Alongside game-specific rates, we also present the p-value for a χ^2 test for equality between the rates in the three games. Finally, we present the non-trading rates across the full set of 15 choices, which are then of course lower because a respondent would have to be a non-trader in each of the three games.

From the results in Table 1, an interesting picture emerges. We first observe that, across all four segments and across both time and cost non-trading, the χ^2 test rejects the null hypothesis of equality in the rates of non-trading across the three games.

For car commuters, we see that the rates of both time and cost non-trading clearly decrease in SP2 compared to SP1, and then again in SP3 compared to SP2. This is in line with having two attributes in SP1 (time & cost), three attributes in SP2 (time, cost and a measure of variability in time), and four attributes in SP3 (time across three conditions and cost).

The picture for other modes is slightly more complex. For time non-trading, we see that, except for other public transport (PT), the rate of non-trading is highest in SP1, while, except for other PT, SP2 and SP3 are similar. For cost non-trading, the rates in SP2 (which has more attributes) are lower than in SP1 across the three modes, but the SP3 rates are essentially the same as the SP1 rates. While this might seem counter-intuitive at first, it should be noted that SP3 here is a crowding game which presents only time and cost for each alternative, albeit that for time, a level of crowding is given. What we thus see is that the level of complexity of SP3 is very similar to that of SP1, and that the rate of those respondents who do not move away from the cheaper of the two options is not impacted by the additional information on crowding.

Overall, we find that Table 1 shows some evidence of higher rates of non-trading in simpler games, and this is confirmed by statistical tests. A formal comparison of the trade-off ranges is not possible given the inclusion of additional attributes in SP2 and SP3. However, the ranges for the attributes were the same across the games, and this would thus mean that the possible utility differences between the alternatives in SP2 and SP3 cover a wider range than in SP1 given the presence of additional attributes, increasing the changes of trading. The actual ranges used were chosen such as to not lead to unreasonably small or large values for the attributes while also covering a wide range of possible boundary VTT measures in SP1.

Table 1: Non-trading in the 2014/2015 GB national VTT data

		Car commuters	Rail commuters	Other PT commuters	Bus commuters
Time non-traders (always fastest)	*SP1*	14.85%	10.19%	6.84%	8%
	SP2	12.82%	4.55%	7.84%	2.86%
	SP3[1]	10.87%	5.52%	0.53%	2.92%
	χ^2 *test p-value*	0.02	$2.7 \cdot 10^{-6}$	$1.1 \cdot 10^{-6}$	0.01
	over three experiments	1.26%	0.22%	0%	0%
Cost non-traders (always cheapest)	*SP1*	17.57%	19.16%	13.86%	22%
	SP2	11.06%	9.06%	5.17%	12%
	SP3[1]	8.34%	19.45%	12.44%	22.22%
	χ^2 *test p-value*	$7.6 \cdot 10^{-10}$	$1.2 \cdot 10^{-11}$	$1.5 \cdot 10^{-4}$	0.01
	over three experiments	2.04%	2.27%	1.06%	3.51%

1) crowding games only for rail, other PT and bus in SP3

3.2. Differences in base valuations across games

The GB study used a joint model for all three SP experiments, but allowing for differences across the three formats, both in the underlying VTT and in the size and sign effects. We return to the latter in Section 3.3. In terms of allowing for differences in the base valuation across the three games, a joint VTT is used and multipliers are estimated for the different games. A base is required, where SP1 was chosen as the base in the GB study, with multipliers estimated for the valuations in the other games. These multipliers are shown in Table 2, each time accompanied by a test of significance against the base value of 1, which would imply no difference from the SP1 valuation. The following observations can be made:

- Except for other PT, the valuation for mean/expected travel time in SP2 is significantly higher than the SP1 valuation.
- For the crowding experiments (SP3), we see that for rail and other PT, the SP1 value is between the lowest and highest crowding values, and is significantly different from both. However, for bus, the SP1 valuation is not significantly different from the SP3 valuation at the lowest crowding level.
- For the congestion experiments (SP3), we see that for car, the SP1 value is not significantly different from the light congestion value, and is significantly higher/lower than the free-flow/heavy congestion values. However, for bus, there is no difference across congestion levels.

While the findings for SP3 are reassuring in that the SP1 values tend to fall between the best and worst conditions, the findings for SP2 are of greater concern. They indicate that, in the presence of the additional travel time reliability attribute, the valuation for travel time is higher, and statistically significantly so, than that for SP1 for three out of the four modes. The reasons for this are of course unclear. However, a reasonable hypothesis is that in the absence of journey quality information, there is greater focus on cost in SP1, leading to lower valuations. In SP3, the source of the higher travel times is made clearer to the respondent and this seems to lead to higher valuations.

Table 2: Multipliers for non-SP1 valuations for commuters in GB study

		est.	robust t ratio (vs 1)
car	SP2 time multiplier	1.60	4.05
	SP3 free-flow multiplier	0.70	-2.26
	SP3 light congestion multiplier	0.98	-0.14
	SP3 heavy congestion multiplier	1.86	2.98
rail	SP2 time multiplier	1.24	2.58
	SP3 lowest crowding multiplier	0.70	-3.07
	SP3 highest crowding multiplier	1.80	4.49
bus	SP2 time multiplier	1.47	2.15
	SP3 lowest crowding multiplier	0.94	-0.34
	SP3 highest crowding multiplier	2.36	3.75
	SP3 free flow multiplier	1.09	0.36
	SP3 slowed down multiplier	1.54	1.50
	SP3 dwell time multiplier	0.75	-0.39
other PT	SP2 time multiplier	1.12	1.27
	SP3 lowest crowding multiplier	0.78	-2.46
	SP3 highest crowding multiplier	1.40	2.74

3.3. Behavioural response to sizes and signs of differences from reference values

A key emphasis in the modelling work conducted for the GB study was the modelling of sign and size effects, i.e. allowing for differences in how respondents react to increases and decreases from a reference value (for time or for cost) and also in how the marginal sensitivity depends on the size of the shift presented. The specific approach used for this relies on the framework developed by de Borger and Fosgerau (2008), hereafter referred to as dBF, which provides all the flexibility we require.

In the dBF framework[6], we define a function that gives the value $v(\Delta x)$ of a change Δx relative to the reference value x_0 of a given attribute, as:

$$v(\Delta x) = S(\Delta x) \cdot \exp(\eta\, S(\Delta x)) \cdot |\Delta x|^{1-\beta-\gamma S(\Delta x)} \qquad (5)$$

with $\Delta x = x - x_0$

where:

- $S(\Delta x)$ is the sign function, defined for $\Delta x \neq 0$ by $S(\Delta x) = \Delta x \,/\, (|\Delta x|)$; for convenience we set $S(0)$ =0.
- η captures sign effects by giving the difference of gain value and loss value from an 'underlying' value. It is explicitly assumed by dBF that gains and losses exactly bracket

[6] This approach is highly advanced and we can only give a brief overview here. Additional detail and interpretation can be found in Hess et al. (2017).

this underlying value. It is expected that $\eta > 0$, so that the value of losses (increases in Δx) is greater than the value of gains.
- β captures size effect by allowing the impact of gains and losses to be non-linear, where, with $\beta > 0$, the marginal value of changes decreases as the change increases, i.e. the value is 'damped'. Generally we anticipate that β should be larger for cost than for time, so that VTT increases as the changes increase, while small time savings have lower monetary value.
- γ allows the non-linearity (size effect) of value to be different for gains and losses.

The 2014/2015 study allowed for sign and size differences in all three games, with the three parameters (η, β and γ) estimated for both time and cost. After extensive exploratory work (cf. Hess et al., 2017), the models were all specified with a multiplicative error structure, where for SP1, a log-WTP space specification was found to be preferable to a simple $U = V \cdot \varepsilon$ specification[7]. All models also allowed for random heterogeneity in the VTT across respondents, using log-uniform distributions.

The presence of these additional terms complicates the VTT calculation, which is now no longer the simple ratio of time and cost coefficients. The value functions v (in equation 5) are defined to have arguments and values denominated in cost units. Thus the cost value of a cost change Δc is given by $v(\Delta c)$, while the cost value of a time change t is given by $v(\theta\Delta t)$, where θ is the 'underlying' value of time.

A simple way to see the derivation of VTT (and other WTP measures) is to think of the values of Δc and Δt that would maintain indifference with the base situation in which $\Delta t = \Delta c = 0$ and the total value is of course zero. Thus when we have a specific value $\Delta t'$, and we have estimated the parameters of the value functions v, we can find the indifference value $\Delta c'$ such that $v(\Delta c') + v(\theta\Delta t') = 0$. The average willingness to pay per unit of time is then $\Delta c'/\Delta t'$.

As discussed in detail in Hess et al. (2017), the actual VTT is then given by

$$VTT = {}^{|\Delta c|}\!\big/_{|\Delta t|} = \theta^{\kappa}|\Delta t|^{\kappa-1}, \tag{6}$$

where $\kappa = \frac{1-\beta_t}{1-\beta_c}$, and Δc is chosen such that $v(\Delta c) = -v(\theta\Delta t)$.

While η and γ present interesting behavioural insights, they drop out of the calculation of the VTT. The same is not true for the size effects represented by β. Only when $\beta_c = \beta_t$ are we in the situation where the VTT is independent of Δt, as the time and cost damping cancel out, i.e. we get that $\kappa = 1$. However, in general the β values will not be equal and VTT is not equal to θ but depends on the size of the time shift.

The most important impact of the dBF parameters in the VTT context is thus the fact that the VTT is no longer independent of the size of time shift considered if κ (given by $\frac{1-\beta_t}{1-\beta_c}$) is different from 1 in Equation (6).

Table 2 presents the values for β and resulting values for κ across all mode-purpose combinations, excluding bus for which no size effects were observed, which can likely be linked to the fact that bus journeys are shorter. For other modes, we see that all 3 values for κ

[7] We have no reason to believe that this difference explains the differences we report below, and the results obtained with $U = V \cdot \varepsilon$ on SP1 were broadly comparable, but with lower fit. Similarly, the differences we report across games were also confirmed by a simpler analysis using models with an additive error structure and without random heterogeneity.

are different from 1 for SP1, while this reduces to 2 for SP2 and 1 for SP3. Table 2 also reports significance levels for the differences in κ between the three different SP games for each segment. These calculations incorporate the correlations between the individual estimates. For comparisons where for one game, κ collapses to 1, the significance level for the difference (against 0) is of course the same as the significance level of a test against 1 for the κ which did not collapse to 1. We see that κ_{SP1} is significantly larger than κ_{SP2} and κ_{SP3} at the 95% level or above for car and other PT, and is not significantly lower than either κ_{SP2} or κ_{SP3} for rail.

These results provide at least some suggestion that the incidence of size effects is greater for SP1 (i.e. simple time-money trade-offs) than for other games. It is also insightful to note that Hess et al. (2017) highlight differences across games in the incidence of significant estimates for η, i.e. showing the presence of asymmetries in the sensitivities to gains and losses. No effects were observed for bus. Looking at car, rail and other PT (tram, underground etc.), we note that, out of the 18 possible η parameters (time and cost for three purposes and three modes), they found significant gain-loss asymmetry in 6 cases for SP1, 5 cases for SP2 and only 1 case for SP3.

Overall, these results provide some indication of differences across games in the role of reference formation, and may point to an influence of the experimental setup. Independently of whether these are real or experimental phenomena, the presence of these effects, especially size effects, poses significant issues in appraisal, and this could be seen as a disadvantage of simple time-money trade-offs such as SP1. We return to this point in Section 3.4 when we look at the implied VTT measures. In addition, returning to the earlier points about consistency with micro-economic frameworks, reduced reference dependence in the more complex scenarios may be seen as an advantage.

Table 3: β and κ values for commuter models for 2014/2015 GB study

	Car		Rail		Other PT	
	est	rob t (0)	est	rob t (0)	est	rob t (0)
$\beta_{t,SP1}$	-0.40	-3.64	-0.21	-3.31	0	
$\beta_{t,SP2}$	-0.16	-2.84	0		0	
$\beta_{t,SP3}$	0		-0.24	-2.51	0	
$\beta_{c,SP1}$	0		0		0.14	2.55
$\beta_{c,SP2}$	0		0.25	3.78	0	
$\beta_{c,SP3}$	0		0		0	
	est	rob t (1)	est	rob t (1)	est	rob t (1)
κ_{SP1}	1.40	3.64	1.21	3.31	1.16	2.19
κ_{SP2}	1.16	2.84	1.33	2.84	1	-
κ_{SP3}	1	-	1.24	2.51	1	-
	est	rob t (0)	est	rob t (0)	est	rob t (0)
κ_{SP1}-κ_{SP2}	0.24	2.47	-0.12	-0.77	0.16	2.19
κ_{SP1}-κ_{SP3}	0.40	3.64	-0.03	-0.27	0.16	2.19
κ_{SP2}-κ_{SP3}	0.16	2.84	0.09	0.54	-	-

3.4. Values for appraisal

We next compare the core VTT measures across the three games for each mode after expansion to national values (cf. Arup et al., 2015; Hess et al., 2017), again focussing on commuters only. The results for this comparison are shown in separate panels for the four modes in Figure 5. Noting that values for κ that are different from 1 arise for car, rail and other PT, the VTT measures are shown for different values of Δt (cf. Equation (6)). This complication does not arise for bus.

Looking first at car, we note the absence of size effects for SP3, and hence the flat profiles for the VTT in the three conditions. We also include in Figure 5 a value weighted according to average conditions in the UK. With κ_{SP1}=1.4 and κ_{SP2}=1.16, we note increasing VTT for SP1 and SP2 (but less so for the latter) with increasing Δt, where for low values of Δt, the SP1 value is below that for free flow conditions, while with high Δt, it is close to that in heavy congestion.

As discussed at length in Hess et al. (2017), the presence of these size effects and their impact on the VTT leads to the difficult situation in which the analyst needs to make a choice for a value of Δt. This has impacts on the overall VTT measures and on the relative values across modes and purposes and is thus not a decision to be taken lightly. The first point to note is that this measure, i.e. Δt, does not refer to the size of a time reduction achieved by a specific transport policy; the reference points against which travellers will compare their travel conditions in the future are of course unknown at present. Rather, it is quite likely the case that the reference dependence effects uncovered in the models are in large part short-term stated preference effects. The choice of a value however remains arbitrary and it is of course not satisfactory that such a decision by an analyst will lead to substantially different VTT results. While we are not denying the possible existence of these effects in real life behaviour, the differences across experiments suggests a certain influence by the design itself and enhances our view that choice contexts that do not lead to significant size effects greatly simplify the use of results.

With the value of $\Delta t = 10$ chosen in the UK appraisal framework, the VTT from SP1 is above the SP3 values weighted to average conditions, while that from SP2 is close to that for high congestion in SP3. This highlights the earlier points about the difficulty faced in interpreting values from simple time-money trade-offs, but also extends it to time-money-reliability trade-offs where the type of time being valued is not explicitly described.

Looking next at rail, we have relatively similar values for κ across the three games, and thus similar non-linearities. We present the simple VTT from SP1 and SP2, along with the VTT in the lowest and highest crowding conditions (out of 10 conditions) from SP3, and a value weighted to average rail crowding in the UK. We note that the value from SP1 is again below that for SP2 (except for very low Δt), while it is also below that for average crowding conditions.

For bus commuters, no size effects were observed and we thus present a single value for each component. We again observe a higher value for SP2 than for SP1, and note that the value for SP1 is below the value for free flow time in SP3, and below the values weighted to either traffic or crowding conditions (noting that some bus users received a crowding game as SP3 while others received a congestion game as SP3).

Finally, for other PT, size effects are only observed in SP1, and this shows that the value for SP1 rises with Δt, going from below that in the lowest crowding conditions, to exceeding average conditions with a value of Δt of below 5, while, at the value of $\Delta t = 10$ used in appraisal, the value from SP1 equals that from SP2 and exceeds that in average conditions.

Across the four modes (and similar results arise with other purposes), we can reach a number of conclusions:

- the VTT from SP2 is generally higher than that from SP1, suggesting that, in the presence of reliability information, respondents place a lower sensitivity on cost than in the absence of reliability information;
- there is no evidence to suggest that the VTT from SP1 corresponds in any way to the values from SP3 weighted to average conditions; and
- the presence of size effects significantly complicates the picture, leading to further issues in choosing values for appraisal, and differences across modes in the relationship between values across different games.

Of course, some difference between SP1 and SP2 could be expected given that the valuations from SP1 relate to a deterministic change in the travel time while the values from SP2 relate to a change in the mean travel time across journeys affected by variability. In reality, time varies across trips, so looking at the mean time makes more sense than the time for a single trip as the actual outcome is not known before the journey, only the distributions of outcomes may be known to some extent. The same cannot be said when comparing SP1 and SP3 which both relate to certain times but still give different results. Overall, this discussion adds empirical evidence to our earlier concerns about what the VTT from SP1, i.e. the simple time-money trade-off, "means".

A simple interpretation could be that the type of time being valued in a simple time-money trade-off is representative of the conditions experienced by the respondent on the reference journey; i.e. if this journey was in heavy congestion, then the direct utility of the time component of the VTT off would relate to time spent in heavy congestion. The 2014/2015 study attempted to explain differences across different respondents in their valuations by linking them to conditions on the reference journey. For example, for rail travellers, the study tested whether respondents who had reported high levels of crowding for their reference journey had higher VTT in SP1 than those who had reported low levels of crowding. Similar tests were made for car travellers, trying to relate the SP1 valuations to the level of congestion experienced on their reference trip. The car results are more reliable as the congestion information was captured using the same metrics as those presented in the SC scenarios, which was not the case for crowding.

Overall, these attempts were not successful[8], i.e. we did not find that the VTT was higher for respondents who had experienced "worse" conditions on their reference journey. A number of reasons are possible for this. Foremost is that the respondents fail to place the hypothetical scenarios within the context of their real-world trip and the conditions experienced on that trip. Additionally, it is of course possible that self-selection is at play and that those respondents who travel on more crowded trains in real life are less sensitive to crowding than other respondents, with a similar argument for those people who drive during congested times. Either way, the finding for SP1 means that we do not learn anything about how valuations differ across different travel conditions from simple time-money trade-offs. Such information can only be obtained by understanding how the same respondent, considering the same reference trip (so that the resource value of time stays constant), reacts to different journey conditions. Taking a value estimated from a simple time-money trade-off may of course tell us the average VTT across travellers in current journey conditions (providing the sample is representative) but we cannot learn anything about how that behaviour might change if journey conditions are changed.

[8] Including in further tests reported in http://www.stephanehess.me.uk/papers/Hess_impact_of_congestion_on_SP_values.pdf

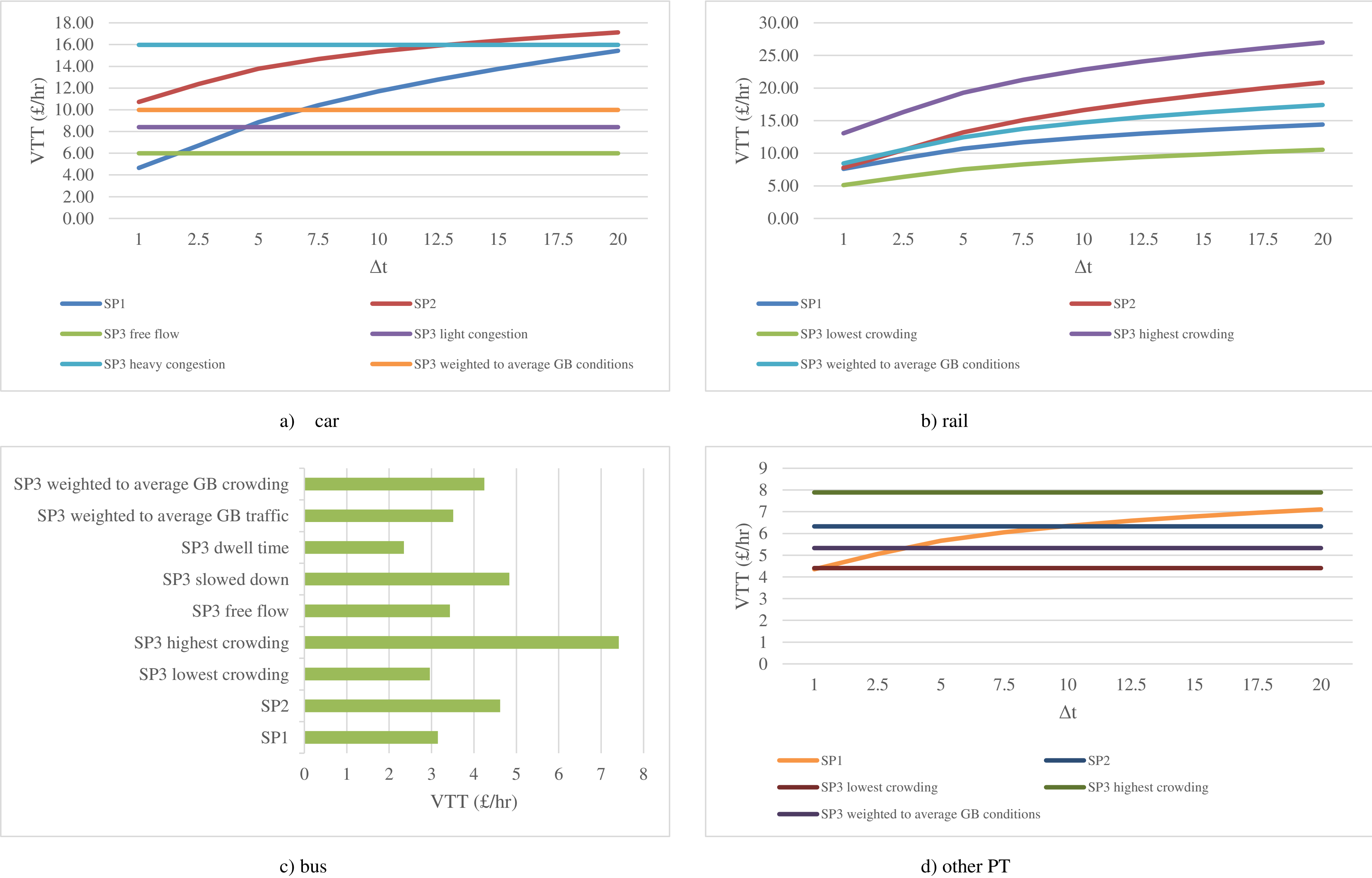

Figure 5: comparison of VTT across games for commuters in 2014/2015 GB study

4. Evidence from other studies

To provide some additional evidence, we now also briefly summarise results from two other national VTT studies, commissioned by the central governments of the countries concerned, and using similar survey approaches to those described above. In particular, they both included a simple time-money trade-off, and the recommended values are always based on that trade-off. The studies all predate the recent British study and our involvement in them was also smaller. These two factors reduce the amount of detail available to us and provided here but brief summaries are included to highlight again that differences across the formats can arise. We do not provide the same level of detail about survey methods and modelling approaches and instead refer the reader to the official reports which are cited below. We do not include the recent Netherlands study (Significance et al., 2013) as the data from the simple time-money trade-offs in that study was merged with the data from more complex games and no differences in sensitivities were allowed for in the final joint models. Similarly, we do not include the 1994 UK study (AHCG, 1996; ITS and Bates, 2003) as the two experiments included in the analysis were both time-money trade-offs, albeit that, in one game, the cost included tolls.

4.1. Danish study (2004 data)

In this 'DATIV' study, data was collected in four stated preference exercises which related to a specific journey made by the respondent (Fosgerau et al., 2007). We again compare the results for a simple time-money trade-off to those from a game involving both in-vehicle and out-of-vehicle components. For car drivers, the components were walking time to/from a car park, searching time for a parking space, free-flow driving and congested driving time, while, for public transport users, the experiment covered in-vehicle time, headway, access/egress time, interchange times as well as cost.

The study introduced the modelling concept of working in log-WTP space for the time-money trade-offs and a multiplicative specification for the more complex game, as also done in the recent GB study. The dBF approach was used to allow for sign and size effects in the time-money trade-offs. However, reference effects were limited to the influence of sign, i.e. no asymmetric size effects. In both experiments, the central VTT value was assumed to be randomly distributed in the population. As in the GB study, the presence of size effects meant that a decision on Δt was required, and a value of 10 minutes was used, which is of course again arbitrary.

Table 4: DATIV values of travel time mean values, all purposes (not working) 2003 DKK per hour, distributions truncated at DKK 1000/hr

	Car driver free-flow	Car passenger free-flow	Bus	Metro	S-Train	Train
VTT from time-money trade-offs	78	52	30	62	35	54
VTT from more complex trade-offs	<50 km: 98 ≥50 km: 78	<50 km: 78 ≥50 km: 98	37	60	54	<50 km: 52 ≥50 km: 183

Note: error margins are not given in the published report.

Although the pattern of VTT in the more complex trade-offs for car driver is unusual, as longer journeys usually have a higher VTT, and the Metro values are contrary to the general trend, it is clear that the more complex trade-offs generally yield a somewhat higher VTT than the simple time-money trade-offs. Since error margins are not quoted in the reports, we are unable to say whether this effect is statistically significant.

4.2 The Norwegian Study (2009 data)

In the Norwegian 2009 VTT study (Ramjerdi et al. 2010), each respondent received 3 choice experiments. First, all respondents received 9 choice tasks from a simple time-money trade-off, followed by two sets of 6 choices involving an additional attribute related to either quality or reliability. The values are summarised in Table 5 and, although the models used differed between the experiments, there is again evidence of different values of time in the simple time-money trade-offs.

Table 5: Values of time (mean values, 2009 NOK per hour)

		Car	Rail	Bus	Air	PT
VTT from time-money trade-offs	Long-distance	148	97	75	192	
	Short-distance	84				54
VTT from congestion game	Long-distance	157				
	Short-distance	47				
VTT from seat availability game	Long-distance					
	Short-distance					36
VTT from mean variance game	Long-distance	255	268	150	522	
	Short-distance	128				67
VTT from scheduling model game[9]	Long-distance	95/ 258	100/ 249	223 / 577	235/ 707	
	Short-distance	35/ 172				31/45

Note: error margins are not given in the study report.

5. Conclusions

This paper has sought to provide a critical assessment of the continued reliance on simple time-money trade-offs in many national VTT studies in Europe, when such approaches are routinely criticised elsewhere as being too abstract and potentially prone to bias.

The aim behind this paper was in no way to invalidate the large amount of high quality methodological work in choice modelling that has made use of such data in recent years, but rather to question its reliability for producing values for appraisal. In this context, we specifically ask the question whether the fact that such approaches have been used many times in the past is a valid justification for continuing to rely on them.

Our initial discussions focussed on complexity. Complex scenarios can provide respondents with a setting that is more in line with their real-life experiences, where a journey is described by many attributes. An argument against complex scenarios has been that respondents cannot process them in surveys, while an argument against simple scenarios has been that when facing scenarios with only a few attributes, the response to them may be different from that in real life settings. Overall, the academic evidence suggests that a) respondents can deal with moderately complex choice scenarios, b) the valuations from more complex scenarios are reliable and c) that analysts can accommodate differences in how respondents relate to these more complex environments. While it is true that the identifying (in the data) and mathematically representing (in the models) complex behavioural phenomena can apparently be easier with simpler data setups, this alone should not be a reason to limit the complexity of the scenarios presented.

Our personal suggestion would then be for applied work used to guide national policy to move towards more complex games, or at the very least to not see simpler approaches as somehow preferable. Of course, this opens up a new question, namely how many additional attributes

[9] The first is the value of travel time relative to time spent at home. The second VTT is the value of travel time relative to time spent at destination after preferred arrival time.

should be included, and what they should be. This can be informed to some extent by focus groups and by gaining a better understanding of which attributes matter for which travellers in making their decisions. Where budgets allow, the inclusion of simpler games for comparison can of course be considered, but this then reduces sample size for the individual games (and increases standard errors) and opens up new issues in terms of which values to use when the results are different.

While concerns about the "realism" of simple time-money trade-offs remain especially for car travel (a quicker route would generally be cheaper), there is of course scope for mitigating this in the survey description. The same cannot be said in relation to concerns about interpretation of such trade-offs by respondents in the absence of a clear definition of the type of time being valued. Empirical support for this point was provided by the results from the recent GB study. In conjunction with results from a number of other recent studies, we can conclude that the VTT from simple time-money trade-offs tend to be lower (sometimes substantially lower) than those from other settings, maybe suggesting that with just two attributes, respondents focus more on cost. This would also be consistent with the findings in Hensher (2004). But in some of the studies, the value from the simple time-money trade-offs is higher for some of the modes.

Of course, we do not know what value is *"right"*, and one could of course argue that rather than the value from simple time-money trade-offs being too low, that from more complex surveys is too high (or vice versa). We do not have an a priori truth and our decision on which value should be used needs to be guided by which one we trust more and which seems more reliable. In this context, a number of other considerations should be taken into account.

Firstly, it is clear that with simple time-money trade-offs, we risk producing a VTT measure that relates to travel conditions which we do not know, and which are subject to unknown interpretations by the respondents. For example, if a scenario is proposed where the driving time is longer than at present, it might be reasonable for a respondent to think this was the result of increased congestion. It also seems likely that the variation in that value across respondents does not reflect the variation in experienced real-world conditions across respondents, as highlighted in the findings in the GB study. Even if we could link the valuations back to experienced real-world conditions, issues with self-selection remain, and we learn nothing about how the same person would react in different conditions. This creates issues in terms of recognising differences in the direct utility of travel time (cf. Section 2.1)[10]. Of course, an inability to link the VTT to travel conditions makes it impossible to forecast changes in the VTT as a function of travel conditions.

[10] In other evidence for this, it is worth noting that the 2008 Swedish VTT study used a method resembling the one used by Ortúzar (2007) to value the perceived security when accessing PT. Respondents were presented with two binary stated choice experiments, differing in the dimensions of in-vehicle travel time, walk time and wait time. In the first experiment, the respondents were asked to have a recently made PT trip in mind. In the second experiment, one of four different physical walking environments was attached to each choice task, presented by coloured drawings. The walking time weights were found to be consistently higher in the second experiment than in the first, even for the nicest and most secure walking environment. This could be due to an over-focus effect on the walking time in the second experiment. However, it is also a support of the conclusion from the UK study, indicating that we cannot learn about the how the value of time depends on travel conditions unless we explicitly vary them in the experiments.

Secondly, the results from the most recent UK study also suggest that there is potentially increased reference dependence and non-linearity in data from simple time-money trade-offs. While these effects are behaviourally interesting, they lead to great complications in using the outputs in appraisal. Even if they are *real-world* effects arising from the framing of the experiment around a short-term setting, and not artefacts from the hypothetical setting, they are a severe problem since welfare analysis relies on long-term stable preferences. Of course, in this context, and notwithstanding data quality issues, there are benefits to the renewed interest in Revealed Preference data. One potential advantage of real world data is that it may be less susceptible to willingness-to-accept(WTA) / willingness-to-pay (WTP) disparities than is the case for Stated Preference data (cf. Sugden, 2009).

Finally, an important point we touched on in the introduction is that appraisal is of course interested not just in the monetary valuation of travel time but also numerous other journey components. If the VTT is obtained from simple time-money trade-offs and then combined with other measures (e.g. value of crowding) from more complex settings, then there is a significant risk that the different values are not compatible with each other. Significant issues with fungibility may arise (cf. Hess et al., 2012) and there might be benefits by instead combining all components into one survey.

As an example, assume that the VTT from a simple time-money trade-off is £10/hr, while the VTT from a game also incorporating interchanges is £15/hr and that this game shows that one interchange is valued in the same way as 15 additional minutes in time. If the appraisal framework is then based on the simple time-money trade-off and uses the VTT of £10/hr, should interchanges be valued at £2.50 or £3.75? The appraisal team for the recent British study (Arup et al., 2015) essentially used the former approach – multipliers for different crowding levels for example were obtained by comparing the VTT for different levels of crowding in SP2 but then applied to the VTT from SP1, disregarding the differences that clearly emerge between SP1 and SP3 as shown in panel b) of Figure 5. All the other concerns expressed in this paper aside, this may well be the strongest argument for including all components of interest in a single experiment. Including all attributes together in a single experiment also has the statistical benefit that the parameters for all attributes are estimated on the basis of the full data. Of course, the question remains what the set of "all components" is and an analyst can likely never be sure to have included all that are relevant while still balancing this against the need to keep respondent burden at an acceptable level.

ACKNOWLEDGEMENTS

The first author acknowledges the financial support by the European Research Council through the consolidator grant 615596-DECISIONS. The last author acknowledges the financial support from VINNOVA grant 2014-03587.

CONTRIBUTIONS

Stephane Hess: literature review, empirical work, manuscript preparation and editing

Andrew Daly: literature review, manuscript preparation and editing

Maria Börjesson: literature review, manuscript preparation and editing

On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Accent and Hague Consulting Group (1996): The value of travel time on UK roads, London, Report for Department of Environment, Transport and the Regions.

Algers, S., J. Lindqvist Dillén, and S. Widlert (1995): 'The national Swedish value of time study', Presented to 'Traffic Days', Aalborg University.

Arup, ITS Leeds and Accent (2015) 'Provision of market research for value of time savings and reliability'. Phase 2 report to the Department for Transport. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/470231/vtts-phase-2-report-issue-august-2015.pdf

Axhausen, K.W., Hess, S., König, A., Abay, G., Bates, J.J. & Bierlaire, M. (2008), Income and distance elasticities of values of travel time savings: New Swiss results, Transport Policy, 15(3), pp. 173-185.

Axhausen, K. W., Schmid, B., Weis, C., 2015. Predicting response rates updated. Arbeitsberichte Verkehrs-und Raumplanung 1063.

Bates, J., Whelan, G., 2001. Size and sign of time savings. Working Paper. Institute of Transport Studies, University of Leeds , Leeds, UK

Bell, D.E., Keeney, R.L., Raiffa, H., 1977. Conflicting Objectives in Decisions. John Wiley & Sons.

Borger, B. de and Fosgerau, M. (2008) The trade-off between money and travel time: A test of the theory of reference-dependent preferences, J. of Urban Economics, Vol. 64, pp. 101-115.

Börjesson, M. & Algers, S. (2011), Internet actually reduced response error: Properties of internet and telephone data collection methods in a stated choice value of time study context, Journal of Choice Modelling, 4(2), pp. 1-19.

Börjesson, M., Fosgerau, M. and Algers, S. (2012), Catching the tail: Empirical identification of the distribution of the value of travel time. Transportation Research Part A: Policy and Practice, vol. 46, no. 2, pp. 378-391, 2012.

Börjesson, M. (2012) "Valuing perceived insecurity associated with use of and access to public transport," Transport Policy, vol. 22, pp. 1-10, 2012.

Börjesson, M., Eliasson, J., 2014. Experiences from the Swedish Value of Time study. Transportation Research Part A: Policy and Practice 59, 144–158. https://doi.org/10.1016/j.tra.2013.10.022

Börjesson, M. and Fosgerau, M. (2015) "Response time patterns in a stated choice experiment," Journal of Choice Modelling, vol. 14, pp. 48-58, 2015.

Brownstone, David, and Kenneth A. Small (2005) "Valuing time and reliability: assessing the evidence from road pricing demonstrations." Transportation Research Part A: Policy and Practice 39.4 (2005): 279-293.

Bwambale, A., Choudhury, C.F. & Hess, S. (2019), Modelling long-distance route choice using mobile phone call detail record data: A case study of Senegal, Transportmetrica A, 15(2), pp. 1543-1568.

Burge, P., Rohr, C., Vuk, G. and Bates, J. (2004) Review of international experience in VOT study design, presented to European Transport Conference.

Caussade, S., Ortúzar, J. de D., Rizzi, L.I. and Hensher, D.A. (2005) "Assessing the influence of design dimensions on stated choice experiment estimates". Transportation Research B, 39, pp. 621-640.

Cantillo, V., Heydecker, B., Ortúzar, J. de D., 2006. A discrete choice model incorporating thresholds for perception in attribute values. Transportation Research Part B: Methodological 40, 807–825. https://doi.org/10.1016/j.trb.2005.11.002

Chintakayala, P.K, Hess, S., Rose, J.M. and Wardman, M.R. (2010), Effects of Stated Choice Design Dimensions on Model Estimates, in Hess, S. & Daly, A., Choice Modelling: the state-of-the-art and the state-of-practice; Proceedings from the Inaugural International Choice Modelling Conference, Emerald, Bingley, pp. 195-216.

Daly, A. J. and Zachary, S. (1975) *Commuters' Values of Time*, Local Government OR Unit Report T55, January; available at http://www.alogit.com/papers/CommutersValues.pdf, accessed 29 July 2016.

Daly, A., Tsang, F. and Rohr, C. (2014) The value of small time savings for non-business travel, in Transport Economics and Policy, Vol. 48, 2, May.

DeSerpa, A. C. (1971). A Theory of the Economics of Time. The Economic Journal, 81(324), 828–846. https://doi.org/10.2307/2230320

Ehreke, I., Hess, S., Weis, C. & Axhausen, K.W. (2015), Reliability in the German value of time study, Transportation Research Record, 2495, pp. 14-22.

Fischhoff, B., 2013. Cognitive processes in stated preference methods [WWW Document]. Risk Analysis and Human Behavior. https://doi.org/10.4324/9780203140710-8

Fosgerau, M. (2006). Investigating the distribution of the value of travel time savings. Transportation Research B, 40(8), 688–707.

Fosgerau, M. (2007). Using nonparametrics to specify a model to measure the value of travel time. Transportation Research A, 41(9), 842–856.

Fosgerau, M., & Bierlaire, M. (2007). A practical test for the choice of mixing distribution in discrete choice models. Transportation Research Part B, 41(7), 784–794.

Fosgerau, M., Hjorth, K. and Lyk-Jensen, S. (2007) The Danish Value of Time Study, DTU Report 5, Final Report; Note 5, Results for Experiment 1; Note 6, Results for Experiment 2; all available at http://www.dtu.dk/english/Service/Phonebook/Person?fr=101&id=38996&mr=100&qt=dtupublicationquery&tab=2, accessed 30 July 2016.

Habib, Khandker Nurul, Ana Sasic, Claude Weis and Kay W. Axhausen. (2013). Investigating the nonlinear relationship between transportation system performance and daily activity–travel scheduling behaviour. Transportation research. Part A, General 49: 342-357.

Hague Consulting Group (1990): The Netherlands Value of Time Study: Final Report, The Hague: Dienst Verkeerskunde, Rijkswaterstaat.

Hague Consulting Group (1998): Value of Dutch Travel Time Savings in 1997: Final Report, The Hague, Adviesdienst Verkeer en Vervoer, Rijkswaterstaat, HCG Report 6098.

Hammond, K.R., Adelman, L., (1976). Science, Values, and Human Judgment. Science 194, 389–396.

Hensher, D.A. (2004) Identifying the Influence of Stated Choice Design Dimensionality on Willingness to Pay for Travel Time Savings, Journal of Transport Economics and Policy, 38 (3), 425-446.

Hensher, D. A., (2006). How do respondents handle stated choice experiments? – attribute processing strategies under varying information load. Journal of Applied Econometrics 21, 861-878.

Hensher, D.A. and Rose, J. (2007) Development of Commuter and non-Commuter Mode Choice Models for the Assessment of New Public Transport Infrastructure Projects: A Case Study, Transportation Research A, 41 (5), 428-433

Hensher, D. A., 2008. Joint estimation of process and outcome in choice experiments and implications for willingness to pay. Journal of Transport Economics and Policy, 42 (2), 297-322.

Hess, S., Greene, E.R., Falzarano, C.S. & Muriello, M. (2011), Pay to drive in my bus lane: a stated choice analysis for the proposed Lincoln Tunnel HOT lane into Manhattan, Transport Policy 18(5), pp. 649-656.

Hess, S., Rose, J.M. & Hensher, D.A. (2008), Asymmetric Preference Formation in Willingness to Pay Estimates in Discrete Choice Models, Transportation Research Part E, 44(5), pp. 847-863.

Hess, S., Rose, J.M. & Polak, J.W. (2010), Non-trading, lexicographic and inconsistent behaviour in stated choice data, Transportation Research Part D, 15(7), pp. 405-417.

Hess, S. (2014), Impact of unimportant attributes in stated choice surveys, European Journal of Transport and Infrastructure Research, 14(4), pp.349-361.

Hess, S., Orr, S. & Sheldon, R. (2012), Consistency and Fungibility of Monetary Valuations in Transport, Transportation Research Part A 46(10), pp.1507-1516.

Hess, S., Daly, A., Dekker, T., Ojeda Cabral, M. & Batley, R.P. (2017), A framework for capturing heterogeneity, heteroskedasticity, non-linearity, reference dependence and design artefacts in value of time research, Transportation Research Part B, 96, pp. 126–149.

Hess, S., Daly, A.J. & Batley, R. (2018), Revisiting consistency with random utility maximisation: theory and implications for practical work, Theory and Decision, 84(2), pp. 181-204.

Hjorth, K., Fosgerau, M., (2011). Loss Aversion and Individual Characteristics. Environmental and Resource Economics 49, 573–596. https://doi.org/10.1007/s10640-010-9455-5 Beesley, M. E. (1965): 'The value of time spent in travelling: some new evidence', Economica, 32, 174–85.

Hossan, Md Sakoat, Hamidreza Asgari, and Xia Jin. "Investigating preference heterogeneity in Value of Time (VOT) and Value of Reliability (VOR) estimation for managed lanes." Transportation Research Part A: Policy and Practice 94 (2016): 638-649.

Hultkrantz, L. and Mortazavi, R. (2001) Anomalies in the Value of Travel-time Changes, Journal of Transport Economics and Policy, 35, 285–300.

Jacoby, J., Speller, D.E., Kohn, C.A., 1974. Brand Choice Behavior as a Function of Information Load. Journal of Marketing Research 11, 63–69. https://doi.org/10.2307/3150994

Jara-Díaz, S. (2007) Transport economic theory. Emerald Group Publishing Limited.

Jara-Díaz, S. R., & Guevara, C. A. (2003). Behind the Subjective Value of Travel Time Savings. Journal of Transport Economics and Policy, 37, 29–46.

Keller, K.L., Staelin, R., 1987. Effects of Quality and Quantity of Information on Decision Effectiveness. J Consum Res 14, 200–213. https://doi.org/10.1086/209106

Lam, Terence C., and Kenneth A. Small. "The value of time and reliability: measurement from a value pricing experiment." Transportation Research Part E: Logistics and Transportation Review 37.2-3 (2001): 231-251.

Lichtenstein, S., Slovic, P., Fischhoff, B., Layman, M., Combs, B., 1978. Judged frequency of lethal events. Journal of Experimental Psychology: Human Learning and Memory 4, 551–578. https://doi.org/10.1037/0278-7393.4.6.551

Malhotra, N.K., 1982. Information Load and Consumer Decision Making. J Consum Res 8, 419–430. https://doi.org/10.1086/208882

MVA, ITS Leeds, and TSU Oxford (1987): 'The value of travel time savings', Policy Journals, Newbury, Berkshire.

Ortúzar, J.D. (2007): Valuation of intangibles with stated preference methods: A view from the fringe of the empire. Oslo Workshop on Valuation Methods in Transport Planning, March 19-20, 2007.

Ramjerdi, F., Flügel, S., Samstad, H., & Killi, M. (2010). Value of time, safety and environment in passenger transport ? Time - Transportøkonomisk institutt. Retrieved from http://www.toi.no/article29726-29.html

Rouwendal, J., de Blaeij, A., Rietveld, P., & Verhoef, E. (2010). The information content of a stated choice experiment: A new method and its application to the value of a statistical life. *Transportation Research Part B: Methodological*, *44*(1), 136–151.

Schmid, Basil, and Kay W. Axhausen. Investigating suppressed demand effects in a Post Car World: A latent variable random effects Poisson (LVREP) approach. Paper presented at the *17th Swiss Transport Research Conference (STRC 2017)*, Ascona, Switzerland, May 17-19, 2017

Significance, VU Amsterdam and Bates, J. (2013) Values of time and reliability in passenger and freight transport in The Netherlands, report for Ministry of Infrastructure and the Environment.

Small, Kenneth A., Clifford Winston, and Jia Yan. "Uncovering the distribution of motorists' preferences for travel time and reliability." Econometrica 73.4 (2005): 1367-1382.

Staelin, R., Payne, J.W., 1976. Studies of the information-seeking behavior of consumers, in: Cognition and Social Behavior. Lawrence Erlbaum, Oxford, England, pp. xiii, 290–xiii, 290.

Sugden, R., 2009. "Market simulation and the provision of public goods: A non-paternalistic response to anomalies in environmental evaluation," Journal of Environmental Economics and Management, Elsevier, vol. 57(1), pages 87-103, January.

Thaler, R.H., 1999. Mental accounting matters. Journal of Behavioral Decision Making 12, 183–206.

Tribe, L.H., 1972. Policy Science: Analysis or Ideology? Philosophy & Public Affairs 2, 66–110.

Tversky, A., 1969. Intransitivity of preferences. Psychological Review 76, 31–48. https://doi.org/10.1037/h0026750

Tversky, A., Kahneman, D., 1981. The framing of decisions and the psychology of choice. Science 211, 453–458. https://doi.org/10.1126/science.7455683

Varela, J.M.L., Daly, A. and Börjesson, M. (2018) Quantifying errors in travel time and cost by latent variables, Transportation Research B, 117, pp. 520-541.

Vrtic, Milenko, Nadine Schüssler, Alexander Erath and Kay W. Axhausen. (2007). Route, mode and departure time choice behaviour in the presence of mobility pricing. Paper presented at the 86th Annual Meeting of the Transportation Research Board, January 21–25

Weis, Claude, Christoph Dobler and Kay W. Axhausen. 2013. A stated adaptation approach to surveying activity scheduling decisions. In Transport survey methods : best practice for decision making, Johanna Zmud, 569-589. Bingley: Emerald, November 14-18.

Widlert, S. (1994), Stated Preference Studies – The Design Affects the Results, 7th IATBR conference, Chile.

Wilson, T.D., Schooler, J.W., 1991. Thinking too much: introspection can reduce the quality of preferences and decisions. J Pers Soc Psychol 60, 181–192. https://doi.org/10.1037//0022-3514.60.2.181