Paper:
Orr, S, Hess, S and Sheldon, R (2012) *Consistency and fungibility of monetary valuations in transport: an empirical analysis of framing and mental accounting effects.* Transportation Research Part A: Policy and Practice, 46 (10). 1507 – 1516.

http://dx.doi.org/10.1016/j.tra.2012.05.020

Consistency and fungibility of monetary valuations in transport:

an empirical analysis of framing and mental accounting effects

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

Shepley Orr
Centre for Transport Studies
University College London
s.orr@ucl.ac.uk

Rob Sheldon
Accent
Rob.Sheldon@accent-mr.com

The author names are arranged alphabetically

Abstract

Governments around the world use monetised values of transport externalities to undertake project appraisal and cost-benefit analysis. However, because different types of benefits are monetised (e.g., travel time savings, preventing statistical fatalities, reliability, etc.) the question naturally arises as to whether they are consistent. That is, whether a "dollar is a dollar" as welfare economics requires, or whether spending money in one area carries a different disutility from spending money in another area. This would equate to a violation of fungibility, which is the property of a good or a commodity whose individual units are capable of mutual substitution. The view that money is not fungible is explained in behavioural economics through theories of framing and mental accounting. This paper describes the results of a stated choice experiment designed to test the fungibility and consistency of monetary valuations in transport. From a nationally representative sample, we elicit direct values for the three pairwise trade-offs between travel time, travel cost, and safety. We then show that in the context of our analysis, any trade-offs inferred on the basis of other trade-offs, as is common practice (e.g. inferring a safety *vs* time trade-off on the basis of monetary valuations for time and safety), produces biased results, suggesting that the assumption of fungibility does not hold. Specifically, we find that time is valued more highly when valued directly by cost than when traded with safety, and the reverse is true for safety.

Keywords: fungibility; consistency; value of time; value of safety; trade-offs; stated choice; behavioural economics

1. Introduction and statement of the research question

Preference elicitation methods are used throughout the world to produce monetary valuations which in turn are used to underpin policy application. In the United Kingdom (UK), for example, the Department for Transport (DfT) currently uses a number of such valuations for appraising projects. The same is the case for national transport authorities in numerous other countries, and extensive use of such measures is also made by regional or metropolitan transport planning authorities, or indeed by transport operators. While the present paper focuses on two specific such measures, it should be stressed that these have been chosen to illustrate the hypothesis put forward in this paper which is felt by the authors to be more generally applicable.

The first valuation tool utilised in this paper is the monetised valuation of travel time (VTT). The second is the monetised value of preventing a statistical fatality (VPF). When undertaking project appraisal or constructing a business case which might have beneficial impacts on travel time or safety, these values are utilised to construct economic impact assessment. In the UK, these figures are currently reflected in the DfT Transport Analysis Guidance (also known as webtag), with DfT (2009a) for the VTT, and DfT (2009b) for the VPF.

Both the VTT and VPF measures currently used are based on studies employing willingness-to-pay survey methodology to derive the economic valuation, and were published in reports for the DfT (and the then Department of Environment, Transport and the Regions). The VTT figure is derived from Mackie et al (2003), and the VPF figure is derived from Chilton et al (1998) and supporting research by the same research team for the health and safety executive (Chilton et al, 2000), where the figure is also discussed and uprated to current prices in DfT (2007).

The existence of two economic values for cost-benefit analysis which have been separately determined through preference elicitation approaches begs an immediate question: are the monetary valuations consistent? This question is both of methodological interest and of enormous policy relevance. Methodologically, the idea that there might be different types of “mental accounts” or processes of “choice bracketing” would be an interesting explanation for having potentially inconsistent valuations of safety and time (see Thaler, 1990, 1999; Heath and Soll, 1996; Read, Loewenstein and Rabin, 1999; Sloman, 2004). The policy relevance is obvious: if safety and time are valued differently, it could mean that decisions made by the DfT are *relatively* either under- or over-valuing time or safety.

The issue of the possible lack of fungibility also has broader implications. Indeed, while the majority of studies will aim to directly produce estimates of the trade-offs at interest from the data at hand, this is not always the case (or possible). As an example, studies looking at reliability may focus their data collection solely on trade-offs between mean travel time and travel time variability, and then, on the basis of the estimated reliability ratio, infer a monetary valuation of travel time variability by means of an *existing* value of time measure.

To test whether the valuations of the two goods are consistent entails a simple test of consistency. This involves first eliciting the willingness-to-pay estimates for safety £(S) and time £(T) from which a ratio of £(S)/£(T) can be constructed. An experiment is then developed to directly value safety with respect to time through a series of trading opportunities designed to construct a marginal rate of substitution (MRS) between safety and time, i.e. S/T. The test then involves establishing whether the ratio £(S)/£(T) is equal to S/T.

This exploration can then lead to a second area of methodological interest. If the valuations are not consistent, what is the "direction" and "size" of the inconsistency? That is, is safety valued more highly with money than when traded against time, or vice versa? And, if so, by what margin?

The remainder of this paper is organised as follows. The following section talks about the general issue of fungibility of money and mental accounting. This is followed by a discussion about survey design and data collection and the methodological framework used for testing the fungibility assumption. Section 4 presents the findings of the empirical results, with the conclusions of the research being discussed in Section 5 .

2. The Fungibility of Money and Mental Accounting

The assumption of fungibility is crucial to the valuation of any good through the use of money. Money is the ultimate fungible resource, which is a feature of any economic textbook. To address the underpinning of any potential problem of consistency, we employ the idea of "mental accounting" (Thaler, 1985, 1990, 1999; Shafir and Thaler, 2006). In this theory, the use of money can be viewed as a fungible resource which is used consistently to purchase goods *within* a single category, but not necessarily *between* categories. The theory of mental accounting is similar to related ideas in behavioural economics such as framing and is also known as choice bracketing (Loewenstein, Rabin and Read, 1999).

To take the case of mental accounting, consider an early example from Thaler (1985):

> *"Mr. X is up $50 in a monthly poker game. He has a queen high flush and calls a $10 bet. Mr. Y owns 100 (worth $100 – ed.) shares of IBM which went up ½ today and is even in the poker game. He has a king high flush but he folds. When X wins, Y thinks to himself, 'If I had been up $50 I would have called too.'"* (1985: 199).

In this example, Y has (at least) two mental accounts. One is his accounting within the poker game, the other is, say, all other income. Even though Y is ahead in the "overall income" account because of his shares in IBM going up, he does not take this into account in his decision to take a risk in the poker game. This is because in his "poker game" mental account, he is only breaking even and cannot afford to take the risk to call, even though he has an excellent hand and is up $50 for the day, just as is X.

This theory obviously draws on the general phenomenon of framing of choices. Our hypothesis is that subjects will potentially have different mental accounts for money spent on different transport externalities, say in our case safety and on time. For example, one possibility would be that, when trading money against safety, subjects may isolate their choice such that they see safety as something which should be purchased with a great deal of thought and may not compare the costs of their safety costs with all other relevant opportunity costs. Trading money with time however is far more frequent and is less likely to be subject to "mental accounting" effects. So trading money against safety is relatively unfamiliar, while trading money against time is highly familiar, however trading safety against time is highly *un*familiar. This could arguably lead respondents to frame their valuations differently when trading time against safety and create different and unfamiliar mental accounts, and could hence lead to different valuations. More specifically, the results of choice versus valuation types of task lead to a standard preference reversal. And indeed, the presence of the price attribute or of valuing versus choosing is known to create a variety of framing effects and other biases in behavioural decision theory and economics (Loomes, forthcoming) and in stated choice experiments (Gyrd-Hansen and Skjoldborg, 2008).

To see how the theory of framing and mental accounts is also related to the preference reversal phenomenon consider the following example (Slovic and Lichtenstien, 1969; discussed in Tversky and Thaler, 1990). In this, a subject has two bets:

The P-bet: win $30 with 90% probability, and zero otherwise.

The $-bet: win $100 with 30% probability and zero otherwise

The general phenomenon of preference reversal occurs because subjects are given two different tasks, a *valuation* task and a *choice* task. In the choice task, a subject is asked which of the two bets they would prefer. In the valuation task, subjects are asked how much they would be willing to pay to play out each bet. When subjects are given the choice between the P- and $-bet, a large majority choose the P bet. That is, they indicate that they prefer the P-bet which is the risk-averse option. However, when asked to value the two bets, a large majority value the $-bet higher, indicating that they prefer the $-bet which is the risk-seeking option. It is suggested that subjects frame the acts of *choosing* and *valuing* differently (Tversky and Thaler, 1990). The theory states that when asked to *value* the two bets by providing a willingness-to-pay, the subjects focus on money, and hence the bet with the highest possible outcome (the $100 in the $-bet), even though it is more risk-seeking.

In the context of the present study, a similar phenomenon could arise, where, when comparing money against safety, the subject may tend to value safety only somewhat highly because safety is purchased regularly in a variety of forms (eg., insurance, fire alarms, helmets for cycling, etc.). If forms of safety are highly comparable and substitutable, it is also likely that money will be relatively fungible. When valuing time, mental accounting effects may again be very minimal, again because people have a great deal of market experience trading money and time. Both of these forms of trading, money versus time and money versus safety, are frequent enough that subjects are likely to be aware of the opportunity costs of spending and will have little in the way of mental accounting effects. Such choices thus possibly bring subjects back down to “the real world”, where trade-offs are made all the time and opportunity costs are clearly weighted against the *overall* costs of choices, what Loewenstein, Rabin and Read (1999) call “broad bracketing”. However, a choice between safety and time may induce yet another different pattern of reasoning and is arguably far less familiar from market experience and more susceptible to mental accounting and framing effects. Furthermore, in this context, respondents are not “valuing” safety (or time) with money, but effectively *choosing* between amounts of safety versus time.

A possibility is that, when choosing between safety and time, respondents will see safety as far more important than time and hence over-state their true value of safety (or conversely under-state their true value of time). The empirical results presented in this paper are in line with such a hypothesis. However, it should be said that while we find clear evidence of a lack of fungibility, we cannot with certainty say why preferences are found to be non-transitive across different comparisons. Nevertheless, such a different valuation would imply that if the trade-off between safety and time is inferred on the basis of trade-offs between time and money, and trade-offs between safety and money, a biased valuation of the safety *vs* time trade-off may be obtained. The same reasoning applies for any other of the three trade-offs when calculated on the basis of the remaining two trade-offs rather than being obtained directly from a relevant experiment.

A further important issue may arise when attempting to infer a third trade-off on the basis of two separately collected trade-offs that share a common factor. Indeed, in the above DfT example, the time *vs* money and safety *vs* money trade-offs are obtained from two unrelated studies, collected at different times, in different locations. Clearly, further differences may be expected in this case as a result of differences in the samples used, however carefully quotas were applied. In the present paper, we eliminate this further source of differences by using data collected from the same respondents in the same survey.

3. Survey design and data collection

This section describes the experimental framework used to test the assumption of fungibility in the context of monetary valuations for time and safety. Our framework is based on a survey which yields for each respondent independently obtained trade-offs between time and money, safety and money, and time and safety. From these three independent valuations, relative valuations can be obtained that allow us to test our assumptions.

The survey work consisted of three separate stated choice components, trading time against cost, safety against cost, and safety against time. In each case, binary choice scenarios were used, with the two alternatives described by only two attributes. Looking for example at the case of time against money, the respondent would be faced with a choice between a cheaper but slower alternative and a faster but more expensive alternative. The choice of such a simplistic design approach is in line with a large share of current practice in the field, allowing us to draw conclusions that are relevant to the context of the guidance provided to policy makers. Additionally, focussing each scenario on only two separate attributes avoids issues with the possible impact of other considered attributes which may otherwise further affect the degree of fungibility.

The order of the three survey components was randomised across respondents, as was the order of alternatives within choice sets, and the order of the choice sets within each set of scenarios. This approach should help minimise the impact of any ordering effects on our empirical results.

In common with a growing trend in stated choice work in transport, the scenarios presented were framed around a current “reference trip”. In order to have as much control over establishing a commonality between journeys for all subjects, respondents were selected who had made a rail journey during the previous two months between London and over 60 station locations where the rail journey will have been scheduled to have taken approximately one hour (either originating in London or one of the over 60 stations).

This reference trip was then taken as the status quo in a five level design, which made use of two decreases, a *no change* level, and two increases, where the changes were of the order of +/-20% and +/-10%. For example, a subject who had a journey of 60 minutes which cost £10 would face a combination of attribute levels of 48 minutes (-20%), 54 minutes (-10%), 60 minutes (no change), 66 minutes (+10%) or 72 minutes (+20%) for time and £8 (-20%), £9 (-10%), £10 (no change), £11 (10%) or £12 (20%) for cost.

Calculating the levels for safety was not done on an individual-specific basis however. This was partly because individualising frequency of trips to control for baseline levels of risk to each individual would have involved extremely minute probabilities which individuals have great difficulty in comprehending. Additionally, it has long been recognised (e.g. Jones-Lee, Hammerton and Philips, 1985) that published data is not disaggregated at a fine enough grain to allow for possible variance in journeys on different services as injury statistics are publicly available only at the national level.

To overcome the difficulty of communicating probabilities to respondents we instead provided the information that would go into defining a numerator and denominator without providing that ratio itself. Specifically, we provided information on the total number of injuries per annum (the putative numerator) and the average number of rail journeys made in and out of London on an average weekday. We derived these numbers from the following information, using statistics for 2008. There are 12,308 minor injuries per annum on the UK railway (RSSB, 2009), while there are some 2.7 million rail journeys per weekday, approximately 1 million of which are to and from London (DfT, 2010). We calculated the number of injuries by assuming that a proportionate number of injuries of the national average take place to and from London. Therefore we simply multiplied the 12,308 minor injuries by the ratio of 1m/2.7m (i.e. London-specific journeys over all national rail journeys per day). In order to make this number more comprehensible we rounded to an average number of annual injuries to and from London of 4,000/annum. The levels of the safety attribute were therefore 3,200 (-20%), 3,600 (-10%), 4,000 (no change), 4,400 (+10%) and 4,800 (+20%).

The actual stated choice experiments used in each set of scenarios were based on a D-efficient design (see e.g. Bliemer & Rose, 2009) that encouraged trading between the two attributes, maintained attribute level balance as far as possible, and avoided any choice situations with a dominated alternative. In other words, in a scenario involving time *vs* money trade-offs, one option would always be faster, while the other would always be cheaper (using comparisons against each other rather than the base), where, as mentioned above, the ordering was randomised. A design with 30 rows was used, divided into six blocks of five rows, using orthogonal blocking, and with the blocks distributed evenly across the sample of respondents, as well as being distributed independently across the three stated choice components.

As an illustration, Figure 1 shows an example of a time vs money choice scenario, Figure 2 uses a safety vs cost scenario and Figure 3 a safety vs time scenario. These three examples are all for a respondent on a London to Cambridge journey, with a base return fare of £31.80, and a base travel time of 50 minutes.

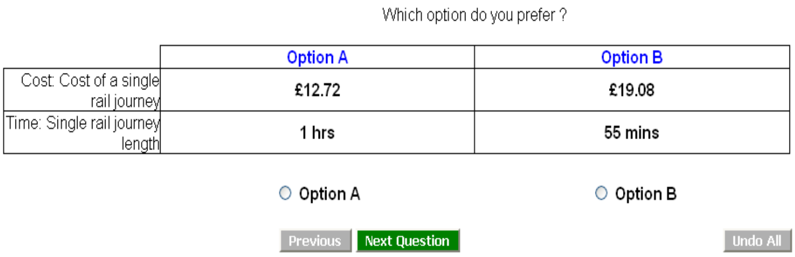

Figure 1: Example of time vs money scenario

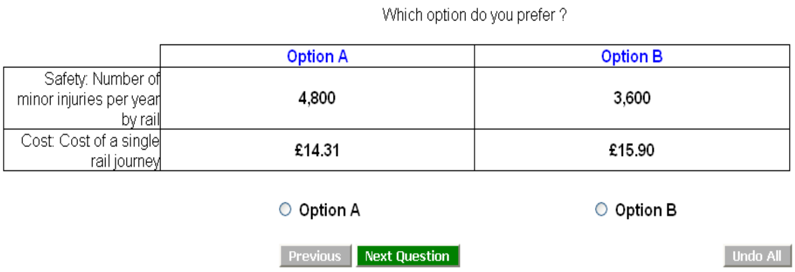

Figure 2: Example of safety vs money scenario

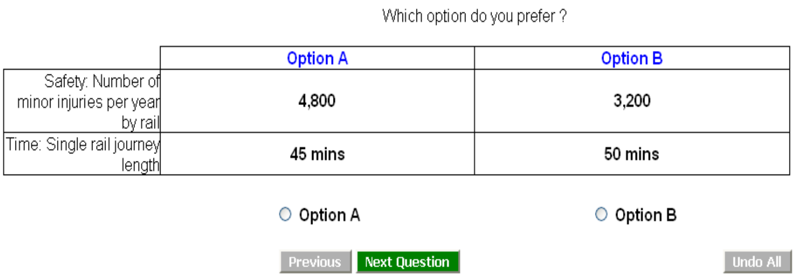

Figure 3: Example of safety vs time scenario

The interviews were conducted through an internet survey, using an established online data collection firm and using an existing respondent panel. This provides a cost effective means of collecting data, and is typical for a growing share of academic and public sector research using stated choice surveys. Subjects were recruited to fill quotas such that the sample would be nationally representative in terms of age and gender. Given a sample of 397 in total, the demographic information is as follows. For gender, there were 198 males (49.9%) and 199 females (50.1%). In terms of age, 183 subjects (46.1%) were between 18 and 34, 205 subjects (51.6%) were between the ages of 35 and 64, with 9 subjects (2.3%) being over 65. Clearly, our sampling approach is not purely random and the resulting data may thus not be nationally representative. This was however never the aim, with our intent being to test a hypothesis, namely exploring the issue of fungibility by establishing whether the same respondent evaluated a series of tasks differently rather than exploring what the different actual values are. We have no a priori reason to suspect that any sampling bias introduced with our specific data collection approach is likely to influence our results in one particular way.

4. Empirical analysis

4.1. Modelling methodology

The standard 'toolkit' for obtaining monetary valuations from data collected using stated choice surveys is to use mathematical structures belonging to the family of random utility models[1]. These models explain the choice amongst a set of mutually exclusive alternatives on the basis of the concept of utility maximisation. The utility of a given alternative is a function of the observed characteristics of the alternative (e.g. time, cost, etc) and the sensitivities or tastes of the respondent. A respondent is assumed to choose the alternative with the highest utility in each choice set. The sensitivities of the respondent are not observed and need to be estimated, where the outputs from model estimation are those parameters that best explain the choices observed in the data.

The main interest in the present study was to look into the validity or otherwise of the fungibility assumption in the context of a simple survey. From this perspective, a very basic modelling approach was used, based on simple Multinomial Logit (MNL) models. The MNL model is based on the assumption that the unobserved components of utility follow a type I extreme value distribution across alternatives and observations, and we write the utility for alternative *i* in choice task *t* for respondent *n* as:

$$U_{int} = V_{int} + \varepsilon_{int} \qquad [1]$$

where V_{int} and ε_{int} represent the deterministic and random components of utility, respectively. The deterministic component of utility is given by:

$$V_{int} = f(x_{int}, \beta) \qquad [2]$$

where β *is the vector of sensitivities that is to be estimated from the data. In most applications, a linear-in-parameters specification is used.*

In the MNL model, the probability of respondent *n* choosing alternative *i* (out of *J* different alternatives) is then given by:

$$P_{int} = exp(V_{int}) / \sum_j exp(V_{jnt}) \qquad [3]$$

where $j=1,...,J$.

As mentioned above, the key output from a model of this type are the estimates of the vector of parameters β. Random utility models are based on the concept of respondents trading off attributes against one another, and the primary interest is consequently on relative values of the different components of β. In particular, we wish to establish the relative impact on the utility of changes in two attributes, say x_1 and x_2. This is made possible by calculating:

$$T = \frac{\partial V}{\partial x_1} \Big/ \frac{\partial V}{\partial x_2} \qquad [4]$$

[1] See Train (2009) for a thorough introduction to random utility modelling.

i.e. the ratio of marginal utilities. With a linear-in-attributes specification, i.e. $V_{int}= \beta' x_{int}$, we would have that T= β_1 / β_2. If T is larger than 1, it would imply that a one unit change in x_1 has a greater impact than a one unit change in x_2. If attribute x_2 is the cost attribute, then T will give the monetary valuation of a one unit change in x_1.

Initial model estimations showed that, with the present data, the significant variation in the base fare across respondents warranted the use of a logarithmic transform for the fare attribute, thus leading to decreasing marginal fare sensitivity, in line with many other studies and official guidelines (cf. Daly, 2010). This implies that as fare increases, the impact of each additional unit increase in fare becomes smaller. Consequently, it also means that the impact of a one unit increase in fare is smaller at a higher base fare. A separate analysis showed that the use of a logarithmic transform for the cost attribute had no impact on the findings in terms of fungibility when compared to a purely linear model, but led to better modelling performance and more robust results (details available on request). A linear specification was used for the travel time and safety coefficients. In the context of looking at sample population level valuations, no attempts were made to incorporate socio-demographic interactions or to allow for random taste heterogeneity.

In summary, the specification used for an alternative in the time vs money scenarios (say T-M) is thus given by:

$$V_{int,T\text{-}M} = \beta_{time} T_{int} + \beta_{log\text{-}cost} ln(C_{int}) \qquad [5]$$

where T_{int} is the travel time for alternative i in choice situation t for respondent n, while C_{int} gives the associated cost. The estimated parameters β_T and $\beta_{log\text{-}C}$ give the marginal impacts of unit increases in the travel time and in the natural logarithm of travel cost.

Corresponding specifications for the safety vs time (say S-T) and safety vs money (say S-M) scenarios are given by:

$$V_{int,S\text{-}T} = \beta_{time} T_{int} + \beta_{safety} S_{int} \qquad [6]$$

and

$$V_{int,S\text{-}M} = \beta_{safety} S_{int} + \beta_{log\text{-}cost} ln(C_{int}) \qquad [7]$$

where S_{int} is the safety attribute for alternative (in 1000s of injuries) i as faced by respondent n in choice situation t, and β_S is the associated marginal utility coefficient.

With this specification, we can calculate three separate trade-offs directly from our estimates. Using Equation [4], it can be seen that, in the time vs money scenario, we have that the monetary valuation of a one minute change in travel time is given by:

$$VTT = \beta_{time} / \beta_{log\text{-}cost} \cdot \text{cost} \qquad [8]$$

where, due to the log-transform on cost, the value of time is now a function of cost, and increases with cost. From the time vs safety experiments, we can work out the relative sensitivity to safety and time, given by:

$$T_{S\text{-}M} = \beta_{safety} / \beta_{time} \qquad [9]$$

which is the value in minutes of a reduction in the number of injuries by 1,000.

Finally, from the safety vs cost scenarios, we can work out the monetary valuation of a reduction in the number of injuries by 1,000, given by:

$$VS = \beta_{safety} / \beta_{log\text{-}ost} \cdot cost \qquad [10]$$

which once again increases with cost.

Four different models were estimated on the data. We first estimated separate models for each of the three stated choice components, using the utility specifications from Equation [5] to Equation [7]. We then also estimated a joint model on the entire dataset. In this model, it is important to recognise that the relative weight of the error in Equation [1] may vary across the three types of scenarios as some choices may be more deterministic from the analyst's perspective (i.e. easier to explain on the basis of the estimated parameters). Such scale differences[2] can be accommodated by estimated separate scale parameters for the different types of experiments, namely replacing Equations [5] to [7] by:

$$V_{int,T\text{-}M} = \mu_{T\text{-}M}[\beta_{time} T_{int} + \beta_{log\text{-}cost} \ln(C_{int})] \qquad [11]$$

$$V_{int,S\text{-}T} = \mu_{S\text{-}T}[\beta_{time} T_{int} + \beta_{safety} S_{int}] \qquad [12]$$

$$V_{int,S\text{-}M} = \mu_{S\text{-}M}[\beta_{safety} S_{int} + \beta_{log\text{-}cost} \ln(C_{int})] \qquad [13]$$

where, for identification reasons, we fix $\mu_{T\text{-}M}$ to a value of 1, meaning that the remaining two scale parameters related to the impact of the error in the safety vs time scenarios and safety vs money scenarios relative to that in the time vs money scenarios. In these joint models, the estimation of each coefficient is informed by the data from two out of the three sets of scenarios, e.g. the travel time coefficient is informed by the time vs money trade-offs and by the time vs safety trade-offs.

All models were estimated using BIOGEME (Bierlaire, 2003). Given the repeated choice nature of the data, an appropriately specified sandwich estimator for the covariance matrix was used with a view to correcting any downwards bias in the estimated standard error (cf. Daly & Hess, 2011).

4.2. Main estimation results

The main estimation results are summarised in Table 1. The results show the expected negative impacts of increases in travel time, the logarithm of cost, and the number of injuries, with all effects attaining high levels of statistical significance. Looking at the results for the three individual models, we observe significantly higher fit (in terms of adj. ρ^2) for the model estimated on the time vs cost data than for the remaining two components, with a further small decrease for the safety *vs* cost data. Additionally, it is clear that the scale in the safety *vs* time and the safety *vs* cost data is significantly lower than that in the time *vs* cost data, a situation that is reflected in the direct scale estimates in the joint model. These results would suggest a lower degree of error in the model for the time *vs* cost component (i.e. a more deterministic choice process), which is consistent with the notion that respondents are familiar with trading time against cost, where this is not necessarily the case for the survey components involving safety.

[2] Scale is inversely proportional to the variance of the error term.

Table 1: Main estimation results

	Time vs cost		Safety vs time		Safety vs cost		Joint	
Observations	1,985		1,985		1,985		5,955	
Log-likelihood	-901.89		-1,184.39		-1,217.76		-3,308.82	
Parameters	2		2		2		5	
adj. ρ^2	0.3431		0.1377		0.1135		0.1972	
	est.	**t-rat.**	**est.**	**t-rat.**	**est.**	**t-rat.**	**est.**	**t-rat.**
β_{time}	-0.0634	-12.93	-0.0255	-6.06	-		-0.0604	-11.92
$\beta_{log\text{-}cost}$	-8.59	-17.71	-		-3.6	-12.01	-8.63	-17.55
β_{safety}	-		-1.09	-12.98	-0.801	-11.16	-2.12	-9.68
$\mu_{T\text{-}M}$	-		-		-		1	-
$\mu_{S\text{-}T}$	-		-		-		0.517	-9.51*
$\mu_{S\text{-}C}$	-		-		-		0.399	-15.61*

* t-ratio for scale parameters calculated in relation to a base value of 1

As a next step, we calculate trade-offs on the basis of the estimation results, with final values and computed t-ratios reported in Table 2, where it should be noted that while the monetary valuation of time reductions is given in pence per minute, the corresponding measure for safety is given in £ per 1,000 injuries. For the two trade-offs involving cost, the valuations interact with the actual fare level, and the estimates reported here are for a base fare of £15, which is the mean fare in the estimation sample. The monetary valuations at a base of £10 would be two thirds of the values reported here, and this would give us a valuation of travel time roughly in line with the recommended UK values (DfT, 2009a).

Table 2 also shows the trade-offs from the joint model, where it can be noted that each trade-off now makes use of data from all three types of scenario, with e.g. the time vs cost trade-off making use of the time vs money and safety vs money scenarios for $\beta_{log\text{-}cost}$, and the time vs money and safety vs time scenarios for β_{time}. We can observe that while the valuation of travel time from the joint model is within around five percent of the estimate from the separate model, the differences are larger for the remaining two trade-offs. The results from the joint model are arguably more representative, being based on a larger sample as well as with each coefficient coming from two different scenarios.

Table 2: Estimated trade-offs (at a base cost of £15)

	Time vs cost		Safety vs time		Safety vs cost		Joint	
	est.	**t-rat.**	**est.**	**t-rat.**	**est.**	**t-rat.**	**est.**	**t-rat.**
VTT - time vs cost (p/min)	11.07	15.77	-		-		10.50	15.37
$T_{S\text{-}M}$ - safety vs time V(min/1000	-		42.75	6.53	-		35.10	9.61
VS - safety vs cost (£/1000 acc)	-		-		3.34	9.81	3.68	10.63

4.3. Test of fungibility

In this section, we now explicitly test the assumption of fungibility. That is, we for example assume that we are in a position where we have results from a study trading time against cost, and from a study trading safety against time. We thus obtain the valuation of travel time as well as the relative sensitivity to safety and time. From this, we wish to infer the monetary valuation of reductions in injuries.

The calculations required to obtain these results are straightforward. In particular, with β_{time}, $\beta_{log\text{-}cost}$ and β_{safety} representing the marginal utility coefficients, we have:

$$v_1 = \beta_{time} / \beta_{log\text{-}cost} \cdot cost \quad [14]$$

$$v_2 = \beta_{safety} / \beta_{time} \quad [15]$$

$$v_3 = \beta_{safety} / \beta_{log\text{-}ost} \cdot cost \quad [16]$$

i.e. using Equations [8] to [10].

It can then easily be seen that in order to infer a given trade-off on the basis of the remaining two trade-offs, we have:

$$v_1^* = v_3/v_2 \quad [17]$$

$$v_2^* = v_3/v_1 \quad [18]$$

$$v_3^* = v_2 v_1 \quad [19]$$

There is clearly no guarantee that the inferred trade-offs are equivalent to the estimated trade-offs. In fact, it can be seen that this would only be the case if the estimates from the individual models were equivalent to those from the joint model, except for a difference in scale. The percentage bias in a given inferred trade-offs against the directly estimated trade-off can be calculated as:

$$b_1 = (v_1^* - v_1)/v_1 \quad [20]$$

$$b_2 = (v_2^* - v_2)/v_2 \quad [21]$$

$$b_3 = (v_3^* - v_3)/v_3 \quad [22]$$

A closer inspection also reveals that $b_1 = b_2$.

The results of this calculation are shown in Table 3. We once again show the estimated trade-offs, this time with 95% confidence intervals. We then show the inferred trade-offs, along with computed standard errors and t-ratios. A first observation is that we see a decrease in the t-ratios for the two monetary trade-offs, with an increase in the t-ratio for the trade-off that involves safety and time. This was to be expected, given that this latter trade-off had the lowest t-ratio in estimation, meaning that inferring it on the basis of measures with a lower error will also indicate a higher level of statistical robustness.

In terms of actual bias, we observe the expected equal bias for the first two inferred trade-offs (see earlier point). The actual levels of bias are worryingly high, at -29.47% for the first two trade-offs, and +41.79% for the third trade-off. The results also show that the inferred trade-offs for time *vs* cost and safety *vs* cost fall outside of the 95% confidence intervals for the actual estimates of the trade-offs, with the inferred trade-off for the safety *vs* time trade-off falling just above the lower limit of the 95% confidence interval for the actual estimate. This, in conjunction with the size of the bias, sheds serious doubts on the validity of the fungibility assumption in this study.

Table 3: Test of fungibility assumption

	from estimates				
	est.	s.e.	t-rat.	5^{th} perc.	95^{th} perc.
VTT - time vs cost (p/min)	11.07	0.70	15.77	9.70	12.45
$T_{S\text{-}M}$ - safety vs time V(min/1000 acc)	42.75	6.54	6.53	29.92	55.57
VS - safety vs cost (£/1000 acc)	3.34	0.34	9.81	2.67	4.00

	Inferred						
	est.	s.e.	t-rat.	5^{th} perc.	95^{th} perc.	bias	asy. CDF point
VTT - time vs cost (p/min)	7.81	1.44	5.44	4.99	10.62	-29.47%	1.67E-06
$T_{S\text{-}M}$ - safety vs time V(min/1000 acc)	30.15	3.62	8.33	23.05	37.24	-29.47%	2.71E-02
VS - safety vs cost (£/1000 acc)	4.73	0.78	6.03	3.20	6.27	41.79%	1.00E+00

It is also possible to compare the inferred trade-offs to those obtained from the joint model, with results reported in Table 4. The errors from this calculation are lower than those in Table 3, which is to be expected as the joint model estimates provide a more overall measure, with each trade-off being influenced by data from all three survey components (one involving the two concerned coefficients and one component each involving only one of the two concerned coefficients). Nevertheless, the errors remain large, and once again only the inferred trade-off for safety against time lies within the 95% confidence interval from the 'true' value. Clearly, the equality between b_1 and b_2 no longer applies as the inferred values are calculated from the estimates of the individual models and the bias is calculated against the estimates from the joint model.

Table 4: Comparison with joint model

	from joint model				
	est.	s.e.	t-rat.	5^{th} perc.	95^{th} perc.
VTT - time vs cost (p/min)	10.50	0.68	15.37	9.16	11.84
$T_{S\text{-}M}$ - safety vs time V(min/1000 acc)	35.10	3.65	9.61	27.94	42.26
VS - safety vs cost (£/1000 acc)	3.68	0.35	10.63	3.01	4.36

	Inferred						
	est.	s.e.	t-rat.	5^{th} perc.	95^{th} perc.	Bias	asy. CDF point
VTT - time vs cost (p/min)	7.81	1.44	5.44	4.99	10.62	-25.63%	4.10E-05
$T_{S\text{-}M}$ - safety vs time V(min/1000 acc)	30.15	3.62	8.33	23.05	37.24	-14.11%	8.75E-02
VS - safety vs cost (£/1000 acc)	4.73	0.78	6.03	3.20	6.27	28.43%	9.99E-01

In the context of the present paper, it is clearly the earlier comparison that is of the greatest interest, given that the need for fungibility only arises in the case of studies where one specific trade-off cannot be calculated directly from model estimates. Whichever of the two comparisons is used, our results are however certainly cause for concern, with the same applying to results from a purely linear-in-attributes specifications (not reported here). While our results are obviously just for a single dataset and more evidence is required, the results themselves are reliable as they are from a highly controlled environment without any external factors that could affect fungibility.

5. Conclusion

This paper has presented evidence from a study looking at trade-offs between time, safety, and money, with each respondent facing three sets of stated choice scenarios, based on the three possible pairs of attributes. The interest of the study was to test for fungibility, i.e. whether a trade-off for one pair of attributes could be reliably obtained on the basis of the two remaining trade-offs. This is not only of interest from a behavioural economics perspective, but is also important in many areas where there is a desire to monetise valuations obtained from surveys that do not have a cost component (e.g. there may attempts to monetise a trade-off between travel time and reliability on the basis of an *established* value of time). To as far as possible eliminate external influences and allow us to provide a controlled environment for our test, the three valuations were obtained in a single study, and in a single survey sitting.

The results of our study show a clear lack of fungibility in monetary valuations. The conclusions as to a finding of inconsistency which violates the axioms of rational choice/expected utility theory are, however, difficult to draw. What we do know is that when valued directly by using money, time is valued at a higher level than safety compared to the value of time when valued directly with safety. To see why, compare the results from Table 3. The estimated or direct valuation of time/cost is 11.07 pence/minute, while the inferred, indirect valuation of time then *decreases* to 7.81 pence/minute. However, the estimated or direct valuation of safety/cost is £3.34 per 1,000 injuries, while the inferred, indirect valuation of safety increases to £4.73 per 1,000 injuries.

If we allow that the direct valuation is more likely the true valuation, then the process of benefit transfers from other monetary valuations is inadvisable. This is our primary policy conclusion. At the level of our methodological conclusions we require a cognitive mechanism or set of mechanisms which makes the valuation of time higher when compared directly with money and lower (or less important) when traded directly against safety. We argue that these are mechanisms concerning mental accounting and framing (or choice bracketing) of goods, and that the way that these goods are traded in a variety of contexts is partly what determines what types of framing are employed by respondents.

There are two further directions for research which follow from the present research. It would be interesting to compare the results obtained from stated choice experiments with results obtained from the contingent valuation method (CVM). Contingent valuation is the method used for the valuation of rail and road safety in the UK (Jones-Lee and Loomes, 2003), and it could be argued that the methodology typically creates more of a focussing effect in choice, as a package of goods is evaluated and traded *solely* against money. And this single, or “one-off” valuation using CVM would also be usefully contrasted with a SC exercise in which all three attributes of time, safety and cost were traded, and an analysis in which we explicitly allow for referencing effects and asymmetrical preference formation (see e.g. Hess et al., 2008). Additionally, the present paper has made no attempts to investigate the *causes* for the lack of fungibility, and future work should investigate this, e.g. whether the violation of fungibility is stronger in some segments than others. Furthermore, it would be of interest to study the impact of model specification on fungibility, for example in terms of the treatment of heterogeneity. Finally, the results presented here clearly are related to just a single dataset, albeit one from a very controlled experiment in which the number of factors possibly influencing the validity of the fungibility assumption should be low. Further evidence from other datasets would be desirable, including making use of wider sets of attributes.

ACKNOWLEDGMENTS

The authors are grateful for comments at previous presentations of this paper at the European Transport Conference 2010 and at the Department for Transport. The first author acknowledges the support of the Leverhulme Trust in the form of a Leverhulme Early Career Fellowship. The authors are grateful to an anonymous referee for comments that substantially improved the quality of the paper.

REFERENCES

Bierlaire, M. (2003). BIOGEME: A free package for the estimation of discrete choice models , Proceedings of the 3rd Swiss Transportation Research Conference, Ascona, Switzerland.

Bliemer, M. and Rose, JM. (2009). "Designing Stated Choice Experiments: State of the Art" in *The Expanding Sphere of Travel Behaviour Research: Selected Papers from the 11th International Conference on Travel Behaviour Research*, ed. R. Kitamura, T. Yoshii and T. Yamamoto, Emerald Group Publishing Limited , Bingley, United Kingdom, pp. 499-538

Chilton, S., Covey, J., Hopkins, L., Jones-Lee, M., Loomes, G., Pidgeon, N. & Spencer, A. (1998) New research results on the valuation of preventing fatal road accident casualties, In: Road Accidents Great Britain 1997, The Casualty Report, Department of the Environment Transport and the Regions.

Chilton, S., Dolan, P., Jones-Lee, M., Loomes, G., Robinson, A., Carthy, T., Covey, J., Spencer, A , Hopkins L., Pidgeon, N., Beattie, J. (2000). *Valuation of Benefits of Health and Safety Control: Final Report, Health and Safety Executive.* Available to download from the HSE website at: http://www.hse.gov.uk/research/misc/crr pdf/2000/crr00273.pdf

Daly, A. (2010). Cost Damping in Travel Demand Models Report of a study for the Department for Transport. http://www.dft.gov.uk/pgr/economics/rdg/costdamping/, also TAG 3.10.2d UK Department for Transport transport analysis guidance, www.dft.gov.uk/webtag/documents/expert/pdf/unit3.10.2d.pdf

Daly, A.J. & Hess, S. (2011), Simple Approaches for Random Utility Modelling with Panel Data, paper presented at the 90th Annual Meeting of the Transportation Research Board, Washington, D.C.

DfT (2009a). Value of Time and Operating Costs, TAG 3.5.6, UK Department for Transport transport analysis guidance, http://www.dft.gov.uk/webtag/documents/expert/pdf/unit3.5.6.pdf

DfT (2009b). Accidents Sub-objective, TAG 3.4.1, UK Department for Transport transport analysis guidance, http://www.dft.gov.uk/webtag/documents/expert/pdf/unit3.4.1.pdf

DfT (2007), Highways Economics Note No.1 (HEN 1), UK Department for Transport, London.

DfT (2010). *Transport Statistics Great Britain 2009*. Transport Stationary Office, UK Department for Transport, London TSO, http://www.dft.gov.uk/adobepdf/162469/221412/217792/4212241/transportstatisticgreatbrit.pdf

Gyrd-Hansen, D. and Skjoldborg, U.S. (2008). "The price proxy in discrete choice experiments: Issues of relevance for future research". Pp. 175-194 in M. Ryan, K. Gerard and M. Amaya-Amaya (Eds.) *Using Discrete Choice Experiments to Value Health and Health Care*. Dordrecht: Springer.

Heath, C. and Soll, J.B. (1996). "Mental budgeting and consumer decisions", *Journal of Consumer Research*, 23: 40-52.

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

Jones-Lee, M., Hammerton, W. and Philips, P.R. (1985). "The value of safety: Results of a national sample survey". *Economic Journal*, 95: 49-72.

Jones-Lee, M. and Loomes, G., (2003). Valuation of safety. In: Hensher, D.A., Button, K.J. (Eds.), *Handbook of Transport and the Environment*, Handbooks in Transport, vol. 4. Elsevier, Oxford, pp. 451–462.

Lichtenstein, S. and Slovic, P. (1971). "Reversals of preference between bids and choices in gambling situations, " *Journal of Experimental Psychology*, 89: 46-55.

Loomes, G. (forthcoming) "Modelling choice and valuations in decision experiments", *Psychological Review*.

Loomes, G., Starmer, C. and Sugden, R., 2010. Preference reversals and disparities between willingness to pay and willingness to accept in repeated markets. *Journal of Economic Psychology*, 31: 374-387.

P.J Mackie, M. Wardman, A.S Fowkes, G. Whelan, J Nellthorp and J Bates, (2003). Values of Travel Time Savings in the UK. Institute for Transport Studies, University of Leeds, Working Paper 567.

RSSB (2009). *Annual Safety Performance Report 2008*. Railway Safety and Standards Board (RSSB), Report no longer on RSSB website, but see press release with relevant information here: http://www.rssb.co.uk/SiteCollectionDocuments/press/2009/ASPR%202008%20press%20release.pdf

Read, D., Loewenstein, G. and Rabin, M. (1999). "Choice bracketing". *Journal of Risk and Uncertainty*, 19: 171-197.

Shafir, E. and Thaler, R.H. (2006). "Invest now, drink later, spend never: On the mental accounting of delayed consumption." *Journal of Economic Psychology*, 5: 694-712.

Sloman, D. (2004). "Framing, loss aversion and mental accounting", pp. 379-98 in D. Koehler and N. Harvey (Eds.) *Blackwell Handbook of Judgement and Decision Making*. Oxford: Blackwell.

Thaler, R.H. (1985). "Mental accounting and consumer choice". *Marketing Science*, 4: 199-214.

Thaler, R.H. (1990). “Savings, mental accounting and fungibility”. *Journal of Economic Perspectives*, 4: 193-205.

Thaler, R.H. (1999). “Mental accounting matters”. *Journal of Behavioral Decision Making*, 12: 183-206.

Train, K.E. (2009). *Discrete Choice Methods with Simulation (second edition)*, Cambridge University Press, Cambridge, MA.

Tversky, A. and Thaler, R.H. (1990). “Anomalies: Preference reversals”, Journal of Economic Perspectives, 4: 201-211.