WORKING PAPER
ITS-WP-04-12

Identifying the Influence of Stated Choice Design Dimensionality on Willingness to Pay for Travel Time Savings

By

David A Hensher

April, 2004

ISSN 1440-3501

INSTITUTE OF TRANSPORT STUDIES
The Australian Key Centre in Transport Management

The University of Sydney and Monash University

Established under the Australian Research Council's Key Centre Program.

NUMBER: Working Paper ITS-WP-04-12

TITLE: Identifying the Influences of Stated Choice Design Dimensionality to Pay for Travel Time Savings

ABSTRACT: This paper explores the influence of the dimensions of stated choice (SC) designs on the value of travel time savings. Utilising principles of experimental design, 16 choice designs are embedded within a global design in which we vary the number of choice sets, the number of alternatives in each choice set, the number of attributes per alternative, the number of levels of each attribute and the range of attribute levels. A mixed logit model is estimated in which design dimensions are interacted with the attribute parameters to explore the influence of these dimensions on willingness to pay (WTP) for travel time savings. The evidence in the context of a sample of respondents in Sydney choosing amongst trip attribute bundles for their car commuting trip suggests that design dimensionality does influence variations in WTP; with higher overall mean values of travel time savings associated with more complex designs (in terms of the number of items to process). Given evidence external to the SC design that most individuals tend to add up the components of travel time in processing bundles of attributes, we interacted aggregate travel time with a choice complexity index as a more general test of the directional impact on VTTS due to the net effect of all positive and negative design influences. The prior finding was confirmed. Importantly, the evidence should not be taken to suggest that simpler designs (i.e., less items) are preferred, but rather that individuals process the information in SC designs in ways that appear to support at least two hypotheses, one related to coping strategies and the other related to behavioural relevance. We recommend, however, that the selection of the design specifications (e.g., relevant attributes and their range) should be guided by, at least, the specific application setting.

KEY WORDS: ***Stated choice designs, design of designs, willingness to pay, CAPI, behavioural response***

AUTHOR: David A Hensher

CONTACT: Institute of Transport Studies (Sydney & Monash)
The Australian Key Centre in Transport Management, C37 The University of Sydney NSW 2006, Australia

Telephone: +61 9351 0071
Facsimile: +61 9351 0088
E-mail: itsinfo@its.usyd.edu.au
Internet: http://www.its.usyd.edu.au

DATE: April, 2004

Research funded under the Australian Research Council Large Grants Scheme, Grant A00103962

1. Introduction

Stated choice (SC) methods are used extensively to reveal the willingness to pay (WTP) for specific attributes. Within the SC setting, sampled agents typically assess a number of alternatives defined by a set of attributes, each of which is offered as a level drawn from a pre-specified set of levels and range of levels, and are asked to choose the most preferred alternative (including the choice to not choose any of the offered alternatives). This assessment is repeated a number of times up to the total number of choice sets that are being offered. The data is then subject to discrete choice modelling using tools such as multinomial logit and mixed logit to establish the amount that an individual is willing to pay (in the current context for travel time savings) for changes in the levels of specific attributes associated with one or more alternatives. The mixed logit model gives us the capability of parameterising unobserved heterogeneity to derive distributions of WTP for travel time savings[1].

Stated choice experiments typically are based on a pre-specified design plan in respect of the number of attributes (including their levels and range), the number of alternatives in a choice set and the number of choice sets to be assessed. While some studies allow for variations in some of these design dimensions, it is more common for all sampled agents to be given the exact same number of attributes, alternatives and choice sets. Without any variation in the dimensionality of the design, it is not possible to assess what influence the design per se has on WTP. Does the design impact in some systematic or non-systematic way on the parameters associated with each attribute and hence on WTP?

Although attention has been paid to design optimisation in the published literature, particularly optimisation of parameter efficiency, it has not adequately been recognised that one cannot optimise choice experiments without understanding and incorporating the likely effects of the nature and complexity of the experiment itself on model parameters and hence behavioural outputs such as WTP (Louviere and Hensher, 2001; Kamakura et. al., 1996).

The focus on choice task complexity is especially interesting when viewed more broadly under what we call the *information processing strategy* (IPS) of a decision maker. Individual's use a range of IPS's according to their capability to process, which is linked to cognitive capability, commitment to effort etc. It is also related to the risk spectrum they wish to operate under ranging from risk aversion to risk proneness. The greater the risk aversion, the smaller the variance in IPS. The variability in risk is often defined by constructs such as habit formation and variety seeking, both of which suggest mechanisms used to satisfy the individual's commitment of effort and cognitive abilities. If we knew what role these constructs played in behavioural response we could design an SC experiment tailored to a specific IPS[2]. Our challenge herein becomes the inverse – to have a

[1] MNL models can also provide distributions by including higher-order polynomials (e.g., quadratics) and interactions with covariates.
[2] Such an SC experiment has some similarities to an adaptive choice experiment in which alternative behavioural choice response segments are identified as a way of recognising decision rules such as 'hard-core

sufficiently wide ranging set of SC experiments that enable us to identify the role that design dimensionality (as a contributor to task complexity) has on the choice response and hence the role of specific attributes (from which the value of travel time savings is revealed). DeShazo and Fermo (2001) for example show that an increase in the quantity of information provided increases the variance with which individuals make their choices, but that if one increases the number of alternatives in a choice set up to a threshold number, the variance decreases and then increases. White et al., (1998) show that the attribute range can significantly alter parameter estimates and that it can (and should) be separated from the effects due to variability within an attribute (across the sample). However the caveat is that manipulating the range on attributes that are not key influences on choice response may have little or no effect on the parameter estimates (i.e., they are simply ignored). Additional non-design information can assist in revealing the IPS, such as the inclusion/exclusion plan for each attribute as well as an aggregation plan (eg the adding up of attributes such as components of travel time) (see Hensher 2004 for further details).

A key message is that choice or response variability is a *behavioural phenomenon*, and is an outcome of a choice experiment as much as observed choices and/or model preference parameters or specifications. Design dimensionality needs to be allowed for in the specification of the utility expressions associated with each alternative. This can be incorporated in a number of ways including its treatment in the observed set of influences or as a conditioning effect on the unobserved influences. Previous studies, while adding to our knowledge of design influences, have concluded that there is a great deal still to learn about the behavioural implications of the design of choice experiments. This is the challenge of this paper.

In the current empirical context, we wish to establish whether the mean and the full distribution for VTTS can be shown to vary systematically with the dimensionality of the SC design and whether there is any evidence of directional implication as the design has more individual elements to evaluate (as one indicator of task complexity). This paper is organised as follows. We begin with a design plan referred to as the *Design of Designs* (DoD) including details of the full set of designs. The empirical context is briefly described, followed by an overview of mixed logit and the results of a multinomial and mixed logit choice models. The WTP distributions and the influence of design dimensionality (DD) are then presented, followed by conclusions and directions for further research.

loyal', 'brand-type', IIA-type and product or service form. This was considered by Kamakura et al., (1996) as a finite mixture of nested logits (brand and product), latent class (for hard-core) and multinomial logit (IIA) models.

2. The Design Plan

Using the context of a car commuter trip, we proposed five design dimensions that are cited in the literature (e.g., Ohler et al., 2000; White et al., 1998; DeShazo and Fermo, 2001; Dellaert et al., 1999; Brazell and Louviere, 1998) as the key dimensions of stated choice experiments and which are likely to have the greatest contextual influence on choice response and WTP. These dimensions are summarised in Table 1 together with the combination offered within each of the16 stated choice (SC) experiments[3].

The 16 SC designs are embedded in one overall design, each with 32 rows. A row is a choice set comprising a number of alternatives. Each respondent is given 16 rows (i.e., choice sets), with an additional blocking variable[4] (with two levels) used to determine sets of 16 rows. Each run of the design determines the specification of a choice experiment that has two versions. For example, the first run might have 15 choice sets of three alternatives each presenting four attributes at three levels.

Table 1 The Sub-Designs of the Overall Design

Number of choice sets	Number of alternatives	Number of attributes	Number of levels of attributes	Range of attribute levels
15	3	4	3	Base
12	3	4	4	Wider than base
15	2	5	2	Wider than base
9	2	5	4	Base
6	2	3	3	Wider than base
15	2	3	4	Narrower than base
6	3	6	2	Narrower than base
9	4	3	4	Wider than base
15	4	6	4	Base
6	4	6	3	Wider than base
6	3	5	4	Narrower than base
9	4	4	2	Narrower than base
12	3	6	2	Base
12	2	3	3	Narrower than base
9	2	4	2	Base
12	4	5	3	Narrower than base

Six attributes have been selected based on earlier studies (Hensher, 2000, 2001). They are: free flow time (FFT), slowed down time (SDT), stop/start time (SST), trip time variability (TTV), toll cost (TLC), and running cost (RC) (based on c/litre, litres/100km). Given that the 'number of attributes' dimension has four levels, we have selected the following

[3] Other possible elements might have been included but we selected those that most analysts have raised as possible sources of response bias. We excluded the ordering of attributes.
[4] A blocking variable in essentially another 'attribute' in a design that is used to allocate sub sets of choice sets from the fractional set to each respondent.

combinations of the six attributes, noting that the aggregated attributes are combinations of existing attributes[5]:

- *designs with three attributes:* total time (free flow + slowed down + stop/start time), trip time variability, total costs (toll + running cost)
- *designs with four attributes:* free flow time, congestion time (slowed down + stop/start), trip time variability, total costs
- *designs with five attributes:* free flow time, slowed down time, stop/start time, trip time variability, total costs
- *designs with six attributes:* free flow time, slowed down time, stop/start time, trip time variability, toll cost, running cost

The aggregation of the different time and costs components allows one to systematically account for the effect of the design dimensions on the willingness-to-pay estimates.

We have selected a generic design (ie unlabelled alternatives) for a number of reasons, including the avoidance of any confoundment with labelling. A generic task would ensure that the effect of changing the number of alternatives is due to just that (namely, increasing that number) rather than to the labelling of the alternatives themselves. The design was also made smaller by restricting estimable effects to linear effects only.

The specific design pivots off of the attribute levels associated with a current car-commuting trip. In selecting the car commuter setting, we assumed that all commuters would have to undertake a trip to or from work. Although opportunities do exist for telecommuting and other forms of distributive work, we reasonably rejected a scenario in which a respondent might not undertake a trip (i.e., the no choice option). This also simplifies the task and ensures one less source of confoundment, since the ‘no choice’ alternative is not described by any of the design attributes.

The designs are computer-generated. They aim at minimising the correlations between attributes and maximising the amount of information captured by each choice task. We maximised the determinant of the covariance matrix, which is itself a function of the estimated attribute parameters (within the experimental design literature this is known as D-optimality – see Hensher et al (2004)). Insights from past studies (Hensher 2001a,b) determined their approximate values.

The levels applied to the choice task differ depending on the range of attribute levels as well as on the number of levels for each attribute. The levels are variations from the attribute value of a recent trip. The variations used in the choice tasks are summarised in Table 2. The number of attribute levels and the range of these levels are identical within each of the 16 designs defined by the master plan. They only vary across designs. Each sampled commuter is given a varying number of choice sets, but the number of attributes

[5] This is an important point because we did not want the analysis to be confounded by extra attribute dimensions.

and alternatives remain fixed. Variation in the number of attributes and alternatives occurs across commuters.

Table 2 The Attribute Profiles for the Entire Design

(units = %)	Base range			Wider range			Narrower range		
Levels:	2	3	4	2	3	4	2	3	4
Free flow time	± 20	-20, 0, +20	-20,-10, +10,+20	-20, +40	-20,+10,+40	-20, 0,+20,+40	± 5	-5, 0,+5	-5, -2.5, +2.5, +5
Slow down time	± 40	-40, 0, +40	-40,-20, +20,+40	-30, +60	-30,+15,+60	-30, 0,+30,+60	± 20	-20, 0, +20	-20, -2.5, +2.5, +20
Stop/start time	± 40	-40, 0, +40	-40,-20, +20,+40	-30, +60	-30,+15,+60	-30, 0,+30,+60	± 20	-20, 0, +20	-20, -2.5, +2.5, +20
Slow down-stop/start time	± 40	-40, 0, +40	-40,-20, +20,+40	-30, +60	-30,+15,+60	-30, 0,+30,+60	± 20	-20, 0, +20	-20, -2.5, +2.5, +20
Total travel time	± 40	-40, 0, +40	-40,-20, +20,+40	-30, +60	-30,+15,+60	-30, 0,+30,+60	± 20	-20, 0, +20	-20, -2.5, +2.5, +20
Uncertainty of travel time	± 40	-40, 0, +40	-40,-20, +20,+40	-30, +60	-30,+15,+60	-30, 0,+30,+60	± 20	-20, 0, +20	-20, -2.5, +2.5, +20
Running costs	± 20	-20, 0, +20	-20,-10, +10,+20	-20, +40	-20,+10,+40	-20, 0,+20,+40	± 5	-5, 0,+5	-5, -2.5, +2.5, +5
Toll costs	± 20	-20, 0, +20	-20,-10, +10,+20	-20, +40	-20,+10,+40	-20, 0,+20,+40	± 5	-5, 0,+5	-5, -2.5, +2.5, +5
Total costs	± 20	-20, 0, +20	-20,-10, +10,+20	-20, +40	-20,+10,+40	-20, 0,+20,+40	± 5	-5, 0,+5	-5, -2.5, +2.5, +5

The design dimensions are translated into SC screens as illustrated in Figure 1.

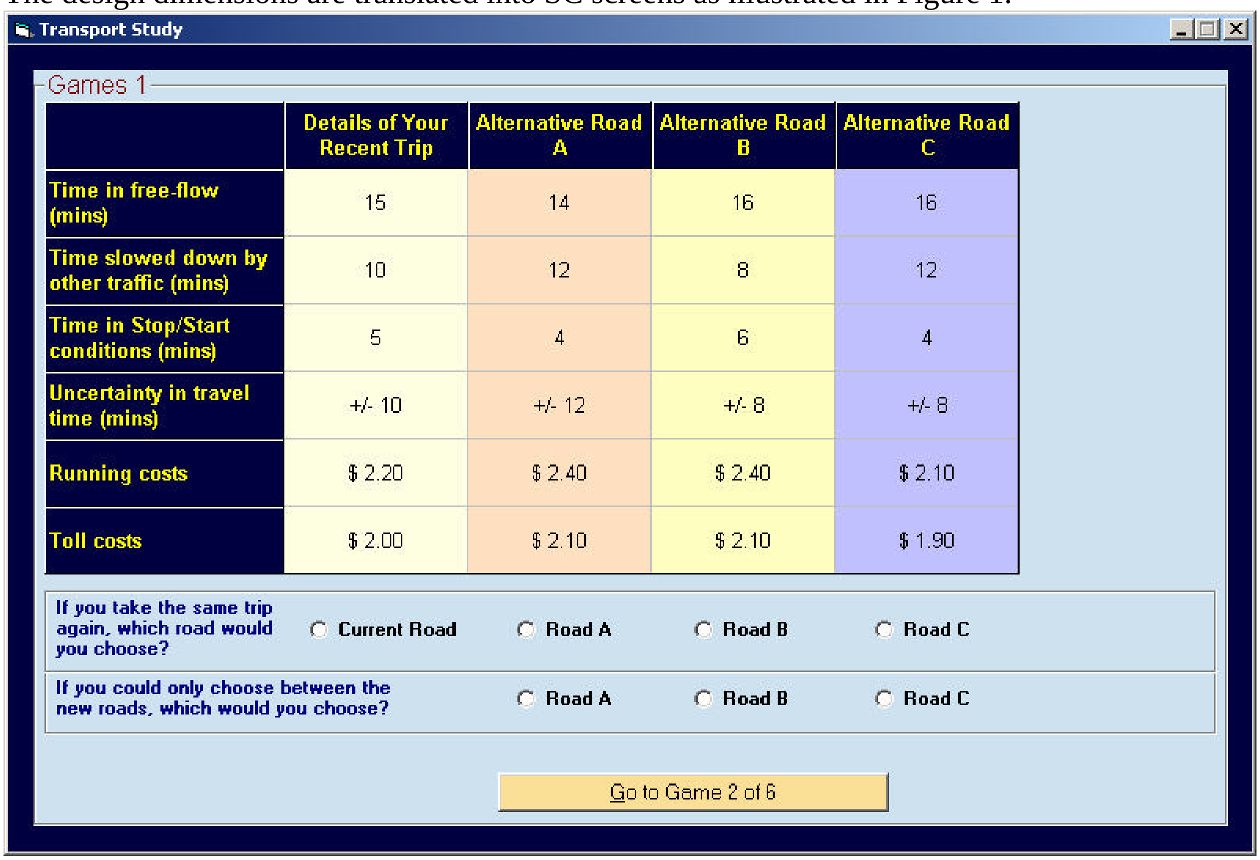

Figure 1 An example of a stated choice screen

3. Overview of the Sample Data

514 face-to-face computer-aided personal interviews (CAPI) (using nine interviewers, with two observed by a supervisor) were undertaken in the Sydney metropolitan area between 19 October and 23 November 2002.[6] 419 of the 514 surveys were useable. 12 rejected surveys were abandoned during the data collection phase due to errors made by interviewers (who entered the data incorrectly onto the CAPI on behalf of the commuter). The balance of the 83 that were eliminated had problems due to interviewers using the same respondent ID[7] and/or unacceptable data on the current commuter trip that produced questionable SC responses.[8] 138 telephone validations were attempted and 60 were completed with 100% assurance of the completed CAPI. A summary of the call analysis is given in Hensher (in press).

Sampling was stratified random according to geographical location of the household. Screening questions established eligibility in respect of commuting by car. Quotas were imposed for three trip lengths: less than 30 minutes (256), 30-60 minutes (190) and 60-90 minutes (60). The entire Sydney metropolitan area was covered. A summary of the mean and standard deviation for each attribute in each design is presented in Table 3.

Table 3 Descriptive Statistics on Design Attributes

(standard deviation in brackets)

Attribute	3 Attribute Design	4 Attribute Design	5 Attribute Design	6 Attribute Design
Free flow time		17.8 (2-72)	17.1 (0-63)	18.2 (2-84)
Slowed time			10.7 (0-64)	8.6 (0-70)
Stop/start time			8.9 (0-72)	10.5 (0-98)
Uncertainty time	19.1 (0-96)	17.7 (0-112)	19.3 (0-90)	19.5 (0-147)
Slowed/stop/start time		17.6 (0-84)		
Total time	37.7 (1-144)			
Running cost				2.2 (.2-8.2)
Toll cost				1.36 (0-7.8)
Total cost	3.0 (.1-16)	2.7 (.2-16.4)	2.7 (.2-14.6)	

[6] This was preceded by a pilot survey of 36 commuters, which was sufficiently large (after expansion of choice sets) to enable estimation of multinomial logit models to at least assess the parameter estimates in respect of sign and relative magnitude (on marginal effects).

[7] An unexpected problem was encountered after the survey was completed. Interviewers periodically copied their completed surveys onto zip diskettes or went back to the survey firm who downloaded the files from the laptops. When the interviewers returned to the field, two of them duplicated the same respondent id. As a consequence when we merged the choice data (one row per alternative per choice set) with the socio-economic and contextual data (1 row per respondent) to format automatically for choice modelling the duplicated id's caused a false mapping. This can be avoided in future studies by assigning a unique random number to each interview.

[8] In particular, some respondents gave travel times that were clearly not achievable given the origin, destination and hence trip distance. In future CAPI surveys we will build in a test for travel time given distance travelled such as a maximum speed.

4. Embedding Design Dimensionality in a Mixed Logit Framework

A number of model frameworks offer interesting ways of incorporating SC design dimensionality. These include latent class multinomial logit (LCML) (Greene and Hensher, 2003), mixed logit (ML) (McFadden and Train, 2000; Hensher and Greene; 2003; Train, 2003), and covariance heterogeneity (nested) logit (CHL). All specifications (to varying degrees), including multinomial logit (MNL), are capable of revealing preference heterogeneity attributable to variations in the dimensions of the SC design. We focus on mixed logit as the most general of the choice model specifications. Mixed logit is increasingly used to estimate choice models. There are a number of useful summaries of the method (such as Train (2003) and Hensher, Rose and Greene (2004)) and so we will limit the detail to a summary overview.

We assume that a sampled individual (q=1,…,Q) faces a choice among J alternatives in each of T choice situations. Individual q is assumed to consider the full set of offered alternatives in choice situation t and to choose the alternative with the highest utility. The utility associated with each alternative j as evaluated by each individual q in choice situation t, is represented in a discrete choice model by a utility expression of the general form in (1).

$$\begin{aligned} U_{jtq} &= \sum_{k=1}^{K} \beta_{qk} x_{jtqk} + \varepsilon_{jtq} \\ &= \mathrm{b}_q' \mathbf{x}_{jtq} + \varepsilon_{jtq} \end{aligned}, \qquad (1)$$

where $\mathbf{x}_{jtq}$ is the full vector of explanatory variables that are observed by the analyst, including attributes of the alternatives, socio-economic characteristics of the individual and descriptors of the decision context and choice task itself in choice situation t. The components b_q and ε_{jtq} are not observed by the analyst and are treated as stochastic influences.

Within the logit context we impose the familiar condition that ε_{jtq} is independent and identically distributed (IID) extreme value type 1 across individuals, alternatives and choice situations. The IID assumption is most restrictive in that it does not allow for the error components of different alternatives to be correlated. One way to take this into account is to introduce into the utility function through b_q additional stochastic elements that may be heteroskedastic and correlated across alternatives. Thus,

$$\mathrm{b}_q = \mathrm{b} + \mathrm{D}\mathbf{z}_q + \mathrm{G}\mathbf{v}_q = \mathrm{b} + \mathrm{D}\mathbf{z}_q + \mathrm{h}_q, \qquad (2)$$

or $\beta_{qk} = \beta_k + \mathrm{d}_k'\mathbf{z}_q + \eta_{qk}$ where η_{qk} is a random term whose distribution over individuals depends in general on underlying parameters, and $\mathbf{z}_q$ is observed data. ε_{jtq} is the same IID random term with zero mean that appeared in the model before. Note that since b_q may contain alternative specific constants, η_{qk} may also vary across choices and, in addition, may thus induce correlation across choices.

The Mixed Logit class of models assumes a general distribution for η_{qk} and an IID extreme value type 1 distribution for ε_{jtq}. That is, η_{qk} can take on different distributional forms such as normal, lognormal, uniform or triangular. Denote the joint density of [η_{q1}, η_{q2},..., η_{qK}] by $f(\mathrm{h}_q|\mathrm{W},\mathbf{z}_q)$ where the elements of W are the underlying parameters of the distribution of b_q (b,D,G) and $\mathbf{z}_q$ is observed data specific to the individual, such as sociodemographic characteristics and h_q denotes a vector of *K* random components in the set of utility functions in addition to the *J* random elements in e_{tq}.

For a given value of h_q, the *conditional* probability for choice *j* is logit, since the remaining error term is IID extreme value:

$$L_{jq}(\mathrm{b}_q|\mathbf{X}_q,\mathrm{h}_q) = \exp(\mathrm{b}_q\text{¢}\mathbf{x}_{jq}) / \textstyle\sum_j \exp(\mathrm{b}_q\text{¢}\mathbf{x}_{jq}). \tag{3}$$

Equation (3) is the simple multinomial logit model, but with the proviso that, for each sampled individual, we have additional information defined by h_q. The probability is conditional on h_q, that is, on $\mathbf{v}_q$ (and $\mathbf{z}_q$). This additional information influences the choice outcome. The *unconditional* choice probability is the expected value of the logit probability over all the possible values of b_q, weighted by the density of b_q. From (2), we see that this probability density is induced by the random component in the model for β_q (Hensher and Greene, 2003). Thus, the unconditional probability is

$$P_{jq}(\mathbf{X}_q,\mathbf{z}_q,\mathrm{W}) = \int_{\mathrm{b}_q} L_{jq}(\mathrm{b}_q \mid \mathbf{X}_q, \mathrm{h}_q)\, f(\mathrm{h}_q \mid \mathbf{z}_q, \mathrm{W})\, d\mathrm{h}_q\,, \tag{4}$$

Where $\mathrm{b}_q = \mathrm{b} + \mathrm{D}\mathbf{z}_q + \mathrm{h}_q$. Thus, the *unconditional* probability that individual *q* will choose alternative *j* given the specific characteristics of their choice set and the underlying model parameters is equal to the expected value of the conditional probability as it ranges over the possible values of b_q. The random variation in b_q is induced by the random vector h_q, hence that is the variable of integration in (4).

Models of this form are called *mixed logit* because the choice probability P_{jq} is a mixture of logits with *f* as the mixing distribution. The probabilities will not exhibit the questionable independence from irrelevant alternatives property (IIA), and different substitution patterns may be obtained by appropriate specifications of *f*. This is handled through the random parameters, specifying each element of b_q associated with an attribute of an alternative as having both a mean and a standard deviation (i.e., it is treated as a random parameter instead of a fixed parameter). The standard deviation of an element of the b_q parameter vector, which we denote σ_k, accommodates the presence of unobservable preference heterogeneity in the sampled population (i.e., allows for individuals within the sampled population to have different b_q as opposed to a single b representing the entire sample population).

5. Mixed Logit Model Results

A series of mixed logit models were estimated to establish evidence on the role of design dimensionality (DD), with the final results presented in Table 5. Each attribute in Table 5 is associated with a subset of the 16 alternatives as defined in Table 4. In arriving at the final models we investigated a large number of ways of representing DD. These included interacting every attribute associated with each alternative with each design dimension (except the attribute per se), creating an overall choice complexity index (CCI) in which the ML utilities were regressed against all design dimensions to establish a parameterised CCI, and decomposing the mean parameters associated with random parameters by one or more design dimensions.

Table 4 Time-Defined attributes and Design Allocation

Design Identifier	Time components (excluding time variability which appears in every design)	Cost components	Number of attributes in design
2	Free flow, slowed down, stop-start	Total cost	5
3	Free flow, slowed down, stop-start	Total cost	5
10	Free flow, slowed down, stop-start	Total cost	5
15	Free flow, slowed down, stop-start	Total cost	5
6	Free flow, slowed down, stop-start	Run cost, toll cost	6
8	Free flow, slowed down, stop-start	Run cost, toll cost	6
9	Free flow, slowed down, stop-start	Run cost, toll cost	6
12	Free flow, slowed down, stop-start	Run cost, toll cost	6
0	Free flow, slowed down-stop-start	Total cost	4
1	Free flow, slowed down-stop-start	Total cost	4
11	Free flow, slowed down-stop-start	Total cost	4
14	Free flow, slowed down-stop-start	Total cost	4
4, 5,7,13	Total time	Total cost	3

All stand-alone travel time attributes were found to be highly statistically significant random parameters in all models as were the fixed parameter costs attributes. For the random parameters a triangular distribution was selected and constrained to ensure that the sign of the WTP for travel time savings was non-negative (see Hensher and Greene, 2003). The second mixed logit model (ML2) includes an additional set of variables defining the interaction between trip attributes and design dimensions; some of which have fixed parameters while others have a random parameter specification. These interactions were established from an extensive assessment of all possible *linear* interactions between a design dimension and a trip attribute. A constrained triangular distribution is also selected.

The preferred model involved the inclusion of each trip attribute as a stand alone effect plus an interaction of each trip attribute with design dimensions. This is more informative than a composite choice complexity index (which has another useful role - see below) and offers the best intuitive interpretation. However the number of statistically significant interactions between attributes of alternatives and design dimensions (i.e., heterogeneity around the trip attribute parameter distribution decomposed by design dimensionality) was very small, limited to six interactions out of the 54 investigated[9]. One interaction was statistically significant for free flow time (number of levels of the attribute), one for slowed down time (number of choice sets), one for stop/start time (narrow range of an attribute compared to base plus wide range), one for slowed/stop/start time (number of levels) and two for total travel time (number of alternatives and number of choice sets). Two of the interactions were best represented through random parameters (free flow time by number of levels and slowed down time by number of choice sets) giving a random parameter specification for both the stand-alone trip attribute and its interaction with design dimensionality.

The signs of the interaction parameters varied, with four positive signs and two negative signs; suggesting that the directional influence on the WTP of a specific trip attribute varies. In particular, the positive sign for free flow interacted with number of levels suggest that the WTP for savings in free flow time decreases as the number of levels of the free flow time increases in a design. Similarly, as the number of choice sets increase, the WTP for savings in slowed down time decreases, as does the WTP for slowed down plus stop/start time savings as the number of levels increases. In contrast, the WTP for stop/start time increases as the range of an attribute narrows, and the WTP for total time savings increases as the number of choice sets increases. Finally the WTP to save stop/start time increases as the number of levels increases. Although statistically significant design dependencies have been identified, the key issue is whether the influence on WTP is sufficiently large to merit close attention.

Table 5 Mixed Logit Choice Models with Design Dimension Contrasts (4,593 observations). Time is in minutes, cost is in dollars. (100 Halton draws) (t-statistics in parenthesis)

Attribute		Base Models		DD Models	
	Design identifier (see Table 5)	MNL	ML1	MNL2	ML2
Constant	5-20	-.5780 (-14.9)	-.7182 (16.3)		
Constant	1-5			.6341 (17.1)	.8099 (18.6)
*Travel time attributes as random parameters * :*					
Free flow time	2-4, 6-8, 10-12, 14-16, 18-20	-.1452 (-19.2)	-.1927 (-17.9)	-.2151 (-7.9)	-.5113 (-13.8)
Slowed time	3,4, 7,8, 11,12,15,16,19,20	-.1004 (-11.2)	-.1239 (-10.9)	-.1229 (-13.1)	-.1642 (-12.2)
Stop/start time	3,4, 7,8, 11,12,15,16,19,20	-.1314 (-14.1)	-.1654 (-13.3)	-.1229 (-13.1)	-.1642 (-12.2)
Uncertainty time	All	-.0166 (-4.9)	-.0313 (-6.4)	-.0182 (-5.4)	-.02522 (-6.1)
Slowed/stop/start time	2, 6,10,14,18	-.1444 (-18.6)	-.1839 (-16.5)	-.2777 (-8.0)	-.2818 (-7.7)
Total time	1,5,9,13,17	-.1734 (-22.9)	-.2349 (-18.5)	-.0326 (-.84)	-.1777 (-4.3)
Cost attributes as fixed parameters:					
Running cost	4,8,12,16,20	-.8023 (-7.7)	-.9518 (-7.8)	-.7002 (-6.6)	-.9352 (-7.2)
Toll cost	4,8,12,16,20	-1.617 (-24.0)	-1.9146 (-23.9)	-1.095 (-15.2)	-1.2569 (-15.3)
Total cost	1-3,5-7,9-11,13-15,17-19	-.9069 (-16.3)	-1.1307 (-16.1)	-.7817 (-14.1)	-1.0268 (-14.1)

[9] The evaluation of the full set of interactions involved an extensive amount of analysis. There were nine interactions between a design dimension and each trip attribute, with range being assessed as two design dimensions- narrow and wide.

*Interactions as random parameters *:*					
Free flow* # nlvls	6-8, 10-12, 14-16, 18-20			.02287 (2.7)	.09213 (9.2)
Slowed time * #choice sets	6-8, 10-12, 14-16, 18-20			.00206 (2.2)	.00286 (2.5)
Interactions as fixed parameters:					
Stop/start * # narrow range	6-8, 10-12, 14-16, 18-20			-.09815 (-2.7)	-.09426 (-2.4)
Slowed/stop/start time*nlvs	2, 6,10,14,18			.04461 (4.1)	.02893 (2.86)
Total time * # alts	5,9,13,17			.01796 (2.6)	.03212 (5.0)
Total time * # choice sets	5,9,13,17			-.02211 (-7.9_	-.01672 (-6.0)
Pseudo-R^2		.6564	.6612	.6624	.6669
Log-Likelihood		-4723.4	-4658.5	-4640.5	-4577.8

Notes: * indicates that ML1 and two random parameters have a triangular distribution in which the mean = standard deviation except for the last two variables in ML2 (see Hensher and Greene, 2003 for a justification).

6. Relationship Between WTP for Time Savings and Design Dimensionality

To investigate the potential influence of design dimensionality on the WTP for travel time savings, we derived the full distribution of values of travel time savings (VTTS) for each of the components of travel time associated with each design (for ML1 and ML2), and investigated the contribution of design dimensions. Given that each attribute does not appear in every design (see Table 4), we derived the distributions for subsets of designs. The defining segmentation criterion is the number of attributes and hence this does not appear in the regression models. We have derived the VTTS based on the running cost parameter where cost is decomposed into running and toll cost, and total cost where it is a single parameter[10].

The formulae used to produce the set of WTP for travel time distributions are of the general form:

VTTS (in dollars per hour) = 60×(<mean parameter of time> + <standard deviation of parameter of time>×<triangular distribution> + <mean parameter of interaction>×<interaction variable> + <standard deviation of parameter for an interaction>×<interaction variable> × <triangular distribution>)/<mean parameter of cost>. For example,

$$ML2\_ff = 60\times(-0.51132 + 0.51132\times T + 0.09213\times nlvls + 0.09213\times nlvls\times T)/-0.935$$

where T is a random draw from a triangular distribution.

[10] The specific relativities in the empirical between trip time components are strictly not comparable, which is of no import for the current study.

The results are summarised in Table 6 (except for MNL2). The MNL results are of passing interest, supporting the accumulating evidence from many studies that MNL models tend to overestimate the mean WTP, explained in large measure by the inability of MNL to separate out the dimensionality of influences to the same extent as more general models, with the mean confounding itself with variance. The contrast between ML1 and ML2 is particularly interesting, with no particular directional trend in the mean in the presence of accounting for design dimensionality. A test of the difference between the two WTP results (for large samples – see Hensher et. al. 2004) suggests that every WTP is statistically significant at the 95% confidence level on the Z-test (last column of Table 6). *Thus we have strong evidence that design specification does matter*. What is the extent of the influence given the common practice of selecting a mean WTP in most applications? We need to identify what differences in mean WTP occur as we vary design dimensionality in ML2, and how the analyst might take such evidence into account when comparing WTP estimates between designs. This is not a procedure for establishing the 'correct' mean WTP since this is unlikely to be known. The best we can do is to establish the increments of WTP when using different designs.

Table 6 WTP for travel time savings ($ per person hour). $AUD2002.Standard deviations are in brackets. We use running cost as cost denominator where appropriate and total cost for attributes denoted with an asterisk.

	Base Models		DD Models	
Attribute	MNL	ML1	ML2	**Z**
Free flow time	10.85	6.10 (3.5)	5.21 (12.2)	**4.752**
Slowed time	7.51	3.94 (2.25)	4.62 (2.5)	**-138.429**
Stop/start time	9.83	5.26 (3.01)	6.81 (4.7)	**-18.821**
Uncertainty time	1.25	0.99 (0.57)	1.07 (0.6)	**-6.551**
Slowed/stop/start time*	9.56	4.87 (2.81)	2.38 (4.1)	**33.950**
Total time*	11.45	6.22 (3.60)	7.42 (3.8)	**-15.536**

There are a number of possible ways of presenting the incremental influence on WTP of design dimensions. These include varying each DD, holding other DD's constant or selecting DD packages that 'represent' what might best be described as 'popular' design configurations. We have opted for the latter and selected 6 design DD packages, presented in Table 7, based on statistically significant design dimensions.

Table 7 Evaluative Design Dimension Packages

	# alternatives	# choice sets	# levels	Range (N,B,W)
DD-P1	2	6	2	Base
DD-P2	2	9	2	Base
DD-P3	3	12	3	Base
DD-P4	3	12	4	Base
DD-P5	4	15	4	Wider
DD-P6	4	15	4	Narrower

The results are summarised in Table 8. For each trip attribute we establish the numerical change in the mean WTP as the level of the relevant design dimensions is varied. There is clear evidence that the design dimensionality does have a non-marginal influence on the mean VTTS, however the influence is differential, being greatest (in absolute terms) for free flow time as a consequence of varying the number of levels from two to three to four, and least for slowed down time (or stop-start time). The mean VTTS approximately doubles (or halves) from DD-P1 through to DD-P6 for the other two attributes. These differences in mean VTTS will have a significant influence on the time benefits in transport projects. Before a reader rushes to a conclusion that this might throw stated choice methods into disarray, we must point out that such variations are not unique to SC data and are indeed common in revealed preference data. What we have in SC data however is a clearer statement on why these variations occur[11].

Table 8 Implications of DD packages on mean WTP ($/person hour). Standard deviations in brackets are of the assumed distribution of** b **and not of the estimate of mean** b**. Note: the number of items is the no. of alternatives * no. of attributes per alternative, the latter varying according to designs given in Table 4

	No. of items to process per choice set	Free flow time	Slowed down time*	Stop start time*	Slowed down plus stop start time	Total time
DD-P1	6 - 12	1.71 (11.1)	4.1 (3.1)	4.1 (3.1)	3.95 (3.8)	5.95 (2.5)
DD-P2	6 - 12	4. 28 (12.2)	3.5 (3.1)	3.5 (3.1)	3.95 (3.8)	8.35 (2.5)
DD-P3	9 - 18	4.28 (12.2)	3.02 (3.1)	3.02 (3.1)	2.57 (3.9)	9.22 (2.5)
DD-P4	9 - 18	10.11 (11.2)	3.02 (3.1)	3.02 (3.1)	1.19 (3.9)	9.22 (2.5)
DD-P5	12 - 24	10.11 (11.2)	2.47 (3.1)	2.47 (3.1)	1.19 (3.9)	10.07 (2.5)
DD-P6	12 - 24	10.11 (11.2)	8.52 (3.1)	8.52 (3.1)	1.19 (3.9)	10.07 (2.5)
DD impacts		*nlvls*	*chest, ntb*	*chest, ntb*	*nlvls*	*Nalts1, chset*

* The parameters for these two attributes are generic.

The suggestion that specific design dimensions have a differential influence on the WTP across the trip attributes is an important finding, complicating the search for a design strategy that is the most appealing. If we had found that the mean WTP did not vary across the DD packages (1-6) on all attributes, then we would have concluded that the selection of a design within the considered set is of no behavioural consequence. We cannot suggest this, but neither can we suggest a preferred design strategy on strictly statistical criteria. All we can say is that some design features do have a statistically significant influence on mean VTTS. The ones we have identified are:

- A narrower attribute range for slowed down and stop-start travel time increases the mean VTTS. This is consistent with earlier evidence in Hensher (in press).
- More attribute levels reduce the mean VTTS of slowed down plus stop-start time but increases the mean VTTS for free flow time.
- The number of choice sets appears to have a relatively small influence on mean VTTS for slowed down time and stop start time compared to the attribute range.
- Increasing both the number of alternatives per choices set and the number of choice sets increases the mean VTTS for total time; with the greatest increase associated with the number of choice sets.

[11] Hensher (2003a) has investigated, within an MNL setting, whether the mean, standard deviation, minimum and maximum levels of an attribute impact on the mean VTTS and indeed there is evidence that one or more of these data descriptors do systematically influence mean VTTS for slowed-down time, combined slowed down and stop-start time, and total time. Such relationship might also be expected from revealed preference data where different samples expose variations in such descriptors.

We can view designs P1 to P6 as increasing in 'complexity' in the strictly additive sense of the numbers of pieces of information to cognitively process (noting that P5 and P6 only differ on the range of attribute levels)[12]. What we observe at the mean is that VTTS for overall travel time increases when a design contains more individual items to process. However while this also applies to free flow time the opposite direction for mean VTTS is observed for the non-free flow components of travel time. This evidence still presents a challenge in deciding on how to choose an appropriate design in future studies. If we can show that attributes are aggregated (or even possibly ignored) then we might have an argument for assessing the variation in VTTS in terms of weighted and unweighted mean total travel time, ignoring the directional impact of specific components of travel time.

A series of supplementary questions were asked to establish the additivity of attributes in the processing of the choice sets. These questions sought out which set of attributes were added up as part of the assessment process The evidence summarised in Table 9 suggests that there is a substantial amount of aggregation of the travel time components in evaluation of the alternatives, with over 75% of the respondent's aggregating all travel time dimensions as part of the way they process the attribute information. What we appear to be seeing is a simplifying information processing strategy (IPS) where many respondents contrast aggregate time with aggregate cost. The information summarised in Table 9 provides the basis for exploring alternative IPS's.

Table 9 Summary of Attribute Role and Treatment of Additivity in Respondent's Processing of SC Screens (proportion of relevant observations). Blank cells mean not applicable.

	Proportion of sample who added up components of:	
Design	Time	Cost
0	.781	
1	.794	
2	.829	
3	.853	
4		
5		
6	.758	.636
7		
8	.839	.613
9	.871	.677
10	.793	
11	.900	
12	.800	.760
13		
14	.893	
15	.750	

Given the strong cognitive evidence for treating travel time as a single attribute, when we convert the components of VTTS into a weighted average VTTS (based on the mix of time components) we find that the mean VTTS increases from a low of $2.94 per person hour for DD-P1 through to $9.30 per person hour for DD-P6 (with intermediary values of $3.88,

[12] In ongoing research we are investigating ways in which task complexity might vary (in some cognitive processing sense) as a result of the correlation structure of the data and the mean and standard deviation of each attribute both within and between alternatives and choice sets.

\$3.95, \$6.46 and \$6.18[13]). This is the same directional impact as that for the total travel time. The weighted average values have an overall partial correlation with the total time VTTS of 0.73. This delivers clear evidence that the mean VTTS systematically increases as we increase the number of elements to process in the design. The simple linear OLS models for the two sets of evidence (t-values in brackets) are:

$$\text{Mean VTTS (weighted average)} = \underset{(-0.490)}{-1.1663} + \underset{(2.86)}{0.3710}\times\text{no. of items};\ r^2 = 0.671$$

$$\text{Mean VTTS (total time)} = \underset{(4.27)}{1.5880} + \underset{(5.41)}{0.10966}\times\text{no. of items};\ r^2 = 0.881.$$

The evidence on systematic and directional variation of mean VTTS associated with overall design dimensionality is encouraging to the extent that it signals some potential relationships that can be used (if reinforced by other studies) to adjust overall VTTS when making an appropriate comparison between different designs, but it is too early to claim generalisable findings. In future studies, what might be the criteria to identify the design space within which application values can be determined? Is it the revealed preference space?

To further investigate the directional influence of design complexity on VTTS we can make use of the more general test based on a global choice complexity index (CCI) introduced in Hensher (in press). CCI is based on the systematic relationship between design dimensionality and the (relative) utility associated with a specific alternative and individual[14]:

$$CCI = \underset{(-12.6)}{-4.789} \underset{(-3.7)}{-0.2099}\times nalts \underset{(-13.4)}{- 0.6411}\times natts \underset{(1.7)}{+0.0243}\times chest \underset{(-6.6)}{- 0.3629}\times nlvls \underset{(-9.5)}{-1.083}\times wtb \underset{(-11.1)}{- 1.234}\times ntb$$

(Adjusted r^2 = 0.0198, t-values in parenthesis) See Table 1 for definitions of explanatory variables.

After establishing the role of specific design dimensions on CCI, we specify a simplified model that aggregates travel time components (in line with what most respondents indicate they do in processing the choice alternatives) and add in an interaction between travel time and CCI. This allows the positive and negative effects of complexity to be averaged in a plausible way, with the resulting influence on VTTS providing a more general statistical test without taint of data mining in order to give confidence in the findings based on the limited set of significant interactions effects in Table 5. We expect a statistically significant negative parameter estimate for the interaction to enable the general test to support our finding of VTTS increasing as the SC design increases in complexity (as measured by the number of ‘levels’ associated with each dimension of the design). The parameter estimate

[13] The strictly increasing VTTS with task complexity as defined by the number of items to process is qualified by fluctuation associated with the range of attributes (with the wider range lowering the mean VTTS compared to the base and narrower range). The upward trend however remains statistically significant.

[14] Although we have used a CCI in Hensher (2003), a referee saw merit in using CCI within the simplified model in which all time attributes are aggregated, as a more general test of increasing VTTS as more items are included in the design dimensionality.

for CCI interacted with total travel time is -0.000047 (t-value of -3.2) which supports the earlier finding that aggregated VTTS increases as the complexity of a design increases.

The cognitive processes used to evaluate attribute trade-offs are complex, both in a revealed preference (RP) and a stated choice context. The difference between RP and SC is that the latter is a *constructed* setting whereas the former is a *reported (perceived)* setting influenced heavily by the elements of the choice context that the analyst seeks from the respondent (e.g. the specific attribute levels of the chosen and non-chosen alternatives). While we should not assume that RP VTTS is ‘correct’ (because if it is then one questions the basis for SC studies), if we can establish how much information is actually processed by individuals in making decisions in real choice markets (not necessarily captured in RP studies), especially the range of attribute levels and the rules for inclusion/exclusion and aggregation, then we have some contextual guidelines to assist in putting boundaries on the dimensionality of SC experiments. This representation of the space in which the behavioural VTTS most likely resides is an appealing role for real market data (in contrast to RP data per se), with SC data serving the important role of providing suitable (bounded) variability for establishing robust parameters for the inputs into WTP.

In the current study we have set out the framework within which future studies might proceed to establish the boundaries for real world applications. The SC design can then be constrained to satisfy the boundary conditions and designed to ensure that the bounded variability offered in the choice experiment is sufficient to deliver robust and efficient parameter estimates to derive behaviourally suitable VTTS.

7. Conclusions

This paper has taken a close look at the evidence on the influence of design dimensionality on the willingness to pay for travel time savings. What we are observing from a systematic assessment of the impact of stated choice experiment designs on key behavioural outputs, within a domain that stretches the main design dimensions over the limits of popular practice, is that design differences do have behavioural, and statistically significant, influences on the distribution of WTP for travel time savings (in its various states of disaggregation).

The evidence cannot be used to conclude that a specific design specification is preferred (especially that we should keep our designs to simple dimensionality such as DD-P1). Rather we have to recognise that WTP evidence will be influenced by the design implemented, with the current evidence suggesting higher mean WTP for total travel time savings for designs with more items to process. Researchers will have to use supplementary criteria to decide on what is an appropriate design dimensionality. These will include evidence from real markets on what information individuals process (such as the specific number of alternatives and the range of levels of attributes). There remains much to investigate.

Acknowledgements

The extensive comments of a referee and Steven Morrison are greatly appreciated as are suggestions from Juan de Dios Ortuzar, John Rose, Phil Goodwin, Tim Hau and Tony Fowkes.

References

Brazell, J.D. and Louviere, J.J. (1998) Length effects in conjoint choice experiments and surveys: an explanation based on cumulative cognitive burden, Department of Marketing, The University of Sydney, July (mimeo).

Brownstone, D. (2001) Discrete choice modelling for transportation, in Hensher, D.A. (ed.) *Travel Behaviour Research: The Leading Edge*, Pergamon Press, Oxford, 97-124.

Brownstone, D. and K. Train (1999) Forecasting new product penetration with flexible substitution patterns. *Journal of Econometrics*, 89, 109-129

Dellaert, B.G.C., Brazell, J.D. and Louviere, J.J. (1999) The effect of attribute variation on consumer choice consistency, *Marketing Letters*, 10, 139-147.

DeShazo, J.R. and G. Fermo (2001) Designing choice sets for stated preference methods: the effects of complexity on choice consistency, *Journal of Environmental Economics and Management*,

Greene, W.H. and Hensher, D.A. (2003) A latest class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research B*, 37, 681-698.

Hensher, D.A. (2001a) Measurement of the valuation of travel time savings, *Journal of Transport Economics and Policy* (Special Issue in Honour of Michael Beesley), 35 (1), 71-98.

Hensher, D.A. (2001b) The valuation of commuter travel time savings for car drivers in New Zealand: evaluating alternative model specifications, *Transportation*, 28 (2), 101-118.

Hensher, D.A. (2003) Revealing differences in value of travel time savings: An MNL Diagnostic for 16 Designs, *Institute of Transport Studies*, The University of Sydney, June.

Hensher, D.A. (2004) How do Respondents Handle Stated Choice Experiments? - Information processing strategies under varying information load, DoD Paper #5, Institute of Transport Studies, The University of Sydney, March.

Hensher, D.A. (in press) Revealing differences in willingness to pay due to the dimensionality of stated choice designs: an initial assessment, *Environmental and Resource Economics*.

Hensher, D.A. and Greene, W.H. (2003) Mixed logit models: state of practice, *Transportation*, 30 (2), 133-176.

Hensher, D.A., Louviere, J.J. and Swait, J. (1999) Combining sources of preference data, *Journal of Econometrics*, 89, 197-221.

Hensher, D.A., Rose, J. and Greene, W.H. (2004) *Applied Choice Analysis*: *A Primer* Cambridge University Press, Cambridge.

Kamakura, W.A., Byung-Do K. and Lee J. (1996) Modeling consumer choice processes: preference and structural heterogeneity, *Marketing Science,* 15 (2), 152-172.

Louviere, J.J. and Hensher, D.A. (2001) Combining Sources of Preference Data, in Hensher, D.A. (ed.) The Leading Edge of Travel Behaviour Research, Pergamon Press, Oxford, 125-144.

Louviere, J.J., Hensher, D.A. and Swait, J.F. (2000) *Stated Choice Methods and Analysis*, Cambridge University Press, Cambridge (Reprinted 2003).

Louviere, J., Carson, R., Ainslie, A., Cameron, T., DeShazo, J.R., Hensher, D., Kohn, R., Marley, T., and Street, D. (2002) Dissecting the random component of utility, Workshop Report for the Asilomar Invitational Choice Symposium, California, June *Marketing Letters*, 13 (3), 163-176.

McFadden, D. and Train K. (2000) Mixed MNL models for discrete response, *Journal of Applied Econometrics*, 15, 447-470.

Ohler, T., Li. A., Louviere, J. and J. Swait (2000) Attribute range effects in binary response tasks, *Marketing Letters*, 11, 3 249-260.

Swait J., and Adamowicz W. (1996) The effect of choice environment and task demands on consumer behaviour: discriminating between contribution and confusion, working paper, *Department of Rural Economy*, University of Alberta.

Swait, J. and Adamowicz, W. (2001) The influence of task complexity on consumer choice: a latent class model of decision strategy switching, *Journal of Consumer Research*, 28, 135-148.

Train, K. (2003), *Discrete Choice Methods with Simulation*, Cambridge University Press, Cambridge.

White, P.J., Johnson, R.D. and Louviere, J.J. (1998) The effect of attribute range and variance on weighted estimates, unpublished paper, *Department of Marketing*, The University of Sydney.