Constructing Efficient Stated Choice Experimental Designs

John M. Rose [1]

Michiel C.J. Bliemer [1, 2]

[1] The University of Sydney,
Faculty of Business and Economics,
Institute of Transport & Logistics Studies,
Newtown, NSW 2006, Australia
E-mail: johnr@itls.usyd.edu.au
michielb@itls.usyd.edu.au

[2] Delft University of Technology,
Faculty of Civil Engineering and Geosciences,
Department of Transport & Planning,
P.O. Box 5048, 2600 GA Delft, The Netherlands
E-mail: m.c.j.bliemer@tudelft.nl

Abstract

Stated choice experiments are often used in transportation studies for estimating and forecasting behavior of travelers, road authorities, etc. This kind of experiments rely on underlying experimental designs. Whilst orthogonal designs are mainstream for practitioners, many researchers now realize that so-called efficient designs are able to produce more efficient data in the sense that more reliable parameter estimates can be achieved with an equal or lower sample size. This paper describes several processes for generating stated choice experiments and is intended to give an overview of the current state-of-the-art. Different methods are described.

Keywords
Stated choice experiments, experimental design, orthogonal designs, efficient designs

1. Introduction

The purpose behind conducting experiments is to determine the independent influence of different variables (attributes or factors depending on the literature cited) on some observed outcome. In stated choice (SC) studies, this translates into the desire to determine the influence of the design attributes upon the choices that are observed to be made by sampled respondents undertaking the experiment. However, an acknowledged limitation of SC studies is that unless the number of person specific observations captured in a survey is extremely large, it is necessary to pool the responses obtained from multiple respondents in order to produce statistically reliable parameter estimates. As such, SC studies typically consist of numerous respondents being asked to complete a number of choice tasks in which they are asked to select one or more alternatives from amongst a finite set of alternatives. In each task, the alternatives, whether labeled or unlabeled[1], are typically defined on a number of different attribute dimensions, each of which are further described by pre-specified levels drawn from some underlying *experimental design*. The number of choice tasks each respondent is asked to undertake will generally be up to the total number of choice situations drawn from the experimental design. Consequently, an archetypal SC experiment might require choice data be collected on 200 respondents, each of whom are observed to make eight choices, thus producing a total of 1600 choice observations.

Exactly how analysts distribute the levels of the design attributes over the course of an experiment (which typically is via the underlying experimental design), may play a big part in whether or not an independent assessment of each attribute's contribution to the choices observed to have been made by sampled respondents can be determined. Further, the allocation of the attribute levels within the experimental design may also impact upon the statistical power of the experiment insofar as its ability to detect statistical relationships that may exist within the data. This ability is related to the sample size of the study and given a large enough sample, the statistical power of an experimental design may not matter. Nevertheless, for sample sizes more commonly used in practice, the ability to retrieve statistically significant parameter estimates may be severely compromised given the selection of a relatively *poor* design. What constitutes a poor design is the focus of this paper, however, at this stage it may be worth noting that there may exist a trade-off between the ability of a design to allow for an independent determination of the impact each design attribute has in a SC experiment (at least insofar as how independence is thought of in a traditional sense) and the ability of the design to

[1] Labeled choice experiments involve studies where the names of the alternatives on offer convey meaning to the respondents beyond the order in which they are shown to respondents (e.g., the alternatives may be labeled as car, bus and train). In unlabeled choice experiments, the names of the alternatives are only meaningful in so far as they relate the order of the alternative as shown to the respondent (e.g., Option A, Option B, etc.). In the later case, each alternative may actually represent a car or a bus or a train in terms of the attribute levels shown to the respondent, but the fact that the alternative resembles one of these modes is not explicitly stated to the respondent. An exception to this rule exists where the different alternatives are treated as an attribute in the experiment. Also, in many SC experiments, a type or brand of alternative is often mentioned in the scenario descriptor of the task. In such cases, all the alternatives represent different versions of the same type or brand (e.g., Option A, Option B, etc., represent different alternative buses).

detect statistically significant relationships. The experimental design chosen by the analyst therefore plays a critical role in SC studies.

Conceptually, an experimental design may be viewed as nothing more than a matrix of values that is used to determine what goes where in a SC survey. The values that populate the matrix represent the attribute levels that will be used in the SC survey, whereas the columns and rows of the matrix represent the choice situations, attributes and alternatives of the experiment. The actual layout of the design matrix is often set out in one of two ways. Some researchers set up the experimental design matrix such that each row represents a different choice situation and each column a different attribute within the experiment (see e.g., Bliemer and Rose 2006; Rose and Bliemer 2007). In such cases, groups of columns form different alternatives within each choice task. Other researchers however set-up the design matrix such that each row of the matrix represents an individual alternative and each column a different attribute (see e.g., Carlsson and Martinsson 2002; Huber and Zwerina 1996; Kanninen 2002; Kessels et al. 2006; Sándor and Wedel 2001, 2002). In these cases, multiple rows are grouped together to form individual choice situations. Independent of how the matrix is set out, the experimental design performs the same function; that being the allocation of attribute levels to choice tasks, as shown in Figure 1. We discuss later why researchers set up the experimental design matrices differently.

car		train	
Time	Cost	Time	Cost
20	5	30	4
25	6	25	5
25	5	30	3
…	…	…	…
…	…	…	…
…	…	…	…

→

	Time	Cost
car	20	5
train	30	4
car	25	6
train	25	5
car	25	5
train	30	3
car	…	…
train	…	…

→

	car	train
Travel time (mins)	20	30
Fuel costs / fare ($)	5	4
	car	train
Travel time (mins)	25	25
Fuel costs / fare ($)	6	5
	car	train
Travel time (mins)	25	30
Fuel costs / fare ($)	5	3

Figure 1: From experimental design to choice situation construction

Given the above, the primary question for those generating experimental designs for SC studies is simply ‘how best to allocate the attribute levels to the design matrix’. Traditionally, researchers have relied on the principle of orthogonality to populate the choice situations shown to respondents (see Louviere et al. 2000 for a review of orthogonal designs). The past decade, however, has seen fundamental changes in the methods employed to construct experimental designs underlying SC experiments. Previous widely held assumptions once thought central to the generation of SC experiments, such as the principle of orthogonality, have been shown to be largely inappropriate, resulting in new and innovative methods being developed to allocate the attribute levels to the design matrix (see e.g., Bliemer and Rose 2006; Carlsson and Martinsson 2002; Ferrini and Scarpa 2007; Huber and Zwerina 1996; Kanninen 2002; Kessels et al. 2006; Sándor and Wedel 2001, 2002, 2005).

In this paper, we set out to describe these new methods and demonstrate some of the problems with how SC experiments have been constructed in the past. In particular, we focus on the issue of orthogonality as a design principle for SC experiments and demonstrate its inappropriateness as a design criterion for generating SC designs for all but a limited number of cases.

2. Orthogonal Designs

The orthogonality of an experimental design relates to the correlation structure between the attributes of the design.[2] By forcing the design attributes (i.e., columns of the design matrix) to be orthogonal, the analyst ensures that the attributes of the experiment are statistically independent. As such, orthogonal experimental designs theoretically allow for an independent determination of each attributes influence upon the observed choices within SC data. It is important to note however, that the orthogonality (or otherwise) of a design suggests nothing about whether two or more attributes are cognitively correlated in the minds of the respondents (e.g., price and quality attributes). As such, orthogonality is purely a statistical property of the design and not a behavioral property imposed upon the experiment.

To generate an orthogonal design, it is often useful, but by no means necessary, to start by generating the full factorial design. A full factorial design is a design that contains choice situations spanning all possible attribute level combinations. As such, the full factorial design will result in the maximum number of choice situations that can be produced, not allowing for repeated choice situations. The ultimate size of a full factorial design in terms of the number of choice situations is dependent on the number of attributes and attributes levels required for the study (where attributes common across alternatives are considered as independent attributes). Mathematically, a full factorial design will produce $\prod_{k=1}^{K} L_k$ choice situations, where L_k is the number of levels assigned to attribute k (e.g., a design with five attributes, two described with two attribute levels, one with three level dimensions and two with four levels will produce a full factorial with $2\times2\times3\times4\times4 = 192$ choice situations).

Full factorial designs will be orthogonal not only in terms of the main effects of designs (i.e., the individual attributes) but also in terms of all interaction terms (i.e., the influence of two or more attribute columns multiplied together). Table 1 shows a full factorial design having three attributes with two levels each along side the correlation matrix for the same design.[3] From the correlation matrix, it can be seen that all main and interaction effects for the design are uncorrelated.

[2] In some cases, this definition of an orthogonal design may be relaxed to define orthogonality as occurring when all attribute correlations are zero within alternatives but not necessarily between alternatives; see Louviere *et al*. (2000) for a discussion on sequential versus simultaneous generation of orthogonal designs.
[3] The attribute level values used in the design are based on orthogonal coding in this example. More details on different types of coding can be found in Bliemer and Rose (2006).

Table 1: Full factorial design with correlation matrix

	Main effects			Interaction effects			
S	A	B	C	AB	AC	BC	ABC
1	-1	-1	-1	1	1	1	-1
2	1	-1	-1	-1	-1	1	1
3	-1	1	-1	-1	1	-1	1
4	-1	-1	1	1	-1	-1	1
5	1	1	-1	1	-1	-1	-1
6	1	-1	1	-1	1	-1	-1
7	-1	1	1	-1	-1	1	-1
8	1	1	1	1	1	1	1

Design correlation matrix

	A	*B*	*C*	*AB*	*AC*	*BC*	*ABC*
A	1						
B	0	1					
C	0	0	1				
AB	0	0	0	1			
AC	0	0	0	0	1		
BC	0	0	0	0	0	1	
ABC	0	0	0	0	0	0	1

In many cases, the full factorial design will produce too many choice situations for any one individual respondent to handle (e.g., 192 choice tasks would be regarded by most researchers as being too large a number for any one individual to have to complete in one survey). In such cases, the analyst may opt to take a subset of choice situations from the full factorial design and use these in construing the choice survey. Such designs are known as fractional factorial designs. There exist a number of different methods to generate fractional factorial designs which retain orthogonality in the attributes (see e.g., Galilea and Ortúzar 2005), however, by far the simplest method remains to first generate the full factorial design and then select subsets of choice tasks corresponding to the levels of higher order interaction terms[4]. This is shown in Table 2, where the fractions are generated from the ABC interaction term.

Table 2: Fractional factorial designs generated from a full factorial design

	Fraction 1						
	Main effects			Interaction effects			
S	A	B	C	AB	AC	BC	ABC
1	-1	-1	-1	1	1	1	-1
2	1	1	-1	1	-1	-1	-1
3	1	-1	1	-1	1	-1	-1
4	-1	1	1	-1	-1	1	-1

Correlation matrix for fraction 1

	A	*B*	*C*	*AB*	*AC*	*BC*
A	1					
B	0	1				
C	0	0	1			
AB	0	0	-1	1		
AC	0	-1	0	0	1	
BC	-1	0	0	0	0	1

	Fraction 2						
	Main effects			Interaction effects			
S	A	B	C	AB	AC	BC	ABC
5	1	-1	-1	-1	-1	1	1
6	-1	1	-1	-1	1	-1	1
7	-1	-1	1	1	-1	-1	1
8	1	1	1	1	1	1	1

Correlation matrix for fraction 2

	A	*B*	*C*	*AB*	*AC*	*BC*
A	1					
B	0	1				
C	0	0	1			
AB	0	0	1	1		
AC	0	1	0	0	1	
BC	1	0	0	0	0	1

As shown in Table 2, moving from a full factorial to an orthogonal fractional factorial design will result in some form of correlation structure being formed within the design.

[4] This holds only when the design is generated using effects coding (see e.g., Bliemer and Rose 2006).

As such, orthogonal fractional factorial designs can only be said to be truly orthogonal in terms of some of the effects of the design. The designs in Table 2 are only orthogonal in the main effects, as can be observed from the correlation matrix. Selecting different interaction terms to generate the fractional factorial designs will result in different correlation structures being observed. By selecting higher order interaction terms, the main effects of the fractional factorial designs generated will *generally* remain orthogonal to all other main effects, with only the interaction terms being confounded. Selecting lower order interaction terms to form the fractional factorial designs may result in some or all of the main effects being correlated, and is hence is commonly avoided where possible.

Independent of the confoundment structure induced within the design, the number of fractions produced will be equal to the number of values in the interaction term(s) selected to generate the fractional factorial designs. In the above example, the ABC interaction term took two values (i.e., -1 and 1), hence resulting in two orthogonal fractions being formed. Typically, once identified, the analyst would then select one of the fractions which would then be given to all of the respondents participating in the study. Alternatively, the analyst may allocate all of the fractions amongst the respondents so that the entire full factorial design is used, with subsets of respondents reviewing each fraction. Such a process is known as *blocking* of the design.

The use of interaction terms to create orthogonal fractional factorial designs can be repeated within already existing fractional factorial designs, such that fractions of fractional factorials are formed, thus producing either even smaller designs, or blocks within the fractional factorial designs. In this way, the analyst may maintain orthogonality (at least for some effects), whilst reducing the number of choice tasks that respondents are asked to review. For further discussion of how to generate orthogonal fractional factorial designs, see Louviere et al. (2000) or Bliemer and Rose (2006).

3. Reasons for Using Orthogonal Designs

Aside from the fact that orthogonal designs allow for an independent estimation of the influence of each design attribute on choice, two other reasons lie behind the common use of orthogonal designs in practice. The first reason is that they are generally easy to construct or obtain (either from software packages or academic papers), although only for a limited number of combinations of attribute levels. Secondly, the common use of orthogonal designs in SC studies is largely a result of historical impetus. In the past, the experimental design literature has been primarily concerned with linear models (such as linear regression models), where the orthogonality of data is considered important. In linear regression models, this is because (a) orthogonality ensures that the model will not suffer from multicollinearity, and (b) orthogonality is thought to minimize the variances of the parameter estimates, which are taken from the variance-covariance (VC) matrix of the model. The VC matrix of a linear regression model is given in equation (1).

$$VC = \sigma^2 \left[X'X \right]^{-1}, \qquad (1)$$

where σ^2 is the model variance, and X is the matrix of attribute levels in the design or in the data to be used in estimation. Fixing the model variance (which simply acts as a scaling factor), the elements of the VC matrix for linear regression models are minimized when the X matrix is orthogonal. A design that results in a model where the elements contained within the VC matrix are minimized is preferable, for two reasons. Firstly, such a design will produce the smallest possible standard errors (i.e., square roots of the variances), and hence maximize the *t*-ratios produced from that model and secondly, an orthogonal design (or data set) will produce zero-off diagonals in the models VC matrix, thus ensuring that the parameter estimates are unconfounded with one another (i.e., no multicollinearity).

As such, orthogonal designs, at least in relation to linear models, meet the two criteria for a good design mentioned in the introduction; they allow for an independent determination of each attributes contribution on the dependent variable and they maximize the power of the design to detect statistically significant relationships (i.e., maximize the *t*-ratios at any given sample size). The question however is whether for discrete choice models, do orthogonal designs produce the same properties? Before we address this question, we first discuss several problems that often occur in practice between the mapping of design orthogonality to data orthogonality.

4. Problems with Orthogonal Designs in Practice

The primary argument for using orthogonal fractional factorial designs is the ability of such designs to produce unconfounded estimates of the population parameters due to the enforced statistical independence between the attributes contained within the design. However, parameters are estimated from data sets underlined by SC experiments, not from the designs themselves. Unfortunately, only under exceptional circumstances will orthogonality be preserved within the data used to estimate discrete choice models, even if the experimental design used to construct the study is itself orthogonal. Indeed, with regards to choice data sets, one would expect orthogonality to be the exception rather than the rule (see however Lanscar et al. (2006) for an example where orthogonality has been transferred from the design through to the data). We offer three reasons for this statement.

Firstly, the principle of orthogonality as we have described it relates solely to the columns of the design matrix being uncorrelated with one another. In cases where respondents review the complete orthogonal matrix, this orthogonality will be preserved through to the data set. When respondents review subsets of the matrix however, problems can occur and orthogonality lost. If the subsets (or blocks) of the design are not replicated evenly over the survey and hence certain blocks are either over or under represented within the data, orthogonality will generally be lost. Simply put, one cannot (i) add or remove rows of the design and/or (ii) replicate unevenly rows of the design over multiple respondents, and retain orthogonality within the data set. Note that the removal of columns from the design will not impact on the orthogonality of the design however. Thus, the onus is on the researcher to ensure that in allocating the choice tasks to respondents, that each choice task is equally represented in the final data. In cases of non-response or where the number of respondents in the study does not allow for each block to be equally distributed over the sample, this may be difficult to achieve (this last point is often missed by the literature, as it has implications on sampling and sample sizes that is rarely, if ever, discussed).

Secondly, it is typical in many choice studies to collect data on non-design attributes such as socio-demographic and contextual variables. In such cases, unless some form of strict sampling is imposed, any covariates within the data set will unlikely be orthogonal, not only amongst themselves, but also with the design attributes. For example, if age, gender, and income are added as variables in some form of analysis, correlations are not only likely to exist for these variables, but given that the variables described are constant over all choice situations within individual respondents, correlations between these variables and other attributes of the design are also likely to exist.

Finally, enforcement of orthogonality as a design principle does not ensure against the production of behaviorally implausible choice situations within the survey. Often, analysts after generating a design will review the final survey and locate choice situations in which they believe the attribute level combination of a particular alternative in a choice situation are such that that alternative has a probability of one of being chosen (i.e., that alternative dominates all other alternatives on offer in terms of preference). In such cases, no information is gained in terms of the possible trade-offs between the attributes of the alternatives. In other cases, analysts may locate choice situations whereby certain combinations of attributes are formed which may not be plausible in reality and which thus detract from the realism of the choice tasks. In these cases, analyst typically reject the choice situations (i.e., delete that row or combination of rows of the design), thus ensuring that the design and data will no longer be orthogonal (for a discussion of the benefits and costs of such strategies, see e.g., Lanscar and Louviere 2006).

Given the above, a carefully determined orthogonal design is likely to produce non-orthogonal data in practice. As such, the question arises as to how important orthogonality is to stated choice experiments. In the next section efficient designs will be introduced, which seem to be outperforming orthogonal designs easily, although such designs have not been used much in practice yet.

4. Efficiency Choice Designs

A number of researchers have begun to query the appropriateness of orthogonal designs for use in SC studies. Generally, the argument against the use of orthogonal designs is that the property of orthogonality may run counter to many of the desirable properties of the econometric models typically used to analyze SC data (i.e., logit and probit models). Indeed, whilst orthogonality may be an important criterion to determine independent effects in linear models, discrete choice models are not linear (Train, 2003). Unlike linear models, the asymptotic variance-covariance (AVC) matrix of discrete choice models are derived by taking the negative inverse of the expected second derivatives of the log-likelihood function of the model (see e.g., McFadden 1974 or Bliemer and Rose 2006), and are not equivalent to Equation (1).

For example, consider the log-likelihood function of the MNL[5] model given as Equation (2).

$$L_N(X,Y,\beta)=\sum_{n=1}^{N}\sum_{s=1}^{S}\sum_{j=1}^{J} y_{jsn}\log P_{jsn}(X,\beta), \qquad (2)$$

where N represents the number of respondents, each facing S choice situations with J alternatives, $X=[x_{jksn}]$ represents the choice data (attribute levels) for each respondent, choice situation, alternative, and attribute k, β the parameter values, and y_{jsn} the outcomes of the survey such that y_{jsn} equals one if respondent n chooses alternative j in choice situation s and is zero otherwise. Given the attribute levels and the parameters, it is possible to calculate the choice probabilities, P_{jsn}. Let V_{jsn} represent the observed component of utility for alternative j in choice situation s and ε_{jsn} the unobserved (by the analyst) component. The utility U_{jsn} perceived in choice situation s by each respondent n is given by

$$U_{jsn}(X,\beta)=V_{jsn}(X,\beta)+\varepsilon_{jsn}, \qquad (3)$$

where $V_{jsn}(X,\beta)=\sum_{k=1}^{K_j}\beta_k x_{jksn}$ assuming linear additive utility.

Given that ε_{jsn} are unobserved by the analyst, it is necessary to make assumptions about how they are distributed across the population. Depending on the assumptions of ε_{jsn}, different model types result (e.g., MNL, NL, or ML model, see e.g., Train, 2003). The most well-known assumption is that all ε_{jsn} are independently and identically extreme value type I distributed, which yields the following MNL probabilities P_{jsn} for selecting alternative j in choice situation s (McFadden, 1974):

$$P_{jsn}(X,\beta)=\frac{\exp\left(V_{jsn}(X,\beta)\right)}{\sum_{i=1}^{J}\exp\left(V_{isn}(X,\beta)\right)}. \qquad (4)$$

Fortunately, the zero-one choice index, y_{jsn}, in Equation (2) *often* drops out when taking the inverse of the second derivatives of the log-likelihood functions of many of the models estimated on discrete choice data and hence is typically not a major concern for the analyst trying to derive the expected AVC matrix for a SC study[6]. The choice probabilities are a concern however, requiring knowledge not only of the design (i.e., X

---

[5] The Log-likelihood of other discrete choice models may be similarly written out, with the discussion that follows applying equally.

[6] For example, see Rose and Bliemer (2005) with regards to the MNL model and Bliemer et al. (2007) for the nested logit model. For the mixed logit model assuming independent choice situations, the choice index also is not required for the calculation of the expected AVC matrix, however it is required when the choice situations are considered not to be independent within respondents; see Bliemer and Rose (2007).

matrix) but also the parameter estimates (Equations (3) and (4)). Advanced knowledge of the parameter estimates would negate the necessity for conducting the study, however even partial knowledge may be of assistance (previous literature, pilot studies, or theory may assist in determining the priors; e.g., from theory we would expect in most instances that a price parameter would be negative). The assumption of the parameter priors allows for a better allocation of the attribute levels to the design given an expectation of the values contained within the AVC matrix for that design, without which the performance of the design once data has been collected cannot be easily determined.

Independent of where the parameter priors are derived from, two methods exist for generating the expected AVC matrix for SC designs. Firstly, the analyst may rely on Monte Carlo (MC) simulations. Here, the analyst first generates a design from which simulated choice data is constructed. Next, models are estimated on the simulated data. In doing so, the analyst may choose either to construct a single large sample and estimate a model based on this sample to obtain an estimate of the designs AVC matrix or alternatively construct many smaller samples and average the AVC matrices over many model estimations to obtain an expected AVC matrix for the design. In either case, the use of MC simulations may be time consuming to construct and analyze, particularly when econometrically complex models are involved. A second method to obtain the expected AVC matrix of a design is to analytically derive the AVC matrix by taking the second derivatives of the log-likelihood function of the model to be estimated. In models where the choice index, y_{jsn}, is no longer present after taking the inverse of the second derivatives of the log-likelihood function of the model to be estimated (e.g., in case of the MNL model), no estimation is required. The analyst simply generates the design assuming a single respondent[7], applies the expected parameter estimates to the design (i.e., priors) and calculates the expected AVC matrix for the design.

Once the expected AVC matrix is constructed for a design, it is necessary to determine the design's efficiency. Rather than work with each element within the expected AVC matrix of a design, many prefer to work with summary measures. A number of summary measures have been proposed within the literature, however the most predominately used measure appears to be the D-*error* statistic. The D-error statistic is calculated simply by taking the determinant of the AVC matrix assuming a single respondent, Ω_1, and scaling this value by the number of parameters, *K*. Minimizing the D-*error* statistic corresponds to minimizing, *on average*, the elements contained within the expected AVC matrix. Designs which minimize the D-*error* statistic are therefore called D-*efficient* designs.

Minor variations of the D-*error* statistics have been proposed within the literature. Early work in the area assumed all parameters were zero, resulting in D_z-*efficient* designs (see Equation 5). Subsequent work assumed non-zero priors that were known with certainty which were termed D_p-*efficient* designs (see Equation 6). More recently, researchers have begun to examine efficient designs where the true population parameters are not known

[7] This is possible as McFadden showed that the AVC matrices of discrete choice models are asymptotically divisible by *N*. Hence, if you construct the AVC matrix for a design assuming a single respondent, one can theoretically calculate the AVC matrix for that same design given any sample size, *N*, simply by dividing the elements in the matrix by *N*.

with certainty. In such cases, prior parameter estimates are drawn from Bayesian parameter distributions (with parameters θ). As such, these designs are known as Bayesian or D_b-*efficient* designs (see Equation 7).

$$D_z - error = \det\left(\Omega_1\left(X,0\right)\right)^{1/K}, \qquad (5)$$

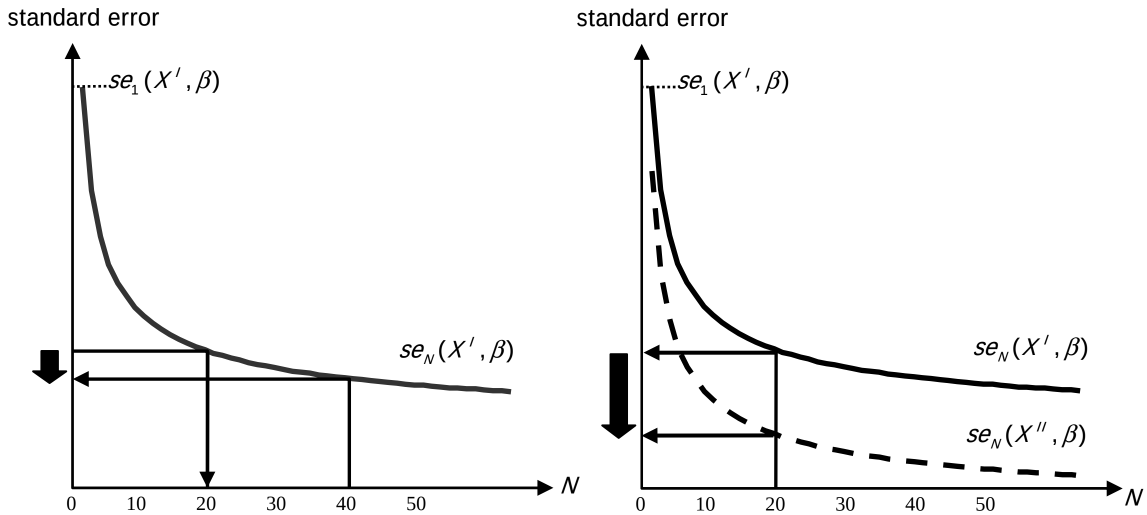
$$D_p - error = \det\left(\Omega_1\left(X,\beta\right)\right)^{1/K}, \qquad (6)$$

$$D_b - error = \int_{\tilde{\beta}} \det\left(\Omega_1\left(X,\tilde{\beta}\right)\right)^{1/K} \phi\left(\tilde{\beta} \mid \theta\right) d\tilde{\beta}. \qquad (7)$$

As noted, alternative measures have been proposed within the literature, such as *A*-error (for a summary of these and other proposed measures, see e.g., Ferrini and Scarpa 2007, Kessels et al. 2006, Rose and Scarpa 2007).

Independent of the efficiency measure used, attempts to minimize the elements of the expected AVC matrix for a design will result in the expected asymptotic standard errors of the design also being minimized (i.e., the square roots of the diagonals of the AVC matrix). If such a minimization translates from the design to models estimated on SC data collected from surveys using the designs, the analyst simultaneously achieves three goals. Firstly, for a given sample size, the smaller the asymptotic standard errors achieved, the smaller the width of the confidence intervals observed around the parameters estimates will be. Secondly, by minimizing the asymptotic standard errors, the asymptotic *t*-ratios for each of the parameters will be maximized, thus producing more reliable study results. Thirdly, the analyst is able to minimize the sample size, *N*, and still maintain a minimum value for the asymptotic *t*-ratios. As shown by McFadden (1974), the AVC matrices of discrete choice models are divisible by *N*, and as such, the asymptotic standard errors are divisible by the square root of *N*. Thus, diminishing returns in terms of the statistical significance of the parameter estimates are observed from each additional respondent added to a survey.

Figure 1(a) reveals the consequences of investing in larger sample sizes for a given design X^{I}. Whilst initial gains can be achieved in terms of improvements to the expected asymptotic standard errors achieved from models estimated based on the design from adding more respondents, such improvements occur at a diminishing rate until each additional respondent added will have only a marginal impact on the expected asymptotic standard errors. Hence, increasing the sample size beyond a certain limit will typically have little impact upon the statistical significance of the parameter estimates achieved from SC studies. Figure 1(b) reveals the impact for a given set of population parameters of investing in a better design X^{II} (i.e., more efficient design). Typically, larger decreases in the standard error can be achieved by investing in finding a more efficient design than by investing in a larger samples.



(a) investing in larger samples ***(b) investing in better design***

Figure 1: Comparison of investing in larger sample sizes versus more efficient designs

5. Optimal Orthogonal Choice Designs

A second stream of researchers has independently continued to examine orthogonal designs for use in SC studies. These researchers have identified an alternative optimality criteria to that used in generating efficient SC designs. As well as maintaining orthogonality, these researchers suggest that SC experiments should be constructed such that attributes common across alternatives should never take the same level over the experiment (see e.g., Burgers and Street, 2005; Street and Burgess, 2004; Street et al., 2001, 2005). Such designs are known as D-*optimal* designs. The argument for using this approach is that respondents are forced to trade on all attributes in the experiment, whilst the orthogonality of the design ensures that independent influence each attribute has upon choice can be determined. Optimality under this definition differs from that of D-*efficient* designs, in that D-*optimal* designs attempt to maximize attribute level differences whereas D-*efficient* designs attempt to minimize the elements that are likely to be contained within the AVC matrices of models estimated from data collected using the design. As such, a D-*optimal* design need not be optimal in terms of the criteria set out for D-*efficient* designs, with the opposite also being true. Indeed, the two optimality criteria are likely to be incompatible with one another for all but a small number of cases. Note that for constructing D-optimal designs, no prior parameters are used (i.e.., we assume the priors are all zeros), as one concentrates on the attribute level differences, hence efficiency will be lost in practice since the parameters are typically not equal to zero.

Optimal orthogonal choice designs suffer from a number of issues which has not been widely discussed within the literature. Firstly, these designs may only be constructed for unlabeled SC experiments. Labeled choice experiments where attributes may not be common across alternatives, or where attribute levels may differ for common attributes are not possible for such designs, as such designs are not covered by the definition of optimality offered. Secondly, these designs may promote certain forms of behavioral response, such as lexicographic choice behavior. By forcing each attribute to be different across alternatives, a particularly dominant attribute level may govern the entire experiment[8].

6. Optimal Choice Probability Designs

Simply put, any constraint one places on a design will impact upon the overall efficiency of that design. Orthogonality represents one such constraint. A second constraint often imposed on designs is attribute level balance. Attribute level balance occurs when each level of an attribute is forced to occur an equal number of times in the design. This constraint is imposed so that each point in preference space (represented by the attribute levels) is covered an equal number of times. The attribute level balance constraint is often imposed on efficient designs, although this need not be the case. Typically, when this constraint is relaxed, a minimum number of times each level must appear is imposed, otherwise the levels of the design will tend to all go to the extremes of the attribute level range, thus not allowing for tests of non-linearity in preference.

Kanninen (2002) has offered a compromise which we call *optimal choice probability* designs. In these designs, K-1 attribute levels are first generated for each J alternatives, typically using an orthogonal or optimal orthogonal approach. The last K^{th} attribute for each alternative is then generated as a continuous variable. The values of these continuous variables are chosen such that the choice probabilities take certain values that minimize the elements of the AVC matrix under the assumption of non-zero prior parameters. Toner et al. (1999) has analytically derived a similar rule for optimal choice probabilities. Kanninen (2002) and Johnson et al. (2006) have determined the desirable probabilities for a limited number of designs, although non-linear programming may be used to determine these for a wider number of designs.

One concern with these designs however is that they are optimal for the parameter priors assumed. Whilst this is true of other designs, the values that the continuous variables take are particular for the priors assumed, and may change markedly given different parameter estimates. Thus, Kanninen (2002) recommends a continual process of updating the design once data is collected and more likely parameter estimates obtained.

7. Case Studies

In this section, we construct four different designs, corresponding to the four methods discussed in Sections 2, 4, 5 and 6. For purposes of comparison, we limit the case study to designs with only two alternatives described by four attributes each with two attribute levels.

---

[8] For example, the authors once constructed a survey where the two alternatives represented different potential dates. One attribute in the experiment was that the potential date either had children or did not. Because the design required that one potential date always had children whilst the other did not, problems arose, particularly with younger respondents, who always selected the date without children. This occurred to the point where no information could be gained on the other attributes of the design.

7.1 Orthogonal Design

Consider the orthogonal fractional factorial design shown in Table 3. Assuming the design was used in an unlabeled SC study involving a single respondent and an MNL model with generic parameters estimated on the resulting data. Table 4 shows the expected AVC matrix for the design under two different parameter prior assumptions. Table 4(a) shows the expected AVC matrix of the design under the assumption that all the parameter estimates from the model are simultaneously equal to zero. Note that under this assumption the off-diagonals of the matrix (i.e., the covariances) are all zero. In Table 4(b), we construct the AVC matrix for the same design where we now assume that the parameter estimates to be non zeros (β_1 = -0.8; β_2 = -0.9; β_3 = 0.8; β_4 = -0.6). In this example, the off-diagonals of the AVC matrix are no longer equal to zero and the diagonal elements are much greater than when all the parameters were assumed to be zero. For the above example, the D_z-error (i.e., all parameters are zero) is 0.2981 and D_p-error is 0.7231.The two measures are not directly comparable as they were constructed under different parameter prior assumptions. Nevertheless, the values are comparable to other equivalent designs, which we construct in Section 7.2.

Table 3: Example Orthogonal Design

	Alternative A				Alternative B			
S	**X_1**	**X_2**	**X_3**	**X_4**	**X_1**	**X_2**	**X_3**	**X_4**
1	0	1	1	5	0	1	1	10
2	1	1	0	10	0	1	0	10
3	0	0	0	5	1	1	0	5
4	1	1	1	5	0	0	0	5
5	1	0	0	10	0	1	1	5
6	1	0	1	5	1	1	0	10
7	0	0	1	10	0	0	0	5
8	0	1	1	10	1	1	1	5
9	0	1	0	10	1	0	0	10
10	1	1	0	5	1	0	1	5
11	0	0	0	5	0	0	1	10
12	1	0	1	10	1	0	1	10

Table 4: AVC Matrices for Table 3 Design under different parameter assumptions

(a) All parameters equal zero

	X_1	**X_2**	**X_3**	**X_4**
X_1	**0.67**	0	0	0
X_2	0	**0.67**	0	0
X_3	0	0	**0.67**	0
X_4	0	0	0	**0.03**

(b) Non-zero parameter priors

	X_1	**X_2**	**X_3**	**X_4**
X_1	**1.19**	0.03	-0.37	0.08
X_2	0.03	**1.06**	-0.09	0.06
X_3	-0.37	-0.09	**2.08**	-0.26
X_4	0.08	0.06	-0.26	**0.15**

This simple example demonstrates an important concept in the generation of SC designs, and in particular the use orthogonal experimental designs for use in SC studies. Put simply, orthogonal designs are constructed with the sole criteria being that the correlations of the design attributes (and possibly the interaction terms) are all zero. The orthogonality of a design does not say anything about the expected AVC matrix of that design when one assumes non-zero priors.

7.2 Efficient Design

For purposes of comparison, we generate an efficient D_p-efficient design assuming the non-zero prior parameters used in the previous section. This design is shown in Table 5. Table 6 shows the AVC matrix for the design assuming all parameters are simultaneously zero as well as assuming the prior parameter estimates used in constructing the design.

Table 5: Example Orthogonal Design

	Alternative A				Alternative B			
S	**X_1**	**X_2**	**X_3**	**X_4**	**X_1**	**X_2**	**X_3**	**X_4**
1	1	1	1	5	0	0	0	5
2	0	1	0	5	1	0	1	5
3	0	0	1	10	1	1	0	5
4	0	1	1	10	1	0	0	5
5	1	0	0	10	0	1	1	10
6	1	1	0	5	0	0	1	10
7	1	0	1	5	0	1	0	5
8	0	0	0	10	1	1	1	10
9	1	0	0	5	0	1	1	5
10	0	1	1	10	1	0	0	10
11	1	0	1	10	0	1	0	10
12	0	1	0	5	1	0	1	10

Table 6: AVC Matrices for Table 5 Design under different parameter assumptions

(a) All parameters equal zero

	X_1	**X_2**	**X_3**	**X_4**
X_1	**0.53**	0.28	-0.20	0.12
X_2	0.28	**0.53**	-0.20	0.12
X_3	-0.20	-0.20	**0.60**	-0.16
X_4	0.12	0.12	-0.16	**0.10**

(b) Non-zero parameter priors

	X_1	**X_2**	**X_3**	**X_4**
X_1	**0.72**	0.40	-0.26	0.21
X_2	0.40	**0.76**	-0.29	0.23
X_3	-0.26	-0.29	**0.73**	-0.22
X_4	0.21	0.23	-0.22	**0.17**

The efficient design given in Table 5 produces D_z-error of 0.2515 (compared to 0.2981 for the orthogonal design) and D_p-error of 0.3447 (compared to 0.7231). For this example, the D-*efficient* design outperforms the orthogonal design under both parameter prior scenarios considered. Indeed, examination of the AVC matrix reveals that the efficient design produces lower variances than that of the orthogonal design for all but the last attribute. As such, the efficient design would require either a lower sample size (in this example, about half the sample size under the assumption of non-zero priors) to retrieve asymptotically significant parameter estimates for these attributes, or for a fixed sample size, produce larger asymptotic *t*-ratios.

7.3 Optimal Orthogonal Design

The design shown in Table 7 was generated using the principles outlined in Street et al. (2005). To construct the design, an orthogonal design was generated for the first alternative which was then used to generate the second alternative. The design generation process ensures that the attribute levels within each alternative are orthogonal but are correlated between alternatives. Applying the same prior parameter scenarios as used in

Sections 7.1 and 7.2, the AVC matrix for the design can be estimated. The AVC matrix for these two scenarios is given in Table 8.

Under the assumption of zero parameter priors, the design performs better than both the orthogonal and efficient design, producing a D_z-error of 0.1491. However, when the parameter estimates are not zero, the design efficiency increases to 0.6443, 4.3 times greater than that of zero-prior scenario and much larger than the D_p-error of the efficiency design (0.3447). For the equivalent scenarios, the ratio between the D_p-error and the D_z-error for the orthogonal and efficient designs are 2.43 and 1.37 respectively. Whilst these ratios occurred off of a larger base, the optimal orthogonal design appears to have lost efficiency at a greater rate than the other designs, a phenomenon that has been witnessed to have occurred on other occasions.

Table 7: Example Orthogonal Design

	Alternative A				Alternative B			
S	X_1	X_2	X_3	X_4	X_1	X_2	X_3	X_4
1	0	1	1	10	1	0	0	5
2	1	1	0	5	0	0	1	10
3	0	0	0	10	1	1	1	5
4	1	1	1	10	0	0	0	5
5	1	0	0	5	0	1	1	10
6	1	0	1	10	0	1	0	5
7	0	0	1	5	1	1	0	10
8	0	1	1	5	1	0	0	10
9	0	1	0	5	1	0	1	10
10	1	1	0	10	0	0	1	5
11	0	0	0	10	1	1	1	5
12	1	0	1	5	0	1	0	10

Table 8: AVC Matrices for Table 7 Design under different parameter assumptions

(a) All parameters equal zero

	X_1	X_2	X_3	X_4
X_1	**0.33**	0	0	0
X_2	0	**0.33**	0	0
X_3	0	0	**0.33**	0
X_4	0	0	0	**0.01**

(b) Non-zero parameter priors

	X_1	X_2	X_3	X_4
X_1	**1.60**	0.42	-0.43	0.22
X_2	0.42	**1.74**	-0.42	0.25
X_3	-0.43	-0.42	**1.60**	-0.22
X_4	0.22	0.25	-0.22	**0.11**

7.4 Optimal Choice Probability Design

In the final case study, we construct an optimal stated choice design. We begin by using the same design as that generated for the optimal orthogonal design, however we now allow the final attribute for each of the two alternatives to be continuous rather than drawn from fixed levels (this was done using Solver in Microsoft Excel). The final design is given in Table 9. The choice probabilities are 0.26 and 0.74 for the two alternatives (not necessarily in that order) in all twelve choice situations, consistent with the optimal choice probabilities computed by Johnson et al. (2006).

The same prior parameter scenarios used previously are applied to the design. Based on these, the AVC matrix for each scenario is generated, which we report in Table 10. The D_z-error for the design is 0.2527 whilst the D_p-error is 0.3280. These values are comparable to that of the D-efficient design, representing a marginal improvement when non-zero parameter estimates are assumed. Nevertheless, the diagonal elements for the AVC matrix of the design shown in Table 10(b) are far smaller under the non-zero parameter scenario, suggesting that the design would require either smaller sample sizes for statistical significance to be detectable or larger asymptotic *t*-ratios for a fixed sample size.

Table 9: Example Orthogonal Design

	Alternative A				Alternative B			
S	**X_1**	**X_2**	**X_3**	**X_4**	**X_1**	**X_2**	**X_3**	**X_4**
1	0	1	1	8.95	1	0	0	6.05
2	1	1	0	3.79	0	0	1	6.21
3	0	0	0	11.62	1	1	1	8.38
4	1	1	1	10.12	0	0	0	9.88
5	1	0	0	6.05	0	1	1	8.95
6	1	0	1	9.12	0	1	0	5.88
7	0	0	1	8.71	1	1	0	6.29
8	0	1	1	7.21	1	0	0	7.79
9	0	1	0	5.12	1	0	1	4.88
10	1	1	0	7.05	0	0	1	12.95
11	0	0	0	9.12	1	1	1	5.88
12	1	0	1	6.62	0	1	0	3.38

Table 10: AVC Matrices for Table 9 Design under different parameter assumptions

(a) All parameters equal zero

	X_1	**X_2**	**X_3**	**X_4**
X_1	**0.53**	0.22	-0.20	0.15
X_2	0.22	**0.58**	-0.22	0.17
X_3	-0.20	-0.22	**0.53**	-0.15
X_4	0.15	0.17	-0.15	**0.11**

(b) Non-zero parameter priors

	X_1	**X_2**	**X_3**	**X_4**
X_1	**0.69**	0.29	-0.25	0.19
X_2	0.29	**0.75**	-0.29	0.21
X_3	-0.25	-0.29	**0.69**	-0.19
X_4	0.19	0.21	-0.19	**0.14**

8. Conclusion

Given large enough samples, the underlying experimental design should not matter in terms of statistical power. Any design, even one with a poor or even random allocation of the attribute levels, will theoretically return statistically significant population parameters given a sufficiently large enough sample. Thus, the generation of efficient choice designs are primarily relevant to studies involving small finite samples. The aim of such designs is therefore to produce statistically significant parameter estimates given as small a sample as is possible.

In this paper, we have outlined four different methods for generating SC experiments. The first method, orthogonal designs, represents that used mainly within the literature. The predominant use of this method is as a result of historical impetuous, however as argued in this paper, may have unappealing impacts upon model results, particularly in smaller samples. The second method examined, efficient choice designs, represents an alternative approach to the use of orthogonal designs. Efficient SC designs are generated in such a way as to tie the design to the econometric form. These designs are constructed with the aim of minimizing the elements of the AVC matrix for the design, and hence maximize the asymptotic *t*-ratios for each of the parameter estimates. A number of issues surround the use of these designs. Firstly, the generation process requires the assumption of prior parameter estimates. Whilst these may come from prior research or from pilot studies, the requirement for making assumptions about the parameters has been somewhat controversial. The use of Bayesian prior parameter distributions has helped alleviate some these concerns, however many researchers still query the approach. Secondly, the AVC matrix for a design is specific to the econometric model assumed. Where the final model is unknown, it may be necessary to examine the AVC matrix for different model types and use some form of efficiency averaging over the models. Unfortunately, this has not yet been done in practice, hence research is required into the appropriateness of this suggestion.

The third method discussed herein was included for completeness only. The use of optimal orthogonal choice designs has been widely promoted in some literature (particularly marketing and health economics), however as shown in this paper, should only be considered in cases where the parameter estimates are expected all to be zero (in which case why do the study at all), or where no prior information on the parameter estimates is available. Furthermore, these designs can only be constructed for a limited range of models. Worryingly, despite the above and the other limitations of the method as discussed, the method remains prevalent in many literature.

The final method discussed, optimal choice probability designs, represents a new and exciting approach. Whilst theoretically, such a method will always produce the most optimal outcome for SC experiments in terms of sample size requirements and statistical efficiency, the current lack of software capable of generating such designs is a limiting factor. Further, far more research is required in terms of how to generate these designs for a wider class of problems, such a those requiring a labeled choice experiment.

References

Bliemer, M.C.J., Rose, J.M. and Hensher, D.A. (2007) Efficient stated choice experiments allowing for differences in error variances across subsets of alternatives. Submitted for publication to *Transportation Research B*.

Bliemer, M.C.J. and Rose, J.M. (2006) Designing Stated Choice Experiments: State-of-the-art, paper presented at the 11th International Conference on Travel Behaviour Research, Kyoto, Japan, August.

Burgess, L. and Street, D.J. (2005) Optimal designs for choice experiments with asymmetric attributes, *Journal of Statistical Planning and Inference*, 134, 288-301.

Carlsson, F. and Martinsson, P. (2002) Design Techniques for Stated Preference Methods in Health Economics. *Health Economics* **12**, 281-294.

Ferrini, S. and Scarpa, R. (2007) Designs with a-priori information for nonmarket valuation with choice-experiments: a Monte Carlo study, Submitted to *Journal of Environmental Economics and Management*, **53**, 342-363.

Galilea, P. and Ortúzar, J.deD.(2005) Valuing noise level reductions in a residential location context. *Transportation Research D* **10**, 305-322.

Huber, J. and Zwerina, K. (1996) The Importance of utility Balance and Efficient Choice Designs, *Journal of Marketing Research*, 33 (August), 307-317.

Johnson, F.R., Kanninen, B.J. and Bingham, M. (2006) Experimental Design For Stated Choice Studies, in Kanninen, B.J. (Ed.) *Valuing Environmental Amenities Using Stated Choice Studies: A Common Sense Approach to Theory and Practice*, Springer, the Netherlands, p159-202.

Kanninen, B.J. (2002) Optimal Design for Multinomial Choice Experiments, *Journal of Marketing Research*, 39 (May), 214-217.

Kessels, R., Goos, P. and Vandebroek, M. (2006) A comparison of criteria to design efficient choice experiments, *Journal of Marketing Research*, **43**(3), 409-419.

Lanscar, E. and Louviere J.J. (2006) Deleting 'irrational' responses from discrete choice experiments: a case of investigating or imposing preferences?, *Health Economics*, **15**, 797–811.

Lanscar, E., Louviere, J.J. and Flynn, T. (2006) Several methods to investigate relative attribute impact in stated preference experiments, *Social Science and Medicine*, **64**(8), 1738-1753.

Louviere, J.J. and Hensher, D.A. and Swait, J.D. (2000) Stated Choice Methods: Analysis and Application, Cambridge University Press, Cambridge.

McFadden, D. (1974) Conditional logit analysis of qualitative choice behaviour, In Zarembka, P. (ed.), *Frontiers in Econometrics*, Academic Press, New York, 105-142.

Rose, J.M. and Bliemer, M.C.J. (2005) Constructing Efficient Choice Experiments, ITLS working paper: ITLS-WP-05-07, April.

Rose, J.M. and Bliemer, M.C.J. (2006) Designing Efficient Data for Stated Choice: Accounting for Socio-demographic and contextual effects in designing stated choice experiments, paper presented at the 11th International Conference on Travel Behaviour Research, Kyoto, Japan, August.

Rose, J.M. and Bliemer, M.C.J. (2007) Stated Preference Experimental Design Strategies, in Hensher, D.A. and Button, K. (eds) *Handbook of Transport Modelling*, 2nd edition, 2007.

Rose, J.M. and Scarpa, R. (2007) Designs Efficiency for Non-market Valuation with Choice Modelling: How to Measure It, What to Report and Why, University Of Waikato: Working paper in Economics 21/07.

Sándor, Z. and Wedel, M. (2001) Designing Conjoint Choice Experiments Using Managers' Prior Beliefs, *Journal of Marketing Research*, **38** (November), 430-444.

Sándor, Z. and Wedel, M. (2002) Profile Construction in Experimental Choice Designs for Mixed Logit Models, *Marketing Science,* **21**(4), 455-475.

Sándor, Z. and Wedel, M. (2005) Heterogeneous Conjoint Choice Designs, *Journal of Marketing Research*, **42**, 210-218.

Street, D.J., Bunch, D.S. and Moore, B.J. (2001) Optimal designs for 2^k paired comparison experiments, *Communications in Statistics, Theory, and Methods*, 30, 2149-2171.

Street, D.J., and Burgess, L. (2004) Optimal and near-optimal pairs for the estimation of effects in 2-level choice experiments, *Journal of Statistical Planning and Inference*, 118, 185-199.

Street, D.J., Burgess, L., and Louviere, J.J. (2005) Quick and easy choice sets: Constructing optimal and nearly optimal stated choice experiments, *International Journal of Research in Marketing*, 22, 459-470.

Toner, J.P., Clark, S.D., Grant-Muller, S.M. and Fowkes, A.S. (1999) Anything you can do, we can do better: a provocative introduction to a new approach to stated preference design, *WCTR Proceedings*, **3**, Antwerp, 107-120.