

# Optimal Designs for Stated Choice Experiments that Incorporate Position Effects

Stephen Bush, Deborah Street and Leonie Burgess

Department of Mathematical Sciences

University of Technology Sydney

PO Box 123 Broadway NSW 2007, Australia

Tel.: +61-2-9514-2243

stephen.bush@uts.edu.au

Key Words: Paired comparisons; Multiple comparisons; Bradley–Terry model; Multinomial logit model; Davidson–Beaver position effects model.

## **Abstract**

Davidson and Beaver (1977) extended the Bradley–Terry model to incorporate the possible effect of position within a choice set on the choices made in paired comparisons experiments. In this paper we further extend the Davidson and Beaver result to choice sets of any size. Under a mild restriction we show that designs optimal for the multinomial logit model are still optimal when position effects are included in the model. We also show how designs balanced for carry-over effects of all orders can be used to construct designs with a diagonal information matrix for attribute effects. The theoretical results in this paper assume that we assume the null hypothesis of equal merits, but also discuss the consequences of unequal

merits using an example.

## 1 Introduction

Discrete choice experiments (DCEs) have been used in areas such as health economics, transportation, marketing, and public policy to model decision making behaviour. Louviere et al. (2000) and Train (2003) provide a comprehensive introduction to the area.

In a DCE we present a series of choice sets to each respondent. Each choice set contains a number of options from which the respondent is asked to choose the option that they think is best. We assume that each choice set contains the same number of options, and that each respondent is presented with the same series of choice sets.

One area that has not received much attention in the DCE literature is how to design for and model the structure of the options within a choice set. That is, how does the position of an option within a choice set affect the probability that the option is selected? Where this problem has been considered, the options have usually been labelled. Chrzan (1994) reports on three studies which between them investigate the importance of choice set order, order of items within choice sets and order of attributes within items. He concludes:

Choice set order (Study 1) influences attribute utilities but neither to a practically important extent nor in a predictable pattern. Attribute order (Study 3) influences utilities in choice-based conjoint analysis but, as for ratings-based conjoint analysis, in no predictable pattern. Profile [item] order (within choice sets) did not influence utilities for generic attributes in the “branded” profile toaster design used in Study 2 but produced statistically and practically significant (but unpatterned) effects for brands.

van der Waerden et al. (2006) also found that the order within the choice set was significant when running experiments with branded alternatives, while Wickelmaier and Choisel (2006) found that order was important for 7 of the 9 attributes that they investigated in a generic DCE, and always favoured the second position.

The most extensive investigations into position effects occur in the literature on paired comparisons experiments, where choice sets have two options each. According to David (1988), the possibility that the order of presentation might influence the selections made was raised by Fechner as early as 1860. The question of how to design to balance for possible position effects was considered by various authors whose work David summarises. David goes on to say that it appears to be sufficient to balance for position effects “unless the effects are large or are of interest in themselves” (p. 143). When the effects are of interest, Davidson and Beaver (1977) propose a modification of the Bradley–Terry model that incorporates the order that the options are presented within each pair.

Position effects are also used in other areas. DCEs formed from a number of paired comparisons are clearly related to tournaments, with “position” corresponding to playing “at home” or “away”. DCEs in which choice sets have  $m$  objects in them, and in which the relative position of the objects will be incorporated into the model, are closely related to block designs that are balanced for carry-over effects of all orders. Such designs were developed by Williams (1949) and Bugelski (1949) for animal feeding trials and modifications of these designs have also been used to design taste-testing experiments by Wakeling and MacFie (1995).

Position effects have also been considered in the context of questionnaires. Both question order and order of response categories within multiple choice questions have been established to influence the conclusions drawn (see Kalton et al. (1978) and Schuman et.al (1981) for example).

In this paper we aim to develop a model that incorporates position effects into DCEs with a fixed, but arbitrary, number of options in each choice set. In the next section we introduce some definitions and notation that we need. In Section 3 we introduce an extension to the multinomial logit model to incorporate position effects, based on the work in Davidson and Beaver (1977). In Section 4 we prove results that give optimal designs for the estimation of main effects of the attributes plus contrasts of the position effects when this extended model is used, and attributes may take any number of levels. We also use an example to

investigate the efficiency of the designs that are optimal under the null hypothesis of equal merits when the merits are unequal. In Section 5 we prove results that give optimal designs for the estimation of main effects plus two-factor interactions of the attributes and contrasts of the position effects when this model is used and all attributes are binary. In Section 6 we consider an alternative design approach based on the designs that are balanced for carry-over effects of all orders.

## 2 Definitions and Notation

In this section we introduce some concepts and notation that will be useful when discussing the design and analysis of DCEs with position effects. We begin by introducing some basic notation, then we introduce the multinomial logit model and the Davidson-Beaver position effects model. We conclude this section by discussing how we can modify the design properties discussed in Street and Burgess (2007) to accommodate position effects.

In a DCE we present a collection of  $N$  *choice sets* to each of the  $s$  respondents. We say that each choice set contains  $m$  *options*. For each option, we present an item that is described by  $k$  *attributes*. These attributes are properties of the item that we would like to test to see if they affect the selections made. Attribute  $q$  may take one of  $\ell_q$  *levels*, labelled by  $0, 1, \dots, \ell_q - 1$ . Then there are  $L = \prod_{q=1}^k \ell_q$  distinct items, each of which are described as a  $k$ -tuple of attribute levels.

If the DCE is set up as above, then we may use the multinomial logit model (MNL model) to estimate the attribute effects using the selections made by the respondents. Under the MNL model, the probability that item  $T_i$  is selected from the unordered choice set  $C = \{T_{i_1}, T_{i_2}, \dots, T_{i_m}\}$  is

$$P(T_i | \{T_{i_1}, T_{i_2}, \dots, T_{i_m}\}) = \frac{\pi_i}{\sum_{a=1}^m \pi_{i_a}},$$

where  $\pi_{i_a}$  is the *merit* of the item  $T_{i_a}$ . We are usually interested in estimating contrasts of the entries in  $\boldsymbol{\gamma} = \ln(\boldsymbol{\pi})$ , where  $\boldsymbol{\pi}$  contains the merits of each of the  $L$  items. We then express

the *utility* of an item  $T_i$  for respondent  $\alpha$  as  $U_{i\alpha} = \gamma_i + \epsilon_{i\alpha}$ . In choice sets of size  $m = 2$  the MNL model coincides with the Bradley–Terry model (Bradley and Terry (1952)).

The Davidson–Beaver model extends the Bradley–Terry model to incorporate position effects. In this model we multiply the merit of an item,  $\pi_i$  say, by a parameter  $\psi_a$  to incorporate the effect of the item being presented in position  $a$  of the choice set<sup>1</sup>. So the probability that item  $T_{i_1}$  is selected from the ordered choice set  $C = (T_{i_1}, T_{i_2})$  is

$$P(T_{i_1} | (T_{i_1}, T_{i_2})) = \frac{\psi_1 \pi_{i_1}}{\psi_1 \pi_{i_1} + \psi_2 \pi_{i_2}},$$

and the probability that  $T_{i_2}$  is selected from the ordered choice set  $C = (T_{i_1}, T_{i_2})$  is

$$P(T_{i_2} | (T_{i_1}, T_{i_2})) = \frac{\psi_2 \pi_{i_2}}{\psi_1 \pi_{i_1} + \psi_2 \pi_{i_2}}.$$

In this situation, we can express the utility of item  $T_i$  when presented in position  $a$  of the choice set, as  $U_{i\alpha} = \tau_a + \gamma_i + \epsilon_{i\alpha}$ , where  $\tau_a = \ln(\psi_a)$ . So the position effect acts as an additional effect in the model independently of the attribute effects.

To discuss designs where position is important we need to modify the way we describe the designs. For example, a choice set with items 1, 2 and 3 (in that order) will be different from the choice set with items 2, 3 and 1 (in that order). For the optimal designs described in Street and Burgess (2007), these two choice sets are equivalent, but if position effects are of interest they are no longer equivalent. So we need to extend the family of competing designs. We still use the  $D$ –optimality criterion to assess designs, so we are searching for the design that maximises the determinant of the Fisher information matrix.

Since the models used here are nonlinear in their parameters, we also need to specify a prior distribution for the parameters in order to compare designs. We assume a point prior distribution with all merits equal to 1. Other design criteria, and other priors, have been used when designing choice experiments; see Kessels et al. (2006) for a discussion. In

---

<sup>1</sup>In fact, Davidson and Beaver (1977) assume that  $\psi_1 = 1$  and  $\psi_2 = \psi$ , thereby reducing the number of position parameters to one. To make the generalisation more intuitive, we do not make this assumption and estimate contrasts of the position main effects instead.

Section 4, we look at an example and we find that the design which is optimal under the null hypothesis of equal merits is also optimal for some other values of  $\pi$ .

In this paper we will partition the set of all possible ordered choice sets of size  $m$  by using the set of differences between the items in the  $m$ -set. This difference vector generalises the difference vector introduced by Burgess and Street (2005) so that it contains not only the difference between the elements but also the location of that difference.

Consider the ordered  $m$ -set  $G = (\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_m)$ , where  $\mathbf{g}_a = (g_{a,1}, g_{a,2}, \dots, g_{a,k})$ . We can use this  $m$ -set to describe a choice set. In particular, if  $\mathbf{g}_1 = \mathbf{0}$ , we call  $G$  a *starter choice set*. To describe  $G$ , we define  $\mathbf{d}_{a,b}$  for each pair of positions  $a$  and  $b$  to be a vector of length  $k$  with a 0 in position  $q$  if  $g_{a,q} = g_{b,q}$  and a 1 in position  $q$  otherwise. We call  $\mathbf{d}_{a,b}$  a *difference*, and collect the differences for each pair of entries in  $G$  to form an *ordered difference vector*,  $\mathbf{v}_G = (\mathbf{d}_{1,2}, \mathbf{d}_{1,3}, \dots, \mathbf{d}_{m-1,m})$ .

The set of all possible ordered choice sets with  $m$  distinct items gives rise to several possible ordered difference vectors. We denote the set of these ordered difference vectors by  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_J\}$ , where there are  $J$  distinct ordered difference vectors in total. For the class of competing designs we assume that all choice sets with a particular ordered difference vector appear equally often in the experiment. The  $m$ -sets associated with a particular ordered difference vector can themselves be partitioned into sets such that all of the  $m$ -sets within a set of the partition can be written as the sum of a  $k$ -vector of levels and an ordered  $m$ -set with  $\mathbf{g}_1 = (00 \dots 0)$ . Since the elements in an  $m$ -set are ordered, this representation is unique. We let  $P_{\mathbf{v}_j}$  be the set of all starter choice sets with difference vector  $\mathbf{v}_j$ .

Thus our class of competing designs consists of all designs that are constructed from all of the starter choice sets in one or more  $P_{\mathbf{v}_j}$ . This is similar to the idea of difference families (or supplementary difference sets) used to construct block designs (see Abel (2006)) and is also closely related to the idea of a starter design to which are added elements from a set of generators, as described in Burgess and Street (2005). Here each set of generators in Burgess and Street (2005) corresponds to a starter choice set and the starting design is the complete factorial. We have chosen to change the focus of our discussion from starting designs to

starter choice sets since the order of the elements within each choice set is important when we include position effects in the model, and as yet we have no results about the behaviour of choice designs that arise from the addition of elements from a fractional design, even if it is regular and of known resolution.

Consider the  $L$  choice sets that arise from starter choice set  $G$ . The choice set with  $-\mathbf{g}_a$  in the first position of the choice set will be the only choice set with starter choice set  $G$  which has  $00\dots 0$  in position  $a$  of the choice set, since  $\mathbf{g}_a + \mathbf{x} = \mathbf{0}$  if and only if  $\mathbf{x} = -\mathbf{g}_a$ . It follows that, for each starter choice set,  $00\dots 0$  will appear in each position of the choice set once.

Finally, we define a series of constants that describe the choice experiment, as did Burgess and Street (2005). Let  $i_{\mathbf{v}_j}$  indicate whether or not all choice sets with ordered difference vector  $\mathbf{v}_j$  appear in the experiment. We also let  $c_{\mathbf{v}_j,a}$  be the number of choice sets containing the item  $00\dots 0$  in position  $a$  of the choice set and with ordered difference vector  $\mathbf{v}_j$ , and let  $x_{\mathbf{v}_j;\mathbf{d},a,b}$  be the number of times the difference  $\mathbf{d} = (d_1, \dots, d_k)$  appears as the difference between the items in positions  $a$  and  $b$  in the ordered difference vector  $\mathbf{v}_j$  (i.e.  $T_{i_a} + \mathbf{d} = T_{i_b}$ ). Finally, let  $y_{\mathbf{d},a,b}$  be the proportion of all choice sets that contain a particular pair with difference  $\mathbf{d}$  in positions  $a$  and  $b$  of the choice set, so

$$y_{\mathbf{d},a,b} = \frac{1}{N \prod_{q=1}^k (l_q - 1)^{d_q}} \sum_{\mathbf{v}_j} c_{\mathbf{v}_j,a} i_{\mathbf{v}_j} x_{\mathbf{v}_j;\mathbf{d},a,b}. \quad (1)$$

We illustrate this terminology in Example 1.

**Example 1.** Consider an experiment with two 2-level attributes, and with choice sets of size 3. An example of a possible design for such an experiment is given in Table 1. There are  $J = 6$  possible ordered difference vectors, which are shown in Table 2. The first entry in each difference vector is the difference between the first and second items in the choice set, the second entry is the difference between the first and third items in the choice set, and the third entry is the difference between the second and third items in the choice set.

The experiment in Table 1 contains all choice sets with ordered difference vector  $\mathbf{v}_1$  and no others. Therefore  $i_{\mathbf{v}_1} = 1$ , and  $i_{\mathbf{v}_j} = 0$  for all of the other difference vectors. The item  $00$

Option 1	Option 2	Option 3
0 0	0 1	1 0
0 1	0 0	1 1
1 0	1 1	0 0
1 1	1 0	0 1

Table 1: An example of a design with two 2-level attributes.

$\mathbf{v}_1$	(01, 10, 11)
$\mathbf{v}_2$	(01, 11, 10)
$\mathbf{v}_3$	(10, 01, 11)
$\mathbf{v}_4$	(10, 11, 01)
$\mathbf{v}_5$	(11, 01, 10)
$\mathbf{v}_6$	(11, 10, 01)

Table 2: Possible ordered difference vectors for the experiment in Example 1.



appears in each position once, and hence  $c_{\mathbf{v}_1,1} = 1$ ,  $c_{\mathbf{v}_1,2} = 1$ , and  $c_{\mathbf{v}_1,3} = 1$ . Since the choice sets with difference vector  $\mathbf{v}_1$  have difference (01) between positions 1 and 2 of the choice set,  $x_{\mathbf{v}_1;(01),1,2} = 1$ . None of the choice sets have difference (00), (10), or (11) between positions 1 and 2 of the choice set, so  $x_{\mathbf{v}_1;(00),1,2} = x_{\mathbf{v}_1;(10),1,2} = x_{\mathbf{v}_1;(11),1,2} = 0$ . Looking at the other pairs of positions, we have  $x_{\mathbf{v}_1;(10),1,3} = x_{\mathbf{v}_1;(11),2,3} = 1$ , and all other  $x_{\mathbf{v}_1;\mathbf{d},a,b} = 0$ . Since each pair with difference (01) appears as a difference between positions 1 and 2 of the choice set in exactly once choice set we have  $y_{(01),1,2} = \frac{1}{4}$ , as there are four choice sets in total. Similarly  $y_{(10),1,3} = \frac{1}{4}$  and  $y_{(11),2,3} = \frac{1}{4}$ . The remaining  $y_{\mathbf{d},a,b}$  terms are all equal to 0 since  $x_{\mathbf{v}_1;\mathbf{d},a,b} = 0$  in each case.  $\square$

### 3 The Generalised Davidson–Beaver Position Effects Model

In this section we consider a generalisation of the MNL model so that it accommodates position effects; choice sets can be of any fixed size. This generalisation is analogous to Davidson and Beaver’s generalisation of the Bradley–Terry model. We first set up the model and then give the information matrix for the estimation of the parameters in the model.

In the Davidson–Beaver position effects model we multiply the merit of the item in position  $a$  of the choice set by an effect,  $\psi_a$ , that incorporates the effect of position. For an arbitrary choice set size  $m$ , we define  $\psi_1, \psi_2, \dots, \psi_m$  to be the effect of an item appearing in positions 1, 2,  $\dots, m$  respectively on the probability of selection. We then multiply the merit of the item in position  $a$  of the choice set by  $\psi_a$  in the same way as the Davidson–Beaver position effects model. Then the probability of choosing an item  $T_i$ , which is presented in position  $a$  of the ordered choice set  $C = (T_{i_1}, T_{i_2}, \dots, T_{i_m})$ , so  $T_i = T_{i_a}$ , is

$$P(T_{i_a}|C) = \frac{\psi_a \pi_{i_a}}{\sum_{b=1}^m \psi_b \pi_{i_b}}.$$

To ensure identifiability we impose the constraint  $\prod_{a=1}^m \psi_a = 1$ . We call this model the *generalised Davidson–Beaver position effects model*. For respondent  $\alpha$ , the probability density

function for the response to the ordered choice set  $C = (T_{i_1}, T_{i_2}, \dots, T_{i_m})$  is

$$f_{C,\alpha}(\mathbf{w}_{C,\alpha}, \boldsymbol{\pi}, \boldsymbol{\psi}) = \frac{\prod_{a=1}^m (\psi_a \pi_{i_a})^{w_{i_a|C,\alpha}}}{\left(\sum_{b=1}^m \psi_b \pi_{i_b}\right)^{n_C}},$$

where  $w_{i_a|C,\alpha}$  is an indicator variable that equals 1 if the item in position  $a$  of the choice set is selected and 0 otherwise,  $\mathbf{w}_{C,\alpha}$  is a vector containing the  $w_{i_a|C,\alpha}$  terms for each item,  $n_C$  is the number of times choice set  $C$  appears in the experiment, and  $\boldsymbol{\psi} = (\psi_1, \psi_2, \dots, \psi_m)$ .

Following El-Helbawy et al. (1994), we let  $\Lambda(\boldsymbol{\pi}, \boldsymbol{\psi})$  be the information matrix for  $\sqrt{sN}\hat{\boldsymbol{\gamma}}$  and  $\sqrt{sN}\hat{\boldsymbol{\psi}}$ . Thus  $\Lambda(\boldsymbol{\pi}, \boldsymbol{\psi})$  contains minus the expected values of the second derivatives of the log-density function, where the differentiation is with respect to the entries in  $\boldsymbol{\gamma}$  and the entries in  $\boldsymbol{\psi}$ . Then we partition  $\Lambda(\boldsymbol{\pi}, \boldsymbol{\psi})$  into four blocks

$$\Lambda(\boldsymbol{\pi}, \boldsymbol{\psi}) = \begin{bmatrix} \Lambda_{\gamma\gamma}(\boldsymbol{\pi}, \boldsymbol{\psi}) & \Lambda_{\psi\gamma}(\boldsymbol{\pi}, \boldsymbol{\psi}) \\ \Lambda_{\gamma\psi}(\boldsymbol{\pi}, \boldsymbol{\psi}) & \Lambda_{\psi\psi}(\boldsymbol{\pi}, \boldsymbol{\psi}) \end{bmatrix}.$$

$\Lambda_{\gamma\gamma}(\boldsymbol{\pi}, \boldsymbol{\psi})$  is an  $L \times L$  matrix that contains minus the expected value of the second derivatives of the log-density function with respect to two entries in  $\boldsymbol{\gamma}$ .  $\Lambda_{\psi\psi}(\boldsymbol{\pi}, \boldsymbol{\psi})$  is an  $m \times m$  matrix that contains minus the expected value of the second derivatives of the log-density function with respect to two entries in  $\boldsymbol{\psi}$ .  $\Lambda_{\gamma\psi}(\boldsymbol{\pi}, \boldsymbol{\psi})$  and  $\Lambda_{\psi\gamma}(\boldsymbol{\pi}, \boldsymbol{\psi})$  contains minus the expected value of the second derivatives of the log-density function with respect to one entry in  $\boldsymbol{\gamma}$  and one entry in  $\boldsymbol{\psi}$ .

El-Helbawy and Bradley (1978) states that, under some mild regularity conditions, the  $(i, j)^{\text{th}}$  entry of the information matrix without position effects is

$$\Lambda(\boldsymbol{\pi})_{i,j} = \sum_C \frac{n_C}{N} \mathcal{E}_{\boldsymbol{\pi}} \left( \frac{\partial \ln(f_{C,\alpha}(\boldsymbol{\pi}, \mathbf{w}))}{\partial \pi_i} \frac{\partial \ln(f_{C,\alpha}(\boldsymbol{\pi}, \mathbf{w}))}{\partial \pi_j} \right) \pi_i \pi_j.$$

Then by differentiating the log-density function, and substituting the expectations, variances and covariances of the entries in  $\mathbf{w}_{C,\alpha}$ , we obtain

$$\Lambda_{\gamma\gamma}(\boldsymbol{\pi}, \boldsymbol{\psi})_{ij} = \sum_{C|T_i, T_j \in C} \frac{n_C}{N} \frac{-\psi_{a_i} \psi_{a_j} \pi_i \pi_j}{\left(\sum_{b=1}^m \psi_b \pi_{i_b}\right)^2},$$

$$\Lambda_{\gamma\gamma}(\boldsymbol{\pi}, \boldsymbol{\psi})_{ii} = \sum_{C|T_i \in C} \frac{n_C \psi_{a_i} \pi_i \left( \left( \sum_{b=1}^m \psi_b \pi_{i_b} \right) - \psi_{a_i} \pi_i \right)}{N \left( \sum_{b=1}^m \psi_b \pi_{i_b} \right)^2},$$

$$\Lambda_{\gamma\psi}(\boldsymbol{\pi}, \boldsymbol{\psi})_{ia} = \sum_{C|T_i \in C} \frac{n_C \pi_i \left( \frac{\delta_{T_i \text{ in pos } a} \left( \sum_{b \neq a} \psi_b \pi_{i_b} \right)}{\left( \sum_{b=1}^m \psi_b \pi_{i_b} \right)^2} - \frac{(1 - \delta_{T_i \text{ in pos } a}) \psi_{a_i} \pi_{i_a}}{\left( \sum_{b=1}^m \psi_b \pi_{i_b} \right)^2} \right),$$

$$\Lambda_{\psi\psi}(\boldsymbol{\pi}, \boldsymbol{\psi})_{a_1 a_2} = \sum_C \frac{n_C}{N} \frac{-\pi_{i_{a_1}} \pi_{i_{a_2}}}{\left( \sum_{b=1}^m \psi_b \pi_{i_b} \right)^2}, \text{ and}$$

$$\Lambda_{\psi\psi}(\boldsymbol{\pi}, \boldsymbol{\psi})_{aa} = \sum_C \frac{n_C}{N} \frac{\pi_{i_a} \left( \left( \sum_{b=1}^m \psi_b \pi_{i_b} \right) - \psi_a \right)}{\psi_a \left( \sum_{b=1}^m \psi_b \pi_{i_b} \right)^2},$$

where  $\psi_{a_i}$  is the position effect parameter for the position that  $T_i$  occupies in choice set  $C$ , and  $\delta_{T_i \text{ in pos } a}$  is an indicator variable that equals 1 if item  $T_i$  appears in position  $a$  of choice set  $C$  and is 0 otherwise.

If we assume, as did Davidson and Beaver (1977), the null hypothesis of equal merits for each of the items and that the entries in  $\boldsymbol{\psi}$  are left unspecified, then  $\Lambda(\boldsymbol{\pi}, \boldsymbol{\psi})$  simplifies. That is, if we assume that  $\boldsymbol{\pi} = \boldsymbol{j} = \boldsymbol{\pi}_0$ , where  $\boldsymbol{j}$  is a vector of 1s of length  $L$ , we obtain

$$\Lambda_{\gamma\gamma}(\boldsymbol{\pi}_0, \boldsymbol{\psi})_{ij} = -\frac{1}{\Psi_1} \sum_{a=1}^m \sum_{b \neq a}^m \psi_a \psi_b \lambda_{T_i \text{ in pos } a, T_j \text{ in pos } b},$$

$$\Lambda_{\gamma\gamma}(\boldsymbol{\pi}_0, \boldsymbol{\psi})_{ii} = \frac{1}{\Psi_1} \sum_{a=1}^m \psi_a \left( \sum_{b=1}^m \psi_b - \psi_a \right) \lambda_{T_i \text{ in pos } a},$$

$$\Lambda_{\gamma\psi}(\boldsymbol{\pi}_0, \boldsymbol{\psi})_{ia} = \frac{1}{\Psi_1} \sum_{b \neq a} \psi_b (\lambda_{T_i \text{ in pos } a} - \lambda_{T_i \text{ in pos } b}),$$

$$\Lambda_{\psi\psi}(\boldsymbol{\pi}_0, \boldsymbol{\psi})_{a_1 a_2} = -\frac{1}{\Psi_1}, \text{ and}$$

$$\Lambda_{\psi\psi}(\boldsymbol{\pi}_0, \boldsymbol{\psi})_{aa} = \frac{\left( \sum_{b=1}^m \psi_b \right) - \psi_a}{\psi_a \Psi_1},$$

where  $\Psi_1 = \left( \sum_{b=1}^m \psi_b \right)^2$ ,  $\lambda_{T_i \text{ in pos } a} = \frac{n_C}{N} \times \delta_{T_i \text{ in pos } a}$ , and  $\lambda_{T_i \text{ in pos } a, T_j \text{ in pos } b} = \lambda_{T_i \text{ in pos } a} \times \delta_{T_j \text{ in pos } b}$ . We notice that under the null hypothesis the entries in  $\Lambda_{\psi\psi}(\boldsymbol{\pi}_0, \boldsymbol{\psi})$  depend only on the entries in  $\boldsymbol{\psi}$ . Therefore  $\Lambda_{\psi\psi}(\boldsymbol{\pi}_0, \boldsymbol{\psi})$  is independent of the design, for a fixed choice set size.

## 4 Optimal Designs for Attribute Main Effects and Position Effects

In this section we prove results that give optimal designs when the generalised Davidson–Beaver position effects model is used. We show that, under a mild restriction, the optimal designs for the estimation of the main effects of the attributes using the MNL model are also optimal for the generalised Davidson–Beaver position effects model for the corresponding effects, under the null hypothesis of equal merits.

We are usually interested in the estimation of contrasts of the entries in  $\boldsymbol{\gamma}$  and  $\boldsymbol{\psi}$ , such as the attribute main effects. Thus we define  $B$  to be a matrix of contrast coefficients such that  $B(\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_L, \boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_m)^T$  are the effects that we are interested in estimating. In this paper we will not estimate any contrasts that involve both entries in  $\boldsymbol{\gamma}$  and entries in  $\boldsymbol{\psi}$ . Thus we have

$$B = \begin{bmatrix} B_\gamma & \mathbf{0} \\ \mathbf{0} & B_\psi \end{bmatrix},$$

where  $B_\gamma$  contains the coefficients of the contrasts of the entries in  $\boldsymbol{\gamma}$  and  $B_\psi$  contains the coefficients of the contrasts of the entries in  $\boldsymbol{\psi}$ . We let

$$B_\gamma = \begin{bmatrix} B_1 \\ \vdots \\ B_k \end{bmatrix}, \text{ where } B_q = \begin{bmatrix} \mathbf{b}_{q_1} \\ \vdots \\ \mathbf{b}_{q_{\ell_q-1}} \end{bmatrix},$$

and  $\mathbf{b}_{q_j}$  is a row vector that contains the contrast coefficients of the  $j^{\text{th}}$  contrast of the main effect of the  $q^{\text{th}}$  attribute. Let  $B_{q_j,x}$  be the entry in the  $j^{\text{th}}$  contrast for the main effect of attribute  $q$  corresponding to the  $x^{\text{th}}$  level of this attribute, and let  $B_{q_j,[i]}$  be the entry in the  $j^{\text{th}}$  contrast for the main effect of attribute  $q$  corresponding to the level of  $T_i$  for attribute  $q$ .

We let  $C(\boldsymbol{\pi}, \boldsymbol{\psi})$  be the information matrix for the estimation of the contrasts in  $B_\gamma \boldsymbol{\gamma}$  and  $B_\psi \boldsymbol{\psi}$ . From the definitions above,  $C(\boldsymbol{\pi}, \boldsymbol{\psi}) = B\Lambda(\boldsymbol{\pi}, \boldsymbol{\psi})B^T$ , and we partition  $C(\boldsymbol{\pi}, \boldsymbol{\psi})$  in the

same way as we partitioned  $\Lambda(\boldsymbol{\pi}, \boldsymbol{\psi})$  to obtain

$$C(\boldsymbol{\pi}, \boldsymbol{\psi}) = \begin{bmatrix} B_\gamma \Lambda_{\gamma\gamma}(\boldsymbol{\pi}, \boldsymbol{\psi}) B_\gamma^T & B_\gamma \Lambda_{\gamma\psi}(\boldsymbol{\pi}, \boldsymbol{\psi}) B_\psi^T \\ B_\psi \Lambda_{\psi\gamma}(\boldsymbol{\pi}, \boldsymbol{\psi}) B_\gamma^T & B_\psi \Lambda_{\psi\psi}(\boldsymbol{\pi}, \boldsymbol{\psi}) B_\psi^T \end{bmatrix} = \begin{bmatrix} C_{\gamma\gamma}(\boldsymbol{\pi}, \boldsymbol{\psi}) & C_{\gamma\psi}(\boldsymbol{\pi}, \boldsymbol{\psi}) \\ C_{\psi\gamma}(\boldsymbol{\pi}, \boldsymbol{\psi}) & C_{\psi\psi}(\boldsymbol{\pi}, \boldsymbol{\psi}) \end{bmatrix}.$$

In the next result we give a design constraint that allows the main effects of the attributes to be estimated independently of the contrasts of the position effects under the null hypothesis of equal merits.

**Lemma 1.** *Let  $B_\gamma$  be the main effects contrasts. Then under the null hypothesis of equal merits,  $C_{\gamma\psi}(\boldsymbol{\pi}, \boldsymbol{\psi}) = \mathbf{0}$  if each of the levels of each attribute appears in each position of the DCE equally often.*

*Proof.* We consider a generic term in the product of the first two matrices in the expression for  $C_{\gamma\psi}(\boldsymbol{\pi}, \boldsymbol{\psi})$ ,  $B_\gamma \Lambda_{\gamma\psi}(\boldsymbol{\pi}, \boldsymbol{\psi})$ . The rows of this matrix are labelled by the main effects of the attributes, and the columns are labelled by the positions in a choice set. Consider the entry of  $B_\gamma \Lambda_{\gamma\psi}(\boldsymbol{\pi}, \boldsymbol{\psi})$  corresponding to the  $j^{\text{th}}$  contrast for the main effect of the  $q^{\text{th}}$  attribute and position  $a$  of the choice set. We have

$$\begin{aligned} (B_\gamma \Lambda_{\gamma\psi}(\boldsymbol{\pi}_0, \boldsymbol{\psi}))_{q_j a} &= \frac{1}{\Psi_1} \sum_{i=1}^L \sum_{b \neq a} \psi_b B_{q_j, [i]} (\lambda_{T_i \text{ in pos } a} - \lambda_{T_i \text{ in pos } b}) \\ &= \frac{1}{\Psi_1} \sum_{x=0}^{\ell_q-1} \sum_{b \neq a} \psi_b B_{q_j, x} \left( \sum_{C|\text{att } q=x \text{ in pos } a} \lambda_C - \sum_{C|\text{att } q=x \text{ in pos } b} \lambda_C \right). \end{aligned}$$

Then  $(B_\gamma \Lambda_{\gamma\psi}(\boldsymbol{\pi}_0, \boldsymbol{\psi}))_{q_j a} = 0$  if  $\lambda_{\text{att } q=x \text{ in pos } a} - \lambda_{\text{att } q=x \text{ in pos } b} = 0$  for all attribute levels  $0 \leq x \leq \ell_q - 1$  and  $b \neq a$ . If this is the case for all attributes then  $B_\gamma \Lambda_{\gamma\psi}(\boldsymbol{\pi}, \boldsymbol{\psi}) = \mathbf{0}$ , and thus  $C_{\gamma\psi}(\boldsymbol{\pi}, \boldsymbol{\psi}) = \mathbf{0}$ .  $\square$

The next result expresses  $\Lambda_{\gamma\gamma}(\boldsymbol{\pi}_0, \boldsymbol{\psi})$  in terms of the ordered difference vectors introduced in Section 2. At this point it is also necessary to incorporate our knowledge about which pairs of items have a given difference so we define  $D_{\mathbf{d}}$  to be an  $L \times L$   $(0, 1)$  matrix with a 1 in position  $(i, j)$  if and only if items  $T_i$  and  $T_j$  have difference  $\mathbf{d}$ .

**Theorem 1.** Under the null hypothesis of equal merits,

$$\Lambda_{\gamma\gamma}(\boldsymbol{\pi}_0, \boldsymbol{\psi}) = \frac{\Psi_2}{\Psi_1} z I_L - \frac{1}{\Psi_1} \sum_{\mathbf{d}} \sum_{a=1}^m \sum_{b \neq a}^m \psi_a \psi_b y_{\mathbf{d},a,b} D_{\mathbf{d}},$$

where  $\Psi_2 = \sum_{a=1}^m \sum_{b \neq a}^m \psi_a \psi_b$ , and  $z = \frac{1}{N} \sum_{\mathbf{v}_j} c_{\mathbf{v}_j} i_{\mathbf{v}_j}$ .

*Proof.* We can write  $\lambda_{T_i}$  in pos  $a = \frac{1}{N} \sum_j c_{\mathbf{v}_j} i_{\mathbf{v}_j} = z$ . Hence

$$\Lambda_{\gamma\gamma}(\boldsymbol{\pi}_0, \boldsymbol{\psi})_{ii} = \frac{1}{N\Psi_1} \sum_{\mathbf{v}_j} \sum_{a=1}^m \left( c_{\mathbf{v}_j} i_{\mathbf{v}_j} \psi_a \left( \sum_{b=1}^m \psi_b - \psi_a \right) \right) = \frac{\Psi_2}{\Psi_1} \times z.$$

To express  $\Lambda_{\gamma\gamma}(\boldsymbol{\pi}, \boldsymbol{\psi})_{ij}$  in terms of the difference vectors used we first need to determine the number of choice sets that have  $T_i$  in position  $a$  and  $T_j$  in position  $b$ .

As each of the  $L$   $k$ -tuples is added in turn to the starter choice sets, there are  $Lc_{\mathbf{v}_j}$  possible choice sets with difference vector  $\mathbf{v}_j$ , and there are  $Lc_{\mathbf{v}_j} i_{\mathbf{v}_j}$  choice sets with difference vector  $\mathbf{v}_j$  in the experiment. Hence there are  $L \sum_{\mathbf{v}_j} c_{\mathbf{v}_j} i_{\mathbf{v}_j}$  choice sets in the experiment in total. It follows that the number of choice sets in the DCE with difference  $\mathbf{d}$  between positions  $a$  and  $b$  of the choice set is  $L \sum_{\mathbf{v}_j} c_{\mathbf{v}_j} i_{\mathbf{v}_j} x_{\mathbf{v}_j; \mathbf{d}, a, b}$ . Thus we need to determine the number of pairs of items with difference  $\mathbf{d}$ .

How many  $T_j$  exist with difference  $\mathbf{d}$  from  $T_i$ ? If  $d_q = 0$  then all such  $T_j$  have the same level for attribute  $q$  as  $T_i$  has. If  $d_q = 1$  then any such  $T_j$  must not have the same level for attribute  $q$  as  $T_i$  has. So there are  $\ell_q - 1$  possible entries in position  $q$  and so the number of items with difference  $\mathbf{d}$  from  $T_i$  is  $\Gamma_{\mathbf{d}} = \prod_{q=1}^k (\ell_q - 1)^{d_q}$ .

If items  $T_i$  and  $T_j$  have difference  $\mathbf{d}$ , the proportion of choice sets in the experiment that contain  $T_i$  in position  $a$  and  $T_j$  in position  $b$  is

$$y_{\mathbf{d}, a, b} = \frac{1}{N\Gamma_{\mathbf{d}}} \sum_{\mathbf{v}_j} c_{\mathbf{v}_j} i_{\mathbf{v}_j} x_{\mathbf{v}_j; \mathbf{d}, a, b}.$$

Hence the matrix containing the off-diagonal entries of  $\Lambda_{\gamma\gamma}(\boldsymbol{\pi}, \boldsymbol{\psi})$  is

$$-\frac{1}{\Psi_1} \sum_{a=1}^m \sum_{b \neq a}^m \psi_a \psi_b \sum_{\mathbf{d}} y_{\mathbf{d}, a, b} D_{\mathbf{d}}.$$

The result follows. □

We use this expression for  $\Lambda_{\gamma\gamma}(\boldsymbol{\pi}_0, \boldsymbol{\psi})$  to show that  $C_{\gamma\gamma}(\boldsymbol{\pi}_0, \boldsymbol{\psi})$  is block diagonal when main effects and position effects are of interest.

**Theorem 2.** *Let  $B_\gamma\boldsymbol{\gamma}$  be the main effects contrasts of the attribute effects. Then  $C_{\gamma\gamma}(\boldsymbol{\pi}_0, \boldsymbol{\psi})$  is block diagonal when the generalised Davidson–Beaver position effects model is used.*

*Proof.* Let  $P_{\ell_q, e_q}$  be an  $\ell_q \times \ell_q$  (0, 1) matrix with a 1 in position  $(t_1, t_2)$  if the difference between the two levels is  $t_2 - t_1 = e_q$ . Then  $P_{\ell_1, e_1} \otimes P_{\ell_2, e_2} \otimes \dots \otimes P_{\ell_k, e_k}$  will give the pairs that have  $T_2 - T_1 = (e_1, e_2, \dots, e_k)$ . Let  $\alpha_{\mathbf{e}, a, b}$  be the number of times  $\mathbf{e} = (e_1, e_2, \dots, e_k)$  appears as a difference between the items in positions  $a$  and  $b$  of the choice set. Then

$$C_{\gamma\gamma}(\boldsymbol{\pi}_0, \boldsymbol{\psi}) = \frac{1}{N\Psi_1} \left[ \left( \sum_{e_1} \dots \sum_{e_k} \sum_{a \neq b} \alpha_{\mathbf{e}, a, b} \psi_a \psi_b \right) B_\gamma \left( P_{\ell_1, 0} \otimes P_{\ell_2, 0} \otimes \dots \otimes P_{\ell_k, 0} \right) B_\gamma^T \right. \\ \left. - \sum_{e_1} \dots \sum_{e_k} \sum_{a \neq b} \alpha_{\mathbf{e}, a, b} \psi_a \psi_b B_\gamma \left( P_{\ell_1, e_1} \otimes P_{\ell_2, e_2} \otimes \dots \otimes P_{\ell_k, e_k} \right) B_\gamma^T \right].$$

However Corollary 6.4.1 of Street and Burgess (2007) shows that both

$$B_\gamma \left( P_{\ell_1, 0} \otimes P_{\ell_2, 0} \otimes \dots \otimes P_{\ell_k, 0} \right) B_\gamma^T \text{ and } B_\gamma \left( P_{\ell_1, e_1} \otimes P_{\ell_2, e_2} \otimes \dots \otimes P_{\ell_k, e_k} \right) B_\gamma^T$$

are block diagonal matrices, so  $C_{\gamma\gamma}(\boldsymbol{\pi}_0, \boldsymbol{\psi})$  is also block diagonal.  $\square$

This theorem allows us to consider only the block diagonal entries of  $C_{\gamma\gamma}(\boldsymbol{\pi}_0, \boldsymbol{\psi})$ , which correspond to the main effects for a single attribute. In addition, Lemma 1 states that if each of the levels of each attribute appear in each position of the DCE equally often then  $C_{\gamma\psi}(\boldsymbol{\pi}_0, \boldsymbol{\psi}) = \mathbf{0}$ , and therefore  $C(\boldsymbol{\pi}_0, \boldsymbol{\psi})$  is block diagonal.

The next theorem gives an expression for the block diagonal entry of  $C_{\gamma\gamma}(\boldsymbol{\pi}_0, \boldsymbol{\psi})$  which corresponds to the main effects of attribute  $q$ .

**Theorem 3.** *Under the null hypothesis of equal merits, the block diagonal entry of the information matrix corresponding to the main effect of attribute  $q$  is*

$$\frac{\ell_q}{N\Psi_1(\ell_q - 1)} \sum_{\mathbf{v}_j} c_{\mathbf{v}_j} i_{\mathbf{v}_j} \sum_{a \neq b} \psi_a \psi_b \sum_{\mathbf{d} | d_q = 1} x_{\mathbf{v}_j; \mathbf{d}, a, b} I_{\ell_q - 1}.$$

*Proof.* Since

$$B_q D_{\mathbf{d}} B_q^T = \frac{\Gamma_{\mathbf{d}}(-1)^{d_q}}{(\ell_q - 1)^{d_q}} I_{\ell_q - 1},$$

(Burgess and Street (2005)), the  $q^{\text{th}}$  block of the block diagonal matrix  $C_{\gamma\gamma}(\boldsymbol{\pi}_0, \boldsymbol{\psi})$  is given by

$$\begin{aligned} B_q \Lambda_{\gamma\gamma}(\boldsymbol{\pi}_0, \boldsymbol{\psi}) B_q^T &= B_q \left[ \frac{\Psi_2}{\Psi_1} z I_L - \frac{1}{\Psi_1} \sum_{\mathbf{d}} D_{\mathbf{d}} \sum_{a \neq b} y_{\mathbf{d}, a, b} \psi_a \psi_b \right] B_q^T \\ &= \frac{\Psi_2}{\Psi_1} z I_{\ell_q - 1} - \frac{1}{\Psi_1} \sum_{\mathbf{d}} \sum_{a \neq b} y_{\mathbf{d}, a, b} \psi_a \psi_b \frac{\Gamma_{\mathbf{d}}(-1)^{d_q}}{(\ell_q - 1)^{d_q}} I_{\ell_q - 1}. \end{aligned}$$

By substituting in the expressions for  $z$  and  $y_{\mathbf{d}, a, b}$ , we obtain

$$\begin{aligned} B_q \Lambda_{\gamma\gamma}(\boldsymbol{\pi}_0, \boldsymbol{\psi}) B_q^T &= \frac{1}{\Psi_1} \sum_{\mathbf{d}} \sum_{a \neq b} \psi_a \psi_b y_{\mathbf{d}, a, b} \frac{\Gamma_{\mathbf{d}}((\ell_q - 1)^{d_q} - (-1)^{d_q})}{(\ell_q - 1)^{d_q}} I_{\ell_q - 1} \\ &= \frac{\ell_q}{N \Psi_1 (\ell_q - 1)} \sum_j c_{\mathbf{v}_j} i_{\mathbf{v}_j} \sum_{a \neq b} \psi_a \psi_b \sum_{\mathbf{d} | d_q = 1} x_{\mathbf{v}_j; \mathbf{d}, a, b} I_{\ell_q - 1}, \end{aligned}$$

as required. □

Using the result in Theorem 3, the determinant of  $C(\boldsymbol{\pi}_0, \boldsymbol{\psi})$  is

$$\det(C(\boldsymbol{\pi}_0, \boldsymbol{\psi})) = \prod_{q=1}^k \left[ \frac{\ell_q}{N \Psi_1 (\ell_q - 1)} \sum_{\mathbf{v}_j} c_{\mathbf{v}_j} i_{\mathbf{v}_j} \sum_{a \neq b} \psi_a \psi_b \sum_{\mathbf{d} | d_q = 1} x_{\mathbf{v}_j; \mathbf{d}, a, b} \right]^{\ell_q - 1} \times \det(C_{\psi\psi}(\boldsymbol{\pi}_0, \boldsymbol{\psi})),$$

where  $\det(C_{\psi\psi}(\boldsymbol{\pi}_0, \boldsymbol{\psi}))$  depends on  $m$  but is independent of the design chosen.

We use this expression to extend the result in Theorem 1 of Burgess and Street (2005) to find the optimum value of  $\det(C(\boldsymbol{\pi}_0, \boldsymbol{\psi}))$  when the generalised Davidson–Beaver position effects model is used.

**Theorem 4.** *The  $D$ -optimal design for the estimation of main effects of the attributes and contrasts of the position effects is given by the set of choice sets where at least one difference vector  $\mathbf{v}_j$  has a non-zero  $i_{\mathbf{v}_j}$ , each pair of positions contains each non-zero difference equally*



often, and for each  $\mathbf{v}_j$  present, and for each attribute  $q$ , the sum of the differences is equal to

$$S_q = \begin{cases} (m^2 - 1)/4, & \ell_q = 2 \text{ and } m \text{ is odd,} \\ m^2/4, & \ell_q = 2 \text{ and } m \text{ is even,} \\ (m^2 - (\ell_q x^2 + 2xy + y))/2, & 2 < \ell_q < m, \\ m(m - 1)/2, & \ell_q \geq m, \end{cases}$$

where positive integers  $x$  and  $y$  satisfy the equation  $m = \ell_q x + y$  for  $0 \leq y < \ell_q - 1$ . The maximum possible value for the determinant of the information matrix is

$$\det(C(\boldsymbol{\pi}_0, \boldsymbol{\psi})_{\text{OPT}}) = \prod_{q=1}^k \left[ \frac{2S_q \ell_q \Psi_2}{Lm(m-1)\Psi_1(\ell_q-1)} \right]^{\ell_q-1} \times \det(C_{\boldsymbol{\psi}\boldsymbol{\psi}}(\boldsymbol{\pi}_0, \boldsymbol{\psi})).$$

*Proof.* To maximise  $\det(C(\boldsymbol{\pi}_0, \boldsymbol{\psi}))$ , we must maximise

$$\prod_{q=1}^k \left[ \frac{\ell_q}{N\Psi_1(\ell_q-1)} \sum_{\mathbf{v}_j} c_{\mathbf{v}_j} i_{\mathbf{v}_j} \sum_{a \neq b} \psi_a \psi_b \sum_{\mathbf{d}|d_q=1} x_{\mathbf{v}_j; \mathbf{d}, a, b} \right]^{\ell_q-1},$$

and so we must maximise

$$\frac{\ell_q}{N\Psi_1(\ell_q-1)} \sum_{\mathbf{v}_j} c_{\mathbf{v}_j} i_{\mathbf{v}_j} \sum_{a \neq b} \psi_a \psi_b \sum_{\mathbf{d}|d_q=1} x_{\mathbf{v}_j; \mathbf{d}, a, b},$$

for each  $q$ . Given our assumption that each pair of positions contains each non-zero difference equally often we obtain

$$\sum_{a=1}^m \sum_{b \neq a} \psi_a \psi_b \sum_{\mathbf{d}|d_q=1} x_{\mathbf{v}_j; \mathbf{d}, a, b} = \frac{2}{m(m-1)} \sum_{\mathbf{d}|d_q=1} x_{\mathbf{v}_j; \mathbf{d}} \times \Psi_2,$$

where  $\Psi_2$  is independent of the design used. By substitution, we obtain

$$\sum_{\mathbf{v}_j} c_{\mathbf{v}_j} i_{\mathbf{v}_j} \sum_{a \neq b} \psi_a \psi_b \sum_{\mathbf{d}|d_q=1} x_{\mathbf{v}_j; \mathbf{d}, a, b} = \frac{2\Psi_2}{m(m-1)} \sum_{\mathbf{v}_j} c_{\mathbf{v}_j} i_{\mathbf{v}_j} \sum_{\mathbf{d}|d_q=1} x_{\mathbf{v}_j; \mathbf{d}}.$$

Theorem 1 in Burgess and Street (2005) shows that  $\sum_{\mathbf{d}|d_q=1} x_{\mathbf{v}_j; \mathbf{d}}$  is maximised when it is equal to  $S_q$ . By observing this result, and that  $\sum_{\mathbf{v}_j} c_{\mathbf{v}_j} i_{\mathbf{v}_j} = \frac{N}{L}$ , we have

$$\frac{1}{N} \sum_{\mathbf{v}_j} c_{\mathbf{v}_j} i_{\mathbf{v}_j} \sum_{\mathbf{d}|d_q=1} x_{\mathbf{v}_j; \mathbf{d}} = \frac{S_q}{L},$$

and hence

$$\det(C_{\gamma\gamma}(\boldsymbol{\pi}_0, \boldsymbol{\psi})_{\text{OPT}}) = \prod_{q=1}^k \left[ \frac{2S_q \ell_q \Psi_2}{Lm(m-1)\Psi_1(\ell_q-1)} \right]^{\ell_q-1}.$$

Since all of the ordered choice sets with a particular difference vector appear in the DCE equally often, each of the levels for each attribute will appear in each position equally often, and hence  $C_{\gamma\psi}(\boldsymbol{\pi}_0, \boldsymbol{\psi}) = \mathbf{0}$  by Lemma 1. For a given  $m$ ,  $C_{\psi\psi}(\boldsymbol{\pi}_0, \boldsymbol{\psi})$  is constant across all designs, and thus

$$\det(C(\boldsymbol{\pi}_0, \boldsymbol{\psi})_{\text{OPT}}) = \prod_{q=1}^k \left[ \frac{2S_q \ell_q \Psi_2}{Lm(m-1)\Psi_1(\ell_q-1)} \right]^{\ell_q-1} \times \det(C_{\psi\psi}(\boldsymbol{\pi}_0, \boldsymbol{\psi})),$$

as required.  $\square$

The expression in Theorem 4 allows us to determine whether other designs are optimal for the estimation of the attribute main effects and contrasts of the position effects when using the generalised Davidson–Beaver position effects model. Since the number of choice sets obtained from this construction can be very large, the next result gives optimal designs with fewer choice sets. This characterisation is based on Theorem 3 of Burgess and Street (2005).

**Theorem 5.** *Consider the collection of starter choice sets  $G_f = \{\mathbf{g}_{f,1} = \mathbf{0}, \mathbf{g}_{f,2}, \dots, \mathbf{g}_{f,m}\}$ , for  $f = 1, \dots, \zeta$ , where  $\mathbf{g}_{f,i} \neq \mathbf{g}_{f,j}$  for  $i \neq j$ . Let  $\mathbf{g}_{f,i} = (g_{f,i,1}, g_{f,i,2}, \dots, g_{f,i,k})$ ,  $i = 1, \dots, m$ . Suppose that the multiset of differences for attribute  $q$  from positions  $a$  and  $b$ , which is  $\{\pm(g_{f,a,q} - g_{f,b,q}) \mid f = 1, \dots, \zeta\}$ , contains each non-zero difference modulo  $\ell_q$  equally often. Then the ordered choice sets obtained by adding each element of the complete factorial in turn to  $G_f$ , for  $f = 1, \dots, \zeta$ , are optimal for the estimation of main effects of the attributes and contrasts of the position effects, provided that there are as few zero differences as possible in each multiset.*

*Proof.* We know

$$B_q \Lambda_{\gamma\gamma}(\boldsymbol{\pi}_0, \boldsymbol{\psi}) B_q^T = \frac{\ell_q}{N \Psi_1(\ell_q - 1)} \sum_{\mathbf{v}_j} c_{\mathbf{v}_j} i_{\mathbf{v}_j} \sum_{a=1}^m \sum_{b \neq a} \sum_{\mathbf{d} \mid d_q=1} \psi_a \psi_b x_{\mathbf{v}_j; \mathbf{d}, a, b} I_{\ell_q-1},$$

from Theorem 3, substituting for  $y_{\mathbf{d},a,b}$ , and simplifying. Using the assumption that the multiset of differences for any two positions,  $a$  and  $b$  say, contains each non-zero difference equally often, we obtain

$$\sum_{a=1}^m \sum_{b \neq a} x_{\mathbf{v}_j; \mathbf{d},a,b} \psi_a \psi_b = \frac{2}{m(m-1)} \alpha_{q,a,b} \Psi_2.$$

Let  $G_f$  have difference vector  $\mathbf{v}_f = (\mathbf{d}_{f,1,2}, \mathbf{d}_{f,1,3}, \dots, \mathbf{d}_{f,m-1,m})$  with  $\mathbf{d}_{f,a,b} = (d_{f,a,b,1}, \dots, d_{f,a,b,k})$ . Then  $\alpha_{q,a,b} = \sum_{f=1}^{\zeta} d_{f,a,b,q}$ . Substituting for  $\alpha_{q,a,b}$  in  $B_q \Lambda_{\gamma\gamma}(\boldsymbol{\pi}_0, \boldsymbol{\psi}) B_q^T$ , and simplifying, gives

$$B_q \Lambda_{\gamma\gamma}(\boldsymbol{\pi}_0, \boldsymbol{\psi}) B_q^T = \frac{2\ell_q \alpha_{q,a,b} \Psi_2}{N \Psi_1 m(m-1)} I_{\ell_q-1}.$$

We see that the determinant of this block will be maximised when  $\alpha_{q,a,b}$  is maximised.

Each non-zero difference must occur in each pair of positions equally often. Thus, for attribute  $q$ , each of the  $\ell_q - 1$  non-zero differences appears as a difference  $\alpha_{q,a,b}$  times between positions  $a$  and  $b$  in the collection of starter choice sets. Since each item is added to each starter choice set in turn, there are  $L(\ell_q - 1)\alpha_{q,a,b}$  non-zero differences for attribute  $q$  in the choice experiment. Counting the differences in the  $N$  choice sets we have

$$L(\ell_q - 1)\alpha_{q,a,b} = S_q N,$$

and rearranging we get

$$\frac{2\ell_q \alpha_{q,a,b} \Psi_2}{N \Psi_1 m(m-1)} = \frac{2\ell_q S_q \Psi_2}{L \Psi_1 m(m-1)(\ell_q - 1)}.$$

The right-hand side of the equation above is the same as the entry corresponding to attribute  $q$  in the expression in Theorem 4. Thus this design is optimal for the estimation of main effects of the attributes and contrasts of the position effects.  $\square$

We now illustrate this theorem with an example.

**Example 2.** Consider an experiment with two 5-level attributes and with choice sets of size 3. Suppose that we construct a design using the starter choice sets (00, 11, 22) and (00, 22, 44). Then the multiset of differences for the first attribute from positions 1 and 2 is

$\{\pm 1, \pm 2\} \equiv \{1, 2, 3, 4\}$ . Each non-zero difference appears once. Similarly, it is straightforward to check that each non-zero difference appears once for each pair of positions for both of the attributes.

The information matrix for the estimation of main effects of the attributes plus contrasts of the position effects has determinant

$$\det(C(\boldsymbol{\pi}_0, \boldsymbol{\psi})) = \frac{(\psi_1\psi_2 + \psi_1\psi_3 + \psi_2\psi_3)^8}{300000000\psi_1\psi_2\psi_3(\psi_1 + \psi_2 + \psi_3)^{17}}.$$

We now compare the determinant above to the optimal determinant as given by Theorem 4. Since  $\ell_1 = \ell_2 = 5$  and  $m = 3$ , we have  $S_1 = S_2 = 3$ , and  $\det(C_{\psi\psi}(\boldsymbol{\pi}_0, \boldsymbol{\psi})) = 1/(3\psi_1\psi_2\psi_3(\psi_1 + \psi_2 + \psi_3))$ . Then

$$\begin{aligned} \prod_{q=1}^k \left( \frac{2S_q \ell_q \Psi_2}{Lm(m-1)\Psi_1(\ell_q-1)} \right)^{\ell_q-1} &= \prod_{q=1}^2 \left( \frac{2 \times 3 \times 5 \times \Psi_2}{25 \times 3 \times 2 \times \Psi_1 \times 4} \right)^4 \\ &= \frac{\Psi_2^8}{20^8 \Psi_1^8}. \end{aligned}$$

Then the optimum value of the determinant of the information matrix for the estimation of main effects of the attributes and contrasts of the position effects is

$$\begin{aligned} \det(C(\boldsymbol{\pi}_0, \boldsymbol{\psi})_{\text{OPT}}) &= \frac{\Psi_2^8}{20^8 \Psi_1^8} \times \frac{1}{3\psi_1\psi_2\psi_3(\psi_1 + \psi_2 + \psi_3)} \\ &= \frac{(\psi_1\psi_2 + \psi_1\psi_3 + \psi_2\psi_3)^8}{300000000\psi_1\psi_2\psi_3(\psi_1 + \psi_2 + \psi_3)^{17}}. \end{aligned}$$

Since this is equal to  $\det(C(\boldsymbol{\pi}_0, \boldsymbol{\psi}))$  for the design constructed from two starter choice sets, this design is optimal for the estimation of main effects of the attributes plus contrasts of the position effects when the generalised Davidson–Beaver position effects model is used.  $\square$

It is difficult to be specific about the minimum number of starter choice sets required in Theorem 5. The example used two starter choice sets and gave 50 choice sets in total, whereas 3600 choice sets would be required using the construction in Theorem 4, since there are 144 starter choice sets with difference (11, 11, 11). If all the  $\ell_q \geq m$  are equal, then at most  $\ell_q - 1$  starter choice sets are required for the construction in Theorem 5. It is clearly

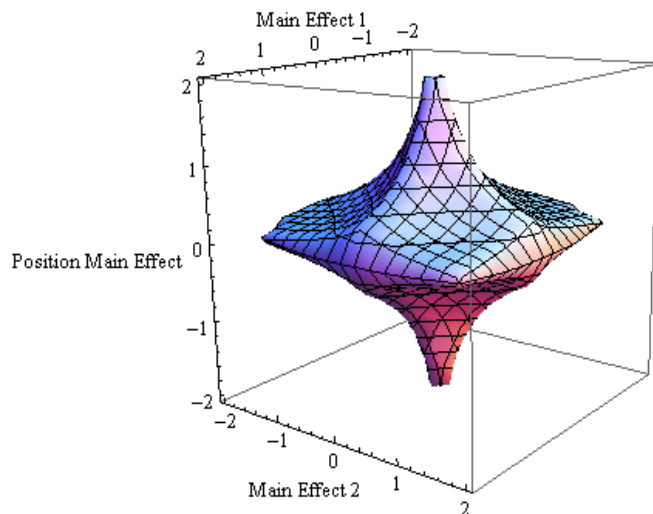


Figure 1: The range of parameter values where the design  $\{(00, 11), (11, 00), (01, 10), (10, 01)\}$  is optimal.

impractical to present that many choice sets to respondents, so we need to consider methods for constructing smaller designs.

One way to obtain a smaller design is to construct a small near-optimal design, which will contain fewer choice sets. Methods for doing this are given in Chapter 8 of Street and Burgess (2007). The other method that is used in practice is to partition the choice sets into *versions* of a suitable size and present respondents with one of the versions. We would then conduct analysis on the aggregate results.

We conclude this section with a small example that illustrates the performance of the designs investigated here for non-zero attribute effects.

**Example 3.** Consider an experiment with two 2-level attributes and choice sets of size 2. The design that is optimal by Theorem 5 is  $\{(00, 11), (11, 00), (01, 10), (10, 01)\}$ . By calculating  $\det(C(\boldsymbol{\pi}, \boldsymbol{\psi}))$  for various values of the attribute main effects and the position main effect for each of the possible designs, we can find a region of parameter values where the optimal design under the null hypothesis is still optimal. This region is given in Figure 1.

We can also look at the  $D$ -efficiency of the above design for particular sets of parameter values. Figure 2(a), (b) and (c) give contour plots of the  $D$ -efficiencies for the design above for a range of main effect values when the position main effect is 0, 1 and 2 respectively.

We can see that the design that is optimal under the null hypothesis is still optimal for reasonably small effect sizes. As the position effect increases, the region where this design is still optimal becomes smaller. We can see that the  $D$ -efficiency decreases as the main effects become larger in absolute value. This decrease is faster when the position effect is also larger. □

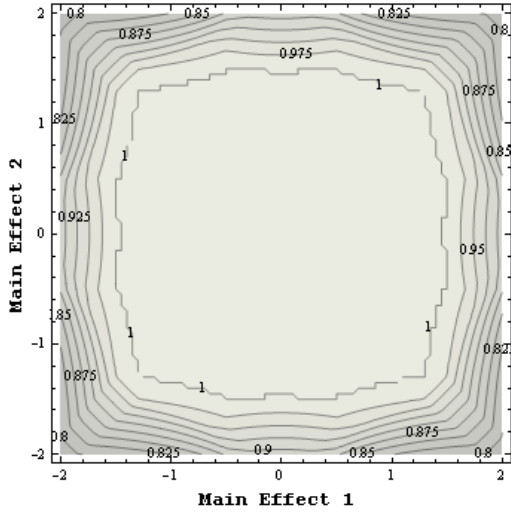
## 5 Optimal Designs for Attribute Main Effects plus Two-Factor Interactions and Position Effects

In this section we find optimal  $2^k$  designs for the estimation of main effects plus two-factor interactions of the attributes, and contrasts of the position effects, when  $m \geq 2$ . Burgess and Street (2003) prove results that gave such designs in the absence of position effects.

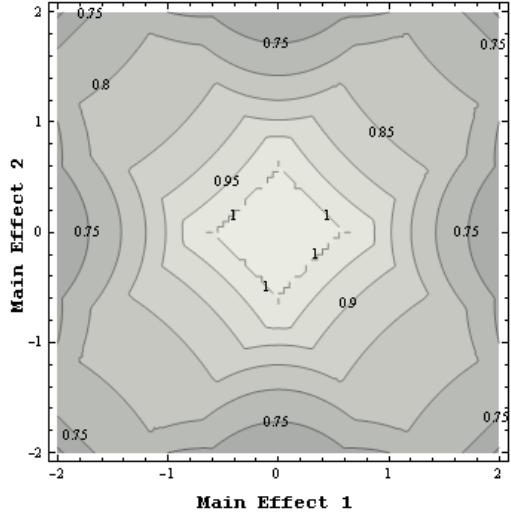
In this section, we generalise the class of competing designs considered in Burgess and Street (2003) to take position into account. The unordered difference vectors in Burgess and Street (2003) are for binary attributes only, and contain the number of attributes that differ in their levels for each pair of items in the choice set. Thus we define an entry of the ordered difference vector corresponding to the positions  $a$  and  $b$  to be the number of attributes in which the items in these positions differ. Hence the entries in the ordered difference vector  $\mathbf{v}$  are now the sum of the entries in each  $\mathbf{d}_{a,b}$ .

We let  $x_{\mathbf{v}_j;i,a,b}$  be the number of times that there are  $i$  attributes at different levels between the items in positions  $a$  and  $b$  of the choice set. Thus we have  $x_{\mathbf{v}_j;i,a,b} = \sum_{\mathbf{d} | \sum_q d_q = i} x_{\mathbf{v}_j;\mathbf{d},a,b}$ .

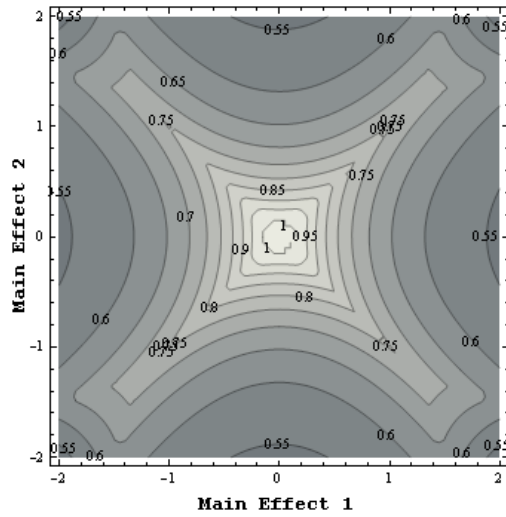
Next we express  $\Lambda_{\gamma\gamma}(\boldsymbol{\pi}_0, \boldsymbol{\psi})$  in terms of  $x_{\mathbf{v}_j;i,a,b}$ ,  $\mathbf{c}_{\mathbf{v}_j}$ ,  $i_{\mathbf{v}_j}$ , and some  $D_{k,i}$  matrices. We let  $D_{k,i}$  be a  $(0, 1)$  matrix of order  $2^k$  with a 1 in position  $(x, y)$  if the items  $T_x$  and  $T_y$  differ in the levels of  $i$  attributes. Street et al. (2001) and Burgess and Street (2003) have used these  $D_{k,i}$  matrices to obtain useful expressions for the determinant of the information matrix



(a)



(b)



(c)

Figure 2: The  $D$ -efficiencies for the design  $\{(00, 11), (11, 00), (01, 10), (10, 01)\}$  when the position effect is 0 (a), 1 (b), and 2 (c).

when the MNL model is used.

We begin by finding the proportion of choice sets where item  $T_i$  appears in position  $a$  of the choice set and item  $T_j$  appears in position  $b$  of the choice set, denoted by  $\lambda_{T_i \text{ in pos } a, T_j \text{ in pos } b}$ . In Section 4 we showed that the number of choice sets with difference  $\mathbf{d}$  for the pair of positions  $a$  and  $b$  was equal to  $L \sum_{\mathbf{v}_j} c_{\mathbf{v}_j} i_{\mathbf{v}_j} x_{\mathbf{v}_j; \mathbf{d}, a, b}$ . Thus the number of choice sets where the items in positions  $a$  and  $b$  differ in the levels of  $i$  of the attributes is

$$L \sum_{\mathbf{d} | \sum_q d_q = i} \sum_{\mathbf{v}_j} c_{\mathbf{v}_j} i_{\mathbf{v}_j} x_{\mathbf{v}_j; \mathbf{d}, a, b} = L \sum_{\mathbf{v}_j} c_{\mathbf{v}_j} i_{\mathbf{v}_j} x_{\mathbf{v}_j; i, a, b}.$$

There are  $\binom{k}{i}$  items that differ from  $T_j$  in the levels of  $i$  of the attributes, since we can select  $i$  of the  $k$  attributes to differ. Then the proportion of choice sets that contain a particular pair of items that differ in the levels of  $i$  of the attributes is

$$y_{i, a, b} = \frac{1}{N} \binom{k}{i}^{-1} \sum_{\mathbf{v}_j} c_{\mathbf{v}_j} i_{\mathbf{v}_j} x_{\mathbf{v}_j; i, a, b},$$

which, after simplification, gives

$$\Lambda_{\gamma\gamma}(\boldsymbol{\pi}_0, \boldsymbol{\psi}) = \frac{1}{\Psi_1} \sum_{a=1}^m \sum_{b \neq a} \sum_{i=1}^k \psi_a \psi_b y_{i, a, b} \left( \binom{k}{i} I_L - D_{k, i} \right).$$

Street et al. (2001) show that for a  $2^k$  experiment

$$B_{k, M} D_{k, i} = \left[ \binom{k-1}{i} - \binom{k-1}{i-1} \right] B_{k, M}, \text{ and}$$

$$B_{k, T} D_{k, i} = \left[ \binom{k-2}{i} - 2 \binom{k-2}{i-1} + \binom{k-2}{i-2} \right] B_{k, T},$$

where  $B_{k, M}$  is a matrix that containing the contrast coefficients for the attribute main effects and  $B_{k, T}$  is a matrix containing the contrast coefficients for the attribute two-factor interactions in a  $2^k$  experiment. Then

$$B_{k, M} \Lambda_{\gamma\gamma}(\boldsymbol{\pi}_0, \boldsymbol{\psi}) B_{k, M} = \frac{2}{\Psi_1} \sum_{a=1}^m \sum_{b \neq a} \sum_{i=1}^k \psi_a \psi_b y_{i, a, b} \binom{k-1}{i-1} I_k, \text{ and}$$

$$B_{k, T} \Lambda_{\gamma\gamma}(\boldsymbol{\pi}_0, \boldsymbol{\psi}) B_{k, T} = \frac{4}{\Psi_1} \sum_{a=1}^m \sum_{b \neq a} \sum_{i=1}^k \psi_a \psi_b y_{i, a, b} \binom{k-2}{i-1} I_{k(k-1)/2},$$



since

$$\binom{k}{i} - \binom{k-1}{i} + \binom{k-1}{i-1} = 2\binom{k-1}{i-1}, \text{ and}$$

$$\binom{k}{i} - \binom{k-2}{i} + 2\binom{k-2}{i-1} - \binom{k-2}{i-2} = 4\binom{k-2}{i-1}.$$

This also gives  $B_{k,M}\Lambda_{\gamma\gamma}(\boldsymbol{\pi}_0, \boldsymbol{\psi})B_{k,T} = \mathbf{0}$ , and  $B_{k,T}\Lambda_{\gamma\gamma}(\boldsymbol{\pi}_0, \boldsymbol{\psi})B_{k,M} = \mathbf{0}$ . Thus  $C_{\gamma\gamma}(\boldsymbol{\pi}_0, \boldsymbol{\psi})$  is a diagonal matrix.

The following result gives conditions for  $C_{\gamma\psi}(\boldsymbol{\pi}_0, \boldsymbol{\psi})$  to be  $\mathbf{0}$ , when attribute main effects plus two-factor interactions and contrasts of the position effects are of interest.

**Lemma 2.** *Let  $B_{\gamma\boldsymbol{\gamma}}$  be the main effects contrasts and the two-factor interaction contrasts. Then under the null hypothesis of equal merits  $C_{\gamma\psi}(\boldsymbol{\pi}_0, \boldsymbol{\psi}) = \mathbf{0}$  if each pair of levels for each pair of attributes appears equally often in each position of the choice set.*

*Proof.* Consider a generic term in  $B_{\gamma}\Lambda_{\gamma\psi}(\boldsymbol{\pi}_0, \boldsymbol{\psi})$ . The rows of this matrix are labelled by the main effects and the two-factor interactions of the attributes. The columns of this matrix are labelled by positions in the DCE.

Consider the entry of  $B_{\gamma}\Lambda_{\gamma\psi}(\boldsymbol{\pi}_0, \boldsymbol{\psi})$  where the row corresponds to the  $j^{\text{th}}$  component of the main effect of attribute  $q$  and the column corresponds to the  $a^{\text{th}}$  position of the choice set. By Lemma 1 we have  $(B_{\gamma}\Lambda_{\gamma\psi}(\boldsymbol{\pi}_0, \boldsymbol{\psi}))_{qj a} = 0$  since each pair of levels appears equally often in each pair of positions, and hence there is equal replication of levels for each attribute in each position.

Now suppose that the row corresponds to the  $j^{\text{th}}$  component of the two-factor interaction between attributes  $q_1$  and  $q_2$ , and that the column corresponds to position  $a$  of the choice

set. Then

$$\begin{aligned}
& (B_\gamma \Lambda_{\gamma\psi}(\boldsymbol{\pi}_0, \boldsymbol{\psi}))_{(q_1 q_2)j a} \\
&= \frac{1}{\Psi_1} \sum_{i=1}^L \sum_{b \neq a} \psi_b B_{(q_1 q_2)j, [i]} \left( \sum_{C|T_i \text{ in pos } a} \lambda_C - \sum_{C|T_i \text{ in pos } b} \lambda_C \right) \\
&= \frac{1}{\Psi_1} \sum_{x_1=0}^{\ell_{q_1}-1} \sum_{x_2=0}^{\ell_{q_2}-1} \sum_{b \neq a} \psi_b B_{(q_1 q_2)j, (x_1 x_2)} \left( \sum_{\substack{C|\text{att } q_1=x_1 \text{ and} \\ \text{att } q_2=x_2 \text{ in pos } a}} \lambda_C - \sum_{\substack{C|\text{att } q_1=x_1 \text{ and} \\ \text{att } q_2=x_2 \text{ in pos } b}} \lambda_C \right).
\end{aligned}$$

Therefore  $(B_\gamma \Lambda_{\gamma\psi}(\boldsymbol{\pi}_0, \boldsymbol{\psi}))_{(q_1 q_2)j a}$  will equal 0 if, for all  $x_1$  and  $x_2$ ,

$$\lambda_{\text{att } q_1=x_1, \text{att } q_2=x_2 \text{ in pos } a} - \lambda_{\text{att } q_1=x_1, \text{att } q_2=x_2 \text{ in pos } b} = 0,$$

for  $b \neq a$ . If this is true for all pairs of attributes  $q_1$  and  $q_2$  then  $B_\gamma \Lambda_{\gamma\psi}(\boldsymbol{\pi}_0, \boldsymbol{\psi}) = \mathbf{0}$ . Hence  $C_{\gamma\psi}(\boldsymbol{\pi}_0, \boldsymbol{\psi})$  will equal  $\mathbf{0}$ , as required.  $\square$

By assuming that the conditions of Lemma 2 apply, the determinant of the information matrix for the estimation of main effects plus two-factor interactions of the attributes and contrasts of the position effects is

$$\begin{aligned}
\det(C(\boldsymbol{\pi}_0, \boldsymbol{\psi})) &= \left[ \frac{2}{\Psi_1} \sum_{a=1}^m \sum_{b \neq a} \sum_{i=1}^k \psi_a \psi_b y_{i,a,b} \binom{k-1}{i-1} \right]^k \\
&\quad \times \left[ \frac{4}{\Psi_1} \sum_{a=1}^m \sum_{b \neq a} \sum_{i=1}^k \psi_a \psi_b y_{i,a,b} \binom{k-2}{i-1} \right]^{k(k-1)/2} \times \det(C_{\psi\psi}(\boldsymbol{\pi}_0, \boldsymbol{\psi})).
\end{aligned}$$

We now use the same method as in Burgess and Street (2003) to find the maximum value of this determinant, and hence the  $D$ -optimal design. We assume that each difference appears in each position equally often, so  $y_{i,a,b} = y_i$  for all  $1 \leq a, b \leq m$ . Then  $\sum_{a=1}^m \sum_{b \neq a} \psi_a \psi_b y_{i,a,b} = \Psi_2 y_i$ , and thus

$$\begin{aligned}
& \det(C(\boldsymbol{\pi}_0, \boldsymbol{\psi})) \\
&= \left( \frac{\Psi_2}{\Psi_1} \right)^{k + \frac{k(k-1)}{2}} \times \left[ 2 \sum_{i=1}^k y_i \binom{k-1}{i-1} \right]^k \times \left[ 4 \sum_{i=1}^k y_i \binom{k-2}{i-1} \right]^{\frac{k(k-1)}{2}} \times \det(C_{\psi\psi}(\boldsymbol{\pi}_0, \boldsymbol{\psi})).
\end{aligned}$$

**Theorem 6.** *Under the null hypothesis of equal merits, the  $D$ -optimal design for the estimation of main effects plus two-factor interactions of the attributes and contrasts of the position effects is given by designs where*

$$y_i = y_{i,a,b} = \begin{cases} \frac{1}{2^k} \binom{k+1}{k/2}^{-1} & \text{where } k \text{ is even and } i = k/2, k/2 + 1, \\ \frac{1}{2^k} \binom{k}{(k+1)/2}^{-1} & \text{where } k \text{ is odd and } i = (k+1)/2, \\ 0 & \text{otherwise,} \end{cases}$$

for all  $1 \leq a, b \leq m$ .

*Proof.* The proof of this theorem is very similar to the proof of Theorem 2 in Burgess and Street (2003). Since  $(\Psi_2/\Psi_1)^{k+k(k-1)/2} \times \det(C_{\psi\psi}(\boldsymbol{\pi}_0, \boldsymbol{\psi}))$  is a function of the entries in  $\boldsymbol{\psi}$  only, and thus independent of the selection of the design for a given choice set size, we need to maximise

$$\left[ 2 \sum_{i=1}^k y_i \binom{k-1}{i-1} \right]^k \times \left[ 4 \sum_{i=1}^k y_i \binom{k-2}{i-1} \right]^{k(k-1)/2},$$

subject to the constraint  $\frac{1}{2^k} \sum_{i=1}^k \binom{k}{i} y_i = 1$ .

As in the proof of Theorem 2 of Burgess and Street (2003), we can restate this problem as maximising the function  $f = AB^{(k-1)/2}$  subject to the constraint  $\sum_{i=1}^k \binom{k}{i} x_i = 1$ , where  $A = 2^k \sum_{i=1}^k \binom{k-1}{i-1} x_i$ ,  $B = 2^k \sum_{i=1}^k \binom{k-2}{i-1} x_i$ , and  $x_i = 2^k y_i$ . Street et al. (2001) show that this constrained function is maximised when

$$x_i = \begin{cases} \binom{k+1}{k/2}^{-1} & \text{where } k \text{ is even and } i = k/2, k/2 + 1, \\ \binom{k}{(k+1)/2}^{-1} & \text{where } k \text{ is odd and } i = (k+1)/2, \\ 0 & \text{otherwise.} \end{cases}$$

Then the result follows by substituting  $y_i = x_i/2^k$ . □

We now illustrate this theorem with an example.

**Example 4.** Consider an experiment with three 2-level attributes and with choice sets of size 2. Then there are three possible ordered difference vectors that could be used to describe the choice sets,  $\mathbf{v}_1 = (1)$ ,  $\mathbf{v}_2 = (2)$ , and  $\mathbf{v}_3 = (3)$ . According to Theorem 6, the

design consisting of all ordered choice sets with  $(k + 1)/2 = 2$  attributes at different levels is optimal for the estimation of main effects plus two-factor interactions of the attributes and contrasts of the position effects. This experiment will consist of 24 choice sets. Since  $C(\boldsymbol{\pi}_0, \boldsymbol{\psi}) = B\Lambda(\boldsymbol{\pi}_0, \boldsymbol{\psi})B^T$  and, using the results in Section 3, we have

$$\det(C(\boldsymbol{\pi}_0, \boldsymbol{\psi})) = \frac{\psi_1^5 \psi_2^5}{1458(\psi_1 + \psi_2)^{12}}.$$

According to Theorem 6, the value of  $y_i$  that gives the optimal design is  $\frac{1}{2^k} \binom{k}{(k+1)/2}^{-1} = \frac{1}{24}$  when  $i = 2$  and 0 when  $i \neq 2$ , and thus

$$\begin{aligned} \det(C(\boldsymbol{\pi}_0, \boldsymbol{\psi}))_{\text{OPT}} &= \left(\frac{\Psi_1}{\Psi_2}\right)^6 \left(2 \times \frac{1}{24} \times \binom{2}{1}\right)^3 \left(4 \times \frac{1}{12} \times \binom{1}{1}\right)^3 \times \frac{1}{2\psi_1\psi_2} \\ &= \frac{\psi_1^5 \psi_2^5}{1458(\psi_1 + \psi_2)^{12}}, \end{aligned}$$

showing that the design is optimal. □

## 6 A Design Approach Based on Williams Designs

A natural candidate for the design of choice experiments when position effects are of interest is a Latin square balanced for carry-over effects of all orders. A *carry-over effect* is the effect that a previously presented item has on the response to the item currently being considered. The *order* of a carry-over effect refers to the number of previously presented items that are considered to have an effect on the item currently being considered. Wakeling and MacFie (1995) provide a good discussion on the use of Latin squares in taste-testing experiments where carry-over effects are of interest. The authors use *Williams designs* (Williams (1949)) to balance for carry-over effects of all orders.

The problem of constructing designs for the estimation of position effects in DCEs is similar to the problem of constructing designs for the estimation of carry-over effects of all orders. In both cases, the selection of items that are to be presented in any pair of positions is important. In this section, we consider how Williams designs could be used for the estimation of main effects of the attributes plus contrasts of the position effects. Note that these designs are not in the class of competing designs considered earlier in this paper.

To obtain a factorial representation for the items presented in the experiment, one design strategy would be to use a Williams design to determine which items form the choice sets, and an *orthogonal array* (OA) to describe each of the items in the Williams design in terms of attributes. Williams (1949) states that if the first column of a set of mutually orthogonal Latin squares is in standard order, then each pair of items will appear in each pair of positions exactly once, so Lemma 1 holds for this construction. We could possibly use other designs that are balanced for some carry-over effects, such as complete Latin squares or equineighbourhood balanced incomplete block designs, but unless the design is balanced for carry-over effects of all orders, Lemma 1 does not necessarily hold. The next example illustrates the use of a Williams design and an OA to construct a choice design.

**Example 5.** Consider the Williams design in Table 3(a) and the OA in Table 3(b). If we label the rows of the OA with 0, 1, 2, and 3, and replace the entries in the Williams designs with the corresponding row of the OA then we obtain the choice sets in Table 4.  $\square$

(a)				(b)										
0	1	2	3	0	2	3	1	0	3	1	2	0	0	0
1	0	3	2	1	3	2	0	1	2	0	3	0	1	1
2	3	0	1	2	0	1	3	2	1	3	0	1	0	1
3	2	1	0	3	1	0	2	3	0	2	1	1	1	0

Table 3: A Williams design with 4 columns (a) and a 4 run orthogonal array (b).

Since each pair of items in the OA appears in each pair of positions exactly once, we can reorder the rows and columns of  $\Lambda_{\gamma\gamma}(\boldsymbol{\pi}_0, \boldsymbol{\psi})$  to give

$$\Lambda_{\gamma\gamma}(\boldsymbol{\pi}_0, \boldsymbol{\psi}) = \frac{\Psi_2}{N\Psi_1} \begin{bmatrix} (m-1)I_m - J_m & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix},$$

Option 1	Option 2	Option 3	Option 4
0 0 0	0 1 1	1 0 1	1 1 0
0 1 1	0 0 0	1 1 0	1 0 1
1 0 1	1 1 0	0 0 0	0 1 1
1 1 0	1 0 1	0 1 1	0 0 0
0 0 0	1 0 1	1 1 0	0 1 1
0 1 1	1 1 0	1 0 1	0 0 0
1 0 1	0 0 0	0 1 1	1 1 0
1 1 0	0 1 1	0 0 0	1 0 1
0 0 0	1 1 0	1 0 1	0 1 1
0 1 1	1 0 1	0 0 0	1 1 0
1 0 1	0 1 1	1 1 0	0 0 0
1 1 0	0 0 0	1 0 1	0 1 1

Table 4: The  $2^3$  choice experiment from Example 5

where the  $\mathbf{0}$  entries correspond to the items in the complete factorial that do not appear in the OA, and  $J_m$  is an  $m \times m$  matrix of 1s.

Since each pair of items appears in each pair of positions once, and the items form an OA, the conditions of Lemma 1 are satisfied. Thus  $C_{\gamma\psi}(\boldsymbol{\pi}_0, \boldsymbol{\psi}) = \mathbf{0}$ . Using the results in Bush (2009) we obtain

$$\begin{aligned}
& C(\boldsymbol{\pi}_0, \boldsymbol{\psi}) \\
&= \begin{bmatrix} B_F & B_{\bar{F}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & B_\psi \end{bmatrix} \times \begin{bmatrix} \frac{\Psi_2}{N\Psi_1}((m-1)I_m - J_m) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \Lambda_{\psi\psi}(\boldsymbol{\pi}_0, \boldsymbol{\psi}) \end{bmatrix} \times \begin{bmatrix} B_F^T & \mathbf{0} \\ B_{\bar{F}}^T & \mathbf{0} \\ \mathbf{0} & B_\psi^T \end{bmatrix} \\
&= \begin{bmatrix} \frac{\Psi_2}{N\Psi_1} B_F((m-1)I_m - J_m) B_F^T & \mathbf{0} \\ \mathbf{0} & C_{\psi\psi}(\boldsymbol{\pi}_0, \boldsymbol{\psi}) \end{bmatrix} \\
&= \begin{bmatrix} \frac{L\Psi_2}{N\Psi_1} I_p & \mathbf{0} \\ \mathbf{0} & C_{\psi\psi}(\boldsymbol{\pi}_0, \boldsymbol{\psi}) \end{bmatrix},
\end{aligned}$$

where  $p = \sum_{q=1}^k (\ell_q - 1)$ ,  $B_F$  contains the contrast coefficients for the items in the OA, and  $B_{\bar{F}}$  contains the contrast coefficients for the remaining items in the complete factorial.

If we use this type of design, then the number of items in the OA is the number of items in each component Latin square. In a Williams design, there are  $m - 1$  mutually orthogonal  $m \times m$  Latin squares. So  $N = m(m - 1)$ .

The existence of such a design depends on the existence of a Williams design of order  $m$  and an OA with  $m$  items. For large  $m$  this approach needs to be modified. One way to do this is to use a balanced incomplete block design (BIBD) to select subsets of the items from the OA and then construct a Williams design on the items in each block. The next example illustrates this construction.

**Example 6.** Suppose that we would like to use the items in an OA with 8 runs, but want 4 options in each choice set. Then by using the Williams design in Table 3(a) and a BIBD(8, 14, 7, 4, 3), we obtain a DCE with 12 choice sets from each of the 14 blocks, and so

$N = 168$ . The first 36 choice sets are shown in Table 5, with the defining entries from the BIBD in bold. □

<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>0</b>	<b>1</b>	<b>2</b>	<b>4</b>	<b>0</b>	<b>1</b>	<b>5</b>	<b>6</b>
1	0	3	2	1	0	4	2	1	0	6	5
2	3	0	1	2	4	0	1	5	6	0	1
3	2	1	0	4	2	1	0	6	5	1	0
0	2	3	1	0	2	4	1	0	5	6	1
1	3	2	0	1	4	2	0	1	6	5	0
2	0	1	3	2	0	1	4	5	0	1	6
3	1	0	2	4	1	0	2	6	1	0	5
0	3	2	1	0	4	2	1	0	6	5	1
1	2	0	3	1	2	0	4	1	5	0	6
2	1	3	0	2	1	4	0	5	1	6	0
3	0	2	1	4	0	2	1	6	0	5	1

Table 5: The first 36 choice sets in the design constructed in Example 6

The DCEs constructed using only an OA and a Williams design are special cases of the class of designs described here. In this case, the BIBD is a single block containing one copy of each of the items. It follows that the properties of the DCEs that incorporate a BIBD in this way are similar to those which use only an OA and a Williams design. Each pair of items in a particular block will appear in each pair of positions of the Williams design developed from that block exactly once. Since we use a  $\text{BIBD}(v, b, r, m, \lambda)$  to allocate items to the entries in each Williams design, each pair of items will appear in the same block  $\lambda$  times. Therefore each pair of items appears in each pair of positions  $\lambda$  times across the DCE. Thus

$$\Lambda_{\gamma\gamma}(\boldsymbol{\pi}_0, \boldsymbol{\psi}) = \frac{\lambda\Psi_2}{N\Psi_1} \begin{bmatrix} r(m-1)I_v - J_v & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix},$$



where, once again, the  $\mathbf{0}$  entries correspond to the items that do not appear in the OA. The number of rows and columns in  $I$  and  $J$  is equal to the number of runs in the OA, the value of  $v$  in the BIBD. Then, by using the results in Bush (2009) for OAs, we obtain

$$C(\boldsymbol{\pi}_0, \boldsymbol{\psi}) = \begin{bmatrix} \frac{L\lambda(r(m-1)+1)\Psi_2}{Nv\Psi_1} I_{\prod_{q=1}^k (\ell_q-1)} & \mathbf{0} \\ \mathbf{0} & C_{\boldsymbol{\psi}\boldsymbol{\psi}}(\boldsymbol{\pi}_0, \boldsymbol{\psi}) \end{bmatrix}.$$

If we use this type of design, then each block gives rise to  $m(m-1)$  choice sets of size  $m$ . So there are  $N = bm(m-1)$  choice sets in total. As with the previous design strategy, this strategy relies on the existence of three designs. We require a Williams design of order  $m$ , an orthogonal array with  $v$  runs, and a BIBD( $v, b, r, m, \lambda$ ).

In general we cannot be explicit about the efficiencies of these designs. If, however,  $\ell_q \geq m$  for all  $q$  then  $S_q = \frac{m(m-1)}{2}$  for all  $q$ . Thus

$$\det(C(\boldsymbol{\pi}_0, \boldsymbol{\psi})) = \prod_{q=1}^k \left[ \frac{L\lambda(r(m-1)+1)\Psi_2}{vN\Psi_1} \right]^{\ell_q-1} \times \det(C_{\boldsymbol{\psi}\boldsymbol{\psi}}(\boldsymbol{\pi}_0, \boldsymbol{\psi})),$$

and

$$\det(C(\boldsymbol{\pi}_0, \boldsymbol{\psi})_{\text{OPT}}) = \prod_{q=1}^k \left[ \frac{2S_q \ell_q \Psi_2}{Lm(m-1)\Psi_1(\ell_q-1)} \right]^{\ell_q-1} \times \det(C_{\boldsymbol{\psi}\boldsymbol{\psi}}(\boldsymbol{\pi}_0, \boldsymbol{\psi})),$$

and so the efficiency of the designs from this construction is

$$\prod_{q=1}^k \left[ \frac{L^2(\ell_q-1)\lambda(r(m-1)+1)}{v^2 \ell_q r(m-1)} \right]^{\ell_q-1}.$$

In particular if  $k = 2$ ,  $m = 3$  and  $\ell_q = 5$ , there is a BIBD(25, 100, 12, 3, 1) from which we construct a DCE with  $100 \times 6$  choice sets that is 86.4% efficient when compared to the 50 choice sets given in Example 2 for this situation.

## 7 Discussion

We conclude by considering some practical considerations arising from the assumptions made about the designs that we construct in this paper.

In Theorem 5 we make the assumption that, for each pair of positions, each non-zero difference needs to appear equally often. When there is a high level of so-called donkey voting, where a respondent always selects the item in a particular position of the choice set, this assumption will not necessarily give the most efficient design. For instance, consider an experiment with  $k = 2$ ,  $\ell_1 = \ell_2 = 2$  and  $m = 3$ , where 90% of respondents always select the first option. The design that is optimal by Theorem 5, which is constructed using three starter choice sets,  $(00, 01, 11)$ ,  $(00, 10, 01)$ , and  $(00, 11, 10)$ , is 92.8% as efficient as the design constructed using only one starter choice set,  $(00, 11, 10)$ .

In Section 4, we mentioned two approaches that can be used to obtain small designs. Both of these methods introduce some level of confounding between the effects. In the first method this is a consequence of choosing a near-optimal design rather than an optimal design. In the latter method the confounding arises from the heterogeneity in the preferences of the respondents. Further research could investigate the merits of treating the versions as blocks both when designing the experiment and analysing the results.

## Acknowledgements

We thank the referees who provided helpful comments that improved the presentation of this paper.

## References

- Abel, J.R. (2006). Difference Families. In: Colbourn, C. Dinitz, J. ed., *Handbook of Combinatorial Designs*, Second Edition. Boca Raton: Chapman & Hall/CRC Press.
- Bradley, R.A., El-Helbawy, A.T. (1976). Treatment contrasts in paired comparisons: Basic procedures with applications to factorials. *Biometrika* **63**:255–262.
- Bradley, R.A., Terry, M.E., (1952). Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons. *Biometrika* 39(3/4), 324–345.

- Bugelski, B.R. (1949). A note on Grant's discussion of the Latin square principle in the design of experiments. *Psychological Bulletin* **46**:49–50.
- Burgess, L.B., Street, D.J. (2003). Optimal designs for  $2^k$  choice experiments. *Communications in Statistics – Theory and Methods* **32**:2185–2206.
- Burgess, L.B., Street, D.J. (2005). Optimal designs for choice experiments with asymmetric attributes. *Journal of Statistical Planning and Inference* **134**:288–301.
- Bush, S. (2009). Optimal Designs for Stated Choice Experiments when using a Fraction Factorial Starting Design. Preprint.
- Chrzan, K. (1994). Three Kinds of Order Effects in Choice-Based Conjoint Analysis. *Marketing Letters* **5**:165–172.
- David, H.A. (1988). *The Method of Paired Comparisons*. London: Charles Griffin & Co Ltd.
- Davidson, R.R., Beaver, R.J. (1977). On Extending the Bradley-Terry Model to Incorporate Within-Pair Order Effects. *Biometrics* **33**:693–702.
- El-Helbawy, A.T., Ahmed, E.A., Alharbey, A.H. (1994). Optimal Designs for Assymetrical Factorial Paired Comparison Experiments. *Communications in Statistics - Simulation and Computation* **22**:663–681.
- El-Helbawy, A.T., Bradley, R.A. (1978). Treatment Contrasts in Paired Comparisons: Large-Sample Results, Applications, and Some Optimal Designs. *Journal of the American Statistical Association* **73**: 831–839.
- Kalton, G., Collins, M., Brook, L. (1978). Experiments in Wording Opinion Questions. *Applied Statistics* **27**:149–161.
- Kessels, R., Goos, P., and Vandebroek, M. (2006). A Comparison of Criteria to Design Efficient Choice Experiments. *Journal of Marketing Research* **43**:409–419.

- Louviere, J.J., Hensher, D.A., Swait, J.D. (2000). *Stated Choice Methods*. Cambridge: Cambridge University Press.
- Schuman, H., Presser, P. (1981). *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Content*. New York: Academic Press.
- Street, D.J., Bunch, D.S., Moore, B.J. (2001). Optimal Designs for  $2^k$  Paired Comparison Experiments. *Communications in Statistics – Theory and Methods* **30**:2149–2171.
- Street, D.J., Burgess, L. (2007). *The Construction of Optimal Stated Choice Experiments: Theory and Methods*. Hoboken: Wiley.
- Train, K.E. (2003). *Discrete Choice Methods with Simulation*. New York: Cambridge University Press.
- van der Waerden, P., Borgers, A., Timmermans, H., Bérénos, M. (2006). Order Effects in Stated-Choice Experiments: Study of Transport Mode Choice Decisions. *Transportation Research Record* **1985**:12–18.
- Wakeling, I.N., MacFie, J.H. (1995). Designing Consumer Trials Balanced for First and Higher Orders of Carry-Over Effect when only a Subset of  $k$  Samples from  $t$  may be Tested. *Food Quality and Preference* **6**:299–308.
- Wickelmaier, F., Choisel, S. (2006). Modeling within-pair order effects in paired-comparison judgments. In: Kornbrot, D. E. Msetfi, R. M. MacRae, A. W. ed., *Proceedings of the 22nd Annual Meeting of the International Society for Psychophysics*. The ISP.
- Williams, E.J. (1949). Experimental Designs Balanced for the Estimation of Residual Effects of Treatments. *Australian Journal of Scientific Research Series A* **2**:149–168.