# Semi-Bayesian D-Optimal Choice Designs for Mixtures

Aiste Ruseckaite*

Master Thesis

Erasmus University Rotterdam

Academic supervisor: Prof. Dr. Peter Goos

Co-reader: Dr. Dennis Fok

August, 2012

*Student ID: 357347

**Abstract**

In order to acquire insight into consumer preferences for products and services that are described by certain attributes, choice experiments are employed. For efficiency, this should be done by means of an optimal experimental design, which gives the most precise estimates for the parameters in the corresponding statistical model. Sometimes attributes of products and services can be mixtures of ingredients. Although mixture models are commonly used in industrial experiments, they have never been introduced in choice modeling. This master thesis aims at introducing mixtures in the choice context, since often consumer products and services can be described as mixtures of ingredients. An algorithm to construct semi-Bayesian D-optimal experimental designs is presented for the multinomial logit model when choices are based on a mixture of ingredients. The resulting designs are D-optimal and based on a mixture coordinate-exchange algorithm. Further, some features of them are discussed. It is shown that designs, when prior parameter values required for choice models are not assumed to be zero, differ from the utility neutral designs, where such an assumption is made. We also show that semi-Bayesian designs differ from and perform better than locally optimal ones (and the utility neutral designs) for most of the time. As often it is difficult to obtain accurate prior parameter values, parameter misspecification is also investigated. It is demonstrated that monotonous misspecifications in true parameters do not distort the outcome, and might help to design more robust designs.

*Keywords*: Bayesian design; Choice experiments; D-optimality; Experimental design; Halton sequences; Mixture coordinate-exchange algorithm; Mixture experiment; Multinomial logit model; Optimal design

**Acknowledgements**

The quality of this master thesis was greatly enhanced by the gracious assistance and guidance of Prof. Dr. Peter Goos and Dr. Dennis Fok. I am heartily thankful to both of them for their interest and support during the process of this thesis.

# Contents

# 1 Introduction

The easiest way to figure out consumer preferences for different products and services is asking what they think they would choose in certain situations and obtaining *stated preference* data. It is also the only means to obtain data on products that are not on the market yet. However, surveys have the drawback of being expensive and not widely available. They also trade higher amount of information (at higher costs) against the decreasing quality of that information, as respondents tend to get bored and tired if the process of filling out a questionnaire is too long. In order to obtain consumer preferences information, choice experiments are employed.

Choice experiments intend to capture consumer preferences for different attributes of various goods and precisely predict future purchasing behaviour. By using them, real marketplace choices and decisions can be modeled and applied to evaluate market demand (Carson et al., 1994). In a choice experiment, a product or service is characterized by a combination of attribute levels called a profile or an alternative. Respondents then choose one from a group of profiles called a choice set. They repeat this task for several other choice sets presented to them. All submitted choice sets make up the experimental design (Kessels et al., 2009). Such a choice experiment allows us to estimate the importance of each attribute and its levels based on the respondents' preferences.

It is important to know which experiments are optimal to carry out as they are expensive and might be of complicated nature to implement. An optimal experimental design is one which provides the most precise estimates of the parameters of the choice model. Optimal designs can reduce experimental costs by allowing econometric models to be used with as few experimental observations as possible. The more efficient design allows a researcher to reduce the number of questions asked in the survey or the number of respondents involved in the process. In case of a non-optimal design, a greater number of observations (and, thus, greater expense) is needed to estimate the parameters with the same precision as with an optimal design. In order to design a choice experiment optimally, one has to select the choice sets that result in accurately estimated model parameters and precise predictions. Thus, the design of the choice experiment determines what model(s) can be used with what levels of precision (Kessels et al., 2011). Techniques for finding such optimal tailor-made experiments are available in the literature.

Up to now, standard choice models are typically applied in the choice modeling literature. Meanwhile, to model the dependence of the product quality on ingredient proportions in agricultural and industrial statistics, experiments with mixtures are common (Cornell, 2002). In general, a mixture experiment involves varying the proportions of $k$ ($k \geq 2$) mixture components and measuring one or more response variables. When constraints on the proportions are imposed (such as, lower and/or upper bounds on the $k$ components and/or on $M$ linear combinations of components), we have a constrained mixture experiment (Piepel, et al., 2005). Consumer products and

services are also often based on a mixture of ingredients. Examples include:

- taste experiments involving the mixing of different blends to determine optimal proportions for each ingredient
- choices between transportation modes depend on different types/mixtures and amounts of costs (toll, fuel) and travel times (congested, free flow)
- cake formulations when using flour, sugar, water, baking powder, and shortening (Cornell, 2002)
- tobacco blends which consist of flue-cured tobacco, burley, turkish blend, and processed tobacco (Cornell, 2002)

Thus, mixture models perfectly fit within the context of consumer products and services to measure how the attributes of them jointly affect consumer preferences.

The aim of this master thesis is to introduce mixture models in the choice context and develop an algorithm to set up optimal choice experiments involving mixtures. It is relevant both scientifically and practically as mixture models have not yet been used in the context of choice experiments and are present only in industrial setting, and as it may be seen from the reasoning above, it might be a convenient way to handle some sort of choice problems in the field.

In the next section, what has been done in the fields of choice modeling, optimal choice experiments and mixture models is briefly reviewed, and the multinomial logit model for choice experiments, mixture models and optimal design criteria are introduced. Then, the algorithm itself is presented. The two main steps are outlined and separately discussed in detail by describing all substeps needed and techniques used. Then, benchmark approach for the designs constructed in this master thesis is introduced. Finally, the developed algorithm is implemented for several different experimental settings and results are discussed. The last section contains a conclusion.

# 2 Models and Design Criteria

In this section, we review literature and previous research first. Then, we describe multinomial logit and mixture models. Finally, we introduce design criteria used in optimal design theory and discuss the D-optimality criterion, that is chosen to be used in this master thesis, in more detail.

## 2.1 Literature and Previous Research

Choice experiments are popular in marketing as they can imitate consumer actual behavior. Quite a few authors work on developing efficient algorithms for designing optimal experimental designs. However, so far, mixture experiments, choice experiments and optimal designs have not been combined into one setting and have mostly been analyzed separately. For instance, Goos et al. (2012) propose a new approach for modeling the impact of investments in different types of media advertising, namely, mixture-amount modeling. The model separates the impact of the media mix from the impact of the total amount of advertising effort. Piepel et al. (2005) combine both mixture experiments and an optimal design idea in a unique and challenging nuclear waste glass constrained mixture experiment design problem. They propose a new coordinate-exchange algorithm for mixture experiments without using candidate points, as in this application it was impossible to generate and store the huge number of them needed for 19 or 21 mixture components. Goos and Donev (2006, 2007) describe an algorithmic approach to designing blocked experiments involving mixture components, when fixed and/or random blocks are present, where blocks represent groups of mixture blends where each group of block is assumed to differ from the other groups or blocks by an additive constant.

On the other hand, choice experiments and optimal design concepts have been considered, among others, by Huber and Zwerina (1996), who investigate the usefulness of reasonable prior values for the coefficients in the model for designing efficient choice designs; Sándor and Wedel (2001), who provide more efficient designs for stated choice experiments based on prior information about parameter values and uncertainty associated with them; Kessels et al. (2006), who elaborate on the G- and V-optimality criteria (which together with D- and A-criteria are used to design optimal experiments in optimal design theory, and are briefly discussed in subsequent sections) for the multinomial logit model to design efficient choice experiments, and compare their prediction performances with those of the D- and A-optimality criteria. Further, Kessels et al. (2009) present a much faster algorithm for generating Bayesian optimal designs for D-, A-, G- and V-optimality criteria, while simultaneously improving the statistical efficiency of the designs.

All papers written so far study either mixture experiments or choice experiments and not the two techniques together. There also exist quite a few articles that involve optimal design ideas.

However, mixture models have not been introduced in choice modeling yet and have been used only in an industrial context. This master thesis aims at contributing to the existing literature and combining all three techniques into one setting by considering mixture experiments, multinomial logit models for data from choice experiments, and an optimal design idea.

As there are many different concepts prevalent in optimal design theory, it is necessary to outline the ones that will be used in this master thesis. In a mixture choice experiment, a product or service is represented by a combination of *ingredient proportions* that sum up to one and make up an *alternative*. A group of alternatives presented to a respondent is called a *choice set*. Every row in a design matrix $\mathbf{X}$ constitutes an alternative for a certain product or service.

## 2.2 Multinomial Logit Model

To model discrete choices among different alternatives, the multinomial logit (MNL) model is commonly used. It relies on random utility theory, in which the utility of each alternative $j$, $j = 1, \ldots, J$, in choice set $s$, $s = 1, \ldots, S$, perceived by respondent is a linear function of observed alternative specific characteristics plus an additive error term,

$$U_{js} = \mathbf{x}'_{js}\boldsymbol{\beta} + \epsilon_{js},$$

where $\mathbf{x}_{js}$ is a $k \times 1$ vector containing the attribute levels of alternative $j$ in choice set $s$, and $\boldsymbol{\beta}$ is a $k \times 1$ vector of parameter values representing the effects of the attribute proportions on the utility. The alternative $j$ in a choice set $s$ is chosen if it gives the highest utility, and, thus, the probability that respondent chooses alternative $j$ in choice set $s$ is

$$p_{js} = P\{U_{js} = \max\{U_{1s}, \ldots, U_{Js}\}\} = P\{\mathbf{x}'_{js}\boldsymbol{\beta} + \epsilon_{js} > \max_{k=1,\ldots,J,k\neq j}\{\mathbf{x}'_{ks}\boldsymbol{\beta} + \epsilon_{ks}\}\}.$$

The stochastic components $\epsilon_{js}$ are assumed to be mutually independent and to follow a so-called log Weibull distribution (also known as a type I extreme value distribution). In this case, the distribution function of each $\epsilon_{js}$ is given by

$$F(t) = \exp(-\exp(-t)).$$

Under these assumptions, the MNL probability that respondent chooses alternative $j$ in choice set $s$ becomes

$$p_{js} = \frac{\exp(\mathbf{x}'_{js}\boldsymbol{\beta})}{\sum_{t=1}^{J}\exp(\mathbf{x}'_{ts}\boldsymbol{\beta})}, \tag{1}$$

where $\boldsymbol{\beta}$ can be estimated using maximum likelihood. Because the same parameter vector $\boldsymbol{\beta}$

is attached to every respondent, it is assumed in this model that people's preferences for the attribute levels are homogeneous across the population (Kessels et al., 2011; Verbeek, 2008).

## 2.3 Mixture Models

In order to model choice as a function of mixtures of ingredients, a mixture model is required, which has been developed in application areas other than choice modeling and applied only in the context of linear models. It is a type of regression model in which the $k$ explanatory variables $x_i$ ($0 \leq x_i \leq 1$) are the proportions of ingredients. The *mixture constraint*

$$\sum_{i=1}^{k} x_i = x_1 + x_2 + \cdots + x_k = 1 \tag{2}$$

has a substantial impact on the models that can be fitted. The first major consequence of the mixture constraint is that the linear model cannot contain an intercept. Otherwise, the model's parameters cannot be estimated uniquely. If we embed mixtures in MNL models, modeling becomes even more complicated than in the industrial context, as the ingredient proportions of the mixture sum to 1. In a MNL model, it is not enough to leave an intercept out. In addition, one of the $x$'s should be dropped out of the random utility model as well, in order to ensure identification. It is important to stress that the algorithm for designing an optimal experimental design (to be developed in this thesis) should be indifferent to which one of them is selected to be left out.

Another consequence of the mixture constraint is that all cross-products of proportions, $x_i x_j$, and the squares $x_i^2$ should not be included simultaneously as this also leads to perfect collinearity. To see this, note that

$$x_i^2 = x_i(1 - \sum_{\substack{j=1 \\ j \neq i}}^{k} x_j) = x_i - \sum_{\substack{j=1 \\ j \neq i}}^{k} x_i x_j,$$

for every proportion $x_i$. Thus, the square of a proportion is a linear combination of that proportion and its cross-products with all other $k-1$ proportions composing the mixture. The same applies to higher powers of ingredient proportions, too.

Often, additional constraints on the proportions of ingredients are imposed, such as, lower and/or upper bounds on the $k$ components,

$$L_i \leq x_i \leq U_i.$$

However, in some cases it is not difficult to redefine the design problem for pseudocomponents, which are linear transformations of the original mixture variables, and for which the constraints

in Equation (2) and $0 \leq x_i \leq 1$ still apply (Goos and Donev, 2006). Constraints on $M$ linear combinations of components can also be introduced

$$C_m \leq \sum_{i=1}^{k} A_{mi} x_i \leq D_m,$$

where $m = 1, \ldots, M$. In this thesis, however, we do not consider this kind of constraints.

The mixture constraint given in Equation (2) and the aforementioned consequences naturally lead to the family of Scheffé mixture models (Goos et al., 2012). The first-order Scheffé model for a continuous dependent variable is given by

$$y = \sum_{i=1}^{k} \beta_i x_i + \epsilon, \tag{3}$$

whereas the second-order Scheffé model is given by

$$y = \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \beta_{ij} x_i x_j + \epsilon. \tag{4}$$

The so-called special-cubic model can be written as

$$y = \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \beta_{ij} x_i x_j + \sum_{i=1}^{k-2} \sum_{j=i+1}^{k-1} \sum_{k=j+1}^{k} \beta_{ijk} x_i x_j x_k + \epsilon. \tag{5}$$

The interpretation of a model coefficient $\beta_i$ in Equation (3) is the expected response if $x_i$ is 100%, i.e., if a product consists of an ingredient $i$ only. We cannot interpret it as the effect of an ingredient $i$, since changing the proportion $x_i$ requires at least one other proportion to be changed as well. Otherwise, the mixture constraint in Equation (2) is violated. It is therefore difficult to interpret individual parameters $\beta_i$.

If we expect interaction effects like synergism (interaction of ingredients such that the total effect is greater than the sum of the individual effects) or antagonism (interaction of ingredients such that the total effect is smaller than the sum of the individual effects), we should use the second- or third-order model in Equation (4) or Equation (5). However, the numbers of terms in the second- or third-order Scheffé mixture models increase rapidly with the number of proportions $k$ (Goos et al., 2012; Scheffé, 1958). As a result, estimating these models requires a larger number of observations.

## 2.4 Design Criteria

In order to design an optimal experimental design, which results in accurately estimated parameter values, a target function is required. In optimal design theory, there exist different design criteria, namely, D-, A-, G-, and V-optimality criteria, which all are functions of the Fisher information matrix on the parameters. The D- and A-optimality criteria have been developed to attain precise estimation of the parameters $\boldsymbol{\beta}$, while the G- and V-optimality criteria are concerned with accurate response predictions. The A-optimality criterion aims at designs that minimize the trace of the variance-covariance matrix. The G- and V-optimality criteria are defined with respect to a design region $\chi$ consisting of all possible choice sets that can be composed from some candidate profiles. The G-optimal design minimizes the maximum prediction variance over the design region $\chi$, and the V-optimal design minimizes the average prediction variance over this region (Kessels et al., 2009).

This master thesis focuses on the D-optimality criterion, which is the most commonly used one in practice and performs well in terms of other criteria, too (Goos, 2002). The D-optimality criterion seeks to maximize the determinant of the information matrix, or to minimize its inverse, the determinant of the variance-covariance matrix of the parameter estimators.

For the MNL model, the total information matrix $\mathbf{I}$ is obtained as the sum of the information matrices of the $S$ choice sets $\mathbf{I}_s$, and is described as

$$\mathbf{I}(\mathbf{X}, \boldsymbol{\beta}) = \sum_{s=1}^{S} \mathbf{I}_s(\mathbf{X}, \boldsymbol{\beta}) = \sum_{s=1}^{S} \mathbf{X}_s'(\mathbf{P}_s - \mathbf{p}_s\mathbf{p}_s')\mathbf{X}_s, \tag{6}$$

with $\mathbf{X}_s = [\mathbf{x}_{js}']_{j=1,\ldots,J}$, $\mathbf{p}_s = [p_{1s}, \ldots, p_{Js}]'$, and $\mathbf{P}_s = \text{diag}[p_{1s}, \ldots, p_{Js}]$ (Kessels et al., 2006). From Equation (6), it can be seen that the information matrix depends on the parameter values through the probabilities, which are unknown before the analysis. If one vector for the unknown parameter values is taken, the resulting design from minimizing the inverse of the determinant of the information matrix in Equation (6) is called a locally optimal[1] design, as it is optimal for only one parameter vector.

However, if we take a prior distribution $\pi(\boldsymbol{\beta})$ of possible parameter values, the criterion expression for the design matrix $\mathbf{X} = [\mathbf{x}_{js}']_{j=1,\ldots,J;s=1,\ldots,S}$ for estimating $\boldsymbol{\beta}$ in the MNL model in Equation (1) becomes

$$\text{D}_\text{B} = \int_{\mathbf{R}^\text{P}} \{\det(\mathbf{I}^{-1}(\mathbf{X}, \boldsymbol{\beta}))\}^{1/p} \pi(\boldsymbol{\beta}) \text{d}\boldsymbol{\beta}, \tag{7}$$

[1] In this thesis, there are two different, but valid meanings of the term *locally optimal*:

- The algorithm can give *locally optimal* designs, because it is a heuristic optimization algorithm
- The designs are called *locally optimal*, because they are optimal for only one $\boldsymbol{\beta}$ vector

where the exponent 1/p ensures that it is independent of the dimension $p$ of the parameter vector $\boldsymbol{\beta}$. Minimizing this function over $\mathbf{X}$ for the prior distribution of parameter values, $\pi(\boldsymbol{\beta})$, results in the $D_B$-optimal design (Kessels et al., 2006, 2011). The criterion is denoted by $D_B$, and the approach is referred to as a semi-Bayesian approach rather than a Bayesian approach, since it does not involve the posterior distribution based on Bayes's theorem. If a degenerate distribution (i.e., the distribution which takes a single value) for $\pi(\boldsymbol{\beta})$ is assumed, we denote the criterion value by D, as no Bayesian approach is applied in this case.

# 3 Algorithmic Approach

In this section we present an algorithm that minimizes Equation (7) through a local search, in order to design an efficient experimental design for multinomial logit model that contains mixtures.

The use of more efficient designs leads to an expectation that a lower number of respondents will be needed to produce statistically significant parameter estimates when compared to less efficient designs. For instance, if the D-error of one design is 50% of the D-error of another design, it means that the design is twice as good and requires 50% fewer respondents to obtain parameter estimates, which are just as accurate as in the less efficient design.

Two complicating issues arise when obtaining an optimal design for mixtures. First, a starting design, which is feasible in the sense that its proportions satisfy the mixture constraint in Equation (2), is required. Second, an ingredient proportion value cannot be changed independently of the other proportions in a design. If one proportion changes, then at least one other one must change, in order to maintain the sum of the mixture ingredient proportions equal to one.

The algorithm for designing an efficient experimental design for multinomial logit model for mixtures is based on two steps: (1) a feasible starting design has to be generated, and (2) the mixture coordinate-exchange algorithm is applied to improve the starting design. Steps (1) and (2) are subsequently repeated user-specified number of times. The repetition may help to avoid a locally optimal, but poor design. There are many substeps and issues that arise throughout the whole process. They all are described in the following subsections, where the techniques used in each of the steps are discussed.

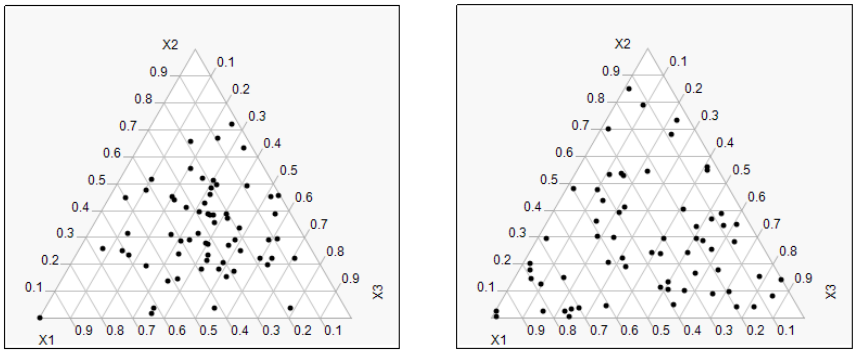
## 3.1 Generating a Starting Design

To obtain the starting design, we sample proportions uniformly from the unit simplex. Specifically, a uniform sample from the set $C = \{(c_1, c_2, \ldots, c_K) | 0 \leq c_i \leq 1, c_1 + c_2 + \cdots + c_K = 1\}$, where $K$ is the dimension of the simplex, is required. One way to obtain such a starting design could be by randomly generating numbers, which are uniformly distributed in the interval (0,1), for every ingredient proportion. Then, for every alternative, the random values have to be divided by the sum of all generated ingredient proportions for that alternative. The division is performed to obtain proportions that sum up to one. The problem with this approach is that it generates non-uniformly distributed points on the simplex.

An elegant way of obtaining a starting design with more evenly spread points is by generating independent and identically distributed random samples from an exponential distribution. It is performed as follows. First, numbers $x_i$ are sampled from (0, 1) uniformly, and values equal to $-\ln(x_i)$ are returned. This is done for $k$ samples (i.e., for every ingredient proportion), and the

resulting values are normalized by dividing each result by the sum of all proportions for a certain alternative. The resulting list of numbers is a uniform sample from the simplex (Geomblog, 2012).

How much better the latter technique covers the design region than the former one, we present in Figure 1, where two designs are plotted. The triangular graph that is used to plot designs involving three ingredients is called a ternary plot. It has lines parallel to the three sides of an equilateral triangle (see Figure 1 again). The vertices of the simplex (or triangle) represent the single-component mixtures and are denoted by $x_i = 1$, $x_j = 0$ for $i, j = 1, 2$, and 3, $i \neq j$. The interior points of the triangle represent mixtures in which none of the three components is absent, that is, $x_1 > 0$, $x_2 > 0$, and $x_3 > 0$. The centroid of the triangle corresponds to the mixture with equal proportions $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ from each of the components (Cornell, 2002).




(a) Samples, drawn pseudo-randomly from the standard uniform distribution on the interval (0, 1) and normalized afterwards

(b) Samples, generated pseudo-randomly from an exponential distribution and normalized afterwards

Figure 1: Starting designs obtained by two different sampling techniques

In Figure 1, two starting designs for a design problem consisting of 3 ingredients, 64 alternatives and 4 choices in a choice set are generated using two different methods discussed above. It can clearly be seen that the second approach generates points which are much more evenly spread on the simplex.

### 3.2 Improving the Starting Design

To improve the starting design, a mixture coordinate-exchange algorithm is used. The algorithm starts with the first ingredient proportion of the initial design and optimizes it using the method of Brent (1973), which is a one-dimensional optimization algorithm, based on a combination of golden section search and successive parabolic interpolation. The proportion is optimized in such a way, that the pairwise ratios of the remaining ingredient proportions remain fixed and the $D_B$-criterion value is minimized. Next, another proportion is optimized.

If we denote a reference mixture by $\mathbf{s} = (s_1, s_2, \ldots, s_k)$, the proportions of a point $\mathbf{x} = (x_1, x_2, \ldots, x_k)$ for a $\Delta_i$ change in the $i$th ingredient are

$$x_i = s_i + \Delta_i,$$

and

$$x_j = s_j - \frac{\Delta_i s_j}{1 - s_i} \qquad \text{for all} \quad j = 1, 2, \ldots, i-1, i+1, \ldots, k.$$

Changing proportions in such a way, which is known as exchanges along Cox effect directions (Cornell, 2002; Piepel, 1982), helps to overcome the issue that proportions in a mixture model cannot be changed independently. This makes the method different from the Meyer and Nachtsheim (1995) coordinate-exchange algorithm. If the minimal value of the objective function is smaller than the current minimum, then the current minimum is replaced and the alternative's current proportions for the ingredients are replaced by the new proportions, corresponding to the new optimum. The process proceeds till all ingredients in all rows of the starting design matrix have been considered for optimization. If any improvements are made, the entire process is repeated, starting with the first ingredient proportion for the first alternative in the new current design. The algorithm stops when no improvements have been performed in a complete pass through all the ingredients in every row of a design matrix $\mathbf{X}$.

In order to avoid ending up at a local minimum, it is advisable to repeat the entire mixture coordinate-exchange algorithm using many random starting designs. Although this does not guarantee convergence to a global optimum, it reduces the chance of finding a locally optimal design (Piepel et al., 2005).

### 3.3 The Prior Distribution

The complicating issue in the search of an optimal design is the fact that probabilistic choice models are nonlinear in the parameters, i.e., the information matrix in Equation (6) depends on the unknown parameters through the probabilities. As a consequence, researchers have to assume values for the parameters before deriving the experimental design. When a single vector is taken for the parameter values, the resulting design is locally optimal. A special case of a locally optimal design is the design for which zero parameter values are assumed. This transforms the nonlinear design problem for the MNL model, described in Equation (1), into a linear one. This assumption causes the probabilities $p_{js}$ of all $J$ alternatives in a choice set $s$ to be equal to $1/J$, which reflects a situation where respondents have no preference for any of the alternatives in a choice set. Such designs are called utility neutral optimal designs. Utility neutral designs are used as a benchmark in this thesis. They are discussed in detail in Section 4.

An alternative to opting for locally optimal designs is to rely on Bayesian techniques, where a prior distribution for the parameters is assumed. Recently it has become popular to rely on the semi-Bayesian approach that was introduced in the marketing literature by Sándor and Wedel (2001), and has been widely used for discrete choice experiments (Bliemer et al., 2009; Kessels et al., 2006, 2009, 2011, among others). It is a more robust strategy as it averages a design criterion over a prior distribution of likely parameter values, $\pi(\boldsymbol{\beta})$. As Sándor and Wedel (2001) showed, even if priors are misspecified (or not well specified), more efficient experimental designs are obtained, since the prior uncertainty is taken into account. A semi-Bayesian $D_B$-optimal design minimizes Equation (7) for the assumed prior parameter distribution.

However, in order to choose a good $\pi(\boldsymbol{\beta})$, identification of parameters has to be considered. When mixtures are embedded in MNL models, all proportions of different ingredients sum up to one. As a result, due to the identification issue, parameters for ingredient proportions cannot be estimated independently, but only with respect to a base parameter.

Let us consider a numerical example for a problem with three ingredients. In such a case, prior parameter values for the ingredient proportions, $\tau_i$, $i = 1, 2, 3$, are generated with respect to the last parameter, $\tau_3$, and we have

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} \tau_1 - \tau_3 \\ \tau_2 - \tau_3 \end{pmatrix}, \quad \boldsymbol{\beta} \sim \pi(\boldsymbol{\beta}), \tag{8}$$

where $\pi(\boldsymbol{\beta})$ is a prior distribution over which the $D_B$-criterion value in Equation (7) is integrated.

It is easier to start by specifying a prior distribution for all the parameters, $\tau_i$, $i = 1, 2, 3$, though. Usually it is chosen to be a multivariate normal distribution. There are different ways to get mean and variance-covariance matrix for the prior distribution of parameter values $\tau_i$, $i = 1, 2, 3$, in Equation (8). Huber and Zwerina (1996) argue that a set of reasonable and useful priors might be obtained through a small pilot test that provides coefficients in need. An educated guess of experienced managers may also be an option. Any other way can be used to obtain prior mean values for the parameters and associated uncertainty with them, too. If a prior distribution is assumed to be a multivariate normal distribution with mean $\boldsymbol{\tau}_0 = (\tau_{01}, \tau_{02}, \tau_{03})'$, and variance-covariance matrix a diagonal matrix with ones on the diagonal, we have

$$\begin{pmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \end{pmatrix} \sim \mathcal{N}\left( \begin{pmatrix} \tau_{01} \\ \tau_{02} \\ \tau_{03} \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right) = \mathcal{N}(\boldsymbol{\tau} | \boldsymbol{\tau}_0, \boldsymbol{\Sigma}_0') = \pi(\boldsymbol{\tau}),$$

which, according to the reasoning above, boils down to

$$\begin{pmatrix} \tau_1 - \tau_3 \\ \tau_2 - \tau_3 \end{pmatrix} \sim \mathcal{N}\left( \begin{pmatrix} \tau_{01} - \tau_{03} \\ \tau_{02} - \tau_{03} \end{pmatrix}, \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \right),$$

and, eventually, we get

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} \tau_1 - \tau_3 \\ \tau_2 - \tau_3 \end{pmatrix} \sim \mathcal{N}\left( \begin{pmatrix} \tau_{01} - \tau_{03} \\ \tau_{02} - \tau_{03} \end{pmatrix}, \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \right) = \mathcal{N}(\boldsymbol{\beta}|\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_0) = \pi(\boldsymbol{\beta}).$$

In such a way the values for the parameters are generated not independently but with respect to one of the parameters, namely, the last one, $\tau_3$, and identification of the parameters is assured.

In the subsequent parts of the thesis, the initial distribution for the parameter values (i.e., the one which can be obtained, say, by interviewing managers) is denoted by $\pi(\boldsymbol{\tau}) = \mathcal{N}(\boldsymbol{\tau}|\boldsymbol{\tau}_0, \boldsymbol{\Sigma}_0')$, while the one, which it boils down to, is denoted by $\pi(\boldsymbol{\beta}) = \mathcal{N}(\boldsymbol{\beta}|\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_0)$.

### 3.4 A Multi-Dimensional Integral for Design Selection

The integral that serves as the design selection criterion is often high dimensional and has to be evaluated many times during the search for an optimal design. If the prior parameter distribution is taken to be a multivariate one with mean $\boldsymbol{\beta}_0$ and variance-covariance matrix $\boldsymbol{\Sigma}_0$, the expectation of a design selection criterion represented in Equation (7) becomes

$$\mathrm{D_B} = \int_{\mathbf{R^p}} \{\det(\mathbf{I}^{-1}(\mathbf{X}, \boldsymbol{\beta}))\}^{1/p} \pi(\boldsymbol{\beta}) \mathrm{d}\boldsymbol{\beta}$$

$$= \int_{\mathbf{R^p}} \{\det(\mathbf{I}^{-1}(\mathbf{X}, \boldsymbol{\beta}))\}^{1/p} (2\pi)^{-\frac{p}{2}} |\boldsymbol{\Sigma}_0|^{-\frac{1}{2}} e^{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)' \boldsymbol{\Sigma}_0^{-1} (\boldsymbol{\beta}-\boldsymbol{\beta}_0)} \mathrm{d}\boldsymbol{\beta}, \tag{9}$$

where $p$ represents the number of parameters.

Computation of the integral in Equation (9) is complex, since it cannot be evaluated analytically. Hence, it has to be approximated numerically: draws for $\boldsymbol{\beta}$ are taken from a multivariate normal prior distribution and the values of the integrand in $\mathrm{D_B}$ are averaged over all draws. In order to sample from a multivariate normal distribution, we transform univariate standard normal draws into the multivariate normal ones. This transformation is done by using Cholesky decomposition of the prior distribution's covariance matrix: $\boldsymbol{\Sigma}_0 = \mathbf{D}\mathbf{D}'$. It leads to $\boldsymbol{\beta} = \boldsymbol{\beta}_0 + \mathbf{D}\boldsymbol{\nu}$, where the vector $\boldsymbol{\nu}$ has elements drawn from independent standard normal distributions. We have

$$(\boldsymbol{\beta} - \boldsymbol{\beta}_0)' \boldsymbol{\Sigma}_0^{-1} (\boldsymbol{\beta} - \boldsymbol{\beta}_0) = \boldsymbol{\nu}' \mathbf{D}' (\mathbf{D}')^{-1} \mathbf{D}^{-1} \mathbf{D} \boldsymbol{\nu} = \boldsymbol{\nu}' \boldsymbol{\nu} = \sum_{j=1}^{p} \nu_j^2,$$

where $\nu_j$ is the $j$th element of $\boldsymbol{\nu}$, and

$$\mathrm{d}\boldsymbol{\beta} = |\mathbf{D}|\mathrm{d}\boldsymbol{\nu} = |\boldsymbol{\Sigma}_0|^{\frac{1}{2}}\mathrm{d}\boldsymbol{\nu}.$$

Equation (9) can then be written as

$$\begin{aligned} \mathrm{D_B} &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \{\det(\mathbf{I}^{-1}(\mathbf{X}, \boldsymbol{\beta}_0 + \mathbf{D}\boldsymbol{\nu}))\}^{1/p} (2\pi)^{-\frac{p}{2}} \prod_{j=1}^{p} e^{-\frac{\nu_j^2}{2}} \mathrm{d}\nu_1 \dots \mathrm{d}\nu_p \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \{\det(\mathbf{I}^{-1}(\mathbf{X}, \boldsymbol{\beta}_0 + \mathbf{D}\boldsymbol{\nu}))\}^{1/p} \prod_{j=1}^{p} \phi(\nu_j) \mathrm{d}\nu_1 \dots \mathrm{d}\nu_p, \end{aligned} \tag{10}$$

where $\phi(\nu_j) = (2\pi)^{-\frac{1}{2}} e^{-\frac{\nu_j^2}{2}}$ is the density function corresponding to a univariate standard normal distribution.

To approximate the integral in Equation (10), $R$ draws are taken from $\phi(\nu_j)$ for each of the $p$ parameters. The $r$th draw for the $j$th parameter, $\nu_j^r$, is computed as $\nu_j^r = \Phi^{-1}(u_j^r)$, where $\Phi(\cdot)$ is the standard normal cumulative distribution function and $u_j^r$, $r = 1, 2, \dots, R$, is a set of points sampled from a uniform distribution on the interval (0, 1). The integral in Equation (10) is then approximated by

$$\mathrm{D_B} = \frac{1}{R} \sum_{r=1}^{R} \{\det(\mathbf{I}^{-1}(\mathbf{X}, \boldsymbol{\beta}_0 + \mathbf{D}\Phi^{-1}(\mathbf{u}^r))\}^{1/p} = \frac{1}{R} \sum_{r=1}^{R} \{\det(\mathbf{I}^{-1}(\mathbf{X}, \boldsymbol{\beta}_0 + \mathbf{D}\boldsymbol{\nu}^r)\}^{1/p},$$

where $\boldsymbol{\nu}^r = \Phi^{-1}(\mathbf{u}^r) = [\Phi^{-1}(u_1^r), \dots, \Phi^{-1}(u_p^r)]'$ (Yu et al., 2010). From this expression, one can clearly see that, as the number of draws $R$ increases, the computation time required to evaluate the $\mathrm{D_B}$-criterion value also increases. This increase is linear with respect to the number of samples as the function has to be evaluated for each of the $R$ draws. To make the algorithm computationally less intensive, the draws $(\mathbf{u}^r)_{r=1}^R$ are obtained using Halton sequences, described in the next section.

As Cholesky decomposition is as twice as fast in calculating a determinant of a symmetric, positive definite matrix, it is chosen to use it when calculating the determinant in Equation (7). If $\mathbf{A}$ has real entries and is symmetric and positive definite, then it can be decomposed as

$$\mathbf{A} = \mathbf{L}\mathbf{L}^T,$$

where $\mathbf{L}$ is a lower triangular matrix with strictly positive diagonal entries, and $\mathbf{L}^T$ denotes the transpose of $\mathbf{L}$. The determinant in Equation (7) is equal then to the square of the product of the diagonal elements of the Cholesky factor $\mathbf{L}$ (Weisstein, 2012).

### 3.5 Systematic Sampling Using Halton Sequences

In the semi-Bayesian approach, numerous draws are required from the prior parameter distribution for constructing efficient stated choice designs. The semi-Bayesian $D_B$-criterion value is calculated as the average of all the $D_B$-error values over the draws. The most common way to take the draws is the Pseudo-Monte Carlo sampling. However, this is a very non-systematic approach as the draws are sampled independently of each other which might make the samples to be unevenly scattered. As a result, this method leads to a large variability in the results, especially when the number of random draws is small, as different sets of them are likely to produce different coverage of the distribution space (Bliemer et al., 2009). To reduce this lack of stability, researchers often use a large number of samples. However, even though this method is easy to implement, the computation time for the expected $D_B$-error increases linearly with the required number of draws for evaluating the multi-dimensional integral (Yu et al., 2010).

To make the sampling more systematic, it is performed using deterministic numbers called Quasi-Monte Carlo (QMC) samples rather than computer-generated pseudo-random numbers. Examples of such approaches are (1) Halton sequences, (2) Faure sequences, (3) modified Latin hypercube sampling, (4) extensible shifted lattice points, (5) a Gauss-Hermite quadrature approach, and (6) a method using spherical-radial transformations. In this master thesis, Halton sequences are chosen. QMC samples are more evenly scattered throughout the integration domain which helps to improve the accuracy of the integral approximation (Yu et al., 2010). The main advantage of opting for a systematic sampling scheme, such as Halton sequences, is that many fewer draws, and, hence, much smaller computing times, are required to compute the integral.

Halton sequences (Halton, 1960) are constructed according to a deterministic method which is based on prime numbers and where a different prime number (base) is utilized for every dimension. In each dimension, the $n$th element in the Halton sequence based on a prime number $b$ is generated by expanding $n$ in terms of the base $b$, according to the formula

$$n = \sum_{s=0}^{m} i_s b^s = i_0 b^0 + i_1 b^1 + i_2 b^2 + \cdots + i_m b^m,$$

where $i_s \in \{0, \ldots, b-1\}$ $(s = 0, 1, \ldots, m)$ and so choice of $i_s$ depends on $n$, and $m$ is chosen so that $b^m \leq n < b^{m+1}$. The $n$th element of the one-dimensional Halton sequence based on prime $b$ is then obtained as

$$\phi_b(n) = \sum_{s=0}^{m} i_s b^{-(s+1)} = i_0 b^{-1} + i_1 b^{-2} + \cdots + i_m b^{-(m+1)}. \tag{11}$$

Equation (11) gives a Halton sequence which is uniformly distributed in $(0, 1)$. A $p$-dimensional

Halton sequence is constructed by combining $p$ one-dimensional sequences based on $p$ consecutive primes $b_1, b_2, \ldots b_p$. The $n$th $p$-dimensional Halton draw, $\mathbf{x}_n$, is

$$\mathbf{x}_n = \left(\phi_{b_1}(n), \phi_{b_2}(n), \ldots, \phi_{b_p}(n)\right), \quad n = 1, 2, \ldots.$$

As an example, consider the sequence of $R = 7$ points being computed by taking base 2. For $n = 4$, $m$ is determined to be equal to 2, as $2^2 \leq 4 < 2^{2+1}$. We then can express the integer $n = 4$ as $0 \times 2^0 + 0 \times 2^1 + 1 \times 2^2$. The fourth element of the sequence is, hence, equal to $\phi_2(4) = 1 \times 2^{-3} = 1/8$. The first element of the sequence is obtained by $\phi_2(1) = 1 \times 2^{-1} = 1/2$, while the remaining ones are equal to $\phi_2(2) = 0 \times 2^{-1} + 1 \times 2^{-2} = 1/4$, $\phi_2(3) = 3/4$, $\phi_2(5) = 5/8$, $\phi_2(6) = 3/8$ and $\phi_2(7) = 7/8$, respectively (Yu et al., 2010). It is clear that the points are generated in a very systematic way and cover the (0, 1) interval evenly.

# 4 Benchmark Approach

In this section, designs that are taken as benchmark designs in this master thesis are discussed.

The approach that we develop in this thesis can be complex, as the choice model is non-linear in parameters. This makes it necessary to assume prior parameter values and uncertainty about them, which is not simple. In order to avoid such difficulties encountered by non-linear choice models, the assumption that $\boldsymbol{\beta} = \mathbf{0}$ could be used, which makes the design problem linear and simplifies it. In such cases, standard software for generating optimal designs can be used to construct a design.

The most similar designs among them to the problems analyzed here are known as mixture experiments in blocks. Block designs for mixture experiments are groups of mixture blends where each group or block is assumed to differ from the other groups or blocks by an additive constant, which captures variation across "trials". Examples of blocks in an industrial environment are different vendors supplying the raw material, or different shifts of plant personnel running the experiments, or separate technicians and/or laboratories performing the experiments (Cornell, 2002). In our choice setting, the blocks can be seen as choice sets.

In a general setting with fixed blocking variables and a linear regression model, the statistical model corresponding to a blocked mixture experiment can be written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon}, \tag{12}$$

where $\mathbf{X}$ is the $(J \times S) \times p$ design matrix corresponding to the components of the mixture, and $\mathbf{Z}$ is the design matrix corresponding to the indicator variables for the fixed blocks. The vectors $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\epsilon}$ represent the mixture variable coefficients, the fixed block effects, and the random errors, respectively (Goos and Donev, 2007).

Kessels et al. (2011) explain that blocked designs for the linear model are essentially the same as utility neutral choice designs, and that optimizing these designs gives exactly the same result. Thus, under one scenario, namely, when $\boldsymbol{\beta} = \mathbf{0}$, optimal designs for blocked experiments and utility neutral choice experiments are equal. This gives us a possibility to compare two solutions and check to what extent our algorithm is able to replicate the results obtained by standard techniques.

Consider the utility neutral design problem for 3 ingredients, 18 alternatives and one choice set (so no blocks are present) for a special-cubic model (i.e., when all cross terms are included in the model). The D-optimal design for a linear model is constructed by JMP®[2] software and depicted in Figure 2. It is in accordance with the well known fact that, if there is one block or choice set,

[2] www.jmp.com

D-optimal $J \times S$-alternative designs, where $J$ is the number of choice sets in the design, and $S$ is the number of choices in a choice set, for models that can be described by Scheffé canonical polynomial models have a minimum support, i.e., they are designs which have observations at as many distinct combinations of the experimental variables as there are parameters in the statistical model to be estimated. The points are replicated $\text{int}(J \times S)/p$ or $\text{int}(J \times S)/p + 1$ times, where $p$ is the number of columns in the design matrix $\mathbf{X}$. For the first-order Scheffé model, the D-optimal points are the $k$ vertices of the simplex. For the second-order model, the $k + k(k-1)/2$ D-optimal points are the vertices and the edge midpoints (Goos and Donev, 2006). In case of the special-cubic model with three ingredients, the middle point of the simplex is added to the optimal ones. By having a symmetric design with observations at the corners of the design region and at or close to the midpoints of the edges, D-optimal designs cover the entire experimental region well.

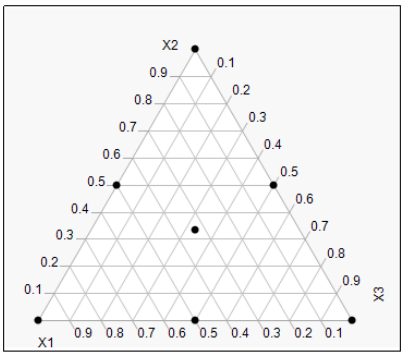
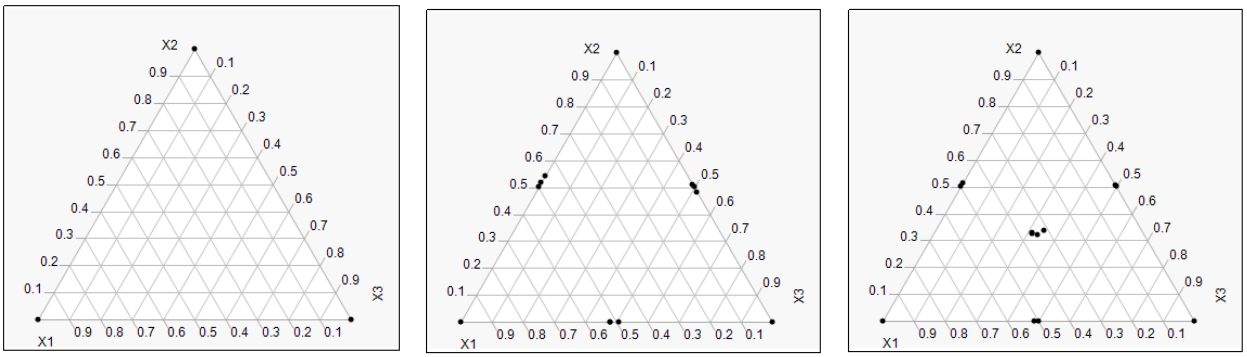



Figure 2: Utility neutral optimal design for a special-cubic model with 3 ingredients, 18 alternatives and one choice set

The same results should be obtained when using the algorithm developed in this master thesis. To test this, we consider the utility neutral design as before with 3 ingredients, 18 alternatives and one choice set. All three possible scenarios are investigated, namely, when only main effects are taken into account, and when two- and three-ingredient interactions are included into the model. The design is optimized for 10,000 random starts. We present the designs obtained for every case mentioned before in Figures 3a, 3b, and 3b, respectively. Unfortunately, the algorithm is not accurate enough to arrive at single points on edges and in the middle of the simplex for 10,000 random starts. It is also not fast enough to be tried with highly greater numbers of them. However, convergence to the optimal points can be seen when increasing the number of random starts. The exact points of the designs in Figure 3 are given in Tables 15, 16, and 17 in the Appendix.

To get an insight about the speed of the algorithm, it took around 39 minutes to optimize the design for 3 ingredients, 18 alternatives and one choice set with two- and three-ingredient




(a) Main effects only

(b) Main effects and two-ingredient interactions

(c) Main effects, two- and three-ingredient interactions

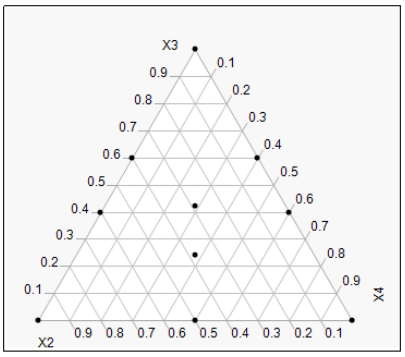
Figure 3: Utility neutral optimal designs for 10,000 random starts for a model with 3 ingredients, 18 alternatives and one choice set when the number of ingredient cross terms included in the model differs

interactions for 10,000 random starts, when utility neutral design was assumed. All computations were performed in MATLAB 7.12.0 using an ASUS laptop with 2.20 GHz Intel Core i7 processor and 4 GB RAM.

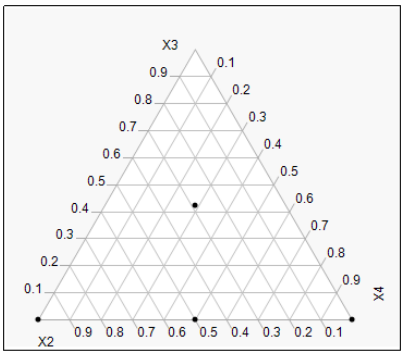
However, when we consider blocks (or choice sets) in the design, it is possible that using vertices, midpoints and centroids is no longer optimal. There are no theoretical results that say what blocked designs should look like, and the same goes for the utility neutral designs.

When we assume experimental designs with blocks (or choice sets), it is interesting to investigate the division into choice sets. Both JMP and our algorithm provide this. Consider the next example, which divides the optimal design points into choice sets. The design we consider is the utility neutral design for 3 ingredients, 12 alternatives and 4 choices in a choice set (hence, fixed blocks of size 4 are assumed), when two-ingredient and three-ingredient interaction terms are also included in the model. Figures 4b - 4d give the three choice sets of the optimal design, optimized using the approach of this thesis, while Figure 4a plots all the optimal points in one figure. By analyzing these plots we are able to see which mixtures in an optimal experiment are grouped together into choice sets. The points of the design can be seen in Table 13 in the Appendix.

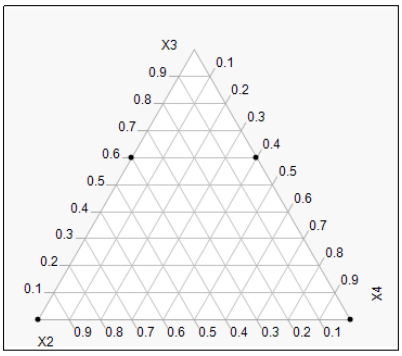
The algorithm presented in this thesis is able to replicate results of a standard software (namely, JMP), as it can be seen from the examples above. However, it can be used only for linear mixture (and the utility neutral choice) models, and when it is not the case, such techniques cannot be applied any more. Then, the algorithm presented in this thesis has to be employed.



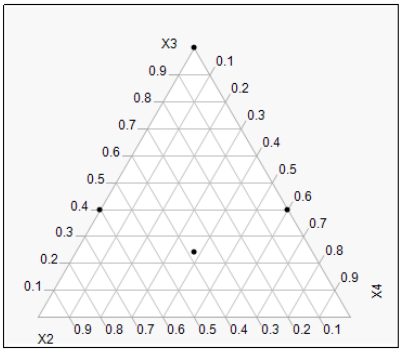

(a) All choice sets




(b) First choice set




(c) Second choice set




(d) Third choice set

Figure 4: Utility neutral optimal design for 3 ingredients, 12 alternatives and 4 choices in a choice set, when division into choice sets is ignored and also when alternatives are divided into three choice sets, when the utility neutral design is assumed

# 5 Results

In this section, we consider a number of different settings for experimental designs and discuss the results. First, we present some basic features for the designs analyzed. Then, we show that D-optimal designs when prior parameter values are not assumed to be zero differ substantially from the utility neutral designs. Further, we show that semi-Bayesian designs differ from and perform better than locally optimal ones (and the utility neutral designs) for most of the time. Since it is often difficult to obtain reasonable prior values for the parameters, their misspecification is also investigated.

Sections 5.1-5.3 consider locally optimal experimental designs, while Section 5.4 and Section 5.5 compare semi-Bayesian designs to the locally optimal ones.

## 5.1 Basic Features

In this section, we investigate how the D-value is affected by different characteristics of the experiment for utility neutral optimal experimental designs. For this reason, we vary some of them while generating optimal experiments. First, the total number of alternatives has been increased from 12 to 24 for four different design problems. The results can be seen in Table 1. The last four columns in the table represent the four different designs. The first column gives the number of alternatives. The notations for the column names should be read as follows, 2/2/2 int. means that the design is optimized for 2 ingredients, 2 choices in a choice set when two-ingredient interactions are included. The designs are optimized for 100 random starts assuming the utility neutral model. It is clear that when the number of alternatives increases, the D-value decreases, meaning that we are able to generate more efficient optimal designs, which is rather intuitive, as we obtain more information by adding extra observations. The pattern stays the same for many other design variations. The dashes for the design with 3 ingredients, 3 choices in a choice set and main effects only (column 3) mean that such designs cannot be generated, as the number of choices in a choice set does not divide the corresponding numbers of alternatives evenly.

| Number of alternatives | 2/2/2 int. | 2/4/2 int. | 3/3/1 int. | 3/4/2 int. |
|---|---|---|---|---|
| 12 | 2.7267 | 3.8474 | 1.1907 | 10.5596 |
| 16 | 2.2466 | 3.1765 | - | 8.4183 |
| 20 | 1.9240 | 2.7357 | - | 6.9376 |
| 24 | 1.7099 | 2.4239 | 0.7501 | 5.9467 |

Table 1: D-criterion value for different number of alternatives in the utility neutral design, 100 random starts

A similar intuitive pattern should be present when the number of random starts is increased.

The designs from Table 1 are taken and optimized for 1,000 random starts instead of 100. Table 2 shows the resulting D-values. The reasoning holds true, the D-values for 1,000 random starts are lower than the ones when 100 random starts are considered. The exception is the design with 3 ingredients, 3 choices in a choice set and main effects only; the D-values for it are equal for both cases. But they are never larger than the ones for 100 random starts. Besides, the pattern discussed previously is also the case here, the D-values decrease with the number of alternatives in the design.

| Number of alternatives | 2/2/2 int. | 2/4/2 int. | 3/3/1 int. | 3/4/2 int. |
|---|---|---|---|---|
| 12 | 2.7263 | 3.8474 | 1.1907 | 10.5481 |
| 16 | 2.2445 | 3.1633 | - | 8.3939 |
| 20 | 1.9235 | 2.7216 | - | 6.9182 |
| 24 | 1.7038 | 2.4211 | 0.7501 | 5.9003 |

Table 2: D-criterion value for different number of alternatives in the utility neutral design, 1,000 random starts

By comparing D-criterion values in Tables 1 and 2, we can notice that they tend to converge and the difference in them is pretty small when we move from 100 random starts to 1,000. Thus, it might be interesting to explore the change in D-values when the number of random starts is increased. Figure 5 provides such information, where we have the logarithm of the number of random starts on the horizontal axis and the D-value on the vertical axis. This investigation is performed for experimental problem consisting of 3 ingredients, 24 alternatives and 4 choices in a choice set, when two- and three-ingredient interactions are included in the model and the utility neutral design is assumed. From Figure 5 it is clear that the largest decrease in D-criterion appears at the beginning, when we move from one random start to a little bit higher numbers of them. However, this decrease is not that salient when we have a larger number of random starts and D-values tend to converge. The exact D-values obtained are provided in Table 3.

| # of random starts | D-value | # of random starts | D-value |
|---|---|---|---|
| 1 | 6.0146 | 20 | 5.9437 |
| 2 | 5.9906 | 50 | 5.9302 |
| 3 | 5.9840 | 100 | 5.9173 |
| 4 | 5.9565 | 1,000 | 5.9003 |
| 5 | 5.9551 | 10,000 | 5.8930 |
| 10 | 5.9544 | 100,000 | 5.8864 |
| 15 | 5.9466 | | |

Table 3: D-criterion value for a different number of random starts

If the number of alternatives is held constant at 24 alternatives and only the choice set size is

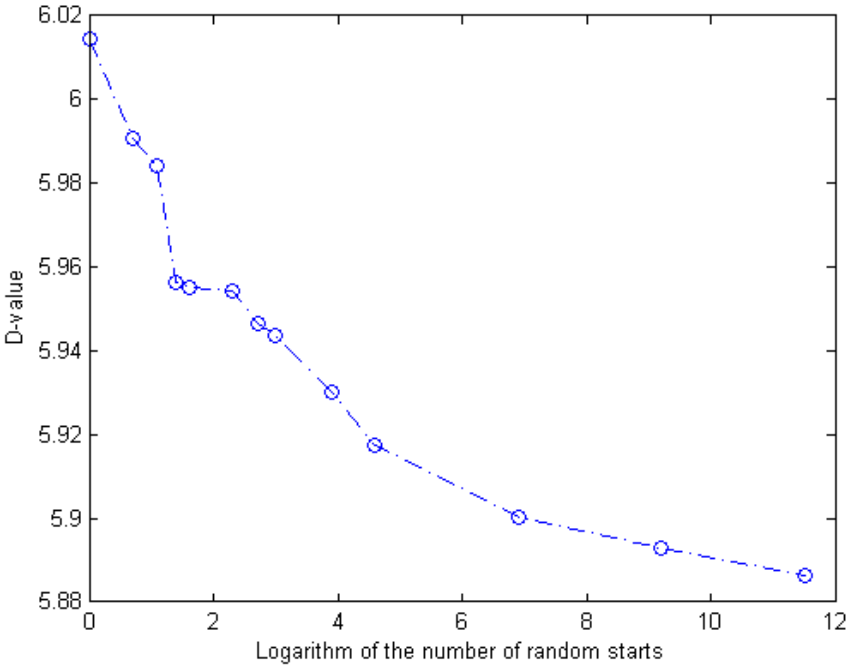



Figure 5: D-criterion values for a different number of random starts

| | D-value | | D-value |
|---|---|---|---|
| 2/2/2 int. | 1.7071 | 3/2/3 int. | 11.4188 |
| 2/3/2 int. | 1.8906 | 3/3/3 int. | 12.5391 |
| 2/4/2 int. | 2.4243 | 3/4/3 int. | 14.6056 |
| 2/6/2 int. | 3.2001 | 3/6/3 int. | 19.1001 |
| 2/8/2 int. | 3.8433 | 3/8/3 int. | 23.7783 |
| 2/12/2 int. | 5.0412 | 3/12/3 int. | 35.4455 |

Table 4: D-criterion value for different sizes of a choice set, 24 alternatives, 100 random starts, when the utility neutral design is assumed

varied (2, 3, 4, 6, 8 or 12 choices in a set), the D-value increases. An example of this is given in Table 4, where 12 designs are considered. Their names are coded in the same manner as before, and they are optimized for 100 random starts assuming the utility neutral design. More efficient designs are obtained for smaller choice sets. For the extreme case, where only one choice set with 24 choices in it is assumed, the D-values are very high, 8.2655 for the two-ingredient problem and 63.9157 for the three-ingredient problem.

The result obtained means that we have more information in smaller sets, which can be explained by the following example. Assume that we have three alternatives, A, B, and C, and two different choice designs. The first one consists of two choice sets of size two (A and B, B and C), and the second one of only one choice set, which is larger and contains all three alternatives

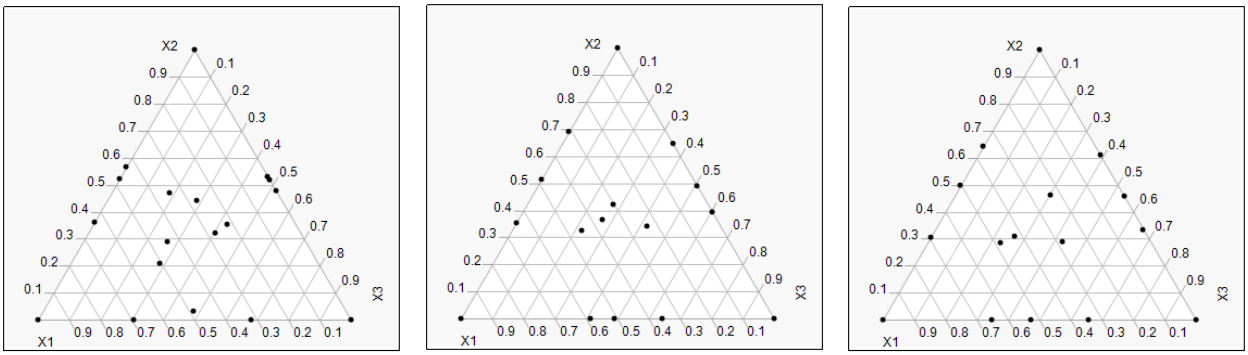
(A, B, C). Let us say that we get the following information after the experiment, there have been A and B chosen from the two choice sets, respectively, under the first design, and A, under the second one. The results imply that A is preferred over B, and B is preferred over C for the first design, while A is preferred over both B and C for the second one. We have no information what relationship exists between B and C under the second scenario. Meanwhile, under the first scenario, from $A \succ B$ and $B \succ C$, we get $A \succ C$. Thus, there is more information under the first experimental design, where we have smaller choice sets, what could explain why the designs obtained for smaller choice sets are more efficient.

## 5.2 Comparison With Utility Neutral Designs

The algorithm presented in this master thesis can generate optimal experiments not only for utility neutral designs but also for locally optimal designs, i.e., the designs, where parameter values $\boldsymbol{\tau}$ are not assumed to be zero. As it is always complicating to find accurate prior parameter values, it is important to investigate how the utility neutral designs differ from the locally optimal ones. This is done in this section.

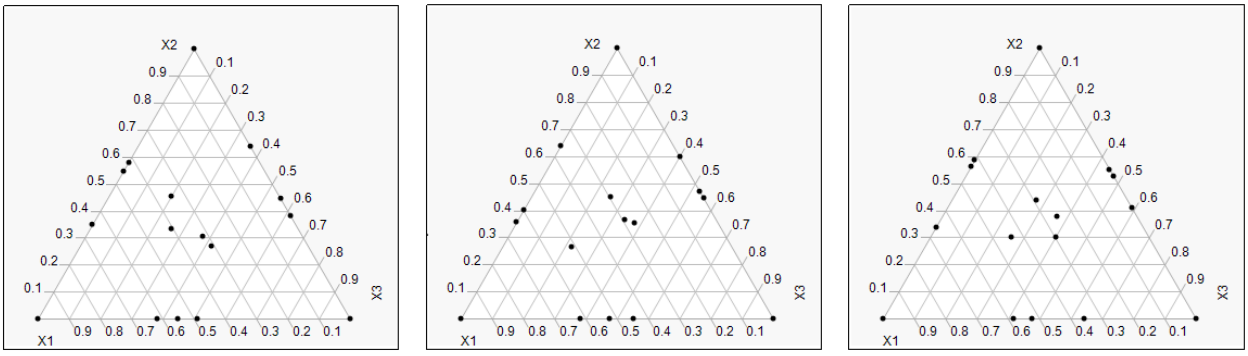
The way in which the utility neutral design differs in look from locally optimal designs can be seen by comparing Figure 6c (the utility neutral design) and Figure 7 (Figures 6a and 6b give designs for smaller numbers of random starts for the same design as in Figure 6c). Designs in Figures 6c and 7 are generated using 100,000 random starts for 3 ingredients, 20 alternatives, and 2 choices in a choice set. Two-ingredient and three-ingredient interactions are included in the model. The design in Figure 6c is obtained by assuming a linear model, while the ones in Figures 7a, 7b, and 7c are locally optimal ones obtained by replacing one, two and all elements in the parameter vector $\boldsymbol{\tau}$ by 1.4, respectively. The exact optimal points can be inspected in Tables 18, 19, and 20 in the Appendix. In Figures 8 and 9 the designs obtained by a more "drastic" move of the parameter vector $\boldsymbol{\tau}$ are displayed, i.e., instead of 1.4 in the aforementioned example, the values of 5 and 10 are used, respectively. The exact optimal points for the three designs under two different situations can be inspected in Tables 21, 22, and 23 for the value of 5, and in Tables 24, 25, and 26 for the value of 10 in the Appendix.

The further parameter vectors from zero, the more distinct in look the resulting designs are. The observed pattern is that with $\boldsymbol{\tau}$ getting further from $\mathbf{0}$ design points become more scattered and, thus, more "important", what is logical, as by assigning higher values to the parameters representing different alternatives, we are not imposing equal preferences over the alternatives any more. This higher scatter is observed both on the edges and in the middle of the design simplex. Designs get less and less symmetric. Also, more points move towards the center of the experimental region. By subsequently inspecting Figures 6c, 7a, 8a, and 9a, which are obtained




(a) 100 random starts (b) 10,000 random starts (c) 100,000 random starts

Figure 6: Utility neutral designs for a different number of random starts, for the design problem with 3 ingredients, 20 alternatives, and 2 choices in a choice set

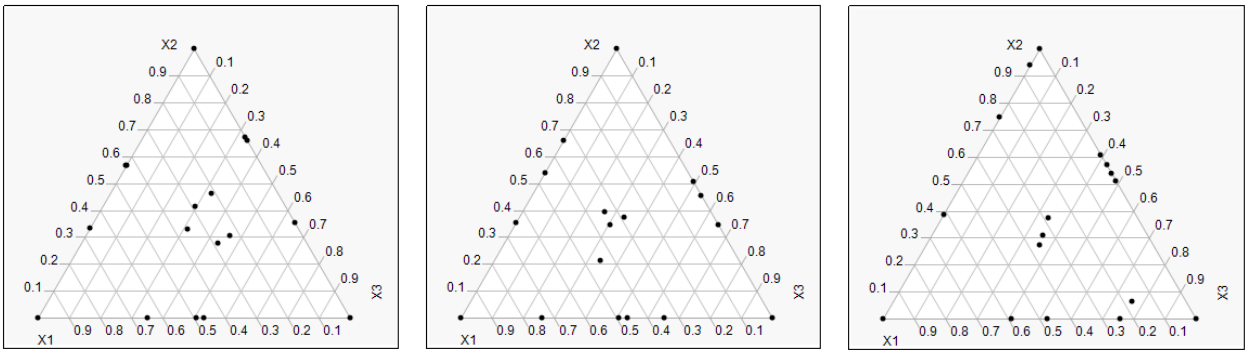



(a) $\boldsymbol{\tau} = (1.4, 0, 0, 0, 0, 0, 0)$ (b) $\boldsymbol{\tau} = (1.4, 1.4, 0, 0, 0, 0, 0)$ (c) $\boldsymbol{\tau} = (1.4, 1.4, 1.4, 1.4, 1.4, 1.4, 1.4)$

Figure 7: Designs for 100,000 random starts when $\boldsymbol{\tau}$ is being moved away from $\mathbf{0}$, for the design problem with 3 ingredients, 20 alternatives, and 2 choices in a choice set

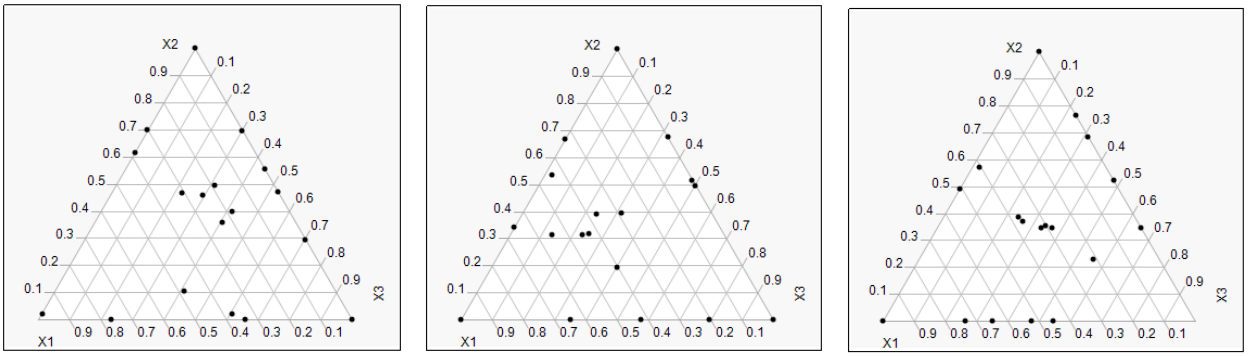
by moving the first element in $\boldsymbol{\tau}$ further and further from 0, it is clearly seen that optimal points move to the right on the simplex. It seems rather intuitive, as by increasing the value for the first element in parameter vector $\boldsymbol{\tau}$ we let the first ingredient dominate. As a result, the optimal design found proposes those mixtures, where the proportion for the first ingredient is lower (i.e., the right part of the experimental region), so we can learn more from such experiments. Moreover, for the designs with parameter vectors which are furthest away from the zero vector (i.e., when $\boldsymbol{\tau} = (5, 5, 5, 5, 5, 5, 5)$ and $\boldsymbol{\tau} = (10, 10, 10, 10, 10, 10, 10)$), one corner point is not part of the optimal design any more.

For an even more extreme case, when zero elements in $\boldsymbol{\tau}$ are exchanged by 30, the designs are given in Figure 10. In this instance, the corner point representing (1, 0, 0) is not optimal any longer for the design with $\boldsymbol{\tau} = (30, 0, 0, 0, 0, 0, 0)$, and all the optimal points move to the right,




(a) $\boldsymbol{\tau} = (5, 0, 0, 0, 0, 0, 0)$ (b) $\boldsymbol{\tau} = (5, 5, 0, 0, 0, 0, 0)$ (c) $\boldsymbol{\tau} = (5, 5, 5, 5, 5, 5, 5)$

Figure 8: Designs for 100,000 random starts when $\boldsymbol{\tau}$ is being moved away from $\mathbf{0}$ more drastically, for the design problem with 3 ingredients, 20 alternatives, and 2 choices in a choice set

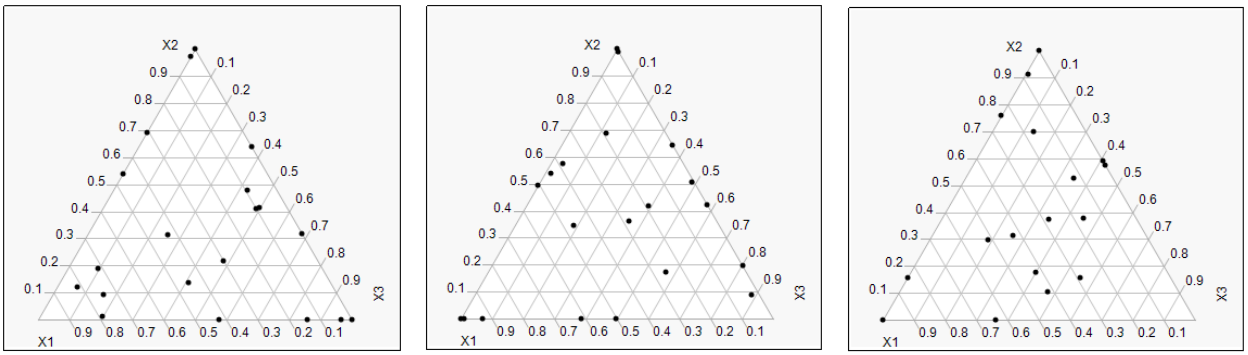



(a) $\boldsymbol{\tau} = (10, 0, 0, 0, 0, 0, 0)$ (b) $\boldsymbol{\tau} = (10, 10, 0, 0, 0, 0, 0)$ (c) $\boldsymbol{\tau} = (10, 10, 10, 10, 10, 10, 10)$

Figure 9: Designs for 100,000 random starts when $\boldsymbol{\tau}$ is being moved away from $\mathbf{0}$ more drastically, for the design problem with 3 ingredients, 20 alternatives, and 2 choices in a choice set

giving even lower proportion values to the first ingredient, what follows from the reasoning above. When the second proportion is also allowed to dominate together with the first one by choosing $\boldsymbol{\tau} = (30, 30, 0, 0, 0, 0, 0)$ vector for parameter values, the corner point (0, 0, 1) is also not among the optimal ones any more. The exact design points for the three cases are given in Tables 27, 28, and 29 in the Appendix.

The reasoning that when a certain ingredient is dominating, the according proportion is lower for it in the optimal design is only partially supported, when actual numbers are considered. Table 5 gives certain characteristics of the distribution of the proportions when different parameter values are assumed. The first column gives measures of characteristics when the parameter vector consists of zeros only, except for the first parameter, the value for which varies from 1.4 to 30, while in the second column the first two elements of the vector are varied from 1.4 to 30. For every




(a) $\boldsymbol{\tau} = (30, 0, 0, 0, 0, 0, 0)$ (b) $\boldsymbol{\tau} = (30, 30, 0, 0, 0, 0, 0)$ (c) $\boldsymbol{\tau} = (30, 30, 30, 30, 30, 30, 30)$

Figure 10: Designs for 100,000 random starts when $\boldsymbol{\tau}$ is being moved away from $\mathbf{0}$ more drastically, for the design problem with 3 ingredients, 20 alternatives, and 2 choices in a choice set

parameter value that is different from zero (i.e., 1.4, 5, 10, 30), three different characteristics are provided, namely, the mean of an ingredient proportion in the optimal design, and the percentage of proportions that are lower than 0.5 and 0.4. Only the first ingredient is considered for the parameter vector when only one element in it is different from zero, and all three ingredients in the second case.

| | | $(\tau, 0, 0, 0, 0, 0, 0)$ | $(\tau, \tau, 0, 0, 0, 0, 0)$ | | |
|---|---|---|---|---|---|
| $\tau$ | | $x_1$ | $x_1$ | $x_2$ | $x_3$ |
| 1.4 | mean | 0.3276 | 0.3169 | 0.3661 | 0.3169 |
| | $< 0.5$ | 75% | 80% | 70% | 70% |
| | $< 0.4$ | 55% | 60% | 55% | 55% |
| 5 | mean | 0.3255 | 0.3190 | 0.3618 | 0.3191 |
| | $< 0.5$ | 80% | 75% | 75% | 70% |
| | $< 0.4$ | 60% | 60% | 55% | 60% |
| 10 | mean | 0.2389 | 0.3458 | 0.3384 | 0.3158 |
| | $< 0.5$ | 90% | 70% | 75% | 70% |
| | $< 0.4$ | 85% | 65% | 60% | 65% |
| 30 | mean | 0.2958 | 0.3352 | 0.3280 | 0.3368 |
| | $< 0.5$ | 80% | 70% | 65% | 80% |
| | $< 0.4$ | 60% | 70% | 65% | 65% |

Table 5: Characteristics of distributions of ingredient proportions in optimal designs for different parameter values

From Table 5 we can see that for the parameter vector $(\tau, 0, 0, 0, 0, 0, 0)$, where $\tau$ is varied from 1.4 to 30, the mean value of the first ingredient tends to decrease, and the proportions of ingredient values below both 0.4 and 0.5 tend to increase. However, it is not always so when

$(\tau, \tau, 0, 0, 0, 0, 0)$, where $\tau$ varies from 1.4 to 30, is considered. To get insight about the first two ingredients, it is easier to investigate the third one, since if proportions for the first two decrease, the proportion for the third one should increase. The mean proportion for the third ingredient increases when we move from $\tau = 1.4$ to $\tau = 5$ and to $\tau = 30$. The percentage of proportions of the first two ingredients smaller than 0.4 or 0.5 does not always increase with every increase in $\tau$ considered.

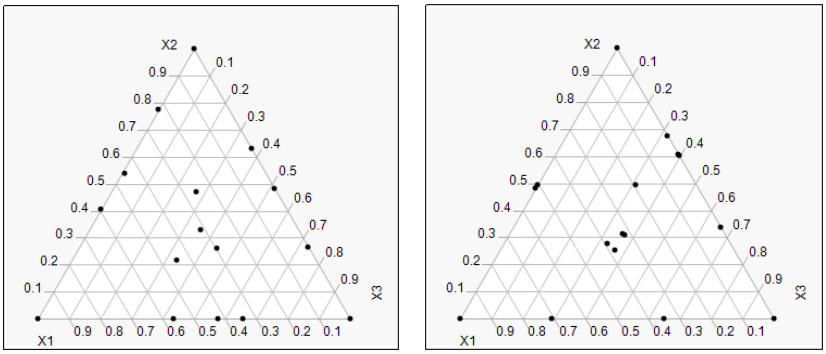
The same characteristics for the utility neutral optimal design are given in Table 6. By comparing Tables 5 and 6, we can see that when we move the first coordinate further from zero, the average value and the number of observations smaller than 0.4 and 0.5 for it decrease. The first coordinate in such a case gets lower and lower values. But it is not always so when we change the first two coordinates. However, we can explain this by some choice sets that "investigate" a trade-off between the first two ingredients in the experiment. Nevertheless, even though the reasoning outlined above does not hold perfectly when all points in optimal designs are taken into account, this pattern is observed in the graphs provided earlier in this section.

| | $(0,0,0,0,0,0,0)$ | | |
|---|---|---|---|
| | $x_1$ | $x_2$ | $x_3$ |
| mean | 0.3521 | 0.3199 | 0.3279 |
| $<0.5$ | 60% | 75% | 65% |
| $<0.4$ | 55% | 60% | 55% |

Table 6: Characteristics of distributions of ingredient proportions in the utility neutral optimal design

It can also be interesting to strengthen the reasoning above by investigating how the optimal points are allocated when one of the interaction terms dominates in the choice problem. Let us compare locally optimal designs under two situations, namely, when $\boldsymbol{\tau} = (0, 0, 0, 0, 0, 0, 0)$ and $\boldsymbol{\tau} = (0, 0, 0, 10, 0, 0, 0)$ for the design problem consisting of 3 ingredients, 20 alternatives and 2 choices in a choice set. In the latter situation we make the interaction of the first two ingredients dominate. The optimal designs are plotted in Figure 11, from where it can be seen that optimal points move slightly down and to the right in the second graph, meaning that lower values for the first and second proportions are optimal now. To get more insight into this, values of all the proportions can be seen in Tables 30 and 31 in the Appendix. If we calculate the mean proportions for both ingredients for the first design problem and for the second one, we get 0.3455, 0.3323 and 0.3020, 0.3428, respectively. Thus, on average, the first proportion decreases by approximately 13%, while the second one slightly increases by 3%.

Assuming utility neutral designs might seem attractive, as it helps to reduce computation time




(a) $\boldsymbol{\tau} = (0, 0, 0, 0, 0, 0, 0)$ (b) $\boldsymbol{\tau} = (0, 0, 0, 10, 0, 0, 0)$

Figure 11: Designs for 100,000 random starts for different $\boldsymbol{\tau}$s for the design problem with 3 ingredients, 20 alternatives, and 2 choices in a choice set

and avoids the necessity to choose a prior distribution for the parameters. However, it must be kept in mind that the situation described by them is rather unrealistic, as it is hard to believe that respondents have equal preferences over alternatives. It is widely known that, for instance, consumers prefer low prices to high ones, certain brands are consistently desired over others, and that other features of a product are equally liked by consumers (Huber and Zwerina, 1996). In the context of this thesis, not all ingredients might be equally important to the respondents. Moreover, after investigating utility neutral designs and designs when consumers are assumed not to be indifferent between alternatives in a choice set, it is obvious that differences between them exist. The pattern observed in general is the further the prior mean from zero, the larger the differences in look and performance of the design. Thus, preferences that consumers have for ingredients of goods and services should clearly be accounted for in the model when designing optimal experiments, and it should not be assumed that utility neutral designs do not differ from non-linear ones and, thus, capture patterns in an equally good way.

## 5.3 Investigating Parameter Misspecification

It is of importance to obtain reasonable prior parameter values required for the optimization, what is not always easy to do. The claim made by Huber and Zwerina (1996) that it is better to be wrong about priors than forget them altogether might seem rather extreme. Strong misspecifications are, of course, not what they mean, and authors have in mind monotonous misspecifications. In this section, the loss in efficiency, when assuming different values for parameters and "misspecifying" them in two different ways is investigated. Designs considered are locally optimal ones. Monotonous misspecifications that we analyze are misspecification in scale and misspecification in skewness, which are obtained by multiplying parameter values by certain constants. Both of

them are introduced below.

The same four experimental problems as analyzed in Section 5.2 are used; and again they are optimized for 100 random starts. They all are given in successive rows in Tables 7 and 8. A choice design that is characterized by, say, two ingredients, three choices in a choice set and 24 alternatives is denoted by 2/3/24 in the tables. Then, misspecification in parameters is analyzed. First, true parameter values are assumed and locally the best (i.e., assuming no uncertainty) design for every experimental problem using them is computed. For the experiment with two ingredients and an interaction term, the true parameter vector is the vector $\boldsymbol{\tau} = (-0.8, 1.2, 0.7)$, and for the experimental designs with three ingredients and all interaction terms, it is taken to be $\boldsymbol{\tau} = (0.7, -1.2, 1.1, 0.2, 0.8, 0.45, 1)$. The loss in efficiency from misspecifications that are described later in this section is judged with respect to the D-value of the best design with the true parameter vector $\boldsymbol{\tau}$. The best design for the misspecified parameter values is constructed and its performance evaluated under the true parameters. In such a way obtained D-values are provided in the tables for every misspecification type for every design analyzed.

For the misspecification in scale, three cases are assumed: the utility neutral design, and designs, where true parameter values are multiplied by 0.75 and 1.25, respectively. The results for all these instances are shown in the last three columns in Table 7. The two last columns in Table 8 give the utility neutral design and misspecification in skewness, where the true parameter values are halved, if they are less than zero, or doubled, if they are greater than zero. All the cells in the tables for the designs obtained by misspecifying the parameters give D-values for designs obtained by assuming a certain misspecification type and evaluating them under the true parameter values. The second columns in both tables give the D-value for the best design with the true parameter vector. Notice that the first three columns in both tables are the same.

| | True parameter vector | Utility neutral design | $\boldsymbol{\tau}^{\text{mis}} = \boldsymbol{\tau} \times 0.75$ | $\boldsymbol{\tau}^{\text{mis}} = \boldsymbol{\tau} \times 1.25$ |
|---|---|---|---|---|
| 2/3/24 | 1.9918 | 1.9936 | 1.9933 | 1.9918 |
| 3/2/24 | 12.8645 | 13.2633 | 12.9070 | 12.9027 |
| 3/3/24 | 13.8435 | 14.4173 | 14.2186 | 13.9398 |
| 3/4/64 | 6.8986 | 7.3834 | 6.9728 | 6.9003 |

Table 7: D-efficiency for different types of monotonous misspecifications in scale of parameter values, when $\boldsymbol{\tau} = (-0.8, 1.2, 0.7)$ or $\boldsymbol{\tau} = (0.7, -1.2, 1.1, 0.2, 0.8, 0.45, 1)$

From Table 7 it can be seen that for every experimental case the worst D-values are obtained for the utility neutral design. Misspecified parameters in scale (that is, when true parameters are multiplied by 0.75 or 1.25) result in more efficient designs. Table 8 shows that the same conclusion holds for misspecification in skewness – the utility neutral design performs worst. Increased efficiency by not assuming zeros for prior parameter values obtained by Huber and

| | True parameter vector | Utility neutral design | $\tau_i^{\text{mis}} = \tau_i/2$, if $\tau_i < 0$<br>$\tau_i^{\text{mis}} = \tau_i \times 2$, if $\tau_i > 0$ |
|---|---|---|---|
| 2/3/24 | 1.9918 | 1.9936 | 1.9935 |
| 3/2/24 | 12.8645 | 13.2633 | 12.8984 |
| 3/3/24 | 13.8435 | 14.4173 | 14.0925 |
| 3/4/64 | 6.8986 | 7.3834 | 6.9356 |

Table 8: D-efficiency for monotonous misspecification in skewness of parameter values, when $\boldsymbol{\tau} = (-0.8, 1.2, 0.7)$ or $\boldsymbol{\tau} = (0.7, -1.2, 1.1, 0.2, 0.8, 0.45, 1)$

Zwerina (1996) is larger, however, the instances taken in this example are meant to show that the same pattern holds for the algorithm developed in this master thesis.

To get more pronounced results, higher values in absolute terms for a parameter vector $\boldsymbol{\tau}$ are assumed. Namely, for the two-ingredient design, we take $\boldsymbol{\tau} = (12, 1, -3)$, and for the three-ingredient designs, it is $\boldsymbol{\tau} = (8.12, -3.75, 1.15, 6.24, -5.11, -2.18, 7.87)$. Again, no uncertainty is assumed. The results for misspecifications in scale and skewness are provided in Tables 9 and 10, respectively, where, again, the first three columns in them coincide. The results are indeed more astounding, however, pattern stays the same. That is, the utility neutral designs always perform worst, and monotonous misspecifications prove to be better than the situations when zeros are assumed for the parameter values. An extreme case is the one where the true parameter vector is a vector of zeros. Then the utility neutral design is the best (Huber and Zwerina, 1996).

| | True parameter vector | Utility neutral design | $\boldsymbol{\tau}^{\text{mis}} = \boldsymbol{\tau} \times 0.75$ | $\boldsymbol{\tau}^{\text{mis}} = \boldsymbol{\tau} \times 1.25$ |
|---|---|---|---|---|
| 2/3/24 | 12.1475 | 320.0444 | 12.4139 | 13.4981 |
| 3/2/24 | 30.9677 | 81.6304 | 31.1692 | 32.8406 |
| 3/3/24 | 33.1608 | 164.3541 | 39.3207 | 35.4446 |
| 3/4/64 | 15.4729 | 78.2656 | 15.4729 | 18.1173 |

Table 9: D-efficiency for different types of monotonous misspecifications in scale of parameter values, when $\boldsymbol{\tau} = (12, 1, -3)$ or $\boldsymbol{\tau} = (8.12, -3.75, 1.15, 6.24, -5.11, -2.18, 7.87)$

| | True parameter vector | Utility neutral design | $\tau_i^{\text{mis}} = \tau_i/2$, if $\tau_i < 0$<br>$\tau_i^{\text{mis}} = \tau_i \times 2$, if $\tau_i > 0$ |
|---|---|---|---|
| 2/3/24 | 12.1475 | 320.0444 | 18.9419 |
| 3/2/24 | 30.9677 | 81.6304 | 44.7952 |
| 3/3/24 | 33.1608 | 164.3541 | 47.6454 |
| 3/4/64 | 15.4729 | 78.2656 | 21.5014 |

Table 10: D-efficiency for monotonous misspecification in skewness of parameter values, when $\boldsymbol{\tau} = (12, 1, -3)$ or $\boldsymbol{\tau} = (8.12, -3.75, 1.15, 6.24, -5.11, -2.18, 7.87)$

In order to strengthen our conclusions, the same strategy is implemented for ten draws from

the multivariate normal distribution with prior mean $\boldsymbol{\tau_0} = (2, -1.05, 0.75, 4, 2.14, -0.88, -3)$, and prior variance-covariance matrix, $\boldsymbol{\Sigma}_0'$, which is an identity matrix. We proceed in the same manner as before. To obtain optimal experimental designs, 50 random starts are used. First, every draw is assumed to be a true parameter vector for the design problem consisting of 3 ingredients, 20 alternatives and 2 choices in a choice set, when two-ingredient and three-ingredient interactions are also included in the model. Then, the locally optimal designs for the true parameter vectors (i.e., for every draw out of the ten ones from the multivariate normal distribution) are computed. Further, the utility neutral design and designs when true parameter vectors are misspecified in the three ways discussed above for the model are obtained. It is then compared how much better the designs with monotonously misspecified parameter values perform than the utility neutral design. The minimum, mean, and maximum decrease in the D-value with respect to the utility neutral design (i.e., when $\boldsymbol{\tau_0}$ is assumed to be $\mathbf{0}$) are given in Table 11 for every type of misspecification. It is clear that the conclusion remains the same, it is always better to assume values for the parameters in the model than assuming the vector of zeros, even if they are monotonous misspecified.

| | $\boldsymbol{\tau}^{\text{mis}} = \boldsymbol{\tau_0} \times 0.75$ | $\boldsymbol{\tau}^{\text{mis}} = \boldsymbol{\tau_0} \times 1.25$ | $\tau_i^{\text{mis}} = \tau_{0i}/2$, if $\tau_{0i} < 0$, $\tau_i^{\text{mis}} = \tau_{0i} \times 2$, if $\tau_{0i} > 0$ |
|---|---|---|---|
| min | 3.59% | 4.23% | 1.87% |
| mean | 8.12% | 8.59% | 6.08% |
| max | 14.99% | 15.5% | 11.41% |

Table 11: Decrease in D-value for the three types of monotonous misspecifications with respect to the utility neutral design for the ten true parameter values assumed

Table 12 gives the minimum, mean, and maximum increases in the D-value for the worst misspecification (column 1) and for the utility neutral design (column 2) with respect to the locally optimal designs for the ten true parameter vectors. By the worst misspecification we mean the one which has the highest D-value compared to the other types of misspecifications. From the values it can be seen that even for the worst misspecification the minimum, mean and maximum increases in D-value with respect to the true locally optimal designs are not very high, especially compared to such increases when the utility neutral design is assumed. Nine times out of ten the worst misspecification appeared to be the one in skewness.

Conclusions drawn in this section are reassuring, as parameters can only be roughly estimated before the choice experiment. Sándor and Wedel (2001) argue that pilot testing used in practice to receive prior parameter values has its limitations. First, some design should be already available for this, and second, uncertainty about parameter values obtained from the pilot are not accounted for. Hence, it is not known how good the constructed designs are if the true parameter values

| | increase for the worst misspecification | increase for the utility neutral |
|---|---|---|
| min | 1.82% | 4.73% |
| mean | 3.99% | 10.06% |
| max | 6.14% | 17.56% |

Table 12: Increase in D-value for the worst misspecification and for the utility neutral design with respect to the locally optimal designs for the ten true parameter values assumed

differ from the ones used. It also might help to overcome the issue of over-confidence of managers who set prior values for the parameters in the model. We see that linear misspecifications do not distort the outcome but rather may help to improve the design. Thus, if it is possible to get decent priors for modeling, it is preferable not to leave them out.

### 5.4 The Semi-Bayesian Approach

In this section, we show that semi-Bayesian designs perform better than locally optimal ones for most of the time. As an example we choose the utility neutral design as a locally optimal design, i.e., a design which is constructed assuming zero parameter values for the model. We then proceed as follows. First, we optimize two designs using 30 random starts: a utility neutral design and a semi-Bayesian one. The model contains 3 ingredients, 20 alternatives and 2 choices in a choice set. All interaction terms are included in the model. The semi-Bayesian design is optimized for 128 Halton draws from a multivariate normal distribution with mean $\mathbf{0}$ and variance-covariance matrix which is a diagonal matrix with ones on the diagonal. Then, we take 1,000,000 draws for the parameter vector in the model from a multivariate normal distribution with zero mean and a variance-covariance matrix which is an identity matrix.

The measure of efficiency that we use is $\frac{D_{\text{Semi-Bayesian}}}{D_{\text{Neutral}}}$, where $D_{\text{Semi-Bayesian}}$ represents the D-value for the semi-Bayesian design, and $D_{\text{Neutral}}$ represents the D-value for the utility neutral design when designs are evaluated for a certain draw. Values of it smaller than 1 indicate situations where the semi-Bayesian design is better and vice versa. The idea is to show a distribution of $\frac{D_{\text{Semi-Bayesian}}}{D_{\text{Neutral}}}$ for 1,000,000 draws. Most of the time, the semi-Bayesian design performs better (957,757 times out of the 1,000,000, or 95.78% of the time). Figure 12 gives the distribution of the efficiency measure used (i.e., the ratio $\frac{D_{\text{Semi-Bayesian}}}{D_{\text{Neutral}}}$). We clearly see that the mass of it is concentrated on the left hand side of the graph (most of the values are below 1), meaning that the semi-Bayesian design outperforms the utility neutral one for most of the draws. The same result has been obtained for other than the utility neutral designs, which shows that it is wiser to account for the uncertainty.

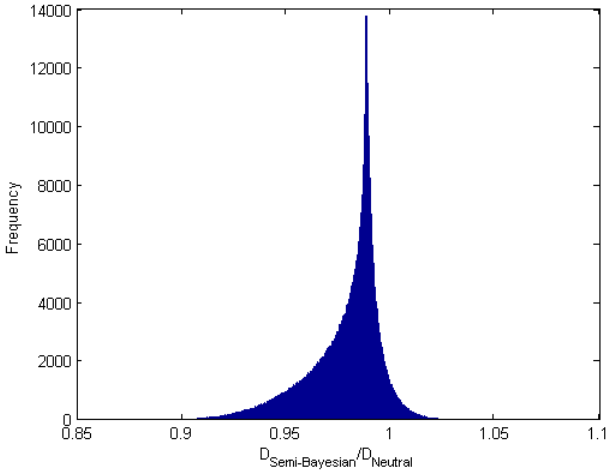


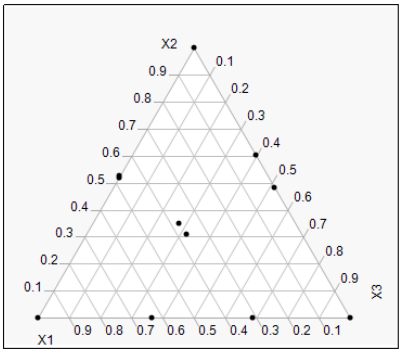

Figure 12: Distribution of $\frac{\mathrm{D_{Semi\text{-}Bayesian}}}{\mathrm{D_{Neutral}}}$

## 5.5 A Detailed Comparison

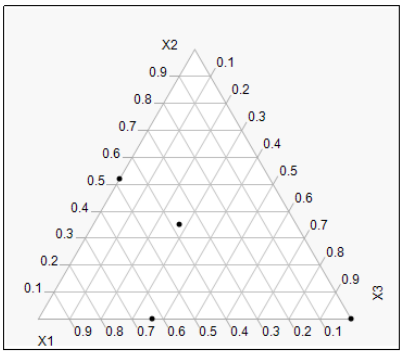
In this section, we show that semi-Bayesian designs differ from benchmark designs. Figure 4 from Section 4 and Figure 13 can serve as a proof that both approaches give different optimal designs. In Figure 4, the utility neutral design for 3 ingredients, 12 alternatives and 4 choices in a choice set, when two-ingredient and three-ingredient interaction terms are also included, is obtained and plotted by JMP. Figure 13 gives the optimal design for the same problem obtained by employing Bayesian approach for 128 draws from the multivariate normal distribution with $\boldsymbol{\tau}_0 = (0, 0, 0, 0, 0, 0, 0)$ and $\boldsymbol{\Sigma}_0'$ being an identity matrix for 1,000 random starts. Draws are obtained using Halton sequences. Two graphs clearly show that two approaches differ and give different optimal experimental designs. The points of two designs can be seen in Tables 13 and 14 in the Appendix.

The following examples support this finding. Figures 14 and 15 give the utility neutral design and the semi-Bayesian design, respectively, for the model consisting of 3 ingredients, 12 alternatives and 2 choices in a choice set. All interaction terms are also included in the model. The semi-Bayesian design is obtained for 128 draws from the multivariate normal distribution with zero mean and variance-covariance matrix being an identity matrix. Halton sequences are used in order to obtain the aforementioned draws. Both designs are optimized for 1,000 random starts. The utility neutral design is obtained by JMP. Again, two approaches give different experimental designs for the same model. The optimal observations for both designs are given in Appendix, in Table 32 for the utility neutral design, and in Table 33 for the semi-Bayesian experimental design.

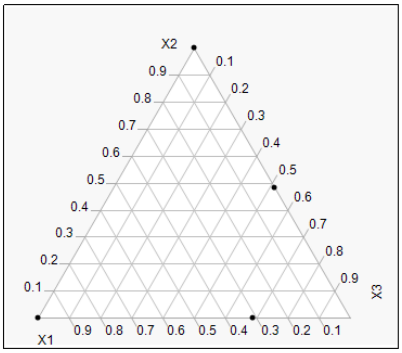
The last example shows two experimental designs given by the two different approaches for the model consisting of 3 ingredients, 12 alternatives and 3 choices in a choice set. All interaction



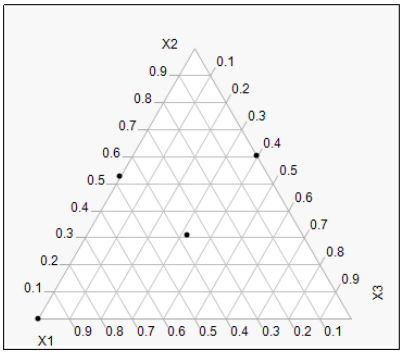

(a) All choice sets




(b) First choice set




(c) Second choice set



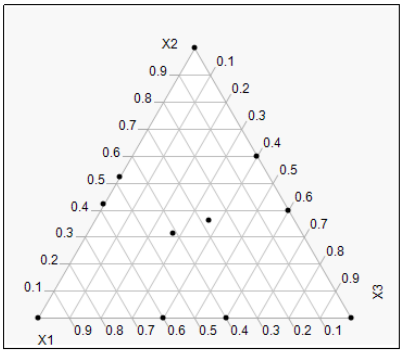

(d) Third choice set

Figure 13: Optimal design for 3 ingredients, 12 alternatives and 4 choices in a choice set, when division into choice sets is ignored and also when alternatives are divided into three choice sets, for Bayesian approach for 128 draws from the multivariate normal distribution with $\boldsymbol{\tau}_0 = (0, 0, 0, 0, 0, 0, 0)$ and $\boldsymbol{\Sigma}_0'$ being an identity matrix

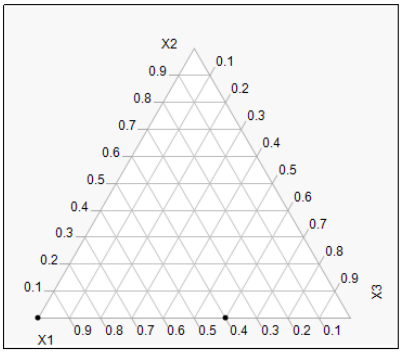
terms are included in the model. Figure 16 gives the locally optimal design assuming the utility neutral design, and Figure 17 gives the optimal semi-Bayesian experimental design, when 128 draws are sampled from the multivariate normal distribution with zero mean and when variance-covariance matrix is an identity matrix. The draws are obtained using Halton sequences. Both the utility neutral design and the semi-Bayesian design are optimized for 1,000 random starts. The utility neutral design is obtained using JMP. The conclusion remains the same, the two approaches, namely, when the utility neutral design is assumed for optimization and when semi-Bayesian approach is used, give different experimental designs. The optimal observations are given in the Appendix, in Tables 34 and 35 for the utility neutral and the semi-Bayesian designs, respectively.

In the previous section, we showed that semi-Bayesian designs perform better than locally optimal (and the utility neutral) designs for most of the time. In this section, we conclude that

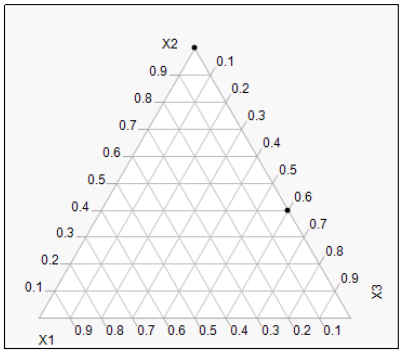
semi-Bayesian designs differ from locally optimal (and the utility neutral) ones. Hence, it is worth using the computationally intensive semi-Bayesian designs instead of the computationally cheap benchmark designs, as results differ for both approaches and they are more robust when the semi-Bayesian approach is employed.



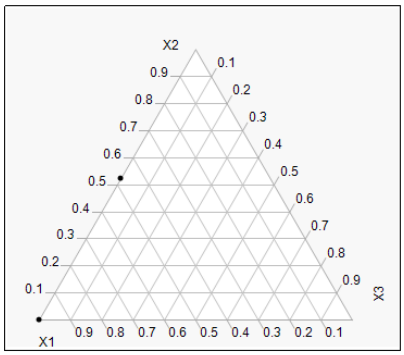

(a) All choice sets



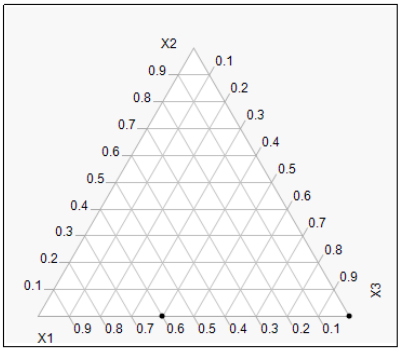

(b) First choice set



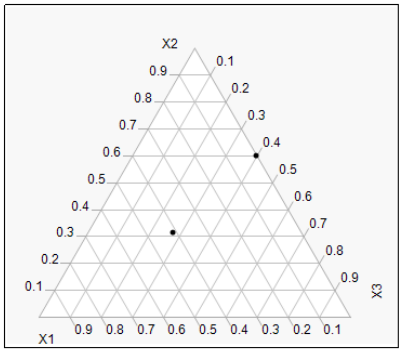

(c) Second choice set



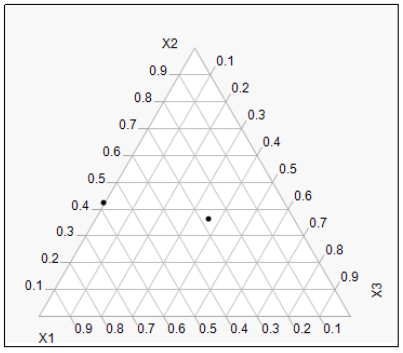

(d) Third choice set




(e) Fourth choice set

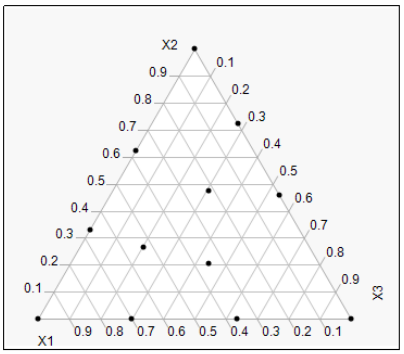



(f) Fifth choice set

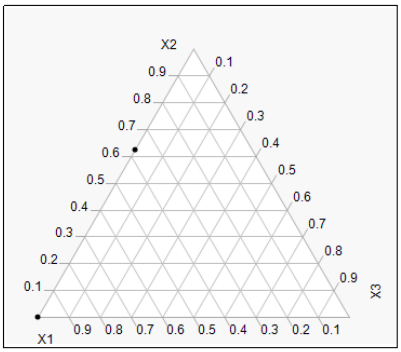



(g) Sixth choice set

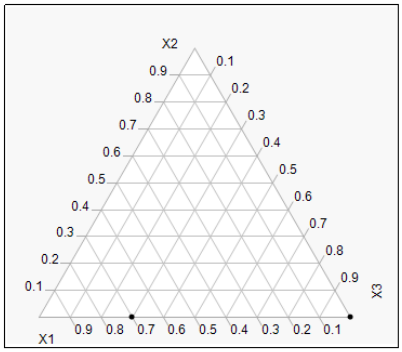
Figure 14: Optimal design for 3 ingredients, 12 alternatives and 2 choices in a choice set, when division into choice sets is ignored and also when alternatives are divided into six choice sets, when the utility neutral design is assumed



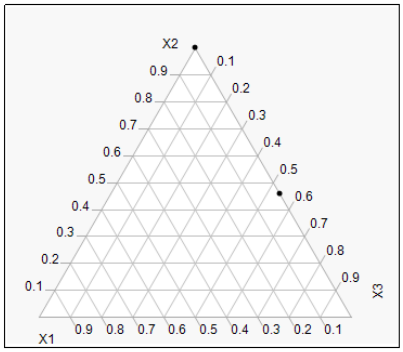

(a) All choice sets



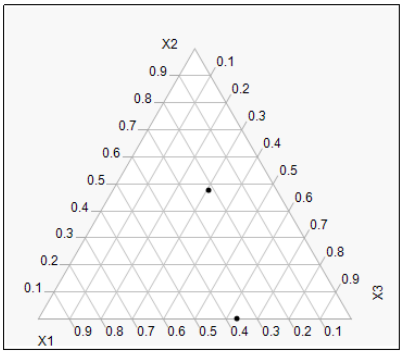

(b) First choice set



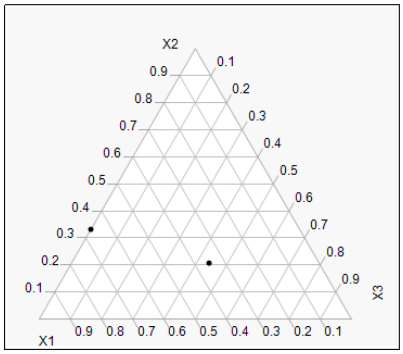

(c) Second choice set



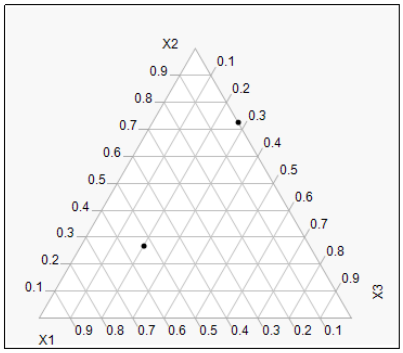

(d) Third choice set




(e) Fourth choice set

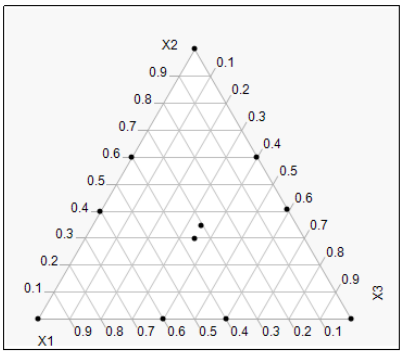



(f) Fifth choice set

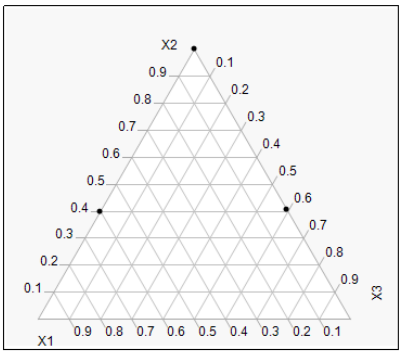



(g) Sixth choice set

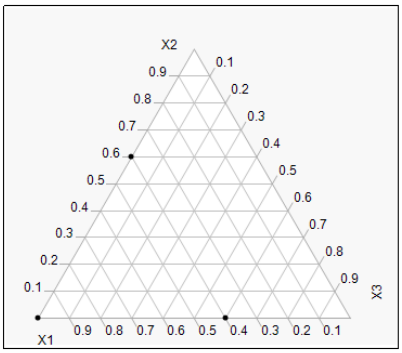
Figure 15: Optimal design for 3 ingredients, 12 alternatives and 2 choices in a choice set, when division into choice sets is ignored and also when alternatives are divided into six choice sets, for Bayesian approach for 128 draws from the multivariate normal distribution with $\boldsymbol{\tau}_0 = (0, 0, 0, 0, 0, 0, 0)$ and $\boldsymbol{\Sigma}_0'$ being an identity matrix



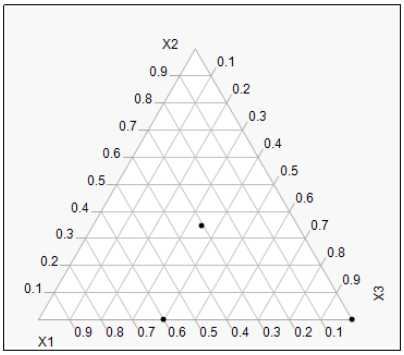

(a) All choice sets



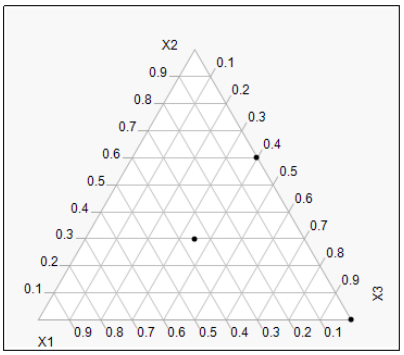

(b) First choice set




(c) Second choice set

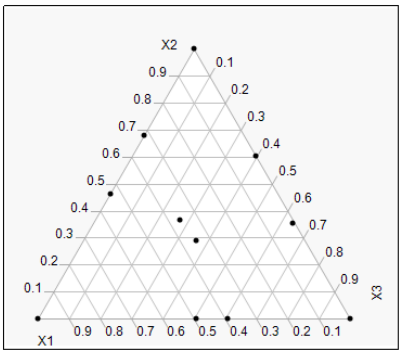



(d) Third choice set

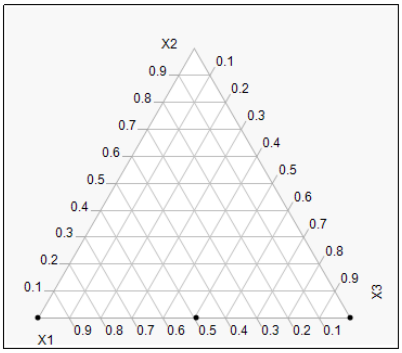



(e) Fourth choice set

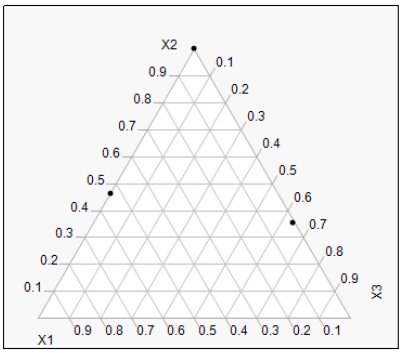
Figure 16: Optimal design for 3 ingredients, 12 alternatives and 3 choices in a choice set, when division into choice sets is ignored and also when alternatives are divided into four choice sets, when the utility neutral design is assumed



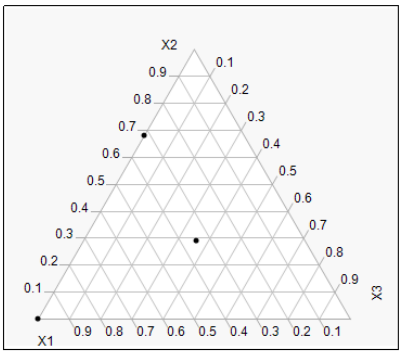

(a) All choice sets



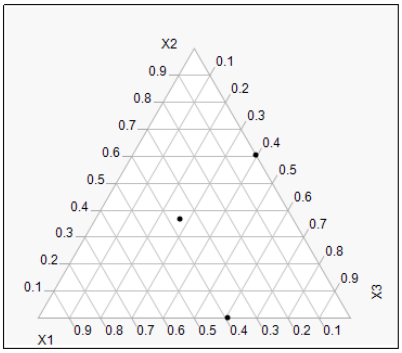

(b) First choice set




(c) Second choice set




(d) Third choice set




(e) Fourth choice set

Figure 17: Optimal design for 3 ingredients, 12 alternatives and 3 choices in a choice set, when division into choice sets is ignored and also when alternatives are divided into four choice sets, for Bayesian approach for 128 draws from the multivariate normal distribution with $\boldsymbol{\tau}_0 = (0, 0, 0, 0, 0, 0, 0)$ and $\boldsymbol{\Sigma}_0'$ being an identity matrix

# 6 Conclusion

In this master thesis we studied a problem of how to construct an efficient experimental design for the multinomial logit model when choices are based on a mixture of ingredients. An algorithm for doing this is developed and analyzed for some different settings. The resulting design is D-optimal. The complication in constructing such designs is that in order to design an experiment that results in efficiently estimated choice model's parameters, those parameters have to be known beforehand, as the information matrix that appears in the D-criterion expression used to optimize a design depends on the parameter values that are not known for a researcher at the beginning of the analysis. The way in which we overcome this circular issue is a semi-Bayesian approach. However, if there is an opportunity to obtain a single decent parameter values vector, it could also be used, as it is demonstrated that monotonous misspecifications in true parameters do not distort the outcome, rather they can even help to design more robust designs.

One way to obtain such parameter values is through interviewing managers, since it is widely believed that they possess relevant knowledge on the behaviour of their customers. If the uncertainty in those parameter values can also be addressed, this might produce even better experimental designs. As Sándor and Wedel (2001) outline, such increased efficiency of the semi-Bayesian design can be decomposed into two components, namely, the efficiency gain due to the use of manager beliefs on the choice probabilities of products and services characterized by a mixture of ingredients, and an improvement due to accommodation of managers' uncertainty about the values elicited about those probabilities in the population. They also state that neglecting this uncertainty is logically inconsistent, because if the values for the parameters in the model were precisely known, no design needs to be generated.

We also show that such designs where prior parameter values are not assumed to be zero differ from the utility neutral ones. Furthermore, semi-Bayesian designs differ from and perform better than locally optimal ones (and the utility neutral designs) for most of the time. As a result, it is important to obtain reasonable prior parameter values, and the utility neutral designs should be better used as a starting point to get them or as a benchmark design.

However, it should be stressed that it cannot be proven that the designs that can be generated using the algorithm presented are strictly optimal. They should not be expected to be so for a couple of reasons. First, design optimality is investigated with respect to the D-optimality criterion, which is not the only one that can be chosen. It is chosen as the criterion in this master thesis, since it is the most commonly used one in practice. However, the method can be very easily extended to employ different criteria. Second, the search is heuristic and, thus, it may happen that the optimum the algorithm arrives at is not a global optimum. A heuristic procedure is employed because an exhaustive search over the entire design space is not feasible in this case,

as proportions of ingredients can obtain extremely many values on the interval (0, 1).

What is also interesting to note is that different designs that are equivalent in D-efficiency can be found during the optimization. Despite the fact that they behave equivalently with respect to D-criterion, they might be different in terms of other criteria. Any of those criteria could then be used as a tie breaker. It is also possible to choose one experimental design from several ones by taking into account the cost of implementing them (Goos and Donev, 2006).

The value that this master thesis adds to the existing literature is the application of mixture models in the choice modeling, that has never been done before. Moreover, the algorithm developed is capable to account for settings different from the utility neutral models, which have mainly been used up to now.

## References

M.C.J. Bliemer, J.M. Rose, and S. Hess. Approximation of Bayesian efficiency in experimental choice designs. *Journal of Choice Modeling*, 1(1):98–127, 2009.

R.P. Brent. *Algorithms for Minimization without Derivatives*. Prentice-Hall, 1973.

R.T. Carson, J.J. Louviere, D.A. Anderson, P. Arabie, D.S. Bunch, D.A. Hensher, R.M. Johnson, W.F. Kuhfeld, D. Steinberg, J. Swait, H. Timmermans, and J.B. Wiley. Experimental analysis of choice. *Marketing Letters*, 5(4):351–367, 1994.

J.A. Cornell. *Experiments with Mixtures*. John Wiley & Sons, Inc, 2002.

The Geomblog. Ruminations on Computational Geometry, Algorithms, Theoretical Computer Science and Life. `http://geomblog.blogspot.be/2005/10/sampling-from-simplex.html`. [Online; accessed 8 June, 2012].

P. Goos. *The Optimal Design of Blocked and Split-Plot Experiments*. Springer, 2002.

P. Goos and A.N. Donev. The D-optimal design of blocked experiments with mixture components. *Journal of Quality Technology*, 38(4):319–332, 2006.

P. Goos and A.N. Donev. D-optimal minimum support mixture designs in blocks. *Metrika*, 65 (1):53–68, 2007.

P. Goos, R. Kessels, P. De Pelsmacker, and N. Dens. Media-mix optimization: A multi-level mixture-amount model approach. Unpublished manuscript, 2012.

J. Halton. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numerische Matematik*, 2(1):84–90, 1960.

J. Huber and K. Zwerina. The importance of utility balance in efficient choice designs. *Journal of Marketing Research*, 33(3):307–317, 1996.

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

R. Kessels, B. Jones, P. Goos, and M. Vandebroek. An efficient algorithm for constructing bayesian optimal choice designs. *Journal of Business & Economic Statistics*, 27(2):279–291, 2009.

R. Kessels, B. Jones, P. Goos, and M. Vandebroek. The usefulness of bayesian optimal designs for discrete choice experiments. *Applied Stochastic Models in Business and Industry*, 27(3): 173–188, 2011.

R.K. Meyer and C.J. Nachtsheim. The coordinate-exchange algorithm for constructing exact optimal experimental designs. *Technometrics*, 37(1):60–69, 1995.

G.F. Piepel. Measuring component effects in constrained mixture experiments. *Technometrics*, 24(1):29–39, 1982.

G.F. Piepel, S.K. Cooley, and B. Jones. Construction of a 21-component layered mixture experiment design using a new mixture coordinate-exchange algorithm. *Quality Engineering*, 17(4): 579–594, 2005.

Z. Sándor and M. Wedel. Designing conjoint choice experiments using managers' prior beliefs. *Journal of Marketing Research*, 38(4):430–444, 2001.

H. Scheffé. Experiments with mixtures. *Journal of the Royal Statistical Society, Ser. B*, (20): 344–360, 1958.

M. Verbeek. *A Guide to Modern Econometrics*. John Wiley & Sons, Ltd, 2008.

E.W. Weisstein. Cholesky Decomposition from *MathWorld* - A Wolfram Web Resource. `http://mathworld.wolfram.com/CholeskyDecomposition.html`. [Online; accessed 31 May, 2012].

J. Yu, P. Goos, and M. Vandebroek. Comparing different sampling schemes for approximating the integrals involved in the efficient design of stated choice experiments. *Transportation Research Part B: Methodological*, 44(10):1268–1289, 2010.

# Appendices

| $x_1$ | $x_2$ | $x_3$ | $x_1$ | $x_2$ | $x_3$ |
|---|---|---|---|---|---|
| 0.50 | 0.00 | 0.50 | 0.00 | 0.43 | 0.57 |
| 0.00 | 0.00 | 1.00 | 0.37 | 0.25 | 0.37 |
| 1.00 | 0.00 | 0.00 | 0.56 | 0.44 | 0.00 |
| 0.29 | 0.41 | 0.29 | 0.00 | 1.00 | 0.00 |
| 0.00 | 0.60 | 0.40 | | | |
| 0.00 | 0.00 | 1.00 | | | |
| 1.00 | 0.00 | 0.00 | | | |
| 0.40 | 0.60 | 0.00 | | | |

Table 13: Optimal points for the locally optimal utility neutral design with 3 ingredients, 12 alternatives and 4 choices in a choice set for 1,000 random starts. Choice sets are separated by dashed lines

| $x_1$ | $x_2$ | $x_3$ | $x_1$ | $x_2$ | $x_3$ |
|---|---|---|---|---|---|
| 0.63 | 0.00 | 0.37 | 0.00 | 0.60 | 0.40 |
| 0.48 | 0.52 | 0.00 | 0.47 | 0.52 | 0.00 |
| 0.37 | 0.35 | 0.28 | 0.37 | 0.31 | 0.32 |
| 0.00 | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 |
| 0.00 | 1.00 | 0.00 | | | |
| 0.00 | 0.48 | 0.52 | | | |
| 0.31 | 0.00 | 0.69 | | | |
| 1.00 | 0.00 | 0.00 | | | |

Table 14: Optimal points for the semi-Bayesian design with 3 ingredients, 12 alternatives and 4 choices in a choice set for 1,000 random starts, when 128 draws are taken from the multivariate normal distribution with $\boldsymbol{\tau}_0 = (0, 0, 0, 0, 0, 0)$ and $\boldsymbol{\Sigma}_0'$ being an identity matrix. Choice sets are separated by dashed lines

| $x_1$ | $x_2$ | $x_3$ | $x_1$ | $x_2$ | $x_3$ |
|---|---|---|---|---|---|
| 0.00 | 0.00 | 1.00 | 0.00 | 1.00 | 0.00 |
| 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| 1.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 |
| 0.00 | 1.00 | 0.00 | 0.00 | 1.00 | 0.00 |
| 0.00 | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 |
| 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| 1.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 |
| 0.00 | 1.00 | 0.00 | 1.00 | 0.00 | 0.00 |
| 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 1.00 |

Table 15: Optimal points for the utility neutral design with 3 ingredients, 18 alternatives, and 9 choices in a choice set for 10,000 random starts, when only main effects are considered. Two choice sets are given in two corresponding columns

| $x_1$ | $x_2$ | $x_3$ | $x_1$ | $x_2$ | $x_3$ |
|---|---|---|---|---|---|
| 0.00 | 1.00 | 0.00 | 0.00 | 1.00 | 0.00 |
| 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| 0.00 | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 |
| 0.46 | 0.54 | 0.00 | 0.50 | 0.50 | 0.00 |
| 0.52 | 0.00 | 0.48 | 0.00 | 1.00 | 0.00 |
| 1.00 | 0.00 | 0.00 | 0.00 | 0.48 | 0.52 |
| 0.00 | 0.51 | 0.49 | 0.00 | 0.00 | 1.00 |
| 0.49 | 0.00 | 0.51 | 0.52 | 0.00 | 0.48 |
| 0.00 | 0.50 | 0.50 | 0.48 | 0.52 | 0.00 |

Table 16: Optimal points for the utility neutral design with 3 ingredients, 18 alternatives, and 9 choices in a choice set for 10,000 random starts, when main effects and two-ingredient interactions are considered. Two choice sets are given in two corresponding columns

| $x_1$ | $x_2$ | $x_3$ | $x_1$ | $x_2$ | $x_3$ |
|---|---|---|---|---|---|
| 0.00 | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 |
| 0.00 | 1.00 | 0.00 | 0.49 | 0.51 | 0.00 |
| 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| 0.49 | 0.51 | 0.00 | 0.36 | 0.33 | 0.32 |
| 0.36 | 0.32 | 0.32 | 0.00 | 1.00 | 0.00 |
| 0.00 | 1.00 | 0.00 | 0.50 | 0.50 | 0.00 |
| 0.51 | 0.00 | 0.49 | 0.50 | 0.00 | 0.50 |
| 0.31 | 0.34 | 0.35 | 0.34 | 0.32 | 0.34 |
| 0.00 | 0.50 | 0.50 | 0.00 | 0.51 | 0.49 |

Table 17: Optimal points for the utility neutral design with 3 ingredients, 18 alternatives, and 9 choices in a choice set for 10,000 random starts, when main effects and two- and three-ingredient interactions are considered. Two choice sets are given in two corresponding columns

| $x_1$ | $x_2$ | $x_3$ | $x_1$ | $x_2$ | $x_3$ |
|---|---|---|---|---|---|
| 1.00 | 0.00 | 0.00 | 0.62 | 0.00 | 0.38 |
| 0.42 | 0.58 | 0.00 | 0.00 | 0.00 | 1.00 |
| 0.00 | 1.00 | 0.00 | 0.31 | 0.27 | 0.42 |
| 0.00 | 0.38 | 0.62 | 0.45 | 0.55 | 0.00 |
| 0.55 | 0.00 | 0.45 | 0.00 | 0.00 | 1.00 |
| 0.34 | 0.45 | 0.20 | 0.00 | 0.64 | 0.36 |
| 1.00 | 0.00 | 0.00 | 0.40 | 0.34 | 0.26 |
| 0.49 | 0.00 | 0.51 | 0.00 | 0.45 | 0.55 |
| 0.65 | 0.35 | 0.00 | 0.00 | 1.00 | 0.00 |
| 0.00 | 1.00 | 0.00 | 0.32 | 0.31 | 0.38 |

Table 18: Optimal points for the design with 3 ingredients, 20 alternatives, and 2 choices in a choice set for 100,000 random starts, when $\boldsymbol{\tau} = (1.4, 0, 0, 0, 0, 0, 0)$. Choice sets are separated by dashed lines

| $x_1$ | $x_2$ | $x_3$ | $x_1$ | $x_2$ | $x_3$ |
|---|---|---|---|---|---|
| 1.00 | 0.00 | 0.00 | 0.29 | 0.37 | 0.34 |
| 0.36 | 0.64 | 0.00 | 0.00 | 1.00 | 0.00 |
| 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| 0.64 | 0.36 | 0.00 | 0.00 | 0.60 | 0.40 |
| 0.00 | 0.00 | 1.00 | 0.45 | 0.00 | 0.55 |
| 0.62 | 0.00 | 0.38 | 1.00 | 0.00 | 0.00 |
| 0.00 | 0.47 | 0.53 | 0.27 | 0.35 | 0.38 |
| 0.51 | 0.27 | 0.22 | 0.60 | 0.40 | 0.00 |
| 0.00 | 0.45 | 0.55 | 0.52 | 0.00 | 0.48 |
| 0.00 | 1.00 | 0.00 | 0.29 | 0.45 | 0.25 |

Table 19: Optimal points for the design with 3 ingredients, 20 alternatives, and 2 choices in a choice set for 100,000 random starts, when $\boldsymbol{\tau} = (1.4, 1.4, 0, 0, 0, 0, 0)$. Choice sets are separated by dashed lines

| $x_1$ | $x_2$ | $x_3$ | $x_1$ | $x_2$ | $x_3$ |
|---|---|---|---|---|---|
| 0.00 | 0.00 | 1.00 | 0.41 | 0.59 | 0.00 |
| 0.58 | 0.00 | 0.42 | 0.30 | 0.30 | 0.40 |
| 1.00 | 0.00 | 0.00 | 0.44 | 0.56 | 0.00 |
| 0.36 | 0.00 | 0.64 | 1.00 | 0.00 | 0.00 |
| 0.00 | 1.00 | 0.00 | 0.00 | 0.53 | 0.47 |
| 0.66 | 0.34 | 0.00 | 0.44 | 0.30 | 0.26 |
| 0.00 | 0.55 | 0.45 | 0.52 | 0.00 | 0.48 |
| 0.00 | 0.00 | 1.00 | 0.29 | 0.44 | 0.27 |
| 0.00 | 0.41 | 0.59 | 1.00 | 0.00 | 0.00 |
| 0.00 | 1.00 | 0.00 | 0.25 | 0.38 | 0.37 |

Table 20: Optimal points for the design with 3 ingredients, 20 alternatives, and 2 choices in a choice set for 100,000 random starts, when $\boldsymbol{\tau} = (1.4, 1.4, 1.4, 1.4, 1.4, 1.4, 1.4)$. Choice sets are separated by dashed lines

| $x_1$ | $x_2$ | $x_3$ | $x_1$ | $x_2$ | $x_3$ |
|---|---|---|---|---|---|
| 0.21 | 0.46 | 0.32 | 0.00 | 0.00 | 1.00 |
| 0.00 | 0.00 | 1.00 | 0.00 | 0.67 | 0.33 |
| 0.00 | 1.00 | 0.00 | 0.36 | 0.33 | 0.31 |
| 0.00 | 0.35 | 0.65 | 0.00 | 0.66 | 0.34 |
| 1.00 | 0.00 | 0.00 | 0.47 | 0.00 | 0.53 |
| 0.67 | 0.33 | 0.00 | 0.43 | 0.57 | 0.00 |
| 0.43 | 0.57 | 0.00 | 0.49 | 0.00 | 0.51 |
| 0.28 | 0.28 | 0.44 | 0.29 | 0.41 | 0.30 |
| 1.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 |
| 0.65 | 0.00 | 0.35 | 0.23 | 0.31 | 0.46 |

Table 21: Optimal points for the design with 3 ingredients, 20 alternatives, and 2 choices in a choice set for 100,000 random starts, when $\boldsymbol{\tau} = (5, 0, 0, 0, 0, 0, 0)$. Choice sets are separated by dashed lines

| $x_1$ | $x_2$ | $x_3$ | $x_1$ | $x_2$ | $x_3$ |
|---|---|---|---|---|---|
| 1.00 | 0.00 | 0.00 | 0.00 | 0.51 | 0.49 |
| 0.34 | 0.66 | 0.00 | 0.45 | 0.21 | 0.34 |
| 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 1.00 |
| 0.00 | 0.35 | 0.65 | 0.35 | 0.00 | 0.65 |
| 0.00 | 1.00 | 0.00 | 0.00 | 1.00 | 0.00 |
| 0.65 | 0.35 | 0.00 | 0.74 | 0.00 | 0.26 |
| 1.00 | 0.00 | 0.00 | 0.49 | 0.00 | 0.51 |
| 0.34 | 0.39 | 0.27 | 0.00 | 0.46 | 0.54 |
| 0.46 | 0.54 | 0.00 | 0.46 | 0.00 | 0.54 |
| 0.35 | 0.35 | 0.30 | 0.29 | 0.37 | 0.34 |

Table 22: Optimal points for the design with 3 ingredients, 20 alternatives, and 2 choices in a choice set for 100,000 random starts, when $\boldsymbol{\tau} = (5, 5, 0, 0, 0, 0, 0)$. Choice sets are separated by dashed lines

| $x_1$ | $x_2$ | $x_3$ | $x_1$ | $x_2$ | $x_3$ |
|---|---|---|---|---|---|
| 0.00 | 0.00 | 1.00 | 0.47 | 0.00 | 0.53 |
| 0.24 | 0.00 | 0.76 | 0.00 | 0.51 | 0.49 |
| 0.00 | 1.00 | 0.00 | 0.59 | 0.00 | 0.41 |
| 0.61 | 0.39 | 0.00 | 0.29 | 0.37 | 0.34 |
| 0.06 | 0.94 | 0.00 | 0.33 | 0.31 | 0.36 |
| 0.36 | 0.27 | 0.36 | 1.00 | 0.00 | 0.00 |
| 1.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 |
| 0.00 | 0.61 | 0.39 | 0.00 | 0.54 | 0.46 |
| 1.00 | 0.00 | 0.00 | 0.00 | 0.57 | 0.43 |
| 0.25 | 0.75 | 0.00 | 0.17 | 0.06 | 0.77 |

Table 23: Optimal points for the design with 3 ingredients, 20 alternatives, and 2 choices in a choice set for 100,000 random starts, when $\boldsymbol{\tau} = (5, 5, 5, 5, 5, 5, 5)$. Choice sets are separated by dashed lines

| $x_1$ | $x_2$ | $x_3$ | $x_1$ | $x_2$ | $x_3$ |
|---|---|---|---|---|---|
| 0.38 | 0.62 | 0.00 | 0.18 | 0.40 | 0.42 |
| 0.48 | 0.11 | 0.41 | 0.00 | 1.00 | 0.00 |
| 0.00 | 0.00 | 1.00 | 0.30 | 0.70 | 0.00 |
| 0.00 | 0.56 | 0.44 | 0.24 | 0.36 | 0.41 |
| 0.00 | 1.00 | 0.00 | 0.31 | 0.47 | 0.23 |
| 0.00 | 0.29 | 0.71 | 0.37 | 0.02 | 0.61 |
| 0.98 | 0.02 | 0.00 | 0.00 | 0.00 | 1.00 |
| 0.77 | 0.00 | 0.23 | 0.00 | 0.70 | 0.30 |
| 0.00 | 0.47 | 0.53 | 0.25 | 0.46 | 0.29 |
| 0.19 | 0.50 | 0.32 | 0.34 | 0.00 | 0.66 |

Table 24: Optimal points for the design with 3 ingredients, 20 alternatives, and 2 choices in a choice set for 100,000 random starts, when $\boldsymbol{\tau} = (10, 0, 0, 0, 0, 0, 0)$. Choice sets are separated by dashed lines

| $x_1$ | $x_2$ | $x_3$ | $x_1$ | $x_2$ | $x_3$ |
|---|---|---|---|---|---|
| 0.00 | 0.00 | 1.00 | 0.00 | 0.50 | 0.50 |
| 0.20 | 0.00 | 0.80 | 0.42 | 0.00 | 0.58 |
| 1.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 |
| 0.33 | 0.67 | 0.00 | 0.55 | 0.31 | 0.13 |
| 0.65 | 0.00 | 0.35 | 0.00 | 0.67 | 0.33 |
| 0.29 | 0.40 | 0.31 | 0.43 | 0.32 | 0.25 |
| 0.66 | 0.34 | 0.00 | 0.45 | 0.31 | 0.23 |
| 0.00 | 1.00 | 0.00 | 1.00 | 0.00 | 0.00 |
| 0.00 | 0.51 | 0.48 | 0.37 | 0.39 | 0.24 |
| 0.40 | 0.19 | 0.41 | 0.44 | 0.54 | 0.02 |

Table 25: Optimal points for the design with 3 ingredients, 20 alternatives, and 2 choices in a choice set for 100,000 random starts, when $\boldsymbol{\tau} = (10, 10, 0, 0, 0, 0, 0)$. Choice sets are separated by dashed lines

| $x_1$ | $x_2$ | $x_3$ | $x_1$ | $x_2$ | $x_3$ |
|---|---|---|---|---|---|
| 0.00 | 0.35 | 0.65 | 0.37 | 0.37 | 0.26 |
| 0.52 | 0.00 | 0.48 | 0.41 | 0.57 | 0.02 |
| 0.00 | 1.00 | 0.00 | 0.30 | 0.36 | 0.34 |
| 0.74 | 0.00 | 0.26 | 1.00 | 0.00 | 0.00 |
| 1.00 | 0.00 | 0.00 | 0.51 | 0.49 | 0.00 |
| 0.00 | 0.76 | 0.24 | 0.37 | 0.39 | 0.24 |
| 0.00 | 0.68 | 0.32 | 0.65 | 0.00 | 0.35 |
| 0.21 | 0.23 | 0.56 | 0.29 | 0.34 | 0.37 |
| 0.00 | 1.00 | 0.00 | 0.00 | 0.52 | 0.48 |
| 0.32 | 0.35 | 0.33 | 0.45 | 0.00 | 0.55 |

Table 26: Optimal points for the design with 3 ingredients, 20 alternatives, and 2 choices in a choice set for 100,000 random starts, when $\boldsymbol{\tau} = (10, 10, 10, 10, 10, 10, 10)$. Choice sets are separated by dashed lines

| $x_1$ | $x_2$ | $x_3$ | $x_1$ | $x_2$ | $x_3$ |
|---|---|---|---|---|---|
| 0.00 | 0.00 | 1.00 | 0.09 | 0.48 | 0.43 |
| 0.00 | 0.64 | 0.36 | 0.04 | 0.00 | 0.96 |
| 0.31 | 0.69 | 0.00 | 0.74 | 0.09 | 0.16 |
| 0.30 | 0.22 | 0.48 | 0.81 | 0.12 | 0.06 |
| 0.71 | 0.19 | 0.10 | 0.03 | 0.97 | 0.00 |
| 0.79 | 0.01 | 0.20 | 0.09 | 0.42 | 0.50 |
| 0.00 | 0.32 | 0.68 | 0.43 | 0.31 | 0.26 |
| 0.00 | 1.00 | 0.00 | 0.42 | 0.00 | 0.58 |
| 0.10 | 0.41 | 0.49 | 0.46 | 0.54 | 0.00 |
| 0.14 | 0.00 | 0.86 | 0.45 | 0.14 | 0.41 |

Table 27: Optimal points for the design with 3 ingredients, 20 alternatives, and 2 choices in a choice set for 100,000 random starts, when $\boldsymbol{\tau} = (30, 0, 0, 0, 0, 0, 0)$. Choice sets are separated by dashed lines

| $x_1$ | $x_2$ | $x_3$ | $x_1$ | $x_2$ | $x_3$ |
|---|---|---|---|---|---|
| 0.99 | 0.00 | 0.01 | 0.19 | 0.69 | 0.12 |
| 0.00 | 0.99 | 0.01 | 0.46 | 0.35 | 0.19 |
| 0.00 | 0.20 | 0.80 | 0.50 | 0.50 | 0.00 |
| 0.03 | 0.09 | 0.89 | 0.00 | 1.00 | 0.00 |
| 0.00 | 0.64 | 0.36 | 1.00 | 0.00 | 0.00 |
| 0.28 | 0.36 | 0.36 | 0.39 | 0.57 | 0.04 |
| 0.00 | 0.42 | 0.58 | 0.44 | 0.54 | 0.02 |
| 0.26 | 0.17 | 0.57 | 0.93 | 0.00 | 0.07 |
| 0.01 | 0.51 | 0.49 | 0.61 | 0.00 | 0.39 |
| 0.50 | 0.00 | 0.50 | 0.19 | 0.42 | 0.39 |

Table 28: Optimal points for the design with 3 ingredients, 20 alternatives, and 2 choices in a choice set for 100,000 random starts, when $\boldsymbol{\tau} = (30, 30, 0, 0, 0, 0, 0)$. Choice sets are separated by dashed lines

| $x_1$ | $x_2$ | $x_3$ | $x_1$ | $x_2$ | $x_3$ |
|---|---|---|---|---|---|
| 0.84 | 0.16 | 0.00 | 0.64 | 0.00 | 0.36 |
| 0.17 | 0.70 | 0.13 | 0.17 | 0.38 | 0.45 |
| 0.43 | 0.31 | 0.26 | 0.00 | 0.58 | 0.42 |
| 0.08 | 0.91 | 0.01 | 0.29 | 0.16 | 0.56 |
| 0.00 | 0.59 | 0.41 | 0.00 | 1.00 | 0.00 |
| 0.42 | 0.10 | 0.47 | 0.13 | 0.53 | 0.35 |
| 0.24 | 0.76 | 0.00 | 0.42 | 0.18 | 0.40 |
| 0.51 | 0.30 | 0.19 | 1.00 | 0.00 | 0.00 |
| 0.28 | 0.37 | 0.35 | 1.00 | 0.00 | 0.00 |
| 1.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 |

Table 29: Optimal points for the design with 3 ingredients, 20 alternatives, and 2 choices in a choice set for 100,000 random starts, when $\boldsymbol{\tau} = (30, 30, 30, 30, 30, 30, 30)$. Choice sets are separated by dashed lines

| $x_1$ | $x_2$ | $x_3$ | $x_1$ | $x_2$ | $x_3$ |
|---|---|---|---|---|---|
| 0.00 | 0.27 | 0.73 | 0.00 | 1.00 | 0.00 |
| 0.00 | 1.00 | 0.00 | 0.60 | 0.40 | 0.00 |
| 1.00 | 0.00 | 0.00 | 0.00 | 0.48 | 0.52 |
| 0.22 | 0.78 | 0.00 | 0.45 | 0.22 | 0.34 |
| 0.56 | 0.00 | 0.44 | 1.00 | 0.00 | 0.00 |
| 0.00 | 0.00 | 1.00 | 0.34 | 0.00 | 0.66 |
| 0.00 | 0.00 | 1.00 | 0.45 | 0.54 | 0.01 |
| 0.00 | 0.63 | 0.37 | 0.29 | 0.26 | 0.44 |
| 0.42 | 0.00 | 0.58 | 1.00 | 0.00 | 0.00 |
| 0.26 | 0.47 | 0.27 | 0.31 | 0.33 | 0.35 |

Table 30: Optimal design for 3 ingredients, 20 alternatives and 2 choices in a choice set, when $\boldsymbol{\tau} = (0, 0, 0, 0, 0, 0, 0)$. Choice sets are separated by dashed lines

| $x_1$ | $x_2$ | $x_3$ | $x_1$ | $x_2$ | $x_3$ |
|---|---|---|---|---|---|
| 1.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 |
| 0.00 | 0.60 | 0.40 | 0.00 | 0.34 | 0.66 |
| 0.00 | 0.00 | 1.00 | 0.00 | 0.61 | 0.39 |
| 0.71 | 0.00 | 0.29 | 0.39 | 0.28 | 0.33 |
| 0.51 | 0.49 | 0.00 | 0.38 | 0.25 | 0.37 |
| 0.33 | 0.31 | 0.36 | 0.00 | 1.00 | 0.00 |
| 0.19 | 0.50 | 0.31 | 0.00 | 0.00 | 1.00 |
| 0.35 | 0.00 | 0.65 | 0.00 | 0.68 | 0.32 |
| 1.00 | 0.00 | 0.00 | 0.52 | 0.48 | 0.00 |
| 0.35 | 0.00 | 0.65 | 0.32 | 0.31 | 0.37 |

Table 31: Optimal design for 3 ingredients, 20 alternatives and 2 choices in a choice set, $\boldsymbol{\tau} = (0, 0, 0, 10, 0, 0, 0)$. Choice sets are separated by dashed lines

| $x_1$ | $x_2$ | $x_3$ | $x_1$ | $x_2$ | $x_3$ |
|---|---|---|---|---|---|
| 0.40 | 0.00 | 0.60 | 0.60 | 0.00 | 0.40 |
| 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| 0.00 | 1.00 | 0.00 | 0.00 | 0.60 | 0.40 |
| 0.00 | 0.40 | 0.60 | 0.41 | 0.32 | 0.27 |
| 1.00 | 0.00 | 0.00 | 0.58 | 0.42 | 0.00 |
| 0.48 | 0.52 | 0.00 | 0.27 | 0.36 | 0.36 |

Table 32: Optimal points for the locally optimal utility neutral design with 3 ingredients, 12 alternatives and 2 choices in a choice set for 1,000 random starts. Choice sets are separated by dashed lines

| $x_1$ | $x_2$ | $x_3$ | $x_1$ | $x_2$ | $x_3$ |
|---|---|---|---|---|---|
| 1.00 | 0.00 | 0.00 | 0.36 | 0.00 | 0.64 |
| 0.38 | 0.62 | 0.00 | 0.22 | 0.47 | 0.31 |
| 0.00 | 0.00 | 1.00 | 0.67 | 0.33 | 0.00 |
| 0.70 | 0.00 | 0.30 | 0.35 | 0.20 | 0.45 |
| 0.00 | 0.46 | 0.54 | 0.53 | 0.27 | 0.20 |
| 0.00 | 1.00 | 0.00 | 0.00 | 0.72 | 0.28 |

Table 33: Optimal points for the semi-Bayesian design with 3 ingredients, 12 alternatives and 2 choices in a choice set for 1,000 random starts, when 128 draws are taken from the multivariate normal distribution with $\boldsymbol{\tau}_0 = (0, 0, 0, 0, 0, 0, 0)$ and $\boldsymbol{\Sigma}_0'$ being an identity matrix. Choice sets are separated by dashed lines

| $x_1$ | $x_2$ | $x_3$ | $x_1$ | $x_2$ | $x_3$ |
|---|---|---|---|---|---|
| 0.60 | 0.40 | 0.00 | 0.00 | 0.00 | 1.00 |
| 0.00 | 0.41 | 0.59 | 0.60 | 0.00 | 0.40 |
| 0.00 | 1.00 | 0.00 | 0.31 | 0.35 | 0.35 |
| 0.40 | 0.00 | 0.60 | 0.35 | 0.30 | 0.35 |
| 0.40 | 0.60 | 0.00 | 0.00 | 0.60 | 0.40 |
| 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |

Table 34: Optimal points for the locally optimal utility neutral design with 3 ingredients, 12 alternatives and 3 choices in a choice set for 1,000 random starts. Choice sets are separated by dashed lines

| $x_1$ | $x_2$ | $x_3$ | $x_1$ | $x_2$ | $x_3$ |
|---|---|---|---|---|---|
| 1.00 | 0.00 | 0.00 | 0.32 | 0.68 | 0.00 |
| 0.00 | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 |
| 0.49 | 0.00 | 0.51 | 0.35 | 0.29 | 0.36 |
| 0.01 | 0.35 | 0.64 | 0.00 | 0.60 | 0.40 |
| 0.54 | 0.46 | 0.00 | 0.36 | 0.37 | 0.27 |
| 0.00 | 1.00 | 0.00 | 0.39 | 0.00 | 0.61 |

Table 35: Optimal points for the semi-Bayesian design with 3 ingredients, 12 alternatives and 3 choices in a choice set for 1,000 random starts, when 128 draws are taken from the multivariate normal distribution with $\boldsymbol{\tau}_0 = (0, 0, 0, 0, 0, 0, 0)$ and $\boldsymbol{\Sigma}_0'$ being an identity matrix. Choice sets are separated by dashed lines