Optimal design for probit choice models with dependent utilities

Ulrike Graßhoff\textsuperscript{a}, Heiko Großmann\textsuperscript{b}, Heinz Holling\textsuperscript{c} and Rainer Schwabe\textsuperscript{b}

\textsuperscript{a}School of Business and Economics, Humboldt University Berlin, Berlin, Germany; \textsuperscript{b}Institute of Mathematical Stochastics, Otto-von-Guericke University Magdeburg, Magdeburg, Germany; \textsuperscript{c}Institute for Psychology, University of Münster, Münster, Germany

ABSTRACT
Discrete choice experiments are a popular method to measure part
worths of economic goods and in health science. These models
include several attributes as explanatory variables. The commonly
used multinomial logit model assumes independent utilities for dif-
ferent choice options. In Graßhoff et al. [Optimal design for discrete
choice experiments. J Statist Plann Inference. 2013;143:167–175] we
pointed out that for such a model designs turn out to be formally
optimal which may comprise choice sets containing identical or
nearly identical options and which are not reasonable for use in
empirical discrete choice studies. To overcome this problem we intro-
duce a novel model based on probit part-worth utilities which can
account for similarities in the alternatives by supposing a depen-
dence structure. For this model we derive locally $D$-optimal designs
which appear to be more reasonable for applications.

1. Introduction
Discrete choice analysis is a popular method for analysing preferences and choices in
economics, as well as in social and health sciences because it closely corresponds to mak-
ing choices in everyday situations. In choice experiments respondents have to repeatedly
choose between different alternatives within a so-called choice set. The alternatives also
called options are defined by the levels of a subset of attributes. It is assumed that respon-
dents choose the alternative with the greatest utility. The expected (overall) utility of an
alternative is usually defined as a linear combination of part-worths assigned to the lev-
els of the attributes of an alternative. The part-worths and the expected (overall) utilities
are estimated from the choices of the respondents by using regression models. We do not
attempt to give a review of the theory of discrete choice experiments here, but refer to
Train [1] and the literature cited therein for a comprehensive description and to Soekhai
et al. [2] for a recent survey on their use in health economics.

Usually choice sets consist of two or three alternatives. When two alternatives are pre-
sented, discrete choice analysis coincides with paired comparisons. Typically, the number
of alternatives is held constant for all choice sets within a discrete choice experiment and all respondents will get the same series of choice sets, so the problem of designing the choice sets has to be considered for one respondent only. Otherwise, when respondents get different blocks of choice sets, potential individual effects have to be taken into account which may go beyond a pure block effect. This would lead to further complications in the design of the choice experiment. Initial results in this direction were found recently by Singh et al. [3] and Großmann [4] for paired comparisons (two alternatives), but we will leave this problem aside, here.

Obviously, application of optimal design principles will be important to efficiently estimate the expected utilities represented by the parameters of the regression models. Usually, multinomial logit models have been applied to estimate the expected utilities. Several authors have developed optimal designs for discrete choice models based on multinomial logit models (see [5], for a review). However, the derived designs do not seem to be suitable for many empirical studies. In particular, when the set of attributes comprises several discrete attributes as well as a further quantitative one (see [6]), locally $D$-optimal designs for such multinomial logit models consist of choice sets with alternatives that are identical in all discrete attributes but differ in the quantitative variable [7]. This counter-intuitive result is closely related to the assumption of ‘Independence from Irrelevant Alternatives’ (IIA), characterizing logit regression. According to the IIA property, the relative choice probabilities, i.e., the probability ratios, of any two alternatives in a choice set are independent of all other alternatives contained in this choice set. The validity of the IIA property is accomplished in the multinomial logit model by the assumption that the utilities of the options follow an extreme value (Gumbel) distribution and that all utilities are mutually independent (see [1, Ch. 3]).

However, the IIA assumption seems to be inadequate for many everyday choice situations as the so-called Red-Bus/Blue-Bus Problem illustrates (see [1, p. 46]): Here, a subject can choose between two alternatives to get to work, say a bicycle and a red bus, each having a choice probability of 1/2 which results in a probability ratio of 1 for the two alternatives. Consider now that there is, in addition, a blue bus available as a third option with identical attribute levels as the red bus, except the attribute colour, and the part-worths for the colours red and blue do not differ. Thus also the expected (overall) utilities of the red and the blue bus are identical. According to the IIA property, the probability ratios do not change and, hence, the choice probabilities for the bicycle, the red bus and the blue bus are then 1/3 for each alternative. Such a change in probabilities for using bicycle or bus does not seem reasonable, as the two alternatives of the red and the blue bus are so similar that not merely their expected utilities but also their actual utilities should be close to each other. The required dependence between the utilities can be achieved by a multinomial probit model for which the correlation structure of the utilities is a substantial ingredient (see [1, Ch. 5]). A correlation of 1 between the utilities for the red and blue bus would then lead to more reasonable choice probabilities of 1/2, 1/4 and 1/4 for the bicycle, the red and the blue bus, respectively, which leaves the probabilities for choosing the bicycle or a bus unchanged. For a more detailed description of how to obtain these probabilities we refer to Section 3.1.

In the present article we will analyse whether a multinomial probit model will also yield counter-intuitive $D$-optimal designs when the utilities are independent. Furthermore, we will derive locally $D$-optimal designs for multinomial probit models allowing for
dependencies between the utilities of the alternatives. These models are based on assumptions which seem to be more realistic for most everyday choice situations.

The paper is organized as follows: In the next section two multinomial probit models will be introduced, one without dependent utilities and the other with dependency between the utilities accounting for similarity between alternatives. In Section 3, we will characterize locally $D$-optimal designs for both models in the case of two alternatives, i.e., paired comparisons. First, the case of models including several qualitative attributes will be considered and then the more general case including a further quantitative attribute. In Section 4, the results derived for paired comparisons will be generalized for both probit models including three options. The last section contains a short discussion of the results. All technical details are deferred to the Appendix.

2. Model description

In a choice experiment individual choices are performed between $m \geq 2$ alternatives $a_j$ of a choice set $A = (a_1, \ldots, a_m)$. Each alternative $a_j = (a_{j1}, \ldots, a_{jk})$, $j = 1, \ldots, m$ is characterized by $K$ attributes, where $a_{jk}$ is the level of the $k$th attribute presented in alternative $j$. These attributes describe (potential) properties of an alternative like size, equipment and price of an apartment or various benefits of a health insurance. The decision behaviour of a respondent can be described by a multinomial response $Y = (Y_1, \ldots, Y_m)^\top$, where $Y_j = Y_j(A) = 1$, if the respondent chooses $a_j$ and $Y_j = 0$ otherwise. By $p = p(A)$ we denote the corresponding vector $p = (p_1, \ldots, p_m)^\top$ of probabilities of preference $p_j = p_j(A) = P(Y_j(A) = 1)$ for choosing the $j$th alternative $a_j$ from the choice set $A$.

The probabilities of preference are assumed to depend on latent utilities $U_j = U_j(a_j)$ for all alternatives $a_1, \ldots, a_m$ within the choice set $A$, and the response is assumed to be obtained by the concept of utility maximization, i.e. $Y_j(A) = 1$, if $U_j(a_j) = \max_i U_i(a_i)$. Note that in general $P(U_i = U_j) = 0$, as the utilities typically have continuous distributions, and, hence, the $Y_j$ are almost surely well defined.

The latent utilities $U_j$ are assumed to decompose additively into the expected (overall) utilities $\mu_j = E(U_j)$ and random deviations $\varepsilon_j = U_j - \mu_j$. In the commonly used multinomial logit choice model these random deviations $\varepsilon_j$ are assumed to be mutually independent and identically distributed according to a standard extreme value (Gumbel) distribution with mean 0 and scaling factor 1. Then the utility differences $U_i - U_j$ follow a logistic distribution with mean $\mu_i - \mu_j$ and scale 1 (see e.g., [1, Ch. 3]). In contrast to that, in the multinomial probit model it is assumed that the joint distribution of the latent utilities and, hence, of the utility differences, is a multivariate normal distribution (see e.g., [1, Ch. 5]). This approach has the notable advantage that utilities need not be mutually independent and their dependence can be modelled by the correlation structure between the alternatives. In particular, we make use of the possibility that not only the expected utilities $\mu_j$, but also the overall utilities $U_j$ themselves may be decomposed according to

$$U_j(a_j) = \sum_{k=1}^{K} U_{jk}(a_{jk})$$

into part-worths $U_{jk}$ for the single attributes. Within each alternative $a_j$, the part-worth utilities $U_{jk} = U_{jk}(a_{jk})$ will be assumed to be independent, normally distributed with mean
part-worths $\mu_{jk} = \mu_{jk}(a_j)$, which depend only on the $k$th attribute each. These mean part-worths $\mu_{jk}(a_j) = f_k(a_j)^\top \beta_k$ are specified by linear effects with known regression function vectors $f_k$ and unknown parameter vectors $\beta_k$ for each attribute $k$ separately. Then the latent utility $U_j(a_j)$ of an alternative $a_j$ has mean $\mu_j = \mu_j(a_j) = f(a_j)^\top \beta$ with joint regression function $f(a_j) = (f_1(a_{j1})^\top, \ldots, f_K(a_{jK})^\top)^\top$ and parameter vector $\beta = (\beta_1^\top, \ldots, \beta_K^\top)^\top$, where $\mu_j = \sum_{k=1}^K \mu_{jk}$. Typically the part-worth regression functions $f_k$ will consist of dummy variables for qualitative factors, or they will be linear, if $a_{jk}$ is quantitative.

For simplification we will assume that all part-worth utilities share a common variance $\sigma_0^2$, i.e., $U_{jk} \sim N(f_k(a_{jk})^\top \beta_k, \sigma_0^2)$, throughout this paper, if not stated otherwise.

In what follows it will be crucial to specify the dependence structure between the $m$ utilities $U_1, \ldots, U_m$. For this we consider two particular models implied by different assumptions on the dependence between the part-worth utilities for an attribute $k$ across the alternatives.

Model I: all $U_{jk}$ and $U_{ik}$ are independent.

This model assumes independence of the part-worth utilities irrespectively whether the attributes of two alternatives differ or not and, thus, results in the standard probit model considered in the literature, which may lead to counter-intuitive results similar to those for the common logit model (cf. [7]) as will be seen later. To avoid these problems a second model is introduced, which accounts for dependence when the same level is presented for an attribute in different alternatives to be compared.

Model II: $U_{jk} = U_{ik}$ if $a_{jk} = a_{ik}$; $U_{jk}$ and $U_{ik}$ are independent, if either $k \neq \ell$ or $a_{jk} \neq a_{ik}$ for $k = \ell$.

In this model it is assumed that the presentation of equal levels for an attribute results in identical part-worth utilities in the alternatives presented together. Hence, in Model II attributes with equal levels ($a_{ik} = a_{jk}$) will not contribute to the decision between alternatives $a_i$ and $a_j$ and the utilities $U_i$ and $U_j$ of the alternatives will become dependent. Here, the condition of independent part-worth utilities for differing levels $a_{jk} \neq a_{ik}$ of the same attribute in the alternatives tacitly assumes that these levels are sufficiently distinct. Model II could be extended by allowing for correlations between similar levels which could lead to even more realistic models, but we will leave this possibility out for simplicity.

Under the assumptions of Model I as well as of Model II the $m$-dimensional vector $U = U(A) = (U_1(a_1), \ldots, U_m(a_m))^\top$ of utilities is multivariate normal with mean $\mu(A) = (\mu_1(a_1), \ldots, \mu_m(a_m))^\top$ and covariance matrix $V(A)$. In both models the utilities have equal variances $\text{Var}(U_j) = \sigma_K^2 = K\sigma_0^2$. While in Model I the utilities are independent such that $V(A) = \sigma_K^2 I_m$, where $I_m$ denotes the $m \times m$ identity matrix, the utilities become correlated in Model II, when identical levels occur for some attributes.

According to the concept of utility maximization, the alternative $j$ will be preferred to the other alternatives if the utility $U_j$ is greater than all other utilities $U_i$, $i \neq j$. This implies for the preference probability

$$p_j = p_j(A) = P(Y_j(A) = 1) = P(U_j(a_j) \geq \max_{i \neq j} U_i(a_i))$$

$$= P(U_i(a_i) - U_j(a_j) \leq 0 \quad \text{for all} \quad i \neq j)$$
For fixed $j$ let $L_j$ be the $(m - 1) \times m$ matrix which transforms the $m$-dimensional vector $U$ of utilities to the $(m - 1)$-dimensional vector $U_{(j)} = (U_i - U_j)_{i=1,...,m,i \neq j}$ of relevant utility differences ($U_{(j)} = L_j U$). Then the $(m - 1)$-dimensional vector $U_{(j)}(A)$ of utility differences $U_i(a_i) - U_j(a_j)$ is multivariate normal with mean vector $\mu_j(A) = L_j \mu(A)$ and covariance matrix $V_j(A) = L_j V(A) L_j^\top$.

Hence, the preference probability

$$p_j(A) = \eta(\mu_j(A), V_j(A))$$

can be written as a function of the mean vector $\mu_j$ and the covariance matrix $V_j$, where $\eta(\mu_j, V_j) = \Phi_{m-1}(0; \mu_j, V_j)$ denotes the distribution function of the $(m - 1)$-dimensional normal variate with mean vector $\mu_j$ and covariance matrix $V_j$ evaluated at $0$.

With this notation we can express the $m$-dimensional mean $E(Y(A)) = p(A)$ of the response $Y(A)$ as

$$E(Y(A)) = \eta_A(F(A)\top \beta),$$

where $F(A) = (f(a_1), \ldots, f(a_m))$ is the $p \times m$-dimensional multivariate regression function,

$$\eta_A(\mu) = (\eta(\mu_1, V_1(A)), \ldots, \eta(\mu_m, V_m(A)))^\top,$$

and the additional dependence of $\eta_A$ on $V(A)$ is indicated by the subscript $A$.

The covariance matrix $\text{Cov}(Y)$ of the response vector $Y$ is given by $\Sigma = \Sigma(A; \beta) = \text{diag}(p) - pp^\top$, where $\text{diag}(p)$ is the $m \times m$ diagonal matrix with diagonal entries $p_j$, $j = 1, \ldots, m$.

Hence, both the mean response vector and the covariance matrix of $Y$ depend on the parameter $\beta$ only through the vector of linear effects $\mu = F(A)\top \beta$ and in addition on the $m$ covariance matrices $V_j = V_j(A)$, which only involve the choice set $A$ presented. Thus the observations may be interpreted as outcomes from an extended multivariate generalized linear model.

In this situation the information for a choice set $A$ can be calculated as

$$M(A; \beta) = \left(\frac{\partial \eta_A}{\partial \beta}\right)^\top \Sigma(A; \beta) \frac{\partial \eta_A}{\partial \beta},$$

where $\partial \eta_A / \partial \beta$ denotes the $m \times p$ functional matrix of partial derivatives of the $m$ components of $\eta_A$ with respect to the $p$ components of $\beta$. Remember that $\Sigma$ as well as $\partial \eta_A / \partial \beta$ depend on $\beta$ only through $F(A)\top \beta$.

The chain rule for the differentiation of multidimensional functions leads to

$$\frac{\partial \eta_A(F(A)\top \beta)}{\partial \beta} = J_{\eta_A}(F(A)\top \beta)F(A)^\top$$

where $J_{\eta_A}(\mu)$ is the Jacobian of the function $\eta_A$ evaluated at $\mu$. Thus the information matrix can be written as

$$M(A; \beta) = F(A) J_{\eta_A}(F(A)\top \beta) \Sigma(A; \beta)^{-1} J_{\eta_A}(F(A)\top \beta)F(A)^\top = F(A) \Lambda(A; \beta) F(A)^\top,$$

where $\Lambda = J_{\eta_A}^\top \Sigma^{-1} J_{\eta_A}$ denotes the $m \times m$ intensity matrix.
To tackle the problem of finding an optimal design, i.e., the best possible selection of choice sets, we will make use of the approximate design theory introduced by Kiefer (see [8]): An approximate design \( \xi \) on the set \( X \) of all choice sets consists of, say, \( n \) different choicesets \( A_i = (a_{i1}, \ldots, a_{im}) \) with weights \( w_i \geq 0 \) and \( \sum_{i=1}^{n} w_i = 1 \), representing the relative frequencies of replications. The normalized per observation information matrix is defined by

\[
M(\xi; \beta) = \sum_{i=1}^{n} w_i M(A_i; \beta) = \sum_{i=1}^{n} w_i F(A_i) A(A_i; \beta) F(A_i)^\top.
\]

Note that for an exact design the usual information matrix equals \( N \) times the normalized one, where \( N \) is the total number of observations (presentations of choice sets).

To measure the quality of a design we will make use of the most common criterion of \( D \)-optimality, i.e., we are looking for designs \( \xi^* \) that are locally \( D \)-optimal at \( \beta \), which maximize the determinant of the information matrix \( M(\xi; \beta) \) (see e.g., [9]).

3. Paired comparisons

First we will focus on the particular case of \( m = 2 \) alternatives, which represents the probit paired comparison model: The choices are performed between two alternatives \( a_1 \) and \( a_2 \) of a pair \( A = (a_1, a_2) \). Because of \( Y_2 = 1 - Y_1 \) and \( p_2 = 1 - p_1 \) we actually have to deal with only one preference probability \( p = p_1 \) for the first alternative in a pair. The mean of the binomial response variable \( Y = Y_1 \) is given by a one-dimensional function \( \eta = \eta_1 \), which leads to an extended generalized linear model with

\[
E(Y(A)) = \eta(\tilde{f}(A)^\top \beta, \sigma^2(A)) = \Phi_0(\tilde{f}(A)^\top \beta / \sigma(A)) ,
\]

where \( \tilde{f}(A) = f(a_2) - f(a_1) \), \( \Phi_0 \) denotes the standard normal distribution function and the variance \( \sigma^2(A) = \text{Var}(U_1(a_1) - U_2(a_2)) \) is the one-dimensional counterpart of the covariance matrix \( V_1(A) \). The variance of the response is given by \( \text{Var}(Y(A)) = p(A)(1 - p(A)) \).

In the present case we have for the derivative

\[
\frac{\partial \eta(\tilde{f}(A)^\top \beta, \sigma^2(A))}{\partial \beta} = \frac{\varphi_0(\tilde{f}(A)^\top \beta / \sigma(A))}{\sigma(A)} \tilde{f}(A)^\top ,
\]

where \( \varphi_0 \) is the density of the standard normal distribution. Hence, the information for a pair \( A \) is given by

\[
M(A; \beta) = \lambda(A; \beta) \tilde{f}(A) \tilde{f}(A)^\top
\]

with intensity function

\[
\lambda(A; \beta) = \varphi_0(\tilde{f}(A)^\top \beta / \sigma(A))^2 / \sigma^2(A) \Phi_0(\tilde{f}(A)^\top \beta / \sigma(A))(1 - \Phi_0(\tilde{f}(A)^\top \beta / \sigma(A))) ,
\]

which depends on \( \beta \) only through the linear component \( \tilde{f}(A)^\top \beta \) and additionally on the scaling factor \( \sigma(A) \).
3.1. Qualitative attributes in the case of indifference

To start, we consider in this subsection the special case $\beta = 0$, which results in equal choice probabilities $p = 1 - p = 1/2$ for any pair of alternatives, which can be interpreted as the situation of indifference.

Under the assumption of Model I we have constant variance $\sigma^2 = \sigma^2(A) = 2K\sigma_0^2$ for all pairs $A$. Then for an approximate design $\xi$ the information matrix

$$M(\xi; 0) = \frac{1}{\pi K\sigma_0^2}M_L(\xi)$$

is proportional to the information matrix $M_L(\xi) = \sum_{i=1}^{n} w_i \tilde{f}(A_i)\tilde{f}(A_i)^\top$ in the corresponding linear paired comparison model (see [10]). As a consequence any $D$-optimal design in the linear paired comparison model is also $D$-optimal in the probit paired comparison model, when all utility terms $U_{1k}$ and $U_{2\ell}$ are assumed to be independent and $\beta = 0$.

Under the assumptions of Model II the comparison depth $d_A$ will play an important role, where $d_A = \#\{k; a_{1k} \neq a_{2k}\}$ is defined as the number of attributes for which the components differ within the pair $A = (a_1, a_2)$. With this notation the variance of the utility difference can be written as $\sigma^2(A) = 2d_A\sigma_0^2$.

To simplify the problem further we consider a setting of $K$ qualitative factors, which may be adjusted to the same number $v_k = v$ of levels 1, $\ldots$, $v$, say, for each attribute $k$. The vector $f = (f_1^\top, \ldots, f_K^\top)^\top$ of part-worth regression functions is chosen according to effects coding. More precisely, $f_k(i) = e_{v-1;i}$, if $i = 1, \ldots, v-1$, where $e_{v-1;i}$ denotes the $i$th unit vector of length $v-1$, and $-1$ is the vector of length $v-1$ with all entries equal to 1 (for more details on this model specifications see [10]).

Since in the present situation the $D$-criterion is invariant with respect to both permutations of the levels for each attribute and to permutations of the attributes themselves, optimal designs can be found within the class of invariant designs which are uniform on the orbits induced by these permutations. These orbits are the sets of pairs with a fixed comparison depth $d \leq K$.

By $\tilde{\xi}_d$ we denote the design which is uniform on the orbit of comparison depth $d$. In particular, for full comparison depth $d = K$ the uniform design $\tilde{\xi}_K$ is the product type design $\xi_0 \otimes \cdots \otimes \xi_0$, where $\tilde{\xi}_0$ is the uniform balanced incomplete block design with blocks of size 2 consisting of the $v(v-1)$ pairs concerning one single attribute. Thus $\tilde{\xi}_K$ is uniform on all $(v(v - 1))^K$ pairs, which have different levels in each attribute. Graßhoff et al. [10] established that the design $\tilde{\xi}_K$ is $D$-optimal in the linear paired comparison model and, thus, it is also optimal in Model I. In that case the optimal information matrix equals $M(\tilde{\xi}_K; 0) = (I_K \otimes M^*)/(\pi K\sigma_0^2)$, where $M^* = \frac{1}{v-1}(I_{v-1} + 1_{v-1}1_{v-1}^\top)$ is the information matrix of the marginal design $\xi_0$ in the single attribute linear paired comparison model and ‘$\otimes$’ is the symbol for the Kronecker product of matrices.

In contrast to Model I, under the assumptions of Model II the intensity $1/(\pi d\sigma_0^2)$ depends on the comparison depth $d$ for pairs belonging to the orbit of comparison depth $d \geq 1$. Simple combinatorial arguments lead to the information matrix

$$M(\tilde{\xi}_d; 0) = \frac{d}{K\pi d\sigma_0^2}(I_K \otimes M^*) = M(\tilde{\xi}_K; 0).$$
Note that for \( d = 0 \) all attributes and, hence, both alternatives and their corresponding utilities completely coincide. Therefore the resulting information is equal to zero \((M(\xi_0; 0) = 0)\).

Since \( M(\xi_d; 0) \) is independent of the comparison depth \( d \geq 1 \), all designs \( \tilde{\xi}_d \) and, in particular, the design \( \tilde{\xi}_K \), which is \( D \)-optimal under Model I, are also \( D \)-optimal under Model II. Furthermore, any convex combination of the designs \( \tilde{\xi}_d, d \geq 1 \), is also \( D \)-optimal under Model II. Note that the number of pairs with comparison depth \( d \geq 1 \) is equal to \((\binom{K}{d})v^K(v - 1)^d\) such that small values of \( d \) are preferable in terms of the number of different pairs required when the number \( v \) of levels is large, while \( d = K \) is to be preferred for \( v = 2 \).

### 3.2. One additional quantitative attribute

We extend the model to the situation investigated by Kanninen [6], which led to counter-intuitive results in the logit model after design optimization for larger choice sets (see [7]) and which caused us to introduce Model II.

The purpose of the present subsection is to provide optimal designs for probit paired comparison models with and without dependence structure in the part-worth utilities before studying larger choice sets. More precisely, we consider a model with pairs \( A = (a_1, a_2) \) of alternatives, where one of the attributes, say the last one, is quantitative and unrestricted and can be interpreted, for example, as a price variable (potentially on a logarithmic scale) and all other attributes are qualitative. Then the set of attributes can be split into two components \( a_j = (x_j^1, t_j)^T \), where \( t_j \in \mathbb{R} \) and \( x_j \) consists of the qualitative attributes. According to the marginal pairs \( x = (x_1, x_2) \) and \( t = (t_1, t_2) \) we can decompose the regression function for \( Y_1 \) as

\[
\tilde{f}(A) = (\tilde{f}_1(x)^T, \tilde{f}_2(t))^T,
\]

where the marginal regression functions are defined by \( \tilde{f}_1(x) = f(x_1) - f(x_2), \tilde{f}_2(t) = t_1 - t_2 \), and for the qualitative attributes the regression function \( f_1 = (f_1^1, \ldots, f_1^K)^T \) is defined as the vector of part-worth regression functions as in Subsection 3.1. The parameter vector \( \beta \) is partitioned accordingly, \( \beta = (\beta_t^1, \beta_t^2)^T \), where \( \beta_t^1 = (\beta_1^T, \ldots, \beta_{1K}^T)^T \) collects the parameters for the part-worths of the qualitative attributes.

Following Kanninen [6] we restrict our investigations for the first component to the setting of \( K \) binary attributes, varying on \( v = 2 \) levels each, i.e., \( x_j = (x_{j1}, \ldots, x_{jk})^T \in \{1, 2\}^K \). Under effects coding the corresponding regression functions are given by \( f_k(x_j) = (f_k(x_{jk}))_{k=1,\ldots,K} \) with \( f_k(1) = 1 \) and \( f_k(2) = -1 \).

The utility \( U_j(a_j) \) in this two component model is generated by partial utilities

\[
U_j(a_j) = U_{j1}(x_j) + U_{j2}(t_j),
\]

where the partial utility \( U_{j1}(x_j) = \sum_{k=1}^K U_{j1k}(x_{jk}) \) of the first component is itself composed of part-worth utilities \( U_{j1k} \) related to the qualitative attributes. These part-worth utilities \( U_{j1k} \) are assumed to be independent within alternatives and to be normally distributed with mean \( f_k(x_{jk})\beta_{1k} \) and constant variance \( \sigma_0^2 \) across the attributes, as in the previous subsection. For the second component we assume a normally distributed part-worth utility \( U_{j2} \) with mean \( \beta_{2} t_j \) and variance \( \sigma_1^2 \geq 0 \), which is independent of the part-worth utilities \( U_{j1k} \) of the first component. Furthermore we will assume throughout that all part-worth
utilities $U_{j2}$ for the second component are independent. Alternatively, for the continuous component a stationary correlation depending on the distance $t_i - t_j$ could be assumed, but will not considered further here. As a special case we may allow for a sharp decision with respect to the quantitative attribute by letting $\sigma_t^2 = 0$, which results in a degenerate utility $U_{j2} \equiv \beta_2 t_j$. This accounts for the possibility that, for example, a smaller value of $t$ is always deemed preferable ($\beta_2 < 0$), such as a lower price or alike.

Optimal designs for such a two component model were first investigated numerically by Kanninen [6] in the binomial logit model. Graßhoff et al. [11] gave explicit proofs for locally D-optimal designs by making use of a canonical transformation introduced by Ford et al. [12] and extended by Sitter and Torsney [13] to the multi-factorial case. We will apply this construction method also to the probit models considered here.

To this end, in a first step the standardized case $\beta_1 = 0$ and $\beta_2 = 1$ is considered. There the intensity function $\lambda$ for a pair $A$ reduces to $\lambda(A; \beta) = \lambda_2((t_1 - t_2)/\sigma(A))/\sigma^2(A)$, where $\lambda_2(z) = \varphi_0(z)^2/(\Phi_0(z)(1 - \Phi_0(z)))$ is the marginal intensity with respect to the quantitative attribute. Hence, the intensity $\lambda(A; \beta)$ depends on the first component $x$ only through the scaling factor $\sigma(A)$.

The situation of independent utilities of Model I results in the standard probit model in the literature: If the part-worth utilities $U_{j1}$ of the first components satisfy the assumptions of Model I, then $\sigma(A) = \sigma_{\text{max}}$ say, attains the same value for all pairs $A$, where $\sigma_{\text{max}}^2 = 2(\sigma_K^2 + \sigma_c^2)$ and, again, $\sigma_K^2 = K\sigma_0^2$. Thus, the intensity function only depends on the linear response through the second component, and the approach described in Graßhoff et al. [11] can be used. As $\beta$ and $\sigma_{\text{max}}$ may be rescaled simultaneously without changing the model, we will assume $\sigma_{\text{max}} = 1$ throughout to ensure identifiability of $\beta$, as is common in the literature. This standardization does not have any effect on design optimality because it only results in a scaling factor for the information matrix. However, in statistical analysis the standardization may cause problems when parameters are compared across models with different numbers of attributes or different correlation structures.

To formalize the result, denote by $\delta_1$ the one-point design at $t = (t_1, t_2)$.

**Theorem 3.1:** Let $z^* > 0$ maximize $\lambda_2(z)^K z^2$ and let $t^*$ satisfy $t_1^* - t_2^* = z^*$. Then the design $\xi^* = \bar{\xi}_K \otimes \delta_1^*$ is locally D-optimal at $\beta = (0, 1)^T$ in the probit paired comparison model with independent part-worth utilities (Model I).

The proof is given in the Appendix. Table 1 lists the optimal values $z^*$ together with the corresponding preference probabilities $p = \Phi_0(z^*)$ for various numbers $K$ of attributes for the first component.

As an example, for $K = 1$, by Theorem 3.1 the locally D-optimal design $\xi^*$ assigns equal weights 1/2 to the pairs of alternatives $((1, c + 1.138), (2, c))$ and $((2, c + 1.138), (1, c))$ for some $c$. Actually, $c$ might differ from pair to pair. Hence, also a design with weights 1/2 on $((1, c_1 + 1.138), (2, c_1))$ and $((2, c_2 + 1.138), (1, c_2))$ would be optimal, where $c_1$ and $c_2$ can be chosen to produce meaningful alternatives.

**Table 1.** Optimal values $z^*$ and preference probabilities $\Phi_0(z^*)$.

| $K$ | 1  | 2  | 3  | 4  | 7  | 10 | 100 |
|-----|----|----|----|----|----|----|-----|
| $z^*$ | 1.138 | 0.938 | 0.816 | 0.732 | 0.581 | 0.497 | 0.165 |
| $\Phi_0(z^*)$ | 0.872 | 0.826 | 0.793 | 0.768 | 0.719 | 0.690 | 0.566 |
Table 2. Optimal fractional factorial design under indifference, \( K = 3 \).

| \( a_1 \) | \( a_2 \) |
| --- | --- |
| \( A_1 \) | (1, 1, 1, \( c_1 + 0.816 \)) \( (2, 2, 2, c_1) \) |
| \( A_2 \) | (1, 2, 2, \( c_2 + 0.816 \)) \( (2, 1, 1, c_2) \) |
| \( A_3 \) | (2, 1, 2, \( c_3 + 0.816 \)) \( (1, 2, 1, c_3) \) |
| \( A_4 \) | (2, 2, 1, \( c_4 + 0.816 \)) \( (1, 1, 2, c_4) \) |

For general \( K \), the locally \( D \)-optimal design \( \xi^* \) assigns equal weights \( 1/2^K \) to the \( 2^K \) pairs which differ in all \( K \) attributes of the first component and which have values \( t^*_1 = c + z^* \) and \( t^*_2 = c \) in the second component for some \( c \). Again, \( c \) may differ from pair to pair. If \( K \) increases, the number \( 2^K \) of different pairs of alternatives becomes prohibitively large for the optimal product-type design \( \xi^* \) of Theorem 3.1. Therefore reduction methods like fractional factorials may be used to obtain the same information matrix as for \( \xi^* \) and, hence, optimal designs with a smaller number of distinct pairs of alternatives. For example, in the case \( K = 3 \), the structure of a \( 2^{4-1} \) fractional factorial yields a locally \( D \)-optimal design which assigns equal weights \( 1/4 \) to the four pairs \( A_i = (a_{i1}, a_{i2}) \) of alternatives shown in Table 2.

Note that optimality is retained when \( z^* \) is replaced by \( -z^* \) in the above designs for all pairs simultaneously. It is also worthwhile mentioning that the particular optimal values \( \pm z^* = \pm 1.138 \) for \( K = 1 \) coincide with the values for the optimal locations in the standard probit regression model (see [12]).

As we will see in Section 4, the model with independent utilities (i.e., Model I for the \( K \) qualitative attributes) may lead to counter-intuitive results, similar to those in the multinomial logit model (see [7]), when choice sets with more than two alternatives are considered. To overcome this problem, we introduce a two component model, where the first component fulfills the assumptions of Model II, but the part-worth utilities \( U_{12} \) and \( U_{22} \) will still be assumed to be independent (potentially degenerate). Then the scaling factor \( \sigma (A) \) is obtained by

\[
\sigma^2(A) = 2(d\sigma_0^2 + \sigma_t^2) = (d\sigma_{\max}^2 + (K - d)2\sigma_t^2)/K
\]

for pairs \( A \) belonging to an orbit of comparison depth \( d \) in the qualitative attributes, where \( \sigma_{\max}^2 = 2(K\sigma_0^2 + \sigma_t^2) \) is the maximal possible variance, which is achieved when \( A \) has comparison depth \( K \). Again we assume \( \sigma_{\max} = 1 \) for reasons of identifiability. Hence, it can be seen that the optimal design \( \xi^* \) of Theorem 3.1 for Model I is also optimal under the assumptions of Model II for any value of \( \sigma_0^2 \) and \( \sigma_t^2 \).

**Theorem 3.2:** If \( z^* > 0 \) maximizes \( \lambda_2(z)^K + 1 z^2 \) and if \( t^* \) satisfies \( t^*_1 - t^*_2 = z^* \) then the design \( \tilde{\xi}_K \otimes \delta_{t^*} \) is locally \( D \)-optimal for the probit paired comparison model with dependent utilities (Model II).

In Theorems 3.1 and 3.2 we focused on the standardized case \( \beta_1 = 0 \) and \( \beta_2 = 1 \). In the following we will indicate how these results can be transferred to arbitrary values for \( \beta_1 \) and \( \beta_2 \). There we have to suppose \( \beta_2 \neq 0 \) in order to guarantee the existence of a finite solution of the design optimization problem. According to Graßhoff et al. [11], locally \( D \)-optimal designs can be constructed by using the concept of canonical transformations (see [12,13]).
The procedure is based on a one-to-one mapping \( g \) defined by \( g(a_j) = (x_j^T, f_1(x_j)^T \beta_1 + t_j \beta_2)^T \) on the alternatives, which transforms the parameters to the case of indifference for the qualitative attributes. The simultaneous transformation \( g(A) = (g(a_1), g(a_2)) \) of both alternatives induces a linear transformation \( \tilde{f}(g(A)) = Q_S f(A) \) of the induced regression functions with

\[
Q_S = \begin{pmatrix} I_K & 0 \\ \beta_1^T & \beta_2 \end{pmatrix}.
\]

If we let \( z_j = f_1(x_j)^T \beta_1 + t_j \beta_2 \) for the unrestricted quantitative component in the transformed model, the information matrix coincides with the standardized situation (\( \beta_1 = 0 \) and \( \beta_2 = 1 \)). Then optimal designs can be obtained by a back transformation of the optimal design \( \tilde{\xi}_K \otimes \delta_t \) for the standardized situation: The induced design defined by \( \xi^*(x, t) = \tilde{\xi}_K(x) \delta_t(g(x, t)) \) is then locally \( D \)-optimal at given \( \beta = (\beta_1^T, \beta_2^T) \), which establishes the following result.

**Corollary 3.1:** Let \( z^* \) maximize \( \lambda_2(z)^{K+1}z^2 \). Denote by \( \tilde{\xi}_{2|1}^* \), the conditional design which is concentrated on \( t^*(x) \) for every pair \( x \), where \( t^*(x) = (t_1^*(x), t_2^*(x)) \) satisfies \( t_1^*(x) - t_2^*(x) = (z^* - f_1(x_1))^T \beta_1 + f_1(x_2)^T \beta_1)/\beta_2 \). Then the combined design \( \xi^* = \tilde{\xi}_K \otimes \xi_{2|1}^* \) is locally \( D \)-optimal under both model assumptions I and II of independent or dependent utilities, respectively.

If is worthwhile mentioning that also in the case of general \( \beta \) the optimal values \( t^*(x) \) for the second component are chosen in such a way that the optimal preference probabilities \( p = P(Y(A) = 1) = \Phi_0(z^*) \) of Table 1 are retained.

But note that the differences \( t_1^*(x) - t_2^*(x) \) of the optimal settings \( t_1^*(x) \) and \( t_2^*(x) \) for the second component depend on the settings \( x = (x_1, x_2) \) of the first component and may, hence, vary across the pairs of alternatives. For example, in the case \( K = 1 \), by Corollary 1 the locally \( D \)-optimal design \( \tilde{\xi}^* \) assigns equal weights 1/2 to the pairs of alternatives \(((1, c_1 + (1.138 - \beta_1)/\beta_2), (2, c_1 + \beta_1/\beta_2))\) and \(((2, c_2 + (1.138 + \beta_1)/\beta_2), (1, c_2 - \beta_1/\beta_2))\).

The back transformation of Corollary 3.1 may also be applied to designs with reduced support. For example, in the case \( K = 3 \), the locally \( D \)-optimal design can be derived from the design under indifference in Table 2. The transformed design then assigns equal weights 1/4 to the four transformed pairs \( A_i = (a_{i1}, a_{i2}) \) of alternatives given in Table 3.

### 4. Choice sets with three alternatives

We turn now to the situation of choice sets with \( m = 3 \) alternatives. In contrast to paired comparisons, there a reduction to one dimension is no longer possible, and we have to...
deal with proper multinomial observations. To compute the preference probabilities \( p_j \) for a choice set \( A = (a_1, a_2, a_3) \) we use the software package \texttt{mvtnorm} implemented in \texttt{R} (see [14,15]) for obtaining the multivariate normal probabilities in the variance terms.

For abbreviation we denote by \( \sigma_{ij}^2(A) = \sigma_{ji}^2(A) \) the diagonal elements \( \text{Var}(U_i - U_j) \) of the covariance matrix \( V_j \) and introduce the standardized mean differences

\[
  z_{ij}(A) = ((f_i(a_i) - f_j(a_j))\top \beta) / \sigma_{ij}(A).
\]

Further let \( \Phi_\varphi \) be the bivariate normal distribution function with location vector zero, scaling parameters one and correlation coefficient \( \varphi \) and denote by \( \varphi_j(A) = \text{corr}(U_i - U_j, U_\ell - U_j) \) the correlation in the covariance matrix \( V_j \). With this notation the preference probabilities can be rewritten as

\[
  p_j(A) = \Phi_{\varphi_j(A)}(z_{ji}(A), z_{j\ell}(A)),
\]

where the indices \( i \) and \( \ell \) denote the alternatives other than \( j \). Then the Jacobian matrix \( J_{\eta_A} \) can be computed as

\[
  J_{\eta_A}(F(A)\top \beta) = \begin{pmatrix}
  h_{12} + h_{13} & -h_{12} & -h_{13} \\
  -h_{21} & h_{21} + h_{23} & -h_{23} \\
  -h_{31} & -h_{32} & h_{31} + h_{32}
\end{pmatrix},
\]

where

\[
  h_{ij} = h_{ij}(A) = \varphi_0(z_{ij})\Phi_0((z_{i\ell} - \varphi_0(z_{ij}))/((1 - \varphi^2_i)^{1/2})/\sigma_{ij}(A)
\]

and \( \ell \) is the index of the third alternative other than \( i \) and \( j \).

4.1. Qualitative attributes in the case of indifference

Also here we first consider the particular case \( \beta = 0 \) of indifference for the setting of \( K \) qualitative attributes as in the corresponding subsection on paired comparisons. However, for simplification we additionally restrict the discussion here to the case of \( \nu = 2 \) levels for each attribute.

As will be seen, the intensity matrix \( \Lambda \) and the assumptions of Model I and II will not be affected under indifference when levels are permuted within attributes and attributes are permuted with each other. Then also the \( D \)-criterion is invariant with respect to these permutations. Hence, as in the paired comparison case, optimal designs can be found within the class of invariant designs, which are uniform on the orbits induced by the permutations.

In order to characterize these orbits, we introduce a multivariate analogue to the concept of comparison depth for paired comparisons. For any choice set \( A = (a_1, a_2, a_3) \) we denote by \( d_{ij} = d_{ij}(A) \) the number of attributes for which the levels of the alternatives \( a_i \) and \( a_j \) differ, i.e. \( d_{ij} \) is the comparison depth of the pair \((a_i, a_j)\). The triple \( d = d(A) = (d_{12}, d_{13}, d_{23}) \) will be called the comparison depth of the choice set. Note that each attribute contributes either zero to the comparison depth if all alternatives coincide in this attribute, or it adds 1 to two components of the comparison depth vector \( d \) if two alternatives are equal and the third one differs in this attribute. Thus it is easy to see that the mean comparison depth \( D = (d_{12} + d_{13} + d_{23})/2 \) satisfies \( D \leq K \).
In the following we will only consider choice sets with full profiles for which the mean comparison depth is maximal \((D = K)\), as choice sets with partial profiles \((D < K)\) for which, at least, one attribute is equal across all alternatives tend to provide less information (see [7], for the logistic case).

All orbits are characterized by their comparison depth \(d\). Because a permutation of the arrangement of the alternatives within a choice set does not affect the corresponding information matrix, an orbit described by the comparison depth \(d\) can be considered as being equivalent to an orbit associated with a permutation of the entries in \(d\): For example in the situation of two identical alternatives the orbit \(d = (K, K, 0)\) indicates that alternative 2 equals alternative 3, whereas on the orbit \(d = (K, 0, K)\) alternative 1 is equal to alternative 3 and on \(d = (0, K, K)\) alternative 1 and 2 are coincide. For each of these cases the remaining alternative differs in all attributes. Hence, without loss of generality we need only consider comparison depths satisfying \(d_{12} \geq d_{13} \geq d_{23}\).

For the uniform design \(\bar{\xi}_d\) on the orbit \(d = (d_{12}, d_{13}, d_{23})\) the information matrix

\[
M(\bar{\xi}\_d) = 4\lambda_d I_K
\]

is a multiple of the identity matrix. The diagonal elements are given by the mean intensity

\[
\lambda_d = \frac{1}{2K} \sum_{j=1}^{3} (d_{jj} + d_{j\ell} - d_{i\ell})\lambda_{jj}(d) = \frac{1}{K} \sum_{j=1}^{3} (K - d_{i\ell})\lambda_{jj}(d),
\]

where again the indices \(i\) and \(\ell\) denote the alternatives other than \(j\) and the \(\lambda_{jj}(d)\) are the diagonal entries of the intensity matrix \(\Lambda(d)\) on the orbit \(d\). Note that for the off-diagonal entries of \(\Lambda\) the relation \(2\lambda_{ij} = \lambda_{i\ell} - \lambda_{ij}\lambda_{jj}\) holds. The determinant \(\det M(\bar{\xi}\_d)\) of the information matrix will then be maximized by the uniform design on the orbit \(d\), which yields the largest value of \(\lambda_d\).

Under the assumption of Model I we observe that the variances \(\sigma_{ij}^2(A) = 2K\sigma_0^2\) for the utility differences \(U_i - U_j\) and the correlations \(\varrho_i(A) = 1/2\) do not depend on the particular choice set \(A\). Additionally, in the present case of indifference the preference probabilities are equal \((p_1 = p_2 = p_3 = 1/3)\) and the intensity matrix amounts to \(\Lambda = 9(3I_3 - I_3 I_3^\top)/(8\pi)\) for every choice set \(A\) and is thus constant within and across the orbits. Hence, for each comparison depth \(d\) the uniform design \(\bar{\xi}_d\) on its orbit has the information matrix \(M(\bar{\xi}\_d) = 9I_K/\pi\), which is independent of the orbit. Consequently, any design \(\bar{\xi}_d\) is \(D\)-optimal as well as any convex combination thereof. This proves the following result.

**Theorem 4.1:** In the case of indifference \((\beta = 0)\), every design which is uniform on orbits with mean comparison depth \(D = K\) is locally \(D\)-optimal under Model I of independent utilities.

Under the assumptions of Model II the variances and correlations of the utility differences may vary with the orbits described by \(d\). For a choice set \(A\) with comparison depth \(d\) we get \(\sigma_{ij}^2(A) = d_{ij}/K\). If additionally \(d_{ij} > 0\) for all \(i\) and \(j\), we obtain for the correlations

\[
\varrho_i(A) = (K - d_{j\ell})/\sqrt{d_{ij}d_{i\ell}} \geq 0.
\]
Consequently, the intensity matrix $\Lambda$ does not vary for the choice sets within an orbit. Then it can be seen that the mean intensity becomes

$$\lambda_d = \frac{1}{4\pi} \sum_{j=1}^{3} \frac{1 + \varrho_j}{p_j},$$

where the preference probabilities are given by

$$p_j = \Phi_{\varrho_j}(0) = \frac{1}{4} + \frac{\arcsin(\varrho_j)}{2\pi}.$$ 

It is worthwhile mentioning that under the assumptions of Model II the individual alternatives need not have equal preference probabilities even in the case of ‘indifference’ ($\beta = 0$) due to the correlations between the utilities.

The situation of a choice set with two identical alternatives with comparison depth $d = (K, K, 0)$ can be covered by the paired comparison case of Section 3. In particular, the red bus/blue bus problem mentioned in the introduction fits in this context with $K = 1$, when the alternatives ‘red bus’ and ‘blue bus’ are considered to be identical and all three alternatives have the same expected utility (‘indifference’). Then we can see that under Model II the preference probabilities are $1/2$ for the bicycle and $1/4$ for each bus, while under Model I the preference probabilities would have been $1/3$ for each of the alternatives ‘bicycle’, ‘red bus’ and ‘blue bus’. Hence, the dependence structure of Model II resolves the red bus/blue bus problem stated in the literature, while just changing the link function (Model I) does not help.

In Table 4 we present the preference probabilities and the values $\det(M(\tilde{\xi}_d))^{1/K}$ of the criterion function together with the corresponding efficiencies $\text{eff}(\tilde{\xi}_d) = (\det(M(\tilde{\xi}_d)))/\det(M(\tilde{\xi}_d^*))^{1/K} = \lambda_d/\lambda_{d^*}$ for $K = 2, \ldots, 7$ attributes and all possible comparison depths $d$ with $d_{12} \geq d_{13} \geq d_{23}$. For each number $K$ of attributes the optimal comparison depths $d^*$ are highlighted in bold.

In the particular situation $d = (K, K, 0)$ the alternatives 2 and 3 are indistinguishable and either of them may be chosen if $U_2 = U_3 > U_1$, which occurs with probability $1/2$ as $U_1$ and $U_2$ are independent and identically distributed. Then for the preference probabilities we have $p_1 = 1/2 = p_2 + p_3$, and the value of the criterion function equals $\det(M(\tilde{\xi}_d))^{1/K} = 8/\pi$.

From Table 4 we can deduce that for $K \leq 7$ the maximal value of $\det(M(\tilde{\xi}_d))$ is achieved for designs that are concentrated on those orbits where the numbers of attributes in which any two alternatives differ are as balanced as possible. It can be shown by convexity arguments that this statement holds true for all $K$ which are multiples of three. In this case, the optimal orbit is specified by $d_{ij} = 2K/3$ for all pairs of alternatives. We conjecture that this result will be valid for any number of attributes $K$.

Note that the efficiencies of the choice sets with identical alternatives (comparison depth $d = (K, K, 0)$) are remarkably low.

### 4.2. One additional quantitative attribute

For a model similar to that introduced by Kanninen [6] we augment the above model with an additional continuous attribute $t$ as in Subsection 3.2. For each alternative the set of
Table 4. Qualitative attributes: characteristics of uniform designs $\bar{\xi}_d$ for all comparison depths $d$ with full profiles under Model II.

| $K$ | $d_{12}$ | $d_{13}$ | $d_{23}$ | $\rho_1^*$ | $\rho_2^*$ | $\rho_3^*$ | $\text{det}(M(\bar{\xi}_d))^{1/K}$ | eff  |
|-----|---------|---------|---------|---------|---------|---------|----------------|-----|
| 2   | 2       | 2       | 0       | 0.500   | (0.500) | 2.546   | 2.546          | 0.610 |
| 2   | 2       | 1       | 1       | 0.375   | 0.375   | 0.250   | 4.171          | 1.000 |
| 3   | 3       | 3       | 0       | 0.500   | (0.500) | 2.546   | 2.546          | 0.939 |
| 3   | 2       | 2       | 1       | 0.402   | 0.348   | 0.250   | 4.154          | 0.967 |
| 3   | 2       | 2       | 2       | 0.333   | 0.333   | 0.333   | 4.297          | 1.000 |
| 4   | 4       | 4       | 0       | 0.500   | (0.500) | 2.546   | 2.546          | 0.950 |
| 4   | 4       | 3       | 1       | 0.417   | 0.333   | 0.250   | 4.131          | 0.966 |
| 4   | 4       | 2       | 2       | 0.375   | 0.375   | 0.250   | 4.171          | 0.975 |
| 4   | 3       | 3       | 2       | 0.366   | 0.317   | 0.317   | 4.278          | 1.000 |
| 5   | 5       | 5       | 0       | 0.500   | (0.500) | 2.546   | 2.546          | 0.950 |
| 5   | 5       | 4       | 1       | 0.426   | 0.324   | 0.250   | 4.111          | 0.960 |
| 5   | 5       | 3       | 2       | 0.391   | 0.359   | 0.250   | 4.165          | 0.973 |
| 5   | 4       | 4       | 2       | 0.385   | 0.308   | 0.308   | 4.249          | 0.992 |
| 5   | 4       | 3       | 3       | 0.348   | 0.348   | 0.304   | 4.282          | 1.000 |
| 6   | 6       | 6       | 0       | 0.500   | (0.500) | 2.546   | 2.546          | 0.950 |
| 6   | 6       | 5       | 1       | 0.433   | 0.317   | 0.250   | 4.094          | 0.953 |
| 6   | 6       | 4       | 2       | 0.402   | 0.348   | 0.250   | 4.154          | 0.967 |
| 6   | 6       | 3       | 3       | 0.375   | 0.375   | 0.250   | 4.171          | 0.971 |
| 6   | 5       | 5       | 2       | 0.398   | 0.301   | 0.301   | 4.223          | 0.983 |
| 6   | 5       | 4       | 3       | 0.367   | 0.336   | 0.297   | 4.267          | 0.993 |
| 6   | 4       | 4       | 4       | 0.333   | 0.333   | 0.333   | 4.297          | 1.000 |
| 7   | 7       | 7       | 0       | 0.500   | (0.500) | 2.546   | 2.546          | 0.950 |
| 7   | 7       | 6       | 1       | 0.438   | 0.312   | 0.250   | 4.079          | 0.951 |
| 7   | 7       | 5       | 2       | 0.410   | 0.340   | 0.250   | 4.143          | 0.966 |
| 7   | 7       | 4       | 3       | 0.386   | 0.364   | 0.250   | 4.168          | 0.972 |
| 7   | 6       | 6       | 2       | 0.407   | 0.297   | 0.297   | 4.201          | 0.979 |
| 7   | 6       | 5       | 3       | 0.380   | 0.328   | 0.292   | 4.249          | 0.990 |
| 7   | 6       | 4       | 4       | 0.355   | 0.355   | 0.290   | 4.263          | 0.994 |
| 7   | 5       | 5       | 4       | 0.352   | 0.324   | 0.324   | 4.291          | 1.000 |

attributes can be split into two components $a_j = (x_j, t_j)$, where $x_j$ consists of the qualitative attributes and $t_j \in \mathbb{R}$. For a choice set $\Lambda = (a_1, a_2, a_3)$ the marginal choice sets are denoted by $\xi = (x_1, x_2, x_3)$ and $t = (t_1, t_2, t_3)$, respectively, and we can split the regression functions accordingly as $F(x, t) = (F_1(x)^T, F_2(t))^T$ with marginal regression functions defined by $F_1(x) = (f(x_1), f(x_2), f(x_3))$ and $F_2(t) = (t_1, t_2, t_3)$. The utilities $U_j(a_j)$ in this two-component model are generated from the part-worth utilities in the same way as in the paired comparison situation.

As before we assume that the first component consists of $K$ qualitative attributes with two levels each. By $\bar{\xi}_{1,d}$ we denote a uniform marginal design on an orbit $d = (d_{12}, d_{13}, d_{23})$ which involves the qualitative attributes only.

We start again with the standardized case, where $\beta_1 = 0$ and $\beta_2 = 1$. There the intensity matrix $\Lambda_1 = \Lambda_1(d(t))$ depends only on the second component $t = (t_1, t_2, t_3)$ and in addition on the scaling factors $\sigma_{ij}$ and the correlations $\rho_i$, which may vary with the orbit $d$.

First we note that for a product type design $\bar{\xi}_{1,d} \otimes \xi_2$ with uniform marginal design $\bar{\xi}_{1,d}$ on the orbit $d = (d_{12}, d_{13}, d_{23})$ and arbitrary marginal design $\xi_2$ on the quantitative attribute the information matrix

$$M(\bar{\xi}_{1,d} \otimes \xi_2) = \begin{pmatrix} 4\int \lambda_{2,d}(t) \xi_2(dt)I_K & 0 \\ 0 & \int m_d(t) \xi_2(dt) \end{pmatrix}$$
is diagonal, where \( m_d(t) = t \Lambda_d(t) t^T \), \( \lambda_{2,d}(t) = \frac{1}{K} \sum_{j=1}^{3} (K - d_{ij}) \lambda_{d,ij}(t) \) is the mean intensity on the orbit \( d \) and \( \lambda_{d,ij} \) is the \( j \)th diagonal element of \( \Lambda_d \). Then the determinant of the information matrix \( M(\xi_1,d \otimes \xi_2) \) becomes

\[
\det(M(\xi_1,d \otimes \xi_2)) = (4f \lambda_{2,d}(t) \xi_2(dt))^K \int m_d(t) \xi_2(dt).
\]

Under the assumptions of Model I all part-worth utilities for the first component are assumed to be independent. Then the variances of the utility differences are again \( \sigma^2_f(A) = 2(K \sigma^2_0 + \sigma^2_t) = \sigma^2_t \). We conjecture that the determinant of the information matrix \( M(\xi_1,d \otimes \xi_2) \) will be maximized by a marginal one point design \( \xi_2 = \delta_t \) for a suitable optimal setting \( t = t^* \) of the second component. Numerically the maximization of the determinant \( \det(M(\xi_1,d \otimes \delta_2)) \) was carried out with respect to \( t \) for \( K \leq 7 \) qualitative attributes and for all possible comparison depths \( d \) with \( d_{12} \geq d_{13} \geq d_{23} \) of full profile \( (D = K) \). There we used the standardized version \( z = (z_1, z_2, z_3) \) with \( z_j = t_j - t_3 \) for the second component, as the choice probabilities are invariant with respect to a shift of location. Because of \( z_3 = 0 \) then only \( z_1 \) and \( z_2 \) have to be optimized.

In Table 5 we present the optimal values \( z^*_1 \) and \( z^*_2 \) for the quantitative attribute, the corresponding choice probabilities \( p^*_j \) and the values \( \det(M(\xi_d))^1/(K+1) \) of the criterion.

| \( K \) | \( d_{12} \) | \( d_{13} \) | \( d_{23} \) | \( z_1^* \) | \( z_2^* \) | \( p_1^* \) | \( p_2^* \) | \( p_3^* \) | \( \det(M)^{1/(K+1)} \) | \( \text{eff} \) |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 0 | 1.26 | 0.00 | 0.827 | 0.087 | 0.087 | 1.344 | 1.000 |
| 2 | 2 | 2 | 0 | 1.07 | 0.00 | 0.769 | 0.116 | 0.116 | 1.609 | 1.000 |
| 2 | 2 | 1 | 1 | 1.33 | 0.55 | 0.741 | 0.199 | 0.060 | 1.504 | 0.935 |
| 3 | 3 | 3 | 0 | 0.96 | 0.00 | 0.731 | 0.134 | 0.134 | 1.801 | 1.000 |
| 3 | 3 | 2 | 1 | 1.21 | 0.55 | 0.698 | 0.231 | 0.071 | 1.720 | 0.955 |
| 3 | 2 | 2 | 2 | 0.88 | 0.00 | 0.702 | 0.149 | 0.149 | 1.547 | 0.859 |
| 4 | 4 | 4 | 0 | 0.88 | 0.00 | 0.702 | 0.149 | 0.149 | 1.947 | 1.000 |
| 4 | 4 | 3 | 1 | 1.12 | 0.54 | 0.667 | 0.252 | 0.081 | 1.881 | 0.966 |
| 4 | 4 | 2 | 2 | 1.19 | 0.75 | 0.632 | 0.305 | 0.063 | 1.848 | 0.949 |
| 4 | 3 | 3 | 2 | 0.82 | 0.00 | 0.679 | 0.161 | 0.161 | 1.740 | 0.894 |
| 5 | 5 | 5 | 0 | 0.83 | 0.00 | 0.681 | 0.159 | 0.159 | 2.060 | 1.000 |
| 5 | 5 | 4 | 1 | 1.06 | 0.54 | 0.643 | 0.269 | 0.088 | 2.006 | 0.973 |
| 5 | 5 | 3 | 2 | 1.12 | 0.75 | 0.603 | 0.327 | 0.069 | 1.978 | 0.960 |
| 5 | 4 | 4 | 2 | 0.77 | 0.00 | 0.659 | 0.170 | 0.170 | 1.886 | 0.915 |
| 5 | 4 | 3 | 3 | 0.92 | 0.39 | 0.628 | 0.256 | 0.116 | 1.823 | 0.885 |
| 6 | 6 | 6 | 0 | 0.78 | 0.00 | 0.663 | 0.168 | 0.168 | 2.152 | 1.000 |
| 6 | 6 | 5 | 1 | 1.01 | 0.53 | 0.626 | 0.280 | 0.094 | 2.105 | 0.978 |
| 6 | 6 | 4 | 2 | 1.07 | 0.74 | 0.586 | 0.340 | 0.074 | 2.081 | 0.967 |
| 6 | 6 | 3 | 3 | 1.05 | 0.93 | 0.514 | 0.422 | 0.064 | 2.069 | 0.962 |
| 6 | 5 | 5 | 2 | 0.73 | 0.00 | 0.643 | 0.178 | 0.178 | 2.001 | 0.930 |
| 6 | 5 | 4 | 3 | 0.88 | 0.38 | 0.614 | 0.263 | 0.122 | 1.947 | 0.905 |
| 6 | 4 | 4 | 4 | 0.67 | 0.00 | 0.618 | 0.191 | 0.191 | 1.857 | 0.863 |
| 7 | 7 | 7 | 0 | 0.75 | 0.00 | 0.651 | 0.174 | 0.174 | 2.227 | 1.000 |
| 7 | 7 | 6 | 1 | 0.97 | 0.52 | 0.612 | 0.288 | 0.100 | 2.185 | 0.981 |
| 7 | 7 | 5 | 2 | 1.02 | 0.72 | 0.572 | 0.348 | 0.080 | 2.165 | 0.972 |
| 7 | 7 | 4 | 3 | 1.02 | 0.87 | 0.522 | 0.408 | 0.070 | 2.154 | 0.967 |
| 7 | 6 | 6 | 2 | 0.70 | 0.00 | 0.631 | 0.185 | 0.185 | 2.095 | 0.941 |
| 7 | 6 | 5 | 3 | 0.85 | 0.37 | 0.605 | 0.268 | 0.127 | 2.047 | 0.919 |
| 7 | 6 | 4 | 4 | 0.90 | 0.62 | 0.553 | 0.347 | 0.100 | 2.023 | 0.908 |
| 7 | 5 | 5 | 4 | 0.65 | 0.00 | 0.610 | 0.195 | 0.195 | 1.968 | 0.884 |
function together with their associated efficiencies $\text{eff}(\tilde{\xi}_d) = (\det(M(\tilde{\xi}_d)))/\det(M(\tilde{\xi}_{d^*}))^{1/(K+1)}$ for $K \leq 7$ attributes and all possible comparison depths $d$ with $d_{12} \geq d_{13} \geq d_{23}$. The optimal comparison depths $d^*$ are highlighted in bold for each $K$. In all cases the maximal value for the determinant is achieved for the design concentrated on the orbits with two identical alternatives. This coincides with the findings in the logistic case observed in Graßhoff et al. [7].

Under the assumptions of Model II for the first component of qualitative attributes the variances of the utility differences between the $i$th and $j$th alternative and the corresponding correlations are given by

$$\sigma_{ij}^2(A) = (d_{ij} + 2(K - d_{ij})\sigma_t^2)/K$$

and

$$\varphi_{ij}(A) = ((K - d_{ij}) - (K - 2d_{ij})\sigma_t^2)/(K\sigma_{ij}\sigma_{jt})$$

for any choice set $A$ of comparison depth $d$. Also here the case $\sigma_t^2 = 0$ represents a sharp decision concerning the quantitative variable $t$. The corresponding numerical results for such a sharp decision are exhibited in Table 6. There we present the optimal values for

**Table 6.** One additional quantitative attribute: optimal values for $z = (z_1, z_2, 0)$, optimal choice probabilities and design characteristics for orbits $d$ under Model II with sharp decision, $K \leq 7$.

| $K$ | $d_{12}$ | $d_{13}$ | $d_{23}$ | $z_1^*$ | $z_2^*$ | $\sigma_{t1}^2$ | $\sigma_{t2}^2$ | $\sigma_{t3}^2$ | $\det(M)^{1/(K+1)}$ | $\text{eff}$ |
|-----|---------|---------|---------|---------|---------|--------------|--------------|--------------|----------------|--------|
| 1   | 2       | 1       | 0       | 1.14    | 0.00    | 0.127 (0.873) | 0.891        | 1.000        |
| 2   | 2       | 1       | 1       | 0.72    | 0.00    | 0.142 (0.826) | 1.109        | 0.540        |
| 2   | 3       | 2       | 1       | 0.77    | 0.49    | 0.159 (0.794) | 1.272        | 0.546        |
| 3   | 2       | 2       | 1       | 0.72    | 0.72    | 0.149 (0.702) | 2.097        | 0.901        |
| 3   | 4       | 3       | 1       | 0.78    | 0.37    | 0.168 (0.767) | 1.398        | 0.551        |
| 4   | 4       | 4       | 1       | 0.77    | 0.29    | 0.178 (0.749) | 1.500        | 0.555        |
| 5   | 5       | 5       | 1       | 0.67    | 0.00    | 0.251 (0.732) | 1.583        | 0.558        |
| 6   | 6       | 6       | 1       | 0.77    | 0.23    | 0.181 (0.724) | 2.273        | 0.963        |
| 6   | 6       | 5       | 1       | 0.77    | 0.23    | 0.247 (0.571) | 2.837        | 1.000        |
| 7   | 7       | 7       | 0       | 0.58    | 0.00    | 0.281 (0.719) | 1.652        | 0.560        |
| 7   | 7       | 6       | 1       | 0.76    | 0.19    | 0.187 (0.550) | 2.948        | 1.000        |
| 7   | 7       | 5       | 2       | 0.64    | 0.32    | 0.200 (0.563) | 2.946        | 0.999        |
$z^*_1$ and $z^*_2$ for the quantitative attribute, the corresponding choice probabilities $p^*_j$ and the values $\det(M(\bar{\xi}_d))^{1/(K+1)}$ of the criterion function together with the associated efficiencies $\text{eff}(\bar{\xi}_d) = (\det(M(\bar{\xi}_d))/\det(M(\bar{\xi}_d^*))^{1/(K+1)}$ for $K \leq 7$ attributes and all possible comparison depths $d$ with $d_{12} \geq d_{13} \geq d_{23}$. The optimal comparison depths $d^*$ are highlighted in bold for each $K$. In all cases ($K \geq 2$) the maximal value for the determinant is achieved for the designs concentrated on the orbits with two alternatives which differ only in one qualitative attribute. However, if the decision is not sharp ($\sigma^2_t > 0$), we found out numerically that other comparison depths may become optimal, where the alternatives differ in more than one qualitative attribute.

Note that similar to Subsection 4.1 in the case of sharp decisions the alternatives $a_2$ and $a_3$ are indistinguishable ($U_2 = U_3$) for choice sets with comparison depth $d = (K, K, 0)$ if $z^*_2 = 0$, i.e. if the quantitative attribute is set to the same level for both alternatives. If in this case $z^*_2 > 0$ (or $z^*_2 < 0$), there will be a strict preference of alternative $a_2$ over $a_3$ (or $a_3$ over $a_2$, respectively) such that essentially we end up in a paired comparison situation for the pair $(a_1, a_2)$ (resp. $(a_1, a_3)$) of alternatives with the same value for the information matrix as specified in Table 6. This may explain why in this situation the efficiencies of choice sets with comparison depth $d = (K, K, 0)$ are so low so that the counter-intuitive result of Model I does not occur.

Although balance may be appealing, in none of the cases $K \geq 3$ reported the most balanced orbit satisfying $\max d_{ij} - \min d_{ij} \leq 1$ results in an optimal design. Instead the optimal design tends to prefer orbits in which two of the alternatives are quite similar.

### 5. Discussion

This paper provides an important extension of previous developments of optimal designs for discrete choice models (for an overview, see [5]). The designs for multinomial discrete choice models derived so far do not seem to be appropriate for many practical purposes due to the IIA property. However, merely changing the link function to a probit one does not alleviate the problem as shown above. Probit models allow for introducing dependencies between the part-worths and, thus, the overall utilities so that many choice situations can be modelled more appropriately. We developed locally $D$-optimal designs assuming choice sets for the relevant situations of either two or three options. Although the $D$-optimal designs derived for the case of indifference may be less important for many choice situations in practice, they can be used as starting points for deriving locally optimal designs for the more general case of arbitrary parameter values. In particular, the concept of canonical transformation [13] can be applied. Future research concerning design for discrete choice models based on probit regression with dependent utilities should include more than two levels for the qualitative attributes and, furthermore, should be extended to several quantitative attributes (see [6]). It would also be of interest to use other optimality criteria than $D$-optimality, such as IMSE-optimality which aims at minimizing the averaged variance of predicted response. Finally, we note that for large numbers of attributes the number of choice sets required for the obtained optimal product-type design is prohibitively large. To overcome this problem concepts of fractional factorials and, more generally, orthogonal arrays can be employed as indicated in the case of two options and three qualitative attributes.
Acknowledgements
The authors wish to express their thanks to Robert Offinger for providing the proof of Lemma A.1.

Disclosure statement
No potential conflict of interest was reported by the author(s).

Funding
Research supported by the Deutsche Forschungsgemeinschaft (DFG) under grant HO 1286/6.

References
[1] Train K. Discrete choice methods with simulation. 2nd ed. Cambridge: Cambridge University Press; 2009.
[2] Soekhai V, de Bekker-Grob EW, Ellis AR, et al. Discrete choice experiments in health economics: past, present and future. Pharmacoeconomics. 2019;37:201–226.
[3] Singh R, Das A, Chai F-S. Optimal paired choice block designs. Statist Sinica. 2019;29:1419–1438.
[4] Großmann H. On the meaning of block effects in paired comparison choice experiments and a relationship with blocked $2^K$ main effects plans. J Statist Plann Inference. 2020;209:76–84.
[5] Großmann H, Schwabe R, Design for discrete choice experiments. In: Dean A, Morris M, Stufken J, et al., editors. Handbook of design and analysis of experiments. Boca Raton: CRC Press, a Chapman & Hall book; 2015. p. 787–832.
[6] Kanninen B. Optimal design for multinomial choice experiments. J Mark Res. 2002;39:214–227.
[7] Graßhoff U, Großmann H, Holling H, et al. Optimal design for discrete choice experiments. J Statist Plann Inference. 2013;143:167–175.
[8] Kiefer J. Optimum experimental designs. J R Stat Soc Ser B. 1959;21:272–304.
[9] Silvey D. Optimal design. London: Chapman & Hall; 1980.
[10] Graßhoff U, Großmann H, Holling H, et al. Optimal design for main effects in linear paired comparison models. J Statist Plann Inference. 2004;126:361–376.
[11] Graßhoff U, Großmann H, Holling H, et al. Design optimality in multifactor generalized linear models in the presence of an unrestricted quantitative factor. J Statist Plann Inference. 2007;137:3883–3893.
[12] Ford I, Torsney B, Wu CFJ. The use of a canonical form in the construction of locally optimal designs for nonlinear problems. J R Stat Soc Ser B. 1992;54:569–583.
[13] Sitter RR, Torsney B, D-optimal designs for generalized linear models. In: Kitsos CP, Müller WG, editors. MODA4-advances in model-oriented data analysis. Heidelberg: Physica; 1995. p. 87–102.
[14] Genz A, Bretz F. Computation of multivariate normal and $t$ probabilities. Berlin: Springer; 2009.
[15] Genz A, Bretz F, Miwa T, et al. mvtnorm: Multivariate normal and $t$ distributions. R package version 1.0-6; 2017. http://CRAN.R-project.org/package=mvtnorm
[16] Gänssler P, Stute P. Wahrscheinlichkeitstheorie. Berlin: Springer; 1977.
[17] Biedermann S, Dette H, Zhu W. Optimal designs for dose-response models with restricted design spaces. J Amer Statist Assoc. 2006;101:747–759.
[18] Schwabe R. Optimum designs for multi-Factor models. New York: Springer; 1996.
Appendix. Proofs

In order to apply the constructions of Graßhoff et al. [11] to the present probit paired comparison situation of Section 3 we make use of the following auxiliary results. We start with some useful inequalities for the normal distribution.

Lemma A.1: (a) \[1 - \Phi_0(z) \geq \left(1 - \frac{z^2 + 7}{8z^2 + 12}\right) \frac{1}{z} \varphi_0(z) \] for \( z \geq 1 \),
(b) \[\Phi_0(z) - \frac{1}{2} \leq \frac{1}{\sqrt{2\pi}} (z - \frac{1}{6} z^3 + \frac{1}{40} z^5) \] for \( 0 \leq z \leq 1 \).

Proof: Assertion (a) follows along the lines for standard lower bounds of the tail probability (see e.g., [16, p. 105]):

Let \( \tau(z) = \left(1 - \frac{z^2 + 7}{8z^2 + 12}\right) \frac{1}{z} \varphi_0(z) \) denote the right-hand side in (a). Then for the derivative \( \tau' \) it holds

\[-\tau'(z) = \left(1 - \frac{2z^6 + 3z^4 + 12z^2 - 15}{z^2(4z^2 + 6)^2}\right) \varphi_0(z) \leq \varphi_0(z) \]

for all \( z \geq 1 \). Hence, for the tail probability

\[1 - \Phi_0(z) = \int_z^\infty \varphi_0(x) dx \geq \int_z^\infty -\tau'(x) dx = \tau(z)\]

as \( \tau(z) \to 0 \) for \( z \to \infty \), which proves (a).

Assertion (b) follows by Taylor expansion up to terms of order five at \( z_0 = 0 \), as the even coefficients are vanishing and the odd coefficients are alternating and decreasing.

Next we derive an auxiliary property of the function \( h(z) = 1/\lambda_2(z) \).

Lemma A.2: Let \( h(z) = \Phi_0(z)(1 - \Phi_0(z))/\varphi_0(z)^2 \). Then the third derivative \( h''' \) satisfies \( h'''(z) > 0 \) for all \( z > 0 \).

Proof: First note that

\[ h'''(z) = (8z^3 + 12z) \frac{\Phi_0(z)(1 - \Phi_0(z))}{\varphi_0(z)^2} - (14z^2 + 10) \frac{\Phi_0(z) - 1/2}{\varphi_0(z)} - 6z \]

and, hence, \( h'''(0) = 0 \) and

\[ \varphi_0(z)^2 h''(z) = 2z^3 + 3z - (8z^3 + 12z)(\Phi_0(z) - 1/2)^2 \]

\[ - (14z^2 + 10) (\Phi_0(z) - 1/2) \varphi_0(z) - 6z \varphi_0(z)^2 \].

For \( z \leq 1 \) by Lemma 1(b) we have \( (\Phi_0(z) - 1/2)^2 \leq \frac{1}{2\pi} \left(z^2 - \frac{1}{3} z^4 + \frac{7}{90} z^6\right) \). Using this, \( e^{-x} \leq 1 - x + \frac{1}{4} x^2 \) for \( 0 \leq x \leq 1 \) applied to \( x = z^2 \) and \( x = \frac{1}{2} z^2 \), respectively, and Lemma A.1(b) we obtain

\[ 2\pi \varphi_0(z)^2 h''(z) \geq 6\pi z + 4\pi z^3 \]

\[ - (8z^3 + 12z) \left( z^2 - \frac{1}{3} z^4 + \frac{7}{90} z^6 \right) \]

\[ - (14z^2 + 10) \left( z - \frac{1}{6} z^3 + \frac{1}{40} z^5 \right) \left( 1 - \frac{1}{2} z^2 + \frac{1}{8} z^4 \right) \]
By the general equivalence theorem (see [9]) the inequality
\[ \int z^2 \max_\delta \psi_0(z) \, dz \geq \int z^2 \max_\delta \psi_0(z) \, dz \]
which proves the assertion for \( z \leq 1 \).

For \( z > 1 \) we use the identity
\[
\frac{\psi_0(z)^2 \lambda''(z)}{2z^3 + 3z} = 1 - \left( 1 - 2 \left( 1 - \Phi_0(z) - \left( 1 - \frac{z^2 + 7}{8z^2 + 12} \right) \frac{1}{z} \psi_0(z) \right)^2 \right.
\]
\[ + \frac{(z^2 - 1)^2 + 24}{(4z^2 + 6)^2 z^2} \psi_0(z)^2. \]

By Lemma A.1 (a) the squared term is bounded by one for all \( z \geq 1 \), while the last expression is positive, which establishes the result. \( \blacksquare \)

**Lemma A.3:** Let \( z^* > 0 \) be the unique maximum of the function \( \lambda_2(z)^p z^2 \). Every design \( \xi^* \) which is concentrated on \( \{-z^*, z^*\} \) maximizes the criterion
\[ \Psi_2(\xi) = \int z^2 \lambda_2(z) \xi(dz) \left( \int \lambda_2(z) \xi(dz) \right)^{p-1}. \]

**Proof:** The proof is similar to the situation of logistic response considered in Graßhoff et al. [11, Lemma 1], and uses an idea of Biedermann et al. [17]:

Let \( \xi^* \) be a \( \Psi_2 \)-optimal design. Denote by \( m_j^* = \int z^j \lambda_2(z) \xi^*(dz) \), \( j = 0, 2 \), the corresponding weighted moments involved in \( \Psi_2 \). The equivalent criterion \( \ln(\Psi_2) \) is concave and its directional derivative at \( \xi^* \) in the direction of the one point design in \( z \) is
\[ \psi_2(z) = \lambda_2(z)(z^2/m_2^* + (p - 1)/m_0^*) - p. \]

By the general equivalence theorem (see [9]) the inequality \( \psi_2(z) \leq 0 \) is satisfied for all \( z \), and the maximum \( \psi_2(z) = 0 \) is attained for \( z \) in the support of \( \xi^* \). Denote further by \( h(z) = 1/\lambda_2(z) \) the inverse intensity function. The above condition can then be rewritten as
\[ g(z) = h(z) - \frac{p - 1}{m_2^*} z^2 \geq 0 \]
for all \( z \), and equality holds for \( z \) in the support of \( \xi^* \). Note that \( g \) is symmetric, \( g(z) \) tends to infinity for \( z \to \infty \), and the third derivative \( g''' = h''' \) has only one root, \( g'''(0) = 0 \), according to Lemma 2. As a consequence \( g \) may have, at most, one local minimum \( z_0 > 0 \), say. Thus, the optimal design \( \xi^* \) is concentrated on \( \{-z_0, z_0\} \) and, hence, \( \Psi_2(\xi^*) = z_0^2 \lambda_2(z_0)^p \), which is maximized by \( z_0 = z^* \). \( \blacksquare \)

For \( K = 1 \) the information matrix of the paired comparison model can be identified with that of a standard probit model with one continuous explanatory variable. In this situation, as a by-product, Lemma 3 gives an analytical proof for the corresponding result of minimal support established numerically in Biedermann et al. [17].

**Proof of Theorem 3.1:** Because only the difference \( t_1 - t_2 \) is involved in the intensity, we consider \( z = Z(t) = t_1 - t_2 \). Let further \( \delta_t \) be the one-point design in \( t \). Then the design \( \delta_t^2 \) induced by \( Z \) is the one-point design \( \delta_2 \) in \( z = Z(t) \). By Lemma 3 the design \( \delta_2 \) maximizes \( \Psi_2(\xi_2) = \int z^2 \lambda_2(z) \xi_2(dz)(\int \lambda_2 \xi_2)^K \). As has been mentioned in Subsection 3.1 the uniform design \( \xi_K \) is \( D \)-optimal in the marginal model associated with the first component \( x \). Due to the orthogonality property \( \int F_1 \xi_K = 0 \) of the uniform design \( \xi_K \) Theorem 2 in Graßhoff et al. [11] applies, which
establishes that for $t^*$ such that $Z(t^*) = z^+$ the product type design $\tilde{x}_K \otimes \delta_{t^*}$ is $D$-optimal for the probit paired comparison model with independent utilities. \[ \square \]

The following result establishes that every design is dominated by a product type design for the model considered in Section 3.2.

**Lemma A.4:** Under the assumptions of Section 3.2, for every design $\xi$ there exists a marginal design $\tilde{\xi}_2$ such that
\[
\det(M(\xi)) \leq \det(M(\tilde{x}_K \otimes \tilde{\xi}_2)).
\]

**Proof:** Let $\tilde{\xi}_1$ be the marginal design of $\xi$ on the first component and denote by $w_d$ the weight of $\tilde{\xi}_1$ on the orbit of comparison depth $d$. The corresponding symmetrized design $\tilde{\xi}$ with respect to permutations of the levels and attributes can be written as a weighted sum $\tilde{\xi} = \sum_{d=1}^K w_d \tilde{\xi}_d \otimes \xi_2$, concentrated on the orbits induced by the comparison depth $d$. Here $\xi_2$ denotes the conditional marginal distribution of $\xi$ for the second component, conditionally on the orbit of comparison depth $d$. Due to the invariance of the $D$-criterion the design $\xi$ is dominated by $\tilde{\xi}$, i.e.,
\[
\det(M(\xi)) \leq \det(M(\tilde{\xi})) \text{ (see e.g., [18, Section 3.2]).}
\]

Denote by $\sigma^2(d)$ the variance associated with comparison depth $d$. The information matrix
\[
M(\tilde{x}_d \otimes \tilde{\xi}_2) = 
\begin{pmatrix}
  c_1(d, \tilde{\xi}_2) I_K \otimes M^* & 0 \\
  0 & c_2(d, \tilde{\xi}_2)
\end{pmatrix}
\]
of a product type design $\tilde{x}_d \otimes \tilde{\xi}_2$ is block diagonal with coefficients
\[
c_1(d, \tilde{\xi}_2) = \int \lambda_2(z/\sigma(d)) \tilde{\xi}_2^Z (dz)
\]
with equality for $d = K$ and
\[
c_2(d, \tilde{\xi}_2) = \int (z/\sigma(d))^2 \lambda_2(z/\sigma(d)) \tilde{\xi}_2^Z (dz),
\]
where $Z(t) = t_1 - t_2$ and $\tilde{\xi}_2^Z$ is the image of $\xi_2$ under $Z$ as in the proof of Theorem 3.1.

Substitute $z = z/\sigma(d)$ and let $\tilde{\xi}_{2,d}$ be the image of $\xi_{2,d}$ under this transformation. Then we obtain $c_1(d, \xi_{2,d}) \leq c_1(K, \tilde{\xi}_{2,d})$ and $c_2(d, \xi_{2,d}) = c_2(K, \tilde{\xi}_{2,d})$. This implies $M(\tilde{x}_d \otimes \tilde{\xi}_{2,d}) \leq M(\tilde{\xi}_K \otimes \tilde{\xi}_{2,d})$ and, consequently,
\[
M(\tilde{\xi}) \leq \sum_{d=1}^K w_d M(\tilde{\xi}_K \otimes \tilde{\xi}_{2,d}) = M(\tilde{\xi}_K \otimes \tilde{\xi}_2),
\]
where $\tilde{\xi}_2$ is defined by $\tilde{\xi}_2 = \sum_{d=1}^K w_d \tilde{\xi}_{2,d}$. This completes the proof. \[ \square \]

**Proof of Theorem 3.2:** Let again $Z(t) = t_1 - t_2$. Since $\det(M(\tilde{x}_K \otimes \tilde{\xi}_2)) = \Psi_2(\tilde{\xi}_2)$ and $\delta_{t^*}$ maximizes $\Psi_2$, the result follows directly from Lemma A.4. \[ \square \]