Decision Field Theory
Equivalence with probit models and guidance for identifiability

Szép, Teodóra; van Cranenburgh, Sander; Chorus, Caspar G.

DOI
10.1016/j.jocm.2022.100358

Publication date
2022

Document Version
Final published version

Published in
Journal of Choice Modelling

Journal of Choice Modelling 45 (2022) 100358

Contents lists available at ScienceDirect

Journal of Choice Modelling

journal homepage: www.elsevier.com/locate/jocm

Research note

Decision Field Theory: Equivalence with probit models and guidance for identifiability

Teodóra Szép *, Sander van Cranenburgh, Caspar G. Chorus

Transport and Logistics Group, Department of Engineering Systems and Services, Delft University of Technology, Netherlands

ARTICLE INFO

Keywords:
Decision Field Theory
Probit
Identification
Distinguishability

ABSTRACT

We examine identifiability and distinguishability in Decision Field Theory (DFT) models and highlight pitfalls and how to avoid them. In the past literature, the models' parameters have been put forward as being able to capture the psychological processes in a decision maker's mind during deliberation. DFT models have been widely used to analyse human decision making behaviour, and many empirical applications in the choice modelling domain rely solely on data concerning the observed final choice. This raises the question if such data are rich enough to allow for the identification of the model's parameters. Insight into identifiability and distinguishability is crucial as it allows the researcher to determine which behavioural and psychological conclusions can or cannot be drawn from the estimated DFT model and how a DFT model can be specified in such a way that resulting parameters have meaningful interpretations. In this paper, we address this issue. To do this, we first show which specifications of DFT are equivalent to conventional probit models. Then, building on this equivalence result, we apply established analytical methods to highlight and explain the identification and distinguishability issues that arise when estimating DFT models on conventional choice data. We find evidence that some of the DFT models' special cases suffer from identifiability issues. Our results warrant caution when DFT models are used to infer psychological processes and human behaviour from conventional choice data, and they help researchers choose the correct specification of DFT models.

1. Introduction

Decision Field Theory (DFT) models are dynamic cognitive *process*[1] models that have been used in mathematical psychology for almost three decades (Busemeyer and Townsend, 1993). They have been used to analyse human response to monetary gambles (Scheibehenne et al., 2009; Hey et al., 2010), risk-taking in sports (Raab and Johnson, 2004), and consumer decisions (Noguchi and Stewart, 2014; Berkowitsch et al., 2014), for instance. Recently several contributions have been made to adapt it to the field of discrete choice analysis (Hancock et al., 2018a,b). DFT models are put forward as a more behaviourally rich alternative to conventional Discrete Choice Models (DCMs) (Busemeyer et al., 2014). The DFT model consists of three main ingredients. Firstly, weight parameters that are associated with attributes are similar to the taste parameters in conventional DCMs. Secondly, psychological parameters represent deliberation processes in the decision maker's mind. Specifically, a memory parameter captures how the previous state of preference for an alternative affects the current one, and a sensitivity parameter captures how the presence

* Corresponding author.
E-mail address: t.szep@tudelft.nl (T. Szép).

[1] Cognitive psychologists use process models to explain and describe mental processes, such as deliberation.

https://doi.org/10.1016/j.jocm.2022.100358
Received 13 September 2021; Received in revised form 10 March 2022; Accepted 6 May 2022
Available online 18 May 2022
1755-5345/

and performance of an alternative affects the decision maker's preference for another alternative. The third ingredient is the timestep parameter that stands for the number of times the decision maker updates their state of mind during deliberation.

Despite the fact that the DFT model has been around for over two decades and has been widely cited and used in an abundance of studies into choice behaviour, its inner workings and econometric properties are not yet fully understood. Since the model in several applications is put forward as being able to capture a psychological (decision-making) *process*, based only on data concerning observed final decision *outcomes*,[2] it is a crucial question whether the parameters that are representing this process are in fact identifiable. The identifiability of a model means that there are no two different sets of parameters capable of being estimated that give the same probability distribution function on any data; this notion has also been called observational equivalence (e.g. Rothenberg, 1971). In the context of choice behaviour, this implies that there are no two sets of parameters that generate the same choice probabilities for choice alternatives in the data set. It is widely accepted that in a case where a model tries to reconstruct meaningful state variables that cannot be measured directly (e.g. memory or preference), identifiability and distinguishability (a closely related concept, meaning two different structures of a model do not produce the same input–output combination) are crucial for drawing meaningful behavioural conclusions from the data (Walter and Pronzato, 1996). In existing DCM literature, conditions for identifiability have been addressed extensively and thoroughly (e.g. Bunch, 1991; Walker, 2002), highlighting that identification issues can lead to biased estimates in choice models and a loss of model fit (Walker, 2002). Therefore, for DFT models to become a viable addition to the choice modeller's toolbox, in-depth understanding on their identifiability is compulsory.

This research note investigates the identifiability and distinguishability of DFT models using analytical derivations. First we show that four DFT specifications can be recast as special cases of a probit model, one of the classic DCMs. More generally, we derive the conditions under which the theoretical equivalence of DFT and probit models holds. This enables us to build on the existing, well-developed techniques of identifiability in Discrete Choice Theory in order to obtain robust results concerning the identifiability of the DFT model. This method (i.e. establishing conditions for equivalence between a process model and a classical DCM, and subsequently using identifiability results from discrete choice theory to obtain corresponding results for the process model) has not been applied previously in existing literature for process models. We show it to be a promising avenue that may help pave the way towards incorporating alternative models from mathematical psychology to the choice modelling domain. Applying our method to DFT, we have found four cases where the probit equivalence holds: (1) when there is just one timestep, (2) when the sensitivity parameter is relatively high, (3) when the memory decay parameter is zero and (4) when there are only two alternatives. If one of these conditions is met, then the DFT model can be considered a probit model with a particular structure of its covariance matrix.[3] Using analytic derivations which capitalize on the probit equivalence, we found that the high sensitivity case is unidentifiable, and is indistinguishable from the zero memory decay case.

The remaining part of the paper is organized as follows. Section 2 introduces our methodology, starting from the parameters DFT model under examination. Then Section 2.2 briefly introduces probit models and their identifiability steps used in this paper. We present our results in Section 3. Here, we first establish the theoretical equivalence between special cases of DFT and probit models, then study the DFT model's identifiability and distinguishability issues using analytical derivations for the special cases. Section 4 discusses the implications of our results in terms of the estimation of DFT models, shows how DFT models in the literature can solve identification issues, and, more generally discusses the making of behavioural inferences based on process models which are estimated on outcome data.

2. Methodology

2.1. Decision field theory parameters

DFT was developed by Busemeyer and Townsend (1992) and over the last few decades it has had several variations applied in different fields for different problems. Its parameters are often put forward as capturing behavioural phenomena such as memory decay (ϕ_2), and competition (through the sensitivity parameter, ϕ_1) and time (through the timestep parameter, t). For instance, Hancock et al. (2018b) writes "details of chosen and unchosen alternatives are often forgotten", "$s_{i,i} < 1$ indicates that memory decays",[4] and "sensitivity parameter, ϕ_1, affects how much alternatives compete with each other". Berkowitsch et al. (2014) writes that "sensitivity parameter ϕ_1 determines the similarity as a function of the distance D between the options in the attribute space" and Busemeyer and Diederich (2002) writes "the popular option tends to get chosen during the early stages when attention is focused on the more important attribute, and the less popular option tends to get chosen during the later stages after attention has switched to the less important attribute" in their studies. Such statements, that imply these parameters capture meaningful concepts, motivate us to find out about parameter identifiability of DFT models on choice data. In this paper we examine the identifiability of these three psychologically interpretable parameters (ϕ_1, ϕ_2, t), based on the model variation that has appeared in the choice modelling domain in recent years, following the developments and notation of Hancock et al. (2018b).[5]

[2] There are exceptions that also take things like eye-tracking data into account (Noguchi and Stewart, 2014).

[3] Note that this is a technical equivalence, as the particular structure on the covariance matrix still uses parameters and notation building on DFT, and not probit.

[4] $s_{i,i}$ is a function of the memory decay parameter, ϕ_1.

[5] For the model details, see Section 2.1 of Hancock et al. (2018b).

2.2. Probit models and identifiability

It has been established that identifiability and distinguishability are crucial criteria for a parametric model in order to draw conclusions from the data (e.g. Rothenberg, 1971; Walter and Pronzato, 1996). It is especially important when the parameters are used to reconstruct meaningful concepts, such as preference, that cannot be measured directly. Probit is one of the most general DCMs (Daganzo, 1979), that often uses meaningful parameters in the covariance matrix structure. Therefore its identifiability has been extensively studied. The equivalence between DFT and probit derived in Section 3 allows us to apply such methods to investigate identifiability in DFT models.

Specifically we build on the most general method (meaning it can be used always) described by Train (2009), where the identifiability of parameters depends on whether they can be computed from the elements of the normalized covariance matrix of utility differences. The steps are the following:

(1) Take the covariance matrix of utility differences;
(2) Normalize it;
(3) Retrieve the estimable parameters from the elements of the normalized covariance matrix of utility differences.

3. Results: equivalence, identifiability and distinguishability

In this section we present our results in the following order. In Section 3.1 we present the specifications needed in a DFT model so that it is equivalent to a heteroskedastic[6] probit model. Then, in Section 3.2 we examine each of these cases by applying the analytical procedure derived for the probit model identifiability by Train (2009). Finally, in Section 3.3 we present our approach to establish distinguishability of the different cases.

3.1. Equivalence of DFT and probit models

First we show the general conditions that must hold in order to get equivalent DFT and probit models. As both of them use the integral of the multivariate normal distribution's probability density function to compute choice probabilities, they can be formulated as:

$$Pr[i] = \int_0^\infty f(X)dX \tag{1}$$

where $f(X)$ stands for the probability density function of the multivariate normal distribution. Variable X in the DFT model is

$$X^{DFT} \sim N(\widetilde{\xi}_i, \Lambda_i) \tag{2}$$

where $\widetilde{\xi}_i$ is the vector of differences between the expected preference value of alternative i and that of each of the other alternatives. Λ_i is the corresponding covariance matrix.

In probit X is

$$X^{probit} \sim N(\widetilde{V}_i, \Omega_i^*) \tag{3}$$

where $\widetilde{V}_i$ is the vector of differences between the systematic utility of alternative i and that of each of the other alternatives. Ω_i^* is the corresponding covariance matrix.

Therefore when

$$\widetilde{\xi}_i = \widetilde{V}_i \tag{4}$$

and

$$\Lambda_i = \Omega_i^* \tag{5}$$

the two models are equivalent. The condition on the covariance matrices (Eq. (5)) can be obtained with structural assumptions on the probit's covariances. Probit provides a flexible framework in which structured covariance matrices are often used to capture interdependencies between alternatives, and DFT's behavioural assumptions result in such covariances. Therefore we can impose the same structure as that of DFT on the elements of the probit covariance matrix. This structural restriction on probit results in a different covariance matrix in each choice scenario (as the covariance matrix of DFT depends on the attribute differences of alternatives), therefore a probit model that is equivalent to DFT is heteroskedastic.

As for the mean, the well-known utility difference of probit (right-hand-side of Eq. (4)) can be written in the linear-additive form of the betas multiplied by the corresponding attribute differences, or formally:

$$V_i - V_j = \sum_m \beta_m (x_{im} - x_{jm}) \qquad \forall j \neq i \tag{6}$$

[6] Note that heteroskedasticity here (contrary to many DCM applications, where the variance of the unobserved factors vary across alternatives) means that the covariance matrix varies across choice scenarios.

Table 1
The different cases of DFT examined in depth. t is the number of timesteps, ϕ_1 is the sensitivity, ϕ_2 is the memory parameter and N is the number of alternatives. α is an arbitrarily small number, D is the distance between alternatives. In Cases 1 and 3 some process parameters drop out of the model (denoted by N/A). N is an integer, all other variables have real values. Parentheses denote open sets.

	t	ϕ_1	ϕ_2	N
Case 1: One timestep	1	N/A	N/A	$N \geq 2$
Case 2: High sensitivity	$\forall t \in (1, \infty)$	$\phi_1 > -\frac{\ln \alpha}{D^2}$	$\forall \phi_2 \in (0, 1)$	$N \geq 2$
Case 3: Zero memory decay	$\forall t \in (1, \infty)$	N/A	0	$N \geq 2$
Case 4: Two alternatives	$\forall t \in (1, \infty)$	$\forall \phi_1 \in \mathbb{R}+$	$\forall \phi_2 \in (0, 1)$	2

In the following subsections we show that the left-hand-side of equation (i.e. condition) (4), in several cases (see Table 1), takes the form of a vector with elements:

$$\xi_i - \xi_j = \pi \sum_m w_m (x_{im} - x_{jm}) \qquad \forall j \neq i \tag{7}$$

π stands for a *scale term*, (that includes the psychological parameters and time) that we multiply the weighted sum of the attribute differences by. For the derivations to find which specifications can be written in this form (Eq. (7)), see Appendix A. Note that here we set the initial preference value (P_0) to zero as in several applications it is not estimated. Its effects are discussed in Section 4. As the probabilities are calculated using the multivariate normal distribution function, we can eliminate this scale term from the mean by multiplying the covariance matrix by the scale term's inverse square, π^{-2}. This means, that in the special cases that we focus on, the mean of the MVN (Eq. (2)) takes the linear additive form of attribute differences multiplied by their weights (similar to probit: Eq. (6)), and the covariance matrix of the MVN (Eq. (2)) is structured by the assumed underlying process of attention wandering, time, and psychological parameters.

Below, we introduce the four cases of DFT which correspond to probit models with linear additive utility and structured, heteroskedastic covariance matrix. These four cases are the following: (1) when there is only one timestep (i.e. updating is omitted), (2) when the sensitivity parameter is relatively high, (3) when there is a constraint on the memory decay parameter, or (4) when there are two alternatives in the choice set. If any of these four conditions apply, a DFT model is equivalent to a heteroskedastic, structured covariance probit model. Table 1 shows the four cases and indicates what they mean for the four parameters.

The following subsection introduces the scale terms for the special cases shown in Table 1. In the first case ($t = 1$, Section 3.2.1) we elaborate on the derivation steps, while in the rest of the cases we only present the final result for π for the sake of brevity. The steps we use are the same as in the first case, which are the following: we take a special case of DFT (restricted number of timesteps, sensitivity or memory parameter), take the difference in preference values ($\tilde{\xi}_i$), and bring it to the form of Eq. (7), so that the scale term can be pointed out. Furthermore, we also present the identifiability analysis in each case.

3.2. Identifiability in DFT's special cases

The identifiability steps of Train (2009) translates as follows when used in DFT models. First we need to transform our DFT model by multiplying the covariance matrix of the preference value differences by the scale term's inverse square. This way, all parameters will be exclusively in the covariance, except for the weights that appear in the preference value differences, in a similar way to betas in probit utilities. If θs are the estimable parameters, Train argues that in order to set the scale of utility, we need to normalize one of these θs by dividing all θs by it. We show that in DFT this step is unnecessary, as the normalization takes place through normalizing the weights (for proof see Appendix B). Once we have the model of interest converted into utility-differences (hereafter referred to as preference-value differences, in accordance with DFT terminology), and it is normalized (it is by definition), we can examine the elements of the covariance matrices. In three alternative cases[7] we have three θs that can be identified (one more than in a conventional probit):

$$\Lambda = \begin{bmatrix} \theta_{1,1} & \theta_{1,2} \\ \cdot & \theta_{2,2} \end{bmatrix}$$

and we examine how these θs relate to the estimable parameters, i.e. whether the psychological parameters and timesteps can be retrieved from these θs.

3.2.1. Case 1: One timestep

The first special case is when there is only one timestep. This case was referred to as being nested within the probit model by Berkowitsch et al. (2014). As we can omit the feedback matrix, DFT's preference value at timestep one (ξ_1) reduces to:

$$\xi_1 = \mu \tag{8}$$

and the corresponding covariance to:

$$\Omega_1 = \Phi \tag{9}$$

[7] We use three alternative cases for illustration purposes, extension to more alternatives is straightforward.

μ stands for the expected valence, that can be written as

$$\mu_i = \sum_m w_m \left(X_{im} - \frac{1}{N-1} \sum_{n \neq i} X_{nm} \right) \tag{10}$$

for the ith alternative. N is the number of alternatives in the choice set. The expected preference value differences between the alternatives is a vector with elements:

$$\begin{aligned}
\xi_i - \xi_j = \mu_i - \mu_j &= \sum_m w_m \left(X_{im} - \frac{1}{N-1} \sum_{n \neq i} X_{nm} \right) - \sum_m w_m \left(X_{jm} - \frac{1}{N-1} \sum_{n \neq j} X_{nm} \right) \\
&= \sum_m w_m \left((X_{im} - X_{jm}) - \frac{1}{N-1} (X_{jm} - X_{im}) \right) \\
&= \sum_m w_m \left((X_{im} - X_{jm}) + \frac{1}{N-1} (X_{im} - X_{jm}) \right)
\end{aligned}$$

$$\xi_i - \xi_j = \sum_m w_m \left(1 + \frac{1}{N-1} \right) (X_{im} - X_{jm}) \tag{11}$$

which, with

$$\pi = \left(1 + \frac{1}{N-1} \right) \tag{12}$$

takes the form of (7). Therefore, it is equivalent to a heteroskedastic, structured covariance probit model. In this case, DFT and probit differ in terms of structure, however no additional parameters (for memory, sensitivity or time) are necessary for them to be equivalent.

In order to establish identifiability in this special case, we take the vectorized form of Λ:

$$\bar{\Lambda} = \pi^{-2} \cdot (L \otimes L) \cdot [(C \times M \otimes C \times M)\bar{\Psi} + \bar{s}] \tag{13}$$

$\bar{\Psi}$ includes the weight parameters (w_1, m assumed to be 2), $\bar{s}$ includes the error term's variance (σ). π contains the number of alternatives (N, in this illustration, being 3) and none of the estimable parameters in this case.

This covariance matrix can be written in the following expanded form:

$$\bar{\Lambda} = \begin{pmatrix} \frac{8\sigma}{9} - (w_1 - 1)\, w_1 \left(X_{1,2}\right)^2 & \frac{4\sigma}{9} - (w_1 - 1)\, w_1 \left(X_{1,2}\right)\left(X_{1,3}\right) \\ . & \frac{8\sigma}{9} - (w_1 - 1)\, w_1 \left(X_{1,3}\right)^2 \end{pmatrix}$$

where

$$X_{i,j} = \left(x_{i,1} - x_{i,2} - x_{j,1} + x_{j,2}\right)$$

There is only one parameter to be identified by this covariance matrix, and that is σ which, for example, can be expressed by the first estimable parameter $\theta_{1,1}$ as:

$$\sigma = \frac{9 \cdot (\theta_{1,1} + (w_1 - 1)\, w_1 \left(X_{1,2}\right)^2)}{8} \tag{14}$$

Therefore the model is identifiable.

3.2.2. Case 2: High sensitivity

The second special case is when the feedback matrix of DFT is diagonal. In the feedback matrix parametrization developed by Hotaling et al. (2010) and used for transport data analysis in several cases (by e.g. Hancock et al. (2018b)), this means that the sensitivity parameter is relatively high (i.e. $e^{-\phi_1 D^2}$ is very close to zero). As D is the distance between alternatives in the multi-attribute space, the sensitivity parameter's size (whether or not it can be considered "high") depends on the data. However, if we establish that it is high, the following applies.

For any number of timesteps and alternatives, the scale term is:

$$\pi_t = \frac{1 - (1 - \phi_2)^t}{\phi_2} \left(1 + \frac{1}{N-1} \right) \tag{15}$$

Multiplying the covariance matrix by π_t^{-2} results in a structured covariance probit model. The covariance matrix includes parameters for time and memory. In this case, the covariance matrix takes the form of:

$$\bar{\Lambda}_t = \pi_t^{-2} \cdot (L \otimes L) \cdot [(I - S \otimes S)^{-1}(I - S^t \otimes S^t) \times \left((C \times M \otimes C \times M)\bar{\Psi} + \bar{s}\right)] \tag{16}$$

which can be expanded to:

$$\Lambda_t = \begin{pmatrix} \frac{\phi_2(((\phi_2-1)^2)^t-1)(8\sigma-9(w_1-1)w_1(X_{1,2})^2)}{9(\phi_2-2)((1-\phi_2)^t-1)^2} & \frac{\phi_2(((\phi_2-1)^2)^t-1)(4\sigma-9(w_1-1)w_1(X_{1,2})(X_{1,3}))}{9(\phi_2-2)((1-\phi_2)^t-1)^2} \\ . & \frac{\phi_2(((\phi_2-1)^2)^t-1)(8\sigma-9(w_1-1)w_1(X_{1,3})^2)}{9(\phi_2-2)((1-\phi_2)^t-1)^2} \end{pmatrix}$$

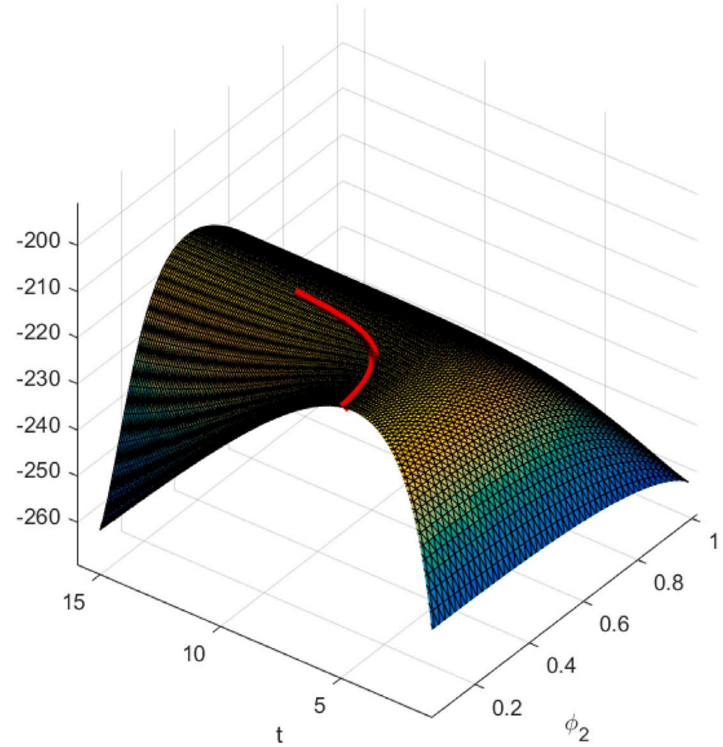

Fig. 1. The two horizontal axes show timesteps and memory parameters, and the vertical axis shows the log-likelihood generated by these combinations, all other parameters being equal. The choice data of this specific plot is based on randomly generated attributes between 0 and 1, 2 alternatives and 2 attributes. The choices are generated with DFT ($w_1 = 0.6899745$; $w_2 = 0.3100255$; $\phi_1 = 3$; $\phi_2 = 0.1$; $s = 1$; $t = 5$). The log-likelihood is calculated with the following parameters being fixed: $w_1 = 0.6899745$; $w_2 = 0.3100255$; $\phi_1 = exp(100)$; $s = 1$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

This can be written in the form of a constant term (across the sample) multiplied with a matrix that includes the error term variance, the weights and the attributes of alternatives in a structured form:

$$\bar{\Lambda}_t = \frac{\phi_2\left(1+\left(1-\phi_2\right)^t\right)}{\left(2-\phi_2\right)\left(1-\left(1-\phi_2\right)^t\right)} \begin{pmatrix} \frac{8\sigma-9(w_1-1)w_1(X_{1,2})^2}{9} & \frac{4\sigma-9(w_1-1)w_1(X_{1,2})(X_{1,3})}{9} \\ . & \frac{8\sigma-9(w_1-1)w_1(X_{1,3})^2}{9} \end{pmatrix} \tag{17}$$

The multiplicative term contains the memory parameter and the timesteps. If we name this term z, we can derive that several combinations of ϕ_2 and t can result in the same z, therefore, these two parameters are not jointly identifiable. Eq. (18) shows t as a function of ϕ_2, z is constant. As this relationship is not dependent on the data, we can state that the parameter combinations generated by this relationship will give the same input–output combination for any data.

$$t = \frac{\ln\left(\frac{z\phi_2-2z+\phi_2}{z\phi_2-2z-\phi_2}\right)}{\ln\left(1-\phi_2\right)} \tag{18}$$

To corroborate this finding, we have plotted the log-likelihood as a function of t and ϕ_2 in Fig. 1. We find that along several combinations of t and ϕ_2 (along the red line) the log-likelihood is flat, and these (t, ϕ_2) pairs satisfy relation (18).

This shape of the log-likelihood as a function of t and ϕ_2 holds for any data, as long as the estimated sensitivity in the model being examined can be considered high. For different data sets, the scales on the axes might differ, but the shape and the flatness is a characteristic that is independent of the data. Therefore, when the sensitivity is high, the memory and timestep parameters are not jointly identifiable. This issue, however, is most likely to be an empirical issue, because DFT models have not been estimated with a sensitivity parameter *fixed* to a high value, it only becomes one through estimation. In these cases the time and memory decay parameters are unidentifiable. Therefore the identifiability issue in this special case is of an empirical nature; and not a theoretical nature. An empirical example can be found in Appendix D.

3.2.3. *Case 3: Zero memory decay*

The third special case is when the feedback matrix is an identity matrix. Using the parametrization of Hotaling et al. (2010), that means that the memory parameter is zero ($\phi_2 = 0$). This also means that the feedback matrix's eigenvalues are equal to 1, therefore the geometric matrix series formula cannot be used as in equations 6a, 6b, 7a, and 7b in Hancock et al. (2018b). Instead, the formulas for ξ and Ω reduce to:

$$\xi_t = \sum_{k=0}^{t-1} S^k \mu = t \cdot \mu \tag{19}$$

$$\Omega_t = \sum_{k=0}^{t-1} [S^k \Phi S^{k'}] = t \cdot \Phi \tag{20}$$

We note that these formulas are the same as in the one timestep case, only multiplied by scalar t. The scale term therefore is:

$$\pi_t = \left(1 + \frac{1}{N-1}\right) \cdot t \tag{21}$$

The vector of the covariance matrix elements is:

$$\bar{\Lambda}_t = \pi_t^{-2} \cdot t \times \left((L \times C \times M \otimes L \times C \times M)\bar{\Psi} + (L \otimes L)\bar{s}\right) \tag{22}$$

which can be written in an extended form as:

$$\bar{\Lambda}_t = \frac{1}{t} \begin{pmatrix} \frac{8\sigma - 9(w_1-1)w_1(X_{1,2})^2}{9} & \frac{4\sigma - 9(w_1-1)w_1(X_{1,2})(X_{1,3})}{9} \\ . & \frac{8\sigma - 9(w_1-1)w_1(X_{1,3})^2}{9} \end{pmatrix}$$

From these three estimable θs, it is straightforward to express σ and t in a similar way to the one timestep case (3.2.1):

$$t = \frac{(w_1 - 1)w_1(X_{1,2}X_{1,3} - X_{1,2,}^2)}{\theta_{1,1} - \theta_{1,2}} \tag{23}$$

and

$$\sigma = \frac{9}{8}(w_1 - 1)w_1 \left(\frac{\theta_{1,1}}{\theta_{1,1} - \theta_{1,2}} X_{1,2}X_{1,3} - \frac{\theta_{1,2}}{\theta_{1,1} - \theta_{1,2}} X_{1,2}^2\right) \tag{24}$$

We conclude that the model is identifiable in this case.

3.3. Distinguishability in DFT's special cases

In order to test distinguishability we compared the covariance matrices we had previously analysed for identifiability. The one timestep case is distinguishable from all the others (i.e. the only way to get equivalent covariance matrices in high sensitivity, zero memory decay or two alternative cases, iff $t = 1$). Similarly, we find that the two alternatives case is distinguishable from the others, as the data-dependent D can only be eliminated if there is only 1 timestep, the memory parameter is zero, or when the sensitivity is relatively high. The last two special cases to compare are the high sensitivity and the zero memory decay cases. Comparing Eqs. (22) and (16) and their expanded forms we can see that both of them are essentially the same matrix, one multiplied by $\frac{1}{t}$ the other multiplied by $\frac{\phi_2(1+(1-\phi_2)^t)}{(2-\phi_2)\left(1-(1-\phi_2)^t\right)}$.

If we now denote the zero memory decay model's timestep parameter with capital T, and make the two multiplicative terms equal, we find that

$$T = \frac{(2-\phi_2)\left(1 - (1-\phi_2)^t\right)}{\phi_2\left(1 + (1-\phi_2)^t\right)} \tag{25}$$

This means, that for any (t, ϕ_2) combination in a high sensitivity model, there is an equivalent zero memory decay model with T timesteps, that generates exactly the same choice probabilities for any input data.

We can solve this for t as well:

$$t = \frac{\ln\left(\frac{2 - \phi_2 - T * \phi_2}{2 - \phi_2 + T * \phi_2}\right)}{\ln(1 - \phi_2)} \tag{26}$$

similar to Eq. (18). To sum up the two consequences of these two formulas (Eqs. (25) and (26)):

- For any combination of ϕ_2 and t in a high sensitivity model, there is a T, that will result in an equivalent zero memory decay model.
- For one zero memory decay model with timestep parameter T, there are several combinations of ϕ_2 and t that will result in equivalent models in a high sensitivity case.

Fig. 2 shows the (t, ϕ_2) combinations with the corresponding Ts that satisfy Eq. (26).

We also examine an empirical example (for details, see Appendix D), which can be summarized as follows. We take an estimated model from the literature (Hancock et al., 2018b, Table 3, model 4), that potentially fits our high sensitivity case (Section 3.2.2 and generate choice probabilities with the estimated parameters (the relevant psychological parameters being $\phi_1 = 142.6043, \phi_2 =$

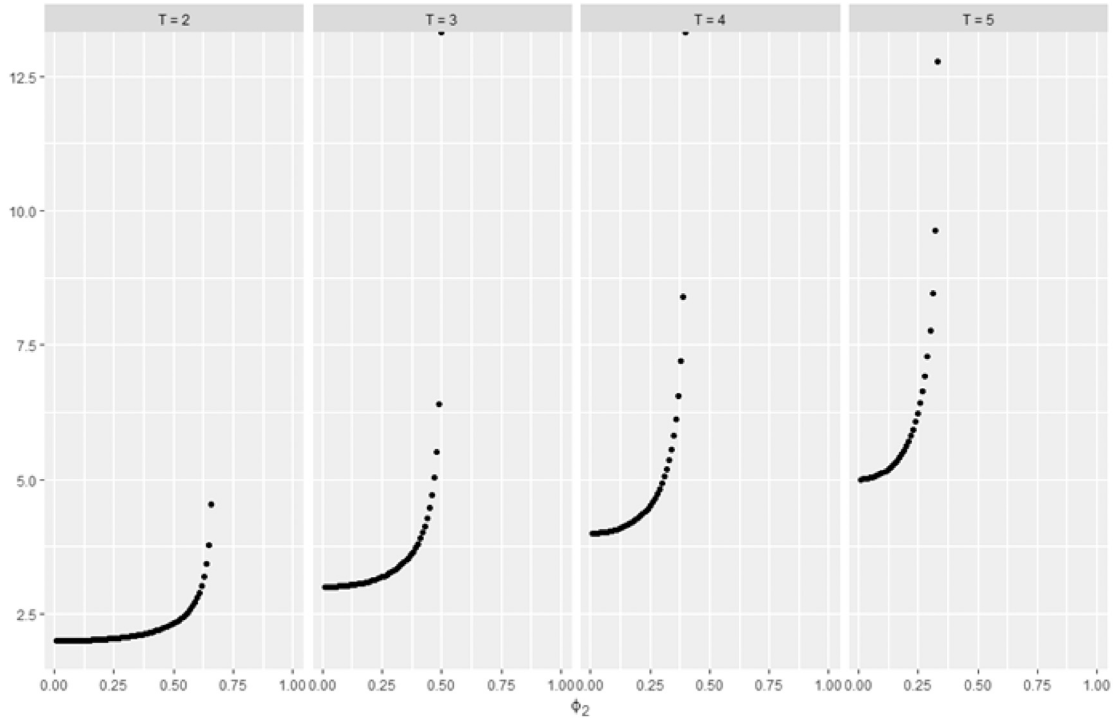

Fig. 2. In each facet we can see that for a zero memory decay model with timestep $T = 2, 3, 4, 5$, which combinations of ϕ_2 and t in a high sensitivity model give the same output. On the x-axis we have the memory (ϕ_2), on the y-axis the timestep parameter (t). The observational equivalence of these combinations is not dependent on the data we use. This corresponds to Fig. 1, where the log-likelihood function is flat along the exponential shape, which we can see here as well.

$0.1835, t = 112.2185$). Then we generate two other parameter sets: one based on Eq. (18) and one based on Eq. (25) (all the non-included parameters are kept the same), and calculate the choice probabilities based on these parameter-sets too. We confirm that for almost the whole dataset, the generated choice probabilities are the same for the three distinct parameter-sets. This means that a parameter-set that has high sensitivity, and other parameter-sets, which correspond to ''better memory'', or even ''perfect memory'' (i.e. $\phi_2 = 0$), and less timesteps during deliberation, are indistinguishable.

4. Conclusion

This research note examines identifiability and distinguishability in Decision Field Theory (DFT) models and highlights two issues. Our study is motivated by the fact that DFT models are increasingly used in the choice modelling field traditionally relying on choice (outcome) data to estimate models while at the same time aiming for more behavioural (process) realism of its models. DFT models are routinely applied as so-called process models, which presumably give insight into the process of decision-making. Such models being estimated based only on outcome data (i.e. the choices made) raises the question of whether process parameters, such as memory decay, can be inferred from the estimated model. Our paper sheds light on this question by using analytical techniques. First, we show that four types of specifications of the DFT model are equivalent to structured covariance probit models. Then we use this equivalence to study the parameter identifiability of DFT models' parameters, based on well-established methods from the field of micro-econometrics (classical discrete choice theory). Specifically, we show that when the DFT model's sensitivity parameter is very high, its memory and time parameters are not jointly identifiable. Furthermore, we establish that the high sensitivity and the zero memory decay specifications of DFT are indistinguishable.

Our main methodological contribution lies in deriving the conditions under which the DFT model is equivalent to probit models, which allows for rigorous analytical methods to examine the identifiability of parameters in several cases. We believe this approach holds potential to be applied to other process models or alternative models proposed in mathematical psychology literature, helping to incorporate such models into the toolbox of choice modellers in general and travel behaviour researchers in particular. The main conclusion of our application of this probit-equivalence based technique, is that when estimating DFT models, it is very important to choose an appropriate model set-up which guarantees a solution to identification issues, or to collect data beyond choice observations, (concerning attention-wandering and decision time, for instance) to avoid misguided behavioural inferences.

Our findings are well aligned with results and intuitions that have already been found and discussed in existing literature. Below, we discuss these earlier insights, and we specify how our results are connected to them, highlighting the contributions of this research note. First we discuss the probit-equivalence-related results, then the identifiability investigations. First Roe et al.

Table 2
Studies that connect DFT's special cases to classic Discrete Choice Models models or identification issues.

Case 1: One timestep	Roe et al. (2001) and Berkowitsch et al. (2014)
Case 2 and 4: High sensitivity and two alternatives	Hancock (2019) (Chapter 6, 2.2) (Sensitivity parameter loses its meaning due to two alternatives, conceptual reasoning)
Case 3 and 4: Zero memory decay and two alternatives	Busemeyer et al. (2006) (not only zero memory, but also two alternative-restriction in the demonstration)

(2001) described the relationship between DFT and the Thurstone model (which was later introduced to economics as the RUM model): with a feedback matrix fixed to identity, at any timestep (one moment in the deliberation process, i.e. $t = 1$) DFT reduces to the Thurstone model. Similarly, Berkowitsch et al. (2014) stated that the probit model is nested within the DFT model, when there is only one timestep. This case can be considered as an identity feedback matrix case, which is constrained by $t = 1$. As such, this specification satisfies the generic conditions we derive in Appendix A (Eqs. (A.7) and (A.8)). To contribute to this strand of literature, we formally derive which structure needs to be imposed on the covariance matrix for the two models to be equivalent. Busemeyer et al. (2006) pointed out the similarity between RUM models and DFT for a special case, where there are only two alternatives and the feedback matrix is identity. Our work extends this finding by showing that these two conditions are sufficient on their own to prove equivalence between DFT and the RUM probit model. Furthermore, we find that the equivalence condition reported in our Appendix A is satisfied by a fourth case: when the feedback matrix is diagonal but not necessarily identity. This is the case when the sensitivity parameter is relatively high and the distance between competing alternatives does not play a role. Our probit-equivalence findings derived in Appendix A are generic in the sense that they are applicable to feedback matrices which are not necessarily in the form of Eq. (8) in Hancock et al. (2018b). To study the identifiability of the parameters we applied this generic result for the above specification as it is most often used in recent literature. This results in four special cases that allow for a formal identifiability analysis based on analytical derivations. Hancock et al. (2021) also presents identification issues related to DFT, our results are complementary to their findings.

We find that if the sensitivity parameter is relatively high, the memory and timestep parameters are unidentifiable. A similar finding also appears in recent literature: Hancock (2019) (Chapter 6, 2.2) found, based on conceptual reasoning, that for two alternatives the sensitivity parameter loses its meaning, and that therefore the process will only depend on the memory and time, which then cannot be identified jointly. We show that this applies to multinomial cases as well, if the feedback matrix is diagonal (i.e. the sensitivity parameter is relatively high). This is important, because such a matrix can be an outcome of an estimation, without imposing it in advance (e.g. Hancock et al. (2018b), Table 3, Model 4). This identifiability problem is therefore primarily an empirical problem. Recent empirical investigations in current literature (Chapter 3 in Hancock (2019)) also showed that in some cases a non-restricted DFT generates similar results to a scaled multinomial logit (MNL) model, where the scale is a function of time. This is in line with our theoretical result that the zero memory decay model and the high sensitivity model are indistinguishable. In both these cases a more complicated version of a DFT model can be reduced to a simpler one with fewer parameters, where the covariance matrix is scaled. The scale can capture how deterministic the choice process is, but it is not possible to establish whether this is a result of more deliberation time or better memory. Our analytical result proves this for two special cases of DFT, while the empirical result of Hancock (2019) also extends to the domain of MNL, and indicates that a non-restrictive DFT might also exhibit the identification problem of memory and time. Table 2 summarizes these results.

In terms of guidance to avoid the mentioned identifiability problems, we propose several solutions that can be applied in the DFT framework. Based on the wide variety of DFT models proposed in existing literature, we provide an overview on the specifications that can ensure that the identification issues discussed above will not arise.

(1) A zero memory decay model where only the timesteps are estimated. This specification was used by Hancock (2019), to demonstrate that the psychological parameters do not always result in an improvement in model fit. Our results show that this specification also serves as a solution to identification issues, when sensitivity is relatively high. This specification comes with the assumption that the decision maker has perfect memory, and that previous preference states matter just as much as the current one.
(2) Assuming that the number of timesteps go to infinity. Berkowitsch et al. (2014) used this specification when estimating their model, to avoid computationally intensive simulations in DFT estimation. Our results suggest that this assumption will also eliminate identification issues when the memory and timestep parameters are not jointly identifiable. Behaviourally, this specification presumes that decision makers make a choice once their preferences have converged.
(3) Including an initial preference state (which implies the behavioural assumption that the decision maker had an initial preference value towards the alternatives in the choice set). Hancock (2019) (Chapter 6) argues the identifiability of this specification in the context of binary choices (which implies in their study that the sensitivity parameter is not playing a role). Our results show that very high sensitivity can also lead to an identifiability problem, and that including an initial preference state ensures that the DFT model cannot be written in the form of Eq. (7). As such this identifiability issue can be avoided.
(4) Scaling of the attributes. When done the right way, this eliminates the issue of relatively high sensitivity, as the relative magnitude depends on the attribute differences of alternatives. Scaling can take several different forms (for an overview see Chapter 4, 3.5 by Hancock (2019)). It has been suggested as a technique to gain better model fit and to avoid the necessity

of a priori knowledge on whether an attribute has a positive or negative effect on the preference value (Hancock, 2019). We showed that identifiability issues can occur due to the interaction between the distance between alternatives and the sensitivity parameter, and that therefore scaling the attribute levels is also a technique to eliminate identification issues. It is important to note however, that even scaled attributes can lead to high sensitivity estimates, as it is the *relative* size of distance and sensitivity that matters.

These solutions each represent different underlying processes; as such, the behavioural conclusions that would be drawn from the resulting model specifications can be very different. This is especially important to keep in mind when one has limited data (for instance when the only observation is the final choice) and when the primary goal is not to find the best model fit or prediction, but rather to actually interpret the parameters and draw behavioural inferences. In this case, our results warrant for caution when estimating the DFT model and interpreting its parameters.

These behavioural notions also urge further exploration of the identifiability of the general DFT model (i.e. without imposing any restrictions on parameters), to gain insight into what steps need to be taken to ensure that the unrestricted DFT model is identifiable. This thorny problem should, preferably, be approached using both empirical and simulated (process) data, while at the same time building on the analytical results provided in this paper. As a first step, collecting empirical data on attention-wandering and deliberation times will help develop suitable DFT-models (e.g. by testing different transformations on the parameters in the estimation or by developing alternative parameterizations for the feedback matrix). Then, to test whether the parameters of the resulting empirically supported models can be recovered without bias, the analytical steps laid out in this paper can be applied, together with analyses based on simulated data.

CRediT authorship contribution statement

Teodóra Szép: Conceptualization, Methodology, Software, Formal analysis, Writing – original draft. **Sander van Cranenburgh:** Conceptualization, Methodology, Writing – review & editing, Supervision. **Caspar G. Chorus:** Conceptualization, Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 724431)

Appendix A. Derivation of probit-like formulas

In order to find DFT specifications where conditions (4) and (5) hold, we need to examine the left-hand-side of condition (4). The aim is to get to Eq. (7), so that the model can be connected to the RUM specification of probit models. For this, we start with equation 6b in Hancock et al. (2018b) assuming $P_0 = 0$, $S^* = (I - S)^{-1}(I - S^t)$ (due to this formulation, we leave subscript t out from the following derivations), and N is the number of alternatives. These reduce preference value equation to:

$$\xi = S^* \cdot \mu \tag{A.1}$$

which in matrix form is:

$$\xi = \begin{bmatrix} S^*_{11} & S^*_{12} & \cdots & S^*_{1N} \\ S^*_{21} & S^*_{22} & & \\ \vdots & & \ddots & \\ S^*_{N1} & & & S^*_{NN} \end{bmatrix} \times \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_N \end{bmatrix}$$

The generic element of the preference differences (i.e. the difference between alternative i and j) is:

$$\xi_i - \xi_j = (S^*_{i1} - S^*_{j1})\mu_1 + \cdots + (S^*_{iN} - S^*_{jN})\mu_N \tag{A.2}$$

Substituting Eq. (10) as the generic element of vector μ,

$$\xi_i - \xi_j = (S^*_{i1} - S^*_{j1}) \sum_m w_m (x_{1m} - \frac{1}{N-1} \sum_{n \neq 1} x_{nm}) + \cdots + (S^*_{iN} - S^*_{jN}) \sum_m w_m (x_{Nm} - \frac{1}{N-1} \sum_{n \neq N} x_{nm}) \tag{A.3}$$

This can be written concisely as:

$$\xi_i - \xi_j = \sum_m w_m \sum_{k=1}^{N} \left((S^*_{ik} - S^*_{jk}) - \frac{1}{N-1} \sum_{l \neq k} (S^*_{il} - S^*_{jl}) \right) x_{km} \tag{A.4}$$

From this equation we gain two conditions that ensure probit-equivalence:

- That the coefficient (multiplicative term) of x_{im} is equal to the opposite of the coefficient of x_{jm} (i.e. attribute m of alternative j, when we examine the difference between alternative i and j) for all m
- The coefficient of x_{km} when $k \neq i, j$ is zero for all m.

The first condition can be formalized as:

$$(S^*_{ii} - S^*_{ji}) - \frac{1}{N-1}\sum_{l \neq i}(S^*_{il} - S^*_{jl}) = -\left((S^*_{ij} - S^*_{jj}) - \frac{1}{N-1}\sum_{l \neq j}(S^*_{il} - S^*_{jl})\right) \tag{A.5}$$

which can be simplified to:

$$\left(1 - \frac{1}{N-1}\right)(S^*_{ii} - S^*_{ji} + S^*_{ij} - S^*_{jj}) - \left(\frac{2}{N-1}\right)\sum_{l \neq i \neq j}(S^*_{il} - S^*_{jl}) = 0 \qquad \forall i, j, l \tag{A.6}$$

If we assume a symmetric matrix with the same elements on the diagonal (which is the case in all DFT applications so far), then the above condition reduces to:

$$\sum_{l \neq i \neq j}(S^*_{il} - S^*_{jl}) = 0 \qquad \forall i, j, l \tag{A.7}$$

meaning the off-diagonal elements of the matrix must be equal.

The second condition, written as a formula is:

$$\left((S^*_{ik} - S^*_{jk}) - \frac{1}{N-1}\sum_{l \neq k}(S^*_{il} - S^*_{jl})\right) = 0 \qquad \forall k \neq i, j \tag{A.8}$$

Which is satisfied once all the off-diagonal elements are equal. In the parametrization we use in this paper, this means either there must be only two alternatives, or the feedback matrix must be diagonal. The reason for this is that the off-diagonal elements are dependent on the attribute distances (therefore the data). The only way they all become equal for any data in a multi-alternative case, is that the sensitivity parameter is very high (relative to the data), or the memory decay is zero.

Appendix B. Normalization

In probit models, we need to normalize the covariance matrix in order to have an identifiable model. In general the problem is that the multivariate normal probability distribution gives the same probability for the left and right hand side of the following equation:

$$\Phi(0|V, \Sigma) = \Phi(0|kV, k^2\Sigma) \tag{B.1}$$

where V is the systematic utility difference between the two alternatives and Σ is the corresponding covariance matrix. In the following we use the 2 alternative, 2 attribute example for illustration. Let us call the following two equations 'model 1':

$$V = \beta_1 X_1 + \beta_2 X_2 \tag{B.2}$$

$$\Sigma = \Omega \tag{B.3}$$

where βs are the taste parameters and Xs are the attribute differences between the two alternatives. Ω is the estimated covariance without further specification.

Following on from Eq. (B.1), we can specify 'model 2' to be equivalent with 'model 1':

$$V' = \beta'_1 X_1 + \beta'_2 X_2 = k(\beta_1 X_1 + \beta_2 X_2) \tag{B.4}$$

$$\Sigma' = \Omega' = k^2\Omega \tag{B.5}$$

This shows, that if the estimable parameters of 'model 2' take the values listed below, the resulting distribution function is equivalent to that of 'model 1'.

$$\beta'_1 = k\beta_1$$

$$\beta'_2 = k\beta_2$$

$$\Omega' = k^2\Omega$$

Without further restrictions, this implies that there are an infinite number of $(\beta_1, \beta_2, \Omega)$ combinations that give the exact same choice probabilities, therefore the model is unidentifiable. The probit model handles it by fixing $\Omega = 1$, then all βs are uniquely defined.

In the following, we show that using the DFT restrictions (two kinds of specifications in particular) on weight parameters will ensure that there is only one k that gives the same choice probabilities, and that is $k = 1$. Therefore the estimable parameters are uniquely defined.

The choice probability in DFT can also be expressed as in Eq. (B.1). In DFT, the taste parameters are named ws, and the estimable parameters as βs. In DFT these two are not (necessarily) the same.

$$V = w_1 X_1 + w_2 X_2 \tag{B.6}$$

$$\Sigma = \Omega \tag{B.7}$$

where

$$w_1 = \frac{exp(\beta_1)}{exp(\beta_1) + exp(\beta_2)} \tag{B.8}$$

and

$$w_2 = \frac{exp(\beta_2)}{exp(\beta_1) + exp(\beta_2)} \tag{B.9}$$

As an infinite number of (β_1, β_2) combinations would give us the same (w_1, w_2), β_1 is normalized to 0.

Therefore, 'DFT model 1' is:

$$V = \frac{1}{1 + exp(\beta_2)} X_1 + \frac{exp(\beta_2)}{1 + exp(\beta_2)} X_2 \tag{B.10}$$

$$\Sigma = \Omega \tag{B.11}$$

and 'DFT model 2' (a model that is equivalent to 'DFT model 1') is:

$$V' = \frac{1}{1 + exp(\beta_2')} X_1 + \frac{exp(\beta_2')}{1 + exp(\beta_2')} X_2 = k(\frac{1}{1 + exp(\beta_2)} X_1 + \frac{exp(\beta_2)}{1 + exp(\beta_2)} X_2) \tag{B.12}$$

$$\Sigma = \Omega' = k^2 \Omega \tag{B.13}$$

In order for this to hold for any X:

$$\frac{1}{1 + exp(\beta_2')} = \frac{k}{1 + exp(\beta_2)} \tag{B.14}$$

from which k can be expressed as:

$$k = \frac{1 + exp(\beta_2)}{1 + exp(\beta_2')} \tag{B.15}$$

from Eq. (B.12) it also follows that:

$$\frac{exp(\beta_2')}{1 + exp(\beta_2')} = k \frac{exp(\beta_2)}{1 + exp(\beta_2)} \tag{B.16}$$

which can be rewritten as:

$$k = \frac{exp(\beta_2')}{exp(\beta_2)} \cdot \frac{1 + exp(\beta_2)}{1 + exp(\beta_2')} \tag{B.17}$$

Substituting Eq. (B.15) gives:

$$1 = \frac{exp(\beta_2')}{exp(\beta_2)} \tag{B.18}$$

$$\beta_2' = \beta_2 \tag{B.19}$$

which also means:

$$k = 1$$

When the weights are the estimated parameters so that $(w_1, w_2 = 1 - w_1)$ (in this case only w_1 is estimated), the following should hold:

$$w_1' = k w_1$$

$$w_2' = k w_2 = k(1 - w_1)$$

also, the condition that $w_1' + w_2' = 1$ must hold, therefore:

$$w_2' = 1 - w_1' = 1 - k w_1$$

merging the latter two equations we find:

$$k(1-w_1) = 1 - kw_1$$

$$k = 1$$

Appendix C. Identification in case 4: Two alternatives

The fourth special case of DFT-probit equivalence is when there are only two alternatives. In this case the scale term is dependent on all psychological parameters and the timesteps as well.

$$\pi_t = \frac{1-\left(1-\phi_2(1-e^{-\phi_1 D^2})\right)^t}{\phi_2(1-e^{-\phi_1 D^2})}\cdot 2 \tag{C.1}$$

The scale term is dependent on the number of alternatives, the distance between the attributes of alternatives (D), sensitivity (ϕ_1), memory (ϕ_2) and also timesteps (t). The covariance matrix consists of a single θ, which is the following:

$$\bar{\Lambda}_i = \theta_i = \frac{\phi_2\left(e^{D_i\phi_1}-1\right)\left(\left(e^{-2D_i\phi_1}\left(\phi_2-\left(\phi_2-1\right)e^{D_i\phi_1}\right)^2\right)^t-1\right)\left(\sigma-2\left(w_1-1\right)w_1\left(X_{1,2,i}\right)^2\right)}{2\left(\left(\phi_2-2\right)e^{D_i\phi_1}-\phi_2\right)\left(\left(\phi_2\left(e^{-D_i\phi_1}-1\right)+1\right)^t-1\right)^2} \tag{C.2}$$

for $i \in 1,\dots,n$, where n is the size of the data. Although the above equation contains four unknown variables (ϕ_1,ϕ_2,t,σ), due to the heteroskedasticity we cannot eliminate the parameter identifiability based on the number of equations. Following up on our previous results (high sensitivity case, where t can be expressed as a function of ϕ_2), we solve the equation for t. In order to express t as a function of the other parameters, first let us reformulate the following term from the numerator:

$$\left(\left(e^{-2D_i\phi_1}\left(\phi_2-\left(\phi_2-1\right)e^{D_i\phi_1}\right)^2\right)^t-1\right) = \left(\phi_2 e^{-D_i\phi_1}-(\phi_2-1)\right)^{2t}-1 \tag{C.3}$$

and the following term from the denominator:

$$\left(\left(\phi_2\left(e^{-D_i\phi_1}-1\right)+1\right)^t-1\right)^2 = \left(\left(\phi_2 e^{-D_i\phi_1}-(\phi_2-1)\right)^t-1\right)^2 \tag{C.4}$$

Let A be defined as:

$$A = \phi_2 e^{-D_i\phi_1}-(\phi_2-1) \tag{C.5}$$

By substituting A and using the well-known identity, $a^2-b^2=(a+b)(a-b)$, we can reformulate Eq. (C.3) as:

$$A^{2t}-1 = (A^t+1)(A^t-1) \tag{C.6}$$

and Eq. (C.4) as:

$$(A^t-1)^2 \tag{C.7}$$

Substituting the above two formulas, and simplifying the numerator and denominator by (A^t-1), Eq. (C.2) becomes:

$$\theta_i = \frac{\phi_2\left(e^{D_i\phi_1}-1\right)\left(\sigma-2\left(w_1-1\right)w_1\left(X_{1,2,i}\right)^2\right)(A^t+1)}{2\left(\left(\phi_2-2\right)e^{D_i\phi_1}-\phi_2\right)(A^t-1)} \tag{C.8}$$

Rearranging this to express A^t we find:

$$A^t = \frac{\phi_2\left(e^{D_i\phi_1}-1\right)\left(\sigma-2\left(w_1-1\right)w_1\left(X_{1,2,i}\right)^2\right)+\theta_i\cdot 2\cdot\left(\left(\phi_2-2\right)e^{D_i\phi_1}-\phi_2\right)}{\theta_i\cdot 2\cdot\left(\left(\phi_2-2\right)e^{D_i\phi_1}-\phi_2\right)-\phi_2\left(e^{D_i\phi_1}-1\right)\left(\sigma-2\left(w_1-1\right)w_1\left(X_{1,2,i}\right)^2\right)} \tag{C.9}$$

from which it follows that:

$$t = \log_A\left(\frac{\phi_2\left(e^{D_i\phi_1}-1\right)\left(\sigma-2\left(w_1-1\right)w_1\left(X_{1,2,i}\right)^2\right)+\theta_i\cdot 2\cdot\left(\left(\phi_2-2\right)e^{D_i\phi_1}-\phi_2\right)}{\theta_i\cdot 2\cdot\left(\left(\phi_2-2\right)e^{D_i\phi_1}-\phi_2\right)-\phi_2\left(e^{D_i\phi_1}-1\right)\left(\sigma-2\left(w_1-1\right)w_1\left(X_{1,2,i}\right)^2\right)}\right) \tag{C.10}$$

We find that t is dependent on D_i, which is not constant across the data. Therefore, with sufficient amount of data, there is no evidence for a theoretical identification problem of time and memory in this case.

Appendix D. Empirical example

The high sensitivity case is not a very common issue in published literature. In this section we investigate an estimated model, where the estimated sensitivity parameter is unusually high (i.e. above hundred). In Hancock et al. (2018b) Table 3, model 4, we see that $\phi_1 = 142.6043$, and the two other psychological parameters are $\phi_2 = 0.1835, t = 112.2185$. After standard score normalization of the data,[8] we generate DFT choice probabilities based on the above parameters. Then, to establish whether there is an identification-issue, we test whether other parameter-sets result in the same choice probabilities. To find such parameter-sets, we use Eqs. (18) and (25).

First, we examine a high sensitivity case with the memory decay parameter set approximately to half its size (i.e. people have "better memory" when deliberating). We set $\phi_2' = 0.09$, and apply Eq. (26) to get $t' = 10.72$. We generate DFT choice probabilities with this new parameter-set ($\phi_2' = 0.09$ and $t' = 10.72$, everything else kept the same as before), and examine the difference compared to the choice probabilities generated by the original parameter-set.

We see that the mean difference in choice probabilities (generated by the original and the newly calculated parameter-set) is 0.00162883 (less than 1 percentage point). Although the highest difference we see in the data is 12 percentage point, out of the total of 3492 data points, in 3380 instances the difference is less than 1 percentage point. This means basically, that for the most part of the data, the original parameter-set and the one corresponding to "better memory" and "less timesteps spent on deliberation" generate the same choice probabilities. Examining the 112 data points where the choice probabilities have larger difference (than 1 percentage point), we find that the average Eucledian distance between the alternatives is 25 times smaller compared to the lower-difference (than 1 percentage point) part of the data. This is because if the distance (D) is very small (i.e. close to 0), then $\exp(-\phi_1 * D^2) \approx 1$. This illustrates why we define high sensitivity as *relatively high sensitivity* in Section 3.2.2; the smaller the distance between alternatives, the higher sensitivity parameter is 'needed' for the model to become unidentifiable.

Next, we test the distinguishability from the zero-memory decay model, applying Eq. (25), and generating DFT choice probabilities with $\phi_2'' = 0$ and $t'' = 9.899183$. The mean choice probability difference from the original estimated model is 0.001776593 (less than 1 percentage point), and out of 3492 data points, in 3373 the difference is less than 1 percentage point.

We can conclude, that although it is not a perfectly-high-sensitivity case, for a substantial part of the data (96.6% of the total sample), the generated choice probabilities are almost the same, whether we assume there is no memory decay at all and the decision is made quickly, or when there is some memory decay and more time is spent on deliberation.

Appendix E. Relation to order and rank conditions

Rank and order conditions (Bunch, 1991; Walker et al., 2007) were also suggested in the literature as ways to find identification issues in probit models.

The order condition states that in a probit model's covariance matrix there are maximum $\frac{J(J-1)}{2} - 1$ identifiable parameters, J being the number of alternatives. In the probit-equivalent DFT specifications this is not applicable as the covariance matrix is structured: it varies across the data.

The rank condition states that in a probit model's covariance matrix there are maximum $Rank(Jacobian(vecu(\Omega_\Delta))) - 1$ identifiable parameters (Walker et al., 2007). In the probit-equivalent DFT specifications this condition should be modified because the scale is already set by the weight normalization (Appendix B); thus 1 should not be subtracted at the end. The rank condition can be used to confirm our findings. For instance, the rank of the jacobian of the vectorized covariance matrix in Eq. (17) is 2, while 3 parameters should be estimated. This is a direct result of the multiplicative term involving two parameters in Eq. (17). As the rank condition takes into account the structure of the covariance matrix (and not just the number of alternatives), it is also suitable for identification analysis in DFT.

References

Axhausen, Kay W, Hess, Stephane, König, Arnd, Abay, Georg, Bates, John J, Bierlaire, Michel, 2008. Income and distance elasticities of values of travel time savings: New swiss results. Transp. Policy 15 (3), 173–185.

Berkowitsch, Nicolas AJ, Scheibehenne, Benjamin, Rieskamp, Jörg, 2014. Rigorously testing multialternative decision field theory against random utility models. J. Exp. Psychol. [Gen.] 143 (3), 1331.

Bunch, David S., 1991. Estimability in the multinomial probit model. Transp. Res. B 25 (1), 1–12.

Busemeyer, Jerome R., Diederich, Adele, 2002. Survey of decision field theory. Math. Social Sci. 43 (3), 345–370.

Busemeyer, Jerome R, Jessup, Ryan K, Johnson, Joseph G, Townsend, James T, 2006. Building bridges between neural models and complex decision making behaviour. Neural Netw. 19 (8), 1047–1058.

Busemeyer, Jerome R., Rieskamp, Jörg, et al., 2014. Psychological research and theories on preferential choice. In: Handbook of Choice Modeling. Edward Elgar Publication, Cheltenham, pp. 49–72.

Busemeyer, Jerome R., Townsend, James T., 1992. Fundamental derivations from decision field theory. Math. Social Sci. 23 (3), 255–282.

Busemeyer, Jerome R., Townsend, James T., 1993. Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment.. Psychol. Rev. 100 (3), 432.

Daganzo, Carlos, 1979. Multinomial Probit: The Theory and its Application to Demand Forecasting. Elsevier.

Hancock, Thomas O., 2019. Travel Behaviour Modelling at the Interface Between Econometrics and Mathematical Psychology (Ph.D. thesis). University of Leeds.

Hancock, Thomas O., Hess, Stephane, Choudhury, Charisma, 2018a. Incorporating response time in a decision field theory model. In: The Transportation Research Board (TRB) 97th Annual Meeting.

Hancock, Thomas O., Hess, Stephane, Choudhury, Charisma F., 2018b. Decision field theory: Improvements to current methodology and comparisons with standard choice modelling techniques. Transp. Res. B 107, 18–40.

Hancock, Thomas O, Hess, Stephane, Marley, Anthony AJ, Choudhury, Charisma F, 2021. An accumulation of preference: two alternative dynamic models for understanding transport choices. Transp. Res. B 149, 250–282.

[8] Swiss route choice data, Axhausen et al. (2008).

Hey, John D., Lotito, Gianna, Maffioletti, Anna, 2010. The descriptive and predictive adequacy of theories of decision making under uncertainty/ambiguity. J. Risk Uncertain. 41 (2), 81–111.

Hotaling, Jared M., Busemeyer, Jerome R., Li, Jiyun, 2010. Theoretical Developments in Decision Field Theory: Comment on Tsetsos, Usher, and Chater (2010).. American Psychological Association.

Noguchi, Takao, Stewart, Neil, 2014. In the attraction, compromise, and similarity effects, alternatives are repeatedly compared in pairs on single dimensions. Cognition 132 (1), 44–56.

Raab, Markus, Johnson, Joseph G., 2004. Individual differences of action orientation for risk taking in sports. Res. Q. Exerc. Sport 75 (3), 326–336.

Roe, Robert M., Busemeyer, Jermone R., Townsend, James T., 2001. Multialternative decision field theory: A dynamic connectionst model of decision making.. Psychol. Rev. 108 (2), 370.

Rothenberg, Thomas J., 1971. Identification in parametric models. Econometrica 39 (3), 577–591.

Scheibehenne, Benjamin, Rieskamp, Jörg, González-Vallejo, Claudia, 2009. Cognitive models of choice: Comparing decision field theory to the proportional difference model. Cogn. Sci. 33 (5), 911–939.

Train, Kenneth E., 2009. Discrete Choice Methods with Simulation. Cambridge University Press.

Walker, Joan, 2002. Mixed logit (or logit kernel) model: Dispelling misconceptions of identification. Transp. Res. Rec. 1805 (1), 86–98.

Walker, Joan L., Ben-Akiva, Moshe, Bolduc, Denis, 2007. Identification of parameters in normal error component logit-mixture (NECLM) models. J. Appl. Econometrics 22 (6), 1095–1125.

Walter, Eric, Pronzato, Luc, 1996. On the identifiability and distinguishability of nonlinear parametric models. Math. Comput. Simulation 42 (2–3), 125–134.