DECISION FIELD THEORY: IMPROVEMENTS TO CURRENT METHODOLOGY AND COMPARISONS WITH STANDARD CHOICE MODELLING TECHNIQUES

Thomas O. Hancock (Corresponding Author)
Institute for Transport Studies & Choice Modelling Centre
University of Leeds
34-40 University Road
Leeds
LS2 9JT
United Kingdom
tstoh@leeds.ac.uk

Stephane Hess
Institute for Transport Studies & Choice Modelling Centre
University of Leeds
S.Hess@its.leeds.ac.uk

Charisma F. Choudhury
Institute for Transport Studies & Choice Modelling Centre
University of Leeds
C.F.Choudhury@leeds.ac.uk

ABSTRACT

There is a growing interest in the travel behaviour modelling community in using alternative methods to capture the behavioural mechanisms that drive our transport choices. The traditional method has been Random Utility Maximisation (RUM) and recent interest has focussed on Random Regret Minimisation (RRM), but there are many other possibilities. Decision Field Theory (DFT), a dynamic model popular in mathematical psychology, has recently been put forward as a rival to RUM but has not yet been investigated in detail or compared against other competing models like RRM. This paper considers arguments in favour of using DFT, reviews how it has been used in transport literature so far and provides theoretical improvements to further the mechanisms behind DFT to better represent general decision making. In particular, we demonstrate how the probability of alternatives can be calculated after any number of timesteps in a DFT model. We then look at how to best operationalise DFT using simulated datasets, finding that it can cope with underlying preferences towards alternatives, can include socio-demographic variables and that it performs best when standard score normalisation is applied to the alternative attribute levels. We also present a detailed comparison of DFT and Multinomial Logit (MNL) models using stated preference route choice datasets and find that DFT achieves significantly better fit in estimation as well as forecasting. We also find that our theoretical improvement provides DFT with much greater flexibility and that there are numerous approaches that can be adopted to incorporate heterogeneity within a DFT model. In particular, random parameters vastly improve the model fit.

KEYWORDS

Decision Field Theory; Choice Modelling; Route Choice

1 AN INTRODUCTION TO DECISION FIELD THEORY

Random Utility Maximisation (RUM) models have dominated the field of choice modelling for over 40 years [McFadden, 2000], particularly in travel behaviour research [Ben-Akiva and Bierlaire, 1999]. Recently, however, there has been increasing interest in using alternative methods to make the models flexible to accommodate departures from behaviours assumed under RUM. A key example in transport research has been Random Regret Minimisation [Chorus et al., 2008, Chorus, 2010], which assumes that decision-makers seek to minimise negative emotions rather than maximising positive ones. Another example comes in the form of Bayesian Belief Networks [Parvaneh et al., 2012], which take a more heuristic approach, looking at an individual's past experiences and expectations about the different alternatives available.

Whilst these new methods both make more of an effort to consider the underlying cognitive processes in decision making, another model, Decision Field Theory [Busemeyer and Townsend, 1992, 1993], was designed purely as a cognitive model to capture the deliberation process in decision making. Decision Field Theory (DFT) is a stochastic-dynamic model of decision-making behaviour, which was expanded to include multi-attribute [Diederich, 1997] and then multi-alternative decision-making [Roe et al., 2001], where it was renamed multi-alternative decision field theory (MDFT)[1].

Due to the psychological roots of DFT [Busemeyer and Diederich, 2002], it has predominantly been used to explain behaviour not typically studied using "traditional" choice models. DFT can theoretically explain similarity, attraction and compromise effects [Roe et al., 2001] and this has largely been the focus of DFT research with many papers looking into how well it can explain these context effects compared to other models [Tsetsos et al., 2010, Trueblood et al., 2013, Noguchi and Stewart, 2014]. It is of course true that RUM models can also be used to test such effects, with notably Nested Logit being used to study the similarity effect [Guevara and Fukushi, 2016] or preference reversals [Batley and Hess, 2016]. However, Decision Field Theory further differentiates from these models by being a dynamic model. This means that it can successfully be used to study risky choices or the effect of time pressure [Busemeyer and Townsend, 1993, Diederich, 1997, Dror et al., 1999]. Despite the success of DFT in explaining time and context effects, it has not often been used to explain riskless choices or decision making in general.

We address this research gap in this paper by providing theoretical improvements to further the mechanisms behind DFT to better represent general decision making, incorporating potential effects of socio-demographic variables and accommodating for heterogeneity. The models are rigorously compared against RUM and RRM, both for estimation and prediction, using simulated and real datasets.

The remainder of this paper is organised as follows. The next section provides a comprehensive review of DFT: how it works, comparisons with other models and arguments in favour of using DFT. Section 3 gives our theoretical improvements for DFT. Section 4 presents the data and looks at our results from using DFT and section 5 presents some conclusions.

[1] Some authors refer to decision field theory as DFT, others use MDFT. We shall henceforth use DFT

2 OVERVIEW OF DECISION FIELD THEORY

Thus far, Berkowitsch et al. [2014] have provided the only comparison of DFT against mainstream choice models. As far as we are aware, DFT has never been compared to RRM or other alternative models from choice modelling, nor have the predictive capabilities of DFT been tested. We do not yet know if specific types of choices will be better explained by DFT or if certain decision-makers may be better represented by a DFT model.

In the following subsection, a summary is provided for the basic mechanisms of DFT. We then consider arguments in support of DFT and look further into how it has been used so far in transport research. We conclude by looking at how DFT has been compared to RUM thus far.

2.1 Mechanisms of Decision Field Theory

Basic mechanism

The main idea behind Decision Field Theory is that each available alternative has a 'preference value', which updates over time. At each step, the current values are multiplied by a 'feedback matrix' before then adding on a valence vector (which can be considered as a utility at a specific moment) at that time. In its most basic form, we have:

$$P_t = S \cdot P_{t-1} + V_t, \tag{1}$$

where P_t is a column matrix containing the current preference values for each alternative at time t, and S is a feedback matrix which contains three parameters (see section 2.1). P_{t-1} is the previous preference vector and P_0 is the initial preference vector. This is often assumed to be $[0,..,0]'$ [Busemeyer and Diederich, 2002]. Finally, V_t is the random valence vector at time t, given by:

$$V_t = C \cdot M \cdot W_t + \varepsilon_t, \tag{2}$$

where C is a contrast matrix, used to compare alternatives against each other, with $c_{i,i} = 1$ and $c_{i,j\neq i} = -1/(n-1)$, where n is the number of alternatives, and M is the attribute matrix. DFT is scale-variant [Busemeyer and Diederich, 2002] and we explore the implications of failing to ensure that the attribute matrix has been appropriately scaled in section 4.3. At each time, t, one attribute is attended to, such that $W_t = [0..1..0]'$ with entry $j = 1$ if and only if attribute j is the attribute currently being attended to. The probability of attending to attribute j is w_j. Since these weights must sum to one, a standard uniform distribution $X \sim U(0,1)$ can be used to select which attribute a decision-maker attends to at each timestep. It is assumed that there is no relationship between the timesteps, which means an attribute could be considered for several consecutive timesteps before the decision-maker considers a different attribute. There is also a random error vector, $\varepsilon_t = [\varepsilon..\varepsilon]'$, with $\varepsilon \sim N(0,s)$ added on to allow for flexibility in the variation of probability values that DFT predicts. The variance for the error, s, is often fixed to 1 [Trueblood et al., 2014] but can also be an estimated parameter.

Calculating expected values

Expanding equation 1 results in:

$$P_1 = S \cdot P_0 + V_1 \tag{3a}$$

$$P_2 = S \cdot (S \cdot P_0 + V_1) + V_2 = S^2 \cdot P_0 + S \cdot V_1 + V_2 \tag{3b}$$

$$\dots \tag{3c}$$

$$P_t = \sum_{k=0}^{t-1} S^k \cdot V_{t-k} + S^t \cdot P_0 \tag{4}$$

The weight vectors w_j are stationary, therefore W_t can be considered a stationary stochastic process. This means that V_t is also a stationary stochastic process with mean $E[V_t]$ and a covariance matrix given by $Cov[V_t]$. Given X, a random vector, and A, a matrix of constants, we have:

$$var(A \cdot X) = A \cdot var(X) \cdot A' \tag{5}$$

We can now use this to calculate the expected valence. We have $E[V_t] = \mu = C \cdot M \cdot w_m$, where $w_m = [w_1, w_2, ..., w_a]'$, and $Cov[V_t] = \Phi = C \cdot M \cdot \Psi \cdot M' \cdot C' + s$, where $\Psi = Cov[W_t]$ and $s = Cov[\varepsilon_t]$. We can then calculate the expected value and covariance of P_t. With S being a constant, $E[P_t]$ reduces to:

$$E[P_t] = \xi_t = \sum_{k=0}^{t-1} S^k \cdot \mu + S^t \cdot P_0 \tag{6a}$$

$$= (I - S)^{-1}(I - S^t) \cdot \mu + S^t \cdot P_0 \tag{6b}$$

Equation 5 also means that we now have:

$$Cov[P_t] = \Omega_t = Cov\left[\sum_{k=0}^{t-1} S^k \cdot V_{t-k} + S^t \cdot P_0\right] \tag{7a}$$

$$= \sum_{k=0}^{t-1} \left[S^k \cdot \Phi \cdot S^{k'}\right] \tag{7b}$$

The feedback matrix

The feedback matrix is fundamental to the performance of DFT and is defined as:

$$S = I - \phi_2 \times \exp(-\phi_1 \times D^2) \tag{8}$$

Where I is an identity matrix, ϕ_1 and ϕ_2 are sensitivity and memory parameters respectively, and D is some measure of distance between the attributes across alternatives. The sensitivity parameter, ϕ_1, affects how much alternatives compete with each other. This allows for the similarity effect to occur [Roe et al., 2001]. The memory parameter, ϕ_2, affects the diagonal entries of the feedback matrix S. The importance of having this parameter is demonstrated by the fact that details of chosen and unchosen alternatives are often forgotten [Mather et al., 2000]. A value of $s_{i,i} < 1$ indicates that memory decays, whereas $s_{i,i} > 1$ indicates that memory grows. Individuals have different working memory capacities [Daneman and Carpenter, 1980] and memories can grow as well as fade [Mather, 2006], an idea that appears in studies on the validity of eyewitness testimony [Flin et al., 1992, Christianson, 1992, Zaragoza and Lane, 1994]. A number of different methods have been used for defining the distance, D, between alternatives in applications of DFT. Roe et al. [2001] have suggested that 'psychological' distances should be used but in application chose distances that took into account the relative position of the alternatives in the multi-attribute evaluation space. The Euclidean distance (the straight-line distance in the multi-attribute evaluation space) has also been used [Qin et al., 2013]. Psychological distances can be used by including a new third parameter within the feedback matrix, w, so that distances between less competitive alternatives increase more slowly, as the Euclidean distance fails to account for the fact that some alternative attributes are more important than others [Hotaling et al., 2010]. Berkowitsch et al. [2015] build on this work by creating a generalised distance function for three or more attributes.

Calculating probabilities

Roe et al. [2001] demonstrate that once we have results for the expected value and the covariance of preference values at time t (ξ_t and Ω_t), we can calculate the probability of choosing alternatives. They show that on the basis of the multivariate central limit theorem, P_t converges to the multivariate normal distribution. Under decision field theory, A is chosen from a set $\{A,B,C\}$ if it has a higher preference value at time t than B and C. It can therefore be calculated as

$$Pr\left[P_t\left[A\right]-P_t\left[B\right]>0\cap P_t\left[A\right]-P_t\left[C\right]>0\right]=\int_{X>0}exp\left[-(X-\Gamma)'\Lambda^{-1}(X-\Gamma)/2\right]/(2\pi|\Lambda|^{0.5})dX \tag{9}$$

with $X=\left[P_t\left[A\right]-P_t\left[B\right],P_t\left[A\right]-P_t\left[C\right]\right]'$, $\Gamma=L\xi_t$, $\Lambda=L\Omega_tL'$ and

$$L=\begin{bmatrix}1 & -1 & 0\\ 1 & 0 & -1\end{bmatrix} \tag{10}$$

L is a matrix comprised of a column vector of 1s and a negative identity matrix of size $n-1$ where n is the number of attributes. The column vector of 1s is placed in the i^{th} column where i is the chosen alternative. We can then use (for example) the pmnorm package in R [Genz, 1992] to calculate the probability of each alternative being chosen.

Simplifying the deliberation stopping process

The 'computationally dissatisfying' process of summing over powers of S (equation 7) can be avoided by assuming that $t\to\infty$ [Berkowitsch et al., 2014]. Therefore, as long as the eigenvalues of S are less than one, $S^t\to\infty$. This reduces equation 6 to:

$$\xi_\infty=(I-S)^{-1}\cdot\mu \tag{11}$$

More importantly, however, is the simplified form of Ω_t. From Appendix B of Berkowitsch et al. [2014], we have:

$$\overline{\Omega_\infty}=(I-Z)^{-1}\cdot\overline{\Phi} \tag{12}$$

Where $\overline{\Phi}$ indicates that Φ has been transformed to a $1 \times n^2$ column vector and Z is a $n^2 \times n^2$ matrix based on S. This means that the laborious time-consuming summation in equation 7 can be avoided, but at the cost of assuming that all decision-makers take infinite response time to make their choices.

2.2 Arguments in favour of Decision Field Theory

There are numerous arguments in favour of using DFT. One of the main strength of DFT is that it is a dynamic model, where each alternative has a 'preference value', which fluctuates stochastically over time. This means that DFT can explain phenomena such as preference reversal [Diederich, 1997], something that static models, such as most RUM models, cannot do.

DFT is a flexible model, with two methods for a decision-maker to come to a conclusion. The decision-makers can stop deliberating either when they reach an internal threshold value for one of the alternatives or when they reach some external factor, such as a response time limit. This is a parallel to 'satisficing' behaviour [Simon, 1957] versus maximising behaviour, a concept that was explored by Schwartz et al. [2002]. Some individuals show satisficing behaviour, meaning they choose one of the alternatives when it is good enough (DFT's internal threshold), whereas others use the full time available to them to try and choose the best alternative, making a decision only when they have to (DFT's external threshold). Krosnick et al. [1996] demonstrated that satisficing behaviour can often occur when participants complete surveys and Wierzbicki [1982] provides one of the first models incorporating satisficing behaviour.

It has also been demonstrated that context effects, which DFT predicts efficiently, may be fundamental to decision making [Trueblood et al., 2013], with similarity, attraction and compromise effects all appearing in a perceptual decision task. Whilst there has not yet been a large impact from neuroscience on economics [Krajbich and Dean, 2015], Busemeyer et al. [2006] suggest that the accumulation of preference, as modelled by the behaviourally derived diffusion models in DFT, closely mimics neural activations in non-human primates during perceptual decision-making tasks. For example, Gold and Shadlen [2000] found evidence of an accumulating balance of sensory information favouring one interpretation over another in the neural circuits that generate and inform a monkey's choice. Ratcliff et al. [2003] similarly found that diffusion models as opposed to Poisson models better matched the evidence accumulation process seen in neural recordings. Schall [2003] adds that it appears that there are separate neurons initiating responses- a parallel to the threshold value within DFT.

DFT is also less of a 'black-box' process than typical RUM models. From a cognitive perspective, basic building blocks of cognition might be shared across a wide range of species and this bottom-top perspective is more in line with both neuroscience and evolutionary biology than the widely used top-down approach [De Waal and Ferrari, 2010]. DFT has a bottom-top perspective, an approach that some researchers believe to be fundamental to understanding individual's choices if we are to truly understand the underlying cognitive processes in decision making [Otter et al., 2008].

To add empirical confirmation, eye-tracking data is most consistent with attribute-and-alternative-wise comparison models [Noguchi and Stewart, 2014], where comparisons are made between pairs of alternatives on single dimensions. This would suggest that DFT is an appropriate model when there are two alternatives available, although empirical confirmations for multinomial alternatives are yet to be explored.

2.3 Transport applications of Decision Field Theory

The number of applications of DFT in transport thus far are limited and mainly theoretical. DFT has been suggested as an appropriate mechanism to explain the dynamics and high variability of choice decisions in congestion situations [Stern, 1999], due to its emphasis on an information-processing approach. Additionally, with some expansion, DFT should also be able to deal with a variety of travel situation effects including situational dynamics, type of travel, cultural habits and societal norms [Stern and Richardson, 2005]. The route choice process of a daily commuter according to DFT has been conceptualised [Stern and Portugali, 1999] and DFT has also been combined with the Queuing Network-Model Human Processor to model a driver's speed control [Zhao et al., 2011]. In an example of actually applying DFT in transport, it was found that given the duration to make a decision, DFT accurately predicted the percentage of participants who chose park and ride, car or bus and subway [Qin et al., 2013]. While these examples demonstrate the potential of DFT in numerous important and relevant applications within transport, they all work with small scale and overly simplified hypothetical studies with limited choice scenarios. Computational limitations of DFT [Otter et al., 2008] have also limited the impact of DFT in the transport literature and there is a distinct research gap in terms of operationalising DFT for full integration in mainstream transport models. For instance, the DFT models tested so far do not account for differences in socio-demographics of the decision-makers, which have been found to have significant effects on RUM and RRM frameworks.

2.4 Decision Field Theory vs Traditional Choice Models

In the only full comparison of DFT with RUM thus far, DFT performed as well as MNL and Probit at predicting consumer product choices made by participants [Berkowitsch et al., 2014]. Additionally, when eliciting context effects, an occurrence of multiple context effects within single participants was found and DFT then performed better than both MNL and Probit, in part due to it being built to cope with such effects. As far as the authors are aware, DFT has never been empirically tested against RRM.

A simplified description of how a DFT model works would be to compare it directly against a MNL or RRM model. Attributes of alternatives, M, are multiplied by W, the relative importance of the different attributes, which are equivalent to the β coefficients of MNL and RRM. We then get V, a valence vector, which can be considered as 'utility at a specific moment' and P, the total preference of alternatives vector, which is equivalent to the utility of alternatives in MNL and total regret in RRM. Whilst DFT models do not produce utilities, we can instead use the total preference of alternatives to calculate the likelihood of alternatives (see equation 4 and section 2.1). Additionally, ϕ_1, the sensitivity parameter, allows for the similarity effect to happen under a DFT model. This means that more similar alternatives compete more with each other, a parallel to the effect the nesting parameter, which captures the correlation across alternatives, has in a nested logit model [Williams, 1977, Daly and Zachary, 1978, McFadden, 1978]. Figure 1 shows a connectionist interpretation of DFT [Roe et al., 2001]. This demonstrates graphically how the total preference of alternatives is calculated (see equations 1 and 2 for mathematical details).

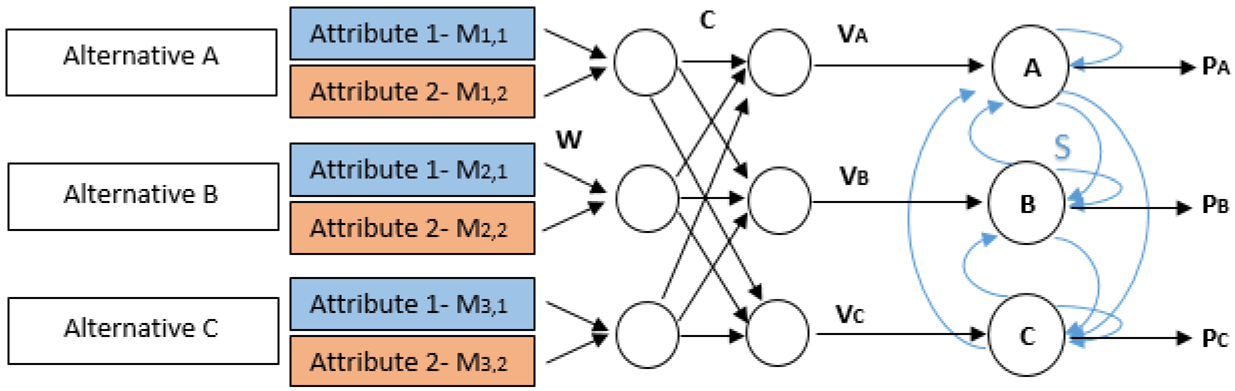

FIGURE 1 : A connectionist interpretation of DFT, adapted from Roe et al. [2001]

2.5 Summary of key DFT applications thus far

Thus far, there have not been many studies that have actually applied Decision Field Theory (particularly within transport). Table 1 provides a summary of some key DFT applications, and highlights the major differences. Only one major application has been published in a transportation journal until now, with most in psychological journals such as Psychological Review. Whereas some applications only use DFT to calculate the probability of alternatives, there are others that fit choice data to calculate the likelihood of multiple choices, as typically done in a choice modelling study. Very few applications estimate all parameters, with some often held constant. Decision field theory has been applied across a variety of types of choices, most often consumer choice, but also decision making in basketball. Most applications fix the number of timesteps, as prior to this paper, there was no closed form expression for calculating the probabilities of more than three alternatives at any timestep (see section 3.1). When only two alternatives are considered, the number of timesteps does not need to be estimated as Busemeyer et al. [2006] demonstrates that the probability of alternatives can be calculated by instead estimating an internal threshold.

While comparisons with RUM are limited, DFT has been compared against a number of different models including the proportional difference model (first introduced by Gonzalez-Vallejo [2002] and compared against DFT by Scheibehenne et al. [2009]), the multiple linear ballistic accumulator [Trueblood et al., 2014] and traditional choice models such as Logit and Probit [Berkowitsch et al., 2014].

TABLE 1 : Some key DFT applications

Authors	Journal	Type of estimation	Parameters	Type of choices	Key assumptions	Key DFT results
Roe et al. (2001)	Psychological Review	Probability of alternatives	Some estimated, some fixed	-	-	Demonstrated how DFT explains context effects
Raab and Johnson (2004)	Research Quarterly for Exercise and Sport	Probability of alternatives	Some estimated, some fixed	Basketball decisions	-	Different initial preferences explained individual choices best
Scheibehenne et al. (2009)	Cognitive Science	Likelihood of multiple choices, by individual	All estimated	Monetary gambles	Two alternatives, so a timestep parameter is not required	DFT performs better than the proportional difference model
Tsetsos et al. (2010)	Psychological Review	Probability of alternatives	Some estimated, some fixed	-	Used steady preference states after a large number of timesteps	DFT performs less well than LCA at explaining context effects
Hotaling et al. (2010)	Psychological Review	Probability of alternatives	Some estimated, some fixed	-	-	DFT obtains more robust predictions with internal stopping rules
Hey et al. (2010)	Journal of Risk and Uncertainty	Likelihood of multiple choices, by individual	All estimated	Monetary gambles	Two alternatives, so a timestep parameter is not required	DFT predicts risky choice better than most other models considered
Qin et al. (2013)	Transportation Research Part F	Probability of alternatives	Some based on questionnaire	Mode choice	Feedback coefficients not estimated	Simulated DFT results match survey results
Trueblood et al. (2014)	Psychological Review	Likelihood of multiple choices across decision makers	Some estimated, some fixed	Likely crime suspects	1001 timesteps	DFT performs less well than MLBA at explaining context effects
Berkowitsch et al. (2014)	Journal of Experimental Psychology	Likelihood of multiple choices across decision makers	All estimated	Consumer decisions	Infinite time steps	DFT performs as well as MNL and Probit
Noguchi and Stewart (2014)	Cognition	Probability of alternatives	Some estimated, some fixed	Consumer decisions	1000 timesteps	Eye-tracking suggests alternatives are compared rather than individually evaluated

3 IMPROVEMENTS TO DECISION FIELD THEORY

The following section provides a method for avoiding the sacrifice by Berkowitsch et al. [2014] whilst simultaneously avoiding computationally intensive simulations. We then present methods for incorporating heterogeneity across and within decision-makers into Decision Field Theory.

3.1 Avoiding the sacrifice of response time being set to infinity

It has been argued that the lack of analytical solutions for DFT means that it has to use computationally intensive simulations [Otter et al., 2008] and should be used with an externally controlled stopping procedure with a large value for response time [Trueblood et al., 2014, Noguchi and Stewart, 2014]. However, Hotaling et al. [2010] argued that the undesirably long fixed stopping times used by Tsetsos et al. [2010] was in part why their DFT model performed worse than their own rival preference accumulation model, the Leaky Competing Accumulator (a model designed to address challenges to previous diffusion, random walk and accumulator models), suggesting that large values for response time should be avoided if possible.

Berkowitsch et al. [2014] avoided arbitrarily setting the number of timesteps by fixing it to infinity, as shown in the previous section. We will now, however, show that as well as being an undesirable sacrifice, this is an unnecessary one. Firstly, we show that the following matrix can be rearranged to a more usable format as follows:

$$\overline{S\Phi S'} = Z\overline{\Phi} \tag{13}$$

where S is the feedback matrix and Φ is the covariance of V_t as before. Again, $\overline{X}$ indicates that matrix X of size $n \times n$ has been reshaped into a column matrix of size $1 \times n^2$. Now if we start with any 3 matrices of size $n \times n$,

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} B = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \dots & b_{nn} \end{bmatrix} C = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \dots & c_{nn} \end{bmatrix} \tag{14}$$

we have that for multiplying matrix A by B, the entry $[AB]_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$. Therefore if we set $ABC = D$, we have entries $[D]_{ij} = \sum_{k=1}^{n} \sum_{l=1}^{n} \left[a_{il} b_{lk} c_{kj}\right]$. Now if we reshape D into a column matrix as before, we have $\overline{D}$ with entries:

$$\left[\overline{D}\right]_{(j-1)n+i} = \sum_{k=1}^{n} \sum_{l=1}^{n} \left[a_{il} b_{lk} c_{kj}\right] \tag{15}$$

Next, we wish to create a new matrix Z of size $n^2 \times n^2$ and to reshape B into a column matrix:

$$Z = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1n^2} \\ z_{21} & z_{22} & \cdots & z_{2n^2} \\ \vdots & \vdots & \ddots & \vdots \\ z_{n^21} & z_{n^22} & \cdots & z_{n^2n^2} \end{bmatrix} \overline{B} = \begin{bmatrix} b_{11} \\ b_{21} \\ \vdots \\ b_{n1} \\ b_{12} \\ \vdots \\ b_{nn} \end{bmatrix} \tag{16}$$

Multiplying these together gives $Z\overline{B}$ with entries $\left[Z\overline{B}\right]_i = \sum_{k=1}^n \sum_{l=1}^n \left[z_{i,(k-1)n+l} b_{lk}\right]$. This gives us

$$\left[Z\overline{B}\right]_{(j-1)n+i} = \sum_{k=1}^n \sum_{l=1}^n \left[z_{(j-1)n+i,(k-1)n+l} b_{lk}\right] \tag{17}$$

Thus for $Z\overline{B} = \overline{D}$, we need only set $z_{(j-1)n+i,(k-1)n+l} = a_{il}\, c_{kj}$. Hence we can rearrange equation (13) to this more useful format by setting $A = S, B = \Phi$ and $C = S'$ and following the above steps to find the new matrix Z. We now wish to show that

$$\overline{S^n \Phi S^{n\prime}} = Z^n \overline{\Phi} \tag{18}$$

To do this, We employ a proof by induction. We have that equation (18) holds when $n = 1$ as we know that equation (13) is true. This means that if we can show that equation (19) holds, then we will have proved that equation (18) holds when $n = [2, 3, 4, ...]$.

$$\overline{S^{n+1} \Phi S^{n+1\prime}} = Z^{n+1} \overline{\Phi} \tag{19}$$

Firstly, we set the matrices $A^n = X$, $C^n = Y$ and $Z^n = W$. Then the elements of the left side matrix of equation (19) are:

$$\left[A^{n+1} B C^{n+1}\right]_{ij} = [AXBCY]_{ij} = \sum_{k=1}^n \sum_{l=1}^n \sum_{r=1}^n \sum_{s=1}^n \left[a_{ir}\, x_{rl}\, b_{lk}\, y_{ks}\, c_{sj}\right] \tag{20}$$

$$\Rightarrow \overline{[AXBCY]}_{(j-1)n+i} = \sum_{k=1}^n \sum_{l=1}^n \sum_{r=1}^n \sum_{s=1}^n \left[a_{ir}\, x_{rl}\, b_{lk}\, y_{ks}\, c_{sj}\right] \tag{21}$$

Now for the right hand side matrix, from the previous result for equation (13) we can set

$$z_{(j-1)n+i,(k-1)n+l} = a_{il}\, c_{kj} \tag{22a}$$

$$w_{(j-1)n+i,(k-1)n+l} = x_{il}\, y_{kj} \tag{22b}$$

and when multiplying these matrices together, we get

$$[ZW]_{uv} = \sum_{r=1}^n \sum_{s=1}^n \left[z_{u,(s-1)n+r}\, w_{(s-1)n+r,v}\right] \tag{23}$$

From before we had

$$\left[Z\overline{B}\right]_i = \sum_{k=1}^{n}\sum_{l=1}^{n}\left[z_{i,(k-1)n+l}b_{lk}\right] \tag{24}$$

so for ZW this becomes

$$\left[ZW\overline{B}\right]_i = \sum_{k=1}^{n}\sum_{l=1}^{n}\left[[ZW]_{i,(k-1)n+l}\,b_{lk}\right] \tag{25}$$

Substituting back in equation (23) and we get

$$\left[ZW\overline{B}\right]_i = \sum_{k=1}^{n}\sum_{l=1}^{n}\left[\sum_{r=1}^{n}\sum_{s=1}^{n}\left[z_{i,(s-1)n+r}\,w_{(s-1)n+r,(k-1)n+l}\right]b_{lk}\right] \tag{26}$$

Finally using equations (22) and rearranging, the right hand side of equation (19) becomes

$$\begin{aligned}
\left[ZW\overline{B}\right]_{(j-1)n+i} &= \sum_{k=1}^{n}\sum_{l=1}^{n}\sum_{r=1}^{n}\sum_{s=1}^{n}\left[z_{(j-1)n+i,(s-1)n+r}\,w_{(s-1)n+r,(k-1)n+l}\,b_{lk}\right] && \text{(27a)}\\
&= \sum_{k=1}^{n}\sum_{l=1}^{n}\sum_{r=1}^{n}\sum_{s=1}^{n}\left[a_{ir}\,c_{sj}\,x_{rl}\,y_{ks}\,b_{lk}\right] && \text{(27b)}\\
&= \overline{[AXBCY]}_{(j-1)n+i} && \text{(27c)}
\end{aligned}$$

Hence, we have that $Z^{n+1}\overline{B} = \overline{A^{n+1}\,B\,C^{n+1}}$ and the induction is complete. Finally, using this result, we can return to equation (7), which now simplifies to become

$$\begin{aligned}
Cov[P_t] = \Omega_t &= \sum_{k=0}^{t-1}\left[S^k \cdot \Phi \cdot S^{k'}\right] && \text{(28a)}\\
&= \sum_{k=0}^{t-1}\left[Z^k \cdot \overline{\Phi}\right] && \text{(28b)}\\
&= (I-Z)^{-1}\,(I-Z^t)\,\overline{\Phi} && \text{(28c)}
\end{aligned}$$

with Z being created from elements of the feedback matrix S by setting $z_{(j-1)n+i,(k-1)n+l} = s_{il}\,s_{jk}$ for $i, j, k, l \in [1:n]$. This means that the time consuming sum calculation has been removed and we can therefore return to having a finite value for t. We thus restore the original core psychological foundations of DFT whilst simultaneously avoiding intensive calculations. [2]

In section 4.2 we compare the results of different versions of DFT, looking at the implications of making this simplification. We compare our version of DFT (DFT-2017), where we estimate the number of timesteps a decision-maker takes to reach a conclusion, against the previous version of DFT (DFT-2014), where decision-makers preferences are assumed to have stabilised over an infinite time. DFT-2014 can be incorporated within DFT-2017, simply by setting the number of timesteps to a high value.

[2]Note that equation 28c becomes equation 12 for $t \to \infty$ if the eigenvalues of the feedback matrix are less than one (or equivalently $\phi_2 > 1$), in which case $S^t \to 0$ [Roe et al., 2001] and hence $Z^t \to 0$

3.2 Alternative specific constant (asc) parameters

One of the strengths of RUM models is their ability to measure baseline preferences of some alternatives through the use of alternative specific constants. DFT can partly accommodate this through P_0, the initial preference vector. Crucially, this leads to just an *initial* preference for this alternative, meaning that the preference disappears as the decision time increases and the number of timesteps becomes high. As DFT-2014 has an infinite number of timesteps, this means that it has no method for accommodating initial preferences. However, Roe et al. [2001] used an additional weight assigned to a zero column matrix, as a way of reflecting that the decision-maker was attending to 'other irrelevant' attributes (which may actually be relevant). We can expand on this idea by having an additional attribute 'looking at other factors favouring alternative x.' This would have attribute levels of y for alternative x, and 0 for all other alternatives. We can either fix y to being a specific value, or allow it to fluctuate by adding it in as another parameter in the same way that alternative specific constants are. Depending on the number of alternatives, more of these additional attributes can be added as required. This gives us two methods for DFT-2017 and one method for DFT-2014 to deal with preferences towards alternatives, all of which are explored in section 4.3.

3.3 Adding heterogeneity

Decision Field Theory has almost always been implemented as a 'one size fits all' model, with an exception being Raab and Johnson [2004], who looked at individual differences in action taking within sport (although this looked at just a single DFT choice scenario, as the attributes were not clearly defined). Scheibehenne et al. [2009] and Hey et al. [2010] also considered individual differences by computing separate DFT models for each decision maker, but as far as we are aware, no studies have thus far fitted a DFT model to multiple decision makers across multiple decisions whilst simultaneously incorporating individual differences. This is surprising given that DFT has psychological origins, where individual differences tend to be better appreciated. Johnson [2006] highlighted the need for DFT to be able to explain individual differences and Liew et al. [2016] found that, as a contradiction to the findings of Berkowitsch et al. [2014], participants rarely showed all three context effects, highlighting the dangers of averaging indiscriminately and not having a method for dealing with individual differences.

We believe that there is no reason that DFT cannot be expanded in exactly the same way that Multinomial Logit (MNL) has been within RUM. A 'Mixed DFT' could incorporate some of the ideas of Mixed Logit [McFadden and Train, 2000]: some parameters could be changed from being fixed to having a mean and a standard deviation instead. Whilst some caution on ranges of the distributions would be required (e.g. positive only weight parameters), there is nothing to suggest that there could not be individual variation in any one the parameters within DFT. In section 4.5 we explore the results of using random parameters in a DFT model as well as the effects of using different distributions for these parameters.

4 EMPIRICAL APPLICATION

4.1 Datasets

In this section we summarise the datasets that we have used to test the explanatory and predictive power of Decision Field Theory against other models as well as finding the best methods to maximise the output from a DFT model.

Simulated dataset A (SD-A)

The first simulated dataset contains 1,000 choice situations, each with two attributes 'A' and 'B', and two alternatives '1' and '2'. The attribute values were drawn from a uniform distribution from 1 to 10. After a random number, $0 < r < 1$, was created for each decision, the probability of choosing alternative 1 was defined as $0.05\,W_A(A_1 - A_2) + 0.05\,W_B(B_1 - B_2) + 0.5 > r$, where W_A and W_B where the weights of the attributes, both set to 0.5 by default. We use this basic dataset with simple choices to test the ability of a DFT model to capture the effect of underlying preferences for an alternative in section 4.3. A preference for (arbitrarily) alternative 1 is added in by defining that for any choice task with a random number of less than a certain value, the decision-maker would always pick alternative 1.

Simulated dataset B (SD-B)

The second dataset contains 8,000 choice situations, each with six attributes 'A' through to 'F' and two alternatives '1' and '2'. Each attribute value was either true or false. An MNL model was used to simulate the choices (with coefficients $\beta_A = -0.6$, $\beta_B = -0.5$, $\beta_C = -0.4$, $\beta_D = -0.3$, $\beta_E = -0.2$ and $\beta_F = -0.1$). The aim of testing this dataset is to see how well DFT copes with binary attributes and to compare it against MNL as detailed in section 4.4.

Simulated dataset C (SD-C)

The third dataset also contains 8,000 choice situations, this time with four attributes- cost (TC), travel time (TT), number of changes (CH) and availability of seating (AS). An MNL model was again used to calculate the probabilities of each alternative being picked (with coefficients $\beta_{TC} = -0.5$, $\beta_{TT} = -0.05$, $\beta_{CH} = -0.5$ and $\beta_{AS} = -0.5$). This time a group difference was added in, such that group '2' attached 3 times more value to β_{AS} (for instance, in real life, the decision of some travellers may be strongly affected by the availability of seating). This dataset could then be used to test the ability of DFT to cope with socio-demographic differences as detailed in section 4.5.

Swiss stated preference dataset (SP-1)

Our first stated preference dataset comes from the Swiss value of time study [Axhausen et al., 2008], and specifically a route choice example for rail users, where 389 participants each completed 9 binary choice tasks described by 4 variables: travel time, travel cost, headway and the number of changes.

UK stated preference dataset (SP-2)

The second stated preference dataset uses 10 choice tasks from each of 368 participants, all of whom are public transport commuters in the UK. Each task involves an invariant reference trip and two hypothetical alternatives. Each alternative is described by travel time, cost, rate of crowded trips, rate of delays (both out of 10 trips), the average length of delays (across delayed trips) and cost of a provision of a delay information service [Hess and Stathopoulos, 2013].

4.2 Differences between different DFT models

In this section we compare DFT-2014 (DFT without a time parameter) against DFT-2017 (DFT with a time parameter). The DFT model with a time parameter uses the method described in section 3.1 whilst the one without follows the method of Berkowitsch et al. [2014], where response time is set to infinity. We also compare these models against simple multinomial logit models and also two versions of random regret minimisation models (the first following the specification of [Chorus, 2010]) and the second following [van Cranenburgh et al., 2015], incorporating μ, a parameter to estimate a profundity of regret). For SP-1, our MNL and RRM models contain five parameters, four for the attributes and one alternative specific constant. SP-2 has an additional attribute and an additional alternative, resulting in seven parameters. The μ-RRM models have six and eight parameters respectively with the addition of the μ parameter. The DFT models have three and four parameters respectively for attributes in SP-1 and SP-2. All DFT models additionally have sensitivity, memory and error parameters (ϕ_1,ϕ_2 and ε) and DFT-2017 models also have a parameter for the number of timesteps.

TABLE 2 : Results from removing the sacrifice of setting response time to infinity

Dataset	Swiss (SP-1)				UK (SP-2)			
	LL	free par.	BIC	timestep estimate	LL	free par.	BIC	timestep estimate
MNL	-1,667.97	5	3,377		-3,721.67	7	7,501	
RRM	-1,667.97	5	3,377		-3,699.49	7	7,456	
μ-RRM	-1,667.97	6	3,405		-3,698.89	8	7,463	
DFT-2014	-1,595.88	6	3,241	∞	-3,676.34	7	7,410	∞
DFT-2017	-1,595.85	7	3,249	122.22	-3,598.87	8	7,263	3.78

From Table 2 we can see that for SP-1, adding the time parameter has very little impact. However, for SP-2, we see that adding a time parameter results in a significant improvement in the model. As SP-1 only has two alternatives, RRM achieves the same result as MNL. For SP-2, RRM and μ-RRM provide significantly better fit than MNL but significantly worse fit than DFT, especially compared to DFT-2017 (see appendix A for full DFT model estimates). It appears that the difference in performance between the different DFT models is due to the estimate for the number of timesteps. When this value is high, DFT-2017 approximately becomes DFT-2014. This results in DFT-2014 and DFT-2017 producing very similar parameter estimates (see Table 14). For SP-2, the estimate for the number of timesteps is small for DFT-2017, resulting in a better fit than DFT-2014.

Whilst the weight estimates are similar (Table 15), the psychological parameters also have somewhat different estimates. However, the minimal impact these parameters have on the preference values (see appendix D) suggests that the difference in goodness of fit is more likely to be a result of the difference in the number of timesteps. For SP-1, DFT has two more parameters than MNL, and it could be argued that BIC values do not penalise this difference enough. However, we find that the best fitting MNL model with an additional two parameters (square root terms for cost and time) has a log-likelihood of $-1,615.79$, which is still significantly worse than DFT.

4.3 Implementation and application of Decision Field Theory

In this section we look at how to best implement and apply a Decision Field Theory model. We consider the implications of the weight parameters having to be greater than zero, look at methods for DFT to incorporate underlying preferences for an alternative and look at the effect of different scaling methods being used on the attribute levels [3].

Implications of Decision Field Theory weight parameters having to be greater than zero

Using the Swiss stated preference dataset (SP-1), it is quickly possible to see the effect of having undesirable attributes in DFT. If a value for an undesirable attribute is positive and high (for example, a large cost), then an appropriate DFT model would factor this in by adding a negative valence to the preference value of an alternative when this attribute is considered. However, the weight parameters in a DFT model cannot be negative, as w_i represents the proportion of time that a decision-maker looks at attribute i. This causes issues when we have 'positive', desirable attributes (such as quality), and 'negative', undesirable attributes (such as travel cost). If the attributes were to be left as they were, then due to DFT being an accumulative model and weights being positive, there would be no way for DFT to reflect that an alternative is more likely to be picked if an attribute level is lower. This means that DFT will have its greatest predictive accuracy when negative attributes are ignored, and their weights are set to zero. Table 3 shows the log-likelihood values of SP-1 under DFT models where some attributes are desirable/undesirable. As all four attributes are undesirable, 'negative' here means that the higher values are less desirable, whereas positive means that they have been reset such that higher, more positive values are more desirable. The table also shows the parameter estimates for the DFT models. As DFT has no clear starting points for estimation, we have to run a number of trials to find a suitable starting point. We set the weight attributes to be equal and use random numbers to set ϕ_1 between 0 and 10, ϕ_2 between 0 and 1, ε between 0 and $1,000$ and t between 0 and 100. We ran 100 trials of this nature and then used the best as the starting point in the R package maxLik [Henningsen and Toomet, 2011]. We found that the inclusion of the third feedback parameter, w, made an insignificant difference to the results of DFT, therefore omitted it in these trials and used Euclidean distances in the feedback matrix. We used standard score normalisation to scale the attributes in this section, but explore scaling methods further in section 4.3.

We can see from Table 3 that if the travel costs are negative (Model 3), the parameter for travel cost, w_{TC}, quickly drops towards zero, reflecting that the DFT model has not used the information as to do so would worsen model fit. An equivalent hindrance on a MNL model, where the beta coefficient for travel cost is fixed to zero, suffers a similar loss in log-likelihood. When headway

[3] From here, DFT-2017 is always used unless otherwise specified

TABLE 3 : Parameter estimates and log-likelihoods for DFT models for positive and negative attributes (using SP-1)

Model	1		2		3		4	
DFT LL	-2,000.61		-1,952.68		-1,724.55		-1,595.85	
Equivalent MNL LL	-2,039.46		-1,976.24		-1,722.97		-1,667.97	
Travel Time (TT)	Negative		Positive		Positive		Positive	
Travel Cost (TC)	Negative		Negative		Negative		Positive	
Headway (HW)	Negative		Negative		Positive		Positive	
Changes (CH)	Positive		Positive		Positive		Positive	
	est	t-ratio	est	t-ratio	est	t-ratio	est	t-ratio
w_{TT}	0.0000	0.00	0.4528	5.26	0.3214	5.86	0.3475	45.31
w_{TC}	0.3973	2.49	0.0000	0.00	0.0005	0.07	0.4676	43.22
w_{HW}	0.0001	0.00	0.0054	0.11	0.2838	15.18	0.0747	12.85
ϕ_1	0.1806	3.20	0.1206	3.08	0.5959	4.59	142.6043	67.97
ϕ_2	0.6645	3.52	0.6748	4.83	0.6012	8.47	0.1835	12.71
ε	11.1728	6.79	9.9875	8.28	1.4430	5.39	0.0017	0.96
t	10.0038	4.00	9.0836	8.42	12.3834	5.72	112.2185	1,805.36

is also negative (model 2), w_{HW} also drops to zero. The value for the error term, ε, has increased significantly. Hotaling et al. [2010] argue that the error term (also known as the noise variance [Roe et al., 2001]) would be higher for more complex tasks, meaning that higher values for ε would be found for less predictable decisions, as demonstrated here. Model 1, where travel time is made negative, obtains further losses in model fit. Perhaps surprisingly, the coefficient for travel cost weight increases away from zero. However, this is due to alternatives with low travel time and high cost being generally preferred to alternatives with high travel time and low cost. This is also reflected in an MNL model with just beta coefficients for travel cost and the number of changes, in which a positive value ($\beta_{TT} = 0.015$, t-ratio$= 2.34$) for travel cost is found.

The memory parameter, ϕ_2, is higher for models with negative attributes, suggesting that DFT predicts that the participant 'forgets' the information they are looking at and that the choice is more down to chance instead, as can be seen by the higher values for ε. This makes sense given that when DFT has negative attributes, preference values increase as the alternative becomes less likely to be picked, meaning that the less the attribute values are used to make the predictions, the better. If all attributes are positive, DFT starts to outperform MNL more significantly. The memory parameter has dropped close to zero as the accumulated preference values are now reflected in the likelihood of an alternative being chosen. In conclusion, this shows that we need to invert all negative attributes such that the higher an attribute value, the more desirable an alternative is. This will improve the performance of a DFT model, but poses problems when we do not know if an attribute is desirable or not. If this is the case, we can simply include the attribute twice, once with the original values and once with the inverted values. The weight for one will quickly drop to zero indicating whether the attribute is positive or negative.

Dealing with underlying preferences

Random Utility Models with a Multinomial Logit framework deal with underlying preferences through alternative specific constants [McFadden and Train, 2000]. These values directly capture market shares. DFT has two methods for capturing shares and dealing with preferences towards an alternative. One is through the initial preference matrix P_0 and the other is by creating a new attribute favouring one of the alternatives. Here we simply add in a dummy variable for the alternatives with a higher value for one of the alternatives. Table 4 displays the results of adding in additional DFT parameters to deal with preferences in simulated dataset A. A parameter pr_1 indicates an initial preference for alternative 1 in P_0 and parameters w_1, w_2 indicate weights for new attributes favouring alternative 1 and 2 respectively.

TABLE 4 : The effect of underlying alternative preferences on DFT models (using SD-A)

Proportion always choosing alternative 1	DFT-2017	additional w_1	additional w_1, w_2	additional pr_1	MNL	pr_1 estimate
0.1	-653.47	-629.86*	-629.86	-628.36	-630.31	23.6
0.2	-652.29	-568.79	-568.80	-565.50	-569.70	55.0
0.3	-652.94	-447.59	-447.59	-444.24	-450.10	124.3
0.4	-667.91	-303.43	-303.43	-299.35	-306.38	216.2
LL of Null model = -693.15 Parameters = 3 (MNL), 5 (DFT-2017). Observations = 1,000						

For the models in Table 4, we set the attribute value difference for the new parameter w_1 to 5 arbitrarily in every case. A value of -629.99 was achieved in case * when a value of 1 is used instead. This value could be set as another parameter, but as the value has not changed significantly we have not explored this further.

Whereas adding in parameter w_1 for a preference of alternative 1 makes a difference, the weight for the preference of alternative 2, w_2, drops to 0. This means that we can treat these parameters equivalently to alternative specific constants in random utility models, where similarly only one parameter would be needed to capture the difference in underlying preferences between two alternatives. We can see from Table 4 that adding in w_1 results in DFT achieving similar LL values to MNL. However, better values are achieved by adding in the parameter pr_1. As the percentage of choices where decision-makers always choose alternative 1 increases, the parameter estimate for pr_1 rises, as does the difference between MNL and DFT LL values.

Using DFT-2014, where the number of timesteps is set to infinity, results in only a few differences for this dataset. This is because the estimate for the number of timesteps is high (see Table 3). Without additional weights, the same log-likelihood values are achieved. However, with one additional weight parameter, the log-likelihood is -447.78 when the proportion always choosing alternative 1 is 0.3, slightly lower than the log-likelihood achieved for DFT-2017. This difference results from the parameter estimate for ϕ_2, the memory parameter, being negative for this model (-0.0004). The estimate for ϕ_2 cannot be negative under a version of DFT where the number of timesteps is infinite, as this results in eigenvalues of the feedback matrix being greater than one, resulting in a preference matrix that never converges.

Scaling of attributes

The most common method for scaling attributes that has been used in previous applications of DFT has been to rescale values to be between two values (unity-based normalisation) [Berkowitsch et al., 2014, Johnson, 2006]. We now consider different methods for scaling the attribute values in dataset SP-1 and the effects this has on the parameter estimates for DFT. The first method we use is unity-based normalisation, where we have a minimum value of 0 and maximum value of 1. We set $a_i = 1 - \frac{a_i - min(a)}{max(a) - min(a)}$ for each of the attributes, ensuring that we set the most desirable attribute level (lower costs and travel times) to be close to 1 and less desirable attribute levels to be close to 0. For the second method, we do not scale the attributes at all, simply setting $a_i = -a_i$. For the third method, we use standard score normalisation and set $a_i = -\frac{a_i - mean(a)}{sd(a)}$. The fourth method employs the same values as the third, with the exception that the travel time values are additionally all multiplied by 10.

Table 5 shows the weight estimates for each attribute in the different DFT models as well as the MNL beta coefficient values for travel time (TT), travel cost (TC), headway (HW) and the number of changes made when travelling by train (CH). As with other departures from RUM, value of time and similar measures cannot be directly calculated under a DFT model. We instead define 'relative importance of time,' as $\frac{w_{TT}/S_{TT}}{w_{TC}/S_{TC}} \times 60$, where S_i is the scale factor used for scaling attribute i, and use the value of time for the relative importance of time under MNL.

TABLE 5 : Parameter estimates (t-ratios in brackets) for DFT models under different types of scaling for SP-1

Model	LL value	TT	TC	HW	CH	Relative importance of travel time CHF/ hour	Relative importance of changes CHF/ change
DFT scale 1	-1,640.34	0.3462 (32.53)	0.5887 (45.23)	0.0261 (12.34)	0.0390	24.43	8.88
DFT scale 2	-1,622.83	0.0565 (17.18)	0.1390 (12.95)	0.0288 (17.83)	0.7757	24.38	5.58
DFT scale 3	-1,595.85	0.3473 (45.32)	0.4678 (43.23)	0.0747 (12.85)	0.1102	21.37	6.57
DFT scale 4	-1,638.43	0.0615 (24.10)	0.6202 (29.02)	0.1330 (13.76)	0.1853	28.54	8.33
MNL	-1,667.97	-0.0598 (-14.04)	-0.1318 (-9.76)	-0.0375 (-20.34)	-1.1528 (-26.56)	27.21	8.74

We can see from Table 5 that DFT produces a higher log-likelihood value than MNL does for SP-1. This difference gradually increases as we change from scaling method 1 through to 3.

It appears that because of the large range of attribute values, some of the information is lost in method 1. The high weight value for the number of changes in scale 2, 0.7767, shows that due to the lack of scaling, the decision-maker has to attend to the number of changes more often for the importance of this attribute to be reflected. The best log-likelihood value under DFT is achieved in scale 3, suggesting that it is important to include information on means and standard deviations when scaling the attribute values before running a DFT model. The relative importance of travel time and the relative importance of changes estimates vary depending on which type of scaling is used. It appears that as the model's log-likelihood value improves, both relative importance values decrease.

Table 5 also shows the impact of multiplying the travel times in SP-1 by 10 in DFT scale 4. The only difference this makes for a RUM Multinomial Logit is that the travel time coefficient becomes exactly 10 times lower. As a contrast, DFT does not equivalently have a simple change of coefficients. Instead, we see that w_{TT} has decreased from 0.3473 to 0.0615. This reflects the fact that to capture the relative importance of time, it has to be attended to less often relative to the other attributes for it to be appropriately incorporated into the model. We see a lower value of log-likelihood, with the increase in relative importance values likely being the cause.

4.4 Differences in results between Decision Field Theory and other models

Exploring the differences between RUM Multinomial Logit and Decision Field Theory probabilities

The different scaling methods for the attribute values for a DFT model has a big impact on the differences between DFT and MNL model probability of alternatives for SP-1. Figure 2 demonstrates that scales 1 and 3 in particular find that when the number of changes and headway is the same for both alternatives, DFT makes a more extreme prediction than MNL, indicated by the grey points on the figure. This does not happen under scale 2, where DFT makes more conservative predictions.

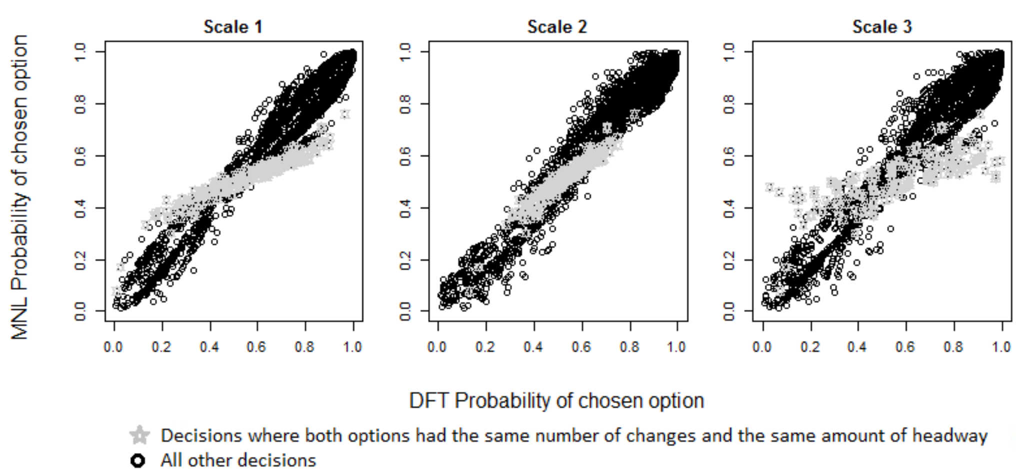

FIGURE 2 : Difference between MNL and DFT probabilities of chosen alternatives for SP-1

Linear regression results on the absolute difference between MNL probabilities and DFT scale 3 probabilities show that this difference is greatest when differences between alternatives are small, but that it decreases as the difference between the number of changes, travel time and headway between the alternatives increases.

This is also the case in simulated dataset A, where similarly, linear regression shows that the absolute difference between MNL and DFT probabilities decreases as the differences between the alternatives increase. Simulated dataset B finds an extremely small average difference between MNL and DFT of $9.51e-05$, with standard deviation 0.0041 and a largest difference of 0.008. This suggests that if alternatives only have true/false attributes, DFT will produce very similar results to MNL.

We also have from figure 3 that the relative importance of travel time has a significant impact on the difference between MNL and DFT. The relative importance of travel time of an alternative is defined here as being positive if the more expensive, faster alternative is chosen and negative if the cheaper, slower alternative is chosen. For example, a value of 50 indicates that the decision-maker is spending 50CHF per hour saved.

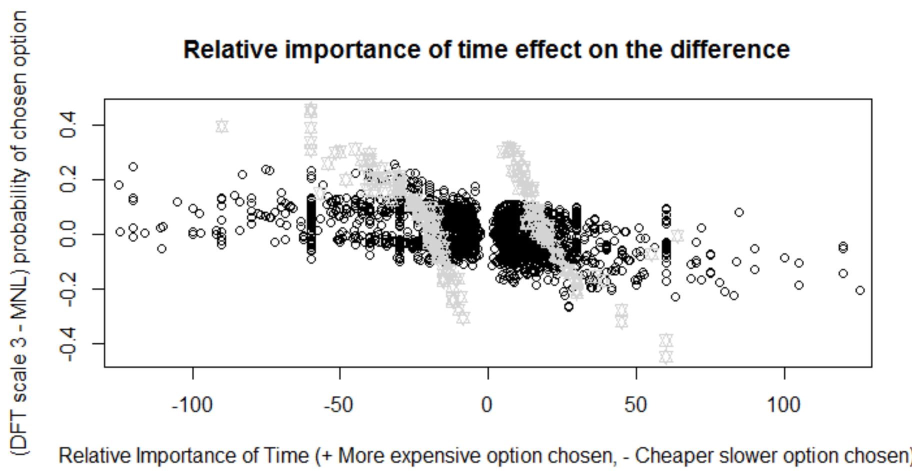

FIGURE 3 : Impact of the relative importance of time on the difference between MNL and DFT probabilities of chosen alternatives for SP-1

From Table 5 we can see that DFT (scale 3) predicts a relative importance of travel time of 21.4 whereas MNL predicts a relative importance of travel time of 27.2. This difference is reflected in figure 3 by the fact that the lower the relative importance of travel time is, the larger the difference between DFT and MNL becomes in favour of DFT. We also see for decisions that are purely a trade-off between time and cost (grey points), the impact of the value of the relative importance of time is larger for DFT.

For SP-2, it appears that DFT gives more importance to the cost of the alternatives than MNL (see figure 4). Whilst both models tend to predict chosen alternatives with a probability of closer to 1 for cheaper alternatives, linear regression confirms that the cheaper the chosen alternative is relative to the unchosen alternatives, the better the fit of DFT in comparison to MNL. Results from linear regression also imply that the reverse is true for other attributes: they have more of an impact on an MNL model.

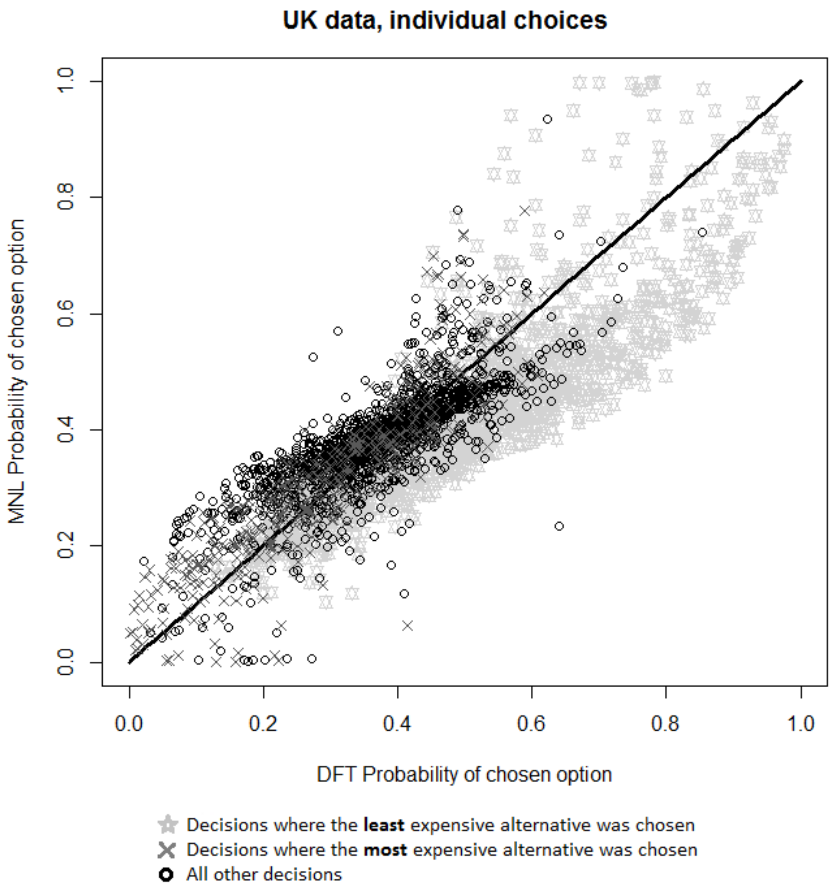

FIGURE 4 : Difference between MNL and DFT probabilities of chosen alternatives for SP-2

We next consider the differences between MNL and DFT from the perspective of individuals (see figure 5). Whilst individuals seem to be fairly evenly distributed for both datasets, linear regression finds that for both, DFT tends to do better for more predictable individuals who contribute high log-likelihood values, whereas MNL provides a better fit for less predictable individuals.

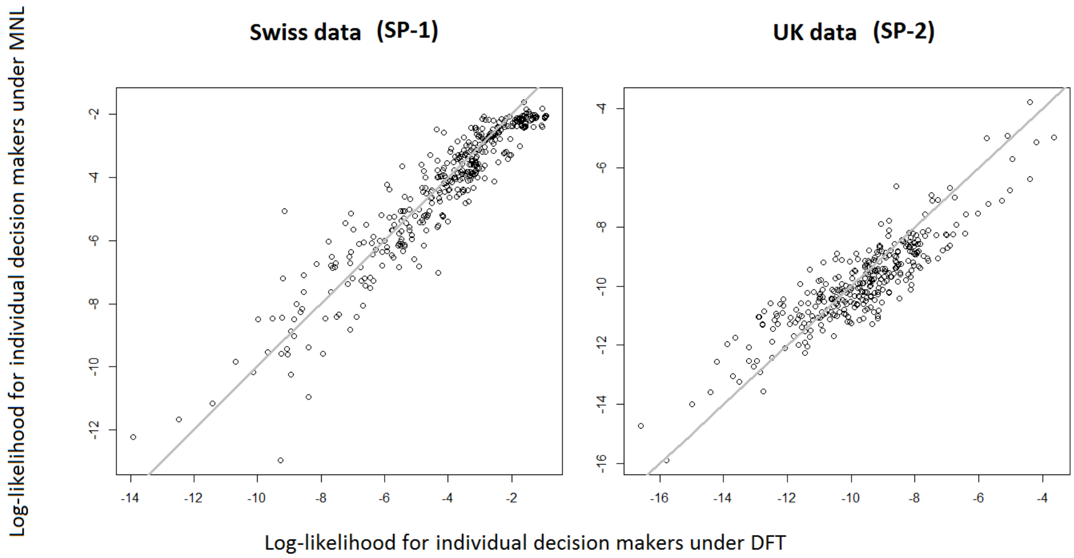

FIGURE 5 : Difference between MNL and DFT log-likelihoods for individuals

Runtime of Decision Field Theory

Decision Field Theory is a relatively slow model to run. Table 6 shows the runtimes for datasets SP-1 and SP-2. The runtimes are normalised relative to the runtime for MNL for SP-1 and SP-2.

TABLE 6 : Relative runtimes of models for SP-1 and SP-2

Model	SP-1 Runtimes (normalised)	SP-2 Runtimes (normalised)
MNL	1.00	1.00
RRM	-	1.58
μ-RRM	-	2.11
DFT-2017	**200.08**	**174.78**
DFT-2014	**318.85**	**138.34**

Whilst DFT-2017 takes longer to run than typical choice models, it is quicker than mixed RRM (see appendix B). Using DFT-2014, with the number of timesteps set to infinity, reduces the runtime for SP-2 but increases it for SP-1. Runtimes for Mixed Decision Field Theory models (see 4.5), which are estimated by R package RSGHB [Dumont et al., 2014], vary vastly depending on the number of iterations set by the coder. A low number of iterations can be used initially to get an approximation of how well a model will work before running a more time-consuming model with more iterations.

4.5 Incorporating Heterogeneity

Using socio-demographic variables in Decision Field Theory

One strength of RUM models is that they are good at using the input of socio-demographic variables to improve model accuracy. As far as we are aware, these factors have never been incorporated into DFT. This idea is explored in this section.

Firstly, we explore the impact of income on the weight parameter for travel cost, w_{TC} for SP-1. Whilst attribute parameters for MNL are independent of each other, this is not the case for DFT weight parameters, as together they sum to 1. We hence define $w_{TC} = w_{TC} + x$, with the other parameters adjusted to $w_i = w_i - x \times \frac{w_i}{1-w_{TC}}$, where x is defined as $x = p_{HI} \times HI$, HI is the household income and p_{HI} is a new parameter defining the strength of the impact of income. For our MNL model, income is included in the utility functions by using $p_{HI} \times HI \times tc$, where tc is the travel cost of the alternative. Table 7 shows the results of including the income parameter on each of the standard models.

TABLE 7 : Log-likelihood values for models with and without income/group difference parameter (using SP-1 and SD-C)

SP-1	Basic model	With income parameter
MNL	-1,667.97	-1,653.61
DFT	-1,595.85	-1,592.35
SD-C	Basic model	With group parameter
MNL	-4,633.90	-4,509.65
DFT	-4,633.63	-4,527.56

Whilst there is some improvement in the result for the DFT model in SP-1, it is not as large as the improvement in the MNL model. This however does not appear to have a significant impact on the differences between MNL and DFT probabilities of chosen alternatives (see figure 6 in appendix C).

We also explore the impact in a deliberately manipulated simulated dataset. Using simulated dataset C, we look at the improvements under MNL and under DFT by including a parameter to control which group the decision belongs to. As before for DFT, we add a factor x to the weight for seating/standing, subtracting this amount proportionally from the other weights. We set $x = p_G$ for group 1, and $x = 0$ for group 2. For MNL, we add $p_G \times S$ onto the utility for both alternatives for group 1, where S is equal to 1 or 0 depending on whether seating is available. Table 7 also shows the results of including a parameter for this group difference on each of the standard models. Once again, it appears that whilst DFT improves with the inclusion of this socio-demographic variable, MNL improves more significantly.

Adjusting psychological parameters in Decision Field Theory

We can additionally make small changes to the psychological parameters, ϕ_1 and ϕ_2, the sensitivity and memory measures, to attempt an improvement in fit in a DFT model. For example, we can adjust the memory parameter depending on how many choice tasks the decision-maker has already completed. We add a new variable, ϕ_3, such that our memory parameter is now $\phi_2 + \phi_3 \times n$, where n is the task number. Alternatively, the sensitivity parameter can be similarly adjusted to be $\phi_1 + \phi_3 \times n$. Results from both of these adjustments are in Table 8.

TABLE 8 : Log-likelihood values for models with and without adjustments to the memory and sensitivity parameters

Model	Swiss (SP-1)	UK (SP-2)
DFT	-1,595.85	-3,598.87
ϕ_1 adjusted	-1,595.13	-3,592.15
ϕ_2 adjusted	-1,595.67	-3,594.78

Whilst the adjustments make very little difference for SP-1, there is a significant effect for SP-2, as only one parameter has been added. This suggests that improving the flexibility of the psychological parameters has the potential to allow for a changing neurological state in the decision-maker.

Adding heterogeneity: Mixed DFT

Our final effort to add heterogeneity to a DFT model is to use random parameters. For both DFT with and without a parameter for the number of timesteps, a significant improvement in fit is found (Table 9). For weight parameters, which cannot be less than zero, we use truncated normal distributions. We trial both normal and truncated normal distributions for the remaining parameters. These are then compared against DFT models with fixed parameters as well as against a DFT model without a time parameter with truncated normal distributions for all parameters. All mixed models are estimated using Dumont et al. [2014]'s R package 'RSGBH'.

TABLE 9 : Log-likelihoods of Mixed Decision Field Theory Models

Model	Time parameter?	Weight parameters	Other parameters	Swiss (SP-1)			UK (SP-2)		
				pars	LL	BIC	pars	LL	BIC
1	yes	fixed	fixed	7	-1,595.85	3,249	8	-3,598.87	7,263
2	no	fixed	fixed	6	-1,595.88	3,241	7	-3,676.34	7,410
3	yes	truncated	normal	14	-1,450.39	3,015	16	-3,156.27	6,444
4	yes	truncated	truncated	14	-1,438.39	2,991	16	-3,140.09	6,412
5	no	truncated	truncated	12	-1,430.41	2,959	14	-3,190.23	6,495

For both datasets, vast gains are made by using random parameters. A better fit is found if truncated normal distributions are used for all parameters rather than just the weight parameters. Using random parameters in Berkowitsch et al. [2014]'s version of DFT results in a lower BIC value for the SP-1 but a much higher one for SP-2. Whilst we do not run mixed multinomial logit or mixed random regret models, Hess et al. [2016] do run these models on the same UK dataset (see appendix B for a table of results). They find that their best fitting model is a mixed μ-RRM model, which achieves a log-likelihood of $-3,174.96$ with a BIC of $6,456.66$. Whilst Mixed DFT has a better fit here, more rigorous trials and replications would be required to test the models against each other fairly.

4.6 Predictive capabilities of Decision Field Theory

As discussed in section 1, previous researchers have only compared the goodness-of-fit of DFT, and its performance in the context of forecasting has not been tested before. We have looked at the predictive capabilities of DFT on both of our route choice stated preference datasets, SP-1 and SP-2. We adopt the method used by Frejinger and Bierlaire [2007], using 80% of the data drawn randomly 5 times for estimation. These estimates are used to calculate probabilities for different choice outcomes for the remaining 20%, and a choice is then assigned probabilistically. We then obtain likelihoods for the forecasted decisions being observed. The results for SP-1 and SP-2 are displayed in Tables 10 and 11, respectively.

Table 12 shows that the likelihood ratio tests obtained comparing DFT against MNL for SP-1 indicate that the DFT model is significantly better for both estimation and forecasting. Whilst DFT is significantly better for forecasting results, the differences are more extreme for estimation results, where DFT produces much higher log-likelihood values. From Table 10 we can see that DFT achieves higher log-likelihood values than μ-RRM in all estimated and forecasted datasets. Likelihood ratio tests show that DFT produces significantly better results than MNL and RRM (which both have one less parameter) too for both estimation and forecasting (see Table 13). Whilst RRM achieves results closer to DFT than MNL does, it is still significantly worse than DFT for all datasets. These results suggest that DFT is an appropriate model for both estimation and forecasting.

TABLE 10 : Log-likelihoods for the estimated and forecasted datasets for MNL and DFT (using SP-1)

Model	Swiss (SP-1)		Dataset 1	Dataset 2	Dataset 3	Dataset 4	Dataset 5
MNL 5 parameters	Estimated 80%	Final LL	-1,356.52	-1,324.59	-1,338.47	-1,333.51	-1,365.84
		$\overline{\rho}^2$	0.298	0.315	0.308	0.310	0.294
	Forecasted 20%	Final LL	-308.25	-319.612	-322.409	-326.311	-342.081
		$\overline{\rho}^2$	0.355	0.332	0.326	0.318	0.286
	Full 100%	Final LL	-1,664.77	-1,644.2	-1,660.88	-1,659.82	-1,707.92
		$\overline{\rho}^2$	0.312	0.320	0.314	0.314	0.294
DFT 7 parameters	Estimated 80%	Final LL	-1,305.57	-1,266.23	-1,281.54	-1,266.97	-1,314.87
		$\overline{\rho}^2$	0.324	0.344	0.336	0.344	0.319
	Forecasted 20%	Final LL	-296.068	-294.684	-290.9	-294.191	-327.112
		$\overline{\rho}^2$	0.376	0.379	0.387	0.380	0.312
	Full 100%	Final LL	-1,601.64	-1,560.91	-1,572.44	-1,561.16	-1,641.98
		$\overline{\rho}^2$	0.337	0.354	0.349	0.354	0.320

TABLE 11 : Log-likelihoods for the estimated and forecasted datasets for MNL, DFT, RRM and μ-RRM (using SP-2)

Model	UK (SP-2)		Dataset 1	Dataset 2	Dataset 3	Dataset 4	Dataset 5
MNL 7 parameters	Estimated 80%	Final LL $\overline{\rho}^2$	-2,992.72 0.073	-2,971.12 0.079	-2,972.05 0.079	-2,986.28 0.075	-2,944.45 0.087
	Forecasted 20%	Final LL $\overline{\rho}^2$	-761.76 0.049	-748.97 0.065	-743.77 0.071	-752.80 0.060	-720.37 0.100
	Full 100%	Final LL $\overline{\rho}^2$	-3,754.48 0.070	-3,720.06 0.078	-3,715.82 0.079	-3,739.08 0.073	-3,664.81 0.092
DFT 8 parameters	Estimated 80%	Final LL $\overline{\rho}^2$	-2,888.03 0.105	-2,873.45 0.109	-2,868.09 0.111	-2,876.55 0.108	-2,876.69 0.108
	Forecasted 20%	Final LL $\overline{\rho}^2$	-704.46 0.119	-717.87 0.102	-702.70 0.121	-704.55 0.119	-681.86 0.147
	Full 100%	Final LL $\overline{\rho}^2$	-3,592.53 0.109	-3,605.89 0.106	-3,570.78 0.115	-3,581.11 0.112	-3,569.88 0.115
RRM 7 parameters	Estimated 80%	Final LL $\overline{\rho}^2$	-2,976.93 0.077	-2,958.28 0.083	-2,955.93 0.084	-2,969.49 0.080	-2,928.01 0.093
	Forecasted 20%	Final LL $\overline{\rho}^2$	-754.66 0.058	-744.29 0.071	-739.96 0.076	-747.77 0.067	-715.22 0.107
	Full 100%	Final LL $\overline{\rho}^2$	-3,731.59 0.075	-3,702.57 0.082	-3,695.89 0.084	-3,717.27 0.079	-3,643.23 0.097
μ-RRM 8 parameters	Estimated 80%	Final LL $\overline{\rho}^2$	-2,976.74 0.077	-2,958.28 0.083	-2,955.63 0.084	-2,969.07 0.080	-2,927.88 0.092
	Forecasted 20%	Final LL $\overline{\rho}^2$	-753.91 0.058	-746.28 0.067	-740.83 0.074	-750.04 0.062	-715.08 0.106
	Full 100%	Final LL $\overline{\rho}^2$	-3,730.65 0.075	-3,704.56 0.082	-3,696.46 0.084	-3,719.11 0.078	-3,642.96 0.097

TABLE 12 : Likelihood Ratio Tests for the estimated and forecasted results of DFT against MNL (using SP-1)

Swiss (SP-1)		MNL/DFT	
		Estimated	Forecast
Dataset 1	T-stat p-value	101.90 3.7E-23***	24.36 2.6E-10***
Dataset 2	T-stat p-value	116.72 2.3E-26***	49.86 7.5E-12***
Dataset 3	T-stat p-value	113.86 9.4E-26***	63.02 1.0E-14***
Dataset 4	T-stat p-value	133.08 6.3E-30***	64.24 5.6E-15***
Dataset 5	T-stat p-value	101.94 3.7E-23***	29.94 1.6E-07***
Signif. codes: ***<0.001 **<0.01 *<0.05 .<0.1			

TABLE 13 : Likelihood Ratio Tests for the estimated and forecasted results of DFT against MNL and RRM (using SP-2)

UK (SP-2)		MNL/DFT		RRM/DFT		μ-RRM/DFT	
		Estimated	Forecast	Estimated	Forecast	Estimated	Forecast
Dataset 1	T-stat p-value	209.39 9.4E-48***	114.60 4.9E-27***	177.81 7.3E-41***	100.40 6.3E-24***	177.43 8.9E-41***	98.90 1.3E-23***
Dataset 2	T-stat p-value	195.34 1.1E-44***	62.21 1.6E-15***	169.65 4.4E-39***	52.84 1.8E-13***	169.65 4.4E-39***	56.82 2.4E-14***
Dataset 3	T-stat p-value	207.92 2.0E-47***	82.15 6.4E-20***	175.70 2.1E-40***	74.52 3.0E-18***	175.09 2.9E-40***	76.26 1.3E-18***
Dataset 4	T-stat p-value	219.45 6.0E-50***	96.49 4.5E-23***	185.88 1.3E-42***	86.44 7.3E-21***	185.03 1.9E-42***	90.98 7.3E-22***
Dataset 5	T-stat p-value	135.51 1.3E-31***	77.02 8.6E-19***	102.64 2.0E-24***	66.73 1.6E-16***	102.38 2.3E-24***	66.45 1.8E-16***
Signif. codes: ***<0.001 **<0.01 *<0.05 .<0.1							

5 CONCLUSIONS

This paper provides theoretical improvements to further the mechanisms behind DFT to better represent general decision making, as well as rigorously comparing DFT against traditional choice models. We also consider multiple mechanisms for incorporating heterogeneity within and across decision-makers within a DFT model.

Prior to our work, there was one comparison between DFT and mainstream choice models [Berkowitsch et al., 2014]. In this paper, we provide further evidence that DFT can be a competitive rival to traditional choice models. Perhaps most significantly, DFT achieves a better model fit than an MNL model in both of our stated preference datasets. DFT also outperforms RRM in SP-2. We demonstrate, for the first time, that DFT can step away from being a 'one-size-fits-all' model and incorporate heterogeneity in a number of different approaches. Whilst only small gains are made with the incorporation of socio-demographic variables into the weight parameters, a vastly significant gain is found with 'Mixed DFT', where the parameters are random with either normal or truncated normal distributions. Large gains are also made with the inclusion of additional weights or parameters to deal with underlying preferences towards alternatives.

Whilst we have made a brief start on the inclusion of socio-demographic variables, future work could explore this much further. For example, it could be possible that income effects are in part captured by the deliberation process in a DFT model and tests could be done to see if there is a relationship between income and any of the psychological parameters in DFT[4]. Additionally, the importance of the psychological parameters needs to be tested. It remains to be seen whether the sensitivity parameter is as efficient at capturing correlation across alternatives as the structural parameter in a nested logit model. It could also be that the psychological parameters are more important in risky choice, but the weight parameters are more important in riskless choice. As we have only tested riskless choice datasets here it could be that socio-demographic effects are easier to find in risky choice. Age, gender and personality have all been found to have an impact on risk-taking behaviour [Lauriola and Levin, 2001, Harris et al., 2006, Mata et al., 2011] and therefore it could be possible that some of these effects are captured in DFT's psychological parameters.

Section 4.4 suggests that the differences between using MNL and DFT to explain an individual's decisions are not vast. However, it would appear that DFT provides slightly more extreme predictions, with more predictable individuals being better explained by DFT in comparison to MNL, and the converse being true for more random individuals. This is perhaps surprising given DFT was originally used mostly for risky choices. Future work could look at whether DFT differentiates more or less than traditional choice models on different kinds of datasets. DFT could also easily be incorporated into one or more of the classes in a latent class structure and thus we could see if individual decision makers are better explained by DFT or another model directly.

This paper provides a method for calculating the probability of alternatives under a decision field theory model whilst simultaneously avoiding computationally intensive simulation and not setting the decision time for decision-makers to infinity. Using this method provides a number of benefits compared to using Berkowitsch et al. [2014]'s method. Whilst our new method provides a better fit

[4] We thank an anonymous reviewer for this suggestion

for only one of the two stated preference datasets we apply it on, it provides a vastly greater amount of flexibility. Firstly, the response time to make a decision can now be simply incorporated into the model: the number of timesteps could vary proportionally to the time taken by the decision-maker[5]. Secondly, the memory parameter can now be negative, reflecting that preferences can inflate as well as deteriorate over time [Mather et al., 2000]. Finally, and perhaps most crucially, initial preferences can play an important role in our version of DFT. This means that, for example, DFT should easily be able to explain a status quo bias. In conclusion, it appears that the restoration of a time parameter in DFT results in a far more realistic psychological model for understanding choice behaviour.

The implementation of Decision Field Theory does not come easily. A standard DFT model takes up to 200 times longer to run than a Multinomial Logit model. Future efforts should look at reducing this runtime as well as removing the scale-variant nature of DFT, as currently we need to know if an attribute is desirable or not before we can incorporate the importance of the attribute into a DFT model. We have shown that a DFT model can be specified to include underlying preferences through the initial preference matrices, while additional weight parameters can be included to capture socio-demographic effects, for example. The outputs from the model provide rich insights into behaviour, but it is clear that traditional measures such as the value of time cannot be obtained from a DFT model. This, however, is typical for departures from RUM, with Dekker [2014] highlighting the difficulties of using value of time measures obtained from Random Regret Minimisation. As is the case with any other departures from RUM, a user thus needs to carefully make a decision of whether the increased behavioural richness of the model and the improvement in fit (both in estimation and in forecasting) is more important than an ability to produce measures for welfare analysis.

In addition, the underlying psychological foundations of DFT may be more suited than a purely microeconomic model at incorporating the increasing number of behavioural and processing indicators that are becoming available. For example, electroencephalogram (EEG) recordings and eye-tracking have already been used to understand and predict choices [Khushaba et al., 2013, Telpaz et al., 2015, Uggeldahl et al., 2016]. Given that we have relaxed the assumption on the number of timesteps for a DFT model, we could now test DFT on a dynamic revealed preference dataset where the attribute levels of an alternative change over time. Response times, EEG, deliberation times or eye-tracking information could be incorporated into a DFT model, to lessen the requirement of estimation for the number of deliberation timesteps at each point as the choice set changes. Testing DFT on such a dataset would enable us to look at the validity of underlying behavioural assumptions of a DFT model. Additional information could also potentially be used to determine whether a decision-maker has come to an internal or external threshold under a DFT model when making a decision. This would allow for a DFT model to predict a decision-maker's level of confidence or uncertainty in their choice. For example, eye-tracking information showing which attributes are considered last could inform how likely a decision-maker is to come to a conclusion through satisficing, when an alternative reaches a certain preference value. Overall, this means that there is much need for further research into DFT, which with its good results for both estimation and forecasting, appears to otherwise be a promising future model for the choice modelling community.

[5] See also Hancock et al. [2018] for examples of how response time can be incorporated into a DFT model

ACKNOWLEDGEMENTS

The authors acknowledge the financial support by the European Research Council through the consolidator grant 615596-DECISIONS. The authors would also like to thank Caspar Chorus, Jerome Busemeyer and three anonymous reviewers for their comments on previous versions of this paper.

REFERENCES

Axhausen, K. W., Hess, S., König, A., Abay, G., Bates, J. J., and Bierlaire, M. (2008). Income and distance elasticities of values of travel time savings: New Swiss results. *Transport Policy*, 15(3):173–185.

Batley, R. and Hess, S. (2016). Testing for regularity and stochastic transitivity using the structural parameter of nested logit. *Transportation Research Part B: Methodological*, 93:355–376.

Ben-Akiva, M. and Bierlaire, M. (1999). Discrete choice methods and their applications to short term travel decisions. In *Handbook of transportation science*, pages 5–33. Springer.

Berkowitsch, N. A., Scheibehenne, B., and Rieskamp, J. (2014). Rigorously testing multialternative decision field theory against random utility models. *Journal of Experimental Psychology: General*, 143(3):1331.

Berkowitsch, N. A., Scheibehenne, B., Rieskamp, J., and Matthäus, M. (2015). A generalized distance function for preferential choices. *British Journal of Mathematical and Statistical Psychology*, 68(2):310–325.

Busemeyer, J. R. and Diederich, A. (2002). Survey of decision field theory. *Mathematical Social Sciences*, 43(3):345–370.

Busemeyer, J. R., Jessup, R. K., Johnson, J. G., and Townsend, J. T. (2006). Building bridges between neural models and complex decision making behaviour. *Neural Networks*, 19(8):1047–1058.

Busemeyer, J. R. and Townsend, J. T. (1992). Fundamental derivations from decision field theory. *Mathematical Social Sciences*, 23(3):255–282.

Busemeyer, J. R. and Townsend, J. T. (1993). Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. *Psychological review*, 100(3):432.

Chorus, C. G. (2010). A new model of random regret minimization. *EJTIR, 10 (2), 2010*.

Chorus, C. G., Arentze, T. A., and Timmermans, H. J. (2008). A random regret-minimization model of travel choice. *Transportation Research Part B: Methodological*, 42(1):1–18.

Christianson, S.-Å. (1992). Emotional stress and eyewitness memory: a critical review. *Psychological bulletin*, 112(2):284.

Daly, A. and Zachary, S. (1978). Improved multiple choice models. *Determinants of travel choice*, 335:357.

Daneman, M. and Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of verbal learning and verbal behavior*, 19(4):450–466.

De Waal, F. B. and Ferrari, P. F. (2010). Towards a bottom-up perspective on animal and human cognition. *Trends in cognitive sciences*, 14(5):201–207.

Dekker, T. (2014). Indifference based value of time measures for random regret minimisation models. *Journal of choice modelling*, 12:10–20.

Diederich, A. (1997). Dynamic stochastic models for decision making under time constraints. *Journal of Mathematical Psychology*, 41(3):260–274.

Dror, I. E., Basola, B., and Busemeyer, J. R. (1999). Decision making under time pressure: An independent test of sequential sampling models. *Memory & cognition*, 27(4):713–725.

Dumont, J., Keller, J., Carpenter, C., and Dumont, M. J. (2014). Package 'rsghb'.

Flin, R., Boon, J., Knox, A., and Bull, R. (1992). The effect of a five-month delay on children's and adults' eyewitness memory. *British Journal of Psychology*, 83(3):323–336.

Frejinger, E. and Bierlaire, M. (2007). Capturing correlation with subnetworks in route choice models. *Transportation Research Part B: Methodological*, 41(3):363–378.

Genz, A. (1992). Numerical computation of multivariate normal probabilities. *Journal of computational and graphical statistics*, 1(2):141–149.

Gold, J. I. and Shadlen, M. N. (2000). Representation of a perceptual decision in developing oculomotor commands. *Nature*, 404(6776):390–394.

Gonzalez-Vallejo, C. (2002). Making trade-offs: A probabilistic and context-sensitive model of choice behavior. *Psychological Review*, 109(1):137.

Guevara, C. A. and Fukushi, M. (2016). Modeling the decoy effect with context-rum models: Diagrammatic analysis and empirical evidence from route choice sp and mode choice rp case studies. *Transportation Research Part B: Methodological*, 93:318–337.

Hancock, T. O., Hess, S., and Choudhury, C. F. (2018). Incorporating response time in a decision field theory model. The Transportation Research Board (TRB) 97th Annual Meeting.

Harris, C. R., Jenkins, M., and Glaser, D. (2006). Gender differences in risk assessment: Why do women take fewer risks than men? *Judgment and Decision Making*, 1(1):48.

Henningsen, A. and Toomet, O. (2011). maxlik: A package for maximum likelihood estimation in r. *Computational Statistics*, 26(3):443–458.

Hess, S., Beck, M., and Crastes dit Sourd, R. (2016). Can a better model specification avoid the need to move away from random utility maximisation? Transportation Research Board (TRB) 96th Annual Meeting.

Hess, S. and Stathopoulos, A. (2013). A mixed random utility - random regret model linking the choice of decision rule to latent character traits. *Journal of choice modelling*, 9:27–38.

Hey, J. D., Lotito, G., and Maffioletti, A. (2010). The descriptive and predictive adequacy of theories of decision making under uncertainty/ambiguity. *Journal of risk and uncertainty*, 41(2):81–111.

Hotaling, J. M., Busemeyer, J. R., and Li, J. (2010). Theoretical developments in decision field theory: comment on Tsetsos, Usher, and Chater (2010). *Psychological review*, 117(4):1294–1298.

Johnson, J. G. (2006). Cognitive modeling of decision making in sports. *Psychology of Sport and Exercise*, 7(6):631–652.

Khushaba, R. N., Wise, C., Kodagoda, S., Louviere, J., Kahn, B. E., and Townsend, C. (2013). Consumer neuroscience: Assessing the brain response to marketing stimuli using electroencephalogram (eeg) and eye tracking. *Expert Systems with Applications*, 40(9):3803–3812.

Krajbich, I. and Dean, M. (2015). How can neuroscience inform economics? *Current Opinion in Behavioral Sciences*, 5:51–57.

Krosnick, J. A., Narayan, S., and Smith, W. R. (1996). Satisficing in surveys: Initial evidence. *New directions for evaluation*, 1996(70):29–44.

Lauriola, M. and Levin, I. P. (2001). Personality traits and risky decision-making in a controlled experimental task: An exploratory study. *Personality and Individual Differences*, 31(2):215–226.

Liew, S. X., Howe, P. D., and Little, D. R. (2016). The appropriacy of averaging in the study of context effects. *Psychonomic bulletin & review*, pages 1–8.

Mata, R., Josef, A. K., Samanez-Larkin, G. R., and Hertwig, R. (2011). Age differences in risky choice: A meta-analysis. *Annals of the New York Academy of Sciences*, 1235(1):18–29.

Mather, M. (2006). Why memories may become more positive as people age. *Memory and emotion: Interdisciplinary perspectives*, pages 135–158.

Mather, M., Shafir, E., and Johnson, M. K. (2000). Misremembrance of options past: Source monitoring and choice. *Psychological Science*, 11(2):132–138.

McFadden, D. (1978). Modeling the choice of residential location. *Transportation Research Record*, (673).

McFadden, D. (2000). Disaggregate behavioral travel demand's rum side. *Travel Behaviour Research*, pages 17–63.

McFadden, D. and Train, K. (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics*, pages 447–470.

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

Otter, T., Johnson, J., Rieskamp, J., Allenby, G. M., Brazell, J. D., Diederich, A., Hutchinson, J. W., MacEachern, S., Ruan, S., and Townsend, J. (2008). Sequential sampling models of choice: Some recent advances. *Marketing letters*, 19(3-4):255–267.

Parvaneh, Z., Arentze, T., and Timmermans, H. (2012). Understanding travelers' behavior in provision of travel information: a bayesian belief approach. *Procedia-Social and Behavioral Sciences*, 54:251–260.

Qin, H., Guan, H., and Wu, Y.-J. (2013). Analysis of park-and-ride decision behavior based on decision field theory. *Transportation research part F: traffic psychology and behaviour*, 18:199–212.

Raab, M. and Johnson, J. G. (2004). Individual differences of action orientation for risk taking in sports. *Research quarterly for exercise and sport*, 75(3):326–336.

Ratcliff, R., Cherian, A., and Segraves, M. (2003). A comparison of macaque behavior and superior colliculus neuronal activity to predictions from models of two-choice decisions. *Journal of neurophysiology*, 90(3):1392–1407.

Roe, R. M., Busemeyer, J. R., and Townsend, J. T. (2001). Multialternative decision field theory: A dynamic connectionst model of decision making. *Psychological review*, 108(2):370.

Schall, J. D. (2003). Neural correlates of decision processes: neural and mental chronometry. *Current opinion in neurobiology*, 13(2):182–186.

Scheibehenne, B., Rieskamp, J., and González-Vallejo, C. (2009). Cognitive models of choice: Comparing decision field theory to the proportional difference model. *Cognitive science*, 33(5):911–939.

Schwartz, B., Ward, A., Monterosso, J., Lyubomirsky, S., White, K., and Lehman, D. R. (2002). Maximizing versus satisficing: happiness is a matter of choice. *Journal of personality and social psychology*, 83(5):1178.

Simon, H. A. (1957). Models of man; social and rational.

Stern, E. (1999). Reactions to congestion under time pressure. *Transportation Research Part C: Emerging Technologies*, 7(2):75–90.

Stern, E. and Portugali, J. (1999). Environmental cognition and decision making in urban navigation. *Wayfinding Behavior: Cognitive Mapping and Other Spatial Processes. John Hopkins Press, Baltimore, MD*, pages 99–118.

Stern, E. and Richardson, H. W. (2005). Behavioural modelling of road users: current research and future needs. *Transport Reviews*, 25(2):159–180.

Telpaz, A., Webb, R., and Levy, D. J. (2015). Using eeg to predict consumers' future choices. *Journal of Marketing Research*, 52(4):511–529.

Trueblood, J. S., Brown, S. D., and Heathcote, A. (2014). The multiattribute linear ballistic accumulator model of context effects in multialternative choice. *Psychological review*, 121(2):179.

Trueblood, J. S., Brown, S. D., Heathcote, A., and Busemeyer, J. R. (2013). Not just for consumers context effects are fundamental to decision making. *Psychological science*, 24(6):901–908.

Tsetsos, K., Usher, M., and Chater, N. (2010). Preference reversal in multiattribute choice. *Psychological review*, 117(4):1275.

Uggeldahl, K., Jacobsen, C., Lundhede, T. H., and Olsen, S. B. (2016). Choice certainty in discrete choice experiments: Will eye tracking provide useful measures? *Journal of choice modelling*, 20:35–48.

van Cranenburgh, S., Guevara, C. A., and Chorus, C. G. (2015). New insights on random regret minimization models. *Transportation Research Part A: Policy and Practice*, 74:91–109.

Wierzbicki, A. P. (1982). A mathematical basis for satisficing decision making. *Mathematical modelling*, 3(5):391–405.

Williams, H. C. (1977). On the formation of travel demand models and economic evaluation measures of user benefit. *Environment and planning A*, 9(3):285–344.

Zaragoza, M. S. and Lane, S. M. (1994). Source misattributions and the suggestibility of eyewitness memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(4):934.

Zhao, G., Wu, C., and Ou, B. (2011). Mathematical modeling of average driver speed control with the integration of queuing network-model human processor and rule-based decision field theory. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 55, pages 856–860. Sage Publications.

APPENDIX A: DFT MODEL ESTIMATES

TABLE 14 : Results for SP-1

Swiss (SP-1)										
Model	1		2		3		4		5	
time par.	yes		no		yes		yes		no	
weights	fixed		fixed		truncated		truncated		truncated	
other pars.	fixed		fixed		normal		truncated		truncated	
pars.	7		6		14		14		12	
LL	-1,595.85		-1,595.88		-1,450.39		-1,438.39		-1,430.41	
BIC	3,249		3,241		3,015		2,991		2,959	
	est	t-ratio	est	t-ratio	est	t-ratio	est	t-ratio	est	t-ratio
$\mu_{wt_{TT}}$	0.3475	45.31	0.3474	45.19	0.3406	8.69	0.3382	13.24	0.3360	12.42
$\mu_{wt_{TC}}$	0.4676	43.22	0.4676	43.26	0.5264	7.76	0.5202	11.14	0.5286	10.41
$\mu_{wt_{HW}}$	0.0747	12.85	0.0747	12.85	0.0598	12.29	0.0612	12.74	0.0597	13.02
μ_{ϕ_1}	142.604	67.97	130.480	62.18	3.5451	4.23	2.0106	7.46	184.649	4.94
μ_{ϕ_2}	0.1835	12.71	0.1834	12.67	-0.0986	-1.93	-0.0301	-0.74	0.0916	8.01
μ_{ε}	0.0017	0.96	0.0020	1.06	0.0001	0.19	0.0027	3.25	0.0015	6.32
μ_t	112.219	1,805.4	-		29.8485	7.40	29.3096	9.49	-	
$\sigma_{wt_{TT}}$	-		-		0.1875	3.57	0.2019	5.16	0.1472	3.30
$\sigma_{wt_{TC}}$	-		-		0.3894	3.45	0.4096	5.35	0.3836	4.23
$\sigma_{wt_{HW}}$	-		-		0.0393	6.42	0.0405	6.83	0.0356	7.40
σ_{ϕ_1}	-		-		1.0818	3.28	0.8760	3.95	41.2838	3.98
σ_{ϕ_2}	-		-		0.1720	2.57	0.0757	4.28	0.0954	4.65
σ_{ε}	-		-		0.0019	2.68	0.0013	2.81	0.0009	5.77
σ_t	-		-		9.3739	5.19	5.7655	3.20	-	

TABLE 15 : Results for SP-2

UK (SP-2)										
Model	1		2		3		4		5	
time par.	yes		no		yes		yes		no	
weights	fixed		fixed		truncated		truncated		truncated	
other pars.	fixed		fixed		normal		truncated		truncated	
pars.	8		7		16		16		14	
LL	-3,598.87		-3,676.34		-3,156.27		-3,140.09		-3,190.23	
BIC	7,263		7,410		6,444		6,412		6,495	
	est	t-ratio	est	t-ratio	est	t-ratio	est	t-ratio	est	t-ratio
$\mu_{wt_{TT}}$	0.1132	5.33	0.1350	11.63	0.1001	7.29	0.1082	7.46	0.0854	10.29
$\mu_{wt_{TC}}$	0.6786	12.72	0.6642	41.98	0.7889	9.53	0.7802	9.41	0.8205	8.85
$\mu_{wt_{CT}}$	0.0691	3.30	0.0816	10.72	0.0339	4.09	0.0320	2.78	0.3361	6.38
$\mu_{wt_{RD}}$	0.0271	3.97	0.0308	3.39	0.0142	2.88	0.0191	5.27	0.0122	2.09
μ_{ϕ_1}	0.1000	13.68	152.5165	102.80	0.0840	3.56	0.0742	5.19	0.1854	3.86
μ_{ϕ_2}	1.9381	180.09	0.4283	9.99	0.7900	12.09	0.8091	7.40	0.6418	4.61
μ_{ε}	0.0393	8.48	0.1543	5.31	0.0300	10.63	0.0448	5.06	0.0872	3.34
μ_t	3.7792	242.14	-		10.7987	14.41	10.1687	9.25	-	
$\sigma_{wt_{TT}}$	-		-		0.0623	4.21	0.0746	5.35	0.0575	5.81
$\sigma_{wt_{TC}}$	-		-		0.8821	9.63	0.7816	8.63	0.8988	6.50
$\sigma_{wt_{CT}}$	-		-		0.0616	3.12	0.0738	5.50	0.0311	6.54
$\sigma_{wt_{RD}}$	-		-		0.2486	4.35	0.0298	4.33	0.0179	5.31
σ_{ϕ_1}	-		-		0.0717	5.09	0.0273	4.80	0.0907	3.81
σ_{ϕ_2}	-		-		0.4196	10.26	0.3342	7.98	0.4683	4.91
σ_{ε}	-		-		0.0131	9.09	0.0527	4.12	0.0822	2.75
σ_t	-		-		3.7301	7.16	3.2715	5.85	-	

The weights used for SP-2 are travel time (TT), cost (TC), rate of crowded trips (CT), rate of delays (RD) and the average length of delays, where this final weight is fixed such that the weights together sum to one. The cost of a provision of a delay information service was found to be insignificant and therefore omitted.

APPENDIX B: MODELS RESULTS FROM HESS ET AL. [2016]

TABLE 16 : Mixed model results for SP-2 from Hess et al. [2016]

UK data:	MNL	RRM	mu-RRM	mixed MNL	mixed RRM	mixed mu-RRM
LL	-3,721.67	-3,699.49	-3,698.89	-3,184.89	-3,205.27	-3,174.96
parameters	7	7	8	12	12	13
BIC	7,500.81	7,456.46	7,463.47	6,468.30	6,509.08	6,456.66
Runtime (normalised)	1.00	1.58	2.11	50.75	316.98	335.69

APPENDIX C: THE IMPACT ON INCLUDING SOCIO-ECONOMIC FACTORS ON THE DIFFERENCES BETWEEN DFT AND MNL

Figure 6 shows the impact of including socio-economic factors (income effects) on the difference between MNL and DFT predicted probabilities of the chosen alternatives for SP-1. Grey points indicate decisions made by individuals who earn less than 80,000 Swiss Francs, whereas black points indicate decisions made by individuals who earn more. It appears that there are no significant differences between the models by including these income effects, despite the fact that the MNL model improved in model fit more significantly.

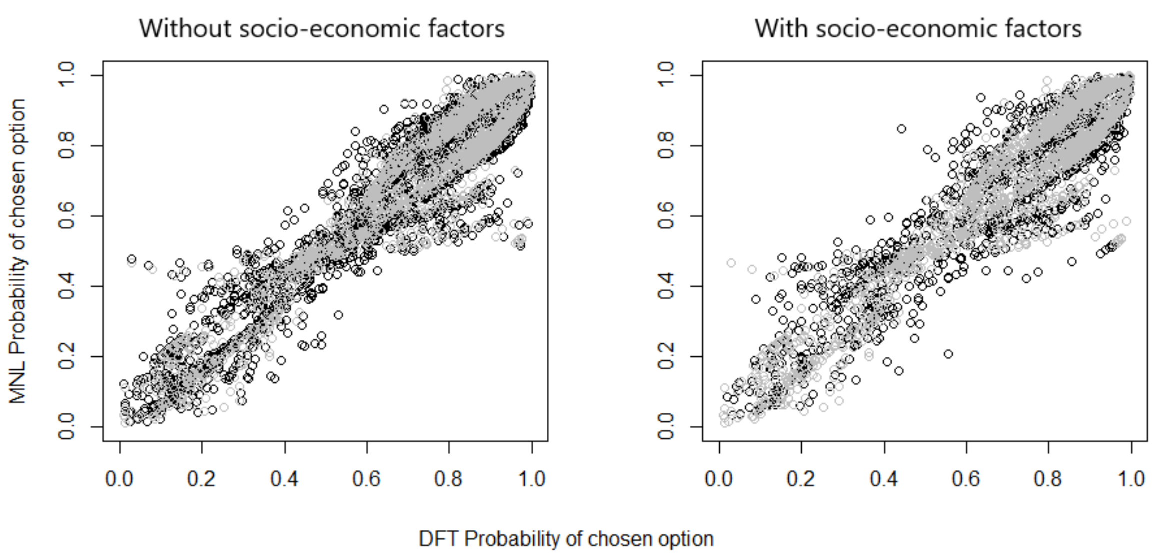

FIGURE 6 : Impact of including income on MNL and DFT models on dataset SP-1

APPENDIX D: NOTES ON DFT PARAMETERS

It should be noted that large differences in ϕ_1 may not have much impact on a decision. For example, suppose we have the following choice task:

	Attribute 1	Attribute 2	Attribute 3
Alternative A	3	4	5
Alternative B	2	4	6
Alternative C	3	7	1

If at some time point we had a preference vector of $P_t = [10,9,8]'$ and a value of 0.05 for ϕ_2, then the following results would be obtained for $S \times P_t$ for the given values of ϕ_1:

ϕ_1	$P_t[1]$	$P_t[2]$	$P_t[3]$
0.1	9.11	7.99	7.09
0.5	9.48	8.45	7.50
1	9.50	8.53	7.58
10	9.50	8.55	7.60

This means that the difference in preference between alternatives is not much impacted by ϕ_1. This could particularly be the case for choice scenarios involving only two alternatives, as shown by the minimal impact adjustments on ϕ_1 and also ϕ_2 had on SP-1 (see Table 8). Future work on DFT could look at the impact of removing these parameters altogether.

Additionally, we can also use this choice task to demonstrate how the timestep and error parameters, t and ε capture distinctly different features of the data. The table below gives the probability of choosing the three alternatives when $wt_1 = 0.3$, $wt_2 = 0.3$, $wt_3 = 0.4$, $\phi_1 = 0.1$ and $\phi_2 = 0.05$:

Probability of choosing alternatives				
t	10	20	10	20
ε	1	1	5	5
Alternative A	0.2807	0.2933	0.3449	0.3628
Alternative B	0.5265	0.5811	0.4721	0.5158
Alternative C	0.1928	0.1255	0.1830	0.1214

Under these conditions, the expected valence, $\mu = [0.3, 0.45, -0.75]'$. This means that with more timesteps, we would expect stronger preferences towards alternatives A and B. Higher values for the number of timesteps indicates that the decision-maker is more likely to consider all of the attributes. This results in the variance of the attribute weights having less impact. Higher values for the error variance ε result in the relative differences between attributes being less significant. (For example, $\varepsilon = \infty$ results in all alternatives being chosen with equal probability). Another way of considering these two parameters psychologically is that they are 'quality' and 'quantity' of information processed. The number of timesteps tells you how much of the information is considered (hence lower values imply less predictable choices) and the error variance tells you how 'distinct' the decision-maker interprets the alternatives (with high values meaning that the decision-maker interprets there being little difference between the alternatives).

Finally, this example can also be used to demonstrate that the scale-variant nature of DFT arises from the variance of the weights. If, for example, attribute 3 values were doubled, then to obtain an equivalent expected valence of $\mu = [0.3, 0.45, -0.75]'$, the weight for attribute 3 would need to be decreased relative to the weights for attributes 1 and 2. Weights of $wt_1 = 0.375$, $wt_2 = 0.375$ and $wt_3 = 0.25$ achieve an expected valence of $\mu = [0.375, 0.5625, -0.9375]'$, exactly 1.25 times the previous μ. However, this would result in a very different value for Ψ, as the variance of the weights has changed. Consequently the probabilities of alternatives would change, despite the relative expected valences remaining the same.