AN ACCUMULATION OF PREFERENCE: TWO ALTERNATIVE DYNAMIC MODELS FOR UNDERSTANDING TRANSPORT CHOICES

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

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

A. A. J. Marley
Department of Psychology, University of Victoria, Victoria, Canada.
Institute for Choice, University of South Australia, Adelaide, Australia.
ajmarley@uvic.ca

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

ABSTRACT

Interest in behavioural realism has gradually led to the introduction of alternatives to random utility models (RUMs) as a paradigm for representing choice behaviour, with notable interest, for example, in random regret minimisation (RRM). These more general models continue to rely on a framework where a single *value function* is calculated for each alternative in each choice setting, and the choice probabilities are calculated by comparing these value functions across alternatives. By contrast, research in mathematical psychology has used a more dynamic approach, where the preference value of each alternative updates over time in a given situation while the decision maker is deliberating about the choice to make. These *accumulator* models are well suited to accommodating a variety of context effects, and have been shown to give good performance for data collected in laboratory-based settings. The present paper considers two such accumulator models, namely decision field theory (DFT) and the multi-attribute linear ballistic accumulator (MLBA), and addresses limitations that have prevented their use in travel behaviour research. The methodological additions include the ability to capture the influence of socio-demographics, the presence of underlying preferences for specific alternatives, and/or the representation of attributes that have opposite effects on choice probabilities. We develop what we believe to be the first in-depth simultaneous comparison of DFT and MLBA with typical discrete choice models, and test both DFT and MLBA on a revealed preference dataset. We find that each model outperforms typical RUM and RRM implementations for both in-sample estimation and out-of-sample prediction, including in a large scale simulation experiment.

1. INTRODUCTION

Whilst mainstream choice modelling has firm economic foundations (McFadden, 1974), the modelling of decision-making behaviour in other fields has been characterised by very different aims and objectives. Since work in the 1970s (Tversky, 1972; Tversky and Kahneman, 1973; Tversky, 1977), the field of behavioural economics has considered choice from an economic viewpoint whilst simultaneously demonstrating that decision-makers are subject to biases, heuristics and context effects that result in choices that are not the most likely under traditional choice models. Choice modellers have long had an interest in increasing the behavioural realism of their models, with recent methodological advances aimed at incorporating alternative behavioural ideas such as random regret minimisation (RRM) (Chorus et al., 2008; Chorus, 2010), heuristics (Swait, 2001), satisficing (González-Valdés and Ortúzar, 2017) and the incorporation of information processing strategies such as attribute non-attendance (for a detailed review, see Hensher (2010)).

Moving away from the traditional random utility maximisation (RUM) framework however entails a number of disadvantages, most notably an inability to easily use the results in welfare analysis (Hess et al., 2018). This means that careful consideration is required before we move to alternative models. In this context, the question then arises whether, if we are willing to move away from RUM, we should move to models that are substantially different from it, rather than still staying within a logit framework as is the case for random regret minimisation. This observation leads us to look further afield in the present paper, and in particular at work in mathematical psychology, where researchers have developed mathematical and computational models that represent context effects such as attraction, compromise and similarity (Roe et al., 2001; Trueblood et al., 2013b; Noguchi and Stewart, 2014) as well as decision-making under time pressure (Busemeyer and Townsend, 1993).

Two models in particular have attracted substantial attention in mathematical psychology, namely Decision Field Theory (DFT) (Busemeyer and Townsend, 1992, 1993; Roe et al., 2001) and the multi-attribute linear ballistic accumulator (MLBA) (Trueblood et al., 2013a, 2014). These models also represent the choice between mutually exclusive alternatives but differ from more traditional discrete choice models in one specific dimension. RUM and RRM models are characterised by their utility and regret functions respectively, which are used to calculate a *single value function* for each alternative, where comparison of this across alternatives then leads to probabilities of a given alternative being chosen. This value function is calculated once per choice situation. On the other hand, DFT and MLBA are members of a broad family of *accumulator* models, where the *cumulative preference* value for an alternative in a single choice context is updated (linearly in MLBA, stochastically in DFT) over time.

Under DFT, the decision maker updates his/her preference for given alternatives by repeated comparisons between them, considering one attribute at a time, where the attribute values of the alternatives in that situation remain constant across these comparisons. Under MLBA, a 'drift rate' is generated at the outset for each alternative; this then drives the updating of the strength of preference of each alternative within the given choice context.

At first, this description of the models might seem similar to choice modelling work that has looked at preferences evolving over a sequence of choices, such as models incorporating value learning (McNair et al., 2012), state dependence (Bruno et al., 2015) or dynamic discrete choice models (Liu and Cirillo, 2018). However, *accumulator* models are structures for internal preference accumulation at the level of every single choice, not models that accumulate evidence over a sequence of choices. The accumulation models thus capture the mental deliberation from the time a particular choice is faced (or stated choice scenario is presented) to the point where the choice is made. The preferences are reset after that, so that the accumulation effect is not carried over to the next choice task, i.e. the accumulation made when considering choice set k does not affect choice set $k+1$ although such extensions are possible too if warranted by the choice context.

As is common with work in mathematical psychology, much of the focus of using DFT and MLBA has been on testing for the presence of specific behavioural phenomena in data collected in lab-based experiments, with little or no focus on prediction or on using data from complex choice scenarios. However, the way in which preferences evolve over time and their inherent ability to accommodate a range of what economists might call behavioural anomalies make these models potentially very appealing for studying more complex decisions, including travel behaviour. This statement is not completely without exceptions. For example, Hawkins et al. (2014) applied the linear ballistic accumulator (LBA, Brown and Heathcote 2008), a precursor to MLBA, to consumer attitudes and patient preferences; Hawkins et al. (2019) provided LBA applications to multi-period stated preference for consumer products; and Berkowitsch et al. (2014) applied DFT to consumer choices for products such as computers, cameras and racing bicycles. In key comparisons against *traditional* choice models, DFT in particular has been found to outperform random utility and random regret based models (e.g. Berkowitsch et al. 2014; Hancock et al. 2018).

Another distinction is that in mathematical psychology, accumulator models are typically used to jointly model the outcome of the choice process and the time it takes to reach that decision. In a traditional choice modelling context, it is the former that is of key interest, and this thus calls for further investigation of the potential advantages of these models in the context of data with choice outcomes only (as opposed to additional information on decision time), as well as investigating any mathematical adaptations required in such cases. In this context, it is worth noting that while some comparisons of MLBA and DFT have concluded that response time data is required to clearly differentiate the models (Evans et al., 2019b), applications within mathematical psychology typically do not study alternatives that are evaluated in qualitatively different ways by different participants and also tend to present tasks in which there are correct/incorrect answers. The work in this paper, by contrast, considers *subjective preferential choice* data, in which the complicated evaluation of alternatives makes choice-only data sufficiently rich to demonstrate differentiation of the predictive capabilities of traditional choice models and the accumulation models from mathematical psychology.

In a travel behaviour context, MLBA has thus far not been used, while the use of DFT to date has been limited due to the computational limitations (Otter et al., 2008). Our previous work on DFT has focused on methodological improvements that have made it possible to rigorously test DFT against typical choice models (Hancock et al., 2018). This motivates us to investigate the suitability of MLBA in modelling travel behaviour as well, as it has been found to outperform DFT in applications in the mathematical psychology literature (Trueblood et al., 2014; Cohen et al., 2017; Turner et al., 2018). Furthermore, there has been increasing attention in transport and choice modelling in general on best-worst datasets (Giergiczny et al., 2013; Rose, 2014; Hawkins et al., 2019) and research in mathematical psychology has shown that linear ballistic accumulators, with each alternative having a mean drift rate equal to an alternative-specific constant, perform well for these datasets (Hawkins et al., 2014).

Beyond simply comparing the two structures, we make a number of methodological improvements to both DFT and MLBA to facilitate their application to rich multi-alternative multi-attribute preferential choice datasets. The key contribution relates to allowing analysts to use DFT with attributes that have opposite effects on choice probabilities, and where this directionality is not known a priori. Previously, DFT models included 'attention weights' which could be used to capture the relative importance of attributes. As these weights must be positive (and sum to one), a priori knowledge is required as to whether an attribute has a positive (e.g. comfort of journey) or negative (e.g. cost) impact on the likelihood of an alternative being chosen. This is particularly an issue for consumer attributes which some decision-makers may like and others dislike, such as the size of a car; we handle such effects with attribute-specific scaling coefficients, meaning that such a priori knowledge is no longer required. We show that these coefficients can also be added to MLBA to capture the relative importance as well as the directionality of different attributes, a feature not typically accounted for in standard MLBA implementations. Further improvements include the ability to capture the influence of socio-demographics and the presence of underlying preferences for specific alternatives, in a manner equivalent to alternative specific constants in typical discrete choice models. We also look in detail at identification issues for both models, and present a number of empirical tests to help inform future applications.

In our empirical work, we offer what we believe to be the first in-depth simultaneous comparison of DFT and MLBA with typical discrete choice models, and also for the first time test both DFT and MLBA on a revealed preference dataset. We find that each model outperforms RUM and RRM implementations for both estimation and out-of-sample prediction across our datasets, and in a large scale simulation experiment.

The remainder of this paper is organised as follows. In the next section, we first provide an overview of the two models in their current form before presenting our various methodological improvements. This is followed by our empirical work on stated choice and revealed preference data, before some further tests on simulated data. The final section summarises the findings and presents some directions for future research.

2. METHODOLOGY: CONTRASTING AND IMPROVING MODELS FROM MATHEMATICAL PSYCHOLOGY

In this section, we first provide an introduction to accumulator models. We then present state-of-the-art implementations before making methodological improvements for both DFT and MLBA.

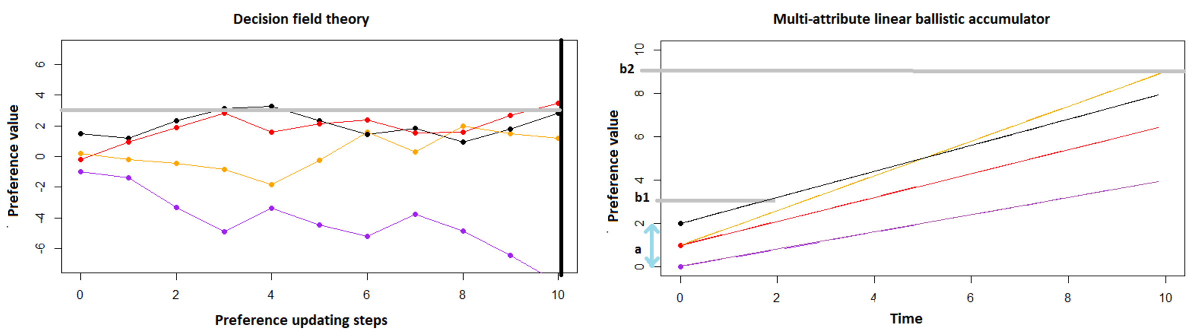
2.1. Introduction to accumulator models

Since the introduction of the drift diffusion model (Ratcliff, 1978), many different variations of accumulation models have been developed by mathematical psychologists (Busemeyer and Townsend, 1992; Usher and McClelland, 2001; Krajbich et al., 2012). Some of these accumulation models fall into the more specific category of sequential sampling models (see Busemeyer et al. (2019) for a review of these models). The idea of a sequential sampling model is that preferences for alternatives update over time depending on what information is being considered. An individual may consider, for example, cost, before then considering travel time. They might make comparisons across alternatives sequentially or randomly. By contrast, mainstream choice models such as random utility or random regret models construct just a single one-off preference or utility value for each alternative given a set of attribute values and then use those values, with error components, to calculate choice probabilities. Critically, sequential sampling models and more generally accumulation models instead assume that these preferences change over the course of the deliberation process whilst the decision-maker is choosing an alternative (even if the attributes of the alternatives stay the same). As already highlighted in the introduction, this preference accumulation is internal and happens at the level of every single choice, i.e. it is not an accumulation over a sequence of choices. Consequently, we can contrast the models with typical discrete choice structures.

Accumulator models aim to 'understand the motivational and cognitive mechanisms that guide the deliberation process involved in decisions' (Busemeyer and Townsend, 1993). The assumed mechanisms have subsequently been shown to resemble those in neural circuits. For example, Gold and Shadlen (2000) found that during a motion perception task, there was an accumulation of sensory evidence in the neural circuits of a monkey's brain, creating a behavioural response when the appropriate amount of information had been received. Furthermore, accumulator models have been developed that explain/describe context effects (Hotaling et al., 2010; Trueblood et al., 2014), capture risky choice behaviour (Busemeyer and Townsend, 1993; Stewart and Simpson, 2008) and can predict preference reversals (Diederich, 2003). Additionally, dynamic models provide a naturalistic method for the modelling of decision making in dynamic choice settings (Holmes et al., 2016).

One popular model from mathematical psychology that can easily be compared to traditional discrete choice models is decision field theory (DFT), first introduced by Busemeyer and Townsend (1992, 1993) and first operationalised in the context of travel behaviour by Hancock et al. (2018). In a DFT model, preference values for the alternatives update stochastically over time. At each moment, a single attribute is compared across alternatives and a valence (momentary preference) is added to the preference value for each alternative. At some point, the decision-maker comes to a conclusion, either as one of the alternatives reaches some threshold (similar to satisficing Kaufman, 1990; Schwartz et al., 2002; González-Valdés and Ortúzar, 2017) or as an external cue forces the decision-maker to make a choice, in which case the decision-maker chooses the alternative with the highest preference value at that moment. As an example, the left panel in Figure 1 demonstrates that different alternatives may be chosen depending on which threshold is used. The alternative that first reaches the internal threshold value is not the one that first reaches the time threshold. Here, it should be noted that the value that evolves over comparison is a *preference value*, rather than a probability, where the latter is calculated from the expectation of the former, a point we will return to later. The horizontal axis is measured in *preference updating steps*, which relate to the number of comparisons between alternatives, on each occasion using one attribute.

FIGURE 1 : An example decision process under both accumulation models



The linear ballistic accumulator model (Brown and Heathcote, 2008) and its multi-attribute version MLBA (Trueblood et al., 2013a, 2014) have a similar accumulation process for the preference of alternatives, but, in contrast with DFT, the updating is not stochastic. Instead, under LBA, decision-makers start with some random amount of initial 'evidence' for each alternative, and a random value of a drift rate, with the evidence in each 'accumulator' then increasing linearly with that sampled rate until one of the accumulators (alternatives) reaches a threshold. This sampled rate depends on the drift rate distributional and functional form, meaning that the preference values for each accumulator grow linearly at some drift rate dependent on the attribute levels of the alternative. A particular form for the drift rates (detailed in Section 2.3) give the MLBA. Depending on the level of the threshold, different alternatives may be chosen. This is demonstrated in the right panel of Figure 1, in which each alternative (accumulator) starts with some random initial value, assumed to be in an interval a, and a different alternative is chosen when the threshold value is b_1 than when it is b_2. The linear drift rates imply that, unlike in DFT, once the alternative with the largest (sampled) drift rate value 'gains the lead' there is no way for another alternative to recover and be chosen. Whilst this would not be the case with a non-linear specification, the current model is specifically linear to allow for simple calculation of the probabilities of alternatives. As with DFT, the value that evolves over time is a preference value, while the horizontal axis in Figure 1 now relates to actual time, given that no additional comparisons are made.

The mathematics underlying MLBA and DFT is vastly different. LBA was specifically designed such that it is 'simple' (Brown and Heathcote, 2008) and mathematically tractable, with MLBA subsequently developed such that it can also accurately capture and predict context effects. The simpler mathematical nature means that the probabilities of alternatives can easily be calculated from a combination of normal and uniform cumulative density functions (see Section 2.3 for a full description of MLBA).

It may be noted that there are numerous other accumulation models from mathematical psychology that seek to explain choice processes and predict choices. However, not all are currently suitable for transitioning into applied choice modelling. Given the complex nature of transport datasets, for example in terms of number of alternatives, even simple implementations will impose large computational costs. This is then further increased if analysts wish to add random heterogeneity in preferences, and models from mathematical psychology thus need to be efficient to run at a basic level if they are to compete. This means that models that do not have analytical solutions for calculating the choice probabilities will likely not be suitable options. For example, the leaky competing accumulator model (LCA, Usher and McClelland 2001) would require two levels of simulation in order to incorporate random parameters. Requirements for computer intensive simulation are also issues for the associative accumulation model (Bhatia, 2013) and the attentional drift diffusion model (Krajbich et al., 2012).

2.2. Decision field theory (DFT)

In this section, we first look at the existing implementation of DFT before making a number of methodological improvements and finally turning to identification issues.

2.2.1. Existing implementation

2.2.1.1 *Theoretical model*

Decision field theory, as an accumulator model, has preference values for each alternative that update over a number of preference updating steps[1] (see left panel in Figure 1). For decision-maker n in choice task (set) s, the vector of preferences after τ preference updating steps is denoted $\boldsymbol{P}_{ns,\tau}$ and is of size J_n where J_n is the number of alternatives in the choice set CS_{ns}. The preferences then update according to:

$$\boldsymbol{P}_{ns,\tau+1} = S_{ns} \cdot \boldsymbol{P}_{ns,\tau} + \boldsymbol{V}_{ns,\tau+1}, \tag{1}$$

where the previous values, $\boldsymbol{P}_{ns,\tau}$, are multiplied by a feedback matrix, S_{ns}, and a valence vector $\boldsymbol{V}_{ns,\tau+1}$ is added. It is worth noting here that DFT is often formulated as a linear difference equation with step size h (see, for example, Busemeyer and Diederich (2002)). We follow Roe et al. (2001) in explicitly using step sizes of $h = 1$. This means that analytical solutions for calculating the probability of choosing each alternative still exist, whilst avoiding the fact that different step sizes (values of h) can result in different outcomes (Evans et al., 2019b). The feedback matrix has two parameters that influence the existence and/or strength of attraction, similarity and compromise effects (Roe et al., 2001; Hotaling et al., 2010; Noguchi and Stewart, 2014), and is defined as:

$$S_{ns} = I_{ns} - \phi_2 \times \exp(-\phi_1 \times D_{ns}^2) \tag{2}$$

where ϕ_1 is a sensitivity parameter, ϕ_2 is a memory parameter, I_{ns} is the identity matrix of size $J_n \times J_n$ and D_{ns} is the distance between alternatives measured with respect to the attribute-levels. Whilst the relative importance of the different attributes can be taken into account with a psychological distance function (Hotaling et al., 2010) and work on new distance functions is possible (e.g. Berkowitsch et al., 2015), the Euclidean distance between the attributes can also be used for simplicity (Qin et al., 2013). The sensitivity parameter, ϕ_1, affects how much the alternatives compete with each other. Values very close to zero results in the distance between the attributes of

[1] Note that whilst it is possible that separate choice tasks could be linked through parameters controlling for learning effects, all of the DFT models in this paper assume that all choice tasks are entirely independent of each other to make them comparable with the RUM and RRM models without state-dependence.

alternatives becoming less important, whereas higher values result in more competition between similar alternatives. The memory parameter (also known as the decay parameter) determines the relative importance of attributes considered towards the end of the decision process relative to those considered at the start. A value of one results in zeros on the diagonals of the feedback matrix, which results in the preference already accumulated becoming irrelevant. As this value tends towards zero, the importance of the already accumulated preference increases.

DFT assumes that, at each preference updating step, the decision-maker compares a single attribute across all of the alternatives. This results in a random valence vector at step τ, $\boldsymbol{V}_{ns,\tau}$, which can be calculated as:

$$\boldsymbol{V}_{ns,\tau} = C_{ns} \cdot M_{ns} \cdot \boldsymbol{W}_{ns,\tau} + \boldsymbol{\varepsilon}_{ns,\tau} \tag{3}$$

where C_{ns} is a contrast matrix used to rescale the values such that they total zero, M_{ns} is the matrix of attribute values dependent on the specific choice task and $\boldsymbol{W}_{ns,\tau} = [0..1..0]'$ with the k^{th} entry, i.e. $\boldsymbol{W}_{ns,\tau,k} = 1$ if and only if attribute x_k is the attribute being attended to by the decision-maker at preference updating step τ. A DFT model thus typically estimates a weight, w_k, for the likelihood of attribute x_k being the single attribute attended to at a given step, where $\sum_k w_k = 1$. There is also a random error vector, $\boldsymbol{\varepsilon}_\tau = [\varepsilon_1..\varepsilon_{J_n}]'$, with $\varepsilon_i \sim N(0,s)$, distributed identically and independently across alternatives, steps, individuals and choice tasks. This allows for flexibility in the range of probability values that DFT predicts. This is in essence an error or noise parameter (Roe et al., 2001), for which higher values would be expected for more complex decision-making tasks (Hotaling et al., 2010).

2.2.1.2 *Calculating choice probabilities*

Under a DFT model, at the conclusion of the deliberation process, the alternative that is chosen is the one with the greatest preference value, regardless of whether the individual stopped deliberating due to a time threshold or due to the preference value for one of the alternatives reaching a preference threshold. Given that most choices do not have a strict time threshold, some applications of DFT calculate the probability for each alternative's preference value reaching a particular threshold first (see examples in Hotaling et al. (2010) and Turner et al. 2018). Given that this probability has no closed-form solution for more than two alternatives, we rely on Roe et al. (2001)'s method to calculate a related set of choice probabilities - namely, the probability for each alternative being chosen after τ preference accumulation steps. This uses the expected value and the covariance of the preference values ($\boldsymbol{\xi}_{ns,\tau}$ and $\Omega_{ns,\tau}$) and results in the stochastic variation being averaged out such that the choice probabilities can be calculated without the computationally heavy simulation that is required for DFT applications by, for example, Turner et al. (2018); Evans et al. (2019b).

To calculate the expected value of the preference values, we must first expand Equation 1, which results in:

$$\boldsymbol{P}_{ns,\tau} = \sum_{r=0}^{\tau-1} S_{ns}^{r} \cdot \boldsymbol{V}_{ns,\tau-r} + S_{ns}^{\tau} \cdot \boldsymbol{P}_{ns,0} \tag{4}$$

where $\boldsymbol{P}_{ns,0} = [\delta_{ns,1}, \delta_{ns,2}, .., \delta_{ns,J_n}]$ is the initial preference vector, with values $\delta_{ns,i} \in \mathbb{R}$. This is often assumed to be a vector of zeros (Busemeyer and Diederich, 2002) but can also be used to capture underlying preferences for different alternatives (Hancock et al., 2018), where $\delta_{ns,i} = \delta_i$, i.e. an estimated parameter that is fixed across individuals and choice tasks.

The attribute weights w_k are stationary, therefore $\boldsymbol{W}_{ns,\tau}$ can be considered a stationary stochastic process. This means that $\boldsymbol{V}_{ns,\tau}$ is also a stationary stochastic process with mean $E[\boldsymbol{V}_{ns,\tau}]$ and a variance covariance matrix given by $Cov[\boldsymbol{V}_{ns,\tau}]$. We let $\boldsymbol{\varepsilon}_{ns,\tau}$ vary according to a normal distribution with mean zero and variance σ_ε^2. Given that $\boldsymbol{\mu}_{ns} = E[\boldsymbol{V}_{ns,\tau}]$, it can be calculated as $\boldsymbol{\mu}_{ns} = C_{ns} \cdot M_{ns} \cdot \boldsymbol{w}$, where $\boldsymbol{w}$ is a vector containing the attribute attention weights, w_k, which corresponds to the probability of each of the attributes being considered. We also have $Cov[\boldsymbol{V}_{ns,\tau}] = \Phi_{ns} = C_{ns} \cdot M_{ns} \cdot \Psi_{ns} \cdot M'_{ns} \cdot C'_{ns} + \sigma_\varepsilon^2 \cdot I$, where $\Psi_{ns} = Cov[\boldsymbol{W}_{ns,\tau}]$ and I is the identity matrix (C_{ns} and M_{ns} are matrices of constants). We can then calculate the expected value and the covariance of $\boldsymbol{P}_{ns,\tau}$. With S_{ns} being a constant, $E[\boldsymbol{P}_{ns,\tau}]$ reduces to:

$$E[\boldsymbol{P}_{ns,\tau}] = \boldsymbol{\xi}_{ns,\tau} = \sum_{r=0}^{\tau-1} S_{ns}^r \cdot \boldsymbol{\mu}_{ns} + S_{ns}^\tau \cdot \boldsymbol{P}_{ns,0} \tag{5a}$$

$$= (I_{ns} - S_{ns})^{-1}(I_{ns} - S_{ns}^\tau) \cdot \boldsymbol{\mu}_{ns} + S_{ns}^\tau \cdot \boldsymbol{P}_{ns,0} \tag{5b}$$

We can also now calculate the covariance of the preference values:

$$Cov[\boldsymbol{P}_{ns,\tau}] = \Omega_{ns,\tau} = Cov\left[\sum_{r=0}^{\tau-1} S_{ns}^r \cdot \boldsymbol{V}_{ns,\tau-r} + S_{ns}^\tau \cdot \boldsymbol{P}_{ns,0}\right] \tag{6a}$$

$$= \sum_{r=0}^{\tau-1} \left[S_{ns}^r \cdot \Phi_{ns} \cdot S_{ns}^{r'}\right] \tag{6b}$$

The resulting calculations are complex, but as shown in our earlier work (Hancock et al., 2018), we can further simplify $Cov[\boldsymbol{P}_{ns,\tau}]$ such that we can avoid the summation. We replace the feedback matrices with a matrix Z_{ns} of size $J_n^2 \times J_n^2$ (where J_n is the number of alternatives) and reshape Φ_{ns} (with entries $p_{i,j}$) into a column matrix, $\overline{\Phi_{ns}}$:

$$Z_{ns} = \begin{bmatrix} z_{1,1} & z_{1,2} & \cdots & z_{1,J_n^2} \\ z_{2,1} & z_{2,2} & \cdots & z_{2,J_n^2} \\ \vdots & \vdots & \ddots & \vdots \\ z_{J_n^2,1} & z_{J_n^2,2} & \cdots & z_{J_n^2,J_n^2} \end{bmatrix}, \overline{\Phi} = \begin{bmatrix} p_{1,1} \\ p_{1,2} \\ \vdots \\ p_{1,J_n} \\ p_{2,1} \\ \vdots \\ p_{J_n,J_n} \end{bmatrix} \tag{7}$$

As S_{ns} is a symmetric matrix, we can then define Z_{ns} by setting it as the Kronecker product of S_{ns} with itself: $z_{i,j} = s_{(i \bmod J_n, j \bmod J_n)} \cdot s_{(\lceil i/J_n \rceil, \lceil j/J_n \rceil)}$. The calculation of the covariance of $\boldsymbol{P}_{ns,\tau}$ now simplifies to:

$$Cov[\boldsymbol{P}_{ns,\tau}] = \Omega_{ns,\tau} = \sum_{r=0}^{\tau-1}\left[S_{ns}^{r}\cdot\Phi_{ns}\cdot S_{ns}^{r'}\right] \tag{8a}$$

$$= \sum_{r=0}^{\tau-1}\left[Z_{ns}^{r}\cdot\overline{\Phi_{ns}}\right] \tag{8b}$$

$$= (I_{ns}-Z_{ns})^{-1}\,(I_{ns}-Z_{ns}^{\tau})\,\overline{\Phi_{ns}} \tag{8c}$$

These succinct forms for $\boldsymbol{\xi}_{ns,\tau}$ and $\Omega_{ns,\tau}$ mean that we can now calculate the probability with which each alternative is chosen. On the basis of the multivariate central limit theorem, $\boldsymbol{P}_{ns,\tau}$ converges to the multivariate normal distribution (Roe et al., 2001). This means that if a time (step) threshold is reached, the choice probability of choosing alternative j from the set CS_{ns} of J_n alternatives at step τ is:

$$\begin{gathered} Prob\left[\boldsymbol{P}_{ns,\tau}[j] = \max_{i\in CS_{ns}} \boldsymbol{P}_{ns,\tau}[i]\right] = \\ \int_{\boldsymbol{X}_{ns,\tau}>0} exp\left[-(\boldsymbol{X}_{ns,\tau}-\boldsymbol{\Gamma}_{ns,\tau})'\Lambda_{ns,\tau}^{-1}(\boldsymbol{X}_{ns,\tau}-\boldsymbol{\Gamma}_{ns,\tau})/2\right]/(2\pi|\Lambda_{ns,\tau}|^{0.5})dX \end{gathered} \tag{9}$$

with $\boldsymbol{X}_{ns,\tau} = [\boldsymbol{P}_{ns,\tau}[j]-\boldsymbol{P}_{ns,\tau}[1],\ldots,\boldsymbol{P}_{ns,\tau}[j]-\boldsymbol{P}_{ns,\tau}[J_n]]'$, the set of differences between the preference for the j^{th} element and the preferences for each respective element $i\in C_{ns}$, $i\neq j$. Then we have $\boldsymbol{\Gamma}_{ns,\tau} = L_{ns}\cdot\boldsymbol{\xi}_{ns,\tau}$ and $\Lambda_{ns,\tau} = L_{ns}\cdot\Omega_{ns,\tau}\cdot L_{ns}'$ where

$$L_{ns} = \begin{bmatrix} 1 & -1 & 0 & \ldots & \ldots & 0 \\ 1 & 0 & -1 & \ddots & & \vdots \\ 1 & \vdots & \ddots & \ddots & \ddots & \vdots \\ 1 & \vdots & & \ddots & -1 & 0 \\ 1 & 0 & \ldots & \ldots & 0 & -1 \end{bmatrix} \tag{10}$$

with L_{ns} being a matrix constructed with a column vector of 1s and a negative identity matrix of size J_n-1 where J_n is the number of alternatives. The column vector of 1s is placed in the j^{th} column where j is the chosen alternative.

2.2.2. *New developments*

2.2.2.1 *Scaling of DFT*

In a typical linear additive RUM or RRM model, changing the units of a single attribute only affects the parameter for that attribute. For example, changing the unit of travel time from minutes to hours results in the corresponding marginal utility component being multiplied by 60, with no impact on other parameters.

DFT on the other hand is scale-variant (Busemeyer and Diederich, 2002; Trueblood et al., 2013a), where the requirements that the attribute weights sum to one ($\sum_k w_k = 1$) means that a change in the scale for one attribute, unless accounted for, might be incorrectly interpreted as a change in the distribution of the attribute weights. The following material illustrates how to extend DFT (and related models) so that importance weights are unchanged by such scale changes (for the proof of how the new scaling parameters result in DFT become scale-invariant, please see Appendix C). Suppose we are studying a choice situation where the two attributes are cost and time, with cost measured in £ sterling and time measured in hours. Let $c_{£}$ (resp., t_h) be the cost in £ (resp., duration in hours) of a particular choice alternative, and assume that we have been able to show that its value is given by:

$$(0.6)c_{£} + (0.4)t_h, \tag{11}$$

where the coefficient (0.6) [resp., (0.4)] measures the importance of cost in £ (resp., time in hours) to the decision maker.[2] The above formulation is fine so long as one continues to work with cost measured in £ and time measured in hours, but will lead the researcher astray if they want to change their measurements to cost measured in Euros and time measured in minutes. Rather than go down that erroneous path, we simply note that there are implicit parameters in Equation 11 that indicate that c (cost) is measured in £ and t (time) is measured in hours. We make those parameters explicit by rewriting Equation 11 as:

$$(0.6)k_{c_{£}}c_{£} + (0.4)k_{t_h}t_h, \tag{12}$$

where $k_{c_{£}} = k_{t_h} = 1$. Such parameters have various names in the (measurement) literature, including *dimensional parameters* (Luce, 1962) and *scale-dependent parameters* (Luce et al. (1990), Section 22.2.4 and page 308). The advantage - in fact, conceptual necessity - of including such parameters is that if we now change our measurement scales - say, from £ to Euros and from hours to minutes - then the importance weights will remain the same (as they should), and the (scale) parameters for cost and time change. For example, with $1£ = 1.18$ Euro and one hour equal 60 minutes, Equation 11 becomes:

$$(0.6)k_{c_E}c_E + (0.4)k_{t_m}t_m, \tag{13}$$

where c_E (resp. t_m) is the cost in Euros (resp., time in minutes) and $k_{c_E} = \frac{1}{1.18}$ and $k_{t_m} = \frac{1}{60}$.

It is possible that scaling functions can be applied to the attributes before they enter Equation 3 (see Section 3.5 or also the 'scaling of attributes' section of Hancock et al. (2018)). However, it is not clear which method should be used a priori, thus we instead define a new scaling method which translates attribute values into 'subjective' values directly. This is achieved by multiplying the values by a vector of attribute-specific scaling coefficients,[3] $\boldsymbol{\beta}_{DFT}$, with properties paralleling the scale-dependent parameters of Equations 11 and 12. This results in the following function for the random valence vector at step τ (dropping the indices for individual and choice task), $\boldsymbol{V}_\tau$:

$$\boldsymbol{V}_\tau = C \cdot M \cdot \beta \cdot \boldsymbol{W}_\tau + \boldsymbol{\varepsilon}_\tau \tag{14}$$

where M is the original attribute matrix, but with each attribute multiplied by its corresponding scaling value from the diagonal matrix β, which has the set of estimated scaling values, $\boldsymbol{\beta}_{DFT}$, on the diagonal entries. With this specification, a decision-maker still attends to a given attribute at random in a given evaluation. In $\boldsymbol{W}_t$, as before, one element is equal to 1 with all others 0.

[2]See Marley et al. (2008), Section 6 for a discussion of the conceptual and practical issues in obtaining such (separable) information on importance and value.

[3]We use the term $\boldsymbol{\beta}_{DFT}$ here as these values correspond to the marginal utility components, β, of RUM, but they cannot be used equivalently in, for example, value of travel time calculations.

Furthermore, we also multiply the attributes by their corresponding scaling coefficients before they are used to estimate the distances between alternatives in the feedback matrix (D_{ns} in Equation 2). The distance between alternatives i and j is thus:

$$D^2_{ns,ij} = \sum_{k=1}^{K} (\beta_k \cdot (x_{ns,i,k} - x_{ns,j,k}))^2 \tag{15}$$

with $x_{ns,k,i}$ the objective value for the k^{th} attribute for alternative i and β_k the relative importance of attribute k. The addition of these scaling parameters results in three possible versions of DFT models:

1. A general model incorporating both attribute attribute-specific scaling coefficients and attribute weights.
2. A model with attribute-specific scaling coefficients only. The attribute weights, w_k, are not estimated, with the modeller simply fixing the weights equal to $1/K$ where K is the number of attributes. This implies that each attribute is just as likely to be attended to at each preference updating step. Whilst this would be unlikely to be the case in the presence of eye-tracking data, it is a reasonable assumption a priori with choice-only data.
3. A model with attribute weights only. The scaling coefficients are all fixed to a value of 1. This is of course equivalent to the original multialternative DFT specification by Roe et al. (2001).

We empirically test these different versions against each other and alternative scaling methods in Section 3.5. The addition of attribute scaling coefficients to DFT results in a number of important benefits.

- First, the revised version of DFT is no longer scale-variant. Changing the unit of travel time from minutes to hours will now impact the estimate for the travel time scaling coefficient only. This means that for each marginal utility coefficient in a RUM model (or a marginal regret coefficient in a RRM model), there is a corresponding attribute scaling coefficient in the DFT model. This allows us to make comparisons across the different models in terms of relative importance of different attributes.
- Second, the attributes are now adjusted accordingly for their relative importance *before* they enter the feedback matrix, meaning that we can calculate an appropriate psychological distance by simply taking Euclidean distances in the calculation of the feedback matrix. Consequently, we do not need a separate ‘dominance’ parameter as defined by Berkowitsch et al. (2015) to take the relative importance of the attributes into account.
- Third, we now separate out the effects of the frequency of considering individual attributes, through the weights, and the importance of different attributes in changing the preference values, through the scaling parameters. Conceptually, only the latter should be influenced by changes in units.

- Fourth, an even more important benefit of the proposed scaling approach relates to the possibility of attributes having opposite impacts on probabilities, i.e. some attributes being desirable and others being undesirable. In the traditional DFT model, an analyst needs to make a priori assumptions about this directionality, and failing to correct for the *sign* of the impact of attributes can have undesired consequences, as illustrated in Table 3 of Hancock et al. (2018). With our new approach, we no longer require a priori knowledge or assumptions about whether an attribute has a positive or negative impact on the likelihood of an alternative being chosen, as the attribute scaling parameters can be estimated to be either positive or negative. This not only results in it being possible to take all attributes into account without any initial adjustments, but would also, in a random coefficients DFT model, allow for the possibility of different signs for a given parameter across different individuals.

- Finally, this new scaling method allows for the importance of attributes to be a function of socio-demographic variables, leading to the possible requirement of alternative specific coefficients for certain attributes, and/or interactions such as income effects. Whilst our previous work (Hancock et al., 2018) demonstrated that DFT could incorporate socio-demographics through adjustments to the attention weights, the new system removes the non-linearity caused by shifts in $Cov[\boldsymbol{W}_{ns,\tau}]$. This means that one could, for example, include β-coefficients that vary substantially across decision-makers.

2.2.2.2 *Identification of parameters*

The literature on DFT (and other mathematical psychology models) often lacks crucial details in relation to model identification. As with any standard choice model, a DFT model requires a normalisation of location and scale. For DFT, this is a result of the multiplication of the expectation, $\boldsymbol{\xi}_\tau$, and covariance, Ω_τ, by L (see Equation 9, noting that throughout this section we drop the indices for individual and choice task). This results in only differences between alternatives, in terms of both expected values and covariances, mattering. In order to estimate the choice probabilities under a DFT model, we require estimates for four 'process parameters' (parameters which have no equivalent measure in a traditional model such as RUM or RRM) which are exclusive to DFT and inform the process by which alternatives accumulate preference. These are ϕ_1 and ϕ_2, the sensitivity and memory parameters respectively, the number of preference updating steps, τ, and the variance of the error term, σ_ε^2. As the choices we investigate in this paper do not have a strictly imposed time threshold, we make no assumptions on the relationship between the number of deliberation steps, τ, and the real world time, t, taken to make the choice.

Theoretical identification

To study identification, we write the expectation of the preference values after τ deliberation steps as $\boldsymbol{\xi}_\tau = \boldsymbol{a} + \boldsymbol{b}$ and the covariance of the preference values as $\Omega_\tau = c\overline{(d+e)}$. Consequently, we have:

- $\boldsymbol{a} = (I - S)^{-1}(I - S^\tau) \cdot \boldsymbol{\mu}$

- $\boldsymbol{b} = S^\tau \cdot \boldsymbol{P}_0$

- $c = (I - Z)^{-1}(I - Z^{\tau})$
- $d = C \cdot \beta \cdot M \cdot \Psi \cdot M' \cdot \beta' \cdot C'$
- $e = \sigma_{\varepsilon}^2 \cdot I.$

An overspecification occurs if we can have two sets of parameters such that exactly the same set of probabilities are generated for the full set of choice scenarios. This occurs if all means change by the same amount, or if the means and square root of the covariance matrix are scaled by the same factor. We thus consider the following three scenarios:

1. $\boldsymbol{a}_1 + \boldsymbol{b}_1 = \delta + \boldsymbol{a}_2 + \boldsymbol{b}_2$
2. $\boldsymbol{a}_1 + \boldsymbol{b}_1 = (\boldsymbol{a}_2 + \boldsymbol{b}_2) \cdot \gamma$
3. $c_1\overline{(d_1 + e_1)} = (c_2\overline{(d_2 + e_2)}) \cdot \gamma^2$

An overspecification of location could exist if some value δ exists such that scenario 1 holds. This is a result of the specification of Equation 9, where $\boldsymbol{\xi}_{\tau}$ is multiplied by L, giving a vector of differences between the expected preference of the chosen alternative and each other alternative. Consequently, the addition of δ to each preference value results in no change in the value of $L\boldsymbol{\xi}_{\tau}$.

To avoid this overspecification, the simplest solution is to fix one of the alternative specific constants (asc) in the initial preference matrix $\boldsymbol{P}_0$. Under DFT, adding the same value to each alternative specific constant results in the same increase in expected preference value for each alternative when $\phi_2 = 0$ (as this results in the feedback matrix becoming an identity matrix, which means there is no impact on $\boldsymbol{P}_0$, see Equation 5).

An overspecification of scale can exist if some value γ exists such that scenarios 2 and 3 both hold. This is a result of the probabilities under DFT being calculated with the use of multivariate normal distributions (see Equation 9). Critically, these scenarios result in $\boldsymbol{\xi}_{\tau,1} = \gamma \cdot \boldsymbol{\xi}_{\tau,2}$ and $\Omega_{\tau,1} = \gamma^2 \cdot \Omega_{\tau,2}$. Then if we have a set of parameters, θ_1 that results in probabilities, Pr_1:

$$Pr_1 = \int_{\boldsymbol{X}_1 > 0} exp\left[-(\boldsymbol{X}_1 - \boldsymbol{\Gamma}_1)'\Lambda_1^{-1}(\boldsymbol{X}_1 - \boldsymbol{\Gamma}_1)/2\right]/(2\pi|\Lambda_1|^{0.5})dX, \tag{16}$$

with the corresponding parameter values θ_2 resulting in probabilities, Pr_2:

$$Pr_2 = \int_{\boldsymbol{X}_2 > 0} exp\left[-(\boldsymbol{X}_2 - \boldsymbol{\Gamma}_2)'\Lambda_2^{-1}(\boldsymbol{X}_2 - \boldsymbol{\Gamma}_1)/2\right]/(2\pi|\Lambda_2|^{0.5})dX, \tag{17}$$

we can now simply observe that the substitutions implied by scenarios 2 and 3, $\boldsymbol{X}_1 = \gamma \cdot \boldsymbol{X}_2$, $\boldsymbol{\Gamma}_1 = \gamma \cdot \boldsymbol{\Gamma}_2$ and $\Lambda_1 = \gamma \cdot \Lambda_2$, result in $Pr_1 = Pr_2$, as the γ terms all drop out of the equation. We thus obtain two distinct sets of parameters θ_1 and θ_2 that result in the same estimated probabilities of observing the full set of choices.

To avoid this overspecification, we need to understand the situations under which scenarios 2 and 3 hold. One example of when this is the case is when:

$$\theta_1 = [\boldsymbol{w}, \boldsymbol{\beta}_{DFT}, \boldsymbol{P}_0, \phi_1, \phi_2, \tau, \sigma_\varepsilon^2] \tag{18}$$

and:

$$\theta_2 = [\boldsymbol{w}, \gamma \cdot \boldsymbol{\beta}_{DFT}, \gamma \cdot \boldsymbol{P}_0, \phi_1/(\gamma^2), \phi_2, \tau, \gamma \cdot \sigma_\varepsilon^2], \tag{19}$$

with $\boldsymbol{\beta}_{DFT} = [\beta_1, \beta_2, .., \beta_k]$, $\boldsymbol{w} = [w_1, w_2, .., w_k]$ and $\boldsymbol{P}_0 = [\delta_1, \delta_2, .., \delta_n]$. To avoid this overspecification, we need to fix a parameter such that $\gamma = 1$ and thus $\theta_1 = \theta_2$. In the absence of a priori knowledge on the directionality of the attributes or the directionality of the underlying preferences towards an alternative, the scaling parameters $\beta_1, \beta_2, .., \beta_k$ and the alternative specific constants should not be fixed. Furthermore, ϕ_2 could have an estimate of zero, resulting in ϕ_1 having no impact in Equations 18 and 19. Consequently, the safest option[4] is to fix $\sigma_\varepsilon^2 = 1$.

It is easy to see a relationship between these two normalisations of location and scale and their corresponding normalisations in a RUM context.

It is also worth noting at this point that whilst in choice-only datasets, the attribute attention weights and attribute scaling parameters may appear to capture the same feature (the relative importance of an attribute), these parameters are not theoretically confounded, as they have differing impacts within the calculation of choice probabilities under DFT. The attribute attention weights do not impact the feedback matrix, whilst the attribute scaling parameters do. This is a direct result of setting the distance between alternatives as $D^2_{ns,ij} = \sum_{k=1}^{K} (\beta_k \cdot (x_{ns,i,k} - x_{ns,j,k}))^2)$ in the definition for the feedback matrix. Additionally, the attribute scaling parameters do not impact $\Psi = Cov[W_\tau]$ (the covariance of the attention weights), which is used directly in the calculation of the covariance of the preference values (see Equation 6, in which Ψ impacts Φ). This is instead calculated solely with the estimates for the attribute attention weights. As a result, only a model containing both will have the full flexibility to capture both effects. We test models with both these features in Section 3.5.

Empirical identification and restrictions

The process parameters in DFT have important behavioural roles. However, DFT models are routinely estimated on data where the only observed outcome is the choice itself, with little information about the process by which that choice was reached. If such process information was available, analysts could use it as additional indicators (i.e. additional dependent variables) in a joint estimation of process and outcome, and this would help inform the values of these parameters. In the absence of such data however, some of the parameters may become partially confounded.

We additionally impose various restrictions to aid empirical identification of parameters under a DFT model. These, together with the restrictions identified in the previous section, are detailed in Table 1.

[4] An analyst should however be aware of the fact that the required stochasticity in a DFT model could be generated by the random attribute attendance alone, meaning that the estimate for $\sigma_\varepsilon^2 = 0$. This would result in the requirement of an alternative normalisation. We explore this possibility in DFT model 3 in Table 4.

TABLE 1 : Restrictions on DFT parameters

No.	Parameter	Description	Restrictions	Reason	Estimated parameter	Relation
1	δ_n	asc for alternative n	=0	Theoretical identification	n/a	n/a
2*	σ_ε^2	error	=1	Theoretical identification	n/a	n/a
3*	β_k	scale for attribute k	fixed	Theoretical identification	n/a	n/a
4**	ϕ_2	decay	=0	Empirical identification	n/a	n/a
5**	ϕ_1	sensitivity	fixed	Empirical identification	n/a	n/a
6	w_i	weights for attributes	$\sum_1^k w_k = 1$	DFT assumption	w_i^*	$w_i = exp(w_i^*)/\sum_1^k exp(w_k^*)$
7	w_1	first attribute weight	$w_1^* = 0$	DFT assumption	n/a	n/a
8**	ϕ_1	sensitivity	> 0	DFT assumption	ϕ_1^*	$\phi_1 = exp(\phi_1^*)$
9	τ	preference updating steps	> 1	DFT assumption	τ^*	$\tau = 1 + exp(\tau^*)$
10*	σ_ε^2	error	≥ 0	Mathematical	σ_ε	$\sigma_\varepsilon^2 = (\sigma_\varepsilon)^2$

*Note that only one of restrictions 2 and 3 should be used, and that restriction 10 should be applied if 2 is not used.
** If the estimate for ϕ_2 is equal to or close to zero, restrictions 4 and 5 may also be required to ensure standard errors can be estimated for the other parameters (with restriction 8 no longer required).

Some of these constraints are necessary to avoid identification issues, while others simply avoid sign issues. The following list details why each restriction is required.

- (1, 2 and 3) There are two theoretical identifications (restriction 1 and either restriction 2 or 3 in Table 1) detailed in the previous section, applied to avoid a theoretical overspecification.
- (4 and 5) There is an empirical identification issue as a result of the decay parameter, ϕ_2. In case there is no impact for the decay parameter, we obtain a value of ϕ_2 equal to or close to zero. This results in the sensitivity parameter, ϕ_1, having no impact on the probabilities of alternatives, thus resulting in an overspecified model. As a consequence, we try two specifications of DFT for each dataset in our empirical applications: one with and one without estimated values for these two feedback parameters (see Tables 4, 5 and 6 in the empirical applications section of the paper).
- (6 and 7) A DFT model assumes that a decision-maker attends one attribute at each preference updating step, which implies that the attribute attention weights must all be positive and sum to one. Estimation of these attention weights is thus aided through the use of logistic transformations and identification is then ensured by fixing one of the adjusted attribute attention weights $w_1^* = 0$ (which is required as adding the same value to every w_i^* results in no change to w_i).
- (8) An exponential transformation ensures that the sensitivity parameter is positive. This restriction results in alternatives that are more similar to each other competing more with each other than alternatives that are less similar, as assumed implicitly by DFT models.
- (9) The number of preference updating steps must exceed a value of one. This can be also be solved through the use of an exponential transformation.
- (10) The noise that is added on at each step to the valence (see Equation 3) is drawn from a normal distribution with mean 0 and variance σ_ε^2. Consequently, as $\sigma_\varepsilon^2 \geq 0$, we instead estimate the standard deviation, σ_ε.

2.3. The multi-attribute linear ballistic accumulator model (MLBA)

In this section, we first look at the existing implementation of MLBA before making a number of methodological improvements and finally turning to identification issues. The latter section here is particularly important given that identifiability issues have been noted for LBA models in the context of simultaneously modelling choice outcome and response time (Evans, 2020).

2.3.1. Existing implementation

2.3.1.1 *Theoretical model*

We begin our discussion by focussing on LBA, i.e. the single attribute model. Under LBA, each alternative has a preference strength that grows linearly towards a threshold (see right panel in Figure 1). The chosen alternative in an LBA model is the first alternative to pass a threshold value, χ. There are two components to the LBA component of the model; the start points and the drift rates.

Start points for each of the alternatives are drawn separately from a uniform distribution $U[0,A]$, where A is estimated. For example, Figure 1 demonstrates what a decision might look like if the start points are drawn from a distribution $U[0,2]$. A different value A_j could be estimated for each alternative j, although it is common practice (Trueblood et al., 2014) to assume that all alternatives have starting values that are drawn using the same estimate A. We thus also assume that the same value A is used for each alternative, in both the theoretical identification (2.3.2.3) and the applications of MLBA in this paper (3).

The differences in the structure of the models in Trueblood et al. (2013a, 2014) illustrate the fact that many forms are possible for the drift rates when the basic LBA is extended to alternatives with multiple attributes. While it is thus reasonable to expect that there might be other suitable drift rate forms beside those of the final MLBA model, we choose to fit versions similar to the mainstream version of MLBA (Trueblood et al., 2014) as this outperforms the first version (described by Trueblood et al. (2013a)) for our choice datasets (see Appendix A).

In the original MLBA work (Trueblood et al., 2014), the model translates attribute values into 'subjective values'. In their example, Trueblood et al. (2014) had two similar attributes: testimony strength of eyewitness P and testimony strength of eyewitness Q. To translate objective values into subjective values, a parameter was introduced such that an 'indifference curve' could be calculated to avoid issues of extremeness aversion (Chernev, 2004), where, for example, values of 50-50 might be preferred to 70-30. This example is however very different from many real-world choice settings, including travel ones, where the individual attributes are not as closely related. To translate objective attribute values into subjective values in this case, we instead require some measure to translate the values appropriately such that the relative importance of the attributes is accounted for. We achieve this by using attribute importance parameters instead of m, the additional parameter in the original specification of MLBA (Trueblood et al., 2014), with more details in Section 2.3.2.2.

MLBA assumes that the drift rate for each alternative is an independent draw from a normal distribution (truncated below zero), where, for individual n, choice task s, and alternative j, we have the drift rate $D_{ns,j}$ given as:

$$D_{ns,j} \sim TN\left(d_{ns,j}, \sigma_{ns,j,\varepsilon}\right) \tag{20}$$

with mean drift rate $d_{ns,j}$ and standard deviation $\sigma_{ns,j,\varepsilon}$. Typically, and in all the applications in this paper, the standard deviation is set to be equal across alternatives, individuals and choice tasks, i.e. $\sigma_{ns,j,\varepsilon} = \sigma_\varepsilon, \forall n,s,j$, but a different value could be estimated for each drift rate (Trueblood et al., 2013b).

In the current version of MLBA, mean drift rates follow the specification used by Trueblood et al. (2014):

$$d_{ns,j} = v_{ns,j} + I_0 \tag{21}$$

where I_0 is a positive constant (which can be specified such that all drift rates have a positive mean) and $v_{ns,j}$ is a value function, similar to random regret minimisation in that it compares an alternative j against all other alternatives i across each attribute x. Specifically, with K attributes and J_n alternatives, we have that:

$$v_{ns,j} = \sum_{i \neq j}^{J_n} \sum_{k=1}^{K} (w_{x_{ns,ij,k}} \cdot (x_{ns,j,k} - x_{ns,i,k})). \tag{22}$$

In this notation, $x_{ns,i,k}$ is the objective value for the k^{th} attribute for alternative i, and $w_{x_{ns,ij,k}}$ is a weight for attribute k and alternative pairing i and j, which relates to the similarity between them,[5]. In particular, the similarity is assumed to be an exponential decaying function of distance, with:

$$w_{x_{ns,ij,k}} = exp(-\lambda \cdot |x_{ns,i,k} - x_{ns,j,k}|) \tag{23}$$

Two different values of λ are used depending on whether the difference between $x_{ns,i,k}$ and $x_{ns,j,k}$ is positive or negative:

$$\lambda = \begin{cases} \lambda_1, & \text{if } x_{ns,j,k} \geq x_{ns,i,k}. \\ \lambda_2, & \text{if } x_{ns,j,k} < x_{ns,i,k}. \end{cases} \tag{24}$$

This feature can capture differences between the subjective similarity between A and B and the subjective similarity between B and A, which may not be equal (Tversky, 1977), with gains and losses regularly having been shown to be treated differently in a transport context (Hess et al., 2008; Masiero and Hensher, 2010; Stathopoulos and Hess, 2012). We consider possible values and interpretations of these parameters further in Appendix D. Additionally, it is worth noting here that work by Terry et al. (2015) suggests that the distributional form of LBA has limited impact. This implies that the formula for the mean drift rates (Equation 22) may be key. We return to this point in Section 2.3.3.

Calculating choice probabilities

If we have values for the drift rates of the alternatives and for the start point and threshold (A and χ respectively), we can calculate the probability of each alternative's accumulator being the first to finish, i.e. for its value function to exceed the threshold χ before any others do (Brown and Heathcote, 2008).

As that the starting evidence is drawn from a uniform distribution $U[0,A]$, the amount of evidence that needs to be accumulated for an alternative to reach the threshold χ is U[χ-A,χ] (assuming $\chi > A$). Given an alternative's drift rate distribution, $D_{ns,j}$, the cumulative distribution function for the time taken for the accumulator associated with alternative j is:

$$F_{ns,j}(t) = Prob\left(\frac{U[\chi - A, \chi]}{D_{ns,j}} < t\right). \tag{25}$$

Brown and Heathcote (2008) demonstrate that for a mean drift rate following a normal distribu-

[5]Note that Equation 22 is equivalent to Equation 3 in Trueblood et al. (2014) but for objective values rather than subjective values, and with a generalisation to multiple alternatives and multiple attributes.

tion[6], this reduces to:

$$F_{ns,j}(t) = 1 + \frac{\chi - A - t \cdot d_{ns,j}}{A} \cdot \Phi\left(\frac{\chi - A - t \cdot d_{ns,j}}{t \cdot \sigma_\varepsilon}\right) - \frac{\chi - t \cdot d_{ns,j}}{A} \cdot \Phi\left(\frac{\chi - t \cdot d_{ns,j}}{t \cdot \sigma_\varepsilon}\right) + \frac{t \cdot \sigma_\varepsilon}{A} \cdot \phi\left(\frac{\chi - A - t \cdot d_{ns,j}}{t \cdot \sigma_\varepsilon}\right) - \frac{t \cdot \sigma_\varepsilon}{A} \cdot \phi\left(\frac{\chi - t \cdot d_{ns,j}}{t \cdot \sigma_\varepsilon}\right), \tag{26}$$

where ϕ and Φ are the standardised normal distribution's density and cumulative density functions, respectively. The associated probability density function is then:

$$f_{ns,j}(t) = \frac{1}{A}\left[- d_{ns,j} \cdot \Phi\left(\frac{\chi - A - t \cdot d_{ns,j}}{t \cdot \sigma_\varepsilon}\right) + d_{ns,j} \cdot \Phi\left(\frac{\chi - t \cdot d_{ns,j}}{t \cdot \sigma_\varepsilon}\right) + \sigma_\varepsilon \cdot \phi\left(\frac{\chi - A - t \cdot d_{ns,j}}{t \cdot \sigma_\varepsilon}\right) - \sigma_\varepsilon \cdot \phi\left(\frac{\chi - t \cdot d_{ns,j}}{t \cdot \sigma_\varepsilon}\right)\right]. \tag{27}$$

To then calculate the probability of a given alternative j being chosen[7], we need to calculate the probability density function of alternative j (given individual n, choice task s, and the set of alternatives CS_{ns}) reaching the threshold χ before all other alternatives $i \in CS_{ns}$, $i \neq j$:

$$Prob(j|CS_{ns}) = \int_0^\infty F_{ns,j}(t)dt = \int_0^\infty f_{ns,j}(t) \prod_{i \neq j}^{J_n} (1 - F_{ns,i}(t))dt. \tag{28}$$

2.3.2. New developments

2.3.2.1 *Incorporating baseline preferences in MLBA*

A key feature of many discrete choice models is the concept of alternative specific constants that capture baseline preferences for specific alternatives. Hancock et al. (2018) discusses in detail how this can be implemented in a DFT model. Here, we extend this to the MLBA model.

In particular, we rewrite Equation 21 as

$$d_{ns,j} = max(0, \delta_{ns,j} + v_{ns,j} + I_0), \tag{29}$$

where $\delta_{ns,j}$ is an additional alternative specific estimated constant capturing a baseline preference for alternative j given choice set s and individual n. Given that we have case studies where each individual completes each choice task a single time (as a contrast to many typical applications of MLBA), we do not have enough data to estimate context-specific baseline preferences, and thus set $\delta_{ns,j} = \delta_j$. The final adjustment we make is to ensure that each mean drift rate has a minimum value of zero (as opposed to fixing I_0 such that this is the case). The use of truncated normals results in it being possible that adding the same constant to each drift rate can result in some more deterministic choices (when the mean drift rates are negative, for example) as well as some less deterministic choices. Adjusting the mean drift rates such that they are at least zero avoids this and aids the estimation of MLBA. For more details of this, please see Appendix B.

[6]We follow the first adjustment made by Heathcote and Love (2012) to translate this for truncated normals.

[7]For full derivations of Equations 26, 27 and 28, refer to Appendix A of Brown and Heathcote (2008).

2.3.2.2 *Incorporating attribute specific weights in MLBA*

An additional limitation of the original implementation of MLBA is in the treatment of the different attributes. Firstly, this applies in terms of directionality, noting that λ_1 is used for a positive difference between objective values $x_{ns,i,k}$ and $x_{ns,j,k}$ independently of whether an increase in attribute x_k increases or decreases the attractiveness of an alternative. This limitation is analogous to the issue with using weight parameters in DFT and would require an analyst to a priori change the sign for undesirable attributes. Secondly, the actual impact of differences between alternatives in a given attribute x_k is constant across attributes. Whilst one possibility is to use different valuation and weighting functions (Cohen et al., 2017), Trueblood et al. (2014) suggest that attribute biases can be dealt with by including attribute-specific 'bias parameters', β_k (an approach analogous to the attribute-specific scaling coefficients that we defined for DFT) in Equation 23, which then becomes:

$$w_{x_{ns,ij,k}} = exp(-\lambda \cdot \beta_k \cdot |x_{ns,i,k} - x_{ns,j,k}|). \tag{30}$$

However, we can relax the limitations of attribute bias and directionality simultaneously by also making an adjustment to the value function (Equation 22), which now takes the same form as that of the original specification with the exception that we add in attribute-specific scaling coefficients, β_k.[8] This results in the value function from Equation 22 being redefined as:

$$v_{ns,j} = \sum_{i \neq j}^{J_n} \sum_{k=1}^{K} (w_{x_{ns,ij,k}} \cdot \beta_k \cdot (x_{ns,j,k} - x_{ns,i,k})). \tag{31}$$

As with the scaling applied to DFT, this change allows us to also make inferences about the relative importance of different attributes in MLBA, as well as incorporating interactions with sociodemographics at the level of individual attributes.

2.3.2.3 *Identification of parameters*

For the estimation of the probability of choosing alternatives in a MLBA model, we require estimates for K attribute scaling parameters, J alternative specific drift rate constants (we drop the indices for individual and choice task in this section) and estimates for six process parameters (A and χ, the start and threshold parameters respectively, a drift rate constant, I_0, a parameter for the standard deviation of the drift rates, σ_ε, and similarity parameters λ_1 and λ_2). Whilst different values for these process parameters could be used for different alternatives, we assume that the same value is used for each alternative (with the exception of the drift rate constants), for the applications in this paper and for the purposes of identification in the sections below. MLBA has closed forms for the choice probabilities and response times. However, using choice-only data requires a number of restrictions, both theoretical and empirical, to aid and improve estimation.

Theoretical identification

For choice-only data (i.e. where no additional process information is available), the start and

[8] Note that for the applications in this paper, we assume that the values for the λ parameters are fixed across attributes. Equivalently, the coefficients β_k are used in both Equations 30 and 31. We relax this assumption when using this function within choice models based on quantum probability in Hancock et al. (2020).

threshold parameters A and χ are perfectly confounded. This is a consequence of the fact that multiplying A and χ by a scale factor f results in no change in the probabilities with which each alternative is chosen. Instead, this simply changes the time that alternative j finishes in from t_j to $f \cdot t_j$. As all alternatives are impacted in the same way, the probability of a given alternative being chosen is not impacted. This can be demonstrated by multiplying A, χ and t by the factor f in the cumulative distribution function for the time taken by an accumulator (Equation 25). The factor f drops out, resulting in identical cumulative distribution functions being given by the set of parameters $\theta_1 = [A, \chi, t]$ and $\theta_2 = [A \cdot f, \chi \cdot f, t \cdot f]$. Consequently, when integrated over $t = 0 \to \infty$ (see Equation 28), the resulting probabilities are equal. This means that we need to fix either A or χ. In the applications in this paper, we choose (in line with previous applications, see Trueblood et al. 2014) to fix $A = 1$.

Additionally, the choice probabilities are unchanged if all mean drift rates d_j and the standard deviation σ_ε are multiplied by the same factor, g. This is possible, for example, with the parameter sets θ_1 and θ_2:

$$\theta_1 = [\boldsymbol{\beta}_{MLBA}, \boldsymbol{\delta}, A, \chi, \sigma_\varepsilon, I_0, \lambda_1, \lambda_2] \tag{32}$$

and

$$\theta_2 = [\boldsymbol{\beta}_{MLBA} \cdot g, \boldsymbol{\delta} \cdot g, A, \chi, \sigma_\varepsilon \cdot g, I_0 \cdot g, \lambda_1 \cdot (1/g), \lambda_2 \cdot (1/g)], \tag{33}$$

with $\boldsymbol{\beta}_{MLBA} = [\beta_1, \beta_2, .., \beta_k]$ and $\boldsymbol{\delta} = [\delta_1, \delta_2, .., \delta_J]$. As with the previous example for changing start and threshold parameters, this effect simply changes the time that alternative j finishes in. Again, all alternatives are impacted in the same way, and thus the choice probabilities remain the same (as we are integrating over $t = 0 \to \infty$), with the only change being that each alternative j now takes $\frac{t_j}{g}$ rather than t_j time to finish. We must then fix either σ_ε or one of the drift rates d_j to ensure identification. As before, we follow Trueblood et al. (2014) in setting $\sigma_\varepsilon = 1$.

Furthermore, in contrast with RUM models, where the choice probabilities only depend on the differences between alternative specific constants δ_j (see Equation 29), each mean drift rate can have a separately identified (alternative specific) constant, as the addition of such constants make the choices more deterministic. This is a result of each alternative taking less time to finish, meaning there is less time for the alternatives to 'overtake' each other. Consequently, the draw for the starting point of an alternative has a greater influence on the probability of that alternative finishing first. Note that if we additionally estimate I_0 (which is desirable as this allows us to determine whether the baseline preferences for the alternatives differ significantly from each other), then one of the constants δ_j must be fixed to ensure identification.

Empirical identification and restrictions

We additionally impose various restrictions to aid empirical identification of parameters under a MLBA model. These, together with the restrictions identified in the previous section, are detailed in Table 2. The following list details why each restriction is required.

- (1) As detailed in the section on including baseline preference parameters for the different

TABLE 2 : Restrictions on the parameters within MLBA

No	Parameter	Description	Restrictions	Reason	Estimated parameter	Relation
1	δ_j	asc for alternative j	=0	Theoretical identification	n/a	n/a
2	A	start	=1	Theoretical identification	n/a	n/a
3	σ_ε	drift rate standard deviation	=1	Theoretical identification	n/a	n/a
4	λ_1	similarity	≥ 0	MLBA assumption	λ_1^*	$\lambda_1 = exp(\lambda_1^*)$
5	λ_2	similarity	≥ 0	MLBA assumption	λ_2^*	$\lambda_2 = exp(\lambda_2^*)$
6	χ	threshold	$\geq$ A	MLBA assumption	χ^*	$\chi = A * (1 + exp(\chi^*))$

alternatives (see Section 2.3.2.1), one of these alternative specific constants must be fixed to avoid a theoretical overspecification.

- (2 and 3) Further theoretical restrictions as discussed in Section 2.3.2.3.
- (4 and 5) Exponential transformations ensure that the lambda parameters are positive. This is required if the similarity between alternatives is to be a decaying function of distance (Trueblood et al., 2014).
- (6) A further restriction is required on the value of the threshold, χ, which must be greater than the start parameter (to avoid the possibility that more than one alternative reaches the threshold before any deliberation has taken place).

2.3.3. Intermediaries between MLBA and econometric choice models

LBA models have four components: start point distributions; thresholds; drift rate distributions; and parametric forms for the parameters (mean, usually) of the drift rate distributions. To date, closed forms have only been derived for uniform start point distributions, where this includes the case of no start point variability. Terry et al. (2015) present closed forms for the choice probabilities generated by several LBA models, each with uniform start point variability, but different drift rate distributions - that is, the probability of an alternative being chosen (Equation 28) can be calculated in closed form for LBA models with distributions other than truncated normals.

We now present notation for the general case with no start point variability, before looking at variations of MLBA by changing the value function or the drift rate distribution. Since threshold parameters are included in psychological models (mainly) to deal with response time effects, we set all thresholds to one in this section (see the theoretical discussion of this in the previous section, 2.3.2.3).

Let n denote the individual; s the current choice task; j an option from the current choice set CS_{ns} for individual n; and t time. Also, let $f_{ns,j}(t)$ (resp., $F_{ns,j}(t)$) denote the drift rate probability density (resp., cumulative density) function. Then, with no start point variability ($A = 0$) and threshold $\chi = 1$, the time taken for the accumulator associated with alternative j simplifies from Equation 25 to:

$$F_{ns,j}(t) = Prob\left(\frac{1}{D_{ns,j}} < t\right). \tag{34}$$

Then, assuming the drift rate distributions, $D_{ns,i}$, are non-negative or truncated below zero, we obtain that the probability of alternative j being chosen given the set of alternatives CS_{ns} is:

$$Prob(j|CS_{ns}) = Prob\left[\frac{1}{D_{ns,j}} = \min_{i \in CS_{ns}} \frac{1}{D_{ns,i}}\right], \tag{35}$$

The following class of LBA models is of particular interest because it leads to choice probabilities that satisfy a *context dependent Luce model* (equivalently, a *context dependent multinomial logit* *model*) - see Marley et al. (2008), and below, for definitions. Assume that the $D_{ns,j}$ are independently distributed with the form $d_{ns,j} \cdot \Delta_j$ where each $d_{ns,j}$ is non-negative, and for every j, $1/\Delta_j$ is Fréchet-distributed with shape and scale parameter equal to one - that is, for $r > 0$,

$$Prob(\frac{1}{\Delta_j} < r) = e^{-(\frac{1}{r})}, \tag{36}$$

Then the cumulative distribution of the time for the accumulator associated with alternative j to reach the threshold of 1 has the form: for $t > 0$,

$$F_{ns,j}(t) = Prob(\frac{1}{d_{ns,j} \cdot \Delta_j} < t), \tag{37}$$

Marley and Regenwetter (2017) demonstrates that the use of this distribution combined with zero start point variability leads to a context-dependent Luce model with:

$$P(j|CS_{ns}) = \frac{d_{ns,j}}{\sum_i^{J_n} d_{ns,i}}, \tag{38}$$

where J_n is the number of alternatives and $d_{ns,j}$ is a non-negative mean drift rate for alternative j that depends on the alternatives in the choice set CS_{ns}. Crucially, this context-dependent Luce model is equivalent to a context-dependent multinomial logit model (Marley et al., 2008) if we define utility $U_{ns,i}$ through: $U_{ns,i} = log(d_{ns,i})$, $i \forall CS_{ns}$.

In addition, if we take MLBA's value function (Equation 31) and adjust it such that it becomes linear (for example, by setting each $w_{x_{ns,ij,k}} = 1$), we obtain the linear value function

$$\begin{aligned} v_{ns,j} &= \sum_{i \neq j}^{J_n} \sum_{k=1}^{K} \beta_k \cdot (x_{ns,j,k} - x_{ns,i,k}) \\ &= \sum_{i=1}^{J_n} \sum_{k=1}^{K} \beta_k \cdot (x_{ns,j,k} - x_{ns,i,k}) \\ &= J_n \sum_{k=1}^{K} \beta_k \cdot (x_{ns,j,k}) - \sum_{i=1}^{J_n} \sum_{k=1}^{K} (\beta_k \cdot x_{ns,i,k}). \end{aligned} \tag{39}$$

If we now assume an LBA with no start point variability ($A = 0$) and Fréchet-distributed drift rates, we can use the above value function to define the mean drift rate by mapping $v_{ns,j}$ by a strictly monotonic increasing function onto the positive reals. In particular, if we define the drift rate by $d_{ns,j} = exp(v_{ns,j})$, and note that the second term in the above expression is a constant for the given choice set CS_{ns}, we obtain the following choice probabilities for option j:

$$P(j|CS_{ns}) = \frac{exp\left[J_n \sum_{k=1}^{K} \beta_k \cdot (x_{ns,j,k})\right]}{\sum_{i=1}^{J_n} exp\left[J_n \sum_{k=1}^{K} \beta_k \cdot (x_{ns,i,k})\right]}. \tag{40}$$

As written, this is a *scale dependent MNL model* when there is a common choice set size across all choice tasks and individuals; it is a *context dependent MNL model* when there is more than one choice set size (i.e. car availability may vary across individuals and thus J_n would not be a constant across the dataset). However, by the equivalence of context dependent MNL models and context dependent Luce models, it is also the latter. In particular, with the former (MNL) interpretation, J_n is the scale (1/variance) of the generating Gumbel distribution.

Table 3 summarises the key difference between the models that we evaluate against each other. An LBA with Fréchet-distributed drift rates, zero start point variability and the drift rate function specified by Equation 39 gives us a model equivalent to MNL. The MLBA model that we test assumes truncated normal drift rate distributions; $A \neq 0$; and weights in the value functions (Equation 31) to translate from objective to subjective values.

TABLE 3 : Context dependent LBA models

		Drift rate distribution, start point variability	
		Fréchet (Eq.37)	Truncated Normal (Eq.28)
		$A=0$	$A=1$
Drift rate function	Eq.31	Specification 1	MLBA
	Eq.39	MNL	Specification 2
	Eq.41	RRM	Specification 3

We can also consider a (LBA) model equivalent to Chorus (2010)'s random regret minimisation model (RRM), with drift rates equal to the exponential of the value function:

$$v_{ns,j} = -\sum_{i \neq j}^{J_n} \sum_{k=1}^{K} ln(1 + exp(\beta_k \cdot (x_{ns,i,k} - x_{ns,j,k}))). \tag{41}$$

In Section 3.6, we test whether it is the *drift rate distribution* (i.e. whether we use truncated normal drift rate distributions or Fréchet-distributed drift rates with start rate parameter $A = 0$) or the *drift rate function* that drives differences between MNL, RRM and MLBA. This test is possible as both MNL and RRM use Equation 38 to generate probabilities of different alternatives and we look at three additional specifications with features of both MLBA and MNL/RRM:

- Specification 1 is effectively a 'context-dependent' logit model. It uses drift rates equal to the exponential of the value function specified by Equation 31 together with Fréchet distributions (Equation 37) and zero start point variability.
- Specification 2 is effectively an MLBA model but with weights $w_{x_{ns,ij,k}} = 1$. It uses truncated normal distributions and ensures mean drifts that are positive through the use of Equation 29 together with values that are generated by the linear value function (Equation 39).
- Specification 3: is a multi-attribute LBA which still treats positive and negative attribute differences differently, but at the cost of no additional parameters. It sets mean drift rates based on Equations 29 and 41.

If our results align with those of Terry et al. (2015), where the type of distribution had little impact on model fit, and MLBA outperforms MNL and RRM, we would also expect to see specification 1 outperform specifications 2 and 3.

3. EMPIRICAL APPLICATIONS ON REVEALED AND STATED CHOICE DATA

In this section, we present empirical results from testing DFT and MLBA on three different datasets, two from stated choice (SC) surveys and one from a revealed preference (RP) survey, where the latter is the first DFT/MLBA application to RP data. For DFT, we investigate the empirical identification points detailed in Table 1. We also evaluate the impact of fixing further DFT/MLBA parameters that are insignificant in the first applications of the models, given that, in the context of choice-only data, some of the process parameters may become unimportant. We also compare the estimation results to typical MNL and RRM models. We finally present an empirical comparison between the different existing specifications of DFT and our proposed new scaling approach, as well as a section testing the impact of different distributions and value functions for LBA, as discussed in Section 2.3.3.

The estimation of DFT and MLBA remains a non-trivial computational task even with our methodological developments, and efficient implementation as well as good starting values are essential. In all of our applications, we use the R packages maxLik (Henningsen and Toomet, 2011) and *Apollo* (Hess and Palma, 2019) for estimation of the likelihood function and the RCPP package together with the Armadillo C++ linear algebra library for fast calculation of the matrices that are required for a DFT model to calculate the choice probabilities. (Eddelbuettel et al., 2011; Sanderson and Curtin, 2016). Additionally, we use an initial parameter search algorithm based on the heuristic for non-linear global optimisation developed by Bierlaire et al. (2010) in an attempt to reduce the risk of convergence to poor local optima or an excessively long estimation process.

3.1. First stated choice survey

Our first dataset is a subset of the data from the Danish value of time study (Fosgerau, 2006). This dataset comes from a typical stated choice survey, where 545 participants faced a total of 4,214 choices between them. The choices were for car drivers and specifically the choice between two different routes, described only by travel cost (in Danish krone, DKK) and travel time (in minutes), where one route is cheaper, but the other is faster. The aim of such a setup is to understand trade-offs between time and money, leading to estimates of the value of travel time (VTT). While very simplistic in nature, this type of dataset is a useful first step in moving from the abstract settings in mathematical psychology towards more complex choices in a transport setting. In all models, we focussed on the time and cost attributes after earlier results confirmed there was no left-right bias that would require the inclusion of alternative specific constants.

Table 4 shows the results for this dataset. Where appropriate, we used the constraints from Table 2 but then report the actual transformed estimates in Table 4, along with the transformed standard errors, obtained using the Delta method (cf. Daly et al., 2012).

We first have two MNL models. We initially have a purely linear specification, with the utility for alternative j for individual n in choice task s is:

$$U_{nsj} = \delta_j + \beta_{TT} \cdot TT_{nsj} + \beta_F \cdot F_{nsj} + \varepsilon_{nsj}, \tag{42}$$

where the β parameters are taste coefficients to be estimated and TT_{nsj} and F_{nsj} give the travel time and fare of the alternative. For the second MNL model, we add terms for the logarithm of time and cost:

$$U_{nsj} = \delta_j + \beta_{TT} \cdot TT_{nsj} + \beta_F \cdot F_{nsj} + \varepsilon_{nsj} + \beta_{LTT} \cdot log(TT_{nsj}) + \beta_{LF} \cdot log(F_{nsj}) + \varepsilon_{nsj}. \tag{43}$$

This latter model offers a significant improvement in fit over the first model, and all four coefficients remain negative, where the significant estimates for the log-time (β_{LTT}) and log-fare (β_{LF}) parameters indicate non-linear sensitivities.

Whilst a number of different parameters within DFT could be fixed to solve the theoretical overspecification issue identified in Section 2.2.2.2, we follow restriction 2 from Table 1 where a priori information about the other attributes is not required. We initially evaluated different DFT models to test the impact of removing the effect of the feedback matrix (model 2 compared to model 1).

The level of competition between alternatives is dependent on their similarity (the Euclidean distance between them) and ϕ_1. In cases with three or more alternatives, context effects are predicted for certain values of the parameter. When there are only two alternatives (as in the Danish dataset), a significant estimate for ϕ_1 means a reduction in the overall preference for more similar pairs of alternatives, resulting in less deterministic choices. Additionally, ϕ_2, the memory parameter, has little meaning when the sequence of attribute attendance is not known. It however also contributes to the level of competition between alternatives as a value of $\phi_2 = 0$ results in the value of ϕ_1 having no impact (see Equation 2). For this dataset, the use of an identity matrix in place of the feedback matrix (model 2) results in an insubstantial loss of model fit. Additionally, we find that if we fix $\sigma_\varepsilon = 0$ (meaning that all of the variation in the DFT model comes from the random attribute attendance), there is again no substantial loss of model fit. Note that with $\sigma_\varepsilon = 0$, we require a further normalisation (see Section 2.2.2.2). Given that we have negative coefficients for the scaling parameters, we follow restriction 3 from Table 4 by setting the first β to -1. As a consequence of the the noise parameter having an insignificant impact on model fit, fixing it to a value of 1 (as configured in models 1 and 2) results in insignificant parameter estimates for the β-coefficients. By appropriately fixing $\sigma_\varepsilon = 0$ and also fixing the first β, we recover significant parameter estimates elsewhere.

Similar to DFT, MLBA has many parameters that have little interpretable output when an analyst has choice data, only, and thus no additional psychometric or process data (such as response time). For example, a decision-maker could make a choice quickly because there is a small difference between the start point and threshold or because there is a large difference between one or more drift rates. Consequently, if we only have choice data some of the MLBA parameters become confounded (see Section 2.3.2.3). As with DFT, we initially test MLBA using a full specification,

TABLE 4 : Estimation results and identifications tests on the first SC dataset, with the bold print highlighting the best performing version (in terms of BIC) of each model

Model		MNL		DFT			MLBA		
Version		1	**2**	1	2	**3**	1	2	**3**
Free Pars.		3	**5**	6	4	**3**	7	6	**6**
Log-likelihood		-2,301.25	**-2,211.69**	-2,015.57	-2,016.25	**-2,016.28**	-2,005.34	-2,036.36	**-2,005.92**
BIC		4,627.54	**4,465.12**	4,081.23	4,065.88	**4,057.60**	4,069.11	4,122.80	**4,061.92**
β_{TT}	est.	-0.1938	-0.1590	-0.3669	-1.2413	-1.0000	-3.2391	-3.0371	-3.2455
	r. t-rat.	-13.53	-8.85	-1.92	-1.31	**fixed**	-16.23	-15.57	-7.18
β_F	est.	-2.4079	-1.7637	-5.8257	-19.8310	-16.0515	-51.2265	-45.1739	-51.2492
	r. t-rat.	-13.51	-9.20	-1.84	-1.28	-25.89	-17.63	-15.18	-7.68
β_{LTT}	est.		-1.0326						
	r. t-rat.		-2.70						
β_{LF}	est.		-1.9001						
	r. t-rat.		-5.39						
δ_1	est.	0.0264	0.0326	0.2004	1.3972	1.1068	0.5037	0.4309	0.5193
	r. t-rat.	1.08	1.29	1.42	1.15	2.51	0.46	1.79	3.36
ϕ_1	est.			0.6172	0.0000	0.0000			
	r. t-rat.			2.06	**fixed**	**fixed**			
ϕ_2	est.			-0.3293	0.0000	0.0000			
	r. t-rat.			-9.77	**fixed**	**fixed**			
σ_ε	est.			1.0000	1.0000	0.0000			
	r. t-rat.			**fixed**	**fixed**	**fixed**			
τ	est.			44.1810	6.6153	6.4712			
	r. t-rat. (*vs* 1)			1.36	11.95	15.25			
χ	est.						1.1936	2.0000	1.1618
	r. t-rat. (*vs* 1)						8.98	**fixed**	6.29
I_0	est.						2.1850	38.8858	1.5062
	r. t-rat.						3.61	20.51	1.81
λ_1	est.						0.0007	0.0025	0.0000
	r. t-rat.						11.51	10.77	**fixed**
λ_2	est.						0.1833	0.0374	0.1965
	r. t-rat.						5.22	12.93	2.93

which implies only setting the values of the start parameter A and the drift rate standard deviation σ_ε to 1.

In model 2, we fix the threshold parameter χ, which is a common approach in mathematical psychology (Trueblood et al., 2014; Cohen et al., 2017; Cataldo and Cohen, 2018) (fixing it to a value of 2 as done in the original MLBA paper Trueblood et al. 2014), but find that this is not appropriate in this case, leading to a substantial loss of fit. On the other hand, our initial estimate for λ_1 is so close to zero that fixing it to a value of zero does not lead to any significant loss of fit (model 3). A value of $\lambda_1 = 0$ implies $w_{x_{ns,ij,k}} = 1$ resulting in mean drift rates that are linear in the positive differences of attributes $x_{ns,j,k} - x_{ns,i,k}$. Previous work where both λ parameters approached zero resulted in applications of MLBA resorting to different valuation and weighting functions (Cohen et al., 2017). In our case, however, only one λ approaches zero, simply meaning that there is a small amount of non-linear sensitivity to attributes.

In terms of model performance using BIC, we see that DFT and MLBA both outperform MNL. The difference in fit between DFT and MLBA is much smaller than between these two models and MNL, with a slight advantage for DFT in terms of BIC, and for MLBA in terms of log-likelihood.

3.2. Second stated choice survey

The second stated choice dataset we consider has a total of 368 participants, each completing 10 choice tasks resulting in 3,680 choices. The participants are all public transport commuters living in the UK. Each task involves the choice between an invariant reference trip and two hypothetical alternatives, where each of the three alternatives is described by travel time (*TT* in Table 5), fare (*LF*), rate of crowded trips, rate of delays (*CR* and *RE*, respectively, both out of 10 trips), the average length of delays (entered into models both as the average extent of delays, *RA*, and as the expected delay, *RB*, by multiplying the length of delays by the rate of delays) and the provision of a delay information service (none used as the base, with parameters for a charged, *ICH*, and free, *IFR*, service). Following earlier results by Hess and Stathopoulos (2013), we applied a log-transform to the fare attribute (described as *LF*).

Table 5 shows the results for the second SC dataset. In the presence of three alternatives, we can now include a RRM model (Chorus, 2010) alongside MNL, where we see fairly similar performance for these two models, with a slight advantage for MNL. All parameters have the expected sign in these models, and we are also able to include two alternative specific constants (ASCs), which results in improvements in log-likelihood of 46 and 61 units respectively for DFT and MLBA (demonstrating the benefits of the new scaling system for DFT from Section 2.2.2.1 and the new developments for MLBA from Section 2.3.2).

For DFT, we follow the same specification tests as on our first SC dataset. However, this time constraining the feedback matrix to be an identity matrix (as in model 2) leads to a significant drop in model fit.[9] This is a direct result of having more than two alternatives, meaning that the feedback matrix is needed for capturing the different similarities between alternatives.

For MLBA, we again show that setting $\chi = 2$ leads to a loss of fit for model 2. Nonetheless, we can set $\chi = 1$ and $I_0 = 0$ (model 3) without a significant loss of fit. With these constraints, we obtain larger values for λ_2 than for λ_1, which means that a greater importance weight is given to positive attribute differences $x_{ns,j,k} \geq x_{ns,i,k}$ compared to negative ones, $x_{ns,j,k} < x_{ns,i,k}$ in the estimation of the mean drift rates. This results in subjective differences corresponding approximately to Graph D in Figure 6: meaning more non-linearity than was observed in SC-1 particularly around differences close to zero.

[9] Note that in this case, a model with $\sigma_\varepsilon = 0$ (as demonstrated to be effective for the first dataset) has a log-likelihood of -3,306.61, thus results in worse fit than a model with $\sigma_\varepsilon = 1$.

TABLE 5 : Estimation results and identifications tests on the second SC dataset, with the bold print highlighting the best performing version (in terms of BIC) of each model

Model		MNL	RRM	DFT		MLBA		
Version		1	1	**1**	2	1	2	**3**
Free Pars.		10	10	**13**	11	14	13	**12**
Log-likelihood		-3,360.43	-3,363.91	**-3,299.82**	-3,327.28	-3,321.57	-3,329.24	**-3,322.26**
BIC		6,802.97	6,809.92	**6,706.38**	6,744.88	6,758.09	6,765.21	**6,743.04**
β_{TT}	est.	-0.0471	-0.0320	-0.3746	-0.1321	-0.0675	-0.0369	-0.0753
	r. t-rat.	-9.50	-9.58	-3.28	-5.20	-2.97	-4.69	-3.76
β_{LF}	est.	-5.9990	-4.1090	-53.9106	-17.5870	-12.2432	-7.0789	-14.0066
	r. t-rat.	-18.87	-17.66	-3.27	-5.69	-1.26	-3.85	-4.79
β_{CR}	est.	-0.2230	-0.1457	-1.7210	-0.6618	-0.3088	-0.1918	-0.3600
	r. t-rat.	-8.58	-8.59	-3.18	-4.45	-3.08	-6.11	-3.92
β_{RA}	est.	-0.1870	-0.1212	-1.2241	-0.5488	-0.1772	-0.0859	-0.1864
	r. t-rat.	-5.96	-5.82	-2.80	-2.98	-4.17	-4.60	-2.08
β_{RE}	est.	-0.0619	-0.0441	-0.7323	-0.1750	-0.1429	-0.0693	-0.1678
	r. t-rat.	-2.64	-3.68	-2.15	-1.15	-3.30	-10.45	-1.22
β_{RB}	est.	-0.0293	-0.0186	-0.1108	-0.0683	-0.0193	-0.0128	-0.0220
	r. t-rat.	-3.25	-3.06	-1.64	-1.77	-0.97	-4.61	-3.47
β_{ICH}	est.	-0.0910	-0.0510	-0.0009	-0.1825	-0.0545	-0.0276	-0.0669
	r. t-rat.	-1.13	-0.95	0.00	-0.98	-1.49	-4.14	-1.67
β_{IFR}	est.	0.3305	0.2179	1.8340	0.7774	0.3023	0.1573	0.3180
	r. t-rat.	4.95	4.85	3.15	4.18	2.54	2.13	4.93
δ_1	est.	0.3902	0.2730	2.2529	2.0355	1.1832	0.6074	1.3043
	r. t-rat.	5.85	4.17	4.86	7.82	1.24	11.51	5.41
δ_2	est.	0.1633	0.1656	1.0726	0.6109	0.3762	0.1741	0.3543
	r. t-rat.	3.30	3.38	3.01	3.30	0.39	4.53	1.13
ϕ_1	est.			0.0003	0.0000			
	r. t-rat.			1.62	**fixed**			
ϕ_2	est.			-0.5656	0.0000			
	r. t-rat.			-4.91	**fixed**			
σ_ε	est.			1.0000	1.0000			
	r. t-rat.			**fixed**	**fixed**			
τ	est.			5.1858	7.5332			
	r. t-rat. (*vs* 1)			10.31	7.65			
χ	est.					1.0007	2.0000	1.0000
	r. t-rat. (*vs* 1)					1.98	**fixed**	**fixed**
I_0	est.					0.6679	2.4447	0.0000
	r. t-rat.					2.37	18.52	**fixed**
λ_1	est.					0.1023	0.2873	0.0959
	r. t-rat.					0.52	3.95	4.09
λ_2	est.					0.7971	5.2198	1.1646
	r. t-rat.					3.29	15.21	3.78

In terms of model performance, we see that each of DFT and MLBA again outperform MNL and also RRM (which, for this dataset, is not identical to MNL). DFT fits better than MLBA, potentially because it is better able to deal with the differential competition between the three alternatives (in contrast to the earlier binary dataset).

3.3. RP data

Whilst both DFT and MLBA have been applied extensively on experimental data and have been shown to accurately explain choices in stated preference surveys, we do not know of applications to revealed preference (RP) data. In this section, we fit MNL, RRM, DFT and MLBA models to our full RP dataset and provide elasticities as well as out-of-sample predictions.

Our RP data comes from the national UK value of travel time study (Arup, ITS Leeds and Accent, 2015). Questionnaires were completed by 2,646 individuals travelling by train from Birmingham, Stoke or Peterborough to London. After extensive data cleaning (see page 164 of Arup, ITS Leeds and Accent 2015), 725 observations were left, with either one or two observations for each of the 578 individuals. This means we have data that is panel data for some individuals, but not others. However, this only impacts the calculation of standard errors and does not effect the results of the models. For every decision recorded, the available alternatives are one or two of Chiltern railways, Northern rail and Midlands railways as well as one of Virgin Trains and East Coast. Travel time, travel cost and the time interval between services (headway) were used to describe the alternatives.

We run a basic MNL model with its specification based on the model developed by Arup, ITS Leeds and Accent (2015); we estimate different travel time coefficients for each of four groups, which we now describe. Individuals are first segmented by travel purpose (employees' business, commute (TT_C in Table 6) or 'other non-work' (TT_O)). Individuals on employees' business where further segmented into those who were very sure (TT_{EB1}) and those who were quite sure (TT_{EB2}) about the attributes of the unchosen alternatives. We also estimated parameters associated with travel cost (TC), and headway (HW). For all three attributes, log values are used (Arup, ITS Leeds and Accent, 2015). Additionally, Arup, ITS Leeds and Accent (2015) use three alternative specific constants for train services run by Chiltern railways (δ_C), Midlands railways (δ_M) and Northern rail (δ_N). Finally, two parameters are incorporated to capture income effects. Travel time coefficients (β_{TT_n}) are calculated for each individual n:

$$\beta_{TT_n} = \beta_{TT_{i,n}} \cdot \left({RI_n}^{\lambda_{inc}} \cdot (1 - z_{miss,n}) + \lambda_{miss} \cdot z_{miss,n} \right) \tag{44}$$

where $\beta_{TT_{i,n}}$ is a travel time coefficient depending on the individual's trip purpose, RI_n is the relative income of the individual, λ_{inc} is an income elasticity on the time sensitivity and λ_{miss} is a multiplier on the time sensitivity used only if the individual did not provide their income in the questionnaire (in which case the dummy variable $z_{miss,n} = 1$). Given the developments of the new scaling system for DFT (Section 2.2.2.1) and scaling parameters for MLBA (Section 2.3.2.2), individual travel time coefficients can be estimated equivalently for DFT and MLBA. Table 6 provides model estimates for these parameters under MNL, RRM, DFT and MLBA.

TABLE 6 : Results, estimates and robust t-ratios from MNL, RRM, DFT and MLBA models on the RP dataset

Model		MNL	RRM	DFT		MLBA		
Version		1	1	1	**2**	1	2	**3**
Free Pars.		11	11	14	**12**	15	14	**12**
Log-likelihood		-370.05	-371.04	-362.53	**-363.31**	-347.84	-351.13	**-351.23**
BIC		812.54	814.52	817.26	**805.66**	794.47	794.48	**781.49**
TT_C	est.	-4.4541	-4.3583	-5.4056	-5.2198	-34.0634	-26.1385	-25.9484
	rob. t-rat.	-3.88	-3.98	-1.43	-2.99	-3.34	-2.57	-1.88
TT_O	est.	-2.0021	-1.7285	-2.3913	-2.4187	-7.7031	-4.6251	-4.3787
	rob. t-rat.	-2.46	-2.40	-1.43	-2.00	-2.02	-3.24	-2.68
TT_{EB1}	est.	-3.7769	-3.5093	-3.6849	-3.6949	-12.3531	-8.2263	-7.8413
	rob. t-rat.	-4.63	-5.04	-1.73	-2.47	-2.37	-5.42	-5.30
TT_{EB2}	est.	-5.7016	-5.2096	-6.2639	-6.1879	-16.7342	-11.6216	-11.2330
	rob. t-rat.	-7.10	-7.45	-1.65	-2.28	-2.95	-7.77	-7.62
TC	est.	-2.2127	-1.8575	-1.9096	-1.8503	-7.5232	-4.8263	-4.8143
	rob. t-rat.	-8.52	-7.96	-2.53	-2.99	-8.46	-6.42	-2.83
HW	est.	-0.1267	-0.1343	-0.1083	-0.1037	-0.0065	-0.0952	-0.1066
	rob. t-rat.	-0.64	-0.84	-0.75	-0.73	-0.08	-0.98	-1.05
δ_C	est.	0.7549	0.6966	2.5746	3.2934	2.0538	1.5538	1.4988
	rob. t-rat.	2.75	2.61	1.58	3.24	2.13	4.86	2.87
δ_M	est.	-0.4882	-0.3740	-0.5472	-0.4284	-0.7663	-0.5365	-0.5925
	rob. t-rat.	-1.86	-1.55	-0.46	-0.41	-1.11	-2.31	-1.24
δ_N	est.	-0.4879	-0.4669	-0.3529	-0.3703	-0.5019	-0.4477	-0.5120
	rob. t-rat.	-1.65	-1.71	-0.30	-0.31	-0.51	-8.85	-1.21
λ_{inc}	est.	0.4563	0.4533	0.5594	0.5411	0.5249	0.5690	0.5849
	rob. t-rat.	4.43	4.37	4.48	4.48	4.10	5.17	3.92
λ_{miss}	est.	0.4844	0.4564	0.8022	0.7948	2.7332	0.6065	0.5829
	rob. t-rat.	1.13	1.04	1.04	1.09	2.29	0.66	0.73
ϕ_1	est.			1.5331	0.0000			
	r. t-rat.			0.58	**fixed**			
ϕ_2	est.			-0.0864	0.0000			
	r. t-rat.			-1.72	**fixed**			
σ_ε	est.			1.0000	1.0000			
	r. t-rat.			**fixed**	**fixed**			
τ	est.			8.2059	8.1592			
	r. t-rat. (*vs* 1)			3.22	3.49			
χ	est.					1.5285	2.1366	2.0000
	r. t-rat. (*vs* 1)					43.18	1.71	**fixed**
I_0	est.					-1.2501	-0.1195	0.0000
	r. t-rat.					-1.28	-4.23	**fixed**
λ_1	est.					0.0774	0.1043	0.1046
	r. t-rat.					6.10	9.07	9.86
λ_2	est.					16.7063	Inf	Inf
	r. t-rat.					1.58	**fixed**	**fixed**

For DFT, we again test two versions. With 118 out of 725 observations having three alternatives available and the rest having only two alternatives available, it is unsurprising that, in line with the results from the first SC dataset, the effect of the feedback parameters being removed (DFT model 2 relative to DFT model 1) has little impact on the log-likelihood.

For MLBA, we see that fixing one of the similarity parameters, λ_2, to infinity (which results in the corresponding weight, $w_{x_{ns,ij,k}} = 0$, when $x_{ns,j,k} < x_{ns,i,k}$) has no impact on model fit (model 2 compared to model 1). This implies that the model fits the data well with only positive differences included in the drift rates. Additionally fixing both I_0 and χ results in an insignificant impact on model fit with a lower BIC value obtained for model 3 compared to model 2.

For the model fit as measured by BIC, MLBA has a lower value than DFT, with both DFT and MLBA outperforming MNL and RRM.

With a view to not just focussing on model fit, Table 7 contrasts the cost and time elasticities on the RP data for the four models. We see that the elasticities for MNL and RRM are quite similar to each other. MLBA obtains visibly higher time and cost elasticities than MNL and RRM. For DFT, the cost and time elasticities are between those for MNL/RRM and MLBA. These results again show that DFT and MLBA offer more significant departures from standard RUM models than RRM does.

TABLE 7 : Cost and time elasticities on RP data

	elasticities	
	cost	time
MNL	-0.537	-0.933
RRM	-0.497	-0.933
DFT	-0.604	-1.017
MLBA	-0.670	-1.278

We finally test all four models for their ability to make out-of-sample predictions. For each of the five data subsets, we take choices corresponding to a random 80% of the individuals in the data to be used for estimation, with the remaining 20% used for validation. We fit each model to each estimation subset and then calculate log-likelihoods for the remaining 20% of the data using the parameter estimates obtained for the first 80%. Table 8 gives the log-likelihoods of the estimation and validation subsets of the data under each model.

We see that DFT and MLBA outperform MNL and RRM across all five subsamples in both estimation and performance on the holdout sample except for DFT in holdout sample 4 and MLBA in holdout sample 1. MNL outperforms RRM in estimation and holdout across all samples, while MLBA typically outperforms DFT. Overall, these findings confirm the results on the full sample.

TABLE 8 : Out-of-sample estimation and holdout log-likelihoods for the RP data

	MNL (11 pars)		RRM (11 pars)	
	estimated	forecast	estimated	forecast
Dataset 1	-302.88	-68.92	-303.22	-69.29
Dataset 2	-298.59	-72.76	-298.64	-73.76
Dataset 3	-296.70	-75.08	-295.66	-77.28
Dataset 4	-302.29	-68.18	-303.49	-68.17
Dataset 5	-296.64	-75.74	-297.66	-75.71
	DFT (12 pars)		MLBA (12 pars)	
	estimated	forecast	estimated	forecast
Dataset 1	-296.90	-67.80	-285.20	-67.93
Dataset 2	-293.80	-70.63	-282.32	-69.93
Dataset 3	-293.12	-71.77	-281.37	-71.38
Dataset 4	-295.41	-68.29	-286.19	-65.94
Dataset 5	-293.23	-72.90	-282.06	-71.76

3.4. Comparison of results

3.4.1. Model fit comparisons

To summarise the results, Table 9 shows the BIC for the final recommended specification for each model type on each dataset. We see that DFT and MLBA consistently offer better performance than MNL and RRM. While DFT marginally outperforms MLBA on the Danish SC data, the differences are more substantial on the remaining two datasets, with DFT performing best on the UK SC data and MLBA best on the RP data.

TABLE 9 : Model fit (BIC) comparison across models and datasets

	MNL	RRM	DFT	MLBA
Danish (SC-1)	4,465.12	4,465.12	4,057.60	4,061.92
UK (SC-2)	6,802.97	6,809.92	6,706.38	6,743.04
RP	812.54	814.52	805.66	781.49

We next consider choice task features that drive the differences in the above model fits. Table 10 shows how the difference in the fare between the two alternatives impacts the log-likelihoods of the respective models for the Danish SC dataset. In this case, we observe a clear difference for choice tasks with a difference of less than 1 DKK, with DFT and MLBA outperforming MNL. For larger differences in fare, there are considerably smaller differences in model fit across the models. This suggests that DFT and MLBA perform better than MNL as a result of better predicting choices in situations where the alternatives are more similar to each other. This is also illustrated in Figure 2, with the top left panel demonstrating that there are considerably more choice tasks in which the difference in fare is less than 1 DKK (represented by grey crosses) above the grey diagonal line (the latter indicates when DFT and MNL give the same probability). In particular, these choice tasks have a greater spread of predicted probabilities under a DFT model, again suggesting that DFT differentiates between similar alternatives more than MNL does.

TABLE 10 : Average log-likelihood contribution per observation in SC-1, with the choice tasks categorised by the difference in fare between the two alternatives.

fare	number of	Dataset (SC-1)		
difference (DKK)	choice tasks	MNL-2	DFT-3	MLBA-3
all	4214	-0.525	-0.478	-0.476
0-1	805	-0.531	-0.372	-0.369
1-2	690	-0.641	-0.633	-0.632
2-4	946	-0.599	-0.572	-0.571
4-10	862	-0.504	-0.492	-0.491
10+	911	-0.373	-0.346	-0.340

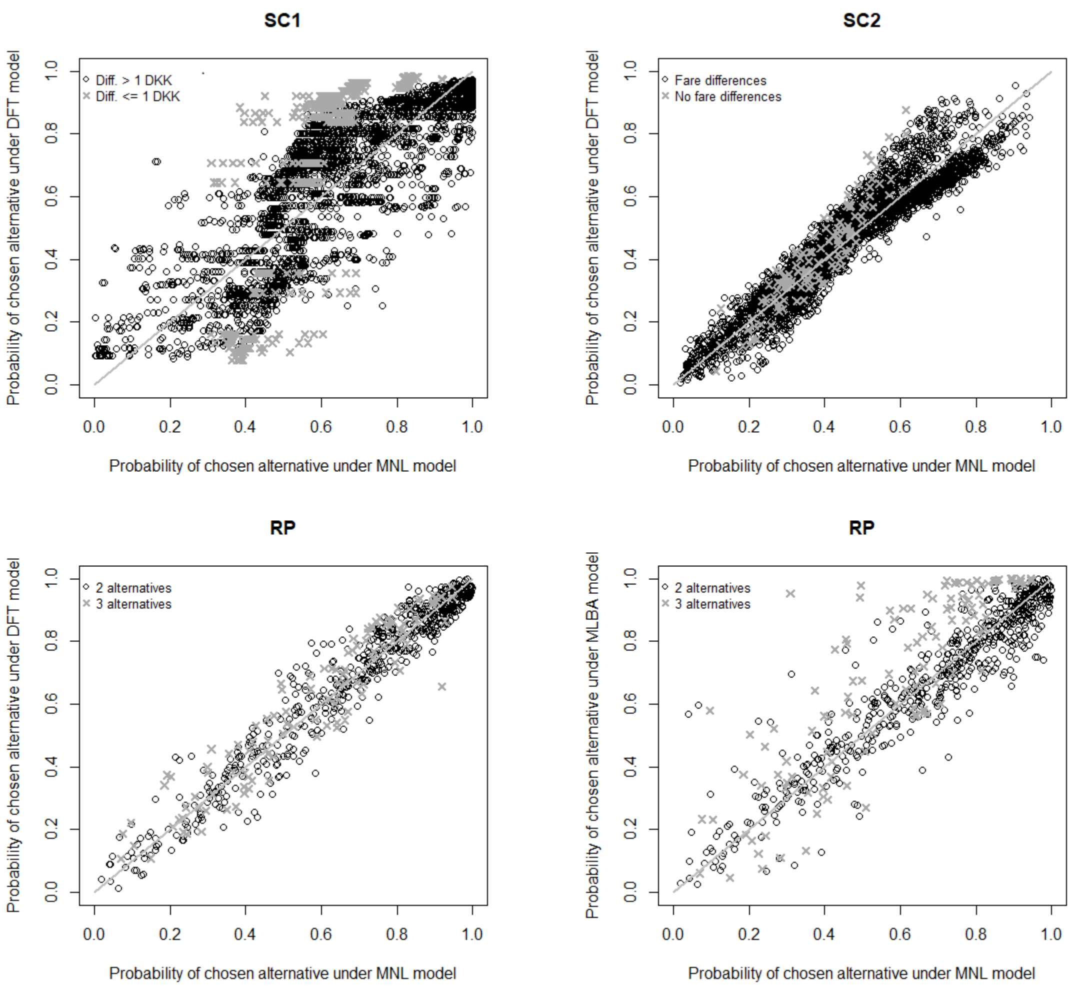
This also appears to be the case for SC-2, with Table 11 showing that there are greater differences in the average log-likelihood contribution for choice tasks with smaller fare differences between the alternatives. However, the difference is less clear cut than for SC-1, with the top right panel of Figure 2 showing only small differences between MNL and DFT.

TABLE 11 : Average log-likelihood contribution per observation in SC-2, with the choice tasks categorised by the difference in fare between the two alternatives. The choice tasks are categorised by the summed percentage of fare differences between pairs of alternatives.

fare	number of	Dataset (SC-2)			
difference	choice tasks	MNL-1	RRM-1	DFT-1	MLBA-3
all	3680	-0.913	-0.914	-0.897	-0.903
0%	127	-1.027	-1.021	-0.977	-1.000
20%	528	-0.941	-0.938	-0.904	-0.920
40%	1425	-0.945	-0.948	-0.928	-0.932
60%	622	-0.874	-0.874	-0.869	-0.874
80%	978	-0.862	-0.864	-0.854	-0.857

For the RP dataset, the driver of the difference in model fit is the relative performance on choice tasks with two or three alternatives. The models have nearly identical log-likelihood contributions for choice tasks with two alternatives (see Table 12). However, DFT and MLBA (in particular) show much smaller contributions to the overall log-likelihood for choice tasks with three alternatives. This is also demonstrated in the lower panels of Figure 2, which shows a cluster of very-well predicted choices especially by MLBA when there are three alternatives.

FIGURE 2 : Comparisons of the predicted probability of the (observed) chosen alternatives under different models for the different datasets. The upper left panel shows a comparison of MNL and DFT for SC-1, with the grey crosses showing choice tasks in which the difference in fare between the alternatives was less than 1 DKK. The upper right panel shows a comparison of MNL and DFT for SC-2, with the grey crosses showing choice tasks where there was no difference in fare across the three alternatives. The lower panels show comparisons of DFT and MLBA against MNL, respectively, for the RP dataset. Grey crosses show choice tasks where there are three alternatives.



3.4.2. Model output comparisons

We now pay closer attention to the model outputs. An additional benefit of the new scaling method we use for DFT is that it allows us to more directly compare parameter estimates across different models, notwithstanding the different meaning of the parameters. This is possible as a result of the new specifications of both MLBA and DFT having attribute-specific scaling coefficients (see Sections 2.2.2.1 and 2.3.2.2, respectively), which have a role analogous to marginal utility coefficients in RUM models. Although these scaling coefficients cannot be directly translated into measures such as the value of travel time, we can calculate ‘relative importance of travel time

TABLE 12 : Average log-likelihood contribution per observation in the RP dataset, with the choice tasks categorised by whether they have 2 or 3 possible alternatives.

number of alternatives	number of choice tasks	Dataset (RP)			
		MNL-1	RRM-1	DFT-1	MLBA-3
all	725	-0.510	-0.512	-0.501	-0.484
2	607	-0.479	-0.479	-0.479	-0.475
3	118	-0.671	-0.682	-0.617	-0.533

with respect to fare' and make comparisons across models (as long as we ensure that the units are equivalent across models). In Table 13, we set the calculated MNL ratios of time and fare parameters to a base rate of 1 (with the rates being based on the MNL value for commuters in the RP dataset). Consequently we can compare whether DFT and MLBA assign more or less importance to travel time with respect to fare.

TABLE 13 : The relative importance of travel time compared to fare parameter coefficients across different models in comparison to MNL, with standard errors provided in parenthesis (calculated with the Delta method, as discussed by Daly et al. 2012).

		MNL	RRM	DFT	MLBA
SP	Danish (SC-1)	1.000 (0.047)	1.000 (0.047)	0.774 (0.019)	0.787 (0.035)
	UK (SC-2)	1.000 (0.105)	0.992 (0.103)	0.885 (0.094)	0.685 (0.090)
RP	Commuters	1.000 (0.263)	1.166 (0.300)	1.401 (0.462)	2.678 (1.131)
	Other Non-Work	0.449 (0.174)	0.462 (0.186)	0.649 (0.193)	0.451 (0.131)
	Employees' Business 1	0.848 (0.180)	0.992 (0.201)	0.992 (0.197)	0.809 (0.330)
	Employees' Business 2	1.280 (0.173)	1.393 (0.202)	1.661 (0.273)	1.159 (0.465)

Across the SP datasets, it appears that MNL tends to assign higher importance to travel time with respect to fare relative to DFT and MLBA, with significant differences found for the Danish dataset. RRM always estimates similar ratios to MNL, while DFT has some similar values to MLBA, with key exceptions being the UK data and commuters in the RP data, for which DFT is more similar to MNL.

3.5. DFT model specifications: alternative scaling methods or weights

In this section, we compare our new method (see Section 2.2.2.1) to scaling methods that have been used in previous DFT applications, where we do this for both SP datasets. As DFT is scale-variant (see Section 2.2.2.1), a failure to appropriately adjust the attribute values can result in inferior model fit (see Table 5 of Hancock et al. 2018). Crucially, all previous methods rely on a priori knowledge of the directionality of the attributes, whereas the new method does not. The different scaling methods tested are listed below.

1. Unity-based normalisation, as used by Berkowitsch et al. (2014), where we rescale the attribute-levels of each attribute, separately, to the range 0 to 1. For an undesirable attribute k, which has value $x_{i,k}$ for alternative i and a set of values X_k across the alternatives in all choice sets in the dataset, we define a normalised attribute value $x'_k = 1 - \frac{x_k - min(X_k)}{max(X_k) - min(X_k)}$, with desirable attribute values being normalised as $x'_k = \frac{x_k - min(X_k)}{max(X_k) - min(X_k)}$.

2. No scaling method other than taking the negative value for all ‘negative’ attributes (as DFT can only capture ‘positive’ effects of attributes as the relative importance weights must be positive - see Section 4.3.1 of Hancock et al. (2018) for an illustration of the results of failing to do this for DFT models)

3. Standard score normalisation, as previously found to be effective for DFT (see results in Hancock et al. 2018), where for undesirable attribute k, which has value $x_{i,k}$ for alternative i and a set of values X_k across the alternatives in all choice sets in the dataset, we define new attribute values $x'_k = -\frac{x_k - mean(X_k)}{sd(X_k)}$. For desirable attributes, $x'_k = \frac{x_k - mean(X_k)}{sd(X_k)}$.

4. Minimum rescaling (dividing each attribute by the smallest value for that attribute across the choice set), as previously shown to be effective for a previous version of MLBA (Trueblood et al., 2013a)

5. Maximum rescaling (dividing each attribute by the largest value for that attribute across the choice set), as previously shown to be effective for a previous version of MLBA (Trueblood et al., 2013a)

6. The new method detailed in Section 2.2.2.1, which removes the scale-variant nature of DFT.

7. The new method detailed in Section 2.2.2.1, with estimated values for attribute scaling coefficients and attribute importance weights.

To estimate the models, we require attribute matrices M_{ns} and attention weights w_k and scaling parameters β_k for attribute k. For methods 1-5, M_{ns} is adjusted depending on the method, w_k are estimated and β_k are fixed to a value of 1. For scaling method 6 and 7, M_{ns} is unchanged and β_k are estimated. The weights w_k are fixed to $1/K$ (where K is the number of attributes) for scaling method 6, but are estimated for scaling method 7. All models adhere to the model identification restrictions detailed in Table 1.

For both datasets, it appears that our new method (model 6 in Table 14) has the best model fit if either the scales or the weights are fixed parameters. This result holds regardless of whether we include DFT’s feedback matrix. Scaling method 6 appears to better capture the impact of the feedback matrix for the UK data, resulting in an improvement of 27 log-likelihood units, whereas this improvement is much smaller with the other scaling methods corresponding to models 1-5. On the other hand, with the Danish data, the feedback matrix is needed for some of the other scaling methods to obtain model fit more in line with the new scaling. Whilst these results support the incorporation of our new scaling method in DFT, the log-likelihood gain here is dataset-dependent. This means that the main benefit of incorporating this method relates to model configuration (i.e. a

TABLE 14 : Log-likelihood (LL) and BIC values for DFT models with different types of scaling methods for the two stated choice datasets.

			Danish (SC-1)					
			with feedback parameters			without feedback parameters		
No.	wt est.	scale est.	free pars.	LL	BIC	free pars.	LL	BIC
1	yes	no	6	-2,018.52	4,087.12	4	-2,018.52	4,070.42
2	yes	no	6	-2,032.70	4,115.48	4	-2,039.64	4,112.67
3	yes	no	6	-2,020.19	4,090.45	4	-2,020.19	4,073.76
4	yes	no	6	-2,111.81	4,273.70	4	-2,111.81	4,257.00
5	yes	no	6	-2,119.33	4,288.73	4	-2,145.57	4,324.53
6	no	yes	6	**-2,015.57**	**4,081.23**	4	**-2,016.25**	**4,065.88**
7	yes	yes	7	-2,012.60	4,083.63	5	-2,012.60	4,066.93
			UK (SC-2)					
			with feedback parameters			without feedback parameters		
No.	wt est.	scale est.	free pars.	LL	BIC	free pars.	LL	BIC
1	yes	no	13	-3,355.59	6,817.91	11	-3,355.97	6,802.26
2	yes	no	13	-3,364.00	6,834.75	11	-3,366.18	6,822.67
3	yes	no	13	-3,337.26	6,781.26	11	-3,343.26	6,776.83
4	yes	no	13	-3,396.28	6,899.30	11	-3,396.89	6,884.10
5	yes	no	13	-3,403.58	6,913.90	11	-3,406.63	6,903.57
6	no	yes	13	**-3,299.82**	**6,706.38**	11	**-3,327.28**	**6,744.88**
7	yes	yes	20	-3,294.12	6,752.44	18	-3,312.30	6,772.39

priori knowledge of the sign of β-coefficients) rather than the model performance, although further applications across a wider variety of choice contexts are required to test how generalisable this finding is.

For both datasets, further improvements in model fit are obtained when both weights and scales are estimated. Whilst this improvement is from just one extra parameter for the Danish dataset, the gain in fit comes at a heavy cost for the UK data, which has 7 extra parameters for a relatively insubstantial gain in model fit (and thus results in a worse BIC). This implies that for basic comparisons of DFT against alternative models, analysts may be better off fixing either the weights or the scales. Finally, these results here additionally demonstrate that scaling methods 4 and 5 lead to relatively poor performance for DFT; these results are compatible with the fact that scaling methods 4 and 5 resulted in MLBA outperforming DFT in previous studies (Trueblood et al., 2013a).

3.6. LBA model specifications

Table 15 shows the log-likelihoods for the models described in Section 2.3.3.

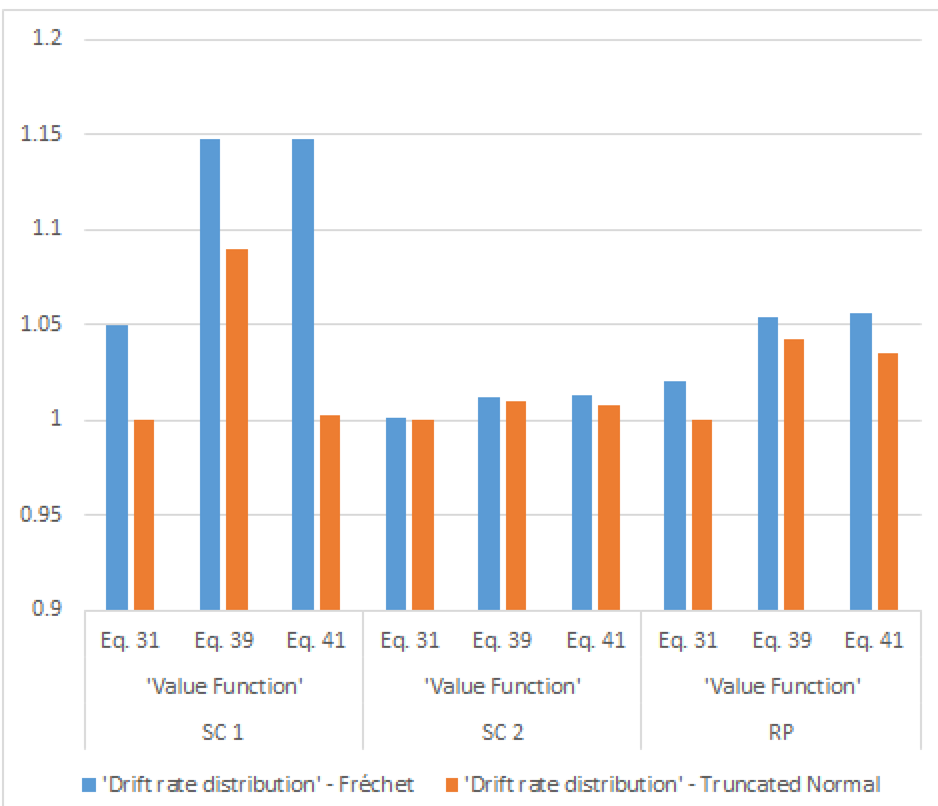
In all cases, standard MLBA has the best model fit. However, the good performance of speci-

TABLE 15 : Log-likelihoods of models in between MNL, RRM and MLBA, where F indicates a Fréchet distribution and TN indicates a Truncated Normal.

Model	Distribution	Value Function	Log-likelihood			BIC		
			first SC	second SC	RP	first SC	second SC	RP
Specification 1	F - Eq.37	Eq.31	-2,105.36	-3,326.43	-358.44	4,252.46	6,751.39	802.50
MNL	F - Eq.37	Eq.39	-2,301.25	-3,360.43	-370.05	4,627.54	6,802.97	812.54
RRM	F - Eq.37	Eq.41	-2,301.25	-3,363.91	-371.04	4,627.54	6,809.92	814.52
MLBA	TN - Eq.28	Eq.31	-2,005.92	-3,322.26	-351.23	4,061.92	6,743.04	781.49
Specification 2	TN - Eq.28	Eq.39	-2,185.19	-3,355.14	-366.11	4,412.10	6,808.81	817.84
Specification 3	TN - Eq.28	Eq.41	-2,009.88	-3,348.16	-363.35	4,061.49	6,794.85	812.31

fication 1 implies that most of the improvement in model fit for MLBA over MNL and RRM comes from its value function rather than the form of its drift rate distributions (vis, truncated normal). This is perhaps unsurprising given that Terry et al. (2015) found little differences in fit between LBA models with different drift rate distributions. The substantial improvement from Specification 3 relative to RRM in the first SC is the one clear exception, where it appears that the choice of distribution does have a distinct impact. This suggests that the benefits obtained through the use of MLBA model are dataset-dependent. Whilst in some cases an analyst may wish to only adopt the value function (Equation 31) for ease of implementation, other cases may require the full model. These results are also displayed in Figure 3, which demonstrates that more substantial model fit differences are found for the first SC and RP dataset.

FIGURE 3 : Relative model fit of alternative structures in comparison to MLBA across the three datasets.



4. SIMULATED DATA EXPERIMENTS

The work in Section 3 has provided initial insights about the potential benefits of DFT and MLBA compared to more traditional structures. Of course, these results are dataset specific and the advantages might be a result of the true (and unobserved) data generation process. In this section, we provide some further evidence based on simulated data, where we have a number of aims. In particular, we test the impacts of considering choices generated by different models, compare the ability of the different accumulator models at capturing various complexities in the data, and finally consider parameter recoverability.

4.1. Generation of simulated data

We use an efficient design to generate $5,000$ mode choice observations where each choice task has four alternatives (car, air, rail and high-speed rail), each described by travel cost (TC) and travel time (TT). Additionally, all alternatives other than car have an access time (AT) attribute.

We then generate choices for four models: an MNL model, a RRM, a DFT, and an MLBA. The aim of this exercise is to see how well each model fits data generated by each other model (including the generating model).

For the MNL model, we define the utility a respondent n obtains from alternative j in choice task s as:

$$U_{nsj} = \delta_j + \delta_{F_j} \cdot z_{F,n} + \beta_{TT} \cdot \alpha_{TT_j} \cdot TT_{nsj} + \beta_{TC} \cdot TC_{nsj} \cdot \alpha_{IE,n} + \beta_{AT} \cdot AT_{nsj} + \varepsilon_{nsj} \tag{45}$$

where δ_j and δ_{F_j} are alternative specific constants, with the latter capturing the difference between male and female participants through the use of an appropriate dummy term, $z_{F,n}$, which takes a value of 1 if individual n is female. TT_{nsj} is the travel time, TC_{nsj} is the travel cost and AT_{nsj} is the access time, all for alternative j in choice situation s for respondent n. There are coefficients for travel cost, access time and mode-specific coefficients for travel time, which are defined as $\beta_{TT} \cdot \alpha_{TT_j}$. A general value β_{TT} is estimated, with appropriate adjustments applied by multiplying by α_{TT_j} for mode j (for identification purposes we fix this coefficient for cars, $\alpha_{TT_{car}} = 1$). We additionally have an income effect, $\alpha_{IE,n}$, which is defined as $\alpha_{IE,n} = (\frac{income_n}{2500})^{\alpha_I}$, where $income_n$ is the income for individual n and α_I is an estimated income elasticity.

These additional coefficients are simple to add to, and estimate in, the various psychological choice models. For the DFT simulated dataset, we incorporate underlying preferences by setting the j^{th} element of $P_{ns,0}$ to $\delta_j + \delta_{F_j} \cdot z_{F,n}$, with this having been effective previously (see results in Hancock et al. 2018). The alternative specific travel time coefficients can be included in DFT and MLBA by multiplication of the attribute values, as we use our new scaling method (see Section 2.2.2.1) which means that these coefficients will have an equivalent impact on the attributes in DFT and MLBA as they would in a RUM model. Finally, in the MLBA models, we incorporate alternative specific constants (δ_j) by adding them to the mean drift rates as in Equation 29. All of the values used for the parameters to generate probabilities for each alternative are given in Table 17.

4.2. Results for simulated data

For each of the four datasets, We now study how well each model performs on data generated with a different model, thus giving an indication of the robustness of each model to the underlying data generation process. Table 16 shows the log-likelihood and BIC values for these comparisons.

TABLE 16 : The log-likelihood and BIC values obtained from models for the simulated datasets

Model	free pars.	dataset							
		MNL		RRM		DFT		MLBA	
		LL	BIC	LL	BIC	LL	BIC	LL	BIC
MNL	13	**-4,842.60**	**9,795.92**	-4,773.88	9,658.48	-4,907.49	9,925.70	-4,991.23	10,093.18
RRM	13	-4,853.38	9,817.48	**-4,727.57**	**9,565.86**	-4,923.43	9,957.58	-5,007.51	10,125.74
DFT	16	-4,847.94	9,832.16	-4,751.80	9,639.88	**-4,853.03**	**9,842.33**	-4,934.57	10,005.42
MLBA	17	-4,845.31	9,835.42	-4,741.07	9,626.92	-4,874.60	9,894.00	**-4,930.17**	**10,005.12**

The main difference between the RRM and MNL data generating models and the MLBA and DFT data generating models is that there are parameters for competition between psychologically similar alternatives in the MLBA and DFT models. From the results, it appears that neither MNL or RRM can capture this competition effect and thus have worse model fits of these models to the datasets. This is also suggested by the fact that the removal of DFT's feedback matrix results in a loss of 10.28 log-likelihood units for the DFT dataset, but only 0.56 units for the MNL dataset. Crucially, DFT and MLBA also outperform RRM on the data set generated by the MNL and outperform MNL for the dataset generated by RRM.

These results are highlighted by the comparison in Figure 4 of the fit of each of the four models to each dataset relative to the fit of the model used for data generation. These comparisons show, for each data set, that DFT and MLBA show much smaller differences in fit relative to the model consistent with the data generating process (DGP) than do the MNL and RRM, suggesting that DFT and MLBA are more robust to potential misspecification. MNL and RRM, by contrast, perform poorly on the datasets generated by DFT and MLBA, with MNL also having poor fit on the data generated by RRM.

4.3. Recovery of parameters from simulated datasets

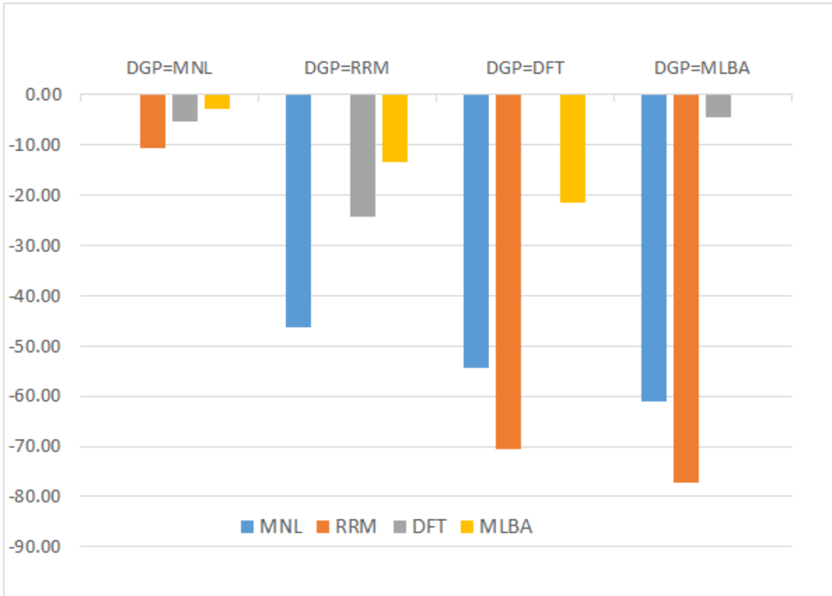
We next consider how well the different models recover the parameter values that were used to generate the simulated datasets for the same model. Table 17 gives the parameters used in simulating the data (labelled as 'setup') as well as the parameters produced in estimation, and the difference between those two. As each model is tested against a dataset generated by the same model, we can test the stability of the parameters. Using our new scaling method allows us to use similar parameter setup values across models, with the exception that parameters are adjusted such that the data generation process has similar amounts of noise across all datasets no matter which model is used to generate the choices.

All four models appear to accurately recover the three β-coefficients associated with the explanatory variables. These appear to be more recoverable than the alternative specific constants, which are more susceptible to noise for DFT and MLBA. All four models, however, additionally

TABLE 17 : Parameter values used to generate datasets and estimates for full models for their respective datasets

	MNL			RRM			DFT			MLBA		
Parameter	Setup	Estimate	Bias	Setup	Estimate	Bias	Setup	Estimate	Bias	Setup	Estimate	Bias
β_{TT}	-0.0050	-0.0044	-12%	-0.0030	-0.0029	-3%	-0.0050	-0.0049	-2%	-0.0030	-0.0037	23%
β_{TC}	-0.0280	-0.0279	0%	-0.0160	-0.0162	1%	-0.0280	-0.0304	9%	-0.0160	-0.0169	6%
β_{AT}	-0.0060	-0.0053	-12%	-0.0040	-0.0046	15%	-0.0060	-0.0057	-5%	-0.0040	-0.0039	-3%
δ_{car}	-0.5000	-0.8238	65%	-0.5000	-0.6965	39%	-0.5000	-1.4179	184%	-0.5000	0.68986	-238%
δ_{air}	-1.5000	-1.8053	20%	-1.5000	-1.8363	22%	-1.5000	-2.7631	84%	-1.5000	-1.5054	0%
δ_{rail}	-1.0000	-0.9036	-10%	-1.0000	-1.0067	1%	-1.0000	-1.8977	90%	-1.0000	0.62732	-163%
$\delta_{car_{fem}}$	-0.5000	-0.4752	-5%	-0.5000	-0.4020	-20%	-0.5000	-0.6257	25%	-0.5000	-0.6254	25%
$\delta_{air_{fem}}$	0.5000	0.6952	39%	0.5000	0.6822	36%	0.5000	0.8528	71%	0.5000	0.57662	15%
$\delta_{rail_{fem}}$	1.0000	1.1188	12%	1.0000	1.1855	19%	1.0000	-0.1264	-113%	1.0000	1.21153	21%
$\beta_{TT_{air}}$	1.2500	1.1041	-12%	1.2500	1.7576	41%	1.2500	1.1109	-11%	1.2500	1.07889	-14%
$\beta_{TT_{rail}}$	2.0000	2.3845	19%	2.0000	2.2579	13%	2.0000	1.9182	-4%	2.0000	2.24317	12%
$\beta_{TT_{hsr}}$	1.5000	1.7723	18%	1.5000	1.3528	-10%	1.5000	1.8306	22%	1.5000	1.17701	-22%
α_I	-0.5000	-0.5106	2%	-0.5000	-0.4962	-1%	-0.5000	-0.3585	-28%	-0.5000	-0.4553	-9%
ϕ_1							0.0500	0.03646	-27%			
ϕ_2							0.1000	0.138	38%			
σ_ε							1.4142	1.4142	**fixed**			
τ							10.0000	8.83029	-12%			
A										1.0000	1.0000	**fixed**
χ										2.0000	2.009	0%
σ_ε										2.0000	2.0000	**fixed**
I_0										10.0000	11.0373	10%
λ_1										0.1000	0.06936	-31%
λ_2										0.2000	0.17905	-10%

FIGURE 4 : Log-likelihood of estimated models compared to model consistent with data generating process (DGP)



perform well at recovering the attribute-specific travel time coefficients. Most importantly for DFT and MLBA, the process parameters are fairly well recovered too.

5. CONCLUSIONS

In this paper, we have considered two alternate accumulator choice models developed in mathematical psychology and compared them against models typically used in discrete choice modelling. The models in question are decision field theory (DFT), a model where the preference strength for each alternative is stochastically updated over time, and the multi-attribute linear ballistic accumulator (MLBA), where the preference strength for each alternative increases linearly and deterministically over time.

We first made a number of methodological developments to improve the suitability of the models for studying travel behaviour and other non-laboratory based choices. For DFT, we implemented a new scaling method on the attributes, which results in a number of benefits such as the modeller not having to know the sign of the impact of the attributes before running the model. This has an immediate benefit for the UK dataset, for which one attribute (whether the delay information service is free) is a desirable attribute while the other attributes are undesirable. A comparison with other available scaling approaches in Section 3.5 also highlights the benefits of this approach. In particular, the new method significantly increases the impact of the feedback matrices on model fit, although the feedback matrix parameters do not influence model fit when there are only two alternatives (see the discussion in Section 3.5). However, regardless of whether the feedback matrix has an impact or not, DFT outperforms MNL and RRM for our SP and RP datasets. Whilst our empirical results suggest that the benefit the new scaling method brings is dataset-dependent, it is clear that future DFT models should always consider its implementation. This is particularly the case when a researcher does not have a priori information on the directionality of an attribute. However, further research is required to establish the benefits of simultaneously estimating scale and weight parameters.

We also considered the impact of including parameters to capture underlying preferences towards specific alternatives in MLBA and DFT. Results from our UK dataset suggest that MLBA and DFT make substantial gains when these parameters are included and can consequently capture status quo biases. We have, however, only considered one method for incorporating preferences in these models. Whilst we add parameters to the drift rate in MLBA, alternative specifications would allow for an adjustment of the starting point A or the threshold χ, such that alternatives have different values for these parameters. It is quite possible that some alternatives may not require as much evidence to be chosen (for example, a commuter's usual route to work), meaning that an MLBA model including alternative specific thresholds may work well. This could be investigated in future research, with, for example, work on accumulator models with collapsing thresholds already proving popular (Bowman et al., 2012; Hawkins et al., 2015; Evans et al., 2019a). We additionally only test MLBA with truncated normal drift rate distributions (with start point variability) and Fréchet-distributed drift rates (with no start point variability, i.e., $A = 0$), while other drift rate distributions could also be considered (Terry et al., 2015). However, results from comparisons of models that have features of MLBA and those that have features of MNL/RRM suggest that the main improvement in model fit for MLBA relative to those standard choice models is due to its value function, and not the form of its (truncated normal) drift rate distributions. This suggests that in some cases, analysts may wish to simply utilise the value function (Equation 31) within a logit model for ease of implementation. Further research is required to understand cases in which moving from this simpler model to MLBA will bring clear advantages, as is the case for our first stated choice study.

The operationalisation of the two models in this paper provides promising results and paves the way for the incorporation of data on the processes of decision-making in these models, such as eye-tracking information, response times and EEG data.

We also investigate in detail the relative importance of different parameters of our models. In particular, we consider a number of important identification restrictions for both DFT and MLBA. Whilst our theoretical identification requirements in Tables 1 and 2 should always be followed, our results show that the empirical identification issues are dataset-specific. For example, the feedback matrix in DFT has an impact in only some cases. Additionally, whilst setting the threshold parameter for MLBA to a particular value does not have a significant impact on its fits for our simulated datasets, it does have an impact for our SP data. The opposite is true for the drift rate constant, I_0, which is important for our simulated datasets but is less important for our SP data. It is possible that the importance of these parameters varies according to how deterministic the data is and further work could test datasets with specified variations in the level of noise. This could help an analyst determine which parameters are important for MLBA for complex choice data.

We tested the models extensively using simulated data, where the findings suggest that DFT and MLBA may be less sensitive to model misspecification (i.e. if the estimated model differs substantially from that used for data generation) than the corresponding RUM and RRM models.

Crucially, both DFT and MLBA outperform MNL and RRM across the two SP datasets and the RP dataset, including in out of sample validation for the latter, which is to the best of our knowledge the first use of either DFT or MLBA on RP data. The good fits of DFT and MLBA to our second stated survey dataset suggest that if there is competition between psychologically similar alternatives (when there are two alternatives that have attributes that are more similar than those of a third alternative), a move towards a choice model with psychological foundations becomes more appealing.

Moving away from RUM has obvious pitfalls, especially in terms of the use of models for welfare analysis (see e.g. Hess et al., 2018). The evidence in this paper suggests that if an analyst is willing to accept these pitfalls, then moving further away from RUM than for example with a RRM model, may be beneficial, and models from mathematical psychology provide an interesting avenue for such work. Of course, more research is needed in terms of additional comparisons, including on larger datasets with more alternatives and attributes. Also, whilst we have considered DFT and MLBA, future research should also consider models from mathematical psychology that do not have closed-form likelihood functions. A large number of models from mathematical psychology such as the drift diffusion model (Wiecki et al., 2013; Ratcliff et al., 2016), the leaky competing accumulator (Usher and McClelland, 2001) and the feed-forward inhibition model (Turner et al., 2016) can be estimated using hierarchical Bayesian estimation combined with probability density approximation (Turner and Sederberg, 2014). This means that there is large scope for further comparisons between psychological and mainstream choice models using hierarchical Bayesian estimation, a method already popular in traditional choice modelling for mixed logit models (Train, 2001; Burda et al., 2008; Dumont et al., 2015; Akinc and Vandebroek, 2018).

Additionally, LBA models (Brown and Heathcote, 2008) have been developed that fit non-stationary data by assuming drift rates that change over time (Holmes et al., 2016); though with relatively simple stimuli. Similar processes could be added to DFT and MLBA, thus allowing for attributes and/or attribute values that change over time. Such extensions would allow DFT and MLBA to be applied to dynamic revealed preference datasets such as the lane merging decisions made by drivers, where typical choice models may not do so well due to their static nature. Complex datasets such as these, as well as datasets with additional process or psychometric measures, would also be useful for further testing the functionality and usefulness of the process parameters within both DFT and MLBA. Additionally, given that in Hancock et al. (2018), we demonstrate that DFT can efficiently incorporate random parameters, it is possible that similar adjustments could also be made for MLBA. All of these potential extensions of DFT and MLBA, combined with the results in this paper, demonstrate that accumulator models such as DFT and MLBA are attractive alternative approaches to random utility models, particularly when it comes to forecasting. It therefore appears that these models, as well as others, may hold significant promise in improving the behavioural realism in choice models, in both transport and beyond.

ACKNOWLEDGEMENTS

The authors acknowledge the financial support by the European Research Council through the consolidator grant 615596-DECISIONS and funding by the Social Sciences and Humanities Research

Council of Canada in the form of Insight grant SSHRC 435-2 to the University of Victoria for Marley. The data used for the revealed preference analysis comes from a study conducted for the UK Department for Transport. The results reported, findings inferred, and views expressed are the responsibility of the authors, and not a statement of the Department's policy or guidance. We would also like to thank Jennifer Trueblood and four anonymous reviewers for their comments on a previous version of this paper.

REFERENCES

Akinc, D. and Vandebroek, M. (2018). Bayesian estimation of mixed logit models: Selecting an appropriate prior for the covariance matrix. *Journal of choice modelling*, 29:133–151.

Arup, ITS Leeds and Accent (2015). Provision of market research for value of time savings and reliability. Phase 2 report to the Department for Transport, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/470231/vtts-phase-2-report-issue-august-2015.pdf.

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.

Bhatia, S. (2013). Associations and the accumulation of preference. *Psychological Review*, 120(3):522.

Bierlaire, M., Thémans, M., and Zufferey, N. (2010). A heuristic for nonlinear global optimization. *INFORMS Journal on Computing*, 22(1):59–70.

Bowman, N. E., Kording, K. P., and Gottfried, J. A. (2012). Temporal integration of olfactory perceptual evidence in human orbitofrontal cortex. *Neuron*, 75(5):916–927.

Brown, S. D. and Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, 57(3):153–178.

Bruno, H., Cebollada, J., and Chintagunta, P. K. (2015). Intra-household heterogeneity and targeting. *Available at SSRN 2892591*.

Burda, M., Harding, M., and Hausman, J. (2008). A bayesian mixed logit–probit model for multinomial choice. *Journal of Econometrics*, 147(2):232–246.

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

Busemeyer, J. R., Gluth, S., Rieskamp, J., and Turner, B. M. (2019). Cognitive and neural bases of multi-attribute, multi-alternative, value-based decisions. *Trends in cognitive sciences*, 23(3):251–263.

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.

Cataldo, A. M. and Cohen, A. L. (2018). Reversing the similarity effect: The effect of presentation format. *Cognition*, 175:141–156.

Chernev, A. (2004). Extremeness aversion and attribute-balance effects in choice. *Journal of consumer research*, 31(2):249–263.

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.

Cohen, A. L., Kang, N., and Leise, T. L. (2017). Multi-attribute, multi-alternative models of choice: Choice, reaction time, and process tracing. *Cognitive Psychology*, 98:45–72.

Daly, A., Hess, S., and de Jong, G. (2012). Calculating errors for measures derived from choice modelling estimates. *Transportation Research Part B*, 46(2):333–341.

Diederich, A. (2003). Mdft account of decision making under time pressure. *Psychonomic Bulletin & Review*, 10(1):157–166.

Dumont, J., Keller, J., and Carpenter, C. (2015). Rsghb: functions for hierarchical bayesian estimation: a flexible approach. *R package version*, 1(2).

Eddelbuettel, D., François, R., Allaire, J., Ushey, K., Kou, Q., Russel, N., Chambers, J., and Bates, D. (2011). Rcpp: Seamless r and c++ integration. *Journal of Statistical Software*, 40(8):1–18.

Evans, N. J. (2020). Same model, different conclusions: An identifiability issue in the linear ballistic accumulator model of decision-making.

Evans, N. J., Hawkins, G. E., and Brown, S. D. (2019a). The role of passing time in decision-making. *Journal of experimental psychology: learning, memory, and cognition*.

Evans, N. J., Holmes, W. R., and Trueblood, J. S. (2019b). Response-time data provide critical constraints on dynamic models of multi-alternative, multi-attribute choice. *Psychonomic Bulletin & Review*, 26(3):901–933.

Fosgerau, M. (2006). Investigating the distribution of the value of travel time savings. *Transportation Research Part B: Methodological*, 40(8):688–707.

Giergiczny, M., Hess, S., Dekker, T., and Chintakayala, P. K. (2013). Testing the consistency (or lack thereof) between choices in best-worst surveys. In *Third International Choice Modelling Conference, Sydney*.

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

González-Valdés, F. and Ortúzar, J. d. D. (2017). The stochastic satisficing model: A bounded rationality discrete choice model. *Journal of Choice Modelling*, 27:74–87.

Hancock, T. O., Broekaert, J., Hess, S., and Choudhury, C. F. (2020). Quantum probability: A new method for modelling travel behaviour. *Transportation Research Part B: Methodological*, 139:165–198.

Hancock, T. O., Hess, S., and Choudhury, C. F. (2018). Decision field theory: Improvements to current methodology and comparisons with standard choice modelling techniques. *Transportation Research Part B: Methodological*, 107:18–40.

Hawkins, G. E., Forstmann, B. U., Wagenmakers, E.-J., Ratcliff, R., and Brown, S. D. (2015). Revisiting the evidence for collapsing boundaries and urgency signals in perceptual decision-making. *Journal of Neuroscience*, 35(6):2476–2484.

Hawkins, G. E., Islam, T., and Marley, A. (2019). Like it or not, you are using one value representation. *Decision*, 6(3):237–260.

Hawkins, G. E., Marley, A., Heathcote, A., Flynn, T. N., Louviere, J. J., and Brown, S. D. (2014). Integrating cognitive process and descriptive models of attitudes and preferences. *Cognitive Science*, 38(4):701–735.

Heathcote, A. and Love, J. (2012). Linear deterministic accumulator models of simple choice. *Frontiers in psychology*, 3:292.

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

Hensher, D. A. (2010). Attribute processing, heuristics and preference construction in choice analysis. In Hess, S. and Daly, A. J., editors, *State-of Art and State-of Practice in Choice Modelling: Proceedings from the Inaugural International Choice Modelling Conference*, chapter 3, pages 35–70. Emerald, Bingley, UK.

Hess, S., Daly, A., and Batley, R. (2018). Revisiting consistency with random utility maximisation: theory and implications for practical work. *Theory and Decision*, 84(2):181–204.

Hess, S. and Palma, D. (2019). Apollo: a flexible, powerful and customisable freeware package for choice model estimation and application, www.apollochoicemodelling.com.

Hess, S., Rose, J. M., and Hensher, D. A. (2008). Asymmetric preference formation in willingness to pay estimates in discrete choice models. *Transportation Research Part E: Logistics and Transportation Review*, 44(5):847–863.

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.

Holmes, W. R., Trueblood, J. S., and Heathcote, A. (2016). A new framework for modeling decisions about changing information: The piecewise linear ballistic accumulator model. *Cognitive Psychology*, 85:1–29.

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.

Kaufman, B. E. (1990). A new theory of satisficing. *Journal of Behavioral Economics*, 19(1):35–51.

Krajbich, I., Lu, D., Camerer, C., and Rangel, A. (2012). The attentional drift-diffusion model extends to simple purchasing decisions. *Frontiers in psychology*, 3:193.

Liu, Y. and Cirillo, C. (2018). A generalized dynamic discrete choice model for green vehicle adoption. *Transportation Research Part A: Policy and Practice*, 114:288–302.

Luce, D., Krantz, D., Suppes, P., and Tversky, A. (1990). Foundations of measurement, vol. iii: Representation, axiomatization, and invariance.

Luce, D. R. (1962). Comments on Rozeboom's criticism of "On the Possible Psychophysical Laws", American Psychological Association.

Marley, A., Flynn, T. N., and Louviere, J. (2008). Probabilistic models of set-dependent and attribute-level best–worst choice. *Journal of Mathematical Psychology*, 52(5):281–296.

Marley, A. and Regenwetter, M. (2017). Choice, preference, and utility: Probabilistic and deterministic representations in batchelder w, colonius h, dzhafarov e, & myung j (eds.), new handbook of mathematical psychology (pp. 374–453).

Masiero, L. and Hensher, D. A. (2010). Analyzing loss aversion and diminishing sensitivity in a freight transport stated choice experiment. *Transportation Research Part A: Policy and Practice*, 44(5):349–358.

McFadden, D. (1974). Conditional Logit Analysis of Qualitative Choice Behaviour. In Frontiers in Econometrics, ed. P. Zarembka (New York: Academic press).

McNair, B. J., Hensher, D. A., and Bennett, J. (2012). Modelling heterogeneity in response behaviour towards a sequence of discrete choice questions: a probabilistic decision process model. *Environmental and Resource Economics*, 51(4):599–616.

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.

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.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85(2):59.

Ratcliff, R., Smith, P. L., Brown, S. D., and McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in cognitive sciences*, 20(4):260–281.

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

Rose, J. M. (2014). Interpreting discrete choice models based on best-worst data: a matter of framing. *The Transportation Research Board (TRB) 93rd Annual Meeting*.

Sanderson, C. and Curtin, R. (2016). Armadillo: a template-based c++ library for linear algebra. *Journal of Open Source Software*.

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.

Stathopoulos, A. and Hess, S. (2012). Revisiting reference point formation, gains–losses asymmetry and non-linear sensitivities with an emphasis on attribute specific treatment. *Transportation Research Part A: Policy and Practice*, 46(10):1673–1689.

Stewart, N. and Simpson, K. (2008). A decision-by-sampling account of decision under risk. *The probabilistic mind: Prospects for Bayesian cognitive science*, pages 261–276.

Swait, J. (2001). A non-compensatory choice model incorporating attribute cutoffs. *Transportation Research Part B: Methodological*, 35(10):903–928.

Terry, A., Marley, A., Barnwal, A., Wagenmakers, E.-J., Heathcote, A., and Brown, S. D. (2015). Generalising the drift rate distribution for linear ballistic accumulators. *Journal of Mathematical Psychology*, 68:49–58.

Train, K. (2001). A comparison of hierarchical bayes and maximum simulated likelihood for mixed logit. *University of California, Berkeley*, pages 1–13.

Trueblood, J., Brown, S., and Heathcote, A. (2013a). The multi-attribute linear ballistic accumulator model of decision-making. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 35.

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):179–205.

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

Turner, B. M., Schley, D. R., Muller, C., and Tsetsos, K. (2018). Competing theories of multialternative, multiattribute preferential choice. *Psychological review*, 125(3):329.

Turner, B. M. and Sederberg, P. B. (2014). A generalized, likelihood-free method for posterior estimation. *Psychonomic bulletin & review*, 21(2):227–250.

Turner, B. M., Sederberg, P. B., and McClelland, J. L. (2016). Bayesian analysis of simulation-based models. *Journal of Mathematical Psychology*, 72:191–199.

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, 79(4):281.

Tversky, A. (1977). Features of similarity. *Psychological Review*, 84(4):327.

Tversky, A. and Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5(2):207–232.

Usher, M. and McClelland, J. L. (2001). The time course of perceptual choice: the leaky, competing accumulator model. *Psychological Review*, 108(3):550.

Wiecki, T. V., Sofer, I., and Frank, M. J. (2013). HDDM: hierarchical bayesian estimation of the drift-diffusion model in python. *Frontiers in neuroinformatics*, 7:14.

A. APPENDIX: ALTERNATIVE VERSIONS OF MLBA

Whilst we use the mainstream version of MLBA (Trueblood et al., 2014) in this paper, it should be noted that the original version of MLBA (Trueblood et al., 2013a) has also not been tested on large-scale consumer choice data. Whilst this version of MLBA, here denoted '$MLBA_0$', uses the same start, threshold and standard deviation for its drift rates, it differs in the specification for the value of the mean drift rate:

$$d_{ns,j} = \frac{10}{1 + exp(-\gamma \cdot v_{ns,j})} \tag{46}$$

where $v_{ns,j}$ is given by a valence function and γ is a logistic parameter. Small values of the logistic parameter γ would result in $exp(-\gamma \cdot v_{ns,j}) \rightarrow 1$, meaning that the valences, $v_{ns,j}$, are less influential and the probabilities of the alternatives become more similar, resulting in a less deterministic choice. The valences are similar to a decision field theory model's valences with the exception that they attempt to additionally capture the comparison process achieved by DFT's feedback matrix. We thus have

$$V_{ns} = C_{ns} \cdot M_{ns} \cdot \boldsymbol{W} \tag{47}$$

where $\boldsymbol{W}$ is a vector comprising of a set of attribute importance weights that sum to 1, M_{ns} is the attribute matrix and C_{ns} is a $J_n \times J_n$ comparison matrix (J_n being the number of alternatives) with diagonal entries of 1 and off-diagonal elements:

$$C_{ns,i,j \neq i} = \frac{exp(-\phi \cdot Dist_{ns,i,j}) - 1}{n-1}. \tag{48}$$

Finally, ϕ is a sensitivity parameter such that high values result in the distance between the attributes of the alternatives becoming insignificant. Low values allow for more similar alternatives to compete more with each other relative to less similar alternatives.

Results from applying the previous version of MLBA to both of the SP datasets and the RP dataset are given in Table 18 below.

TABLE 18 : Comparison of different versions of MLBA

Dataset	$MLBA_0$	MLBA	Difference
Danish	-2,189.78	-2,005.92	-183.86
UK	-3,394.36	-3,322.26	-72.10
RP	-375.24	-351.23	-24.01

From these results, it appears that the old version of MLBA has far inferior fits compared to that of the mainstream MLBA. Consequently, it would appear that modellers should focus on the mainstream version of MLBA.

B. APPENDIX: A MINIMUM MEAN DRIFT RATE

Originally, the mean drift rate in MLBA was specified as:

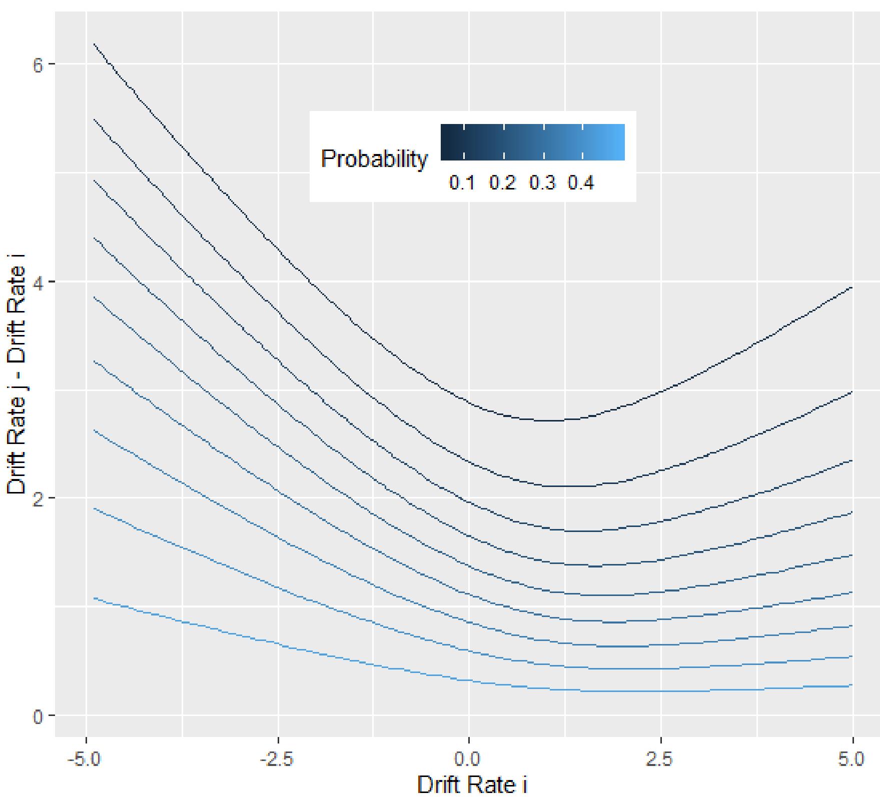
$$d_{ns,j} = v_{ns,j} + I_0. \tag{49}$$

As well as choosing to add an alternative specific component here, we also choose to truncate the mean drift rate such that it is always positive:

$$d_{ns,j} = max(0, \delta_j + v_{ns,j} + I_0), \tag{50}$$

To consider the impact of not restricting mean drift rates to be positive, we consider the probabilities generated by drift rates i and j where these have values of $-5 < i < 5$ and $i < j < i + 10$. Figure 5 gives the probability of choosing alternative i given the use of these drift rate values as well as values of $A = 1$, $\chi = 2$ and $\sigma_\varepsilon = 1$.

FIGURE 5 : The probability of picking alternative i as the mean drift rate values change



In this figure, the contours demonstrate the points at which probabilities remain the same. If these lines were horizontal, it would indicate that for MLBA, only differences in drift rates matter (much as only differences in utility matter in RUM). However, the presence of curved lines indicates that at some points adding I_0 will increase the choice probability of alternative i, whereas it will decrease the probability at other points. This can cause convergence issues for MLBA. We find that fixing the mean drift rates to have a minimum value of zero reduces the impact of this issue and improves estimation of MLBA models.

C. APPENDIX: PROOF OF SCALE-INVARIANCE OF DFT WITH SCALING PARAMETERS

Assume a single choice scenario where we have N alternatives each with K attributes. This gives some attribute matrix M_1. Now suppose we have a set of importance weights $\boldsymbol{w}$ and scaling values $\boldsymbol{\beta}_{DFT}$ that give the diagonal matrix β_1 (these could be theoretical values, or based on estimation). To calculate the probability of choosing each alternative, we require the expectation, $\boldsymbol{\xi}_\tau$:

$$\boldsymbol{\xi}_\tau = (I-S)^{-1} \cdot (1-S^\tau) \cdot C \cdot M_1 \cdot \beta_1 \cdot \boldsymbol{w} + S^\tau \cdot \boldsymbol{P}_0, \tag{51}$$

and covariance, Ω_τ:

$$\Omega_\tau = (I-Z)^{-1}\,(I-Z^\tau)\,\overline{C \cdot M_1 \cdot \beta_1 \cdot \Psi \cdot \beta_1' \cdot M_1' \cdot C' + \sigma_\varepsilon^2 \cdot I}, \tag{52}$$

where the expanded forms show us where the attribute matrix M_1 enters the calculations.

If we now assume that the attribute matrix is adjusted through the use of, for example, different units for the attributes, we obtain a new matrix, M_2. Crucially, M_2 can be written as $M_2 = M_1 \cdot \gamma$, where γ is a matrix created with *dimensional parameters* (see Section 2.2.2.1), arranged in a diagonal matrix such that each attribute k has an associated dimensional parameter γ_k, on the k^{th} diagonal element, which gives the appropriate unit change. With this change, we now have:

$$\boldsymbol{\xi}_\tau' = (I-S)^{-1} \cdot (1-S^\tau) \cdot C \cdot M_1 \cdot \gamma \cdot \beta_2 \cdot \boldsymbol{w} + S^\tau \cdot \boldsymbol{P}_0, \tag{53}$$

and:

$$\Omega_\tau' = (I-Z)^{-1}\,(I-Z^\tau)\,\overline{C \cdot M_1 \cdot \gamma \cdot \beta_2 \cdot \Psi \cdot \beta_2 \cdot \gamma \cdot M_1' \cdot C' + \sigma_\varepsilon^2 \cdot I}, \tag{54}$$

where S, τ, C, $\boldsymbol{w}$, $\boldsymbol{P}_0$, Z, σ_ε^2 and Ψ all remain unchanged from Equations 51 and 52.

As our new version of DFT includes $\beta_{\mathbf{DFT}}$ scaling parameters, we can set $\xi_\tau = \xi_\tau'$ and $\Omega_\tau = \Omega_\tau'$ simply by setting $\beta_2 = \beta_1/\gamma$. This results in β_{DFT} parameters that correspond to marginal utility parameters in a random utility model, with the probabilities of choosing each alternative unchanged.

If we do not have scaling parameters (i.e. $\beta_1 = \beta_2 = I$ in the above equations), DFT becomes scale-variant. This is because the importance weights $\mathbf{w}$ adjust (under estimation) to accommodate the change from M_1 in such a way that the relative importance of the different attributes remains the same. Theoretically, this change would would give us a new vector $\mathbf{w}'$, with elements $w_i' = (w_i/\gamma_i)$. However, there is no guarantee that these new weights satisfy $\sum_1^K w_k' = 1$, as we already have $\sum_1^K w_k = 1$, and the elements of $\gamma \in \mathbb{R}$. More crucially, this changes the value of Ψ (with diagonal elements $\Psi_{ii} = (w_i) \cdot (1-w_i)$ and off-diagonal elements $\Psi_{ij} = -(w_i) \cdot (w_j)$). The non-linearity of Ψ results in different probabilities being generated should $\mathbf{w}$ change.

For example, if we consider the case described in Section 2.2.2.1, we initially have $\mathbf{w} = \begin{pmatrix} 0.6 \\ 0.4 \end{pmatrix}$. Incorporating the scale change into the weights instead of the dimensional parameters gives $\mathbf{w}' = \begin{pmatrix} 0.6/1.18 \\ 0.4/60 \end{pmatrix}$. These values however must sum to 1. Multiplying both by a factor of

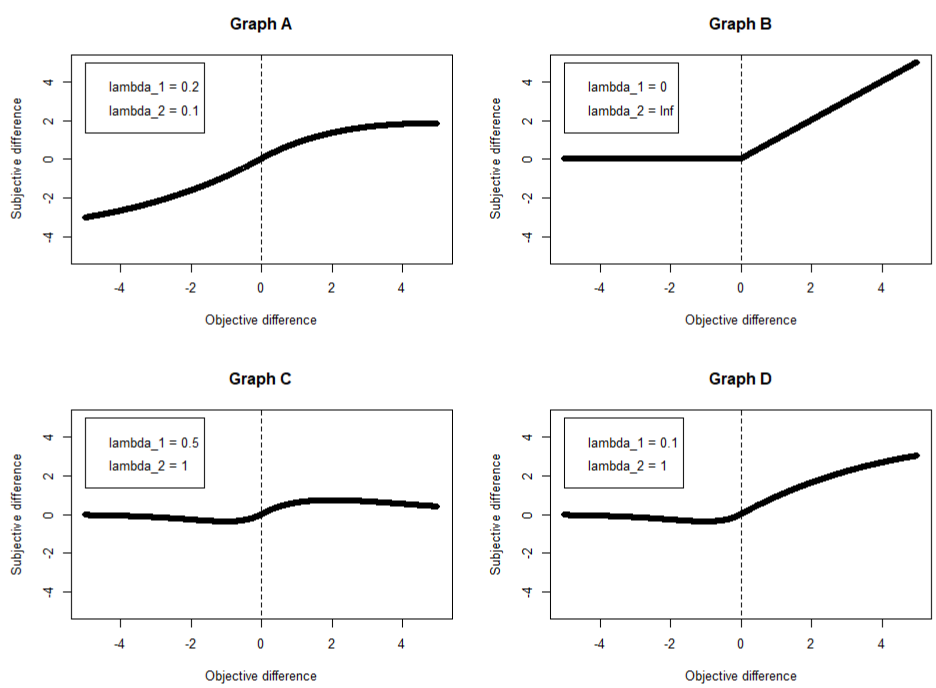
1.94 gives $\mathbf{w}'' = \begin{pmatrix} 0.987 \\ 0.013 \end{pmatrix}$. This implies that the time attribute is only considered in 1% of preference updating steps. As a result, the probabilities of choosing the different alternatives changes significantly if there is a low number of preference updating steps. For example, with a single updating step, a slower but cheaper alternative would be chosen with a probability of 60% before the unit change, but 99% after the unit change.

D. APPENDIX: POSSIBLE INTERPRETATIONS OF MLBA PARAMETERS

In this appendix, we consider possible estimated values of the λ parameters in MLBA (Equation 24) and how they could be interpreted.

In the first case, values of $\lambda_1 = 0.2$ and $\lambda_2 = 0.1$ give us Graph A in Figure 6, which looks similar to a graph of utilities under prospect theory, with decreasing marginal utilities and loss aversion. The ability to capture such effects was the original aim of developing MLBA (Trueblood et al., 2014).

FIGURE 6 : Different possible transformations from objective to subjective differences depending on the value of the lambda parameters under MLBA.



However, a key difference here is that MLBA models count each attribute difference between alternatives twice, as each alternative has a separate evaluation of its drift rate (Equation 22 is calculated separately for each alternative). This means that values of $\lambda_1 = 0$ and $\lambda_2 = \infty$ (Graph B) correspond to each difference only being added to one drift rate (linearly) rather than to two drift rates. Additionally, some values of λ result in psychologically unrealistic subjective differences, with for example $\lambda_1 = 0.5$ and $\lambda_2 = 1$ (Graph C) resulting in only certain (close to zero) differences being added to the drift rates. This may be unrealistic if this is the only contribution to drift rates. However, paired individual estimates for the λ parameters may not always be unrealistic- for example, Graph D with $\lambda_1 = 0.1$ and $\lambda_2 = 1$ (which was unrealistic in Graph C) demonstrates a slightly non-linear evaluation of positive objective differences with an extra cost against the worst of two very similar alternatives (through the negative addition to drift rates for very small negative objective differences).

As these parameters are also what allows MLBA to capture the similarity effect, it is in application difficult to establish what features the model may be capturing. Notably, an MLBA model which does not find a similarity effect nor any non-linear sensitivities will result in λ parameters approaching zero. This issue resulted in previous applications of MLBA resorting to different valuation and weighting functions (Cohen et al., 2017).