# A random utility maximisation model considering the information search process

Nova, Gabriel; Guevara, C. Angelo; Hess, Stephane; Hancock, Thomas O.

**DOI**
10.1016/j.trb.2025.103264

**Publication date**
2025

**Document Version**
Final published version

**Published in**
Transportation Research Part B: Methodological

Transportation Research Part B 199 (2025) 103264

Contents lists available at ScienceDirect

Transportation Research Part B

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

# A random utility maximisation model considering the information search process

Gabriel Nova [a,d], C. Angelo Guevara [a,b,*], Stephane Hess [c], Thomas O. Hancock [c]

[a] *Civil Engineering Department, Universidad de Chile, Santiago, Chile*
[b] *Instituto Sistemas Complejos de Ingeniería (ISCI), Santiago, Chile*
[c] *Choice Modelling Centre and Institute for Transport Studies, University of Leeds, Leeds, United Kingdom*
[d] *Faculty of Technology, Policy and Management, TU Delft, Delft, The Netherlands*

ARTICLE INFO

*Keywords:*
RUM-DFT
Decision field theory
Information search process
Choice modelling

ABSTRACT

Discrete choice analysis aims to understand and predict decision-makers' behaviour, a goal that is crucial across several disciplines, including transportation. This type of analysis has relied predominantly on static representations of preferences, principally through the Random Utility Maximisation (RUM) model, due to its ease of implementation, economic interpretability, and statistical formality. However, this model assumes that individuals possess complete information about all attributes of alternatives and that they can process and recall this information instantaneously, which may not align with actual human behaviour. In contrast, the Decision Field Theory (DFT) model from mathematical psychology explicitly incorporates the repeated scrutiny of attributes and recall effects within the decision-making process, which enables it to model attention weights, but lacks microeconomic interpretability and clear statistical parameter identification. This paper introduces the RUM-DFT model, which seeks to integrate strengths of both approaches. Through Monte Carlo simulations, the proposed model is shown to be able to: (i) recover parameters related to the deliberation process, (ii) replicate the dynamic behaviour of utilities during deliberation as observed in practice, (iii) maintain economic interpretability by estimating coefficients that can be used to calculate the marginal indirect utilities, and (iv) highlight the pitfalls of using a RUM model that disregards the true dynamics of data generation process. The SwissMetro case study is employed also to evaluate the RUM-DFT model using a real-world dataset, demonstrating the viability and superior goodness-of-fit of the proposed model.

## 1. Introduction

Discrete choice models provide a mathematical framework for estimating and predicting the choice behaviour of economic agents, and have been widely used for decades across various areas, such as economics, health, marketing, and transportation (Hess and Daly, 2024).

The main paradigm in discrete choice modelling is Random Utility Maximisation (RUM), favoured for its simplicity in implementation, low computational cost, robust statistical foundation and microeconomic interpretability. RUM models assume that individuals are rational agents who select the alternative with the highest perceived utility within the choice set, after conducting

* Corresponding author at: Civil Engineering Department, Universidad de Chile, Santiago, Chile.
*E-mail addresses:* g.n.nova@tudelft.nl (G. Nova), crguevar@ing.uchile.cl (C.A. Guevara), S.Hess@leeds.ac.uk (S. Hess), T.O.Hancock@leeds.ac.uk (T.O. Hancock).

https://doi.org/10.1016/j.trb.2025.103264
Received 18 October 2023; Received in revised form 5 March 2025; Accepted 28 May 2025
Available online 8 July 2025
0191-2615/

an implicit and instantaneous *in-depth* information search process. Specifically, the process is implicit, as it is unobserved by the researcher; instantaneous, as it does not account for deliberation time; and in-depth, as utilities are built for each alternative by combining their separate attributes to inform the choice decision. Given that RUM models disregard the information search and deliberation processes, they are classified as *static* by Busemeyer and Townsend (1993). This is in contrast with *dynamic* models that explicitly account for these processes.

Static choice models often fail to capture the cognitive complexities inherent in real-world decision-making processes. Overlooking the true decision-making process can result in inconsistent parameter estimators due to endogeneity, stemming from model misspecification, and impede accurate forecasting, which is essential for designing informed policies (Guevara, 2015, 2024).

Cognitive complexities in decision-making are substantial, particularly in transportation contexts. A wealth of literature based on psychological theory shows the relationship between alternative selection and the corresponding decision response time (see, e.g. Brown and Heathcote (2008), Ariely and Zakay (2001) and Stewart et al. (2016a)). Empirical evidence further suggests that decision-makers often revisit attributes multiple times during the deliberation process, as they do not fully recall the information about the choice task (Johnson et al., 2008; Stewart et al., 2016b; Nova and Guevara, 2022). Moreover, individuals tend to filter attributes dynamically, focusing their attention on specific values that are either more difficult to conceptualise or are more important in the utility or preference function (Jacoby, 1975; Payne, 1976; Nova and Guevara, 2022). In the realm of transportation, the influence of cognitive processes becomes even more evident. For instance, Katsikopoulos et al. (2000) demonstrate through experiments that the presentation of information affects drivers' decision-making strategies. Similarly, Agrawal and Peeta (2021) presented a hybrid route choice model that integrates electroencephalography (EEG) signals and measurable variables, allowing them to show that drivers are more likely to change their current route when they are paying high attention to the road and incurring a higher cognitive cost. Furthermore, Ayaz et al. (2023) conclude that network complexity and the mental effort to process information (cognitive load) reduce the likelihood of choosing risky route choices in simpler networks. These studies highlight the importance of incorporating cognitive effects and the information search process in choice, as they enhance understanding of behaviour in real-world environments, improve forecasting, and emphasise the critical role of information presentation in decision outcomes.

In contrast to static models, dynamic probabilistic models explicitly account for the influence of the deliberation process on choice probabilities. In these models, the time spent making a decision directly impacts the final choice, as probabilities evolve over the course of the deliberation process due to the continuous acquisition and processing of information (attributes). This iterative process updates the value of preferences or utilities before the choice is made. Models that include cognitive cost in the information search process, such as the Directed Cognition Model (Gabaix et al., 2006) and the Adaptive Path Choice model have been shown to perform better than compensatory models, such as RUM structures, in complex decision-making contexts (Gao et al., 2011).

A pre-eminent example of a theory that incorporates the dynamic processing of attributes is Decision Field Theory (DFT). Originally developed as a cognitive model to capture the deliberation process in choice making (Busemeyer and Townsend, 1992, 1993), DFT was later extended to a probabilistic-dynamic model capable of handling multiple attributes (Diederich, 1997) and further generalised to account for multiple alternative decision-making scenarios (Roe et al., 2001). In general, DFT models assume that the cognitive decision-making process is an iterative, *breadth-first* information search process, wherein individuals focus on one specific attribute at a time during each step of the deliberation process. This is in contrast with the *depth*-first process of RUM models. The DFT framework allows individuals to re-attend to the same attribute several times, evaluate the differences across alternatives, and update the preference value of the alternatives accordingly. The process continues until a decision-maker reaches a conclusion and makes a choice, either through reaching an internal preference limit (similar to satisficing) or hits an external time threshold $T_{max}$ (e.g. reaching the deadline before for which a choice has to be made). The breath-first information search assumption aligns with empirical results obtained from analysing data that captures decision-making processes, such as eye-tracking or click-tracking data (Noguchi and Stewart, 2014; Sui et al., 2020; Nova and Guevara, 2022).

Recent contributions to DFT theory have enhanced its applicability and competitiveness compared to traditional discrete choice models. For instance, Hancock et al. (2018) refined the underlying mechanisms of the DFT model, enabling it to account for the incorporation of heterogeneity both between and within decision-makers. Furthermore, Hancock et al. (2021) introduced scale parameters into the basic mechanism of the DFT model, which eliminate the requirement to conceptualise a priori parameter values that may affect model estimation and identification. Finally, Hancock et al. (2022) extended the model to include data from eye-tracking processes, to capture attribute attention weights more realistically during the deliberation process. All these contributions have also included empirical work showing that the DFT model fits the data effectively and outperforms conventional static models.

However, the DFT model has notable limitations in areas where RUM models stand out. Specifically, it relies on ad-hoc matrix implementations that are difficult to interpret and apply in practice. Its lack of clarity regarding appropriate model identification further complicates its use. Additionally, the DFT model neither benefits from a robust statistical theoretical framework nor does it align with the micro-economic theory grounded principle of Random Utility Maximisation, rendering parameter interpretation and welfare analysis infeasible.

The combined benefits of current static and dynamic models motivate the need to create a new theoretical framework. The model proposed in this article aims to keep the desirable properties of RUM models whilst also overcoming their limitations concerning the representation of the choice deliberation process. In particular, this could be achieved through the development of a RUM model that reflects cognitive dynamics, including the significant findings regarding the information search process, such as that processes are typically breadth-first, that decision-makers revisit attributes more than once, and that information is filtered.

This work thus introduces a new model: **RUM-DFT**. Alongside making improvements relative to traditional RUM structures, this new model also aims to rectify the DFT model's identification, inference and parameter interpretation limitations. Likewise,

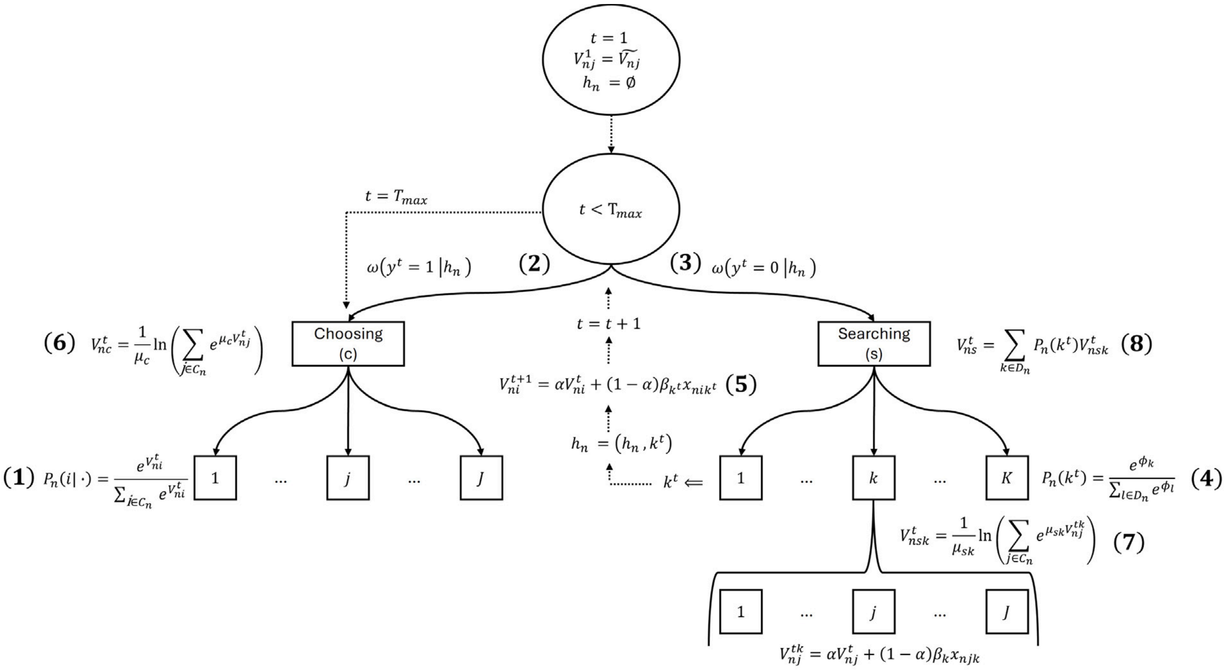


**Fig. 1.** Flowchart of the RUM-DFT model.

the new model includes parameters that allow the deliberation process to be adequately modelled with respect to the degree of the researcher's knowledge of the information search process. Thus, three model specifications are formulated and estimated, differentiated by the extent of this knowledge. The first explicitly includes the sequence of attributes attended during the choice deliberation process up to the point at which a decision is made. The second considers the deliberation time or the number of searches it takes for decision-makers, and the third one only uses traditional stated preference choice data to freely estimate the parameters associated with the attributes and those of the deliberation process.

The remainder of this paper is structured as follows. Section 2 details the assumptions and the formulation of the dynamic utilities of the random utility maximisation model that considers the information search process. Section 3 presents the data and examines the results using the RUM-DFT model on six case studies and Section 4 presents the conclusions.

## 2. RUM-DFT model formulation

The RUM-DFT model proposes that decision-makers are rational agents that carry out a breadth-first information search process in order to iteratively update the value of initial preconceived utilities before making a decision to choose the alternative with the largest utility within a choice set. This means that decision-makers acquire information about alternatives based on one attribute at a time, may consider each given attribute more than once and consecutively, until they make a choice. These assumptions make it possible to represent all the cognitive tasks that are carried out during the deliberation process, such as acquiring, processing, and evaluating the available information until individuals decide to choose an alternative. In this study, we have considered each information search carried out by the decision-maker as the basic discrete unit of the deliberation process, i.e., the sequence of attended attributes. Therefore, the formulation presented here aims to represent the discrete updating of utilities through the attributes considered up to the moment of choice. In future formulations, one could explore the possibility of including the time spent attending to each attribute, which would allow for updates of varying sizes.

### *2.1. General flowchart of the dynamic model*

The general flowchart of the dynamic model is depicted in Fig. 1, including the respective equation numbers on parenthesis. The process is initialised at time $t = 1$ by assigning preconceived utilities $\widetilde{V_{nj}}$ to each alternative $j$. At each step $t$ of the information search process, the decision-maker $n$ decides either to choose (with probability $\omega(.)$) or to update the value of the utilities by attending to an extra attribute $k \in D_n$. This iterative process continues until the individual completes the deliberation and makes a choice at $t = T_n$, either because some convergence tolerance or maximum time $T_{max}$ is reached. $T_n - 1$ represents both the total number of attributes attended during the deliberation and an approximation of the deliberation time. If the decision-maker decides to make a choice at step $T_n$, the alternative $i$ that provides the highest utility within the choice set ($C_n$) is chosen.

To address the intermediate decisions made by individual $n$, we introduce $y^t$ as the discrete decisions, where $y^t = 1$ indicates the decision to choose $(c)$ between alternatives at step $t$ and finish the search process, while $y^t = 0$ denotes the decision to search $(s)$ for information at step $t$, by attending to a specific attribute $k$.

We define the *sequence of attended attributes* $h_n$, as the unique ordered list of attributes $k \in D_n$ attended to by individual $n$ at each step $t$ of the deliberation process, until the final choice of an alternative at $t = T_n$. Note that $h_n$ has dimension $T_n - 1$ when a choice is made. As shown in Fig. 1, $h_n$ is initialised as the empty set. If the individual decides to choose at $t = T_n = 1$, $h_n$ remains empty, the choice is made based on the preconceived utilities $\widetilde{V_{nj}}$ , and the dimension of $h_n$ is 0, i.e., $T_n - 1$. In turn, if the individual decides to choose at $t = T_n = 2$, this means that a search decision was made at $t = 1$ and $h_n$ became $(k^1)$, of dimension $T_n - 1 = 1$. The same occurs for larger values of $T_n$.

The sequence $h_n$ of attended attributes may be obtained in practice from eye-tracking or click-tracking surveys, such as in Nova and Guevara (2022). Instead, when using data from conventional surveys in which $h_n$ is not observed, additional assumptions must be made about the maximum deliberation time performed by respondents to bound this value across the entire sample population, as discussed at the end of Section 2.3.

The conditional probability $P(i|y^t = 1, h_n)$ of choosing alternative $i$, having decided to make a choice at step $t$ and followed a sequence $h_n$ of attended attributes, can be calculated as shown in Eq. (1).

$$P_n(i|y^t = 1, h_n) = \frac{e^{V^t_{ni}(h_n)}}{\sum_j e^{V^t_{nj}(h_n)}} \tag{1}$$

This considers that the decision-maker chooses the alternative $i$ that maximises his/her current utility $U^t_{ni} = V^t_{ni} + \varepsilon^t_{ni}$, which has two components. The first component, $V^t_{ni}$, is known as the systematic utility, which depends on the attributes of the alternatives, according to $h_n$, as explained later in Eq. (5). $\varepsilon^t_{ni}$ is known as the random component and it is assumed to be an iid Extreme Value I (EVI) error, resulting in the Logit choice probability shown in Eq. (1) (Ben-Akiva and Lerman, 1985).

$\omega(y^t = 1|h_n)$ in Fig. 1 is the probability that the decision-maker has decided to choose at step $t$ (i.e. $y^t = 1$), conditional on the sequence of attended attributes $h_n$, which is calculated as follows:

$$\omega(y^t = 1|h_n) = \begin{cases} P(|U^t_{nc} - U^t_{ns}| \leq \delta t^2) & \text{if } t < T_{max} \\ 1 & \text{if } t = T_{max} \end{cases} \tag{2}$$

$$\omega\left(y^t = 0|h_n\right) = 1 - \omega(y^t = 1|h_n) \tag{3}$$

If $t = T_{max}$ in Eq. (2), meaning that the decision deadline has been reached, the individual makes a choice. Otherwise, if $t < T_{max}$, the decision to stop searching and choose depends on the difference between $U^t_{nc}$ —the utility of choosing $(c)$ in step $t$— and $U^t_{ns}$ —the utility to continue searching $(s)$. These utilities will be developed in Section 2.2.

The stopping criterion uses the absolute value of the difference, implying that the decision-making process interprets any sufficiently large discrepancy —whether positive or negative— as evidence of incomplete utility convergence, thereby implying a need for further information search. At each step of the search process, the individual gathers new information about the actual utility they will experience from their choice, suggesting that any significant change — regardless of direction — is worth exploring. This is similar to the idea of collapsing boundaries for preference accumulation, a common feature in choice models based on psychological theory (Voskuilen et al., 2016).

$\delta$ in Eq. (2) is the tolerance parameter that represents the existence of a cognitive cost. By incorporating $\delta t^2$ as the threshold, the model captures the idea that decision-makers grow more tolerant of discrepancies over time. As the search process continues, they accept progressively larger mismatches between $U^t_{nc}$ and $U^t_{ns}$ before halting information acquisition. While $\delta$ is treated as a population parameter in this formulation, it could theoretically vary across individuals to account for heterogeneity in decision-making behaviours, such as differences in patience or cognitive capacity (Kahneman, 2011).

This formulation of $\omega(\cdot)$ allows us to model under the assumption that the decision-maker could decide to choose for two distinct reasons. The first may be due to an external limitation that forces one to choose an alternative in a maximum deliberation time $(T_{max})$, at which time the decision-maker is forced to choose the alternative with the highest preference value (maximising behaviour, Schwartz et al. (2002)). The second reason corresponds to the point at which utility potentially attainable by continue the seach is similar enough to the current utility. This threshold is determined by individual's internal tolerance, leading to stability in the preference levels close to the decision time (Ariely and Zakay, 2001). This implies that $T_n$, and by transitivity $h_n$ cannot be infinite since the internal tolerance prompts the individual to choose at some point.

At step $t$, when the decision-maker decides to perform an information search, the next attribute $k$ to update in the utility values must be selected. This probability is modelled using a Logit model of constants $\phi_k$ $\forall k \in D_n$, as shown in Eq. (4). The constants $\phi_k$ are somehow equivalent to the attention weights used in DFT models, reflecting the degree of focus and potential filtering applied to specific attributes. They are treated as exogenous, meaning they are not derived from the coefficients of the choice model nor influenced by the current estimate of utility. This is done because searching for attributes that maximise the current estimate of utility would prevent the exploration of the true value of the utility and, instead, to focus mostly on those attributes that are more positively valued.

$$P_n\left(k^t\right) = \frac{e^{\phi_k}}{\sum_{l \in D_n} e^{\phi_l}} \tag{4}$$

If the decision-maker opts to continue searching, i.e., $y^t = 0$, and an attribute $k^t$ is chosen at step $t$, the loop proceeds to the next step. As shown in Fig. 1, $t$ is incremented by one $(t = t + 1)$, utilities are updated (as shown in Section 2.2), and the sequence of attended attributes is updated to $h_n = (h_n, k^t)$.

### 2.2. Utility functions and microeconomic interpretability

The RUM-DFT model, like the RUM, considers that decision-maker $n$ chooses the alternative $i$ with the highest random utility within the choice set ($C_n$). However, different from RUM, RUM-DFT proposes a functional form of utilities that simulates the dynamics of updating the preference values of decision-makers considering the information search process they have carried out, which allows for a more realistic and accurate cognitive context. In general, the updated utility obtained from alternative $i$ by individual $n$ at step $t+1$ is:

$$U_{ni}^{t+1} = V_{ni}^{t+1} + \varepsilon_{ni}^{t+1} = \alpha V_{ni}^{t} + (1-\alpha)\beta_{k^t} x_{nik^t} + \varepsilon_{ni}^{t+1}, \tag{5}$$

where $\alpha$ is the memory parameter that represents the influence of current utilities on the next deliberation step $t+1$. In this formulation, $\alpha$ is considered a constant, between 0 and 1, and represents a rate of forgetting that is invariant throughout the deliberation process. This implies that information about attributes that are attended to at the beginning of the attention sequence start to be forgotten as the utility is updated in the subsequent steps. $\beta_{k^t}$ is the parameter of the attribute $k$ attended in step $t$ and $x_{nik^t}$ corresponds to the value of attribute $k$ of alternative $i$ for decision-maker $n$ attended in step $t$. $\varepsilon_{ni}^{t+1}$ is an error term that captures the imperfect knowledge of the researcher or unobserved heterogeneity, which is assumed to be *i.i.d.* EVI.

Considering EVI errors at each update step $t$ requires some discussion. This assumption allows us to account, with closed forms, for intermediate stochastic choices within the scrutiny process. Similar to the nested logit model (Ben-Akiva and Lerman, 1985), this framework also allow, through appropriate use of scales parameters, for the possible correlation between sequences of attended attributes that have common paths. Additionally, it permits the error variance to evolve during the deliberation process, adapting as more attributes are explored. The use of EVI errors in sequential choices has been previously considered in several studies. For example, Ben-Akiva and Lerman (1985) used it in the derivation of the nested logit model; Train (2009), following Rust (1987), used it in the modelling of choices occurring in two consecutive time periods; and Xie (2019) used it to model a case in which attributes are explored sequentially to reduce the variance of unknown factors. This requires making assumptions about the joint distribution of errors, which may be problematic for EVI, but feasible, as in the previously cited articles. More formally, one may consider that the joint distribution of $\varepsilon_{ni}^{t}$ follows a G(M)EV distribution (McFadden, 1978) that fulfils the required features. Alternatively, one may consider that the errors are normally distributed, and the closed form is just viewed as an approximation (Ruud, 1983).

Thus, the utility functional form of this model allows capturing both the phenomenon of forgetting the information of previously considered attributes due to the memory effect, and the explicit incorporation of the attribute value at step $t$ of the deliberation process into the current utility. In this way, $V_{ni}^{t}$ includes the recursive evolution of decision-maker $n$'s utility of alternative $i$ up to step $t$, in which s/he has attended to the sequence of attributes $h_n$ in his/her deliberation process.

The utility of choosing ($c$) and the utility of continuing to search ($s$) can thus be developed. On the one hand, the utility of choosing for decision-maker $n$ at step $t$, $U_{nc}^{t}$, is defined as the expected maximum of all current utilities $U_{ni}^{t}$, across alternatives $i \in C_n$, considering the sequence $h_n$ of attributes attended to up to that step. Following (Ben-Akiva and Lerman, 1985, p. 106), it can be shown that $U_{nc}^{t}$ follows an EVI distribution with location $V_{nc}^{t}$ and scale $\mu_c$, as shown in Eq. (6). Besides, $U_{nc}^{t}$ can be written as $U_{nc}^{t} = V_{nc}^{t} + \varepsilon_{nc}^{t}$, where $\varepsilon_{nc}^{t}$ follows an EVI (0, $\mu_c$) distribution, as shown in Fig. 1.

$$U_{nc}^{t} = \max_{i\in C_n}\left(V_{ni}^{t} + \varepsilon_{ni}^{t}\right) = V_{nc}^{t} + \varepsilon_{nc}^{t} \quad \sim \quad EVI\left(V_{nc}^{t} = \frac{1}{\mu_c} ln\left(\sum_{j\in C_n} e^{\mu_c V_{nj}^{t}(h_n)}\right), \mu_c\right) \tag{6}$$

On the other hand, the decision-maker is assumed to consider the utility of continuing searching, $U_{ns}^{t}$, as a two-level process. At the lower level, $U_{nsk}^{t}$ corresponds to the expected maximum utility that could be attained by attending to the $k_{th}$ attribute for all alternatives at an hypothetical future step $t+1$, considering the complete sequence of attributes attended up to that step, plus the respective $k$th attribute, i.e., the sequence corresponds to $[h_n, k]$, as shown on the lower right of Fig. 1. $U_{nsk}^{t}$ is distributed EVI with scale $\mu_{sk}$ and location (and systematic utility) $V_{nsk}^{t}$, as shown in Eq. (7) and Fig. 1. Note that this calculation assumes the future assessment involves only one additional step, neglecting all future utilities beyond $t+1$, using the same logic utilised by Xie (2019) and Chorus and Bierlaire (2013) in other frameworks.

$$V_{nsk}^{t} = \frac{1}{\mu_{sk}} ln\left(\sum_{j\in C_n} e^{\mu_{sk}\left(\alpha V_{nj}^{t} + (1-\alpha)\beta_k x_{njk}\right)}\right) \tag{7}$$

Then, at the upper level, the utility of continuing the search, $U_{ns}^{t}$ is modeled as a weighted average of the expected maximum utilities attainable by exploring each individual attribute $k$, denoted as $U_{nsk}^{t}$. For practical convenience, $U_{ns}^{t}$ is distributed EVI with scale $\mu_s$ and location (and systematic utility) $V_{ns}^{t}$, as shown in Eq. (8).

$$V_{ns}^{t} = \sum_{k\in D_n} P_n(k^t) V_{nsk}^{t} \tag{8}$$

$$V_{ns}^{t} = \sum_{k\in D_n}\left[\frac{e^{\phi_k}}{\sum_{l\in D_n} e^{\phi_l}} \frac{1}{\mu_{sk}} ln\left(\sum_{j\in C_n} e^{\mu_{sk}\left(\alpha V_{nj}^{t} + (1-\alpha)\beta_k x_{njk}\right)}\right)\right] \tag{9}$$

This nested formulation is based on the assumption that decision-makers perform a breadth-first information search process, which implies that they attend to one attribute at a time to update utilities. That is, they first calculate the maximum utility of attending to a specific attribute, across all alternatives, and then calculate their expected utility across attributes, to construct the future utility. However, this formulation could be varied to represent other information search patterns, such as in-depth, diagonal adjacency, or non-adjacency, as described by Nova and Guevara (2022). Note also that an alternative to Eq. (8) would be to calculate $V_{ns}^t$ as the logsum of $V_{nsk}^t$. However, this approach was discarded as it would systematically bias exploration toward better attributes, which is not behaviorally consistent.

Once we have defined the utility for making a choice ($U_{nc}^t$) and the utility of continue searching to an extra attribute ($U_{ns}^t$), we can formulate the probability that the decision-maker decides to choose $\omega(y^t = 1|h_n)$. Introducing the Eqs. (6), (7) and (8) into the Eq. (2), we have:

$$\omega(y^t = 1|h_n, t < T_{max}) = P\left(\left|V_{nc}^t + \varepsilon_{nc}^t - V_{ns}^t - \varepsilon_{ns}^t\right| \leq \delta t^2\right). \tag{10}$$

Thus, assuming that the difference of the errors of choosing and continue searching are i.i.d. EVI, $\omega(\cdot)$ from Eq. (2) can be shown to be:

$$\omega(y^t = 1|h_n, t < T_{max}) = \frac{1}{1 + e^{-\left(V_{ns}^t - V_{nc}^t + \delta t^2\right)}} - \frac{1}{1 + e^{-\left(V_{ns}^t - V_{nc}^t - \delta t^2\right)}}. \tag{11}$$

Building the RUM-DFT model explicitly from utilities preserves its microeconomic interpretability. This is not the case for the DFT for which one may only devise numerical methods to infer a proxy of the marginal rates of substitution between attributes impacting choices, while it is not possible to build formal welfare measures from them.

Marginal utilities can be calculated from RUM-DFT estimates as shown in Eq. (12), where $1_{(k^l=k)}$ is an indicator function that takes value 1 if attribute $k$ is attended at step $l$. These estimates may then be used to calculate marginal rates of substitution between attributes, a concept commonly used in transportation models to estimate the subjective value of travel time savings (see, e.g., Jara-Díaz and Guevara (2003)). Besides, regarding welfare analysis, since the estimates of the RUM-DFT provide indirect utilities, nothing precludes the use of measures of compensated variation, or marshallian surplus, like the ones proposed by Small and Rosen (1981) in this case.

$$\frac{\partial V_{ni}^{T_n}}{\partial x_{ik}} = \beta_k (1 - \alpha) \sum_{l=1}^{T_n - 1} \alpha^{T_n - l - 1} \mathbf{1}_{(k^l = k)} \tag{12}$$

Note that Eq. (12) implies that, for being able to calculate marginal utilities, attribute $k$ must have been attended to at least once. This also implies that individuals' marginal utilities and willingness-to-pay may vary across choice tasks, depending on the sequence of attended attributes, an heterogeneity that does not occur in static models. The possible implications of this result are left for future research, but it can be conjectured that obtaining a person-specific measure of willingness to pay would require integration over possible sequences.

### *2.3. Calculating the likelihood under various data availability scenarios*

For calculating the likelihood we consider three scenarios for the type of data that might be available for estimation, named as *Information Search* (IS), *Deliberation Time* (DT) and *Stated Choice* (SC). IS corresponds to a case in which the analyst knows the detailed sequence of attributes attended to until the choice is made, with, e.g., eye-trackers. In DT, only deliberation time is recorded instead. SC corresponds to a traditional database in which only the stated choice is recorded.

For the Information Search (IS) case data setting, the whole sequence $h_n$, as well as $T_n$, are known to the researcher. The likelihood $L_{IS}$ of observing the choices $i_n$ across all individuals $n$, and their respective $h_n$, corresponds to Eq. (13). $P_n(i_n, y^{T_n} = 1, h_n)$ is the joint probability of observing the choice $i_n$ for individual $n$, making a decision to choose at step $T_n$, and to have followed the path $h_n$. Under the nesting structure assumed, this probability can be written as the product of three terms. The rightmost term accounts for the likelihood of making $T_n - 1$ decisions to search for information to form the sequence $h_n$ of attributes $k^t$. The middle term accounts for the conditional probability of deciding to choose at $T_n$, given the sequence $h_n$. The leftmost term of Eq. (13) corresponds to the conditional probability of choosing $i_n$, given the path $h_n$ and the decision to make a choice at $T_n$.

$$L_{IS} = \prod_{n=1}^{N} P_n(i_n, y^{T_n} = 1, h_n) = \prod_{n=1}^{N} \left( \frac{e^{V_{ni_n}^{\tau}(h_n)}}{\sum_{j \in D_n} e^{V_{nj}^{\tau}(h_n)}} \omega(y^{T_n} = 1|h_n) \prod_{t=1}^{T_n - 1} \left\{ \omega(y^t = 0|h_n) P_n(k^t) \right\} \right) \tag{13}$$

The model becomes more burdensome to estimate when less information is available. For example, in many cases, instead of having the complete $h_n$, one might only have access to information about Deliberation Time (DT), particularly when choices are collected using computer-assisted tools or web surveys. If DT is homogeneously related to the number of steps $T_n$ or *fixations*, it becomes possible to estimate $T_n$ from DT. Diverse evidence supports this assumption. For instance, Stewart et al. (2016a) and Glöckner and Herbold (2011) found that fixation times on attributes in both simple and complex choice tasks are relatively brief and do not depend entirely on attribute values. Additionally, Nova and Guevara (2022) demonstrate a relationship between the number of attributes displayed in a choice task and the average fixation time across tasks. Based on these findings, it seems plausible that the total number of attended attributes, $T_n$, can be derived as a function of the total time spent on all fixations (DT) and the estimated time per fixation, which remains stable regardless of the attribute value.

If $T_n$ is known, but not the sequence $h_n$, the likelihood $L_{DT}$ of observing the choice of an alternative $i_n$ has to be integrated over the set $H_n^{T_n}$ of all possible sequences of attributes attended $\eta_n \in H_n^{T_n}$ that may have been followed by each individual $n$ up to $T_n$, and its respective probability, as shown in Eq. (14). As an approximation, in the application shown later in Section 3, $P(\eta_n)$ is considered equiprobable to reduce the complexity of the modelling it with a classical approach. It is conjectured that a Bayesian approach may allow us to avoid this simplification, a line of research that is left for the future.

$$L_{DT} = \prod_{n}^{N} \sum_{\eta_n \in H_n^{T_n}} P_n(i_n, y_n^{T_n} = 1, \eta_n) P(\eta_n) \tag{14}$$

The final case corresponds to a situation in which no information about the deliberation process is available, just the Stated Choice (SC). The classical probability approach to determine the likelihood $L_{SC}$ in this case requires integrating not only over the sequences $\eta_n$ but also over all possible deliberation times $\tau_n \in \Theta$, which are assumed to be discrete in Eq. (15), and their respective probabilities.

$$L_{SC} = \prod_{n}^{N} \sum_{\tau_n \in \Theta} \sum_{\eta_n \in H_n^{\tau_n}} P_n(i_n, y^{\tau_n} = 1, \eta_n) P(y_n^{\tau_n} = 1) P(\eta_n) \tag{15}$$

This approach quickly becomes impractical. Two simplifications are explored in the empirical application in Section 3, where $P(\eta_n)$ is assumed to be equiprobable, and $L_{SC}$ is evaluated only for a given set of $\tau$.

## 3. Empirical analysis

This section comprises four subsections. Section 3.1 begins with a simulation of a RUM-DFT data generation process (DGP) to evaluate the proposed model's ability to replicate breadth-first choice behaviour and the deliberation process, both of which are observed in practice and are reproducible using the DFT model. Then, Section 3.2 comprises a Monte Carlo experiment to explore the feasibility of recovering the model parameters of the proposed RUM-DFT model under three scenarios of data availability. Then, in Section 3.3, RUM-DFT is compared with RUM and DFT in terms of model fit and parameter recovery under RUM and DFT DGPs, in additional Monte Carlo experiments. Finally, Section 3.4 compares RUM-DFT, RUM, DFT and two variations of the RRM model (van Cranenburgh et al., 2015), using the *Swiss Metro* (Bierlaire et al., 2001) real stated choice dataset as a test bed. These data and codes have been made available to facilitate replication and future research (Nova et al., 2025).

The simulated dataset for the Monte Carlo experiment used for Sections 3.1 to 3.3 consists of 4000 observations, representing mode choice scenarios for morning commutes. Each scenario includes four alternatives — bus, car, metro, and bike — characterised by two attributes: travel time and travel cost. In a variation of the experiment, waiting time is also included as an additional attribute. Attribute values are drawn from uniform distributions, with ranges aligned to typical values for these modes in Santiago, Chile. The parameter values used for simulated data generation vary across RUM-DFT, RUM, and DFT models, as follows. Travel time ($\beta_{tt}$) and cost ($\beta_c$) coefficients are consistently −2.0 and −0.5, respectively, in RUM-DFT and RUM. Attention weights for travel time ($\phi_t$) and cost ($\phi_c$) are 0.0 and 0.3 in RUM-DFT, while DFT assigns $\phi_c$ a value of −0.738. Memory ($\alpha$) is 0.6, tolerance ($\delta$) is 0.05, and scale parameters ($\mu_c$, $\mu_s$, $\mu_{sk}$) are all 0.1 in RUM-DFT. Alternative Specific Constants (ASC) vary, with initial utilities in DFT for bike, bus, metro, and car at −2.889, −4.473, 0.0, and 0.627, respectively. DFT weights are derived using a logistic transformation, and initial preferences are calibrated to ensure similar modal shares.

### 3.1. Illustration and assessment of RUM-DFT's dynamic Information Search behaviour

This subsection employs simulated data to illustrate and analyse the dynamic behaviour of utilities in the RUM-DFT model, focusing also on how the memory parameter influences the convexity of the model's Log-Likelihood function.

Fig. 2 illustrates the evolution of utilities for a given decision-maker in the choice situation in the Monte Carlo experiment. At each step of the deliberation process, the decision-maker updates utility values by focusing on one attribute at a time. The left panel considers travel time and cost as attributes, while the right panel includes also waiting time. In both cases, the utilities begin with simulated initial preconceived values and evolve based on the information searches conducted by the decision-maker.

When the choice situation consists of four alternatives and two attributes, the decision-maker follows this sequence of attending attributes: $h_n = \{TC, TT, TT, TC\}$, and then chooses the Bus. At step $t = 5 = T_n$, the individual anticipates that future utilities will not deviate significantly from current utilities (i.e., less than the internal tolerance $\delta t^2$), leading to the selection of Bus, as its utility is the highest among the alternatives at that deliberation step. On the other hand, the right panel of Fig. 2 depicts the evolution of utilities for the same decision-maker in the same choice situation but with waiting time added as an additional attribute. This inclusion affects the perception of alternatives and alters the information search process. In this case, at step $t = 7$, the decision-maker again selects Bus, as future utilities are expected to remain stable, and Bus has the highest utility. These graphs illustrate that the model exhibits the desired and expected dynamics: decision-makers revisit attributes, utilities fluctuate based on the attribute attended to, and as the complexity of the choice situation increases, the number of information searches conducted also rises.

Fig. 2 illustrates also the utilities derived from the random utility maximisation (RUM) approach (dashed lines), showing that they remain constant throughout the information search process. Notably, when the decision-maker finalises their choice, the Bus alternative generates the highest utility in both approaches. However, if an external constraint were imposed, forcing the decision-maker to choose earlier, the RUM model would fail to account for the incomplete consideration of available information. Consequently, the alternative with the highest utility could differ between the two approaches. This behaviour aligns with findings by Miletić and van Maanen (2019), which demonstrate how the accuracy of time-internal representations influences choice behaviour when decisions are constrained by a deadline.

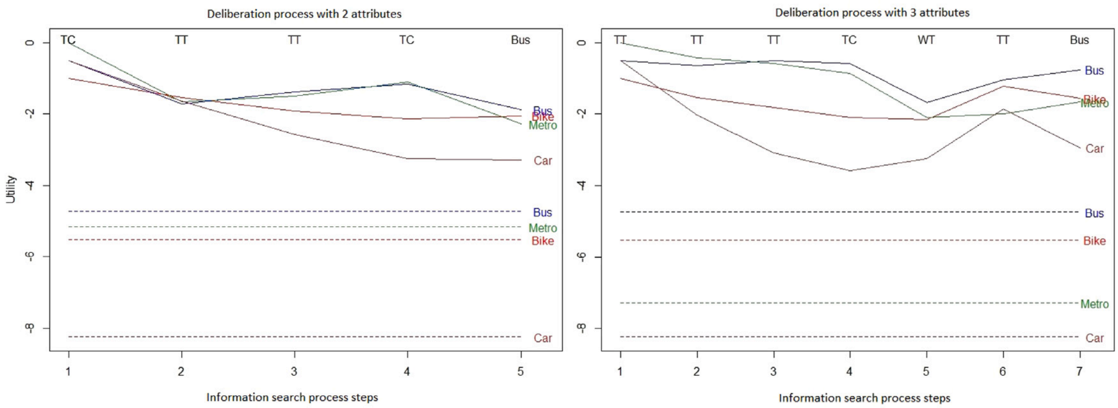


**Fig. 2.** RUM-DFT dynamics of utilities in a decision-makers information search process. Solid lines: RUM-DFT Utilities. Segmented lines: RUM Utilities. Left panel: 4 alternatives and 2 attributes. Right panel: 4 alternatives and 3 attributes.

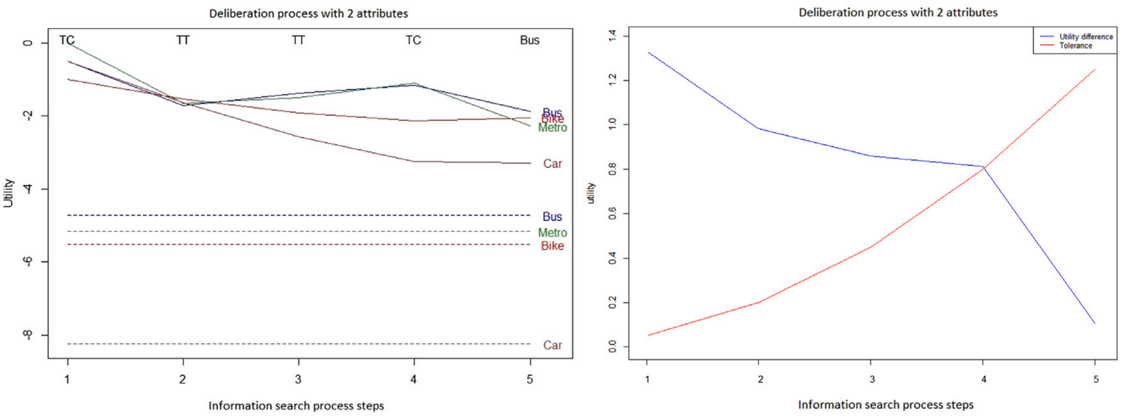


**Fig. 3.** Absolute difference between $U_{nc}^{t}$ and $U_{ns}^{t}$ and tolerance evolution in RUM-DFT Model.

Fig. 3 (right) depicts the absolute difference between $U_{nc}^{t}$ and $U_{ns}^{t}$, alongside the individual's tolerance evolution, $\delta t^2$, throughout the deliberation process. This visualisation illustrates the RUM-DFT Model mechanism. When the gap between present and future utilities falls below the individual's tolerance, the decision-maker ceases the information search and selects an alternative. In the final timestep, it becomes evident that incorporating additional information or attributes would yield negligible utility improvement, incurring a cognitive cost exceeding the individual's tolerance at that point.

Finally, note that the memory factor introduces non-convexity into the log-likelihood function due to the Markovian utility dynamics, where $\alpha$ incorporates information from previous steps. This dependence, evident in the second derivative of the systematic utility with respect to $\alpha$, highlights its relationship with $\alpha$ and the coefficients ($\beta_k$). Careful estimation is required to avoid convergence to local optima. To address the problem, we use a fixed and known memory parameter ($\alpha$) to simplify the model and avoid non-convex local optima. Estimation is done in two stages: first, global optimisation techniques such as DEoptim (Mullen et al., 2011) provide initial parameter estimates, though they are slow and lack standard errors. DEoptim is particularly effective for non-convex problems as it explores the entire parameter space, reducing the risk of being trapped in local optima compared to gradient-based methods. Second, $\alpha$ is fixed, and local optimisation refines the log-likelihood for statistical testing. Simulated data tests confirm this approach yields reliable and meaningful estimates.

### 3.2. Estimation of RUM-DFT under different degrees of available information

This section uses Monte Carlo data to evaluate the RUM-DFT estimation under a RUM-DFT data generation process (DGP), for three information levels: full information search (IS), deliberation time (DT), and stated choice only (SC). The IS case is resolved completely and for DT and SC, exploratory simplifications are tested. For the DT case, RUM-DFT is compared with two alternative DFT models. For SC, it is also compared with RUM, specifically a Multinomial Logit with coefficients for travel time,

**Table 1**
Estimation results of RUM-DFT model IS experiment.

| 20 Simulations | IS: RUM-DFT DGP & full information | | | |
|---|---|---|---|---|
| Parameters | Value ($\beta_{true}$) | Average estimates ($\tilde{\beta}$) | S.E | $t(\tilde{\beta} - \beta_{true})$ |
| Travel time $\beta_{tt}$ | −2.000 | −2.039 | 0.157 | 0.253 |
| Travel cost $\beta_c$ | −0.500 | −0.508 | 0.031 | 0.281 |
| Attention weight travel cost $\phi_c$ | 0.300 | 0.297 | 0.014 | 0.226 |
| Memory $\alpha$ | 0.600 | 0.602 | 0.029 | −0.005 |
| Tolerance $\delta$ | 0.050 | 0.049 | 0.001 | 0.226 |
| Initial utility Bike $U_A^0$ | −1.000 | −0.979 | 0.194 | −0.105 |
| Initial utility Bus $U_B^0$ | −0.500 | −0.462 | 0.153 | 0.002 |
| Initial utility Car $U_D^0$ | −0.500 | −0.503 | 0.157 | −0.001 |
| Scale of current utilities $\mu_c$ | 0.100 | 0.140 | 0.044 | −0.890 |
| Scale of future utilities $\mu_s$ | 0.100 | 0.139 | 0.046 | −0.853 |
| LL | | −20,553.304 | | |
| AIC | | −41,126.608 | | |
| BIC | | −41,189.548 | | |

cost, and alternative-specific constants. The Basic-DFT includes attention weights (Hancock et al., 2018), the Scaled-DFT adds scaling parameters (Hancock et al., 2021), and the Process-DFT incorporates fixation counts and attribute times (Hancock et al., 2022).

*Information Search (IS) Experiment*

Consider first the Information Search (IS) experiment, in which a RUM-DFT *data generation process (DGP)* is considered and the *information* on the sequences of attended attributes $h_n$ is known to the researcher. In this case, 20 simulations and estimates for the database were generated using a RUM-DFT data generation process. With this information, one can explicitly calculate the joint likelihood of observing a sequence $h_n$ and a choice, as shown in Eq. (13). Table 1 shows that the RUM-DFT model estimated in this case can retrieve all parameters associated with the deliberation process, including Memory, Tolerance, and Preconceived Utilities, as shown by the test depicted in the last column, using the habitual 5% threshold.

*Deliberation Time (DT) Experiment*

Now consider the DT experiment in which the DGP is also RUM-DFT, but now only information on the Deliberation Time (DT) is available to the researcher. While $h_n$ is unknown, the deliberation time of each respondent can serve as a proxy for the number of steps $T_n$ in the information search process. In the RUM-DFT-DT model, all possible sequences of attended attributes, $\eta_n \in H_n^{T_n}$, up to the decision-maker's $T_n$, are included in the likelihood calculation, as shown in Eq. (14). To reduce the computational complexity of accounting for these multiple sequences, two simplifications were applied in this exploratory analysis. First, only five simulations (compared to 20 in Table 1) were performed. Second, we assumed that all sequences are equally probable, effectively excluding the probability of a decision-maker following a specific sequence of attended attributes. Consequently, tolerance and scale parameters are not estimated. These simplifications could be addressed by adopting a Bayesian approach, a direction reserved for future research.

This analysis compares the proposed RUM-DFT model with two Process-DFT model variants (Hancock et al., 2021) that incorporate the total number of fixations (representing information search) into the attention weights: Scaled-DFT-1 and Scaled-DFT-2. Scaled-DFT-1 estimates attention weight parameters ($\beta_k$) to capture the relative importance of the two attributes in the choice situation, while Scaled-DFT-2 estimates only the scale parameter ($\gamma_k$), with attention weights fixed at 1. The comparison is reasonable since Process-DFT models aggregate fixations to represent deliberation time. For estimating Scaled-DFT-1 and Scaled-DFT-2, we used the known number of fixations for each choice task and decision-maker. The mean deliberation time was 5.3, with a maximum of 10.

Table 2 presents the estimation results of the RUM-DFT-DT model. According to these results, the parameters estimated using the proposed model closely approximate their true values, notably the memory factor ($\alpha$) and the travel cost coefficient ($\beta_c$). Furthermore, it can be observed that the initially preconceived utilities, while close to those values with which they were generated, present a higher standard deviation compared to the coefficients obtained when the sequence of attended attributes is known. Similarly, when comparing the model fitting between this version and the RUM-DFT-IS model, the RUM-DFT-DT model obtains a higher log-likelihood. This difference arises because the RUM-DFT-DT model excludes intermediate decisions related to the attention process. Therefore, a model that considers only the deliberation time will generate parameters with a higher standard deviation and will always have a higher or equal log-likelihood than an approach that incorporates the actual information search process.

In the same Table 2, the estimation results for two versions of the Process-DFT model are also presented. Under these approaches, there is little benefit in commenting on the estimates of the scale parameters or attention weights since these values do not make microeconomic sense and are not representative of the data generation process. However, in both specifications, it can be observed that the cost attention weight has a lower impact than travel time in the deliberation process, i.e. the attribute associated with time is less considered by decision-makers when updating their preferences. This is because $\gamma_c$ and $\gamma_{tt}$ or $\beta_c$ and $\beta_{tt}$ represent logistic attention parameters. Also, in both Scaled-DFT versions, it can be seen that the sensitivity ($\phi_{sen}$) and memory ($\phi_{mem}$) parameters are not statistically different from zero. First, it is reasonable that Scaled-DFT does not fully capture the first parameter since correlated alternatives were not simulated in the data generation process. Secondly, it is unsurprising that the parameter associated with memory is insignificant since the sequence of attributes attended in this approach is not precisely known. In addition, a slightly

**Table 2**
Estimation results under RUM-DFT data generation process and known DT ($T_n$).

| 5 Simulations | DT: RUM-DFT DGP & DT information | | | |
|---|---|---|---|---|
| Parameters | Value ($\beta_{true}$) | RUM-DFT | Scaled-DFT-1 | Scaled-DFT-2 |
| Travel time $\beta_c$ | −2.000 | −2.029 (0.295) | 0.359 (0.023) | |
| Travel cost $\beta_{tt}$ | −0.500 | −0.560 (0.122) | | |
| Memory $\alpha$ | 0.600 | 0.628 (0.120) | | |
| Attention weight travel cost $\phi_c$ | 0.300 | 0.284 (0.247) | | |
| Attention weight travel cost $\gamma_c$ | | | | −1.001 (0.082) |
| Fixation $\alpha_f$ | | | 0.497 (0.044) | 0.501 (0.035) |
| Sensitivity $\phi_{sen}$ | | | 5.934 (1.229) | 6.087 (0.065) |
| Memory $\phi_{mem}$ | | | −0.024 (0.030) | −0.033 (0.002) |
| Error $\sigma_\epsilon$ | | | 5.020 (0.537) | 4.764 (0.501) |
| Preferences update steps $\tau$ | | | 1.799 ( 0.268) | 1.652 (0.237) |
| Initial utility/Preference Bike $U_A^0$ | −1.000 | −0.784 (0.353) | −1.979 (0.561) | −1.413 (0.509) |
| Initial utility/Preference Bus $U_B^0$ | −0.500 | −0.544 (0.200) | −1.307 (0.509) | −0.634 (0.414) |
| Initial utility/Preference Car $U_D^0$ | −0.500 | −0.519 (0.352) | −0.944 (0.288) | −0.552 (0.449) |
| LL | | −4804.176 | −4830.954 | −4846.360 |
| $\bar{\rho}^2$ | | 0.132 | 0.127 | 0.125 |
| AIC | | 9622.352 | 9677.908 | 9708.72 |
| BIC | | 9666.410 | 9728.260 | 9759.08 |

RUM-DFT estimated with equiprobable $h_n$.

better fit for the first model can be observed compared to the one that estimates the scale. This result is similar to the one found in Hancock et al. (2022) and contrary to the one shown in Hancock et al. (2021), who found that DFT models with estimated scale parameters consistently outperform DFT models with estimated attention weights. In this particular case, there is no significant gain of log-likelihood. Consequently, it is preferable to consider the Scaled-DFT-DT2 model since it avoids the identification and estimation problems of the first model, which are discussed in Hancock et al. (2022).

More significantly, the goodness-of-fit indicators of both Scaled-DFT models have worse performance than the proposed RUM-DFT model when observing the $\bar{\rho}^2$, AIC and BIC. This is expected since the data generation process followed the proposed model's approach. Nevertheless, the memory factor ($\alpha$) can only be significantly recovered using the RUM-DFT-DT model. Likewise, it can reasonably account for the acquisition and processing of the information sought before choice, not purely from the attention weights and the random error vector in the Scaled-DFT-DT2 model. The superiority in the goodness-of-fit and the addition of parameters of the deliberation process are reasonable indications and must be tested in a real database for conclusive results.

*Stated Choice (SC) Experiment*

The third experiment (SC) considers also a RUM-DFT DGP, but neither the sequence of attributes attended nor the deliberation time are known to the researcher, just the stated choice (SC). This makes the optimisation process much more complex, requiring integration over search sequences and deliberation time, in a problem that is in general non-convex. We address this challenge in practice by making three assumptions. First, we set a maximum value of the deliberation time for the whole population ($T_{max}$). For this we explored the impact of considering different levels of it from 5 to 10, finding equivalent results. The outcome with $T_{max} = 5$ is reported in Table 3. The second assumption is to fix the scales of future/present utility nests, and the third, is to assume that the search sequences $h_n$ are equiprobable. The investigation of the impact that these simplifying assumptions may have in practice is left for further research, but it can be speculated that they may be better addressed instead by a Bayesian approach.

Table 3 shows the estimation results of the RUM-DFT model with a maximum deliberation time of 5 steps, the scaled DFT models that do not include fixations, and a Multinomial Logit model. The first thing that stands out is that the proposed model has the best fit compared the other approaches. All parameters associated with the attributes (attention weights in the DFT model) differ significantly from zero in all models. Also, note that the RUM-DFT model provides estimates close in magnitude and sign to the true values. This is important because, although the data generation process (DGP) in this experiment followed the RUM-DFT framework, certain assumptions were necessary to reduce the computational burden. The successful recovery of the population parameters suggests that these assumptions were not overly restrictive.

In the RUM-DFT model, the coefficients for each attribute are associated with the indirect marginal utility of the respective attribute, under a framework of complete information, as expressed in Eq. (12). For this SC framework, where $h_n$ is unknown, calculating a distribution of the marginal utilities, and thus the value of time, would be possible, but would require accounting for the probabilities of deliberation time and the sequence of attribute attendance. A more in-depth analysis of the potential bias in estimating the value of time when using a RUM model that ignores the deliberation process in practice is left for future research.

### 3.3. Assessment under RUM and DFT Data Generation Processes (DGP)

In this section, we provide insights into the fit of the RUM-DFT model and parameter recovery when estimated on simulated mode choices generated under the assumption that the data generation process is Random Utility Maximisation or Decision Field

**Table 3**
Estimation results for RUM-DFT data generation process and Stated Choice (SC).

| 1 Simulation | SC: RUM-DFT DGP & SC information | | | | |
|---|---|---|---|---|---|
| Parameters | Value ($\beta_{true}$) | RUM-DFT | Scaled-DFT-1 | Scaled-DFT-2 | RUM |
| Travel time $\beta_c$ | −2.000 | −2.226 (0.676) | 0.331 (0.025) | | −0.207 (0.009) |
| Travel cost $\beta_{tt}$ | −0.500 | −0.788 (0.499) | | | −0.758 (0.044) |
| Memory $\alpha$ | 0.600 | 0.765 (0.193) | | | |
| Attention weight travel cost $\phi_c$ | 0.300 | −0.038 (0.315) | | | |
| Attention weight travel cost $\gamma_c$ | | | | −1.119 (0.076) | |
| Sensitivity $\phi_{sen}$ | | | 4.976 (2.897) | 2.624 (0.109) | |
| Memory $\phi_{mem}$ | | | 0.105 (0.053) | −0.028 (0.039) | |
| Error $\sigma_\epsilon$ | | | 5.313 (0.622) | 5.230 (0.575) | |
| Preferences update steps $\tau$ | | | 2.111 (0.262) | 2.016 (0.242) | |
| Initial utility/Preference Bike $U_A^0/ASC_A$ | −1.000 | −0.676 (0.423) | −2.217 (1.451) | −1.742 (0.468) | −0.034 (0.060) |
| Initial utility/Preference Bus $U_B^0/ASC_B$ | −0.500 | −0.325 (0.357) | −0.832 (1.015) | −0.742 (0.518) | −0.087 (0.048) |
| Initial utility/Preference Car $U_D^0$ | −0.500 | −0.119 (0.216) | −0.777 (1.283) | −0.948 (0.510) | −0.013 (0.056) |
| LL | | −4817.923 | −4836.140 | −4839.38 | −4853.264 |
| $\overline{\rho}^2$ | | 0.130 | 0.126 | 0.125 | 0.123 |
| AIC | | 9649.846 | 9688.290 | 9692.76 | 9716.528 |
| BIC | | 9693.904 | 9738.64 | 9743.112 | 9747.998 |

RUM-DFT estimated with $T_{max}$=5 and equiprobable $h_n$.

Theory. For a more adequate analysis, the simulated utilities are obtained considering the same values of the attributes of the previous situations, but varying only the decision rule.

For the estimation of the proposed model, it was assumed that the nesting scales of the present and future utilities were equal to each other, and that the maximum deliberation time was 10 steps. Then, the process was developed in two stages. First, the global optimiser DEoptim was used to obtain the value of the memory coefficient, to prevent the model from converging to a local optimum. However, this package does not deliver the standard errors of the estimators. Therefore, a second estimation was performed assuming a known (and fixed) value of the memory factor. Thus, in this second phase, non-convexity can be eliminated, the estimation complexity can be reduced and the standard errors of the estimators can be found. The Apollo package (Hess and Palma, 2019) was used to obtain the results of a Multinomial Logit Model (RUM) and DFT (Hancock et al., 2018) in which the attention weights are estimated on the same database.

From Table 4, several comments can be made. On the one hand, when the data generation process is RUM, the estimated Multinomial Logit model obtains the best fit and also efficiently recovers the true values of the data generation process, which is to be expected. Then, the DFT model follows, and although only the travel time attention weight and the preference update step are statistically significantly different from zero, it still helps to represent the variability of the data. Moreover, as discussed in Hancock et al. (2021), we find that lacking knowledge of the information search process, the Basic-DFT model has difficulty in effectively capturing the parameters that define that process. Finally, the proposed RUM-DFT model presents a slight difference in the information criteria but only gains 7 points in log-likelihood compared to the Basic-DFT model. However, it has a larger number of statistically significant parameters with expected signs, which allows it to adequately represent the information search process. It should be noted that a relevant question arises regarding the trade-off between the high economic interpretability that can be obtained from the RUM-DFT model compared to the slight performance gain of the Basic-DFT model when one wishes to explicitly include the deliberation process and subsequent estimation in discrete choice models.

On the other hand, when the data generation process follows a DFT approach, as shown in Table 4, it becomes evident that the Basic-DFT model exhibits the best fit. Notably, while the estimated Basic-DFT recovers the parameters of the deliberation process used to generate preferences, it struggles to recover the cost attention weight. In contrast, the RUM-DFT model shows only a slight deviation in performance from the DFT approach. Despite this, the proposed model can estimate several parameters, including the time attention weight ($\phi_{cost}$), memory ($\alpha$), cost coefficient ($\beta_{cost}$) and initial car utility ($ASC_B$), all of which are statistically different from zero. Finally, the RUM model demonstrates the poorest fit overall. Although it provides significant results for the travel time and cost coefficients, these estimates deviate significantly from the true values.

In conclusion, the analysis of the simulated database has demonstrated that the RUM-DFT model effectively estimates the parameters related to decision-making deliberation and attributes, given a certain level of understanding of the information search process. The difficulty of the proposed model lies in its estimation complexity which induces a high computational time.

### 3.4. Application to Swissmetro Real Stated Choice (SC) case study

This section compares the RUM-DFT model with five alternative models, using the Stated Choice (SC) case study SwissMetro as a common testbed. The SwissMetro project, proposed in the 1990s but not yet implemented, featured magnetically levitated trains operating within a vacuum tube. The case study data, collected by Bierlaire et al. (2001), considers stated choice (SC) experiments in which each respondent faced nine choice situations involving three alternatives (Train, SwissMetro, and Car) described by two

**Table 4**
Estimation results under RUM and DFT data generation processes.

| DGP | RUM | | | DFT | | |
|---|---|---|---|---|---|---|
| Parameters | RUM-DFT-SC | Basic-DFT | RUM | RUM-DFT-SC | Basic-DFT | RUM |
| Travel time $\beta_c$ | −1.110 | | −0.495* | −0.072* | | −0.272* |
| Travel cost $\beta_{tt}$ | −5.345* | | 1.890* | <0.001 | | −0.573* |
| Memory $\alpha$ | 0.853[a] | | | 0.670* | | |
| Attention weight travel cost $\phi_c$ | −2.037* | | | −1.713* | | |
| Attention weight travel cost $\gamma_c$ | | | | | −0.718 | |
| Attention weight travel time $\gamma_{tt}$ | | 1.263* | | | | |
| Sensitivity $\phi_{sen}$ | | 12.221 | | | 6.706* | |
| Memory $\phi_{mem}$ | | −0.002 | | | 0.044* | |
| Error $\sigma_\epsilon$ | | 6.172 | | | 1.017* | |
| Preferences update steps $\tau$ | | 4.153* | | | 3.182* | |
| Initial utility/Preference Bike $U_A^0$ | −2.158* | −18.510 | −0.579* | −0.286 | −2.883 | −0.047 |
| Initial utility/Preference Bus $U_B^0$ | −0.461 | −0.977 | −0.068 | −0.906* | −4.440 | −0.082 |
| Initial utility/Preference Car $U_D^0$ | 1.147* | 13.690 | 0.379* | −0.079 | 5.259 | 0.031 |
| LL | −3060.886 | −3053.110 | −3037.350 | −4497.16 | −4490.65 | −4501.61 |
| $\bar{\rho}^2$ | 0.447 | 0.448 | 0.451 | 0.188 | 0.189 | 0.187 |
| AIC | 6135.77 | 6122.21 | 6084.69 | 9008.314 | 8997.300 | 9013.220 |
| BIC | 6179.83 | 6172.57 | 6116.16 | 9052.372 | 9047.652 | 9044.690 |

[a] $\alpha$ estimated first with DEoptim. Final estimation fixing $\alpha$ and scales. RUM-DFT estimated with $T_{max}$=5 and equiprobable $h_n$.
* $\equiv p < 0.05$.

**Table 5**
Model estimates applied to real SwissMetro data.

| Parameters | RUM-DFT-SC ($T_{max}$=7) | RUM-DFT-SC ($T_{max}$=10) | RUM | Basic-DFT | $\mu$ RRM | P RRM |
|---|---|---|---|---|---|---|
| Travel time $\beta_c$ | −3.252* | −3.797* | −1.272* | | −0.887* | −0.907* |
| Travel cost $\beta_{tt}$ | −4.107* | −4.821* | −1.153* | | −0.613* | −0.683* |
| Memory $\alpha$ | 0.953[a] | 0.960[2] | | | | |
| Attention weight travel cost $\phi_{tt}$ | 0.483* | 0.486* | | | | |
| Tolerance $\delta$ | 0.257* | 0.258* | | | | |
| Attention weight travel cost $\gamma_c$ | | | | −0.87* | | |
| Attention weight travel cost $\gamma_{tt}$ | | | | −1.251* | | |
| Sensitivity $\phi_{sen}$ | | | | 0.000 | | |
| Memory $\phi_{mem}$ | | | | 0.000 | | |
| Error $\sigma_\epsilon$ | | | | 1.000 | | |
| Preferences update steps $\tau$ | | | | 0.564* | | |
| Initial utility/Preference train $U_A^0$ | −0.914* | −0.898* | −1.168 | −1.120* | −1.151* | −1.213* |
| Initial utility/Preference car $U_C^0$ | −0.008 | −0.008 | −0.250 | −0.076* | −0.268* | −0.424* |
| $\mu$ | | | | | 1.256* | |
| LL | −4,234.849 | −4,234.439 | −4,382.500 | −4,277.170 | −4,373.910 | −4,439.340 |
| $\bar{\rho}^2$ | 0.312 | 0.312 | 0.288 | 0.305 | 0.289 | 0.279 |
| AIC | 8,479.882 | 8,479.820 | 8,772.712 | 8,564.340 | 8,757.820 | 8,886.670 |
| BIC | 8,519.673 | 8,519.611 | 8,799.626 | 8,597.499 | 8,790.980 | 8,913.200 |

[a] $\alpha$ estimated first with DEoptim. Final estimation fixing $\alpha$ and scales.
* $\equiv p < 0.05$.

attributes (Travel time and Travel cost). The subset of data used in this article includes 5607 observations, focusing on individuals whose travel purpose was either studying or working, and who owned a car.

The RUM-DFT model is compared to the RUM, the DFT, and two variations of the Random Regret Minimisation (RRM) model: $\mu$-RRM, and P-RRM (van Cranenburgh et al., 2015). The DFT model estimates the attention weights, with the memory parameter ($\phi_{mem}$) set to 0, the sensitivity ($\phi_{sen}$) to 0, and the error term ($\sigma_\epsilon$) to 1. These settings seem reasonable as those values are typically insignificant when there is no prior knowledge of the information search process (Hancock et al., 2018). The RUM approach corresponds to a Multinomial Logit model and the RRM models are included as they represent an implicit breadth-first closed-form model, potentially competing with the proposed RUM-DFT.

For the SC data, in which there is no information on $h_n$, the estimation process for the RUM-DFT model was conducted in two stages. First, all parameters were obtained using the DEoptim optimiser, which finds the global optimum but does not provide the variance–covariance matrix. In the second stage, the memory parameter was fixed using the first stage estimates, reducing the process's complexity and allowing the standard deviations to be computed for statistical analysis,using a local optimiser. Additionally, the procedure was run twice, with the maximum number of deliberation steps set to seven and ten timesteps, to evaluate the trade-off between computational cost, estimated coefficients, and model fit performance.

The results presented in Table 5 indicate that models incorporating deliberation process parameters — namely, the RUM-DFT and DFT approaches — demonstrate superior performance based on log-likelihood, AIC, and BIC criteria. Among these, the RUM-DFT model provides a better fit and successfully recovers the estimators for coefficients associated with both attributes (time and cost) and the deliberation process (memory and attention weights). Furthermore, these coefficients are found to be statistically significant.

The RUM-DFT model has the best fit by far. In this model, the coefficients $\beta_{tt}$ and $\beta_c$ representing the travel time and cost contribution the utility of the alternatives are significant and have the expected negative sign. The time wight ($\phi_{time}$), of the probability of attending to an attribute in the next step, is significant and has a positive value. This means that decision-makers are more likely to attend to the time-related attribute than to cost in their information search process. The memory factor ($\alpha$) obtained with the global optimiser is statistically different from 0 and within the expected range. Finally, the maximum deliberation time considered in both DFT cases does not significantly modify the values of the coefficients or the log-likelihood. However, they differ considerably in the estimation time to find the value of the memory factor since $K^t$ sequences of extra attended attributes are considered for each $t$ to estimate the coefficients. Therefore, the limitation of this model is the high computational cost involved in its estimation, which strongly depends on the maximum deliberation time previously fixed.

Concerning its counterpart, the basic-DFT model stands out, which obtains significant estimates for the attention weights ($\beta_t$ and $\beta_c$) and the number of preference updating steps ($\tau$), but with a worse model fit. This approach is competitive against RUM-DFT since the estimation process takes less time and allows for the modelling of the deliberation process. However, interpreting basic-DFT's parameters is still unclear if modellers want to make an economic analysis. Then, the $\mu$ RRM model follows in fit; these good fit indicators may be because it is implicitly considered that decision-makers perform an information search in breadth-first and anticipate values to reduce their regret. Finally, the RUM model obtains significant estimates with the expected sign for the $\beta$-coefficients. However, it has a worse fit than the other two models previously discussed.

## 4. Conclusions

This paper introduces RUM-DFT, a discrete choice model that integrates the random utility maximisation (RUM) with decision field theory (DFT). By incorporating realistic information search dynamics from DFT, RUM-DFT addresses RUM's assumption of instantaneous, complete search. Evaluated via Monte Carlo simulations and real data, the model recovers parameters, replicates dynamic behaviours —including attended attribute sequences, deliberation time, and choice—, and retains microeconomic interpretability across scenarios.

The RUM-DFT model shows promising results in the first simulated case study (IS), where the full sequence of attended attributes for all decision-makers is known. In this version, the parameters are recovered and the dynamic utilities behave as expected. Notably, in the RUM-DFT-IS model, the economic interpretation of the parameters associated with attributes corresponds to the marginal utilities as in RUM static models. The use of process tracking surveys, such as the one constructed by Nova and Guevara (2022), is functional for capturing the specific sequences of respondents to be able to estimate the proposed RUM-DFT-IS model.

Concerning the RUM-DFT under DT and SC cases, the proposed model recovers the true values for attribute and deliberation process parameters in most cases, despite having less knowledge about the search process compared to the first case. However, the initially preconceived utilities are less accurately estimated, likely because of the assumptions of equiprobability of the choice paths and fixed scales. Additionally, we showed that these assumptions not only allowed for capturing the data generation process but also helps to reduce noise in estimating initial preconceived utilities.

When the data were generated using RUM or DFT processes, the RUM-DFT model achieved goodness-of-fit metrics comparable to those of the original (data-generating) models. Basic-DFT excels when matching its generation process but lacks generalizability, while RUM-DFT provides robust parameter estimation across paradigms at the cost of modest fit compromises.

Finally, the RUM-DFT model was estimated on the SwissMetro database. The signs, magnitudes and significance of the parameters and goodness-of-fit indicators were compared with the RUM, DFT, $\mu$-RRM and $P$-RRM models. In general, the models that incorporate the assumptions supported in this work (such as that the information search is breadth-first, that respondents attend to the attributes more than once and focus on specific attributes, and that they filter the information) obtain the best-fit indicators. Moreover, the RUM-DFT model specification, which does not include information on the deliberation process, can still significantly estimate the attribute coefficients (time and cost) as well as those associated with the deliberation process (tolerance, attention weights and memory) with reasonable values, successfully outperforming the DFT and RUM approaches. Notably, the proposed approach can capture both types of parameters, unlike the rest of the conventional models.

Several directions for future research emerge from this study. First, there is a need to develop methods that reduce the computational burden associated with accounting for all possible search paths and reliance on simplifying assumptions (e.g., equiprobability). Potential solutions include Maximum Simulated Likelihood with Random Sampling (Guevara et al., 2018) or Bayesian approaches. Second, future work could incorporate the time spent attending to each attribute—measurable via eye-tracking—into the model, as this likely influences search path probabilities. Third, further investigation is needed into how different attribute-attention sequences affect marginal utilities and willingness-to-pay measures. Finally, a systematic assessment of the bias in value-of-time estimates—resulting from models that ignore the deliberation process—would provide valuable insights.

## CRediT authorship contribution statement

**Gabriel Nova:** Writing – original draft, Software, Investigation, Formal analysis, Data curation, Conceptualization. **C. Angelo Guevara:** Writing – original draft, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Stephane Hess:** Writing – review & editing, Validation, Methodology. **Thomas O. Hancock:** Writing – review & editing, Validation, Methodology.

## Acknowledgements

This research was partially funded by ANID FONDECYT 1191104 and 1231584 and ANID PIA/PUENTE AFB230002. Stephane Hess acknowledges the support of the European Research Council through the advanced grant 101020940-SYNERGY.

## Data availability

Data and codes were shared in the Mendeley repository (Nova et al., 2025).

## References

Agrawal, S., Peeta, S., 2021. Hybrid route choice model incorporating latent cognitive effects of real-time travel information using physiological data. Transp. Res. Part F: Traffic Psychol. Behav. 81, 223–239.

Ariely, D., Zakay, D., 2001. A timely account of the role of duration in decision making. Acta Psychol. 108 (2), 187–207.

Ayaz, S.B., Tian, H., Gao, S., Fisher, D.L., 2023. Proactive route choice with real-time information: Learning and effects of network complexity and cognitive load. Transp. Res. Part C: Emerg. Technol. 149, 104035.

Ben-Akiva, M.E., Lerman, S.R., 1985. Discrete Choice Analysis: Theory and Application to Travel Demand, vol. 9, MIT Press.

Bierlaire, M., Axhausen, K., Abay, G., 2001. The acceptance of modal innovation: The case of Swissmetro. In: Swiss Transport Research Conference. (CONF).

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

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

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

Chorus, C., Bierlaire, M., 2013. An empirical comparison of travel choice models that capture preferences for compromise alternatives. Transportation 40 (3), 549–562.

Diederich, A., 1997. Dynamic stochastic models for decision making under time constraints. J. Math. Psych. 41 (3), 260–274.

Gabaix, X., Laibson, D., Moloche, G., Weinberg, S., 2006. Costly information acquisition: Experimental analysis of a boundedly rational model. Am. Econ. Rev. 96 (4), 1043–1068.

Gao, S., Frejinger, E., Ben-Akiva, M., 2011. Cognitive cost in route choice with real-time information: An exploratory analysis. Procedia- Soc. Behav. Sci. 17, 136–149.

Glöckner, A., Herbold, A.-K., 2011. An eye-tracking study on information processing in risky decisions: Evidence for compensatory strategies based on automatic processes. J. Behav. Decis. Mak. 24 (1), 71–98.

Guevara, C.A., 2015. Critical assessment of five methods to correct for endogeneity in discrete-choice models. Transp. Res. Part A: Policy Pr. 82, 240–254.

Guevara, C.A., 2024. Endogeneity in discrete choice models. In: Handbook of Choice Modelling. Edward Elgar Publishing, pp. 668–692.

Guevara, C.A., Tang, Y., Gao, S., 2018. The initial condition problem with complete history dependency in learning models for travel choices. Transp. Res. Part B: Methodol. 117, 850–861.

Hancock, T.O., Hess, S., Choudhury, C.F., 2018. Decision field theory: Improvements to current methodology and comparisons with standard choice modelling techniques. Transp. Res. Part B: Methodol. 107, 18–40.

Hancock, T.O., Hess, S., Choudhury, C.F., 2022. Secret in their eyes': Incorporating eye-tracking and stress indicator data into travel behaviour models. working paper University of Leeds.

Hancock, T.O., Hess, S., Marley, A.A., Choudhury, C.F., 2021. An accumulation of preference: two alternative dynamic models for understanding transport choices. Transp. Res. Part B: Methodol. 149, 250–282.

Hess, S., Daly, A., 2024. Handbook of Choice Modelling. Edward Elgar Publishing.

Hess, S., Palma, D., 2019. Apollo: A flexible, powerful and customisable freeware package for choice model estimation and application. J. Choice Model. 32, 100170.

Jacoby, J., 1975. Perspectives on a consumer information processing research program. Commun. Res. 2 (3), 203–215.

Jara-Díaz, S.R., Guevara, C.A., 2003. Behind the subjective value of travel time savings. J. Transp. Econ. Policy (JTEP) 37 (1), 29–46.

Johnson, E.J., Schulte-Mecklenbeck, M., Willemsen, M.C., 2008. Process Models Deserve Process Data: Comment on Brandstätter, Gigerenzer, and Hertwig (2006). American Psychological Association.

Kahneman, D., 2011. Thinking, fast and slow. Farrar, Straus Giroux.

Katsikopoulos, K.V., Duse-Anthony, Y., Fisher, D.L., Duffy, S.A., 2000. The framing of drivers' route choices when travel time information is provided under varying degrees of cognitive load. Hum. Factors 42 (3), 470–481.

McFadden, D., 1978. Modeling the choice of residential location. Transp. Res. Rec. 672, 72–77.

Miletić, S., van Maanen, L., 2019. Caution in decision-making under time pressure is mediated by timing ability. Cogn. Psychol. 110, 16–29.

Mullen, K.M., Ardia, D., Gil, D.L., Windover, D., Cline, J., 2011. DEoptim: An R package for global optimization by differential evolution. J. Stat. Softw. 40, 1–26.

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

Nova, G., Guevara, C.A., 2022. In depth, breadth-first or both? Characterising the information search process in a public transport SP experiment. Working paper University of Chile.

Nova, G., Guevara, C.A., Hancock, T.O., Hess, S., 2025. Data random utility maximisation model considering the information search process. http://dx.doi.org/10.17632/m4htc543py.1, Mendeley Data, V1.

Payne, J.W., 1976. Task complexity and contingent processing in decision making: An information search and protocol analysis. Organ. Behav. Hum. Perform. 16 (2), 366–387.

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

Rust, J., 1987. Optimal replacement of GMC bus engines: An empirical model of harold zurcher. Econ.: J. Econ. Soc. 999–1033.

Ruud, P.A., 1983. Sufficient conditions for the consistency of maximum likelihood estimation despite misspecification of distribution in multinomial discrete choice models. Econ.: J. Econ. Soc. 225–228.

Schwartz, B., Ward, A., Monterosso, J., Lyubomirsky, S., White, K., Lehman, D.R., 2002. Maximizing versus satisficing: happiness is a matter of choice. J. Pers. Soc. Psychol. 83 (5), 1178.

Small, K.A., Rosen, H.S., 1981. Applied welfare economics with discrete choice models. Econ.: J. Econ. Soc. 105–130.

Stewart, N., Gächter, S., Noguchi, T., Mullett, T.L., 2016a. Eye movements in strategic choice. J. Behav. Decis. Mak. 29 (2–3), 137–156.

Stewart, N., Hermens, F., Matthews, W.J., 2016b. Eye movements in risky choice. J. Behav. Decis. Mak. 29 (2–3), 116–136.
Sui, X.-Y., Liu, H.-Z., Rao, L.-L., 2020. The timing of gaze-contingent decision prompts influences risky choice. Cognition 195, 104077.
Train, K.E., 2009. Discrete Choice Methods with Simulation. Cambridge University Press.
van Cranenburgh, S., Guevara, C.A., Chorus, C.G., 2015. New insights on random regret minimization models. Transp. Res. Part A: Policy Pr. 74, 91–109.
Voskuilen, C., Ratcliff, R., Smith, P.L., 2016. Comparing fixed and collapsing boundary versions of the diffusion model. J. Math. Psych. 73, 59–79.
Xie, Y., 2019. Choice Modeling with Observed and Unobserved Information Search (Ph.D. thesis). Massachusetts Institute of Technology.