Quantitative Marketing and Economics
https://doi.org/10.1007/s11129-025-09292-9

Discrete choice in marketing through the lens of rational inattention

Sergey Turlo[1] · Matteo Fina[1] · Johannes Kasinger[2,3] · Arash Laghaie[4] · Thomas Otter[1,4]

Received: 10 August 2023 / Accepted: 24 December 2024

Abstract
Models derived from random utility theory represent the workhorse methods to learn about consumer preferences from discrete choice data. However, a large body of literature documents various behavioral patterns that cannot be captured by basic random utility models and require different non-unified adjustments to accommodate these patterns. In this article, we discuss strategies how to apply rational inattention theory—which explains a large variety of such departures—to the analysis of discrete choice among multiple alternatives described along multiple attributes. We first review existing applications that make restrictive belief assumptions to obtain choice probabilities in closed multinomial logit form. We then propose a model that allows for general consumer beliefs and demonstrate its empirical identification. Further, we illustrate how this model naturally motivates stylized empirical results that are hard to reconcile from a random utility perspective.

Keywords Choice modeling · Rational inattention · Conjoint analysis · Discrete choice experiments

JEL Classification C00 · C35 · D83 · M31

✉ Sergey Turlo
turlo@marketing.uni-frankfurt.de

Matteo Fina
matteo.fina@its.uni-frankfurt.de

Johannes Kasinger
j.kasinger@tilburguniversity.edu

Arash Laghaie
arash.laghaie@novasbe.pt

Thomas Otter
otter@marketing.uni-frankfurt.de

1 Johann Wolfgang Goethe University Frankfurt, Frankfurt, Germany

2 Tilburg School of Economics and Management, Tilburg University, Tiburg, Netherlands

3 Leibniz Institute for Financial Research (SAFE), Frankfurt, Germany

4 Nova School of Business and Economics, Carcavelos, Portugal

1 Introduction

Discrete choice models based on rational inattention (RI) theory are becoming popular in economics and marketing for studying consumer choices and preferences when information processing is limited but adaptive. This article demonstrates how to apply these models in various multi-attribute, multi-alternative (MAMA) contexts. Unlike the traditional random utility model (RUM), RI-based choice models provide a unified framework to explain phenomena like consideration sets, stake sensitivity, and brand-specific price responses, which previously required ad-hoc adjustments to the RUM. This article aims to support applied researchers by: first, raising awareness of the potential of discrete choice models under RI; second, outlining the steps for implementing RI for empirical research; and third, clarifying the strengths and limitations of different implementations.

Based on foundational ideas from psychology (e.g., Simon & Newell, 1971), rational inattention theory, introduced by Sims (2003) in macroeconomics, suggests that decision-makers (DMs) face cognitive limitations and, therefore, do not take in all available information when making choices. DMs recognize their cognitive limitations and strategically decide how much and what type of costly information to process in each decision scenario. RI suggests that DMs adjust their processing efforts based on prominent, accessible aspects of a choice task, which shape their prior beliefs about unknown factors affecting utility. The adaptive and partial information processing implied by RI motivates a rich set of behaviors, even with standard additively separable utility.

Under RI, probabilistic choice follows from costly and thus imperfect processing of information, i.e., from DMs' residual uncertainty about what the utility maximizing choice alternative is. In contrast, the RUM by McFadden (1974) derives choice probabilities by assuming that the DM acts on a larger information set than observed by the analyst. Unlike RUMs, RI choice models can explain choices from MAMA sets without assuming that only the decision-maker observes certain utility factors. This aligns with earlier research that explained randomness in choice through cognitive processes (e.g., Thurstone, 1927; Quandt, 1956; Louviere et al., 1999). RI adds to this literature by offering a micro-foundation for probabilistic choice based on economic optimization.

The general discrete choice problem under RI lacks a closed form solution. To utilize standard estimation methods, current empirical RI discrete choice models (RI-DCMs) make particular—and potentially unrealistic—assumptions about consumers' prior beliefs, resulting in choice probabilities that follow a closed-form multinomial logit (MNL) function of the underlying utility index. While these models can motivate some deviations from the full information RUM, they face similar conceptual issues as logit models, such as unrealistic substitution patters and exogenous consideration sets.

To address these limitations, we demonstrate how to estimate a RI-DCM under general prior belief assumptions, including rational expectations, and consumer heterogeneity. In this model, alternatives' payoffs are represented by linear utility indices based on preferences and attributes, which may be *simple* or *complex*. DMs can process simple attributes at no cost, whereas processing or integrating utility from complex attributes requires costly effort. DMs have prior beliefs about the complex attributes, assuming rational expectations. In choice experiments, these expectations are shaped by the experimental design. We show that both preference parameters and the distinction between simple and complex attributes can be likelihood identified in this model. The primary advantages of this model over existing empirical RI-DCMs are that i) it can explain a broader range of phenomena that deviate from the RUM-DCM framework and ii) it enables more flexible counterfactual analysis. The drawback of this approach is that it requires a numerical solution to the formal RI problem. Table 1 provides a comparison of existing methods for applying RI to MAMA data.

Using the RI-DCM with general beliefs, we demonstrate how different phenomena contradicting the microeconomic foundation of RUMs in MAMA settings endogenously follow from the optimal deployment of limited cognitive resources. Examples include brand-specific price coefficients (e.g., Carmone & Green, 1981; Sawtooth Software, 1996; Kalra & Goodstein, 1998), separate coefficients for different aspects of price such as, e.g., a coefficient for regular price and one for a price discount or a tax (Guadagni & Little, 1983; Blattberg & Neslin, 1989; Chetty et al., 2009), as well as consideration sets and attribute non-attendance.

Table 1 Comparison of different empirical RI-DCMs

Strategy	Paper	Prior beliefs	Implied choice probabilities	Comments
RI-DCM with choice in closed MNL form	Brown and Jeon (2024)	Beliefs over index of unknown attributes follows Cardell distribution	Equivalent to multinomial logit (additive separability over alternative characteristics, fully compensatory)	Assumed belief distribution has full support on the real line which may be unreasonable, e.g., with prices; does not reproduce certain RI features like consideration sets
	Joo (2023)	Beliefs over all utility components are a function of non-utility components, implicitly defined such that resulting choice probabilities are equivalent to multinomial logit in closed form	Equivalent to multinomial logit (additive separability over alternative characteristics, fully compensatory)	Approach motivates the inclusion of non-utility attributes in the logit index; implied prior distribution is not accessible by the analyst; counterfactuals with respect to beliefs are available only in a restricted fashion; does not reproduce certain RI features like consideration sets
RI-DCM with general beliefs	This paper	Any prior belief distribution over a discrete state space, e.g., rational expectations over choice tasks in a DCE	non-compensatory, not additively separable	Resulting model reproduces all qualitative features of discrete choice under RI, e.g., consideration sets or attribute interactions

In the RI-DCM with general beliefs, predictions of, e.g., consideration sets and attribute non-attendance become implicit functions of the composition of a choice set, reflecting adaptations to prominent features of the set. This flexibility is absent in current empirical RI-DCMs which invoke very specific assumptions about consumer beliefs (e.g., Joo, 2023; Brown & Jeon, 2024). Similarly, this RI-DCM predicts effects of attribute range, number of attribute levels, and the size of choice sets on choice behavior. RUMs typically do not account for these common properties of choice data, at least not comprehensively in one unified model.

Different from extant models of consumer search and learning,[1] RI does not restrict the structure of informative signals that the DM uses. This feature makes RI distinct and more generally applicable than models where all uncertainty is resolved upon search. For example, in the context of discrete choice experiments (DCEs) all relevant pieces of information, i.e., attribute information for all alternatives, are readily presented to DMs. Thus, the distinction between more and less processing of the available information is qualitatively different from the distinction between knowing or not knowing certain product attribute values as typical of search models. This idea resembles the distinction between evaluation costs and search costs in Guo (2021) and Gu and Wang (2022). Consequently, RI-DCMs are particularly important and useful departures from extant models when evaluating and integrating information for an updated overall understanding of a choice situation is decisive and effortful. In contrast, search models are arguably more adequate when knowing or not knowing a particular attribute makes the difference.

Finally, search models that allow for partial learning about the value of available alternatives are more closely related to RI (e.g., Ursu et al., 2020). However, sequential search/learning models applied to MAMA choice likely will be computationally intractable without observing the search/learning sequence. Alas, this sequence that involves mental operations beyond reading attribute information may well be fundamentally unobservable. In addition, the standard assumption of normally distributed prior beliefs and signals in these models does not correspond with the empirical distribution of attributes and their bounded support, especially in DCEs.

In this article, we provide an accessible presentation of the RI theory for studying discrete choice in MAMA settings common in economics and marketing. Our contributions are threefold. First, we outline two general strategies for applying RI to discrete choice MAMA data and discuss their respective advantages and disadvantages. The first strategy, typical for existing RI-DCMs, invokes specific assumptions about consumer beliefs (continuous with full support following a Cardell distribution) that yield choice probabilities in closed MNL form. The full support assumption contradicts observable distributions of attributes. The second strategy can accommodate general consumer beliefs and requires a numerical solver. A researcher choosing between these strategies faces a trade-off between conceptual realism and computational burden. Our second contribution is that this article is, to our knowledge, the first to detail the necessary steps for implementing the second empirical strategy ("RI-DCM with general prior beliefs") while accounting for consumer heterogeneity. Third, we demonstrate that a RI-DCM with general beliefs can effectively reproduce various behavioral

[1] For a recent overview, see Honka et al. (2019).

patterns inconsistent with standard RUM models in MAMA settings parsimoniously. In contrast, existing RI-DCMs that yield choice probabilities in closed MNL form imply the same constraints as MNL derived from RU theory, such as independence of irrelevant alternatives and strictly positive choice probabilities for all alternatives in a set.[2]

The remainder of this paper is organized as follows. Section 2 derives discrete choice among multiple alternatives described along multiple attributes under the RI framework. Section 3 illustrates estimation and empirical identification of the model. Section 4 discusses key features of the RI model and provides illustrative simulations. Section 5 concludes with a discussion and an agenda for future research.

2 A rational inattention model of discrete choice

The basic idea behind RI theory is that DMs face an abundant amount of information and cannot process all of it. However, they are aware of this limitation and decide how to process the available information optimally, trading off costs and benefits of being better informed. This idea was suggested by Sims (2003) to provide a unifying framework for different frictions in macroeconomics. While the original model was developed for continuous action spaces, Matějka and McKay (2015) extend this theory to discrete choices.

Our presentation builds on the discrete choice version of the RI model (Matějka & McKay, 2015; Caplin et al., 2019) and tailors it to the typical MAMA setting. Introducing the model, we first present the RI choice problem and discuss how its various components translate into the MAMA setting. Then, we turn to the problem's solution and cover how the various primitives affect the resulting choice behavior. In particular, this will illustrate the impact of the complexity of a choice task and of the incentives to process information.

To ease the exposition of the various components of the RI framework, we will refer as an example to a DCE where a DM has to choose between a car and an outside option. In this example, the final price paid by the DM consists of two components: i) a list price that is easily evaluated by the DM, and ii) a discount that applies only to specific cars (thus encouraging the purchase of such vehicles). While both components have the same impact on final utility, we assume it is more effortful to find out if and what discount applies to a particular car.

This simple example may align well with existing traditional search models if determining the eligibility of a specific car is a simple search task, e.g., checking whether the discount in monetary terms applies to a specific car model. However, we posit a scenario wherein the eligibility of a specific car hinges upon (a combination of) diverse characteristics. In such a situation, judging the applicability of the discount requires the consumer to collect information from various sources and integrate it to

[2] Matějka and McKay (2015) already pointed out that a random utility model cannot generally capture behavior implied by RI agents.

determine the overall value of the discount. Consequently, it is possible that the DM processes only some parts of the information and, therefore, may arrive at a faulty evaluation of the final price, which in turn leads to choice errors.[3]

2.1 Formal problem and its translation into MAMA settings

We closely follow Caplin et al. (2019) in defining the problem faced by the rationally inattentive DM. There is a finite number of states Ω the DM can learn about.[4] An action a is a mapping from states to utilities. $\mathcal{A}$ denotes the set of all possible actions. The mapping $u : \mathcal{A} \times \Omega \rightarrow \mathbb{R}$ describes the utility from any action in each state. The problem faced by the DM is non-trivial because, typically, different actions are optimal in different states, and the DM is uncertain about the true state. However, as we will explain later, the DM can costly learn about the true state.

The general nature of RI theory provides room for different translations of the framework into the typical MAMA setting in marketing. We naturally impose that actions correspond to different alternatives from which the DM chooses and define states as representing different choice sets characterized by the specific attribute compositions of alternatives available to the DM. Accordingly, in this setting, Ω corresponds to the set of all attainable choice sets in a given choice environment.

Payoffs Similar to the distinction between directly observable attributes and attributes that need to be searched in search models (e.g., Honka et al., 2019; Gardete & Hunter, 2020) or the distinction between attributes that guide consideration and attributes that are only processed upon consideration in two-stage models of choice (e.g., Aribarg et al., 2018), we assume that the subjective value of alternatives is derived from attributes that fall into two categories. The first category consists of *simple* attributes $\mathbf{x}_s$ whose joint valuation is immediate to the DM. The second category comprises *complex* attributes $\mathbf{x}_c$ whose joint valuation and integration with simple attributes requires cognitive effort and time.[5]

We assume additive separability such that the subjective utility of an alternative is given by

$$u(a, \omega) = \mathbf{x}'_{a,s}(\omega)\boldsymbol{\beta}_s + \mathbf{x}'_{a,c}(\omega)\boldsymbol{\beta}_c, \tag{1}$$

[3] There are many more things about a car that are likely payoff relevant to a DM, and it may or may not be effortful to evaluate and integrate them into an overall evaluation. However, this minimal example will help develop basic principles.

[4] An alternative formulation with a continuous state space is given in Matějka and McKay (2015).

[5] Studies that explore the choice process using eye traces have documented an “orientation phase” where the DM acquires partial information about the products, which guide her subsequent information acquisition (e.g., Russo & Leclerc, 1994; Musalem et al., 2021). This orientation phase and subsequent behavior involves both bottom-up and top-down processing (see Corbetta & Shulman, 2002 for a review).

where $\boldsymbol{\beta}_s$ and $\boldsymbol{\beta}_c$ are the respective part-worths of simple and complex attributes.[6] The dependence of $\mathbf{x}_{a,s}$ and $\mathbf{x}_{a,c}$ on ω above highlights that attributes of alternatives change from choice set to choice set.[7]

Before learning, the DM has some beliefs about the value of complex attributes $\mathbf{x}_{a,c}$, which become more precise as the DM processes information. Note that any uncertainty is due to complex attributes. We refer to the portion of utility derived from simple attributes as the "simple utility component" while the portion derived from complex attributes is termed the "complex utility component".

In our example, the list price of a car is a simple attribute, and the discount is a complex attribute that requires time and cognitive effort to process and integrate with the simple list price to arrive at a final price and an assessment of utility. To further illustrate the challenges associated with integrating information, consider the following examples involving two price components: (i) prices are given in currency units and discount values in percentages given as numbers, e.g., 9.90 Euro and a 16% discount, (ii) discount values in currency units, e.g., 9.90 Euro and a 1.58 Euro discount, (iii) discount values mentioned in some way together with the final price of, in this example, 8.32 Euro. Cases (i) and (ii) require the DM to integrate the discount information with the price. However, this exercise arguably is more involved in case (i) than in case (ii) because the DM has to calculate the discount value as part of the integration exercise. The main point, however, is that case (ii)—where both the price and the discount are presented in currency units—is harder than case (iii), where the final price is displayed. This added difficulty stems from the need to integrate two pieces of information by subtracting the discount from the price to determine the overall value.

Prior beliefs The DM's problem is given by a pair (μ, A). Here, $\mu \in \Delta(\Omega)$ is her prior belief over the states of the world, with $\Delta(\Omega)$ being the set of distributions over Ω, and $A \subset \mathcal{A}$ is the set of actions she can choose from. In our illustrative example, a state ω corresponds to a specific choice set characterized by a particular combination of attribute realizations. Since simple attributes are processed at no cost by the DM, each combination of simple attribute realizations $\mathbf{x}_s$ induces a different prior belief distribution $\mu_s \in \Delta(\Omega)$ over possible choice sets $\omega \in \Omega$. In general, these prior beliefs, conditional on costless information, will differ from the unconditional distribution over choice sets. In particular, the DM obtains prior beliefs μ_s by conditioning the distribution over all choice sets on the simple attribute realizations faced in a specific choice set $\widehat{\omega}$. Formally, prior beliefs are given by

$$\mu_s(\omega) = \Pr\left(\omega | \{(\mathbf{x}_{a,s}(\widehat{\omega}))\}_{a \in A}\right)$$

where the distribution over choice sets, $\Pr(\omega)$, is determined by the choice environment.

[6] Additive separability is by no means a necessary but often a natural assumption when, e.g., different price components add up to a total price.

[7] With the present notation, any state or choice set is defined by the configuration of the alternatives: $\omega = \{(\mathbf{x}_{a,s}(\omega), \mathbf{x}_{a,c}(\omega))\}_{a \in A}$.

Table 2 Set of choice sets with respective payoffs and prior beliefs

Choice set Probability of ω_i	ω_1 $\gamma_p(1-\gamma_d)$	ω_2 $\gamma_p\gamma_d$	ω_3 $(1-\gamma_p)(1-\gamma_d)$	ω_4 $(1-\gamma_p)\gamma_d$
Payoffs:				
Inside alternative u_I	$\beta_b - p_L$	$\beta_b - p_L + D$	$\beta_b - p_H$	$\beta_b - p_H + D$
Outside alternative u_O	0	0	0	0
Information set	$p = p_L$,	$p = p_L$,	$p = p_H$,	$p = p_H$,
	$d \in \{0, D\}$	$d \in \{0, D\}$	$d \in \{0, D\}$	$d \in \{0, D\}$
Prior beliefs μ_s:				
$\Pr(u_I = \beta_b - p_L)$	$1-\gamma_d$	$1-\gamma_d$	0	0
$\Pr(u_I = \beta_b - p_L + D)$	γ_d	γ_d	0	0
$\Pr(u_I = \beta_b - p_H)$	0	0	$1-\gamma_d$	$1-\gamma_d$
$\Pr(u_I = \beta_b - p_H + D)$	0	0	γ_d	γ_d

Each column represents a different choice set ω_i. In addition to the payoffs, the objective probability of each choice set, the information set of the DM before any learning takes place, as well as the resulting prior beliefs are displayed. Note that this fully characterizes the DM's prior since for all choice sets $\Pr(u_O = 0|\omega_i) = 1$

Intuitively, one can think of a sequentially updating DM who is aware of the choice environment (e.g., an experimental design) and the implied distribution of attributes over all possible choice sets. Once she observes the realized *simple* attribute values ($\mathbf{x}_s$) in a specific choice set, she forms conditional prior beliefs μ_s, which in turn determine how she processes complex attributes $\mathbf{x}_c$.

Returning to our exemplary DCE, suppose the DM chooses only between a single car brand and an outside option of not buying. The utility of the inside alternative, i.e., the car, is given by $u_I = \beta_b - p + d$ with β_b being the brand coefficient, p being a simple price, and d being a complex discount. For the sake of a minimal example, we assume that brand is a simple attribute as well and that whatever (complex) criteria qualify the car for the discount only contribute to utility through the discount, such that cars with and without the discount have the same brand coefficient β_b.[8] We further assume that the DM has to commit to purchasing the car at a price p, i.e., pay p and only after is reimbursed d, depending on the eligibility of the car. The utility of the outside option is normalized to $u_O = 0$. Further, the experimental design is such that $p \in \{p_L, p_H\}$ with $\Pr(p = p_L) = \gamma_p$ and $d \in \{0, D\}$, $D > 0$ with $\Pr(d = D) = \gamma_d > 0$.[9] In Table 2, the columns represent the possible choice sets of this design. The DM can face four different choice sets as there are two attributes with two levels each, so $|\Omega| = 4$.

The objective probabilities of each choice set ω (prior to the realization of the simple attributes) are displayed in Table 2. Since there are only two possible realizations of simple attributes, characterized by the two levels of the simple price p_L and p_H, there

8 In the notation of Eq. 1, we have the following decomposition into simple and complex contributions to utility: $[1, p]_s[\beta_b, -1]' + d_c \cdot 1$ with a shared price coefficient equal to 1.

9 In our example, the conditional distribution of the complex attribute equals the marginal distribution due to the orthogonal design. In designs with built-in correlations, e.g., with conditional pricing, the realized values of simple attributes will predict the distribution of complex attributes values.

are two different conditional prior beliefs μ_s by design. Each combination of simple attributes results in different prior beliefs. In our example, choice sets ω_1 and ω_2 give rise to the same prior beliefs, as do choice sets ω_3 and ω_4, although the beliefs differ between these two groups (see the last four rows in Table 2).

Information processing Under RI, the DM first chooses what and how much to learn. This process is formally modeled by allowing the DM to choose the distribution of an unrestricted stochastic signal. The signal (usually imperfectly) indicates the true "state of the world". The distribution of the stochastic signal results from the DM's information processing strategy. The DM controls the quality of the signal through her processing effort. More costly effort results in more precise signal distributions.

After observing the realized signal (drawn from the chosen distribution), the DM forms a posterior belief. Given this posterior belief, the DM chooses the alternative a that maximizes her payoff. Consequently, the DM's posterior belief about the true complex attribute(s) levels in a choice set translates into a belief about the optimal action a that then determines the DM's choice.

We follow the established RI literature in using an information cost function that is based on Shannon mutual information. Then, the optimal signal is such that different signal realizations deterministically map into different choice actions a, so that a processing strategy is equivalently characterized by (state) conditional signal distributions *or* (state) conditional choice probabilities $P(a|\omega)$. This probability thus corresponds to the probability that the DM's processing effort translates into a signal that points to choice a in the true state ω, which is then chosen deterministically. Consequently, the DM chooses (and subsequently acts upon) the conditional choice probabilities $P(a|\omega)$ that maximize expression Eq. 2.

$$\sum_{\omega\in\Omega}\mu_s(\omega)\left(\sum_{a\in A}P(a|\omega)u(a,\omega)\right)-\lambda\left[\sum_{\omega\in\Omega}\mu_s(\omega)\left(\sum_{a\in A}P(a|\omega)\ln P(a|\omega)\right)-\sum_{a\in A}P_s(a)\ln P_s(a)\right] \tag{2}$$

The first term in expression Eq. 2 captures the expected utility from choosing conditional choice probabilities $P(a|\omega)$. The sum $\sum_{a\in A}P(a|\omega)u(a,\omega)$ is the expected utility from a particular state ω given $P(a|\omega)$. The sum of all state-specific expected utilities ($\omega\in\Omega$) weighted by the prior probabilities, according to μ_s, yields the expected utility over all possible choice sets, given the realized values of simple attributes in the current choice set. The more weight $P(a|\omega)$ puts on the optimal actions in the respective states, i.e., the higher the quality of the signals, the higher the expected payoffs of the DM.

The second term of expression Eq. 2 formalizes the information processing costs. The term in brackets is known as mutual information that is multiplied by $\lambda > 0$, the unit costs of information. Mutual information measures the expected reduction in choice uncertainty due to information processing. The stronger $P(a|\omega)$ deviates from the *endogenous* unconditional choice probability $P_s(a)=\sum_{\omega\in\Omega}\mu_s(\omega)P(a|\omega)$, putting more weight on the optimal action in a choice set ω, the higher the costs of information processing. Mutual information is based on the concept of Shannon entropy, an uncertainty measure rooted in information theory.[10] It is the difference between the entropy of unconditional choice probabilities $P_s(a)$ (in a choice set characterized by specific realizations of simple attributes), and the expected entropy of conditional choice probabilities, $P(a|\omega)$. The expectation is again formed over the prior distribution of complex attributes, conditioned on realized simple attributes in a particular choice set (μ_s).

Unconditional choice probabilities $P_s(a)$ correspond to the probability of choosing alternative a from a choice set ω, *given* a processing strategy, but *before* actually processing the costly information about the *complex* attributes in a specific choice set in line with this strategy. Therefore, unconditional probabilities are not arbitrary, subjective functions of simple attributes. Instead, they are defined as prior *expectations* over how choice probabilities would change through costly processing of *possible* realizations of complex attributes and are thus constrained by the equality $P_s(a) = \sum_{\omega\in\Omega} \mu_s(\omega)P(a|\omega)$. Constraining the prior to be equal to the expected posterior beliefs makes the RI-DM coherent in a Bayesian sense and guarantees that processing complex information will be associated with positive costs (as mutual information cannot be negative when $P_s(a) = \sum_{\omega\in\Omega} \mu_s(\omega)P(a|\omega)$).

Turning back to the exemplary DCE, a DM will form, based on the simple brand and price that she processes at no cost, beliefs over possible payoffs. Depending on the observed simple attributes, she has different prior beliefs over possible payoffs (see, e.g., columns 2 and 3 in Table 2). This prior belief will affect how she processes the information about the discount, that is, how much time and effort she spends on reading product descriptions, checking discount criteria, or contemplating the economic consequences of her purchase. This, in turn, translates into an expectation of how likely she is to purchase the car ($P_s(a)$) taking into account that the discount may or may not apply as by her conditional prior.[11]

To clarify the difference between conditional and unconditional choice probabilities, $P(a|\omega)$ and $P_s(a)$, suppose it is optimal to choose the car when the discount is high and the outside option when it is low. If processing information is free ($\lambda = 0$), she would optimally perfectly identify the true state and choose the best state-contingent option. Accordingly, the conditional choice probabilities of buying the car will equal one if the discount applies and zero otherwise. Thus, the DM will always choose the utility-maximizing alternative and make no choice errors if $\lambda = 0$. In this case, the unconditional choice probability of buying the car will be equal to the conditional prior belief: $\Pr(d = D) = \gamma_d$.

[10] The entropy of a distribution q on Ω is formally given by $-\sum_{\omega\in\Omega} q(\omega)\ln q(\omega)$.

[11] In the RI framework, this expectation is the result of the aforementioned formal optimization, in reality this may be either the result of determining an optimal decision rule that is applied repeatedly to different choice tasks or it may be learned over time through experience (Maćkowiak et al., 2023). DCEs are an example of the former case because there is no feedback or learning from individual choices in such experiments.

The RI framework is agnostic about whether processing information is hard because it is cognitively difficult to determine the actual value of complex attributes or due to the integration of different attribute realizations into a single value through cognitive processing, e.g., calculations. Moreover, λ may vary across individuals, e.g., due to differences in cognitive ability or prior experience, and across qualitatively different choice environments. In the case of a complex discount, processing the eligibility requirements will be affected by the number of criteria that must be checked, the font size used to describe the discount, and the characteristics of the DM, e.g., prior experience or (intellectual) ability.

2.2 Solution, endogenous consideration sets, and stochastic choice

The solution to the maximization problem of the DM in expression Eq. 2 with respect to $P(a|\omega)$ gives us the state-contingent choice probabilities for all actions as a function of unconditional choice probabilities (Matějka & McKay, 2015):

$$P(a|\omega) = \frac{P_s(a)\exp\{u(a,\omega)/\lambda\}}{\sum_{b\in A} P_s(b)\exp\{u(b,\omega)/\lambda\}}. \tag{3}$$

By rewriting $P_s(a)\exp\{u(a,\omega)/\lambda\}$ as $\exp\{(u(a,\omega)+\lambda\ln P_s(a))/\lambda\}$, Eq. 3 reveals that state-contingent choice probabilities $P(a|\omega)$ are determined by a modified logit formula. In this formula, the payoffs from an action a in a state ω are adjusted by $\lambda \ln P_s(a)$ and divided by the information processing costs λ. The adjustment depends on the unconditional choice probability of a that is a function of prior beliefs conditional on simple attributes (μ_s) and the information processing costs. While choice probabilities depend on the payoffs in a particular state, similar to a RU logit model, they are shifted towards those actions that appear to be more attractive based on prior information. The extent of this shift depends on the magnitude of the processing costs λ. The lower these costs, the stronger the state-contingent choice probabilities will deviate from their unconditional counterparts towards the action that is optimal in a given state ω.

Equation 3 also points to the qualitatively different impact of simple and complex attributes on choice in this model, tailored to a MAMA setting. Because realizations of simple attributes affect both prior beliefs μ_s and choice specific payoffs $u(a,\omega)$, both the unconditional choice probability $P_s(a)$ and the term $\exp\{u(a,\omega)/\lambda\}$ are affected. In contrast, realized values of complex attributes, which require cognitive processing, will affect only the latter component. This reflects that realizations of simple attributes, together with the (conditional) distribution of complex attributes, determine the optimal processing strategy.

Returning to our sample DCE, in choice sets where the simple price is either very large or very small, the costs of making a choice error when choosing based on prior beliefs are small, and so the DM may choose not to process any information about the discount. However, when the difference between prior expected payoffs of available alternatives is small, so that the DM is a priori indifferent between the car and her outside option, the costs of a choice error are rather large. Consequently, obtaining a (more) precise signal over the true value of the discount becomes valuable to the DM.

Sufficient conditions and endogenous consideration sets Caplin et al. (2019) characterize the set of actions chosen with strictly positive probability—the DM's consideration set. They show that choice probabilities $P(a|\omega)$ that fulfill Eq. 3 are optimal in the sense of RI *if and only if*

$$\sum_{\omega \in \Omega} \frac{\mu_s(\omega) \exp\{u(a, \omega)/\lambda\}}{\sum_{b \in A} P_s(b) \exp\{u(b, \omega)/\lambda\}} \leq 1 \tag{4}$$

for all $a \in A$. This condition holds with equality when $P_s(a) > 0$. The set of alternatives with $P_s(a) > 0$ is interpreted as a consideration set.

The reason for the endogenous formation of consideration sets is as follows. Consider the DM's objective in expression Eq. 2, and recall that it is costly to choose state-contingent choice probabilities that deviate from unconditional choice probabilities. By setting $P_s(a') = 0$ for some a', such that in all states $P(a'|\omega) = 0$, the DM incurs lower processing costs since these actions' state-contingent choice probabilities always equal the unconditional choice probabilities.

Intuitively, consideration sets simplify the information processing (and thus save cognitive costs) by reducing the dimensionality of the DM's problem. In typical marketing settings, the endogenous consideration set will be a function of realized simple attributes in a choice set, the (conditional) distribution of complex attributes, and cognitive costs. In our example, the DM may find it optimal to choose on the basis of the simple price alone if the expected gains from processing information about the complex discount is not worth the associated costs of processing.

Endogenous stochastic choice As discussed above, the rationally inattentive DM will typically make choice errors because information processing is costly (and perfectly identifying the state is not optimal due to the convex information cost function). So, there are cases where the signal indicates $d = D$, such that the DM chooses to buy the car in our simplistic example, but where—in hindsight—it would have been optimal not to buy the car as the discount does not apply. Thus, such choice errors are the result of optimal but constrained behavior. They do not appeal to aspects of utility only observed by the DM but rather to the DM's imperfect observations or processing. Hence, the probabilistic nature of choice under RI simply reflects the DM's information processing, which is imperfect because of processing costs.

Note that the larger the information processing costs, the more (or more consequential) choice errors the DM will make, everything else equal. Also, the strategy implied by maximizing the expression in Eq. 2 is adaptive in the sense that the optimal amount of costly information processing, as provided by complex attributes, differs as a function of the values of simple attributes in a choice set. We provide numerical illustrations in Section 4.

3 Empirical identification strategies

In the RI framework, three primitives determine the likelihood of data given the RI model:

- preferences or payoffs,
- information processing costs, and
- prior beliefs.

In general, empirical applications fix two out of these three and learn from data about the remaining one. Therefore, an analyst interested in inferring preferences has to make assumptions about what a state is, what DMs may costly learn, and what they believe about those costly to learn aspects. Since the likelihood in the RI framework is homogeneous of degree zero with respect to payoffs and costs, only their ratio is directly identifiable from the choice data. This follows from the fact that the DM's objective function is homogenous of degree one.

Existing applications make strong, and arguably often unrealistic, assumptions about DMs' prior beliefs to obtain choice probabilities in closed MNL form. However, these approaches, which we present in detail next, a priori rule out many qualitatively distinguishing features of discrete choice under RI, such as endogenous consideration sets. For this reason, we lay out the estimation and identification of an RI-DCM capable of incorporating general prior beliefs—thereby preserving such features—in Section 3.2.

This is followed by an illustration of the empirical identification and Bayesian estimation of the RI-DCM with general beliefs in a hierarchical setting with heterogeneity (Section 3.3). We discuss how the hierarchical structure facilitates inference about heterogeneity of both preferences and information processing costs. Finally, we demonstrate likelihood identification of simple and complex attributes.

3.1 RI-DCMs with choice probabilities in closed MNL form

The key difficulty in solving the RI choice problem is to determine the unconditional choice probabilities $P(a)$ for the individual alternatives a. Typically, the solution is determined by a set of non-linear equations. Once these probabilities are known, the conditional choice probabilities follow immediately from Eq. 3 both for continuous and discrete prior beliefs.

To circumvent this issue, existing approaches with observational data typically choose, sometimes only implicitly, prior beliefs such that the resulting conditional choice probabilities have a closed multinomial logit form. The resulting choice probabilities depend on an additively separable index of observable characteristics of alternatives,

$$P(a|\omega) = \frac{\exp\{\mathbf{x}_a' \boldsymbol{\beta}\}}{\sum_{b \in A} \exp\{\mathbf{x}_b' \boldsymbol{\beta}\}},$$

where $\mathbf{x}_a'$ encode observable characteristics of alternatives and $\boldsymbol{\beta}$ are coefficients estimated from data. Observable characteristics may include non-price attributes and various price components. This expression for choice probabilities is identical to the RU logit with type 1 extreme value (T1-EV) distributed error terms.

The advantage of this strategy is that it translates the RI framework into a tractable logit form, which facilitates estimation. Moreover, it provides a micro-foundation for DCMs that include non-utility components (Joo, 2023). However, a crucial implication of constraining RI in this way is that documented behavioral patterns, such as interactive contributions of attributes or endogenous consideration sets, are ruled out a priori (see Bertoli et al., 2020, and Joo, 2023, for applications under these assumptions). Moreover, assumptions about beliefs implied by this formulation often, if not always, conflict with their rational counterparts, as discussed in more detail below.

Furthermore, existing contributions differ in their assumptions on the nature of simple information. One group of papers such as Brown and Jeon (2024) assume, similar to the proposed operationalization in Section 2, that certain components of utility are simple and thus do not require any processing. They analyze choices for insurance plans and assume that utility from choice options can be linearly decomposed into simple (insurance premiums) and complex payments (out-of-pocket-costs).[12] Other papers, including Joo (2023) and Natan (2021), build on the assumption that there are consideration shifters such as advertising or prior product purchases, that shift prior beliefs but have no impact on consumption utility.[13]

Cardell prior beliefs If prior beliefs over the uncertain utility components ($\{\mathbf{x}'_{a,c}\boldsymbol{\beta}_c\}_{a\in A}$) follow independently and identically distributed Cardell distributions, the solution to the RI problem yields logit choice probabilities with indices that are additively separable in simple (directly observed and processed) and complex (costly processed) components, see Brown and Jeon (2024), Bertoli et al. (2020), and Porcher (2019).

Brown and Jeon (2024) as well as Bertoli et al. (2020) show that the DM then always considers all alternatives, i.e., $P(a) > 0$ for all alternatives $a \in A$, and unconditional choice probabilities are given in closed-form by

$$P(a) = \frac{\exp\{C\mathbf{x}'_{a,s}\boldsymbol{\beta}_s/\lambda\}}{\sum_b \exp\{C\mathbf{x}'_{b,s}\boldsymbol{\beta}_s/\lambda\}}$$

where C is a function of the variance of complex component $\mathbf{x}'_{a,c}\boldsymbol{\beta}_c$ and information processing costs λ.[14] This together with Eq. 3 implies that choice probabilities

[12] A nice feature of the strategy proposed by Brown and Jeon (2024) is that the authors are able to differentiate choices under full information and under the existence of information frictions essentially directly. Since simple and complex utility components both affect costs of insurance, DMs acting under full information will react equally to changes in simple and complex components. However, if DMs react differently to changes in the two utility sources, one can conclude that imperfect information about the component associated with the smaller reaction prevails.

[13] Contrary to what the term "consideration shifters" may suggest, the models in Joo (2023) and Natan (2021) rule out the existence of consideration sets a priori.

[14] Formally, this is related to the following observation. Conditional choice probabilities derived under RI in Eq. 3 resemble choice probabilities obtained from RUM with alternative's utilities given by

$$u(a,\omega) = \mathbf{x}'_{a,s}\boldsymbol{\beta}_s + \mathbf{x}'_{a,c}\boldsymbol{\beta}_c + \log P(a) + \varepsilon_a$$

conditional on a specific choice set for an alternative a read

$$P(a|\omega) = \frac{\exp\{(\mathbf{x}'_{a,s}(C\boldsymbol{\beta}_s + \boldsymbol{\beta}_s) + \mathbf{x}'_{a,c}\boldsymbol{\beta}_c)/\lambda\}}{\sum_b \exp\{(\mathbf{x}'_{b,s}(C\boldsymbol{\beta}_s + \boldsymbol{\beta}_s) + \mathbf{x}'_{b,c}\boldsymbol{\beta}_c)/\lambda\}}.$$

Defining $\widetilde{\boldsymbol{\beta}}_s \equiv C\boldsymbol{\beta}_s + \boldsymbol{\beta}_s$, one can see how this derivation can motivate different coefficients for, e.g., a simple and a complex component of price. However, the resulting model otherwise is indistinguishable from a standard logit RUM.

Prior beliefs consistent with RI choice probabilities in closed MNL form In the application by Joo (2023), each alternative $a \in A$ is characterized by a set of observable consideration shifters $\mathbf{d}_a$, e.g., advertising or shelf placing, that solely affect prior beliefs but not consumption utility.[15] All product attributes are assumed to be complex so that $u(a, \omega) = \mathbf{x}'_{c,a}\boldsymbol{\beta}_c$. Given prior beliefs μ, that depend solely on the informational shifters $\{\mathbf{d}_a\}_{a\in A}$, and information costs λ the DM learns and chooses following the RI framework. Joo (2023) shows that for any combination of alternative specific payoffs $\{u(a, \omega)\}_a$, information costs λ, and strictly positive unconditional choice probabilities $P(a)$, there are prior beliefs μ that are consistent with choice behavior of a rationally inattentive DM.

However, the actual parameterization Joo (2023) brings to the data, i.e., a logit form with an additively separable index of alternative specific attributes $\mathbf{x}_{c,a}$ and information shifters $\mathbf{d}_a$, requires very specific beliefs about the index from alternative specific attributes $\mathbf{x}_{c,a}$. These beliefs are not derived from the objective distribution of this index in the marketplace and are substantially different from this distribution, as already implied by the full support assumption.[16] This limits the formulation proposed in Joo (2023) as a model of rationally inattentive DMs that acquire knowledge about the distribution of alternative specific attributes $\mathbf{x}_{c,a}$ in the marketplace over longer time horizons.

3.2 RI-DCM with general prior beliefs

The RI-DCM with general prior beliefs does not have a closed-form solution, and we need to solve for choice probabilities numerically to compute the likelihood in this model. However, over and above incorporating more realistic assumptions about beliefs, this generalization yields the qualitatively distinctive features of RI discrete choice we previewed in Section 1 and will elaborate on in Section 4. In applications, prior beliefs can be determined, for instance, by assuming rational expectations

where ε_a are identically and independently T1-EV distributed. The Cardell distributed prior beliefs over complex utility components $\mathbf{x}_{a,c}$, together with ε_a, follow the T1-EV distribution. This feature is key for obtaining unconditional choice probabilities in closed form. For a detailed derivation, see Supplemental Appendix A-1 of Brown and Jeon (2024) and Appendices A.1 and A.2 in Bertoli et al. (2020).

[15] Natan (2021) employs a similar identification strategy. In this paper, however, information processing costs λ explicitly depend on the size of the choice set.

[16] While Cardell beliefs result in additive separability as demonstrated by Brown and Jeon (2024), the model in Joo (2023) is not immediately consistent with Cardell beliefs because of the assumption that the outside good payoff is known with certainty.

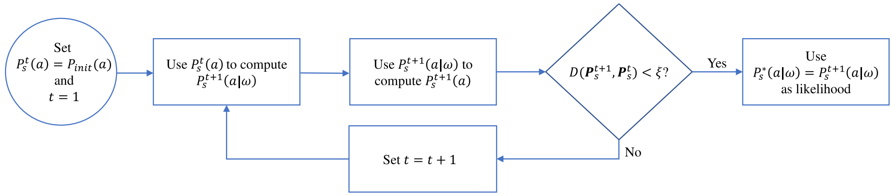

Fig. 1 Flowchart of the Blahut-Arimoto algorithm to compute the likelihood in RI-DCM

about the experimental design (in case of DCE data) or the empirical distribution of attributes (with observational data). However, the model could readily accommodate directly elicited, potentially heterogeneous beliefs about the (conditional) distributions of complex attributes and implied choice sets.

Likelihood computation Conditional on $\boldsymbol{\beta}_s$ and $\boldsymbol{\beta}_c$ the utility of the DM, $u(a, \omega)$, is given by Eq. 1. Given $u(a, \omega)$, $\mu_s(\omega)$, and λ, the likelihood can be obtained by computing the optimal conditional choice probabilities, $P(a|\omega)$, in Eq. 3. As $P(a|\omega)$ is a function of the endogenous unconditional choice probabilities $P_s(a)$ in Eq. 3, we solve for both $P(a|\omega)$ and $P_s(a)$ using the Blahut-Arimoto algorithm (Cover and Thomas, 2006).

The algorithm starts by initializing the $P_s(a)$ and iterates between updating $P_s(a|\omega)$ and $P_s(a)$ until convergence. The first step of each iteration t uses the optimality condition in Eq. 3 to compute $P_s^{t+1}(a|\omega)$ given $P_s^t(a)$ obtained in the past iteration. The second step computes $P_s^{t+1}(a)$ by integrating $P_s^{t+1}(a|\omega)$ over the (conditional) prior belief distribution:

$$P_s^{t+1}(a) = \sum_{\omega\in\Omega} \mu_s(\omega) P_s^{t+1}(a|\omega) = P_s^t(a) \sum_{\omega\in\Omega} \frac{\mu_s(\omega)\exp\{u(a, \omega)/\lambda\}}{\sum_{b\in A} P_s^t(b)\exp\{u(b, \omega)/\lambda\}}. \quad (5)$$

We calculate the distance between unconditional distributions obtained in subsequent steps, P_s^{t+1} and P_s^t, via the Bhattacharyya (1946) distance $D(P_s^{t+1}, P_s^t)$. The algorithm stops when $D(P_s^{t+1}, P_s^t) < \xi$ or when the maximum number of iterations $iter_{max}$ is reached. The converged conditional choice probabilities give us the individual level likelihood $\mathcal{L}(\mu_s(\omega), \boldsymbol{\beta}_s, \boldsymbol{\beta}_c, \lambda) = P_s^*(a|\omega)$.[17] Note that parameters ξ and $iter_{max}$ govern the precision of the numerical solution, and care must be taken in setting their value. Figure 1 shows the flowchart of the algorithm. Inference based on solutions from the Blahut-Arimoto algorithm, and specifically Markov Chain Monte Carlo (MCMC) estimation, is computationally costly as we need solutions for every unique combination of a choice set indexed by simple attribute realizations with model parameters visited by the MCMC.

To speed up computations, without giving up on required precision, we employ two optimization strategies. First, we optimize the initialization of the starting values ($P_{init}(a)$ in Fig. 1) by leveraging (saved) solutions from the current state of the MCMC. At each MCMC draw r, we set $P_{init}(a) = (1 - \gamma)P_s^*(a)|_{r-1} + \gamma/|A|$,

[17] Convergence of the algorithm has been proven by Csiszár (1974).

where $P_s^*(a)|_{r-1}$ denotes the converged unconditional choice probabilities from the last MCMC draw, $\gamma \in (0, 1]$ is a weight parameter, and $|A|$ is the number of alternatives in the choice set.[18] For $r = 1$ we use uniform choice probabilities as initial values. Second, we implement a parallelized version of the Blahut-Arimoto algorithm, which simultaneously computes likelihoods for multiple choice observations.

Identification As only the ratio of costs and preferences can be identified from choice data, preferences $\boldsymbol{\beta}_s$ and $\boldsymbol{\beta}_c$ can be identified by fixing λ and leveraging variations in simple and complex attributes across alternatives and choice sets presented to the DM.[19] However, there may be situations without variation in simple attributes. For example, one could think of "brand" as the only simple attribute. The RI model is still identified in this case, subject to variation in a complex attribute, e.g., total price, in the same way as a simple RUM with alternative specific constants. However, with only one configuration of simple attributes, there is only one optimal processing strategy (see expression 2). If this strategy results in a full consideration set, the model is empirically indistinguishable from a standard RUM. However, note that depending on unobserved heterogeneous preferences, consumers may have different consideration sets that are endogenous to their brand preferences and price sensitivity. However, if the environment is such that not all the brands are available all the time, there will be variation in the conditioning argument to the optimal processing strategy under RI.

Finally, it is important to note that point identification of preferences may not be achieved, since deterministic choice is a possible endogenous outcome under RI. This occurs, for instance, at extreme values of λ. Intuitively, a DM with a very high information acquisition cost does not react to variations in the complex attributes, and one with very low information acquisition costs fully processes the complex attributes, leading to deterministic choices (see Section 4). Both of these cases result in set-identification of preference parameters.[20] In this context, our Bayesian inference framework, illustrated next, will be useful, as the likelihood surface then exhibits flat regions, complicating maximum likelihood estimation.

3.3 Empirical identification with DCE data under preference heterogeneity

We now illustrate the Bayesian estimation and empirical identification of the RI-DCM with general beliefs proposed in Section 3.2, using simulated data, in a "small T, large N" setting, typical of DCEs in marketing. We use standard weakly informative subjective prior settings for the parameters indexing hierarchical prior distributions (see Appendix A.2 for details).

18 The transition kernel of our MCMC naturally results in unconditional distributions (implied by a proposal for a new parameter value) that often are close to those at the current state. The initialization guarantees non-degenerate starting values, i.e., a vector of probabilities with all entries larger than zero and smaller than one. In our simulations, we found $\gamma = 0.2$ to work well.

19 Appendix A.1 illustrates how a difference in processing costs $\Delta\lambda$ can be empirically identified if a DM with invariant preferences makes decisions in different environments.

20 We illustrate set-identification of preferences in Appendix A.3.

Table 3 Posterior means of preference distributions for different model specifications

Model	Brand	Price	Discount	$\|\beta_p/\beta_d\|$	$\|\beta_p/\beta_b\|$	LMD
Data generation	2.50	-1.00	$\equiv -\beta_p$	1.00	0.40	
RI-DCM	2.50	-1.00	$\equiv -\beta_p$	1.00	0.40	-2,530.29
	(0.05)	(0.02)				
RU logit separate	27.61	-9.16	4.44	2.06	0.34	-2,620.36
	(0.69)	(0.22)	(0.13)			
RU logit joint	9.81	-3.95	$\equiv -\beta_p$	1.00	0.40	-4,558.35
	(0.19)	(0.07)				

We report data generating parameters as well as the estimated posterior means of the RI-DCM and the benchmark logit models with and without the constraint $\beta_p = \beta_d$. Standard errors are in parentheses. $|\beta_p/\beta_d|$ and $|\beta_p/\beta_b|$ are the ratios of mean coefficients

We simulate data from the following hierarchical setup. A sample of rationally inattentive DMs ($N = 1,000$) face $T = 20$ choices between an inside and an outside good each. The utility of the inside good to DM j in choice task t is given by $u_{j,t} = \beta_{b,j} + \beta_{p,j}(p_t - d_t)$ where $\beta_{b,j}$ is the brand coefficient, $\beta_{p,j}$ the price coefficient, p_t is the price, and d_t is the discount. The utility of the outside option is normalized to zero: $u_O = 0$. In our simulation, brand and price are simple attributes and thus perceived and processed, i.e., integrated to an overall utility, immediately and at no cost, while the discount requires costly processing.

We first illustrate the case without heterogeneity in processing costs. Here, all individuals have the same processing costs of $\lambda = 0.25$. DMs differ in their structural utility parameters. Preference coefficients are generated from the following distributions: $\beta_b \sim \mathcal{N}(2.5, 0.25)$, $\beta_p \sim \mathcal{N}(-1, 0.04)$. Prices p and discounts d are drawn uniformly and independently from the following sets: $p \in \{2.5, 3, 3.5, 4, 4.5\}$ and $d \in \{0, 0.5, 1, 1.5, 2\}$ such that the resulting design is orthogonal. Individuals know the value of the price p, and for all prices and for any discount level d' the prior beliefs are given by $\Pr(d = d'|p) = 1/5$.

With the simulated data, we fit the RI-DCM and two RU logit specifications: one that allows for separate price and discount coefficients ("RU logit separate") and one with only one coefficient measuring the utility of money ("RU logit joint"), as in the data-generating process. We rely on Rossi's `bayesm`-package for the estimation of the hierarchical RU logit (Rossi et al., 2005). The estimation of the hierarchical RI-DCM employs Metropolis-Hastings steps to update individual-level preference parameters and relies on standard results for updating parameters indexing the hierarchical prior distributions (e.g., Rossi et al., 2005). We obtain the likelihood by solving the problem Eq. 2 for given parameters numerically with the Blahut-Arimoto algorithm, as described in Section 3.2. Without loss of generality, we fix the value of λ to be equal to its true value in estimation. We will revisit this point below.

Table 3 summarizes posterior means and Table 4 reports posterior variances. We see that the estimated RI-DCM nicely recovers data-generating parameters. We use log marginal density (LMD) estimates throughout the paper to compare model fits

Table 4 Posterior variances of preference distributions for different model specifications

Model	Brand	Price	Discount
Data generation	0.25	0.04	$\equiv \text{Var}(\beta_p)$
RI-DCM	0.33	0.09	$\equiv \text{Var}(\beta_p)$
	(0.05)	(0.01)	
RU logit separate	16.00	1.71	1.93
	(4.02)	(0.39)	(0.31)
RU logit joint	4.03	0.71	$\equiv \text{Var}(\beta_p)$
	(0.71)	(0.11)	

We report data generating parameters as well as the estimated posterior variances of the RI-DCM and the benchmark logit models with and without the constraint $\beta_p = \beta_d$. Standard errors are in parentheses

(see Rossi et al., 2005). The inferior fits of the RU logit models, the (still) current benchmark, testify to the empirical identifiability of the RI-DCM (see the last column in Table 3). The RU logit with separate parameters for price and discount fits the data much better than the RU logit with only one price coefficient, i.e., suggesting that different "sources of money" are valued differently. However, here the larger magnitude of the price coefficient, relative to the discount coefficient, simply reflects that rationally inattentive DMs react to the realized discount value adaptively, both as a function of the realized (simple) price that varies across choice sets and as a function of heterogeneous preference coefficients that vary across DMs.

As a consequence, the RU logit also struggles with measuring heterogeneity in preference parameters. For example, the RU logit dramatically overestimates the heterogeneity in the price coefficient (and the discount coefficient, where separately specified). This observation is important given that existing applications of RI to discrete choice with observational data present reinterpretations of RU logit choice conditioned on additively separable indices.

Finally, Table 5 reports elasticities for all models given changes in different discount and price levels. Elasticities are calculated using posterior expected (changes in) choice probabilities and, therefore, incorporate all posterior uncertainty. The columns Discount A and Price A report elasticities resulting from a discount decrease from $d = 1$ to $d = 0.5$, and a price increase from $p = 4$ to $p = 4.5$, respectively. Similarly, columns Discount B and Price B denote scenarios where discounts decrease from

Table 5 Elasticities for different price and discount levels

Model	Discount A	Price A	Discount B	Price B
RI-DCM	0.28	0.52	0.21	0.52
RU logit separate	0.25	0.53	0.29	0.58
RU logit joint	0.38	0.38	0.38	0.38

Columns report elasticities from changes in discount (Discount A and Discount B) and changes in price (Price A and Price B). The elasticities in the first two columns are calculated for an inside good with $p = 4$ and $d = 1$ and in the last two columns for an inside good with $p = 5$ and $d = 2$

$d = 2$ to $d = 1.5$, and prices rise from $p = 5$ to $p = 5.5$. In all instances, absolute and relative changes in total price are the same.

Under RI, price elasticities depend both on the source of the price variation and on the composition of the inside good. The RI-DCM yields that the price elasticity is much smaller (larger) when the change in total price comes through the complex discount (the simple price). The RI-DCM also yields that price elasticity from changing the discount further decreases at the higher simple price (compare columns Discount A and Discount B in Table 5). The reasons are the information friction associated with processing and integrating the complex discount and the stronger prior against the inside alternative when the simple price is larger.

The RU logit separate correctly picks up that price changes from the complex discount result in smaller elasticities than simple price changes. However, this model wrongly suggests that the elasticity from changing the price through the discount increases at the higher simple price. Finally, the RU logit that does not differentiate between price changes through the simple price and the complex discount necessarily overestimates (underestimates) elasticities from changing the discount (changing the simple price).

3.4 Heterogeneous preferences, heterogeneous information processing costs

In general, both processing costs and preferences likely vary across DMs. Hence, it may be theoretically appealing and efficient in a hierarchical model to structure heterogeneity in $\boldsymbol{\beta}$ as residual heterogeneity after considering heterogeneity in λ. Such a decomposition in the context of a hierarchical RI-DCM can be viewed as a micro-founded version of the idea behind Fiebig et al. (2010)'s generalized multinomial logit model that, in addition to preference heterogeneity, captures the heterogeneity in the scale of the error term in a hierarchical RU logit.

As already mentioned, the objective function in the RI framework is homogeneous of degree one with respect to payoffs and costs, and only their ratio is directly identifiable from choice data. However, from a statistical point of view, the combination of continuous heterogeneity in preferences with continuous heterogeneity in information

Table 6 Posterior means of preference distributions for the RI-DCM with and without scale mixture component

Model	Brand	Price	Discount	$\sigma^2_{\log(\lambda)}$	LMD
Data generation	2.50	-1.00	$\equiv -\beta_p$	0.09	
RI-DCM	2.40	-0.95	$\equiv -\beta_p$	0	-2,984.55
	(0.04)	(0.02)			
RI-DCM scale mixture	2.48	-0.99	$\equiv -\beta_p$	0.08	-2,884.90
	(0.06)	(0.02)		(0.01)	

We report data generating parameters as well as the estimated posterior means of the RI-DCM with homogenous and heterogeneous information processing costs. Standard errors are in parentheses. $\sigma^2_{\log(\lambda)}$ is the variance of log information processing costs in the population

Table 7 Posterior variances of preference distributions for the RI-DCM with and without scale mixture component

Model	Brand	Price	Discount
Data generation	0.25	0.04	$\equiv \mathrm{Var}(\beta_p)$
RI-DCM	0.41	0.07	$\equiv \mathrm{Var}(\beta_p)$
	(0.05)	(0.01)	
RI-DCM scale mixture	0.31	0.05	$\equiv \mathrm{Var}(\beta_p)$
	(0.06)	(0.01)	

We report data generating parameters as well as the estimated posterior variances of the RI-DCM with homogenous and heterogeneous information processing costs. Standard errors are in parentheses

costs gives rise to a scale mixture distribution. For a well-known example, the Student t-distribution can be derived as a scale mixture of normal distributions.

Because scale mixtures can be identified and distinguished from their non-mixed counterparts (see, e.g., Choy & Smith, 1997), one can identify heterogeneity in the unit information costs and preferences, as long as one is willing to assume a continuous distribution of preferences that is not a scale mixture. For example, popular semi-parametric distributions such as a mixture of normals are strictly continuous and not scale mixtures. However, again because the objective function defining the RI problem is homogeneous of degree one, the joint distribution of preferences and unit information costs is only identified up to the first moment of the latter distribution.

For illustration, we extend the simulation from Section 3.3 to include heterogeneity in information processing costs with $\log(\lambda_j) \sim \mathcal{N}(\mu_{\log(\lambda)} = -1.4, \sigma^2_{\log(\lambda)} = 0.09)$ in the data generating mechanism. As noted previously, the mean of this distribution is not jointly identified with the mean of the distribution of preference parameters. Hence, we fix $\mu_{\log(\lambda)}$ to the data generating value in estimation and without loss of generality.[21] Tables 6 and 7 document that we recover the joint distribution of preference parameters and information costs subject to fixing the first moment of the latter. Moreover, we see that by not accounting for heterogeneity in information processing costs, one overestimates the preference heterogeneity, here reflected in brand and price variance estimates in the RI-DCM with homogeneous information costs.

Complementing the illustrative simulations here, we conduct a simulation study that varies the number of inside goods (one versus two), simulates data with and without heterogeneity in processing costs (in addition to preferences heterogeneity), and adds a two-stage choice model with consideration sets from screening on price (see, e.g., Gilbride & Allenby, 2004; Pachali et al., 2023) to the model comparison. We simulate 50 data sets in the four data-generating settings. The benchmark models are estimated once with a single price parameter and once with separate parameters for the (simple) price and the (complex) discounts. We summarize results in Tables 15, 16, 17, and 18 and discuss additional details in Appendix A.2.

Not surprisingly, we find that only the RI-DCM recovers data-generating parameters. We also find that (i) we can reliably distinguish the data generating RI-DCMs from the benchmark models, (ii) the benchmark models fare relatively much worse in the

21 Normalization of the information processing costs is analogous to that of the error variance in RU logit models in that it does not affect market share or welfare computations.

larger choice set because of the corresponding increase in number of optimal RI information strategies, and (iii), slightly worse when processing costs are heterogeneous (in addition to preferences heterogeneity).

3.5 On the distinction between simple and complex aspects of a choice task

Different from extant search models, RI motivates information frictions even in situations in which all attribute information is essentially equally accessible, and the DM's challenge is not to resolve values of unknown attributes but to integrate accessible information to overall utility. Hence, another practical challenge of the proposed framework is the identification of simple and complex attributes. In some cases, prior knowledge may be sufficient to classify attributes, possibly as a function of the specifics of a product category under study or the experimental design. In other cases, we envision that the distinction between simple and complex attributes must be empirical.

This distinction is greatly facilitated whenever theory constrains coefficients in the utility function to be equal, such as in the example of different price components. In this case, descriptive models, or even just marginal summaries of the data, can reveal that choice probabilities react more strongly to changes, say, in price component A than in price component B. It follows that price component A is simple relative to price component B, and price component B is complex relative to component A (e.g., Brown & Jeon, 2024).

Obviously, this argument fails when theory allows for different utility coefficients for different attributes. Next, we illustrate by simulation that the distinction between simple and complex attributes is likelihood identified, even in this case. Whereas theoretical results imply that any combination of rationally inattentive behavior and information processing costs can be rationalized with some state-contingent payoffs (Lipnowski and Ravid, 2022), we illustrate that additive linear separability in utility contributions can suffice to distinguish between simple and complex attributes empirically, given Shannon costs and a non-degenerate distribution over states.[22]

One inside good We simulate 2,000 choice tasks, each involving a DM choosing between an inside and an outside good. The inside good is characterized by two linear attributes, x_s and x_c, that additively combine into overall utility. Attribute x_s is simple, and the attribute x_c is complex. Each of the two linear attributes is represented by three levels in the experimental design: $x_s \in \{2, 2.25, 5\}$ and $x_c \in \{1, 1.5, 3\}$. Preferences are given by $\beta_s = -\beta_c = -1$ and information processing cost is set to $\lambda = 0.5$. In this design, the generated data comprises probabilistic and deterministic choices. Attribute combinations defining inside goods are drawn from independent uniform distributions over the discrete attribute support. With the simulated data, we estimate different structural RI-DCMs.

In the first specification, the distinction between simple and complex attributes follows that of the data-generating process. In the second model, we (falsely) reverse what

[22] In the context of search, Abaluck et al. (2022) propose how to identify limited information about an attribute under the assumptions that another attribute is known to be fully processed and all-or-nothing learning.

Table 8 Identification of simple and complex attributes in a design with one inside alternative and one outside alternative

Model	min	25%	50%	75%	max	ML
Correct - Linear	-380.52	-371.39	-370.68	-370.25	-369.94	-369.93
Misspecified - Linear	-1076.47	-1067.22	-1066.28	-1065.71	-1065.11	-1039.15
Misspecified - Linear Out	-417.41	-412.02	-411.30	-410.95	-410.78	-410.70
Misspecified - Categorical	-382.63	-375.21	-374.04	-373.13	-370.54	-369.23

Quartiles as well as the minimum and the maximum of the log-likelihood MCMC draws are reported for the correct linear, the misspecified linear, the misspecified linear with an additional outside parameter, and the misspecified categorical model, respectively. The last column reports the maximum of the likelihood (ML) under an improper prior, i.e., the “frequentist maximum”

is simple and complex in estimation. The third model adds a coefficient for the outside good (equal to zero in the data-generating process). This model isolates the linearity of utility differences between inside goods in attributes as a source of identification. Finally, we estimate a model with completely flexible utility within attributes by coding the two attributes as categorical while misspecifying which attribute is simple and which is complex in estimation. As subjective prior distribution for preference parameters we use $\boldsymbol{\beta} \sim \mathcal{N}(0, 100\mathbf{I})$. We also report the maximum of the log-likelihood under an improper prior, to safeguard against an undue influence of this subjective prior setting.

Table 8 presents quantiles of the log-likelihoods from Markov-Chain-Monte-Carlo (MCMC) estimation and the numerical maxima of the likelihoods implied by the different models. We find that, subject to constraints on the utility function, there is scope for empirical identification of what is simple and complex in a choice task (comparing the first three lines of Table 8). However, once we give up on linearity in attributes (see the last line in Table 8), we can no longer distinguish between simple and complex in this minimal example.[23]

Two inside goods A basic constraint from utility theory, namely that of no cross-effects between alternatives in a model of perfect substitution, does not come into play when there is only one outside good. To showcase identification from this constraint, we extend the simulation described above and include a second inside alternative. The DM chooses between two inside goods and an outside good. The inside goods have two attributes, one simple and one complex with three levels each. The major difference to the one inside good case is that simple attribute realizations of two inside goods now effectively interact in determining the optimal processing strategy. For example, particular realizations of simple attributes may lead to considering both,

[23] In Appendix A.3, we additionally show that we can no longer distinguish between simple and complex attributes based on model fit once processing costs are either small enough or large enough such that (essentially) deterministic choices ensue. When λ becomes sufficiently small, all information is fully processed, and the conceptual and empirical distinction between simple and complex vanishes. When λ becomes sufficiently large, the information in complex attributes is never integrated into the overall evaluation of alternatives, and a model with extreme coefficients for the simple attribute (misspecified as complex) will approach a perfect fit to the data.

Table 9 Identification of simple and complex attributes in a design with two inside and one outside alternative

Model	min	25%	50%	75%	max	ML
Correct - Linear	-715.7	-709.2	-708.6	-707.6	-707.3	-707.1
Misspecified - Categorical	-1589	-1585	-1583	-1582	-1581	-1580.3

Quartiles as well as the minimum and the maximum of the log-likelihood MCMC draws are reported for the correct linear and the misspecified categorical model, respectively. The last column reports the maximum of the likelihood (ML) under an improper prior, i.e., the "frequentist maximum"

only one, or none of the two inside alternatives. Misspecifying the simple attribute that drives the DM's choice of information strategy as complex fails to capture these possibilities, even if we drop the linearity constraint and code all attributes categorically (see Table 9).

4 Features of discrete choice in MAMA contexts under RI

In this section, we illustrate the implications of RI theory for discrete choice in MAMA settings, as common in marketing. We use the RI-DCM with general beliefs and show through simulations how several well-documented phenomena in the discrete choice literature that are difficult to justify in a RU framework naturally follow from RI. We further demonstrate that estimating a standard RU logit can yield misleading conclusions about the behavior of RI agents. This approach follows a common research strategy in that literature. Often, various logit models are estimated and compared in different contexts, e.g., with varying numbers of inside alternatives. This comparison then allows us to identify the moderating effect of context, thereby revealing deviations from the standard RU logit model.

While certain implications have been discussed in prior (theoretical) RI literature, we add to this body of work by exploring additional implications due to the MAMA structure. In general, the presented implications arise from how RI agents translate the MAMA choice environment into the unconditional choice probabilities outlined above. Existing RI-DCMs impose simplifying assumptions on this very process to facilitate estimation at the expense of not capturing the features presented here. In our illustrations, we distinguish between i) *endogenous features* of RI (Section 4.1) and ii) *context effects* (Section 4.2).

Endogenous features arise as a result of the optimal allocation of limited cognitive resources in the RI-DCM. These features can explain empirical phenomena in MAMA settings that cannot be captured by basic RUMs. Previously, researchers have applied diverse, non-unified adjustments to RUMs to address these phenomena. Table 10 outlines important endogenous features of RI and includes examples from studies that have made non-unified modifications to RUMs to accommodate the related phenomena. In Section 4.1, we discuss the following endogenous features of discrete choice under RI:

Table 10 Endogenous characteristics of RI and examples of extant related literature

Feature	Related literature
Stochastic choice due to limited attention	Difficulty of comparison (e.g, Shugan, 1980), imperfect perception (Thurstone, 1927)
Inattention to attributes	Shrinkage estimation (Gilbride et al., 2006; Yegoryan et al., 2020)
Attribute interactions	Information integration theory studies how the formation of overall judgments may depend on attributes in a non-additive way (Anderson, 1981, 1982), brand specific price coefficients (Blattberg et al., 1995)
Inattention to alternatives	Descriptive models studying the impact of advertising (e.g, Terui et al., 2011; Goeree, 2008; Ching et al., 2009; Ching, 2010), brand and shelf space (e.g, Bronnenberg & Vanhonacker, 1996), and price (e.g, Andrews & Srinivasan, 1995); consumer search (Hauser & Wernerfelt, 1990; Roberts & Lattin, 1991) with price uncertainty (e.g, Mehta et al., 2003; Honka, 2014; De los Santos et al., 2012; Honka & Chintagunta, 2017), match value uncertainty (e.g, Kim et al., 2010; Kim et al., 2017; Moraga-González et al., 2023), or multidimensional uncertainty (e.g, Chen & Yao, 2017; Yao et al., 2017)

1. *Stochastic Choice due to Limited Attention:* This feature allows explaining stochastic choice without invoking unobservable utility shifters.
2. *Inattention to Attributes:* Under RI, complex attributes will not receive full attention and their effect on choice is dampened. We show that simple attributes moderate the impact of complex attributes, and we show that this, in turn, explains attribute interactions as well as brand specific price coefficients. Moreover, this implies non-standard substitution patterns at the attribute level both within an individual alternative and across alternatives relative to a RUM with an additively separable utility index.
3. *Inattention to Alternatives:* Under RI, consideration sets are an endogenous outcome. We show how simple attributes drive their formation. Moreover, we illustrate how the choice set itself affects the consideration of individual alternatives.

Relatedly, the discussed context effects demonstrate how variations in the choice environment affect choice behavior under the assumption of RI. In Table 11 we summarize different context effects that have been studied in the literature. In Section 4.2, we illustrate how the following context effects can be reproduced within the RI framework:

1. *Impact of Information Costs and Incentives:* High processing costs decrease the impact of complex attributes on choice, whereas high incentives increase their impact. We show that such variations may result in choice reversals and affect the formation consideration sets, yielding deterministic choice in extreme cases.

Table 11 Effects of choice context variations and examples of extant related literature

Manipulation	Affected RI primitive	Resulting RI behavior	Extant literature
Processing cost effect: Variation in the difficulty in determining the payoffs of an alternative, e.g., through the number of necessary (mathematical) operations	Unit information processing costs λ increase.	Processing costs reduce the impact of complex attributes and the consideration set size. Choice consistency is U-shaped, and choice is deterministic for extreme costs.	In search literature, higher search costs result in smaller consideration sets (e.g., Honka et al., 2019). In marketing, choice task complexity is measured by the amount of information (e.g., Jacoby et al. 1974; Keller & Staelin, 1987) or the structure of the choice set, e.g., the number of differing attributes (Mazzotta & Opaluch, 1995), attribute correlations within a choice set (DeShazo & Fermo, 2002), or entropy (Swait & Adamowicz, 2001). Higher task complexity reduces choice consistency (DeShazo & Fermo, 2002), implies smaller impact of complex attributes (e.g., Chetty et al., 2009), and promotes the use of simple choice rules (Swait & Adamowicz, 2001; Orme, 2019).
Incentives effect: Change in the realization probabilities.	Payoffs are scaled by a factor ρ.	Impact is the inverse of the effect of processing costs.	Incentive-aligned experiments result in better out-of-sample predictions in real-world applications (Ding et al., 2005; Ding, 2007; Dong et al., 2010); Cao and Zhang (2021) report increases in price sensitivity.
Attribute range effect: The difference between the minimum and the maximum attribute levels increases.	Payoffs of some alternatives change.	Impact of affected attribute on choice probabilities increases, and there is a U-shaped relation between choice consistency and the range.	Empirical evidence on the impact on attribute effect is mixed (for an overview, see Bestard & Font, 2021), choice consistency decreases (Dellaert et al., 1999), and it becomes easier to detect non-linearities (Ohler et al., 2000).
Attribute level effect: The number of intermediate levels of an attribute increases.	The set of possible choice sets Ω (with intermediate payoffs) increases.	Depending on the incentives set by the additional levels, the impact may either increase or decrease.	More attribute levels are typically observed to lead to a stronger attribute impact (e.g., Liu et al., 2009), albeit contradicting evidence (e.g., Hensher, 2006) exists.

Table 11 continued

Manipulation	Affected RI primitive	Resulting RI behavior	Extant literature
Attribute correlation effect: Experimental designs that are not orthogonal or that contain attributes that have in the real world different correlation structures.	Prior beliefs μ vary.	Changes in prior beliefs may induce choice reversals and result in (deterministic) choice of inferior alternatives for a significant part of viable beliefs	Consumer search models identify the significance of prior beliefs for (discrete) choice (for a recent overview, see Jindal & Aribarg, 2021). A large marketing literature identifies the impact of price image for brands and stores (e.g., Hamilton & Chernev, 2013; Lourenço et al., 2015; Lombart et al., 2016).
Choice set expansion: The number of available alternatives in a choice set increases.	Number of possible choice sets (states) $\|\Omega\|$ and actions $\|A\|$ increases.	Violation of independence of irrelevant alternatives and monotonicity (e.g., Matějka and McKay, 2015); decrease in choice consistency, estimated coefficients become a function of the experimental design.	Reduced form approaches relating the error term variance to the choice set size identify a U-shaped relationship due to a trade-off between statistical efficiency and choice complexity (DeShazo & Fermo, 2002; Caussade et al., 2005). Estimated coefficients may either increase (e.g., Hensher, 2006) or decrease (e.g., Meißner et al., 2020) in the choice set.

Moreover, we illustrate that the proposed RI-DCM may be useful for bridging from laboratory settings (with low incentives) to real-world markets (where stakes and thus incentives to process information are high).

2. *Attribute Range/Dispersion and Levels Effects:* Under RI, the impact of complex attributes is affected by their distribution. A higher dispersion, for instance, typically results in more information processing and therefore a stronger impact. We illustrate that this feature explains empirically documented variations in choice consistency due to variations in the DCE design.
3. *Price Image and Attribute Correlation:* We demonstrate that both correlations between simple and complex attributes as well as correlations between complex attributes can be accounted for in the proposed model and that these correlations affect choice.
4. *Choice Set Expansions:* In contrast to RUMs, the implied error variance may depend on the choice set size under the proposed RI-DCM. We show that RI choice data looks less consistent, i.e., has a higher error term variance, through the lens of logit when the choice set expands.

Throughout this section, our illustrations build on the example introduced earlier, i.e., a DCE where the DM has to decide whether to purchase a specific car, potentially out of a set of different alternatives including an outside option, with a simple price and a complex discount. We generate data from a utility function where the simple price and the complex discount have the same absolute impact on utility, that is, $\beta_p = -\beta_d$. For illustration purposes, we vary data-generating coefficients or the design of the choice task across the following simulations. The payoff of the outside alternative is normalized to zero, $U_O = 0$, throughout. Unless stated otherwise, we simulate $T = 1,000$ choices in each illustration for statistically reliable inference from singular simulated data samples, and we assume information processing costs of $\lambda = 0.5$.

4.1 Implications of RI for choice in MAMA settings

4.1.1 Stochastic choice due to limited attention

As already mentioned, a major difference between the RUM and the RI-DCM lies in the interpretation of the error term. In RUM, the error term represents utility shifters that are known to the DM but not to the analyst. In contrast, the stochasticity of choice in RI is due to the DM's cognitive constraints. While the RU interpretation may fit applications to observational data in which the data only sparsely reflect the actual choice environment, it lacks appeal in situations where cognitive constraints significantly influence individual decisions. RI-DCM helps link the randomness in choices to the decision-making environment in these situations. In the typical DCE, the analyst fully controls the amount of information provided to the DM, making the RU interpretation unappealing and the RI-DCM potentially very useful.[24]

[24] Some extant research, including the original interpretation of the probit model in Thurstone (1927), proposes that choice errors in DCEs result from cognitive processes (e.g., Louviere et al., 1999). However, unlike RI, this research does not build on an integrated solution to a constrained optimization problem.

4.1.2 Attribute interactions and inattention to attributes

Observed simple attributes shift prior beliefs, influencing the DM's processing of complex attributes. As a result, the impact of observed complex attribute levels on choice depends on the levels of simple attributes in the set, even when the utility function is linearly and additively separable. Existing (statistical) models for attribute inattention, e.g., Gilbride et al. (2006) and Yegoryan et al. (2020), rely on shrinkage estimation and implicitly assume that (in-)attention is constant across different choice sets, in contrast.

For an intuitive example, imagine you are considering buying a car listed at $25,000. If you already know you can buy the same (up to an unknown complex discount level) car for $20,000 elsewhere, you are unlikely to put in the effort to find out the size of the discount at the higher-priced dealer. However, if the car is listed at $20,500, you might invest the effort to determine and account for the discount size, as a significant reduction could make it cheaper than the $20,000 option. For a more formal illustration, consider the effect of the discount on choice for different simple price levels. The complex discount d is either $d = 0$ or $d = 2$ with equal probability, and the simple brand coefficient is always $\beta_b = 2$. In Fig. 2, the left panel displays the choice probabilities of the inside good for the two possible discount levels. The right panel shows the impact of a discount increase from 0 to 2 in this example.

Under RI realized values of simple attributes determine the optimal processing level of complex attributes. If realized simple attribute values indicate that an alternative is either very attractive or unattractive, e.g., due to a very low or very high simple price (relative to the known distribution of complex attributes), then processing the complex attributes is less beneficial. The potential losses from a wrong decision based on prior beliefs are rather small. In the limit, the impact of the realized discount is zero at both very low and very high prices, while it is highest at a price where ex ante the DM is indifferent between the inside and the outside good ($p = 3$).

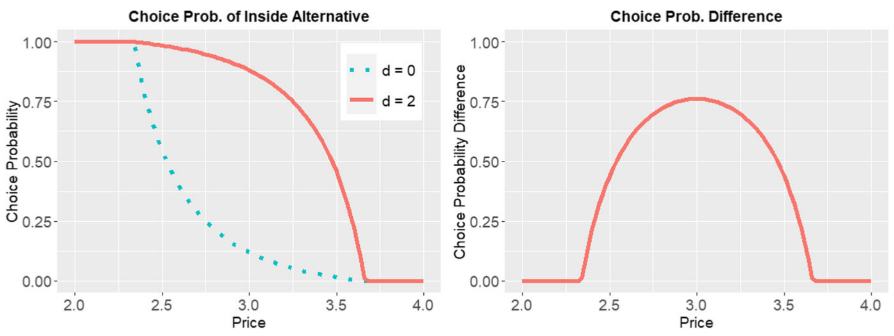

Fig. 2 Hump-shaped impact of a constant discount increase for different (simple) price levels. The left panel displays conditional choice probabilities as a function of the simple price for the possible discount levels. In the right panel, differences in the choice probability of the inside good are displayed when the realized discount increases from $d = 0$ to $d = 2$ for different simple prices p. Notably, the discount effect is zero for extreme prices p and it is highest at $p = 3$ where the DM is indifferent between alternatives given prior beliefs

Table 12 Logit approximation with linear and quadratic interaction terms of simple price and complex discount

Model	Brand	Price	Discount	Discount × Price	Discount × Price^2
Main Effect	15.82	-6.01	2.21		
	(0.47)	(0.18)	(0.08)		
With Interactions	15.20	-5.76	-10.72	8.34	-1.34
	(0.63)	(0.24)	(2.12)	(1.53)	(0.22)

First and third row show coefficient estimates for approximations without and with interaction terms and standard errors are indicated in parentheses below. In the simulated data, price p is drawn uniformly from the interval [2, 4], and d follows $\Pr(d = 0) = \Pr(d = 2) = 0.5$

Note that because of processing costs, there are cases where realized values of the complex discount cease to matter even when neither of the available alternatives is dominant. For example, at a price of $p = 2.25$ the outside (inside) option is the better choice in the absence (presence) of a discount, but the DM optimally chooses not to process the complex discount given the associated costs.

Next, we highlight how fitting a RU logit model to data generated from the RI-DCM may (mis-)lead an analyst into questioning a theory-based utility function. Table 12 illustrates that a main-effects logit model fitted to observations generated from the RI-DCM with $T = 5,000$ choices infers a discount coefficient that is much smaller in absolute value than the price coefficient. A logit model that allows for interactions infers that discount and price do not independently contribute to choice. A researcher pursuing a RU interpretation of these estimates will be left puzzling about the lack of a microeconomic justification. Similarly, RI as a data generating mechanism can motivate brand-specific coefficients in a logit model fitted to RI data. Interestingly, industry researchers generally include brand-specific price coefficients in models fitted to data from DCEs, despite the pushback from academic researchers who call out the lack of an economic rationale for such interactions (see, e.g., Sawtooth Software, 1996).

The interaction effects illustrated in the previous example imply marginal rates of substitution that are fundamentally different from those implied by the RUM. To see this, consider our next example that asks the DM to choose between an inside alternative and an outside good, however, with the modification that the complex discount d is uniformly distributed now on the set $\{0, 0.5, ..., 3.5, 4\}$ and the simple brand coefficient equals $\beta_b = 6$.[25] Fig. 3 displays iso-choice-probability sets, that is, combinations of the discount and the price that result in the same conditional choice probabilities for the RI-DCM (left panel) and the RU logit (right panel) based on the same utility function.

Higher discounts require higher prices in order to keep choice probabilities constant. However, under RI this relationship is non-linear whereas under the RU logit model this relationship is linear with a slope of one. In the RU logit model, choice probabilities are fully determined by the utility indices of the alternatives in the choice set. Under

[25] We increase the number of discount levels and adjust the brand coefficient accordingly for illustration purposes and without loss of generality.

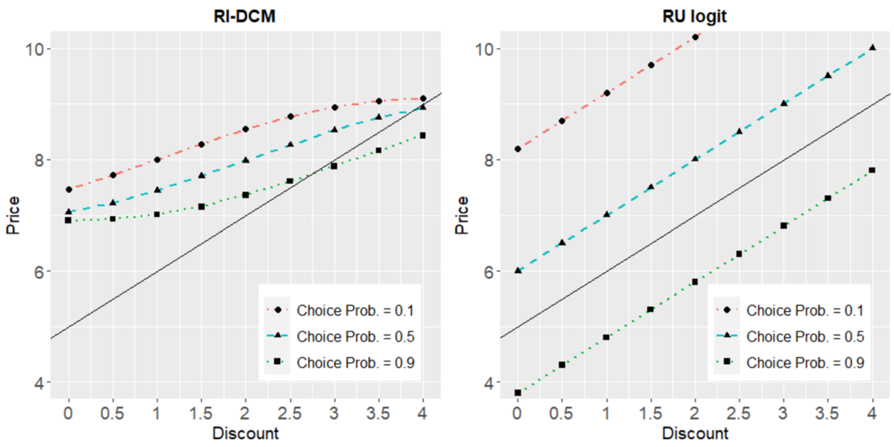

Fig. 3 Iso-choice-probability sets for the RI-DCM and the RU logit. This Figure displays sets of price-discount combinations of the inside good that result in the same conditional choice probabilities for the RI-DCM (left panel) and the RU logit (right panel). The thin solid line is the 45° line. Under the RU logit, the iso-choice probability sets are linear and individual sets are parallel to each other. In contrast, under the RI-DCM, the substitution rate varies depending on the composition of the inside good so that it is non-linear and the individual sets are not parallel. The dots in the left panel indicate actual attribute combinations of the inside good

RI, it matters whether the source of utility is a simple or a complex attribute. As shown in the left panel of Fig. 3, the ratio of changes in the discount and price that keep choice probabilities constant is smaller than one, i.e., an increase of the discount by one unit offsets an increase in the price that is strictly smaller than one under RI even though both price and discount have the same impact on utility. This is explained by the information friction present in the RI-DCM, which makes choice probabilities a function not only of an alternative's utility value but also of the source of that utility.

Under the RI-DCM, there will be prices that are sufficiently low (high) so that the DM chooses deterministically (given a simple price) in the limit. Consequently, with a fixed discount distribution, the iso-choice sets become lower (upper) contour sets with a boundary that is flat in the discount. In contrast, under the RU logit, there are no combinations of finite discounts and prices so that the DM chooses deterministically.

For a final illustration, Fig. 4 displays the conditional choice probability of the inside good for different combinations of the price and the discount for a fixed utility. All points displayed are associated with the same net utility equal to one, $u_I = 1$ where $u_I = \beta_b - \beta_p p + \beta_d d$ with $\beta_b = 6$ and $\beta_p = -\beta_d = -1$. However, going from left to right, we increase the discount and the price simultaneously by the same amount so that $p - 5 = d$ within the same experimental design. Under the RU logit, the choice probability of the inside good remains constant. In contrast, choice probabilities decrease weakly as the price increases under the RI-DCM. Even though the discount is just a negative price in utility terms, the distinction matters under RI if price is simple and the discount is more complex to process and integrate.

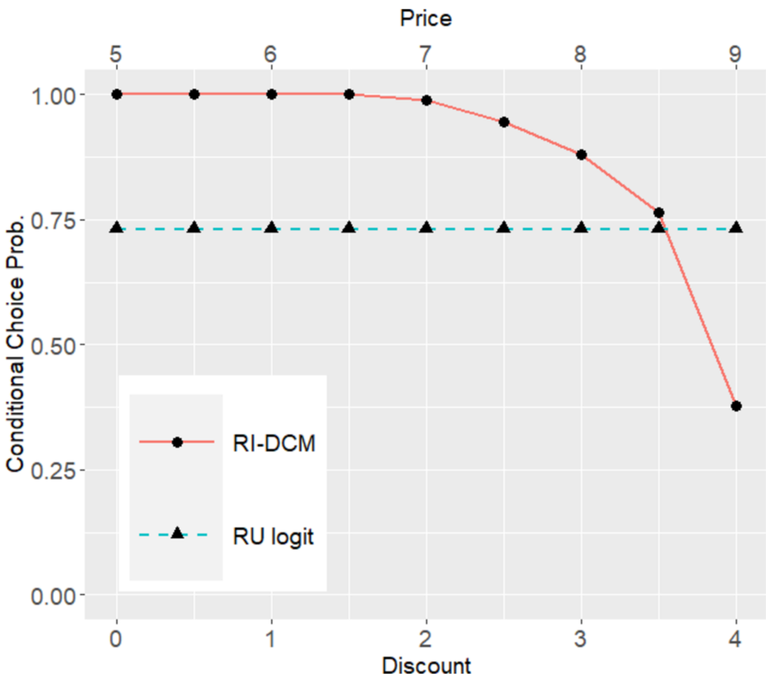

Fig. 4 Conditional choice probabilities for different compositions of the inside alternative under RI-DCM and RU logit. This Figure depicts the conditional choice probability for the inside good in the discount under the constraint that the price increase equals the discount increase with $p - 5 = d$ so that the net utility of the inside good is constant. While choice probability is constant under the RU logit, it is weakly decreasing when both the price and the discount are increasing in the considered variation of the inside good composition. The dots are the conditional choice probabilities of the specific inside good compositions

4.1.3 Inattention to alternatives

A central feature of discrete choice under RI is that consideration sets form endogenously. These sets include only the alternatives that have a strictly positive probability of being chosen. Consideration sets arise as a result of the DM's optimal information strategy. Here, we demonstrate how attributes—specifically the configuration of simple attributes—determine which alternatives are considered.[26]

Figure 5 depicts choice probabilities in a choice set with four inside alternatives $i = 1, ..., 4$ that provide utilities of $u_i = \beta_{b,i} - p_i + d_i$ with $\beta_{b,i} \in \{2, 1.75, 1.5, 1\}$, $p_i = 2$ and a discount d_i with $\Pr(d_i = 0) = \Pr(d_i = 2) = 0.5$. The specific values for $\beta_{b,i}$ are chosen for illustration purposes. While $\beta_{b,4}$ represents the least preferred brand alternative, we demonstrate how the DM reacts qualitatively differently to this brand as the composition of the choice set changes. The top-left panel of Fig. 5

[26] For related illustrative examples of endogenous consideration set formation that do not differentiate between attributes, see Caplin et al. (2019).

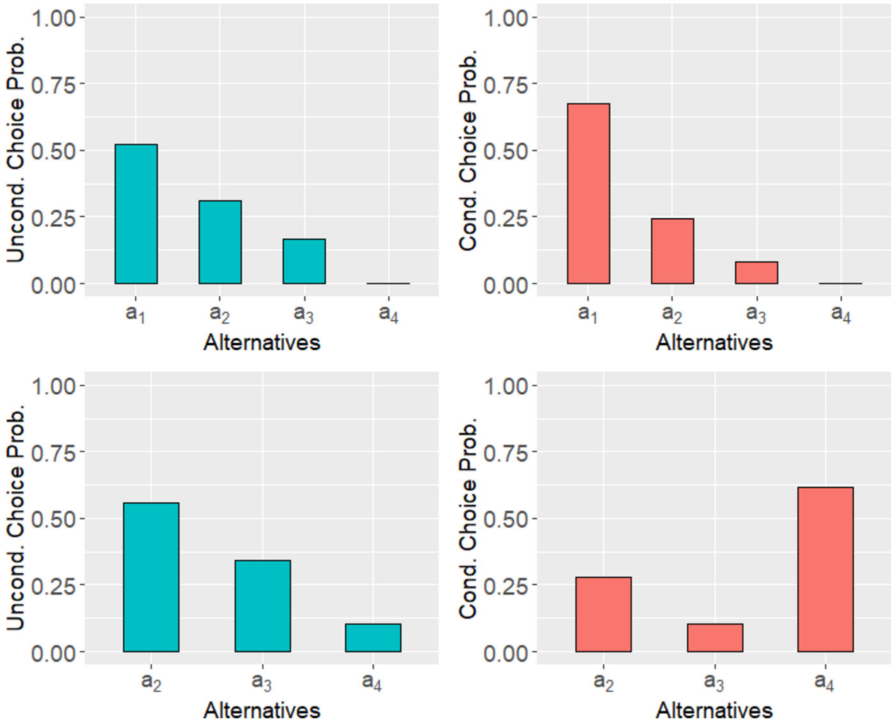

Fig. 5 Consideration set formation. The above panels present choice probabilities of a rationally inattentive DM facing up to four alternatives $\{a_1, a_2, a_3, a_4\}$ ordered from highest to lowest simple brand $\beta_{b,i}$ with identical prices p_i and identically distributed complex discounts d_i. The top-left panel shows the unconditional choice probabilities. The top-right depicts conditional choice probabilities given a choice set where alternative a_4 provides the highest utility. Due to information frictions, even in such a case, alternative a_4 is never chosen. The bottom-left panel exhibits the updated unconditional choice probabilities in a reduced choice set that drops alternative a_1. Finally, the bottom-right panel shows that, as a consequence of updating unconditional choice probabilities to the smaller set, alternative a_4 has the highest conditional choice probability in the state where it delivers the highest payoff

illustrates unconditional choice probabilities that represent the DM's beliefs about how she will choose before any processing of the complex discount has taken place. The top-right panel shows the choice probabilities conditioned on a specific choice set, i.e., realized values of the complex discount (from the analyst's perspective, as the DM will not necessarily learn the exact choice set because of processing costs). The complex discounts in the specific choice set are $d_1 = d_2 = d_3 = 0$ and $d_4 = 2$, implying that a_4 provides the highest utility in the specific choice set ($u_4 = 1 > u_j$ for $j \neq 4$). Yet, as apparent from the figures, the DM does not consider a_4 because excluding a_4 from her endogenous consideration set was (a priori) optimal for the DM given her processing costs and the small prior probability that a_4 is, in fact, optimal.

However, in contrast to extant two-stage models (e.g., Gilbride & Allenby 2004; Goeree, 2008; Terui et al., 2011), RI predicts that a_4 will be considered once a_1 is no longer available to the DM (see the unconditional choice probabilities in the bottom-left panel of Fig. 5). Due to updating unconditional choice probabilities to the smaller set, alternative a_4 has the highest conditional choice probability in the state where it delivers the highest pay-off (bottom-right panel).

Accordingly, RI implies deterministic consideration sets conditional on λ, prior beliefs μ_s and a utility function, while choice conditional on such a consideration set is stochastic. However, because prior beliefs μ_s are a function of realized simple attribute values, consideration sets will generally change from choice set to choice set in ways that cannot be captured by an alternative specific index or decision rule. If the simple information in a choice set does not vary, RI still implies consideration sets that are endogenous to consumers' preferences, priors over complex attributes, and information processing costs. An attractive feature of characterizing consideration sets this way is that the exclusion of, e.g., a brand from consideration in a particular choice set, does not have to be motivated by persistent extreme tastes or screening rules. The latter may not generalize to changes in the set of available brands or the priors over complex attributes.

4.2 Effects of choice environment variations under RI

4.2.1 Impact of information processing costs and incentives

A large body of experimental evidence documents that the complexity of choice tasks and incentives affect choice behavior (e.g., Swait & Adamowicz, 2001; Ding et al., 2005). In RI, both aspects influence attention allocation and, thus, observable choice.

Recall that the costs of information processing in RI are the product of mutual information and the strictly positive unit information cost $\lambda > 0$, see expression Eq. 2. Structurally, characteristics of the (expected) choice task as well as characteristics of the DM relate to λ (e.g., Regier et al., 2014). To continue with our example of a complex discount, processing the eligibility requirements will be affected by the number of criteria that must be checked, or even the font size used to describe the discount. Intuitively, more criteria that need checking or a smaller font size will increase λ. Similarly, a less constrained DM, or more experience with the product or the eligibility criteria, will be reflected in a relatively smaller λ.

Finally, it is possible to cast λ as a function of the incentives offered in a DCE.[27] For example, if an incentivized DCE instructs participants facing N choice tasks that one of the N choices will become an actual transaction, the realization probability of a specific choice is $\rho = 1/N$. With probability $1 - \rho$ the DM's choice is hypothetical, i.e., she does not actually obtain the chosen alternative.[28] From the perspective of the

[27] The implicit assumption here is that a change in incentives does not affect the payoff function or the (subjective) prior distribution over states ω.

[28] Recall that in an incentive-aligned DCE, the DM is endowed with a budget. When the DM chooses the outside option, she retains the endowed budget.

DM, the resulting RI objective function for each choice task is given by

$$\rho\left[\sum_{\omega\in\Omega}\mu_s(\omega)\left(\sum_{a\in A}P(a|\omega)u(a,\omega)\right)-\frac{\lambda}{\rho}\left[\sum_{\omega\in\Omega}\mu_s(\omega)\left(\sum_{a\in A}P(a|\omega)\ln P(a|\omega)\right)\right.\right.$$
$$\left.\left.-\sum_{a\in A}P(a)\ln P(a)\right]\right].$$

This formulation reveals that a higher realization probability has the same impact on choice behavior as a decrease in the information processing costs. For example, a 1% increase in information processing costs λ will be offset by a 1% increase in the realization probability ρ. Thus, one way of interpreting the patterns we illustrate next is through the lens of changing incentives in a DCE.

Consider the case when the DM chooses between an inside alternative, characterized by a simple brand valued at $\beta_b = 1.2$ at a simple price $p = 2$ and a complex discount d with $\Pr(d = 0) = \Pr(d = 2) = 0.5$. Based on expected utility, the DM thus prefers the inside good. Figure 6 illustrates how λ affects attention and choice in this example. The top left panel shows that when λ increases, the processing of complex attributes decreases until the DM learns nothing beyond the known distribution of the complex discount attribute (at $\lambda = \lambda''$), i.e., conditional choice probabilities equal the unconditional choice probabilities. At $\lambda \geq \lambda''$, the DM deterministically chooses the inside option based on prior expectations (bottom left and right panel), which of course implies that choice probabilities no longer change as a function of the complex discount (top right panel). At $\lambda = \lambda' = 0$, the DM perfectly learns the complex discount and deterministically chooses the alternative with the highest utility, and hence maximally reacts to changes in the complex discount value.

Note that changes in λ can impact what is revealed about the DM's preferences. The bottom-left panel of Fig. 6 shows that as λ increases, there is a choice reversal as the DM switches from choosing the outside good (based on learning that the complex discount does not apply) to choosing the inside good (based on prior expectations and without learning the true choice set ω). Finally, the bottom-right panel of Fig. 6 shows that choice probabilities, and here specifically the probability of making a choice error (from the point of view of the analyst who has all information about alternative specific payoffs), can be non-monotonic in the amount of cognitive processing (top-left panel). The non-monotonic relationship here derives from the prior pointing to the payoff maximizing choice in the absence of processing complex information.

To showcase what an analyst taking a RU perspective when analyzing RI choice data may find in this example, we fit logit models to RI choices conditional on different values of λ. Figure 7 summarizes point estimates of logit coefficients for brand, price, and discount, across different simulated RI choice data sets with varying λ.

We see that absolute values of the brand and price coefficients, i.e., the coefficients associated with the simple attributes, first decrease and then increase, while that of the complex discount decreases in λ. The latter effect is immediate, since higher

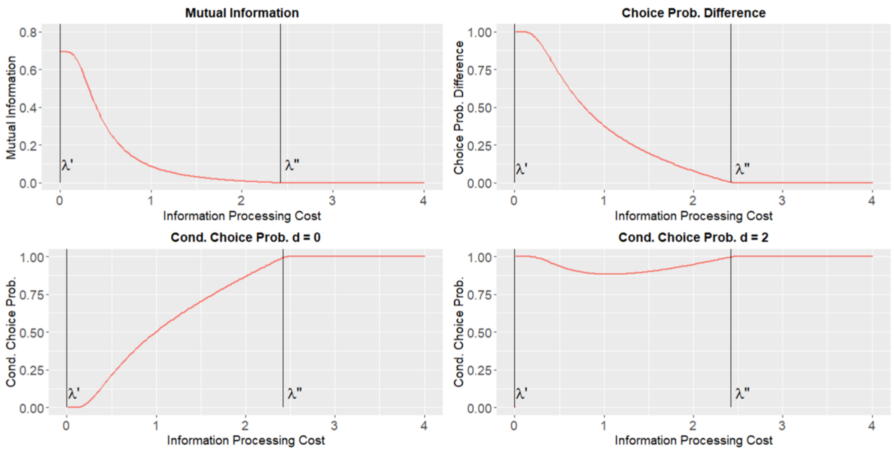

Fig. 6 Impact of information processing costs on attention, discount effect, and choice. The upper left panel shows mutual information as a function of information costs. The upper right panel displays the impact of a fixed increase of the complex discount (from 0 to 2) on conditional choice probability for different levels of λ. The panels in the bottom row show conditional choice probabilities of the inside alternative in choice sets where the discount is 0 (left) and 2 (right). Thus, the inside good provides a lower (higher) payoff in the left (right) panel than the outside option. Note that choice is deterministic for information costs λ equal to zero and larger than λ''

processing costs dampen the effect of the discount, as discussed previously. The rationale for the former pattern is that at small values of λ, the DM processes in most choice sets all available information, and choice becomes nearly deterministic. For intermediate levels of λ, some choice sets (characterized by different simple prices) will motivate more, and some less information processing, causing a higher level of overall stochasticity in the data that is reflected in absolutely smaller brand and price coefficients (from the viewpoint of a RU logit).

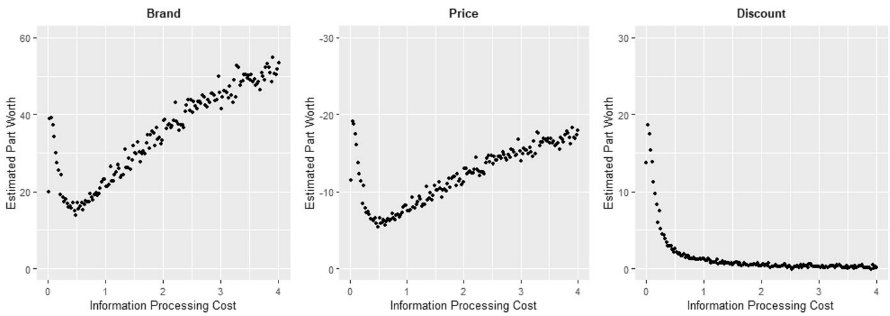

Fig. 7 Logit approximation for different levels of information processing costs. Each panel shows logit estimates of the respective coefficients for varying levels of information processing costs λ. Note that each point is the result of an estimation from simulated data with $T = 1,000$ choice tasks each. The data generating parameters are $\beta_p = -\beta_d = -1$, $\beta_b = 2$, p is uniformly drawn from [2, 4], and d is distributed with $\Pr(d = 0) = \Pr(d = 2) = 0.5$

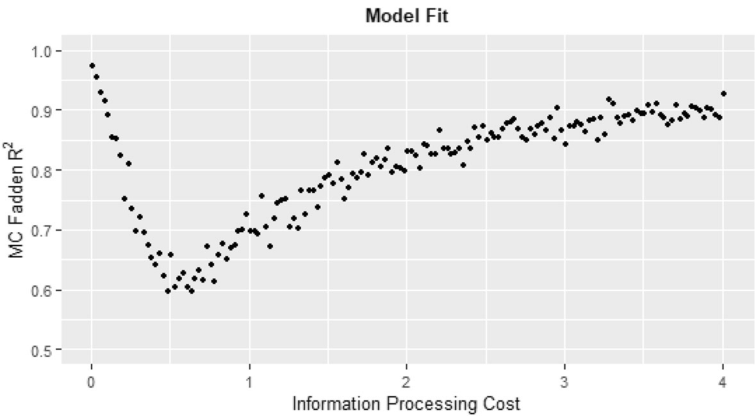

Fig. 8 Choice consistency is non-monotonic in the information processing costs λ. Choice consistency is measured by McFadden's pseudo R-squared. For details, see Domencich and McFadden (1975)

Eventually, as λ increases, realized discount values are ignored, as it becomes too costly to process the corresponding complex eligibility requirements and the discount coefficient approaches zero in the logit fit. However, as the amount of processing of complex information decreases beyond some level, so does the level of stochasticity in the data. Eventually, RI choices are based on prior information only, conditioned on the simple attributes brand and price here, and deterministic. This is reflected in absolutely increasing brand and price coefficients in the logit fits, summarized in Fig. 7. It is common in the choice modeling literature to report the estimated error term variance as a measure of choice consistency (e.g., DeShazo & Fermo, 2002. Figure 8 plots McFadden's pseudo R-squared in relation to λ. It illustrates that RI choices are more deterministic at very low and very high values of λ, and less deterministic at intermediate values.

Figure 9 extends the illustration of RI choice as a function of λ to the case of three alternatives.[29] Alternative $a_i, i = 1, ..., 3$, yields utility $u_i = \beta_{b,i} - p_i + d_i$ with $\beta_{b,1} = 3.5$, $\beta_{b,2} = 3.25$, $\beta_{b,3} = 3$, $p_i = 4$, and d_i are independently distributed according to $\Pr(d_i = 0) = \Pr(d_i = 2) = 0.5$. Based on prior information, alternative a_1 is the best and a_2 is the second best. Figure 9 displays conditional choice probabilities for a choice set ω where alternative a_3 provides the highest payoff based on realized values of the complex discount attribute, illustrating how information costs λ impact the formation of consideration sets.

As λ increases, the number of alternatives chosen with strictly positive unconditional probability first decreases from three to two (at λ''), and eventually results in deterministic choice of a_1 based on prior considerations only (to the right of λ'''). As the information costs increase, the costs of resolving uncertainty about a priori less attractive alternatives outweigh the (expected) benefits. As a consequence, it becomes optimal to ignore such alternatives even if there exist choice sets ω in which the ignored

29 For ease of exposition and without loss of generality, there is no outside option in this example.

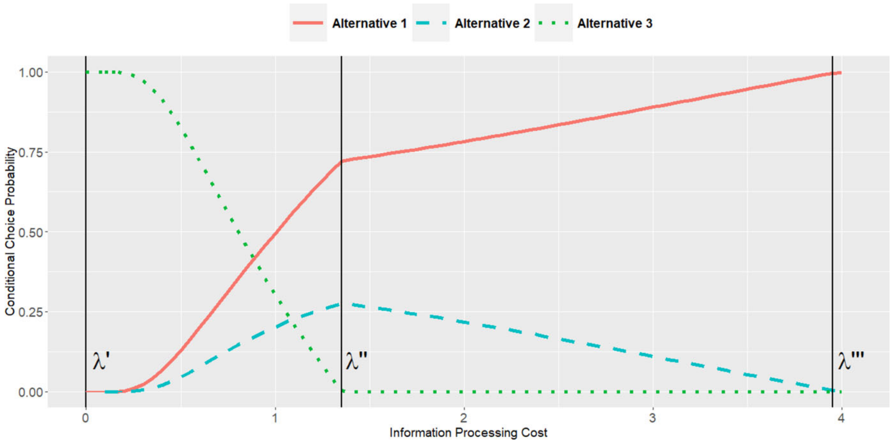

Fig. 9 Impact of information costs on consideration set size. Conditional choice probabilities for three alternatives are displayed as a function of information processing costs λ in the choice set where a_3 is the best alternative. As information costs increase, the DM rationally chooses to ignore alternatives in the choice set. λ', λ'', and λ''' indicate threshold values for which the consideration set size changes. For costs $\lambda' = 0$ choice is deterministic, and the best alternative a_3 is always chosen, however, after considering all alternatives. For costs larger than λ''', choice becomes deterministic again, however now because the DM ignores alternatives a_2 and a_3, regardless of realized complex discount levels

alternatives provide the highest payoff (as depicted in Fig. 9).[30] Together, the illustrations in this section suggest that RI provides a useful basis for bridging across choices under different incentive or difficulty levels.

4.2.2 Attribute range/dispersion and levels effects

Next, we show how the attribute range, typically measured as the difference between the highest and the lowest level of an attribute, or more generally, the dispersion of complex attributes, moderates the impact of a one-unit increase in that attribute on choice. The underlying mechanism is that as the range of the complex attribute increases, the expected gain from identifying its realized value also increases, making processing information more valuable. This ultimately increases the impact of the complex attribute on choice and in contrast to what one would expect when taking a RU perspective.

Figure 10 illustrates this mechanism in our leading car example. We set $\beta_b = 6$ and recall that $\beta_p = -\beta_d = -1$. Here, we study the impact of an increase in the discount from 2 to 3 on conditional choice probabilities for different discount ranges. Both lines in Fig. 10 depict how conditional choice probabilities change when the complex discount increases by one unit for different values of the simple price. The dashed blue line represents the case where the complex discount is drawn from the set $\{1, 2, 3, 4\}$, while in the second case (solid red line) the discount takes values in

30 This way, RI can motivate a positive probability of choosing an alternative that is dominated a posteriori, i.e., after processing (some) complex information (cf. Ruan et al., 2008 who propose a sequential sampling model to model dominance as a form of similarity).

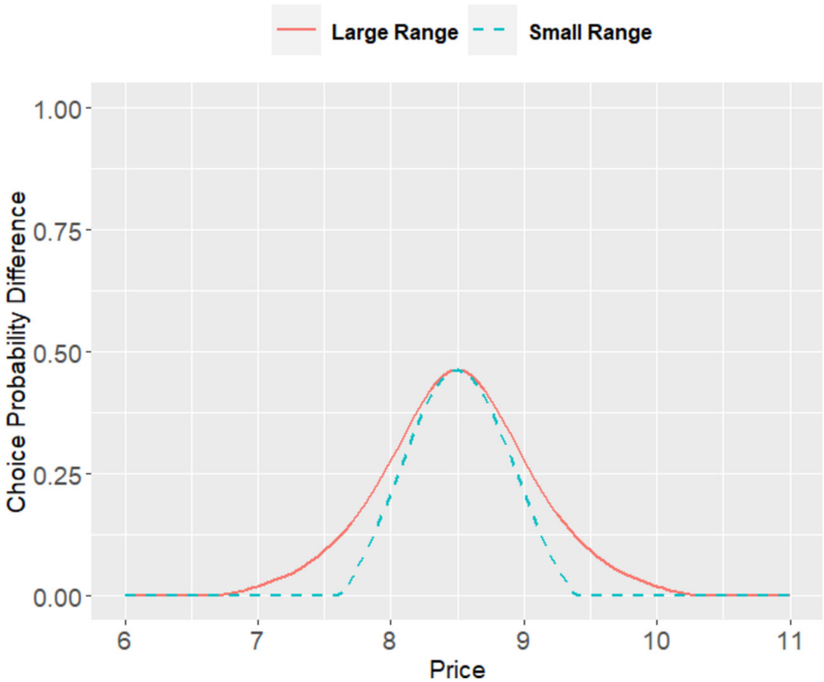

Fig. 10 **Range effects of complex attributes.** This Figure displays conditional choice probability differences of the inside alternative in reaction to an increase of the discount from 2 to 3 given a large range (solid red) and a small range (dashed blue) of the complex discount as a function of the simple price. Specifically, complex attribute levels are $\{1, 2, 3, 4\}$ when the range is small, and they are $\{0, 2, 3, 5\}$ when the range is large

$\{0, 2, 3, 5\}$. In both cases, the DM's beliefs are uniform over the respective support. Figure 10 shows that as the range of the complex discount attribute increases, the impact of a one-unit increase of the discount also increases. Technically, an increase in the attribute range spreads the range of possible payoffs from choosing the inside good further, motivating larger (costly) departures of conditional choice probabilities from their unconditional counterparts as a result of the optimal processing strategy.

To showcase what an analyst taking a RU perspective when analyzing RI choice data may find in this example, we fit logit models to simulated RI choices conditional on different ranges of the complex discount in the experimental design. We generate data sets as follows: $\beta_b = 5$, $\beta_p = -\beta_d = -1$, $p \in [8, 10]$, and d is drawn with equal probability from the binary set $\{4 - x, 4 + x\}$ with $x \in [0.25, 4]$. Figures 11 and 12 summarize the logit estimates as well as McFadden's pseudo R-squared values as a function of the range of the complex discount in a particular experimental design. When the range of the complex discount is small, the estimated brand and price coefficients are absolutely large, and the discount coefficient is, relatively, much smaller. As the range increases, brand and price coefficients become smaller in absolute value, and the inferred discount coefficients tend to increase.

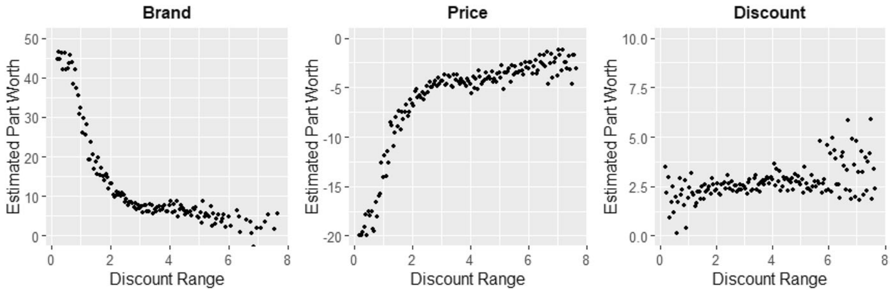

Fig. 11 **Logit estimates for different levels of the discount range.** This Figure displays estimated coefficients of brand, price, and discount for different ranges of the complex attribute. Each point is the result of fitting a logit model with simulated data from $T = 1,000$ choices with data generating parameters $\lambda = 0.5$, $\beta_b = 5$, $\beta_p = -\beta_d = -1$, and p being uniformly drawn from [8, 10]. The discount range, given as the difference between the two discount levels, is varied from 0.5 to 8

Figure 12 exhibits a U-shaped relationship between choice consistency and the range of the complex discount. When the range is very small, the DM makes less costly mistakes when choosing based on prior beliefs, conditioned on the simple attributes brand and price in this example. As a consequence, the DM pays little attention to the realized discount levels and chooses rather consistently based on simple attributes as well as the *expected* discount level only. As the range increases, the mistakes when choosing based on prior beliefs can become rather costly. However, there still are prices at which learning the realized complex discount in addition does not add much value. In the RU logit approximation, the relative importance of the discount increases, however, choice consistency decreases.

Finally, when the discount range becomes so large that fewer and fewer simple prices within the support of the design translate into a good enough choice (in expectation) without knowing the realized discount level, the DM will process the complex discount consistently. This again results in less stochastic data.

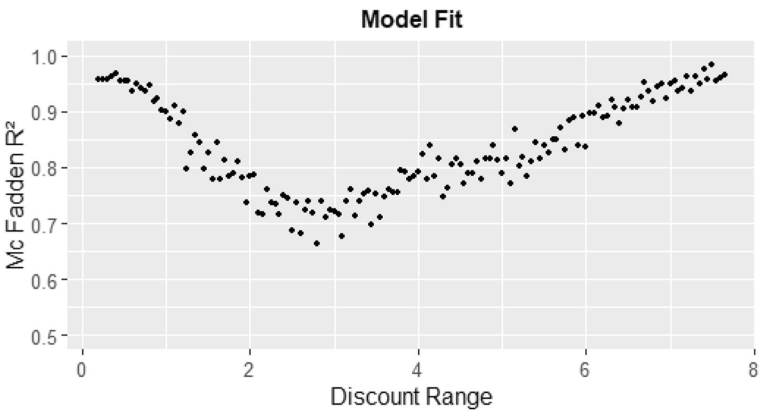

Fig. 12 **Non-monotonic effect of the discount range on choice consistency.** Choice consistency is measured as McFadden's pseudo R-squared. For details, see Domencich and McFadden (1975)

Attribute levels effects Finally, we note that the RI-DCM can explain both positive and negative effects of increasing the number of attribute levels on choice probabilities, as documented in, e.g., Liu et al. (2009) and Hensher (2006), in addition to range effects. The intuition is that "in-between" levels of ordered attributes may increase or decrease the (conditional on realized simple attributes) ex ante uncertainty about relative payoff advantages motivating more or less processing of complex attributes. This intuition applies equally to adding in-between levels to simple and complex attributes.

4.2.3 Price image and attribute correlation

Price images can be modeled as brand-specific prior beliefs about complex price components under RI. Similarly, the expectation that higher simple prices are associated with better (complex) quality aspects (e.g., Erickson & Johansson, 1985) can be modeled via prior beliefs. Here, we provide two examples. First, we vary the correlation between the inside alternative's brand and the discount d. Second, we investigate the case of two complex attributes and vary the correlation between these. Both variations affect the DM's prior beliefs.

In our first example, the DM chooses between an outside good and an inside good that yields $u_I = 1 - p + d$ with $p = 2$ and d being distributed over the binary set $\{0, 2\}$. As a function of the correlation θ between the discount and the brand of the inside good, the DM's prior beliefs become $\Pr(d = 2) = (1+\theta)/2$. Thus, a larger correlation coefficient θ increases the prior probability of a positive discount, and hence the ex ante valuation of the inside good. At the same time, absolutely larger correlations reduce the variance in realized discounts levels and consequently the DM's uncertainty about the payoff from the inside good.

Figure 13 depicts conditional choice probabilities implied by RI as a function of θ. The choice probability of the inside good increases in the correlation parameter θ.

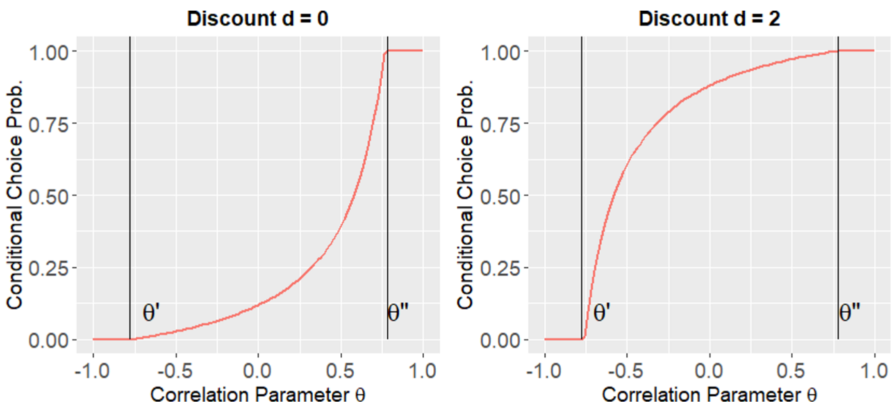

Fig. 13 Impact of brand-specific discount beliefs on choice. Conditional choice probabilities are displayed as a function of the correlation coefficient θ for choice sets with discount levels $d = 0$ (left) and $d = 2$ (right). As θ varies, the prior belief structure of the DM changes, which directly affects optimal choice probabilities. Note that in both choice sets the DM chooses the inside (outside) good deterministically once the correlation θ becomes sufficiently large (small) even while it is still interior

Notably, absolutely large correlations between the complex discount and the inside brand translate into deterministic choice behavior that ignores the actual choice set ω (at $\theta \leq \theta'$ or $\theta \geq \theta''$), well before the choice set becomes deterministic itself (at $\theta = -1$ or $\theta = 1$). Specifically, when θ is small (large) enough, i.e., the prior likelihood of the discount is small (large) enough, the DM will rationally decide not to costly learn the actually realized discount level, but always choose the outside (inside) good.

In our second example, the inside good has two correlated complex discounts d_1 and d_2 that can be learned separately by the DM. The payoff from the inside good is now given by $u_I = 1 - p + (d_1 + d_2)$ with $p = 2.75$ and $d_k \in \{0, 2\}$. The correlation between discounts d_1 and d_2 is θ so that the resulting prior beliefs are

$$\Pr(d_1 = 0, d_2 = 0) = \Pr(d_1 = 1, d_2 = 1) = \frac{1}{4}(1 + \theta) \text{ and}$$

$$\Pr(d_1 = 1, d_2 = 0) = \Pr(d_1 = 0, d_2 = 1) = \frac{1}{4}(1 - \theta).$$

Figure 14 displays the conditional probability of choosing the inside good as a function of the correlation between d_1 and d_2 in all four possible choice sets ω. Overall, as the correlation θ increases, the DM is less likely to choose the inside option. There exist beliefs where the decision maker always chooses deterministically. Specifically, for smaller values of θ, the prior probability of one of the discounts applying increases such that eventually, the inside good will always be chosen (one discount is sufficient to make the inside good utility maximizing given the parameters from above). As θ increases to larger positive values, the discounts are more likely to apply jointly or not at all, so it becomes rational to guard against the loss from choosing inside without any discount. Costly finding that any one of the two discounts is zero, i.e., does not apply, is predictive of no discount applying under this prior. Hence, the DM becomes

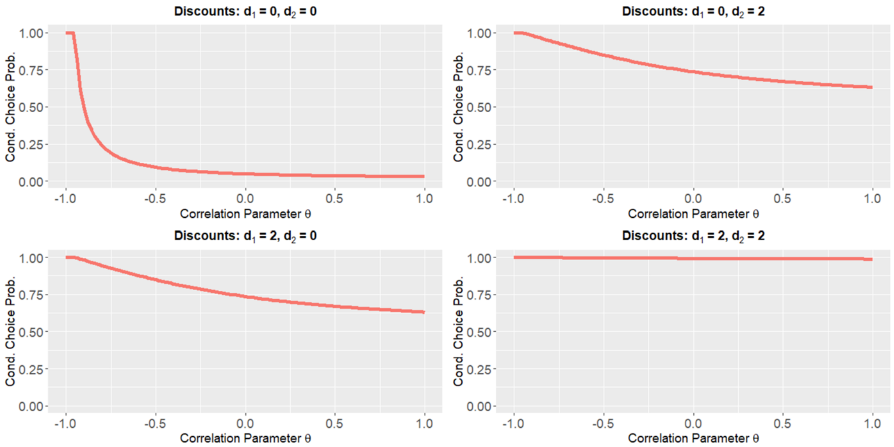

Fig. 14 Impact of correlations between complex discounts on choice. Conditional choice probabilities of the inside good with different complex discount levels are displayed as a function of the correlation parameter θ. The panels differ in the realized values of the discounts d_1 and d_2. Note that in all choice sets the DM chooses the inside good deterministically once the correlations become sufficiently low while still being interior

Table 13 Logit approximations for different choice set sizes

Choice Set Size	Brand 1	Brand 2	Price	Discount	McFadden R^2
2	27.92		-7.71	2.32	0.67
	(2.75)		(0.76)	(0.28)	
3	16.60	16.55	-3.83	1.85	0.60
	(1.12)	(1.12)	(0.26)	(0.13)	

Columns two to four show logit estimates for the corresponding attributes, with standard errors in parentheses for choice sets with two (first row) and three (third row) alternatives. The final column reports corresponding McFadden R-squared (Domencich & McFadden, 1975) as a measure for choice consistency. Note that the added alternative is identical to the already existing inside alternative

less likely to choose the utility-maximizing inside option when only one discount applies under this prior. Our two examples are simple for expositional reasons, but they suggest how the RI framework can account for belief dependencies both within and across choice alternatives. We revisit this point in Section 5.

4.2.4 Choice Set Expansions

Standard RUMs assume that the utility index, as well as the error term variance, are independent of the choice set size. However, there is empirical evidence contradicting both assumptions, e.g., DeShazo and Fermo (2002) or Meißner et al. (2020). An increasing error term variance has been interpreted as the result of a latent cognitive process that captures a DM's struggle with increasing complexity due to a higher information load.

We show that when RI choice data is analyzed from the perspective of a RU logit model, choice behavior appears to become less consistent as the choice set expands in the number of choice alternatives. The reason for this is that as the simple information varies from choice set to choice set, the impact of the complex attributes also varies. As the choice set grows, under fixed (conditional) distributions of complex attributes, the number of implied interactions between simple and complex attributes increases. A standard RUM does not account for such interactions.[31] Therefore, the RU error term variance increases as the number of alternatives in the choice set grows.

Consider the following illustrative simulation. There are choice sets with an outside alternative a_0 that yields a payoff of zero and inside alternatives a_i, $i = 1, 2$ with a payoff of $u_i = 3 - p_i + d_i$. We simulate choice data from a DM who faces choice sets with either one inside alternative, $A_1 = \{a_0, a_1\}$, or two inside alternatives, $A_2 = \{a_0, a_1, a_2\}$. For any choice task, prices are drawn uniformly from the interval $[2, 4]$ and discounts are independently distributed with $\Pr(d_i = 0) = \Pr(d_i = 2) = 0.5$. Then, we fit descriptive logit models to the two data sets to illustrate what a researcher bringing an RU perspective to these data may find. Table 13 summarizes the logit fits.

[31] Even though we use the term interaction here, there is no way of knowing what finite dimensional logit index will reasonably approximate RI choice behavior locally. Obviously, even a reasonable local approximation will be insufficient to answer counterfactual queries that involve DMs' beliefs.

There are two noteworthy observations from this analysis. First, as the choice set expands from A_1 to A_2, all estimated coefficients decrease in absolute value. The intuition is, as discussed above, that in a larger choice set, there are more implied interactions in the data-generating model. In the small choice set, the impact of the complex discount of alternative a_1 is very small for extreme values of the simple price p_1. For instance, when the price p_1 is very small, the DM will always choose a_1 regardless of the discount. However, with a second inside good, the price of the second alternative p_2 will also determine the impact of d_1 on choice. The logit approximation, however, ignores these interactions, and thus the error term variance increases, which is reflected in the shrinking coefficient estimates.

Second, observe that the decrease is stronger in the coefficient of the simple price than in the coefficient of the complex attribute. The reason for this is that as the number of alternatives increases, choice sets in which the complex discount attribute can be rationally ignored for all alternatives become less likely.

Lastly, note that while the logit approximation results in a larger error term variance as a consequence of a choice set expansion, this does not imply that the DM becomes worse-off. In contrast, in expectation the DM under RI benefits from any choice set expansion as long as the DM's prior beliefs coincide with the objective distribution over choice sets. Still, it may be the case that for one particular choice set, characterized by specific attribute realizations, an expansion may lead to a worse welfare outcome.[32] Intuitively, the decision maker can choose to optimally ignore alternatives in a choice set without any additional costs.[33] As such, the RI framework allows the analyst to take a stance on the impact of choice set expansion on welfare, as it relates the observed error to optimal decision-making under cognitive constraints.

5 Conclusion and outlook

In this paper, we have reviewed two strategies for bringing discrete choice under RI to multi-attribute, multi-alternative settings typical of marketing applications. The first strategy relies on specific assumptions about prior beliefs that yield choice probabilities that have a closed MNL form. While this approach allows estimation with established methods, it cannot reproduce a variety of key qualitative features. Therefore, we characterize a RI-DCM that can incorporate general beliefs. We illustrate that a hierarchical version of this model can be calibrated from "small T, large N" data, as typical of marketing applications. Our illustration suggests that such data generated from the proposed RI-DCM will reliably distinguish the data-generating model from a standard hierarchical logit model.

Moreover, we have shown how a series of behavioral patterns—each individually requiring qualitatively different modifications of the standard RU logit—can be

[32] For instance, an added alternative may dominate existing alternatives in most choice sets so that it is much better in expectation. Then it becomes optimal for a DM not to process information about some alternatives in its presence. However, in general, there will still be specific choice sets where some of these alternatives provide a higher payoff than the newly added. In such a choice set, the DM will be worse off (in terms of realized payoffs) after a choice set expansion.

[33] Implicitly, the RI model assumes free disposal (or ignorance) of available information.

jointly nested within a discrete choice model rooted in RI. Consequently, the RI-DCM with general prior beliefs may have the potential to become a new workhorse model for applied researchers. We conclude this article by outlining the open challenges in achieving this as well as promising avenues for future research.

5.1 Open empirical challenges: large state spaces

The RI-DCM with general beliefs for MAMA choice is consistent with the original RI idea. It motivates choice stochasticity from imperfect information processing only. We view this as a decisive advantage over assuming additively separable utility components only observed by the DM, particularly in the context of DCEs where the researcher fully controls the (maximum) information set available to DMs.

Similar to the conditional choice probabilities in Matějka and McKay (2015)'s formulation, the likelihood in the RI-DCM with general beliefs does not have a closed-form solution. We rely on the Blahut-Arimoto algorithm for numerical solutions to the RI choice problem. This algorithm solves a fixed point problem, which makes the estimation procedure computationally costly for datasets with large numbers of respondents or repeated choices per respondent. We discussed how we speed up the computation of the likelihood given by solutions to the RI choice problem in Section 3.2.

However, the number of possible states grows exponentially in the dimensionality of the complex information in a choice set. In high-dimensional problems, solving the RI choice problem using the Blahut-Arimoto algorithm is no longer practically feasible. The theoretical and computational challenges posed by this problem are an active field of research (see e.g., Li et al., 2006; Sutton & Barto, 2018; Gershman & Lai, 2020; Miao & Xing, 2023; Armenter et al., 2024).

5.2 Unobserved attributes – endogeneity

RI motivates stochastic choice behavior from imperfect information processing. Hence, RI rationalizes stochastic choices in environments controlled by the researcher, such as in DCEs without invoking exogenous error terms. In settings outside the lab, e.g., in scanner panel data, it may well be that the researcher has to rely on an incomplete description of the choice environment facing the DM. Hence, there may be two sources of stochasticity in the data: i) incomplete processing of information known to the analyst by the DM and ii) (non-constant) aspects of the choice environment unobserved by the analyst. Conceptually, the latter motivates additional error terms in the model that may represent simple or complex information (to the DM) unobserved by the analyst. Furthermore, these additional error terms will likely be endogenous to observed conditioning arguments in observational data. Methods to account for endogeneity (from the supply side) in models of adaptive information processing by the DM will have to be developed.[34]

[34] In the context of consumer search, Moraga-González et al. (2023) propose a method for dealing with endogeneity based on the generalized method of moments (Nevo, 2000). In the context of discrete choice,

5.3 Use of auxiliary data

RI is usefully agnostic about the exact process by which DMs process complex information. However, within the framework proposed in this paper, RI makes predictions that could be corroborated using so-called process measures (that, of course, pertain to observable aspects of decision making). This is similar to how process measures are used to test, inform, and calibrate cognitive models of choice in psychology (e.g., Krajbich et al., 2010).

As per the proposed model, simple attributes should have priority in processing. Consequently, measuring eye fixations (and their order) could support the role of simple attributes in forming conditional priors or even help identify what is, in fact, simple about a particular choice task in a particular design. Along the same lines, eye traces could corroborate the existence of consideration sets implied by unconditional choice probabilities. Complex attributes of alternatives outside the consideration set are expected to get no eye fixations. Finally, one could consider eliciting responses to questions like "Which alternatives did you reject based on a first inspection, if any?" or "Which alternatives did you inspect more closely before making a choice?" in every choice set to provide a set of fallible consideration indicators as part of a DCE.

Moreover, RI implies that the amount of processing of complex information can be measured by the amount of entropy resolved, the mutual information in a particular choice set. If processing more complex information takes more time, the model predicts longer (shorter) decision times in choice sets where conditional choice probabilities are further away (closer) to their unconditional counterparts. It follows that choice sets that give rise to smaller endogenous consideration sets should require less time than choice sets with larger such sets, on average.

5.4 Consumer belief elicitation and its role in market simulations

Prior beliefs are important for information processing and subsequent choices of DMs under RI. We illustrated this feature previously by showing how changes in beliefs can even result in choice reversals while keeping every other aspect of the choice task fixed. A growing theoretical literature, in particular from industrial organization, demonstrates the significance of prior beliefs for consumer choices and market outcomes under RI. This includes Matějka and McKay (2012), who study the effects of differences in prior beliefs on market equilibria; Boyacı and Akçay (2017), who focus on implications for the optimal pricing of monopolistic firms; and Janssen and Kasinger (2024) who examine the role of consumers' prior beliefs for equilibrium pricing and obfuscation behavior of profit-maximizing firms in a duopoly. Thus, counterfactual analyses under RI require an understanding of beliefs held by the DMs whose behavior is modeled.

In DCEs, researchers have some control over prior beliefs as conveyed to DMs by the experimental design. However, for counterfactuals, e.g., market simulations, a

Petrin and Train (2010) describe a control function approach. However, their method is not immediately applicable, as they assume that the utility index contains an additively separable error term; and the RI-DCM with general beliefs will in general have an implied error term that depends on the simple attributes.

key question is thus how to assess and simulate (likely) beliefs in the market setting. Choosing belief distributions based on tractability in a closed-form logit framework with additively separable indices, as currently standard in empirical applications (see Section 3.1), does not seem to be a satisfactory solution.

In light of a growing empirical RI literature that relies on belief assumptions motivated by analytical convenience, we feel that research into the sensitivity with respect to different assumptions about beliefs will be useful, as well as an integration of methods to study potentially heterogeneous market beliefs empirically. The characterization of an RI-DCM that allows for general prior beliefs in this paper paves the way for this line of research. Other frameworks modeling limited information, such as consumer search models or learning models, face a similar challenge of dealing with typically unobservable (consumer) beliefs as a key building block to empirical analysis and counterfactual computations and are often quite sensitive with respect to assumptions about beliefs (Chintagunta & Nair, 2011).

Notably, some of these contributions have explored the added benefit of eliciting beliefs through auxiliary information rather than relying on purely theoretical assumptions, such as rational expectations. Successful approaches that have been employed in other areas to learn about beliefs use survey-based belief elicitation methods (e.g., Cavallo et al., 2017; Coibion et al., 2018; Armona et al., 2019) or observational data such as clickstream data (e.g., Hu et al., 2019), or eye-tracking data (e.g., Ursu et al., 2024). For a recent overview in marketing, see the literature section in Jindal and Aribarg (2021).

5.5 Alternative information cost functions

There is an ongoing discussion in the economics literature on the question of which attention cost function is appropriate for which choice circumstances. Most of the experimental studies, either explicitly or implicitly, estimate the cost function, which is then used in subsequent analyses. Consequently, much of the experimental literature dealing with RI (implicitly) tests the applicability of different cost functions in a variety of stylized settings. As in the present paper, many results in RI theory have been derived under the assumption of costs that are linear in Shannon mutual information.[35]

For instance, experimental evidence suggests that non-linearities of information costs exist, where subjects pay too little attention to high rewards compared to low rewards (Caplin & Dean, 2013; Dean & Neligh, 2019). Perhaps more important for analyzing MAMA choice, a linear Shannon entropy cost function precludes the concept of perceptual distance, where some states are harder to distinguish than others (Dean & Neligh, 2019; Hébert & Woodford, 2021). As a result of this critique, some papers generalize the linear Shannon cost function or apply different cost functions. For instance, Hébert and Woodford (2021) introduce a neighborhood cost function to

[35] Maćkowiak et al. (2023) discuss what features of RI are robust with respect to the functional form of information processing costs.

address the problem that some states are harder to distinguish than others.[36] A detailed discussion of this mostly theoretical literature is beyond the scope of this paper.[37]

We believe that the choice of cost function could become important for the applications envisioned in this paper as well. For instance, the model formulation based on Shannon costs implies that any two possible choice sets (states ω) within an experimental design are equally hard to distinguish. However, it is not unlikely that the distinction between two choice sets that pose many complex trade-offs may be harder than that between two choice sets that pose fewer trade-offs each. We thus conjecture that neighborhood-based costs that allow to impose that certain sets of choice sets are harder to distinguish than others may eventually become useful. We note that this is related to a finer distinction of levels of processing difficulty than the distinction between simple and complex attributes proposed in this paper.

5.6 Conceptualization of free information

The history of research on choice in marketing, economics, and psychology is rich in empirical and theoretical results about what may be simple and more complex about a particular choice task or set of such tasks. For example, it may be worthwhile to reconsider the literature on heuristics in choice as information (for the analyst) about simple aspects of choice tasks that, in the context of a RI-DCM, may give rise to prior beliefs that guide the amount of processing of more complex aspects of a choice task.[38]

In our suggested specification of the RI-DCM, some attributes are considered simple so that processing them does not require (significant) cognitive effort. Thus, it is these attributes $\mathbf{x}_s$ that determine (conditional) prior beliefs μ_s, which then form a key ingredient to how much processing of complex attributes should occur (see Section 2). This formulation imposes a specific mapping from choice sets into prior beliefs that is not given by RI theory but assumed by the analyst (even if, as we showed, an empirical distinction between simple and complex attributes is possible, in general).

In principle, any aspect of a choice set and the attribute configuration presented therein could be simple information and thus determine prior beliefs. Consider, for example, a DCE design where the discount has values $d \in \{50, ..., 90, 100\}$. A DM may easily recognize that the discount equals 100 due to the substantially different visual stimulus (two vs. three digits). In this case, the DM may assign a positive belief to all choice sets where $d = 100$ and zero to all other.

This raises the conceptual question of which pieces of information in any choice task can be considered free and, consequently, how to map choice sets into (conditional) prior beliefs. Our discussion regarding the empirical identification of simple vs. complex attributes in Section 3.5 should be viewed as a special case of this broader

[36] A related generalization of Shannon costs that allows for alternatives to have different information costs is suggested by Huettner et al. (2019).

[37] For an overview of different information cost functions applied in the behavioral inattention literature, see Gabaix (2019); for a more detailed discussion on entropy-based cost functions, see Dean and Neligh (2019) as well as Maćkowiak et al. (2023).

[38] See also Maćkowiak et al. (2023) who propose the idea that RI provides a model for the formation of heuristics.

consideration. Due to the high dimensionality of this question, it will typically not be viable to give purely empirical answers, and an appeal to theory and previous empirical results is required.

While this is beyond the scope of this paper, it suggests that the RI framework may be able to fruitfully integrate conjectured and empirically demonstrated choice simplification strategies with fully rational behavior that is guided by priors formed on the basis of whatever a DM may easily and immediately process about a choice task. With an eye towards industry-grade applications with many attributes and many alternatives, this is an important part of future research.

A Appendix

A.1 Identification of $\Delta\lambda$ across decision environments under stable preferences

Next, we present estimation results to show that a relative change in information processing costs under stable preferences can be identified (see Table 14). We simulate 2,000 choices for one individual, with one inside and one outside option. The first half of the 2,000 choices are made in an "easy" environment with lower information processing costs λ_{low}, and the second half in a more difficult environment subject to higher information processing costs λ_{high}. The inside good consists of three attributes: two simple and one complex. One simple attribute is a brand intercept, and the remaining attributes have the following levels: simple price $p \in \{2.5, 3, 3.5, 4, 4.5\}$ and complex discount $d \in \{0, 0.5, 1, 1.5, 2\}$. The preference vector is given by $\boldsymbol{\beta} = (\beta_b, \beta_p, \beta_d) = (2.5, -1, 1)$ and information processing costs are $\lambda_{low} = 0.1$ and $\lambda_{high} = 0.3$. In estimation, we fix λ_{high} to the data generating value and jointly estimate λ_{low} and $\boldsymbol{\beta}$. The third row in Table 14 shows that we can recover the data-generating parameters, and the fifth row illustrates the bias from ignoring the difference in processing costs, as well as the poorer fit from doing so.

A.2 Simulation study

The simulation study builds on the set-up used for illustrative simulations in Section 3.3 of the paper. We vary the number of inside goods (one versus two), simulate

Table 14 Posterior means of preferences and information processing costs

Model	Brand	Price	Discount	λ_{low}	λ_{high}	LMD
Data generation	2.50	-1.00	$\equiv -\beta_p$	0.10	0.30	
RI-DCM	2.59	-1.03	$\equiv -\beta_p$	0.10	0.30	-253.38
	(0.10)	(0.04)		(0.01)		
RI-DCM with fixed λ	2.28	-0.91	$\equiv -\beta_p$	0.20	0.20	-298.98
	(0.11)	(0.04)				

In the first model, λ_{high} is fixed to data generating value in estimation. For the second model, the information processing costs are fixed to 0.2

Table 15 Means of posterior means and variances of preference distributions for different model specifications using 50 simulations with one inside and one outside good each and homogenous information processing costs

	Posterior Means			Model Fits	
Model	Brand	Price	Discount	LMD	Δ LMD
Data generation	2.50	-1.00	$\equiv -\beta_p$		
RI-DCM	2.51	-1.00	$\equiv -\beta_p$	-2,465.49	
	(0.02)	(0.01)		(63.88)	
RU logit separate	26.41	-8.71	4.41	-2,630.29	162.17
	(0.84)	(0.32)	(0.19)	(75.46)	(26.47)
RU logit joint	9.94	-4.01	$\equiv -\beta_p$	-4,602.05	2018.19
	(0.14)	(0.09)		(236.44)	(50.34)
BC separate	27.43	-9.08	4.51	-2,608.36	136.66
	(0.89)	(0.51)	(0.25)	(73.41)	(25.72)
BC joint	11.48	-4.13	$\equiv -\beta_p$	-3,160.25	685.37
	(0.99)	(0.16)		(82.37)	(49.87)
	Posterior Variances				
Model	Brand	Price	Discount		
Data generation	0.25	0.04	$\equiv \text{Var}(\beta_p)$		
RI-DCM	0.29	0.05	$\equiv \text{Var}(\beta_p)$		
	(0.04)	(0.01)			
RU logit separate	14.01	1.82	1.12		
	(0.84)	(0.32)	(0.19)		
RU logit joint	5.31	0.70	$\equiv \text{Var}(\beta_p)$		
	(1.01)	(0.30)			
BC separate	20.11	1.73	1.57		
	(1.06)	(0.57)	(0.45)		
BC joint	9.73	0.44	$\equiv \text{Var}(\beta_p)$		
	(0.83)	(0.20)			

Standard deviations are in parentheses. For the BC separate and BC joint model, the means of posterior means for the price threshold are 10.91 and 4.67 with standard deviations of 5.02 and 0.81, respectively. The means of the posterior variances for the price screening thresholds are 8.41 for the BC separate and 1.01 for the BC joint model

data with and without heterogeneity in processing costs (in addition to preferences heterogeneity), and add a two-stage choice model with consideration sets from screening on price (see e.g., Gilbride & Allenby, 2004; Pachali et al., 2023) to the model comparison. We use the following (standard) weakly informative subjective prior parameter settings for the one and the two-inside good cases, respectively: $\{\bar{\boldsymbol{\beta}} \sim \mathcal{N}(0, 100\mathbf{I}), \mathbf{V}_{\boldsymbol{\beta}} \sim IW(4, 1.5\mathbf{I})\}$ and $\{\bar{\boldsymbol{\beta}} \sim \mathcal{N}(0, 100\mathbf{I}), \mathbf{V}_{\boldsymbol{\beta}} \sim IW(5, 2\mathbf{I})\}$. The slightly more informative subjective prior for the hierarchical prior variance of $\{\boldsymbol{\beta}_i\}$, $\mathbf{V}_{\boldsymbol{\beta}}$ is owed to the increased dimensionality of the estimation problem. When we estimate the hierarchical prior variance of $\{\log(\lambda_i)\}$, we use $\sigma^2_{\log(\lambda)} \sim IG(3, 1)$.

We simulate 50 data sets in the four data-generating settings and estimate five models for each: the RI-DCM, two RU logit models, and two models with price-based consideration sets (BC separate and BC joint). The benchmark models are estimated once with a single price parameter and once with separate parameters for the (simple) price and the (complex) discount. Each data replication in this simulation study resamples $N \times T$ choice sets from the design base defined by prices p and discounts d, drawn uniformly and independently from the following sets: $p \in \{2.5, 3, 3.5, 4, 4.5\}$ and $d \in \{0, 0.5, 1, 1.5, 2\}$ and re-samples preferences (and processing costs when

Table 16 Means of posterior means and variances of preference distributions for different model specifications over 50 simulations with two inside goods and one outside good each and homogenous information processing costs

	Posterior Means				Model Fits	
Model	Brand 1	Brand 2	Price	Discount	LMD	Δ LMD
Data generation	2.50	2.50	-1.00	$\equiv -\beta_p$		
RI-DCM	2.54	2.53	-1.03	$\equiv -\beta_p$	-3,807.55	
	(0.05)	(0.05)	(0.03)		(56.43)	
RU logit separate	24.89	24.55	-8.06	4.33	-4,160.46	272.48
	(0.76)	(0.69)	(0.31)	(0.19)	(102.34)	(54.03)
RU logit joint	10.01	9.97	-3.99	$\equiv -\beta_p$	-6,838.41	2940.38
	(0.25)	(0.24)	(0.05)		(107.61)	(60.37)
BC separate	25.19	24.91	-8.41	4.44	-4,196.47	228.54
	(0.69)	(0.62)	(0.33)	(0.10)	(94.31)	(48.01)
BC joint	11.15	11.67	-4.01	$\equiv -\beta_p$	-5,431.81	1549.33
	(0.29)	(0.37)	(0.09)		(81.69)	(50.67)
	Posterior Variances					
Model	Brand 1	Brand 2	Price	Discount		
Data generation	0.25	0.25	0.04	$\equiv \text{Var}(\beta_p)$		
RI-DCM	0.28	0.30	0.05	$\equiv \text{Var}(\beta_p)$		
	(0.05)	(0.06)	(0.01)			
RU logit separate	12.01	11.19	1.97	1.22		
	(0.84)	(0.72)	(0.21)	(0.11)		
RU logit joint	6.31	6.23	0.60	$\equiv \text{Var}(\beta_p)$		
	(0.41)	(0.44)	(0.17)			
BC separate	17.41	16.33	1.70	1.27		
	(3.16)	(3.55)	(0.39)	(0.21)		
BC joint	8.73	8.56	1.04	$\equiv \text{Var}(\beta_p)$		
	(0.83)	(0.81)	(0.20)			

Standard deviations are in parentheses. For the BC separate and BC joint model, the means of posterior means for the price threshold are 12.01 and 4.26 with standard deviations 5.04 and 0.71, respectively. The means of the posterior variances for the price screening thresholds are 10.39 for the BC separate and 0.91 for the BC joint model

Table 17 Means of posterior means and variances of preference distributions for different model specifications over 50 simulations with one inside and one outside good each and heterogeneous information processing costs

	Posterior Means			Model Fits	
Model	Brand	Price	Discount	LMD	Δ LMD
Data generation	2.50	-1.00	$\equiv -\beta_p$		
RI-DCM scale mixture	2.51	-1.01	$\equiv -\beta_p$	-2,444.54	
	(0.04)	(0.03)		(53.41)	
RU logit separate	26.97	-8.92	4.43	-2,616.04	185.58
	(1.51)	(0.47)	(0.10)	(62.66)	(17.39)
RU logit joint	10.07	-4.02	$\equiv -\beta_p$	-4,530.19	2083.55
	(0.69)	(0.21)		(95.38)	(61.37)
BC separate	27.50	-8.84	4.23	-2,603.48	156.47
	(2.31)	(1.04)	(0.68)	(74.31)	(21.71)
BC joint	11.31	-4.40	$\equiv -\beta_p$	-3,171.05	663.49
	(1.62)	(0.60)		(67.46)	(82.94)
	Posterior Variances				
Model	Brand	Price	Discount	$\sigma^2_{\log(\lambda)}$	
Data generation	0.25	0.04	$\equiv \text{Var}(\beta_p)$	0.09	
RI-DCM scale mixture	0.30	0.05	$\equiv \text{Var}(\beta_p)$	0.08	
	(0.05)	(0.01)		(0.02)	
RU logit separate	15.11	1.62	1.73		
	(0.94)	(0.32)	(0.29)		
RU logit joint	6.21	0.73	$\equiv \text{Var}(\beta_p)$		
	(1.54)	(0.14)			
BC separate	14.41	1.84	1.90		
	(2.13)	(0.41)	(0.45)		
BC joint	10.15	0.39	$\equiv \text{Var}(\beta_p)$		
	(1.23)	(0.13)			

Standard deviations are in parentheses. For the BC separate and BC joint model, the means of posterior means for the price threshold are 10.38 and 4.13 with standard deviations 5.26 and 0.80 respectively. The means of the posterior variances for the price screening thresholds are 13.94 for the BC separate and 0.89 for the BC joint model. $\sigma^2_{\log(\lambda)}$ is the variance of log information processing costs in the population

heterogeneous) from population parameters. In each replication, a (fresh) sample of rationally inattentive DMs ($N = 1,000$) face $T = 20$ choices sets.

The utility of inside good i to DM j in choice task t is given by $u_{j,i,t} = \beta_{b,j,i} + \beta_{p,j}(p_{j,i,t} - d_{j,i,t})$ where $\beta_{b,j,i}$ is the brand coefficient, $\beta_{p,j}$ the price coefficient, $p_{j,i,t}$ is the price, and $d_{j,i,t}$ is the discount. The utility of the outside option is normalized to zero: $u_O = 0$. As in the illustrative simulations in Section 3.3, brand and price are simple attributes, while the discount requires costly processing. When processing costs are homogenous, they are set to $\lambda = 0.25$. When heterogeneous, they are generated from $\log(\lambda_j) \sim \mathcal{N}(-1.4, 0.09)$. Preference coefficients are independently generated

Table 18 Means of posterior means of preference distributions for different model specifications over 50 simulations with two inside goods and one outside good each and heterogeneous information processing costs

	Posterior Means				Model Fits	
Model	Brand 1	Brand 2	Price	Discount	LMD	ΔLMD
Data generation	2.50	2.50	-1.00	$\equiv -\beta_p$		
RI-DCM scale mixture	2.52	2.52	-1.02	$\equiv -\beta_p$	-3,800.39	
	(0.04)	(0.04)	(0.02)		(56.51)	
RU logit separate	27.33	27.19	-7.34	4.71	-4,172.86	363.87
	(4.56)	(5.31)	(0.97)	(0.53)	(75.37)	(44.63)
RU logit joint	10.94	10.88	-5.63	$\equiv -\beta_p$	-6,873.34	3,014.79
	(1.34)	(1.57)	(0.81)		(79.31)	(84.28)
BC separate	26.39	26.01	-8.11	5.01	-4,114.34	305.67
	(5.11)	(4.89)	(1.02)	(0.64)	(69.37)	(50.31)
BC joint	10.70	10.17	-5.21	$\equiv -\beta_p$	-5,404.29	1,603.48
	(1.31)	(1.29)	(0.93)		(79.14)	(85.34)
	Posterior Variances					
Model	Brand 1	Brand 2	Price	Discount	$\sigma^2_{\log\lambda}$	
Data generation	0.25	0.25	0.04	$\equiv \text{Var}(\beta_p)$	0.09	
RI-DCM scale mixture	0.29	0.28	0.05	$\equiv \text{Var}(\beta_p)$	0.09	
	(0.06)	(0.06)	(0.01)		(0.01)	
RU logit separate	13.11	9.59	1.62	1.17		
	(0.83)	(0.68)	(0.25)	(0.12)		
RU logit joint	6.95	6.77	0.68	$\equiv \text{Var}(\beta_p)$		
	(0.48)	(0.48)	(0.19)			
BC separate	16.44	15.98	1.65	1.31		
	(3.50)	(3.31)	(0.40)	(0.25)		
BC joint	9.02	8.74	0.64	$\equiv \text{Var}(\beta_p)$		
	(0.88)	(0.80)	(0.23)			

Standard deviations are in parentheses. For the BC separate and BC joint model, the means of posterior means for the price threshold are 13.31 and 4.43 with standard deviations 6.84 and 0.83 respectively. The means of the posterior variances for the price screening thresholds are 16.73 for the BC separate and 0.96 for the BC joint model. $\sigma^2_{\log(\lambda)}$ is the variance of log information processing costs in the population

from the following distributions: $\beta_{b,j,i} \sim \mathcal{N}(2.5, 0.25)$ and $\beta_{p,j} \sim \mathcal{N}(-1, 0.04)$, i.e., the two brands are symmetric in the population in the case of two inside brands.

Tables 15, 16, 17, and 18 report distributions of preference estimates and model fit across data replications in our four data generating settings for each of the five models fit. The last column in the upper tables, ΔLMD, shows log-marginal density differences relative to the RI-DCM across simulations. Positive values indicate that the RI-DCM achieves better model fit.

Not surprisingly, we find that only the RI-DCM recovers data-generating parameters. We also find that (i) we can reliably distinguish the data generating RI-DCMs from all the benchmark models (see columns LMD and ΔLMD in the respective tables), (ii)

Table 19 Quartiles as well as the minimum and the maximum of the log-likelihood MCMC draws for different λ levels for the correct and wrong specification respectively

Model	min	25%	50%	75%	max
$\lambda = 0.01$					
Correct - Linear	-1.42	0.00	0.00	0.00	0.00
Misspecified - Linear	-1.86	0.00	0.00	0.00	0.00
$\lambda = 5$					
Correct - Linear	-4.85	-3.83×10^{-7}	-9.82×10^{-9}	-1.71×10^{-10}	0.00
Wrong - Linear	-1.45	-3.73×10^{-7}	-1.07×10^{-8}	-2.36×10^{-10}	0.00

the approximating benchmark models fare relatively much worse in the larger choice set because of the corresponding increase in the number of optimal RI information strategies[39] (comparing ΔLMD in Tables 15 and 17 to that in Tables 16 and 18, respectively), and (iii), slightly worse when processing costs are heterogeneous, in addition to preferences heterogeneity (comparing ΔLMD in Tables 15 and 16 to that in Table 17 and 18, respectively).

Finally, we can see that benchmark models substantially benefit from including separate coefficients for price and discount and that modeling screening based on price further improves the fit of benchmark models. However, even the combination of separate coefficients for price and discount with price screening fits reliably worse than the RI-DCM. In the case of one inside good only, what is missing from this approximation is the complex interaction between price and discount implied by optimal processing under RI. In the case of two inside goods, consideration of one inside good also depends on simple features of the other inside good. Similarly, the processed contribution of one brand's discount depends on the simple attributes of this brand and that of the other brand.

A.3 Identification in the case of extreme information processing costs

Distinction of simple and complex utility aspects of a choice task Table 19 illustrates that the distinction between simple and complex attributes becomes mute once the data become (essentially) deterministic at very low or very high processing costs λ. We again simulate 2,000 choice tasks with the same design as outlined in Section 3.5 but now with information processing costs $\lambda = 0.01$, making processing complex information essentially free, and $\lambda = 5$, making information processing infeasible. When λ is sufficiently small, all information is fully processed, and the conceptual and

[39] Recall from expression Eq. 2 that optimal information strategies are conditional on realized levels of simple attributes. With one inside brand and five levels of simple price, there are five different optimal processing strategies, depending on the realized simple price in a choice set. With two inside brands and five levels of simple price, there already are $5^2 = 25$ different configurations of simple attributes, giving rise to different optimal processing strategies.

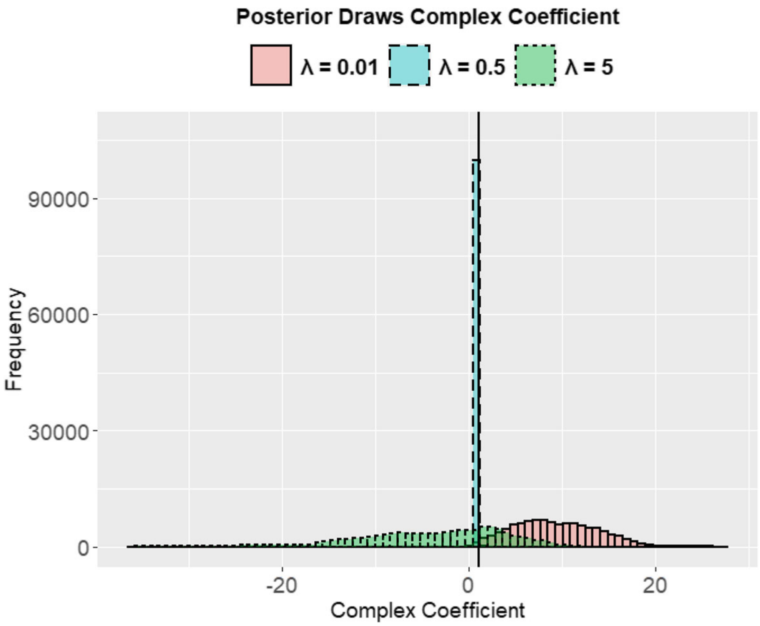

Fig. 15 Histogram of MCMC draws of the complex attribute parameter under different information processing costs λ. This Figure shows histograms of the posterior distribution of complex attribute coefficient based on 100,000 draws from the implied marginal posterior for correctly specified simple and complex attributes. The solid vertical line indicates the data-generating value

empirical distinction between simple and complex vanishes. When λ is sufficiently large, the information in complex attributes is not integrated into the overall evaluation of alternatives. If, in this case, the analyst falsely specifies complex attributes as simple and simple attributes as complex, the estimator will infer extreme utility parameters for the latter relative to the former such that deterministic choice based on simple attributes (falsely assumed to be complex) ensues.

Set identification of preferences Figure 15 illustrates that utility coefficients are only set-identified once choices become (essentially) deterministic, using the example of the complex attribute coefficient in the correctly specified model.

Acknowledgements We thank Ana Martinovici, Dan Bartels, Xiaolin Li, Tamer Boyaci, Robert Zeithammer, Daniel Ackerberg, Jean-Pierre Dubé, Jeremy Fox, Wes Hartmann, Günter Hitsch, Carl Mela, and Stephan Seiler for helpful comments and suggestions. We also benefitted from the comments of seminar participants at the 17th Annual Bass FORMS Conference in Dallas, EMAC 2022 in Budapest, the INFORMS Marketing Science Conference 2022, as well as the Kleinwalsertal Seminars in 2020-2022. We gratefully acknowledge financial support by the German Research Foundation (DFG Grant ID: OT 447/2-1). Arash Laghaie acknowledges funding by Fundação para a Ciência e a Tecnologia (UIDB/00124/2020, UIDP/00124/2020, UID/00124, Nova School of Business and Economics and Social Sciences DataLab - PINFRA/22209/2016), POR Lisboa and POR Norte (Social Sciences DataLab, PINFRA/22209/2016).

Funding Open Access funding enabled and organized by Projekt DEAL.

Data Availability Code for estimation and replication of tables and figures in the paper are available at: https://gitfront.io/r/user-3843061/SqrbQZ5sXtbH/Rational-Inattention-Discrete-Choice/

Declarations

Conflicts of Interest This work was supported by the German Research Foundation (Grant ID: OT 447/2-1). The authors have no further competing interests to declare that are relevant to the content of this article.

References

Abaluck, J., Compiani, G., & Zhang, F. (2022). A method to estimate discrete choice models that is robust to consumer search. Unpublished manuscript, Yale School of Management, University of Chicago, and University of California.

Anderson, N. H. (1981). *Foundations of Information Integration Theory*. Academic Press.

Anderson, N. H. (1982). *Methods of Information Integration Theory*. Academic Press.

Andrews, R. L., & Srinivasan, T. (1995). Studying consideration effects in empirical choice models using scanner panel data. *Journal of Marketing Research, 32*(1), 30–41.

Aribarg, A., Otter, T., Zantedeschi, D., Allenby, G. M., Bentley, T., Curry, D. J., Dotson, M., Henderson, T., Honka, E., Kohli, R., et al. (2018). Advancing non-compensatory choice models in marketing. *Customer Needs and Solutions, 5*(1), 82–92.

Armenter, R., Müller-Itten, M., & Stangebye, Z. R. (2024). Geometric methods for finite rational inattention. *Quantitative Economics, 15*(1), 115–144.

Armona, L., Fuster, A., & Zafar, B. (2019). Home price expectations and behaviour: Evidence from a randomized information experiment. *The Review of Economic Studies, 86*(4), 1371–1410.

Bertoli, S., Moraga, J.F.-H., & Guichard, L. (2020). Rational inattention and migration decisions. *Journal of International Economics, 126*,

Bestard, A. B., & Font, A. R. (2021). Attribute range effects: Preference anomaly or unexplained variance? *Journal of Choice Modelling, 41*, 100321.

Bhattacharyya, A. (1946). On a measure of divergence between two multinomial populations. Sankhyā: *The Indian Journal of Statistics*, pp. 401–406.

Blattberg, R. C., Briesch, R., & Fox, E. J. (1995). How promotions work. *Marketing Science, 14*(3), G122–G132.

Blattberg, R., & Neslin, S. (1989). Sales promotion: The long and the short of it. *Marketing Letters, 1*, 81–97.

Boyacı, T., & Akçay, Y. (2017). Pricing when customers have limited attention. *Management Science, 64*(7), 2995–3014.

Bronnenberg, B. J., & Vanhonacker, W. R. (1996). Limited choice sets, local price response and implied measures of price competition. *Journal of Marketing Research*, pp. 163–173.

Brown, Z. Y., & Jeon, J. (2024). Endogenous information and simplifying insurance choice. *Econometrica, 92*(3), 881–911.

Cao, X., & Zhang, J. (2021). Preference learning and demand forecast. *Marketing Science, 40*(1), 62–79.

Caplin, A., & Dean, M. (2013). Rational inattention and state dependent stochastic choice. Unpublished manuscript, New York University.

Caplin, A., Dean, M., & Leahy, J. (2019). Rational inattention, optimal consideration sets, and stochastic choice. *The Review of Economic Studies, 86*(3), 1061–1094.

Carmone, F. J., & Green, P. E. (1981). Model misspecification in multiattribute parameter estimation. *Journal of Marketing Research, 18*(1), 87–93.

Caussade, S., de Dios Ortúzar, J., Rizzi, L. I., & Hensher, D. A. (2005). Assessing the influence of design dimensions on stated choice experiment estimates. *Transportation Research Part B: Methodological, 39*(7), 621–640.

Cavallo, A., Cruces, G., & Perez-Truglia, R. (2017). Inflation expectations, learning, and supermarket prices: Evidence from survey experiments. *American Economic Journal: Macroeconomics, 9*(3), 1–35.

Chen, Y., & Yao, S. (2017). Sequential search with refinement: Model and application with click-stream data. *Management Science, 63*(12), 4345–4365.

Chetty, R., Looney, A., & Kroft, K. (2009). Salience and taxation: Theory and evidence. *American Economic Review, 99*(4), 1145–1177.

Ching, A. T. (2010). A dynamic oligopoly structural model for the prescription drug market after patent expiration. *International Economic Review, 51*(4), 1175–1207.

Ching, A., Erdem, T., & Keane, M. (2009). The price consideration model of brand choice. *Journal of Applied Econometrics, 24*(3), 393–420.

Chintagunta, P. K., & Nair, H. S. (2011). Structural workshop paper–discrete-choice models of consumer demand in marketing. *Marketing Science, 30*(6), 977–996.

Choy, S., & Smith, A. (1997). Hierarchical models with scale mixtures of normal distributions. *Test, 6*, 205–221.

Coibion, O., Gorodnichenko, Y., & Kumar, S. (2018). How do firms form their expectations? New survey evidence. *American Economic Review, 108*(9), 2671–2713.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience, 3*(3), 201–215.

Cover, T. M., & Thomas, J. A. (2006). Elements of Information Theory, Second Edition. Jon Wiley and Sons, Inc.

Csiszár, I. (1974). On the computation of rate-distortion functions (corresp.). *IEEE Transactions on Information Theory 20*(1), 122–124.

Santos, B. D. L., Hortaçsu, A., & Wildenbeest, M. R. (2012). Testing models of consumer search using data on web browsing and purchasing behavior. *The American Economic Review, 102*(6), 2955–2980.

Dean, M., & Neligh, N. (2019). Experimental tests of rational inattention. Unpublished manuscript, Columbia University and University of Tennessee.

Dellaert, B. G., Brazell, J. D., & Louviere, J. J. (1999). The effect of attribute variation on consumer choice consistency. *Marketing Letters, 10*(2), 139–147.

DeShazo, J., & Fermo, G. (2002). Designing choice sets for stated preference methods: The effects of complexity on choice consistency. *Journal of Environmental Economics and management, 44*(1), 123–143.

Ding, M. (2007). An incentive-aligned mechanism for conjoint analysis. *Journal of Marketing Research, 44*(2), 214–223.

Ding, M., Grewal, R., & Liechty, J. (2005). Incentive-aligned conjoint analysis. *Journal of Marketing Research, 42*(1), 67–82.

Domencich, T. A., & McFadden, D. (1975). Urban travel demand: A behavioral analysis. Contributions to economic analysis. New York: American Elsevier, Amsterdam and Oxford: North-Holland.

Dong, S., Ding, M., & Huber, J. (2010). A simple mechanism to incentive-align conjoint experiments. *International Journal of Research in Marketing, 27*(1), 25–32.

Erickson, G. M., & Johansson, J. K. (1985). The role of price in multi-attribute product evaluations. *Journal of Consumer Research, 12*(2), 195–199.

Fiebig, D. G., Keane, M. P., Louviere, J., & Wasi, N. (2010). The generalized multinomial logit model: Accounting for scale and coefficient heterogeneity. *Marketing Science, 29*(3), 393–421.

Gabaix, X. (2019). Behavioral inattention. In: *Handbook of Behavioral Economics: Applications and Foundations 1*, vol. 2, pp. 261–343. Elsevier.

Gardete, P., & Hunter, M. (2020). Guiding consumers through lemons and peaches: An analysis of the effects of search design activities. Stanford University Graduate School of Business Research Paper No. 3669.

Gershman, S. J., & Lai, L. (2020). The reward-complexity trade-off in schizophrenia. bioRxiv, pp. 2020–11.

Gilbride, T. J., & Allenby, G. M. (2004). A choice model with conjunctive, disjunctive, and compensatory screening rules. *Marketing Science, 23*(3), 391–406.

Gilbride, T. J., Allenby, G. M., & Brazell, J. D. (2006). Models for heterogeneous variable selection. *Journal of Marketing Research, 43*(3), 420–430.

Goeree, M. S. (2008). Limited information and advertising in the U.S. personal computer industry. *Econometrica 76*(5), 1017–1074.

Guadagni, P. M., & Little, J. D. (1983). A logit model of brand choice calibrated on scanner data. *Marketing Science, 2*(3), 203–238.

Guo, L. (2021). Endogenous evaluation and sequential search. *Marketing Science, 40*(3), 413–427.

Gu, C., & Wang, Y. (2022). Consumer online search with partially revealed information. *Management Science, 68*(6), 4215–4235.

Hamilton, R., & Chernev, A. (2013). Low prices are just the beginning: Price image in retail management. *Journal of Marketing, 77*(6), 1–20.

Hauser, J. R., & Wernerfelt, B. (1990). An evaluation cost model of consideration sets. *Journal of Consumer Research, 16*(4), 393–408.

Hébert, B., & Woodford, M. (2021). Neighborhood-based information costs. *American Economic Review, 111*(10), 3225–55.

Hensher, D. A. (2006). How do respondents process stated choice experiments? Attribute consideration under varying information load. *Journal of Applied Econometrics, 21*(6), 861–878.

Honka, E., Hortaçsu, A., & Wildenbeest, M. (2019). Empirical search and consideration sets. In: *Handbook of the Economics of Marketing*, vol. 1, pp. 193–257. Elsevier.

Honka, E. (2014). Quantifying search and switching costs in the US auto insurance industry. *The RAND Journal of Economics, 45*(4), 847–884.

Honka, E., & Chintagunta, P. (2017). Simultaneous or sequential? Search strategies in the us auto insurance industry. *Marketing Science, 36*(1), 21–42.

Hu, M., Dang, C., & Chintagunta, P. K. (2019). Search and learning at a daily deals website. *Marketing Science, 38*(4), 609–642.

Huettner, F., Boyacı, T., & Akçay, Y. (2019). Consumer choice under limited attention when alternatives have different information costs. *Operations Research, 67*(3), 671–699.

Jacoby, J., Speller, D. E., & Kohn, C. A. (1974). Brand choice behavior as a function of information load. *Journal of Marketing Research, 11*(1), 63–69.

Janssen, A., & Kasinger, J. (2024). Obfuscation and rational inattention. *Journal of Industrial Economics, 72*(1), 390–428.

Jindal, P., & Aribarg, A. (2021). The importance of price beliefs in consumer search. *Journal of Marketing Research, 58*(2), 321–342.

Joo, J. (2023). Rational inattention as an empirical framework for discrete choice and consumer-welfare evaluation. *Journal of Marketing Research, 60*(2), 278–298.

Kalra, A., & Goodstein, R. C. (1998). The impact of advertising positioning strategies on consumer price sensitivity. *Journal of Marketing Research, 35*(2), 210–224.

Keller, K. L., & Staelin, R. (1987). Effects of quality and quantity of information on decision effectiveness. *Journal of Consumer Research, 14*(2), 200–213.

Kim, J. B., Albuquerque, P., & Bronnenberg, B. J. (2010). Online demand under limited consumer search. *Marketing Science, 29*(6), 1001–1023.

Kim, J. B., Albuquerque, P., & Bronnenberg, B. J. (2017). The probit choice model under sequential search with an application to online retailing. *Management Science, 63*(11), 3911–3929.

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience, 13*(10), 1292–1298.

Li, L., Walsh, T. J., & Littman, M. L. (2006). Towards a unified theory of state abstraction for MDPs, Citeseer, vol. 1, pp. 3.

Lipnowski, E., & Ravid, D. (2022). Predicting choice from information costs. arXiv preprint arXiv:2205.10434.

Liu, Q., Dean, A., Bakken, D., & Allenby, G. M. (2009). Studying the level-effect in conjoint analysis: An application of efficient experimental designs for hyper-parameter estimation. *Quantitative Marketing and Economics, 7*(1), 69–93.

Lombart, C., Louis, D., & Labbé-Pinlon, B. (2016). Price image consequences. *Journal of Retailing and Consumer Services, 28*, 107–116.

Lourenço, C. J., Gijsbrechts, E., & Paap, R. (2015). The impact of category prices on store price image formation: An empirical analysis. *Journal of Marketing Research, 52*(2), 200–216.

Louviere, J. J., Meyer, R. J., Bunch, D. S., Carson, R., Dellaert, B., Hanemann, W. M., Hensher, D., & Irwin, J. (1999). Combining sources of preference data for modeling complex decision processes. *Marketing Letters, 10*(3), 205–217.

Maćkowiak, B., Matějka, F., & Wiederholt, M. (2023). Rational inattention: A review. *Journal of Economic Literature, 61*(1), 226–273.

Matějka, F., & McKay, A. (2012). Simple market equilibria with rationally inattentive consumers. *American Economic Review, 102*(3), 24–29.

Matějka, F., & McKay, A. (2015). Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review, 105*(1), 272–298.

Mazzotta, M. J., & Opaluch, J. J. (1995). Decision making when choices are complex: A test of heiner's hypothesis. *Land Economics*, pp. 500–515.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics*, pp. 105–142.

Mehta, N., Rajiv, S., & Srinivasan, K. (2003). Price uncertainty and consumer search: A structural model of consideration set formation. *Marketing Science, 22*(1), 58–84.

Meißner, M., Oppewal, H., & Huber, J. (2020). Surprising adaptivity to set size changes in multi-attribute repeated choice tasks. *Journal of Business Research, 111*, 163–175.

Miao, J., & Xing, H. (2023). Dynamic discrete choice under rational inattention. *Economic Theory*, pp. 1–56.

Moraga-González, J. L., Sándor, Z., & Wildenbeest, M. R. (2023). Consumer search and prices in the automobile market. *The Review of Economic Studies, 90*(3), 1394–1440.

Musalem, A., Montoya, R., Meißner, M., & Huber, J. (2021). Components of attentional effort for repeated tasks. *Journal of Behavioral Decision Making, 34*(1), 99–115.

Natan, O. R. (2021). Choice frictions in large assortments. Unpublished manuscript, University of Chicago.

Nevo, A. (2000). A practitioner's guide to estimation of random-coefficients logit models of demand. *Journal of Economics & Management Strategy, 9*(4), 513–548.

Ohler, T., Le, A., Louviere, J., & Swait, J. (2000). Attribute range effects in binary response tasks. *Marketing Letters, 11*(3), 249–260.

Orme, B. (2019). Consistency cutoffs to identify "bad" respondents in CBC, ACBC, and MaxDiff. Sawtooth Software: Technical report.

Pachali, M. J., Kurz, P., & Otter, T. (2023). Omitted budget constraint bias and implications for competitive pricing. *Journal of Marketing Research, 60*(5), 968–986.

Petrin, A., & Train, K. (2010). A control function approach to endogeneity in consumer choice models. *Journal of Marketing Research, 47*(1), 3–13.

Porcher, C. (2019). Migration with costly information. Unpublished manuscript, Princeton University.

Quandt, R. E. (1956). A probabilistic theory of consumer behavior. *The Quarterly Journal of Economics, 70*(4), 507–536.

Regier, D. A., Watson, V., Burnett, H., & Ungar, W. J. (2014). Task complexity and response certainty in discrete choice experiments: An application to drug treatments for juvenile idiopathic arthritis. *Journal of Behavioral and Experimental Economics, 50*, 40–49.

Roberts, J. H., & Lattin, J. M. (1991). Development and testing of a model of consideration set composition. *Journal of Marketing Research, 28*(4), 429–440.

Rossi, P. E., Allenby, G. M., & McCulloch, R. (2005). *Bayesian Statistics and Marketing*. Chichester: Wiley series in probability and statistics.

Ruan, S., MacEachern, S. N., Otter, T., & Dean, A. M. (2008). The dependent poisson race model and modeling dependence in conjoint choice experiments. *Psychometrika, 73*(2), 261.

Russo, J. E., & Leclerc, F. (1994). An eye-fixation analysis of choice processes for consumer nondurables. *Journal of Consumer Research, 21*(2), 274–290.

Sawtooth Software (1996). Using conjoint analysis in pricing studies: Is one price variable enough? Technical report.

Shugan, S. M. (1980). The cost of thinking. *Journal of Consumer Research, 7*(2), 99–111.

Simon, H. A., & Newell, A. (1971). Human problem solving: The state of the theory in 1970. *American Psychologist, 26*(2), 145–159.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics, 50*(3), 665–690.

Sutton, R. S., & Barto A. G. (2018). Reinforcement Learning: An Introduction. MIT press.

Swait, J., & Adamowicz, W. (2001). The influence of task complexity on consumer choice: A latent class model of decision strategy switching. *Journal of Consumer Research, 28*(1), 135–148.

Terui, N., Ban, M., & Allenby, G. M. (2011). The effect of media advertising on brand consideration and choice. *Marketing Science, 30*(1), 74–91.

Thurstone, L. L. (1927). A law of comparative judgement. *Psychological Review, 34*(4), 273–286.

Ursu, R. M., Erdem, T., Wang, Q., & Zhang, Q. (2024). Prior information and consumer search: Evidence from eye tracking. *forthcoming in Management Science*.

Ursu, R. M., Wang, Q., & Chintagunta, P. K. (2020). Search duration. *Marketing Science, 39*(5), 849–871.

Yao, S., Wang, W., & Chen, Y. (2017). TV channel search and commercial breaks. *Journal of Marketing Research, 54*(5), 671–686.

Yegoryan, N., Guhl, D., & Klapper, D. (2020). Inferring attribute non-attendance using eye tracking in choice-based conjoint analysis. *Journal of Business Research, 111*, 290–304.