Discrete Choice in Marketing through the Lens of Rational Inattention*

Sergey Turlo† Matteo Fina‡ Johannes Kasinger§ Arash Laghaie¶
Thomas Otter‖

February 17, 2023

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 suggest how to develop an empirical rational inattention model for the analysis of discrete choice among multiple alternatives described along multiple attributes, as encountered in prototypical discrete choice experiments and choice-based-conjoint analysis in marketing and economics. We then illustrate how this model naturally motivates stylized empirical results that are hard to reconcile from a random utility perspective. Finally, we contrast the proposed approach to extant empirical work that builds on rational inattention.

Keywords: Choice modeling, rational inattention, conjoint analysis
JEL Classification: C00, C35, D83, M31

*We thank Ana Martinovici and Dan Bartels for detailed comments and literature suggestions. We are grateful to audiences from EMAC 2022 in Budapest, the INFORMS Marketing Science Conference 2022, as well as the Kleinwalsertal Seminars in 2020–2022. We acknowledge financial support by the German Research Foundation (DFG).

†Johann Wolfgang Goethe University Frankfurt. E-mail: turlo@marketing.uni-frankfurt.de.
‡Johann Wolfgang Goethe University Frankfurt. E-mail: fina@marketing.uni-frankfurt.de.
§Leibniz Institute for Financial Research (SAFE) & Johann Wolfgang Goethe University Frankfurt. E-mail: kasinger@safe-frankfurt.de.
¶Nova School of Business and Economics. E-mail: arash.laghaie@novasbe.pt.
‖Johann Wolfgang Goethe University Frankfurt. E-mail: otter@marketing.uni-frankfurt.de.

1 Introduction

Discrete choice models (DCM) are widely used in marketing and economics to learn about the preferences of decision-makers (DM) in multi-attribute multi-alternative (MAMA) choice settings. DCMs derived from random utility theory are known as random utility models (RUMs). Under the standard assumption of stable, preexisting utility functions, inference from RUMs is invariant to attribute range, number of attribute levels, the size of choice sets, and other characteristics. However, a large body of empirical literature documents context dependence of inference from RUMs (e.g., McFadden et al., 1999; Leong and Hensher, 2012). And researchers have proposed tailored context dependent adjustments to the utility function (Rooderkerk et al., 2011) or the assumed joint distribution of random aspects of utility (e.g., Dotson et al., 2018) to motivate individual instances of context dependency.

In this paper, we propose a structural DCM rooted in rational inattention (RI) theory and adapted to MAMA settings that motivates context dependent choice behavior. RI theory assumes that information processing is costly and that DMs optimally and adaptively choose what and how much information about alternatives and their payoffs to process.

We illustrate how various phenomena that are at odds with the RUM, like consideration sets and non-attendance to attributes, endogenously follow from the optimal deployment of limited cognitive resources in the suggested model. As such, predictions for consideration sets and attribute non-attendance become implicit functions of the composition of a choice set, reflecting adaptations to easily recognized features of a choice set. We show how our RI-based DCM motivates previously documented effects of attribute range, number of levels, and the size of choice sets on inferences from RUMs. We also clarify the difference between RI and search theory as structural motivations for observable choice behavior.

We view these illustrations as both practically useful and important in a conceptual sense. The RI-DCM is a practical and parsimonious model to capture a variety of stylized findings that previously required separate, non-unified adjustments to RUMs. We show that phenomena that conceptually question the relevance of stable, additively separable utility functions in the context of analyses based on RUMs are perfectly compatible with this assumption once optimization over limited cognitive resources is considered. Noteworthy departures from additive separability in applied research that lack a structural motivation in the context of RUMs include brand-specific price coefficients (e.g., Carmone and Green, 1981; Sawtooth Software, 1996; Kalra and Goodstein, 1998) as well as 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 and Little, 1983; Blattberg and Neslin, 1989; Chetty et al., 2009).

We contrast the proposed model formulation to existing applications of RI both in the context of experimental and observational choice data. Different from the aim of this paper, existing applications using experimental data primarily focus on inferring cognitive costs given known utilities or payoffs, whereas existing applications using observational data invoke specific and potentially unrealistic assumptions to preserve a notion of additive separability in the estimation equations. Paving the way for future empirical applications, we illustrate the estimation and empirical identification of the proposed model in a hierarchical setting with heterogeneity using simulated data. Finally, we discuss open challenges for this model to become a new workhorse in applied research.

In line with basic insights from psychology (e.g., Simon and Newell, 1971), RI theory, developed by Sims (2003) in macroeconomics, assumes that DMs are cognitively constrained and thus will not usually process all available information when making decisions. DMs are aware of their constraint and are able to optimally choose how much and what kind of costly information to process in any given decision context. RI predicts that DMs adapt their processing efforts based on more readily available, salient aspects of a choice task that, in turn, determine their (rational) expectation of how information processing translates into incremental utility from choice.

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 random utility aspects shift the index values of choice alternatives for a utility-maximizing individual. The prevailing interpretation of random utilities is that they capture the valuation of choice relevant information known to the DM, but not to the analyst. This interpretation thus assumes that the DM chooses based on a larger information set than available to the analyst. Different from RUMs, RI based choice models can rationalize choice from MAMA sets without invoking aspects of utility that are only observed by the DM. This is in the tradition of researchers who focused early on explaining the associated randomness with cognitive processes beyond the evaluation of attributes only observed by the DM (e.g., Thurstone, 1927; Quandt, 1956; Louviere et al., 1999).

Over time, researchers proposed various tailored modifications and additions to RUMs to accommodate the aforementioned empirical phenomena, often loosely motivated by informational frictions or cognitive limitations and justified by increased fit and face validity of implied substitution patterns. Consider, for instance, the case of consideration sets. Consideration sets reflect that DMs may not fully evaluate all alternatives in a choice set, where partial evaluations translate into choice probabilities equal to zero. Consideration sets have been motivated from i) (strict) requirements on observed attributes also known as screening rules (e.g., Gilbride and Allenby, 2004), ii) from variables excluded from the utility function such as e.g., advertising that only create awareness (e.g., Goeree, 2008; Terui et al., 2011), and iii) as a result of costly search (e.g., Honka et al., 2019).

Motivations i) and ii) assume that consideration set formation is exogenous, i.e., the consideration set is not motivated from constrained optimization. However, if consideration sets reflect limited cognitive resources and the deployment of these resources follows some optimality calculus, exogenous consideration set formation models can only capture special cases. Similar concerns apply to assuming exogenous and fixed probabilities of processing specific attributes and, more generally, the independent contribution of individual attributes to choice. In contrast, consideration sets from search do result from constrained optimization. However, extant empirical search models rule out choice under partial information about considered alternatives, i.e., assume full information about searched alternatives.

The remainder of this paper is organized as follows. Section 2 introduces the RI framework in the context of discrete choice among alternatives described along multiple attributes. Section 3 discusses existing empirical applications of RI in the context of discrete choice and illustrates estimation and empirical identification in a hierarchical model with heterogeneity using simulated data. Section 4 discusses key features of discrete choice under RI theory when applied to a typical marketing setting and provides illustrative simulations. Section 5 concludes with a discussion and an agenda for future research.

2 Rational Inattention and 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 thus decide optimally how to process the available information, 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. We first present the discrete choice version of 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 how both the complexity of a choice task and the incentives to process information affect choice behavior. While we build on the analytical model introduced by Matějka and McKay (2015) and extended by Caplin et al. (2019) in this section, the adaptation to the MAMA setting, including estimation of hierarchical RI models with heterogeneity in Section 3, as well as the connections between RI and stylized results from choice among multi-attribute alternatives established in the following Section 4 are novel.

To ease the exposition of the various components of the RI framework, we will refer to an exemplary 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.[1]

2.1 Formal Problem and its Translation into MAMA Settings

We next describe the problem faced by the rationally inattentive DM according to Caplin et al. (2019). There is a finite number of states $\omega \in \Omega$ the DM can learn about.[2] An action a is a mapping from states to utilities. $\mathcal{A}$ denotes the set of actions. The mapping $u : \mathcal{A} \times \Omega \to \mathbb{R}$ describes the utility from any action in each state. This problem 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.

[1] There are many more things about a car that are likely payoff relevant to a DM, and may or may not be effortful to evaluate. However, this minimal example will help develop basic principles.

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

Payoffs in DCE The general nature of RI theory provides room for different translations of this framework into the typical MAMA setting in marketing. Unless stated otherwise, we impose that actions correspond to different alternatives from which the DM chooses and states ω represent different choice sets, characterized by the specific attribute compositions of available alternatives, that the DM can face in an experiment.

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 and 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 derives 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 evaluation and integration with simple attributes* requires cognitive effort and time.[3][4]

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

$$u(a, \omega) = \mathbf{x}'_{a,s}(\omega)\beta_s + \mathbf{x}'_{a,c}(\omega)\beta_c, \tag{1}$$

where β_s and β_c are the respective part-worths of simple and complex attributes.[5] 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.[6]

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 the complex attributes. We term the part of utility due to simple attributes the "simple utility component" and the part due to complex attributes as 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 be fully processed.

Prior beliefs In the RI model, 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. Ω corresponds to the set of all attainable choice sets in a given experimental design, $\Delta(\Omega)$ is the set of distributions over Ω, and $A \subset \mathcal{A}$ is the set of actions she can choose from. As stated previously, our formulation in the context of DCEs implies that 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

[3]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 and Leclerc, 1994; Musalem et al., 2021). This orientation phase and subsequent behavior involves both bottom-up and top-down processing (see Corbetta and Shulman, 2002, for a review).

[4]What distinguishes the RI-DCM from a consumer search model is that RI does not impose restrictions on the type and extent of information processing. What distinguishes the RI-DCM from two-stage models of choice is the adaptive nature of implied consideration sets. We will revisit these points in detail in Section 4. We investigate the empirical distinction between simple and complex attributes in Section 5.

[5]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. As we discuss later, RI gives rise to behavior that may appear as if, e.g., different price components are weighted differently or even interact despite homogeneous, additively separable contributions of price components to utility or payoffs.

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

Choice set	ω_1	ω_2	ω_3	ω_4
Probability of ω_i	$\gamma_p(1-\gamma_d)$	$\gamma_p\gamma_d$	$(1-\gamma_p)(1-\gamma_d)$	$(1-\gamma_p)\gamma_d$
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	$p = p_L$, $d \in \{0, D\}$	$p = p_L$, $d \in \{0, D\}$	$p = p_H$, $d \in \{0, D\}$	$p = p_H$, $d \in \{0, D\}$
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

Table 1: **Set of choice sets with respective payoffs and prior beliefs.** 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 $\Pr(u_O = 0|\omega_i) = 1$.

DM, each combination of simple attribute realizations 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 design of the choice experiment.

To clarify more concretely, we return 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 will 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 in this example.[7] We will further assume that the DM has to commit to purchasing the car at the 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 $\Pr(d = D) = \gamma_d > 0$. In Table 1, 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 (prior to the realization of the simple attributes) are displayed in Table 1. However, 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 are two different

[7] In the notation of equation (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.

conditional prior beliefs μ_s by design. Each combination of simple attributes will result in different prior beliefs, which the DM subsequently uses for optimizing attention allocation and choice. In our example, choice sets ω_1 and ω_2, as well as choice sets ω_3 and ω_4 yield the same prior beliefs (see the last four rows in Table 1).

Intuitively, one can think of a sequentially updating DM who is aware of the experimental design and 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 $\mathbf{x}_c$.[8] Because processing of complex attributes takes effort, it may be optimal not to perfectly process these attributes before making a choice.

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 a stochastic signal. The signal (usually imperfectly) indicates the true "state of the world". In our example, the state corresponds to the choice set, defined by the true attribute values. The distribution of the stochastic signal results from the DM's information processing strategy. The DM controls the quality of the signals by her processing effort. More precise signal distributions result from investing more costly effort.

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 belief about the true complex attribute(s) levels in a choice set translates into a belief about what choice action a maximizes her payoff that then determines the choice.

Formally, a processing strategy is characterized by a signal distribution given states ω. And the optimal processing strategy under RI maximizes an expected payoff from information processing (and implied choices) while accounting for information processing costs. Because different realized signals deterministically map into different choice actions a, a processing strategy is equivalently characterized by (state) conditional signal distributions or (state) conditional choice probabilities $P(a|\omega)$. The conditional choice probability $P(a|\omega)$ 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.

DM's objective Under specific assumptions about information processing costs discussed below, RI postulates that DMs choose according to conditional choice probabilities $P(a|\omega)$ that maximize:

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

[8] In this 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.

where $P_s(a) = \sum_{\omega \in \Omega} \mu_s(\omega) P(a|\omega)$ denote the unconditional choice probabilities. The first term 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 are the expected payoffs of the DM. However, there are (processing) costs formalized in the second term of expression (2) that is multiplied by $\lambda > 0$, the unit costs of information. The term in brackets is known as mutual information and measures the expected reduction in choice uncertainty due to optimal information processing. The stronger $P(a|\omega)$ deviates from the unconditional choice probability $P_s(a)$, in the sense of putting more weight on the optimal action in a choice set ω, the higher the costs of information processing.[9]

Unconditional choice probabilities $P_s(a)$ differ between alternatives in a choice set and across choice sets as a function of the realization of simple attributes, as indicated by the subscript (see also Table 1). They correspond to the probability of choosing alternative a from a choice set ω, given an optimal processing strategy but *before actually processing the costly information* as provided by *complex* attributes in a specific choice set. Unconditional probabilities are not arbitrary subjective functions of simple attributes but constrained by the equality $P(a) = \sum_{\omega \in \Omega} \mu_s(\omega) P(a|\omega)$. In other words, unconditional choice probabilities are defined as prior *expectations* over how choice probabilities would change through costly processing of *possible* realizations of costly attributes. It is this constraint that imbues the model with rationality and implies that processing complex information will be associated with positive costs (mutual information cannot be negative).

Turning back to the exemplary DCE, a DM will form, based on the simple brand and price that she processes at no cost, 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.[10] Depending on the observed simple attributes, she has different prior beliefs over possible payoffs (see, e.g., columns 2 and 3 in Table 1). This prior belief will affect how she processes the information about the discount. In choice sets where the price is either very large or very small, she may choose not to process any information about the discount because it is unlikely to affect her choice. However, when the difference between expected payoffs of available alternatives is small, the discount will be the deciding factor for the best alternative. Consequently, obtaining a (more) precise signal over the true value of the discount becomes valuable to the DM.

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

[9] Mutual information is based on the concept of Shannon entropy. 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 choice probabilities implied by an optimal level of processing of complex attributes ($P(a|\omega)$). The expectation is formed over the prior distribution of complex attributes, conditioned on realized simple attributes in a particular choice set (μ_s).

[10] 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., 2021).

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.[11] Thus, the DM will always choose the utility maximizing alternative and make no choice errors if $\lambda = 0$.

However, 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, but where—retrospectively—it would have been optimally not to buy the car as the discount does not apply. The larger the information processing costs the more (or more consequential) choice errors the DM will make, everything else equal. These 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 aspects of utility imperfectly observed or processed by the DM. Hence, under RI, the probabilistic nature of choice simply reflects the DM's information processing, which is imperfect because of processing costs. Also, note that the strategy implied by maximizing the expression in (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.

Finally, 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 both 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, or even 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 Comparative Properties

Matějka and McKay (2015) derive necessary conditions for the solution of the problem. They describe the state-contingent choice probabilities $P(a|\omega)$ for those actions a that are chosen with strictly positive unconditional probability as

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

Upon rewriting $P_s(a)\exp\{u(a,\omega)/\lambda\}$ as $\exp\{(u(a,\omega) + \lambda \log P_s(a))/\lambda\}$, equation (3) shows that state-contingent choice probabilities $P(a|\omega)$ are given by a modified logit formula where $u(a,\omega)$, the payoffs from an action a in a state ω, are shifted by $\lambda \log P_s(a)$ and divided by the information processing costs λ. The latter term 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.[12] Clearly, the extent of this shift depends on the magnitude of the processing

[11] Note that the unconditional choice probability of buying the car will be equal to $\Pr(d = D) = \gamma_d$ in this case.

[12] An alternative way how to view equation (3) is that it is the product of $P_s(a)$, the unconditional or expected choice probabilities of an alternative a, and a weight that explicitly depends on the payoffs in the specific choice

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 a MAMA setting. Because realizations of simple attributes affect both prior beliefs μ_s as well as choice specific payoffs $u(a,\omega)$ both unconditional choice probability $P_s(a)$ as well as the term $\exp\{u(a,\omega)/\lambda\}$ are affected. In contrast, realized values of complex attributes, that 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.

Caplin et al. (2019) extend the results of Matějka and McKay (2015) and characterize the set of actions chosen with strictly positive probability. They show that a choice mapping P is 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 $a \in A$, with equality when $P_s(a) > 0$, and if for all such actions and states equation (3) holds. 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 (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 updating problem from unconditional to conditional choice probabilities. 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. We already note that in contrast to extant two-stage models of consideration, the consideration set implied by RI is a function of the realized simple attributes of all alternatives in a choice set, i.e., cannot be reduced to a decision rule where the marginal considered alternative is defined independent of the context of a specific choice set. We will revisit this point further below.

Further intuition can be gained from considering the DM's objective in expression (2). When $\lambda = 0$, in each state the utility maximizing action a is chosen, i.e., no actions are ruled out before processing all information as information processing is costless in this case. For intermediate values of λ, it may be optimal for the DM to set $P_s(a') = 0$ for some a' so that in all choice sets characterized by the same realizations of simple attributes $P(a'|\omega) = 0$, reducing the DM's processing costs as described above. When λ is sufficiently large, the DM no longer processes any information and thus chooses based on (conditional) prior beliefs only. In the first case, the DM's consideration set size is maximized, while it contains only a single element in the last case. Thus,

set (state) ω. This weight shifts the choice probabilities such that alternatives that provide a relatively high payoff in a specific choice set (state) are more likely to be chosen than expected.

choice behavior is deterministic both when $\lambda = 0$ and when λ is very large. However, choice is stochastic for intermediate values because of the underlying cognitive information processing.[13]

Finally, we note that the RI model of discrete choice does not have a closed-form solution unless one is willing to make restrictive, and arguably often unrealistic assumptions.[14] Formally, this is due to the non-linear optimality conditions implied by equation (3) and condition (4). Below we rely on the Blahut-Arimoto algorithm for numerical solutions to the RI problem (see e.g., Caplin et al., 2019, Section 2.2). We discuss alternative strategies aimed at preserving a closed-form solution in the following Section 3, where we cover recent empirical implementations of the RI model.

3 Empirical Identification Strategies

This section discusses empirical identification strategies under the assumption of rationally inattentive DMs. We first present challenges associated with an empirical application of the RI framework. We then provide an overview and discussion of extant identification strategies in experimental and observational settings (Sections 3.1 and 3.2). This is followed by an illustration of the empirical identification and Bayesian estimation of a hierarchical RI-DCM with heterogeneity (Section 3.3).

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 of the aforementioned primitives and learn from data about the remaining one. In doing so, the analyst has to make assumptions about what a state is and what DMs may costly learn. Note that since the objective function in the RI framework is homogeneous of degree one with respect to payoffs and costs, only their ratio is directly identifiable from the data. Also note that RI covers deterministic behavior such that a particular data set may only set identify preferences, conditional on processing costs and prior beliefs.

Existing applications with observational data make strong, and arguably often unrealistic, assumptions about prior beliefs to conveniently obtain choice probabilities in closed-form without having to numerically solve for unconditional choice probabilities. As we will elaborate in detail below, one can rationalize a logit model with an additively separable utility index as an RI-DCM in this way. However, this computationally appealing solution rules out many qualitatively distinguishing features of an RI-DCM such as endogenous consideration sets a

[13] Also note how consideration under RI usefully distinguishes between prior and posterior dominance of an alternative in a choice set. If an alternative is dominated by another alternative already based on realized simple attributes, given (conditional) distributions of complex attributes, it can never be chosen under RI. However, if an alternative is worse on simple attributes but may or may not be dominated based on realized complex attributes, RI can motivate a positive probability of choosing an alternative that is dominated a posteriori, i.e., after processing (some) complex information.

[14] For example, assuming that all alternatives are ex ante identical (and thus ignoring the notion of easily processed simple attributes differentiating alternatives a priori) leads to logit choice probabilities that are exactly identical to those implied by a standard RUM.

priori (see Section 4 for a demonstration of distinctive implications of an RI-DCM that align with stylized observations in choice data). In contrast, we will assume rational expectations derived from the nature of the experimental design presented to DMs. Extending beyond the experimental realm, rational expectations about product attributes and other marketing variables give rise to distributions with limited support that, under RI, motivate qualitative departures from what can be captured by a logit model with an additively separable utility index. Extant empirical applications using experimental data typically do not estimate preferences. In contrast, we are concerned with the recovery of preference parameters and preference heterogeneity from MAMA choices in DCEs.

3.1 Empirical Identification Strategies with Observational Data

Observational data usually necessitate stronger assumptions for identification as the analyst does not control the decision environment as would be the case in an experimental setting. Existing approaches typically choose, sometimes only implicitly, prior beliefs such that the resulting conditional choice probabilities have a logit form that depends on an additively separable index of observable characteristics of alternatives

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

where $\mathbf{x}_a'$ are some alternative specific variables and β are coefficients estimated from data. These may include non-price attributes and various price components. Note that this expression for choice probabilities is identical to that from the RU logit with type 1 extreme value (T1-EV) distributed error terms. We provide the details of deriving logit choice probabilities with an additively separable index (in attributes and other covariates) under RI further below.

The advantage of this strategy is that it translates the RI framework into a tractable logit form which can then be studied by established analytical methods, facilitating estimation. Moreover, it provides a micro-foundation for DCMs that imply logit choice that depends on an index that includes non-utility components (Joo, 2022). However, a crucial implication of constraining RI in this way is that behavioral patterns implied by RI other belief structures, such as interactive contributions of attributes or endogenous consideration sets, are ruled out a priori (see Bertoli et al. (2020) and Joo (2022) for applications under these assumptions). Moreover, assumptions about beliefs implied by this formulation often, if not always, are at odds with their rational counterparts, more on this point below.[15]

Further, existing contributions differ in their assumptions on the nature of simple information. One group of papers such as Brown and Jeon (2021) assume, similar to our proposed operationalization, 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

[15] A different strategy for dealing with unobservable beliefs is suggested by Caplin et al. (2016) who show how a simple discrete market framework can be exploited to infer unconditional choice probabilities (and thus the effect of prior beliefs) from past market shares. In their model, market shares are freely observable by RI consumers that do not perfectly observe their private preference type. As a result of a steady state condition, market shares mirror the true distribution of preferences in the population, which then allows to disentangle the effects of beliefs and preferences.

(out-of-pocket-costs).[16] Other papers, including Joo (2022) 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.

Logit choice with additive separability under RI If prior beliefs over the uncertain utility component follow a Cardell distribution, the solution to the RI problem yields logit choice probabiliteis with indices that are additively separable in simple (directly observed and processed) and complex (costly processed) components, see Brown and Jeon (2021), Bertoli et al. (2020), and Porcher (2019). Consider a DM who chooses among several alternatives $a \in A$. The utility from choosing alternative a in a state ω is

$$u(a, \omega) = \mathbf{x}'_{a,s}\beta_s + \mathbf{x}'_{a,c}\beta_c.$$

Subscripts s and c mark simple and complex utility components, respectively. Recall that under RI the DM observes and processes simple attributes $\mathbf{x}_s$ at no cost and learns optimally about the contribution of complex attributes $\mathbf{x}_c$ at the cost of λ given preferences β_s and β_c.

Recall from the discussion in Section 2 that the key difficulty in determining the solution of 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 equation (3) both for continuous and discrete prior beliefs.

By assuming that prior beliefs are such that the complex utility components $\mathbf{x}'_{a,c}\beta_c$ are independently and identically distributed according to the Cardell distribution across alternatives, Brown and Jeon (2021) as well as Bertoli et al. (2020) show that the DM 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

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

[16] A nice feature of the strategy proposed by Brown and Jeon (2021) 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.

where C is a function of the variance of complex component $\mathbf{x}'_{a,c}\beta_c$ and information processing costs λ.[17] This together with equation (3) implies that choice probabilities conditional on a specific choice set for an alternative a read

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

In the application by Joo (2022),[18] 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. All product attributes are assumed to be complex so that $u(a,\omega) = \mathbf{x}'_{c,a}\beta_c$. 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 (2022) 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 (2022) brings to the data, i.e., a logit form with an additively separable index of alternative attributes $\mathbf{x}_{c,a}$ and information shifters $\mathbf{d}_a$, requires very specific beliefs about the index from alternative specific attributes $\mathbf{x}_{c,a}$—not derived from the objective distribution of this index in the marketplace—and substantially different from this distribution, as already implied by the full support assumption.[19] This limits the formulation proposed in Joo (2022) as a model of rationally inattentive DMs that acquire knowledge about the distribution of alternative specific attributes $\mathbf{x}_{c,a}$ in the market place over longer time horizons. Another critical aspect of this model formulation is that consumer beliefs are only implicitly determined preventing further investigation and assessment.

3.2 Empirical Identification Strategies with Experimental Data

Controlled experiments allow for a clean identification and formal tests of RI and its predictions. Manipulating primitives that influence the choice likelihood under RI is comparably easy in a controlled environment. Extant experimental studies are highly stylized, though. In a typical choice task there are only few states and actions, participants are explicitly provided with prior beliefs, and they usually need to solve abstract information tasks to determine the state, such as in summing or counting exercises. In the majority of experimental studies, utility is assumed to be linear in monetary payoffs, and thus known to the analyst. These studies then primarily focus

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

$$u(a,\omega) = \mathbf{x}'_{a,s}\beta_s + \mathbf{x}'_{a,c}\beta_c + \log P(a) + \varepsilon_a$$

where ε_a are identically and independently T1-EV distributed. The Cardell distribution is the conjugate prior to T1-EV distribution which implies that the sum of a T1-EV random variable and a Cardell random variable also follows the T1-EV distribution. This feature is key for obtaining unconditional choice probabilities in closed form. For a detailed derivation see Appendix A-1 in Brown and Jeon (2021) and Appendices A.1 and A.2 in Bertoli et al. (2020).

[18] The study by Natan (2021) is based on a similar identification strategy albeit in that study information processing costs λ explicitly depend on the size of the choice set.

[19] While Cardell beliefs result in additive separability as demonstrated by Brown and Jeon (2021), the model in Joo (2022) is not immediately consistent with Cardell beliefs because of the assumption that the outside good payoff is known with certainty.

on estimating information processing costs rather than utility parameters. The resulting data is coined "state-contingent choice data" signifying that the analyst has complete information about the state in contrast to the participants (Caplin and Dean, 2015).

An example for an experimental paper that refers directly to RI theory and nicely illustrates the general structure of such experiments is Dean and Neligh (2019). Subjects are presented with a stylized information task to learn about the uncertain "state of the world". In this study, the states refer to a fraction of red and blue balls. In a subsequent step, subjects choose between different predefined actions. The payoffs of these actions depend on the state. Prior beliefs are fixed by informing subjects about the a priori likelihood of each state. To test a number of RI predictions, the experimenters vary different primitives: i) the set of available actions while keeping the number of states fixed, ii) the state-dependent payoffs to alter incentives, iii) the prior beliefs, or vi) the number of states while keeping the number of actions fixed.

The authors show that subjects adjust their attention in response to changes in incentives, and behavior is consistent with optimal processing of costly information. Moreover, this work shows that increasing the number of actions can increase the choice likelihood of existing alternatives. This constitutes a violation of regularity, a key theoretical difference between RI-DCM and discrete choice under the RUM (Matějka and McKay, 2015).

Other related experimental studies testing predictions of RI theory are presented in Ambuehl et al. (2020) who study the impact of fixed incentives on the selection of rationally inattentive individuals into different transactions. By varying participation fees, and therefore changing the relative payoff structure, they find that higher participation fees disproportionately attract individuals with high information costs. Novák et al. (2021) experimentally test a RI model of "belief polarization", which predicts diverging information acquisition across individuals depending on the payoff of the safe outside option, even if prior beliefs are identical. They provide experimental evidence in line with this prediction in a binary choice task, in which subjects can choose between two predefined informative signal structures.

These two studies are relevant to our paper, as they implicitly study the effects of simple attributes, such as the size of the participation fee or the value of the outside option, on the extent and type information acquisition about a complex attribute. More generally, early studies by Cheremukhin et al. (2015) and Pinkovskiy (2009) find, in accordance with RI, that subjects respond to incentives by processing more information, which is connected to more consistent choices by the subjects.[20]

[20] As part of their study, Cheremukhin et al. (2015) also estimate a general choice (under risk) model, including risk preferences.

3.3 Empirical Identification with DCE Data under Preference Heterogeneity

We conclude this section by illustrating the Bayesian estimation and empirical identification of the proposed RI-DCM (Section 2.1) in a "small T, large N" setting, typical of DCEs in marketing. For this purpose, 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.

For the sake of illustration, 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: $\beta_b \sim \mathcal{N}(2.5, 0.25)$, $\beta_p \sim \mathcal{N}(-1, 0.04)$. In the experimental design, 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 $d \in \{0, 0.5, 1, 1.5, 2\}$ so that the resulting design is orthogonal. Individuals' beliefs are such that they know the value of the price p, and for all prices and for any discount level d' they believe that $\Pr(d = d'|p) = 1/5$.

With the simulated data we fit the RI-DCM and two RU logit specifications: one that allows for separate price and discount coefficients and one with only one coefficient measuring the utility of money, 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. We obtain the likelihood by solving the problem (2) for given parameters numerically with the Blahut-Arimoto algorithm. Without loss of generality, we fix the value of λ to be equal to its true value in estimation.[21]

Table 2 summarizes posterior means and Table 3 reports posterior variances. We see that the estimated RI-DCM nicely recovers data generating parameters. The inferior fits of the RU logit models testify to the empirical identifiability of the RI-DCM relative to the (still) current benchmark model (see the last column in Table 2). 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., suggests 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

[21] On the hand this is without loss of generality because the proposed model only identifies preference and cost parameters up to a multiplicative constant, i.e., the objective function in equation (2) is homogeneous of degree one. On the other hand, one could structure the distribution of heterogeneity such that variation in processing costs is an underlying one-dimensional factor and only higher dimensional variation is (necessarily) pure preference heterogeneity.

Model	Brand	Price	Discount	$\lvert\beta_p/\beta_d\rvert$	$\lvert\beta_p/\beta_b\rvert$	Log Marginal Density
Data generation	2.50	-1.00	$\equiv -\beta_p$	1.00	0.40	
RI-DCM	2.50	-1.01	$\equiv -\beta_p$	1.00	0.40	-2,403.56
	(0.04)	(0.02)				
RU logit separate	27.41	-9.50	4.17	2.28	0.35	-2,520.19
	(0.69)	(0.22)	(0.13)			
RU logit joint	9.91	-4.05	$\equiv -\beta_p$	1.00	0.41	-4,436.71
	(0.19)	(0.07)				

Table 2: **Posterior means of preference distributions for different model specifications.** Standard errors are in parentheses. $|\beta_p/\beta_d|$ is the ratio of mean coefficients.

Model	Brand	Price	Discount
Data generation	0.25	0.04	0.04
RI-DCM	0.35	0.08	$\equiv \mathrm{Var}(\beta_p)$
	(0.05)	(0.01)	
RU logit separate	0.25	0.98	1.69
	(0.20)	(0.71)	(1.36)
RU logit joint	0.45	1.43	$\equiv \mathrm{Var}(\beta_p)$
	(0.03)	(0.21)	

Table 3: **Posterior variances of preference distributions for different model specifications.** Standard errors are in parentheses.

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 (see Section 3.1).

Recall that in the data-generating process, there are multiple sources of heterogeneity. First, there is heterogeneity in preference parameters across DMs. Second, due to the adaptive nature of prior beliefs and choice, as described in Section 2.1, there is non-iid heterogeneity across choice tasks within a DM. Finally, we note that with "small T, large N" data, the random pairing of specific choice sets with specific preferences becomes influential for how measures of heterogeneity from a hierarchical RU-logit depart from the hierarchical RI-DCM.

In the following Section 4, we discuss in detail how the RU logit fails to capture stylized aspects of choice behavior that contribute to the difference in model fit we see here.

4 Features of Discrete Choice under RI

In this section, we first illustrate implications of RI for choice given a fixed experimental design and then discuss how changes in the experimental design affect choice under RI. We show that a number of well documented phenomena in the discrete choice literature that are difficult to motivate in a RU framework naturally follow from choice under RI.[22]

Table 4 provides an overview of the covered endogenous characteristics and examples from extant literature dealing with these. In Table 5 we summarize the different design manipulations

Feature	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)
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 shelfspace (e.g, Bronnenberg and Vanhonacker, 1996), and price (e.g, Andrews and Srinivasan, 1995); consumer search (Hauser and Wernerfelt, 1990; Roberts and Lattin, 1991) with price uncertainty (e.g, Mehta et al., 2003; Honka, 2014; De los Santos et al., 2012; Honka and Chintagunta, 2017), match value uncertainty (e.g, Kim et al., 2010, 2017; Moraga-González et al., 2021), or multi-dimensional uncertainty (e.g, Chen and Yao, 2017; Yao et al., 2017)

Table 4: **Endogenous characteristics of RI and examples of extant related literature.**

that have been studied in the literature. We highlight which aspects of the RI model they affect and how these manipulations impact choice behavior under the assumptions of RI.

Throughout this section, our illustrations build on the leading 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 a DCE

4.1.1 Stochastic Choice due to Limited Attention

As already mentioned previously, 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 often only sparsely reflect the actual choice environment, it lacks appeal in the typical DCE where the analyst fully controls the amount of information provided to the DM.

[22] See Gabaix (2019) and Maćkowiak et al. (2021) for a detailed account of the empirical relevance of inattention in choice behavior.

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., **?**Keller and Staelin, 1987) or the structure of the choice set, e.g., the number of differing attributes (Mazzotta and Opaluch, 1995), attribute correlations within a choice set (DeShazo and Fermo, 2002), or entropy (Swait and Adamowicz, 2001). Higher task complexity reduces choice consistency (DeShazo and Fermo, 2002), implies smaller impact of complex attributes (e.g., Chetty et al., 2009), and the use of simple choice rules (Swait and Adamowicz, 2001; Orme, 2019).
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 and 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.	Depeding on the incentives set by the additional levels the impact may either increase or decrease.	More attributes are typically observed to lead to a stronger attribute impact (e.g., Liu et al., 2009), albeit contradicting evidence (e.g Hensher, 2006) exists.
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 and Aribarg, 2021). A large marketing literature identifies the impact of price image for brands and stores (e.g., Hamilton and 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 of complex attributes increase relative to the coefficients of simple attributes.	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 and 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.

Table 5: **Effects of exogenous design variations and examples of extant related literature.**

Extant Literature There is a history of relating choice errors in discrete choice to cognitive processes (e.g., Louviere et al., 1999) that dates back to the original interpretation of the probit model in Thurstone (1927). Shugan (1980) suggests a model where the comparison of any two products in a choice set is costly. Rather than maximizing utility, the DM aims at choosing the best alternative with a sufficiently high probability that is given by an exogenous parameter. In economics, De Palma et al. (1994) propose a model of a DM who may lack in the ability to choose which translates into choice error once the choice task is sufficiently complex. These authors explicitly link the ability to choose to the DM's inability to conduct a complete comparison of all viable budget allocations. Notably, the source of the error lies in the DM's perception of utilities. However, the error is exogenous given a level of complexity. Under the assumption of a T1-EV distribution, choice probabilities follow the multinomial logit model à la McFadden (1974). Finally, and in contrast to RI, the extant literature motivating stochastic aspects of choice from cognitive limitations is not derived from an integrated solution to a constrained optimization problem.

4.1.2 Attribute Interactions and Inattention to Attributes

Recall that we distinguish between simple and complex attributes. Realizations of simple attributes shift prior beliefs, and hence affect the DM's strategy of costly processing of complex attributes. As such, the contribution of realized levels of complex attributes to choice depends on the realized levels of simple attributes in this set, even if the underlying utility function is linearly additively separable.

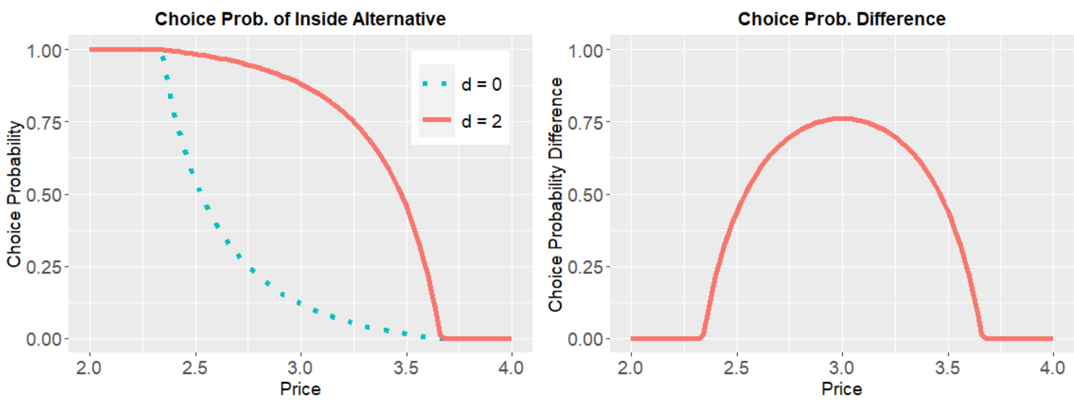

Figure 1: **Hump-shaped impact of a constant discount increase for different (simple) price levels.** The left panel displays conditional choice probabilities as a function of the simple price for the possible discount levels. In the right panel, differences in the choice probability of the inside good 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.

For illustration, consider the effect of the discount on choice for different simple price levels. In Figure 1, the right panel displays the impact of a discount increase from 0 to 2 in a DCE when d is either $d = 0$ or $d = 2$ with equal probability, and the simple brand coefficient is always $\beta_b = 2$. Recall that under RI realized values of simple attributes determine the optimal processing level of complex attributes. In particular, 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 ($p = 3$).

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.

Model	Brand	Price	Discount	Discount $\times$ Price	Discount $\times$ Price^2
Main Effect	15.86	-6.00	2.27		
	(0.48)	(0.18)	(0.08)		
With Interactions	15.20	-5.81	-10.73	8.38	-1.34
	(0.63)	(0.24)	(2.12)	(1.53)	(0.22)

Table 6: **Logit approximation with linear and quadratic interaction terms of simple price and complex discount.** First and third row show coefficient estimates for approximations with and without 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 $\Pr(d = 0) = \Pr(d = 2) = 0.5$.

We close this illustration by highlighting how fitting an RU logit model to data generated from the RI-DCM may (mis-)lead an analyst into questioning a theory-based utility function. Table 6 illustrates that a main-effects logit model fitted to observations generated from the RI-DCM with $N = 5,000$ 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 how to motivate these results. In a similar way, RI as a data generating mechanism can motivate brand-specific coefficients in a logit model fitted to the RI data. Interestingly, industry researchers generally include brand specific price coefficients in models fitted to data from DCEs, despite the push back from academic researchers that call out the lack of an economic rationale for such interactions (see e.g., Sawtooth Software, 1996).[23]

Clearly, 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, as introduced earlier, 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 $\beta_b = 6$. We increase the number of discount levels and adjust the brand coefficient accordingly for illustration purposes and without loss of generality. Figure 2 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

[23]See Blattberg et al. (1995) for an alternative motivation for brand specific price coefficients in a descriptive model.

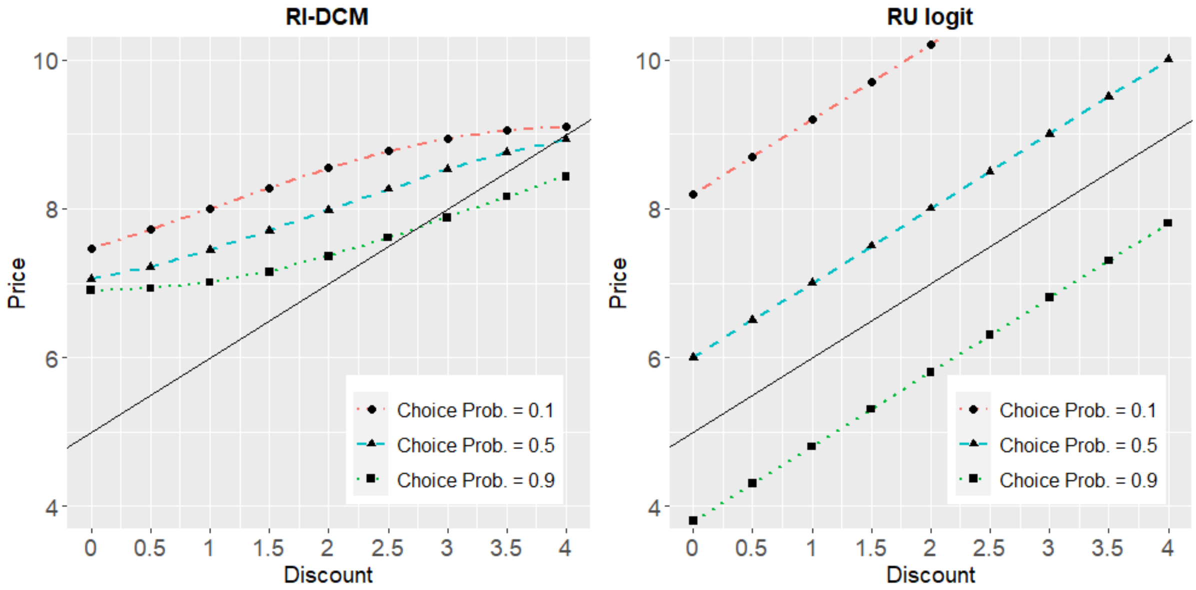

Figure 2: **Iso-choice-probability sets for the RI-DCM and the RU logit.** This Figure displays sets of price-discount combinations of the inside good that result in same conditional choice probabilities for the RI-DCM (left panel) and the RU logit (right panel). The thin solid line is the 45° line through the origin. 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.

linear and has 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 RI, it matters whether the source of utility is a simple or a complex attribute: 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 such that choice probabilities will be a function not only of the utility value of an alternative but also of the source of that utility.

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

For a final illustration Figure 3 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 $p - 5 = d$. 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. This is in line with the observation from Figure 2 that a one unit increase in price requires a discount increase of more than one unit in order to keep choice probabilities constant.

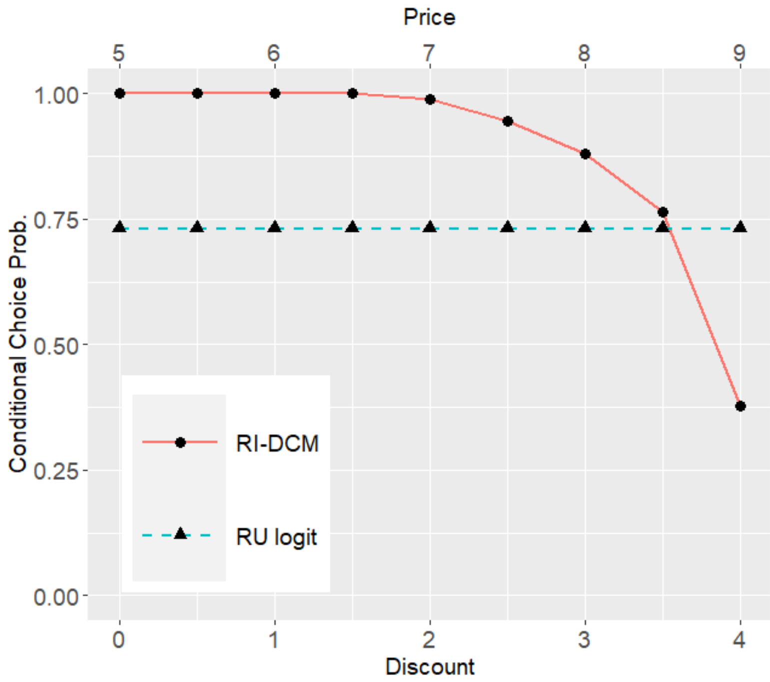

Figure 3: **Conditional choice probabilities for different compositions of the inside alternative under RI-DCM and RU logit.** This Figure depicts the conditional choice probability for the inside good in the discount under the constraint that the price increase equals the discount increase with $p - 5 = d$ 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.

Extant Literature Statistical models for attribute inattention were presented in Gilbride et al. (2006) and Yegoryan et al. (2020) and are based on the idea of shrinkage estimation. A key difference between those models and RI is that the attention allocation in the former is implicitly assumed to be constant across different choice sets whereas in RI the attention allocation adapts to the choice context. As the values of the simple attributes change from choice set to choice set, the optimal deployment of attention adapts. Regarding attribute interactions, information integration theory (Anderson, 1981, 1982) studies how the formation of overall judgments may depend on attributes in a non-additive way.[24] However, despite its name, information integration theory does not address how DMs apply limited cognitive resources adaptively and is not motivated from constrained optimization behavior.

4.1.3 Inattention to Alternatives

A key feature of discrete choice under RI is the endogenous formation of consideration sets which are defined as the set of alternatives that are chosen with strictly positive probability. As mentioned previously, consideration sets arise as a way to simplify the information processing task, and realized simple attribute values in a choice set influence which alternatives are considered.

Figure 4 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 $\beta_{b,i} \in \{2, 1.75, 1.5, 1\}$, $p_i = 2$, a discount d_i with $\Pr(d_i = 0) = \Pr(d_i = 2) = 0.5$, and $\lambda = 0.5$. The specific values for $\beta_{b,i}$ are chosen for illustration

[24] See also Lynch (1985).

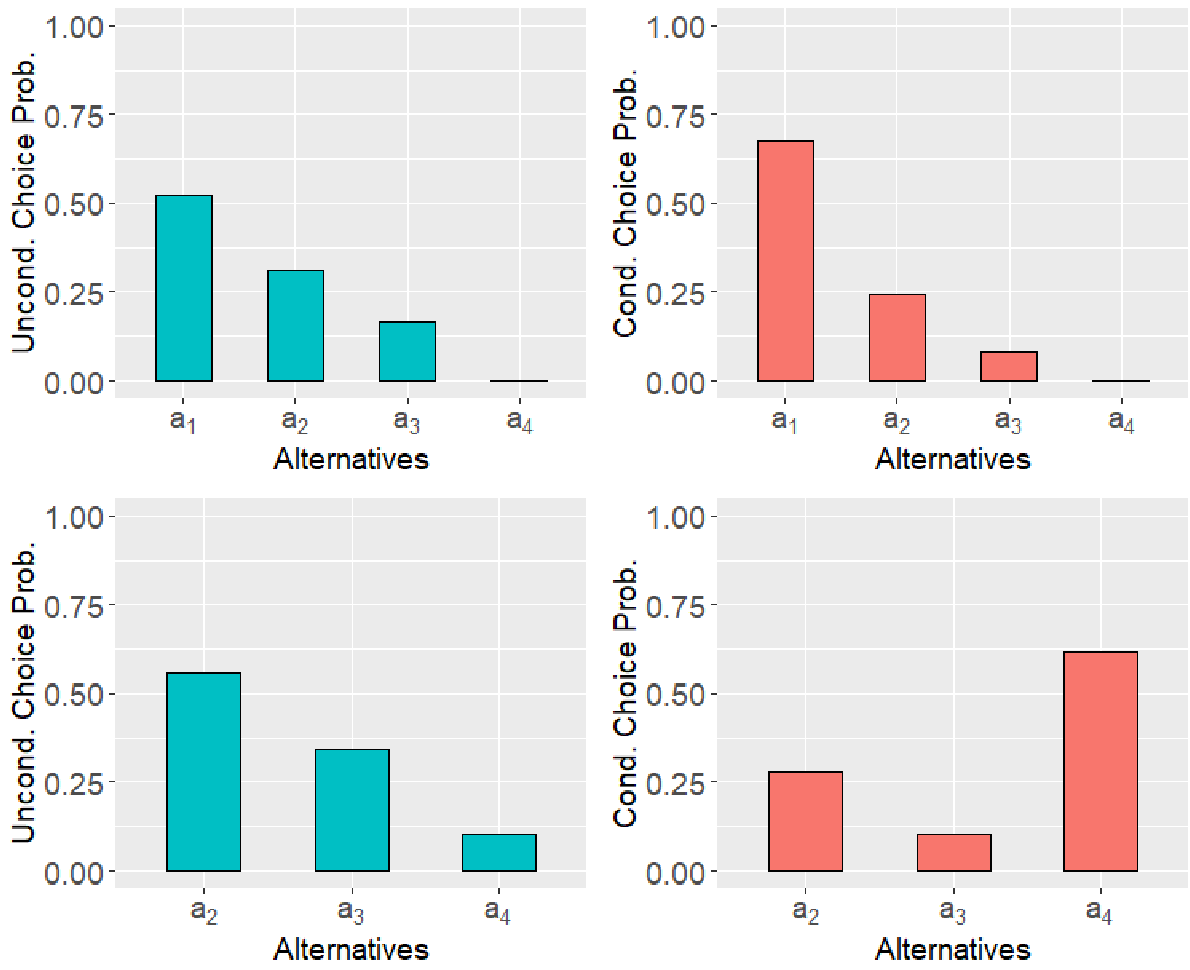

Figure 4: **Consideration set formation.** The above panels present choice probabilities of a rationally inattentive DM facing up to four alternatives $\{a_1, a_2, a_3, a_4\}$ ordered from highest to lowest simple brand $\beta_{b,i}$ with identical prices p_i and identically distributed complex discounts 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.

purposes. While $\beta_{b,4}$ represents the least preferred brand alternative, we can show how DM's react qualitatively differently to this brand as the composition of the choice set changes. The top-left panel of Figure 4 exhibits 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 instead conditions on a specific choice set, i.e., specific realized values of the complex discount (from the perspective of the analyst as the DM will not necessarily find out the exact nature of the choice set because of processing costs). In this choice set, the complex discounts are $d_1 = d_2 = d_3 = 0$ and $d_4 = 2$ so that alternative a_4 provides the highest utility in that choice set as $u_4 = 1 > u_j$ for $j \neq 4$. Even though a_4 provides the highest utility, the DM fails to consider this alternative based on a prior that determined that always excluding a_4 is beneficial given processing costs and the small prior probability that a_4 in fact is the utility maximizing choice in this design.

However, and different from extant two-stage models, RI predicts that a_4 will indeed be considered once a_1 is no longer available in the design (see the unconditional choice probabilities in the bottom-left panel of Figure 4). 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 pay-off (bottom-right panel).

Thus, RI implies deterministic consideration sets conditional on λ, prior beliefs μ_s and a utility function, while choice given such a consideration set is stochastic. However, because prior beliefs μ_s are function of the configuration of realized simple attribute values in a choice set, consideration sets will generally change from choice set to choice set in ways than can not be captured by an alternative specific index or decision rule.

Extant Literature: Descriptive Methods A variety of more descriptive approaches to modeling consideration sets known as two-stage models of choice have been proposed in the literature. For example, Terui et al. (2011); Goeree (2008); Ching et al. (2009) and Ching et al. (2009) model consideration as a function of alternative or brand specific advertising. Bronnenberg and Vanhonacker (1996) model consideration as a function of brand and shelf space and Andrews and Srinivasan (1995) as a function of price. Gilbride and Allenby (2004) model consideration as a flexible function of the attribute configuration of alternatives.[25] RI departs from these approaches in that it explicitly motivates consideration sets from constrained optimization. As a consequence, and again different from these approaches, consideration sets become implicit functions of the configuration of realized simple attribute values in a choice set. An important implication is that the consideration status of an alternative in a choice set can change even if nothing about this particular alternative changed (as we demonstrated in Figure 4).

Extant Literature: Structural Consumer Search Structural models of consideration set formation are typically consumer search models. In these models, the DM chooses among several alternatives that are characterized by two types of attributes. The DM observes a subset of attributes, similar to the simple attributes in the present paper, at no cost. Other attributes, loosely related to the complex attributes in this paper, have values that are ex ante unknown to the DM. However, the DM has some beliefs about these. Common examples of the former are brands and of the latter are prices or match values.

The DM can, at some search or learning costs, resolve the uncertainty from unknown attribute values for individual alternatives. Typically an all-or-nothing (AON) learning protocol is used for tractability. That is, at some fixed search cost the DM becomes perfectly informed about the value of the uncertain attribute of a single alternative. For instance, a consumer can choose to spend the time needed to learn the price of a specific product in an online store. Thus, the DM trades-off the costs and benefits of becoming informed about the uncertain attributes. In settings with more than one alternative, the analyst has to impose whether DMs search sequentially or simultaneously, i.e., whether they can commit to searching a fixed sample, or not.

Early contributions employing search models are Hauser and Wernerfelt (1990), who use a sequential search model, and Roberts and Lattin (1991) leveraging a simultaneous search

[25] See Aribarg et al. (2018) for an overview of this literature.

framework. Recent contributions mostly assume that the source of uncertainty are either prices or match values. In models with price uncertainty, the beliefs over prices are assumed to follow a specific distribution, e.g., normal or T1-EV distribution, for tractability (Mehta et al., 2003; Honka, 2014; De los Santos et al., 2012; Honka and Chintagunta, 2017). Moreover, these models generally assume that there is an error term that shifts utility and that is only observed by the DM but not by the analyst (e.g., Honka, 2014; De los Santos et al., 2012).[26] A different set of papers deals with consumer search in differentiated markets where consumers are uncertain about the match values of the available alternatives, but not about the price (e.g., Kim et al., 2010, 2017; Moraga-González et al., 2021). Some recent work allows for multi-dimensional uncertainty, e.g., simultaneous price and match-value uncertainty (e.g., Chen and Yao, 2017; Yao et al., 2017), albeit under assumptions that allow to apply established solution methods from (sequential) search theory.[27]

An important difference between search models invoking AON learning and choice under RI is that choice is deterministic under AON without invoking aspects of utility that are only observed by the DM. For this reason, an error term as part of the utility index is required to reconcile all possible data with the imposed choice model. Based on whether the DM has to learn the value of the error term, we distinguish two cases: In the first case, the error term is known to the DM before search for other attribute values. The resulting model exhibits stochastic consideration sets and choice that is stochastic given such sets from the analyst's perspective. Moreover, when the error has full support, all alternatives are chosen with a strictly positive probability even at extreme values of simple attributes, e.g., extreme prices. When estimating such a model conditional on observed consideration sets, which is common in online markets, further identification issues may occur.[28]

In the second case, the error term must be learned by the DM at some fixed cost. Then, the resulting model has deterministic consideration sets when search is simultaneous, and stochastic consideration sets when search is sequential, while choice is stochastic in both cases. Moreover, the impact of the ex ante uncertain attribute on choice is the same across all levels of the simple attributes where learning takes place (just because the uncertain attribute is learned perfectly under AON, if at all). This is in contrast to RI where the impact of complex attributes adjusts incrementally to changes in the simple attributes, leaving room for aspects of cognitive processing that extend beyond knowing or not knowing the level of a particular attribute, e.g., the effort of translating and integrating levels into the overall utility (see Figure 1). Another noteworthy characteristic of this variant of the AON model is that the resulting demand function has jump discontinuities in the simple attributes and their coefficients. Finally, a common assumption is that the DM cannot choose an alternative without searching it. In RI, a DM can choose purely on the basis of prior information.

[26] De los Santos et al. (2012) even assume a further "consideration set" error term.

[27] For a recent overview of search models in marketing see Honka et al. (2019).

[28] Suppose the DM searches for prices and observes an error term not known by the analyst. For a simple illustration, suppose that the DM searches one inside good, learns that is has the lowest possible price and still chooses the outside option. Such behavior is impossible under the suggested error term specification with AON learning, while RI allows for choice errors given a consideration set.

4.2 Effects of DCE Design Variations under RI

4.2.1 Impact of Information Processing Costs and Incentives

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 (2)). The parameter λ captures the cognitive costs associated with processing complex attributes and thus of reducing the uncertainty about the alternatives' payoffs.

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.[29] For example, if an incentivized DCE instructs participants facing N choice tasks that one the the N choices becomes an actual transaction, the realization probability of actually obtaining a specific choice is simply $\rho = 1/N$. With probability $1 - \rho$ the DM's choice is hypothetical, i.e., she does not actually obtain the chosen alternative.[30] From the perspective of the DM, the resulting RI objective function for each choice task is given by

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

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 $p = 2$, as well as a complex discount d with $\Pr(d = 0) = \Pr(d = 2) = 0.5$, and an outside option that yields a payoff $u_O = 0$. Based on expected utility, the DM thus prefers the inside good. Figure 5 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 (see the bottom left and right panel of Figure 5), which of course implies that choice probabilities no longer change as a function of the complex discount (see the top right panel of Figure 5). At $\lambda = \lambda' = 0$, the DM perfectly learns the complex discount and

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

[30] Recall that in an incentive-aligned DCE, the DM is endowed with a budget $R > 0$. In case of choosing the outside option, i.e., not purchasing any of the inside alternatives, the DM retains the endowed budget R.

deterministically chooses the alternative with the highest utility, and hence maximally reacts to changes in the complex discount value.

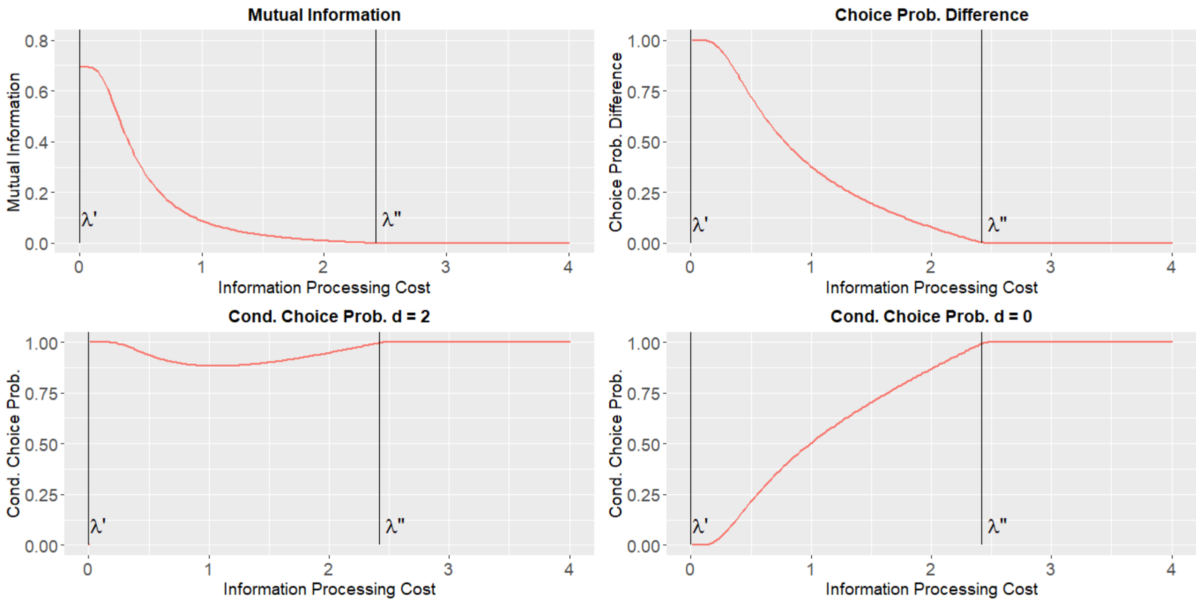

Figure 5: **Impact of information processing costs on attention, discount effect, and choice.** The upper left panel shows mutual information as a function of information costs. The upper right panel displays the impact of a fixed increase of the complex discount (from 0 to 2) on conditional choice probability for different levels of λ. The panels in the bottom row show conditional choice probabilities of the inside alternative in a 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''.

Note that changes in λ can impact what is revealed about the DM's preferences. The bottom-right panel of Figure 5 shows that as λ increases, 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 state ω). Finally, the bottom-left panel of Figure 5 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. 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 6 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 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 (from the viewpoint of a RU logit) that is reflected in absolutely smaller brand and price coefficients.

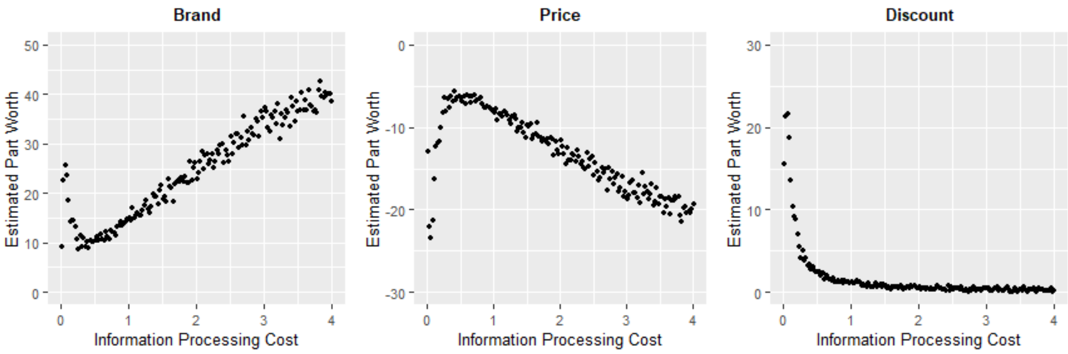

Figure 6: **Logit approximation for different levels of information processing costs.** Each panel shows logit estimates of the respective coefficients for varying levels of information processing costs λ. Note that each point is the result of an estimation from simulated data with $N = 1,000$ choice 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 $\Pr(d = 0) = \Pr(d = 2) = 0.5$.

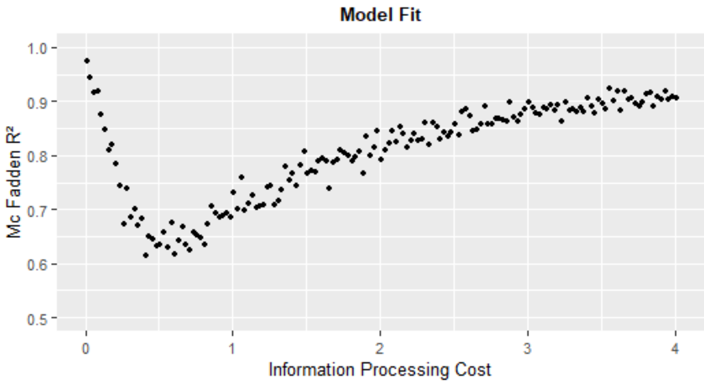

Figure 7: **Choice consistency is non-monotonic in the information processing costs λ.** Choice consistency is measured by McFadden's pseudo R-squared. For details see Domencich and McFadden (1975).

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 Figure 6. It is common in the choice modeling literature to report the estimated error term variance and typically to interpret this as choice consistency (e.g., DeShazo and Fermo, 2002). We follow this practice in Figure 7 which exhibits the corresponding McFadden's pseudo R-squared values. Figure 7 confirms that RI choices are more deterministic at very low and very high values of λ, and less deterministic at intermediate values.

Figure 8 extends the illustration of RI choice as a function of λ to the case of three alternatives.[31] Alternative a_i, $i = 1, ..., 3$, yields utility $u_i = \beta_{b,i} - p + d_i$ 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 8 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

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

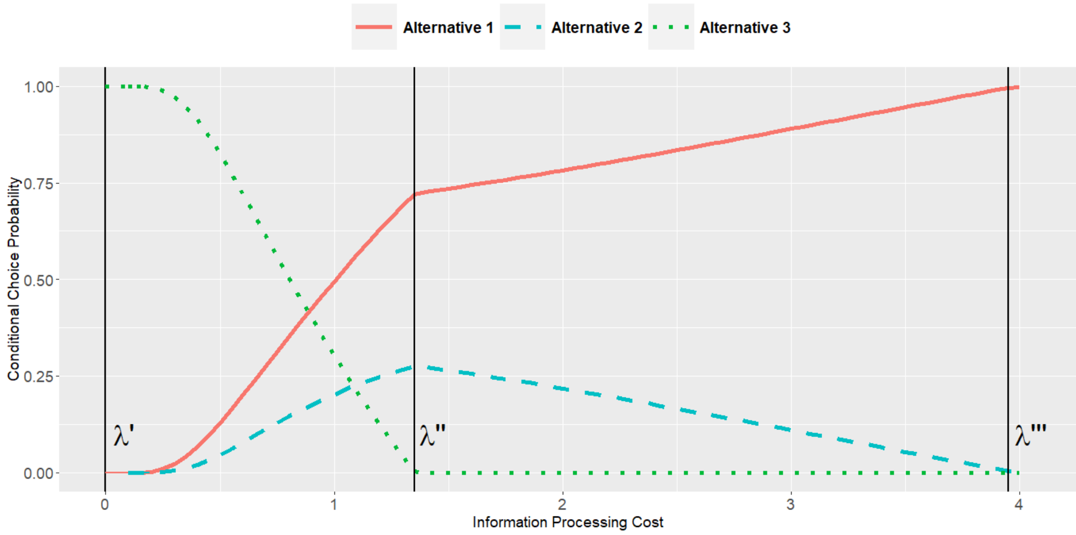

Figure 8: **Impact of information costs on consideration set size.** Conditional choice probabilities for three alternatives are displayed as a function of information processing costs λ. As information costs increase, the DM rationally chooses to ignore alternatives in the choice set. λ', λ'' and λ''' 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.

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 alternatives provide the highest payoff (as depicted in Figure 8). Together, the illustrations in this section suggest that RI provides a useful basis for bridging across choices under different incentive levels or different levels of difficulty of processing complex attributes.

Extant Literature The majority of the extant experimental work suggests that choice becomes less consistent, that is, the scale of the error term increases, as complexity increases (e.g., DeShazo and Fermo, 2002; Dellaert et al., 1999; Mazzotta and Opaluch, 1995). For instance, DeShazo and Fermo (2002) find that most of their complexity measures are negatively correlated with choice consistency, except when complexity is measured by the number of available alternatives. However, Swait and Adamowicz (2001) show that as the complexity in a choice task increases, DMs tend to change the choice rule they apply. In particular, in their experiment brands, which are salient attributes, become more important for choice as the complexity increases. Relatedly, applied researchers mention the possibility that DMs may behave in a consistent fashion when facing complex tasks by applying a "fast" simplifying rule (Orme, 2019).

It seems that the notion of complexity in choice is complex itself as evidenced by the different operationalizations of choice complexity in the literature. Early studies used the number of attributes or the number of alternatives as a measure for complexity (e.g., **?**Keller and Staelin, 1987). An alternative set of complexity measures relates complexity to the structure of the specific choice set faced by the DM given a particular experimental design. Examples include the number of attributes that differ across alternatives of a choice set (Mazzotta and Opaluch, 1995), the correlation of (normalized) attribute levels within and across alternatives in a choice set (DeShazo and Fermo, 2002), or the amount of entropy (Swait and Adamowicz, 2001). In fact, what may make a choice task more or less complex is somewhat endogenous to how a particular researcher thinks about choice.

A basic consensus seems to be that complexity results in the processing of less of the available information, or less effective processing of information. RI then implies that the complexity of a given choice task (as measured by the amount of information processed in that task) is confounded with what the DM gains from processing complex information. RI also implies that a monotonic mapping from realized stochasticity to complexity is likely misleading because of the non-monotonic relationship between the amount of information processed and the stochasticity of the resulting data under RI, in line with the findings in Swait and Adamowicz (2001) and the conjectures about choice simplification in Orme (2019).

However, the RI parameter λ maybe conceived of as a general reflection of the difficulty of extracting and integrating the information in complex attributes into an overall assessment of payoffs. For example, the number of alternatives in a choice task affects the optimal processing strategy, and hence choice behavior under RI, but not the unit cost of information processing. In fact, the information processing cost λ should remain constant in the expansion of the available costly information as long as the DM is able to ignore newly added pieces of information so that that the resulting simplified problem is identical in terms of payoffs and beliefs to the initially smaller choice problem. However, it is certainly possible that simple information may cease to be simple—in the sense of immediately processed to form a prior that guides further processing—as e.g., choice sets expand dramatically (e.g., Natan, 2021).

Finally, the literature on incentive-aligned DCEs finds that incentive-aligned experiments result in better out-of-sample predictions in real-world applications (Ding et al., 2005; Ding, 2007; Dong et al., 2010). The discussion in this subsection provides an information economic based explanation for this observation (see also Cao and Zhang (2021) who propose a structural model of incentive-aligned choice on the basis of a simple framework of individual learning and show that a higher realization probability increases price sensitivity).

4.2.2 Attribute Range/Dispersion and Levels Effects

Here 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 increases such that processing information becomes 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. At the end of this subsection, we briefly discuss the impact of changing the number of levels of an attribute.

Figure 9 illustrates this mechanism in our leading car example. Recall that $\lambda = 0.5$, $\beta_b = 6$ and $\beta_p = -\beta_d = -1$. Here, we study the impact of an increase of the discount from 2 to 3 for different discount ranges on conditional choice probabilities. Both lines in Figure 9 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 $\{0, 2, 3, 5\}$. In both cases, the DM's beliefs are uniform over the respective support. Figure 9 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.

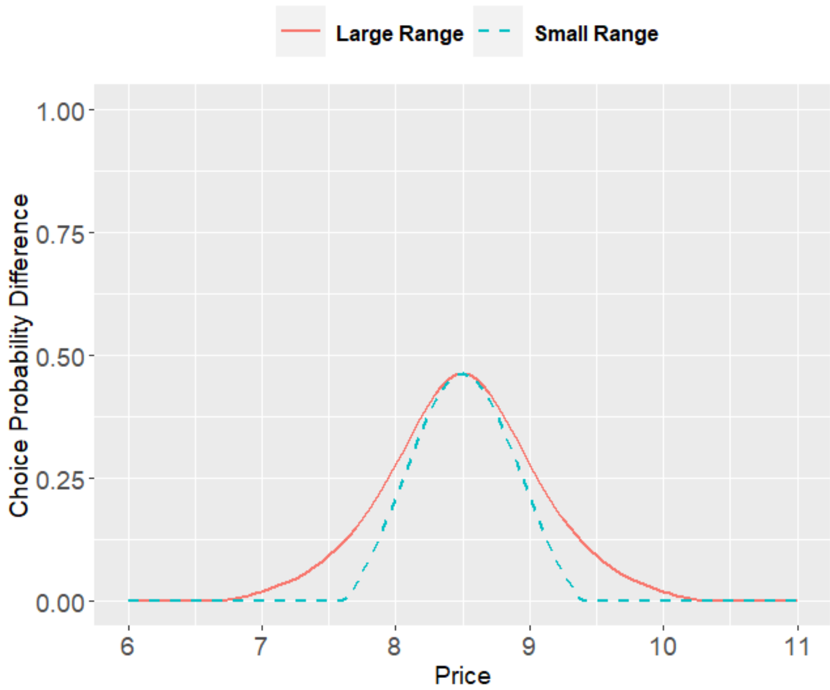

Figure 9: **Range effects of complex attributes.** This figure displays conditional choice probability differences of the inside alternative in reaction to an increase of the discount from 2 to 3 given a large range (solid red) and a small range (dashed blue) of the complex discount as a function of 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.

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: $\lambda = 0.5$, $\beta_b = 8$, $\beta_d = 1$, $\beta_p = -1$, $p \in [8, 10]$, and d is drawn with equal probability from the binary set $\{4 - x, 4 + x\}$ with $x \in [0.1, 4]$. Figures 10 and 11 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 of the complex discount increases, brand and price coefficients become smaller in absolute value, and the inferred discount coefficients tend to increase. Finally, Figure

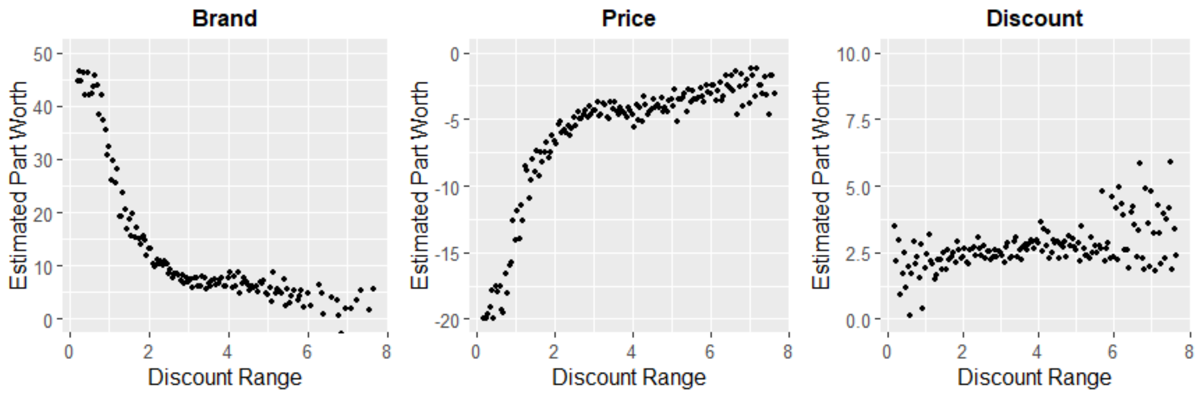

Figure 10: **Logit estimates for different levels of the discount range.** This figure displays estimated coefficients of brand, price, and discount for different ranges of the complex attribute. Each point is the result of fitting a logit model with simulated data from $N = 1,000$ choices with data generating parameters $\lambda = 0.5$, $\beta_b = 4$, $\beta_d = 1$, $\beta_p = -1$, and p being uniformly drawn from $[8, 10]$. The discount range, given as the difference between the two discount levels, is varied from 0.2 to 8.

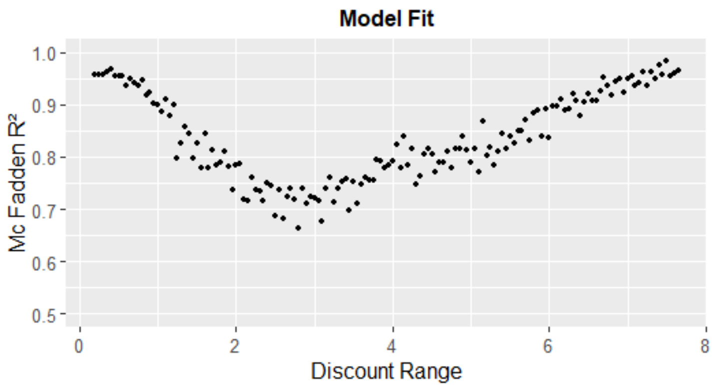

Figure 11: **Non-monotonic effect of the discount range on choice consistency.** Note that choice consistency is measured as McFadden's pseudo R-squared. For details see Domencich and McFadden (1975).

11 exhibits a U-shaped relationship between choice consistency and the range of the complex discount.

When the range of complex discounts is very small, the DM makes smaller mistakes when choosing based on prior probabilities, 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 of the complex discount increases, the mistakes when choosing based on prior probabilities, conditioned on the simple attributes brand and price in this example, can become large. 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 before adding a potentially very large number of interaction effects between price and discount.[32] Such interaction effects would pick up, in a descriptive manner, that the DM strategically pays more or less attention to the discounts with a medium range, depending on realized values of simple attributes (here prices).

[32]The quality and the feasibility of approximations to the RI-DCM based on high-dimensional, non-linear indices in a logit model is beyond the scope of this paper.

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.

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 affects. 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 the corresponding complex attribute. This intuition applies to both adding in-between levels to simple and to complex attributes. For example, if adding an in-between level of a simple attribute, overall, decreases (conditional on realized simple attributes) ex ante uncertainty, the relative importance of the simple attribute will increase when viewed through the lens of a descriptive RU logit model. If doing the same for a complex attribute increases (conditional on realized simple attributes) ex ante uncertainty, the relative importance of the complex attribute will increase in a (descriptive) RU logit model fit to the choice data.

Extant Literature The experimental literature on the sign of range effects, that is whether a wider range of an attribute increases or decreases its impact on choice, is mixed (see Bestard and Font (2021) for a recent overview of the relevant literature). A typical finding is that choice becomes less consistent in the range of attributes (e.g., Dellaert et al., 1999). The proposed explanation is that a wider range makes comparisons more difficult which in turn results in a larger choice error variance. Our simulation results show how the RI-DCM can recreate behavior that resembles parts of the documented empirical choice data, however, the mechanism is different, as it relates to the adaptive processing of complex attributes.

Other work in this domain relates the attribute range to the non-additivity of attributes in the utility index of a choice model. For example, Ohler et al. (2000) show in a binary choice experiment that ranges appear to have little to no effect on the size of attribute coefficients or the error term scale. However, wider attribute ranges are likely to improve the power to detect non-additivity of attributes. This is consistent with the results in Dellaert et al. (1999) since one possible explanation for the choice consistency decrease in that study is that non-linearities play a larger role with wider attribute ranges. It is also consistent with RI in the sense that going from a small range to a medium range of a complex attributes, RI implies interactions between simple and complex attributes in a descriptive model fitted to the RI data.

4.2.3 Price Image and Attribute Correlation

RI theory can incorporate the notion of price images with, e.g., brand specific prior beliefs about complex price components. The mechanism via prior beliefs of course generalizes to all kinds of prior beliefs about complex attributes that may depend on the realization of simple attributes, e.g., that higher simple prices are associated with better (complex) quality aspects in expectation in a DCE (see Erickson and Johansson (1985) for an empirical example).

Here, we illustrate 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 inside and an outside good under the parameters $\lambda = 0.5$, $p = 2$, $\beta_p = -1$, $\beta_b = \beta_d = 1$, and d uniformly 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 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.

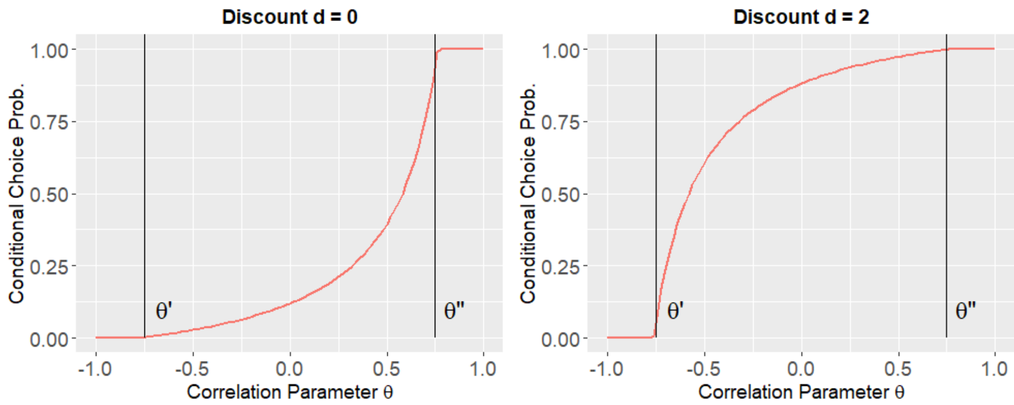

Figure 12: **Impact of brand-specific discount beliefs on choice.** Conditional choice probabilities are displayed as a function of the correlation coefficient θ for choice sets with discount levels $d = 0$ (left) and $d = 2$ (right). As θ varies, the prior belief structure of the DM changes which directly 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.

Figure 12 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. Notably, absolutely large correlations between the complex discount and the inside brand translate into deterministic choice behavior that ignores the actual state ω, and well before that state 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 $p = 2.75$, $d_k \in \{0, 2\}$ and learning costs $\lambda = 0.5$. 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 13 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 or states ω. Overall, as the

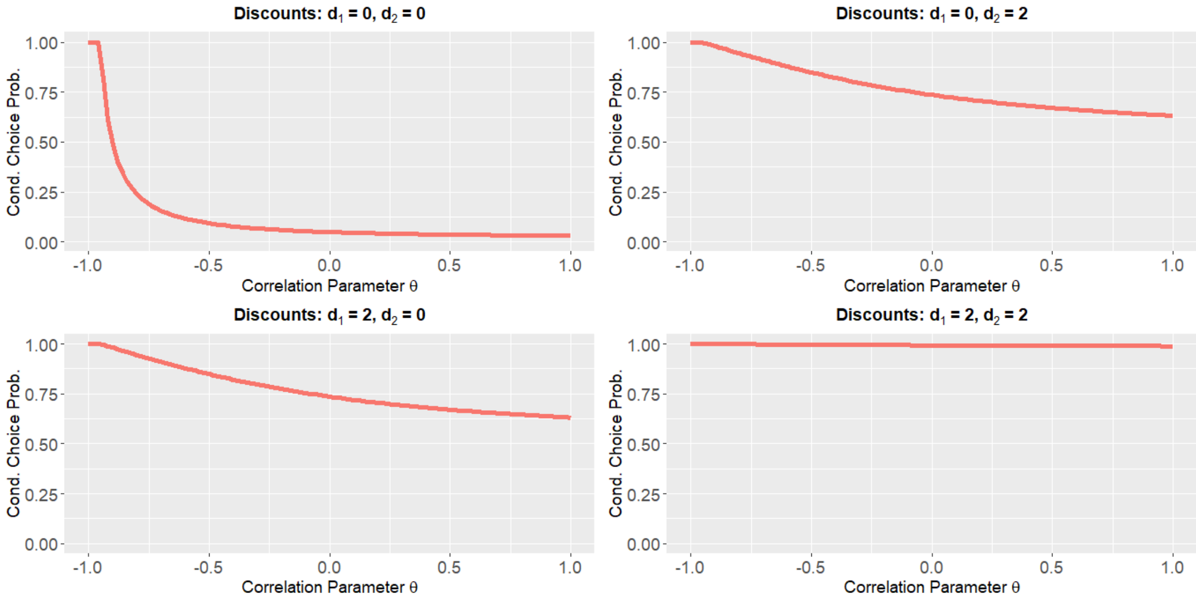

Figure 13: **Impact of brand specific beliefs about discounts on choice.** Conditional choice probabilities of the inside good with different complex discount levels are displayed as a function of the correlation parameter θ. The panels differ in the value of the discounts d_1 and d_2. Note that in all choice sets the DM chooses the inside good deterministically once the correlations becomes sufficiently low while being interior.

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, such that 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 less likely to choose the utility maximizing inside option when only one discount applies under this prior.

Our two examples here are simple for expositional reasons but suggestive of how the RI framework can take account of belief dependencies within and across choice alternatives. Naturally, this raises the question of how the analyst can learn about the beliefs of consumers outside of the realm of beliefs induced by the design of a DCE for counterfactual analyses. We revisit this point in Section 5.

Extant Literature A related class of models accounting for DMs' beliefs are search models where beliefs about the distribution of some characteristics of the available offerings determine whether an alternative is searched. Closest to the present setting are search models with uncertainty over prices. Search models typically make the assumption of rational expectations, that is, consumers are assumed to have the price beliefs that equal the true (equilibrium) distribution of prices in the market. More recent contributions weaken this assumption by allowing for subjective beliefs and augment the analysis with auxiliary data that informs the analyst about (heterogeneous) consumer beliefs (see Jindal and Aribarg (2021) for an overview of the relevant literature).

In marketing, there is an extensive literature studying price image and the related concept of reference prices for brands and stores (see Hamilton and Chernev (2013) for a review). For instance, Lourenço et al. (2015) study how price changes in categories with variations in characteristics, such as discount depth or discount frequency, affect the store price image by combining scanner-panel data with survey data on the price image of visited stores. However, most of this literature investigates the determinants of price image, and papers that study the behavioral implications of price image (beyond intentions) are rare (Lombart et al., 2016), in particular, in the context of DCEs.

Lastly, an older literature questions if orthogonal designs in DCEs and generally conjoint experiments at odds with ecological dependencies between attributes, bias the inference from such experiments (e.g., Moore and Holbrook, 1990). The RI framework usefully allows the researcher to distinguish between beliefs induced by a particular experimental design and beliefs that may apply outside of the realm of a particular experiment.

4.2.4 Choice Set Expansions

Next, 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 is expanded 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 varies as well. 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.[33] 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 $u_i = 3 - p_i + d_i$. The information costs are given by $\lambda = 0.25$. 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 7 summarized the logit fits.

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 no matter the value of the discount. However, with a second inside good,

[33] 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.

No. Alt.	Brand 1	Brand 2	Price	Complex Discount	McFadden R^2
2	14.11		-11.37	4.29	0.69
	(0.41)		(0.30)	(0.12)	
3	12.01	11.88	-9.10	3.98	0.38
	(0.31)	(0.30)	(0.21)	(0.10)	

Table 7: **Logit approximations for different choice set sizes.** 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 (second row) alternatives. The final column reports corresponding McFadden R-squared (Domencich and McFadden, 1975) as a measure for choice consistency. Note that the added alternative is identical to the already existing inside alternative.

also the price of the second alternative p_2 will determine the impact of d_1 on choice. However, the logit approximation 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.[34] Intuitively, the decision maker can choose to optimally ignore alternatives that are inferior to the extant alternatives of a choice set without any additional costs.[35] As such, the RI framework allows the analyst to take a stance on the impact of choice set expansion on welfare since it relates the observed error to optimal decision making under cognitive constraints.

Extant Literature Note first that standard RUMs impose that the error term variance is independent of the size of the choice set. In contrast to this assumption there is empirical evidence showing that the error term variance appears to be increasing in the choice set size which has been interpreted as the result of a latent cognitive process that capture a DM's struggle with an increasing complexity due to a higher information load in the past.

A few extant papers account for the impact of the choice set size by relating it to the scale of the error term (which is the inverse of the error term variance) in a heteroscedastic logit model. In stated choice experiments, both DeShazo and Fermo (2002) as well as Caussade et al. (2005) present evidence of an inverse U-shaped relationship between choice consistency and the

[34] 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.

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

number of alternatives in choice sets with the maximum being attained between three and four alternatives. While we cannot explain the at first increasing choice consistency, RI can motivate decreasing choice consistency as result of a choice set expansion.[36]

For our second observation, i.e., that the choice set expansion decreases the coefficient of the simple price more than the coefficient of the complex attribute (when viewed through the lens of a descriptive RU logit), we refer to that part of the experimental discrete choice literature that shows how the choice set size affects the processing of information. For instance, Hensher (2006) shows that as the number of alternatives increases also the number of attributes considered by DMs increases while the findings in Meißner et al. (2020) indicate by using eye-tracking that DMs adjust their information processing strategy to the number of alternatives in the choice set. In particular, their results suggests that the number of non-attended attributes increases in the number of alternatives. We note that under RI both an increase in attention to attributes as well as a decrease can be optimal depending on realized simple attributes and the conditional distribution of complex attributes.

5 Conclusion and Outlook

In this paper, we have presented how to adapt a RI-DCM to the multi-attribute multi-alternative setting that is typical of marketing applications and, in particular, discrete choice experiments. We illustrated that a hierarchical version of our model can be calibrated from "small T, large N" data, as typical of marketing applications. Our illustration suggests that small T, large N data generated from the proposed RI-DCM will reliably distinguish the data-generating model from a standard hierarchical logit model. Finally, we have shown how a series of behavioral patterns that individually require qualitatively different modifications of the standard RU logit can be jointly nested by a discrete choice model rooted in RI. We conclude this article by providing an overview of open challenges to turning the proposed RI-DCM into a workhorse for applied researchers.

5.1 Open Empirical Challenges: Statistical and Economic Identification

The RI framework we propose for the 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, 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 our model does not have a closed-form solution. We rely on the Blahut-Arimoto algorithm for numerical solutions to the RI choice problem that define the likelihood (see Caplin et al., 2019; Csaba, 2018).[37] The Blahut-Arimoto algorithm solves a fixed point problem. This makes the estimation procedure computationally costly for datasets with large numbers of

[36] There is the possibility that the at first increasing consistency may be a statistical artefact related to properties of the RU multinomial logit likelihood when calibrated based on finite data.

[37] Cover and Thomas (1991) provide an introduction to the Blahut-Arimoto algorithm.

respondents or repeated choices per respondent. Thus, it will be important to compute solutions for repeated measures and respondents in parallel for industry-grade applications. Relatedly, we are currently investigating an estimation strategy that builds on automatic differentiation of individual-level objective functions in connection with Hamiltonian-Monte-Carlo sampling as implemented in `Stan` (Stan Development Team, 2022).

Another practical challenge of the proposed framework is the identification of simple and complex attributes. If the set of simple and complex attributes are known a priori, preferences for both groups of attributes can be identified up to a proportion similar to, e.g., preferences in the RU logit. 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 many applications, however, 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 and Jeon, 2021).

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. While 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 separability in utility contributions can suffice to distinguish between simple and complex attributes empirically.[38]

We simulate 2,000 choice tasks in each of which a DM chooses between an inside and an outside good. The inside good is characterized by two linear attributes that additively combine into overall utility. One attribute is simple, and the other is complex. Each of the two linear attributes is represented by three levels in the experimental design. Attribute combinations defining inside goods are drawn from independent uniform distributions over the discrete attribute support points. With the simulated data, we estimate two 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 is simple and what is complex in estimation. We present the quantiles of the log-likelihoods from Markov-Chain-Monte-Carlo (MCMC) estimation for different data generating levels of λ for both specifications in Table 8.[39] The summaries in Table 8 suggest that there is scope for likelihood-based identification of what are simple and complex attributes from choice data, and specifically for intermediate values of λ (relative to given utility parameters) that render choice probabilistic.

[38]We demonstrate that the RI-DCM with simple and complex attributes can be empirically distinguished from the special case where all attributes are complex (i.e., the RU logit) in Sections 3.3 and 4. Also note that the special case of only simple attributes results in deterministic choice.

[39]The respective log-likelihood traces 10,000 draws (after convergence) are available in Figure 15 in the Appendix.

When λ (relative to given utility parameters) becomes very small or very large, choice becomes deterministic and the data only set-identify utility parameters (see Figure 14). When λ becomes smaller and smaller, all information is fully processed, and the conceptual and empirical distinction between simple and complex vanishes. When λ becomes larger and larger, the information in complex attributes is never integrated into the overall evaluation of alternatives. When the analyst falsely specifies complex attributes as simple and simple attributes as complex in this case, 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.

λ = **0.01**					
Model	min	25%	50%	75%	max
Correct	-5.88	-0.03	-9.77x10^{-4}	-9.46x10^{-6}	-2.5x10^{-13}
Wrong	-5.71	-0.07	-2.37x10^{-3}	-1.29x10^{-5}	-9.55x10^{-15}

λ = **0.5**					
Model	min	25%	50%	75%	max
Correct	-518.18	-510.46	-508.60	-504.05	-503.46
Wrong	-541.94	-538.48	-536.71	-531.15	-530.52

λ = **5**					
Model	min	25%	50%	75%	max
Correct	-4.79	-0.03	-7.88x10^{-3}	-2.7x10^{-3}	-2.43x10^{-4}
Wrong	-5.40	-0.58	-0.03	-8.01x10^{-2}	-4.77x10^{-4}

Table 8: **Quartiles and min/max of Log-Likelihood MCMC draws for different λ levels for the correct and wrong specification respectively.**

While this example builds on the utility function's linearity and additive separability of attributes, we drop the assumption of linearity in Table 9 in the Appendix. We find that even arbitrarily non-linear relationships between attribute levels and utility cannot compensate for misspecifying what is simple and complex at an intermediate level of λ. We thus conjecture that the additive separability in the payoffs is sufficient for identification.

The final open question we bring forward is related to heterogeneity. Specifying hierarchical distributions for the parameters of the model facilitates estimation in the predominant "small T, large N" setting with heterogeneous DMs. Under the assumption of homogeneous rational prior beliefs and homogeneity in what is simple and complex about choices in a specific domain, there are two sources of heterogeneity in the model. One source of heterogeneity is the vector of preferences; the other is the information cost that scales the preference vector. Assuming Shannon costs, we can likelihood-identify the ratio β/λ for each DM in the data. However, in a hierarchical model it may be both theoretically appealing and efficient to structure heterogeneity in β as residual heterogeneity after taking heterogeneity in λ into account. 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.

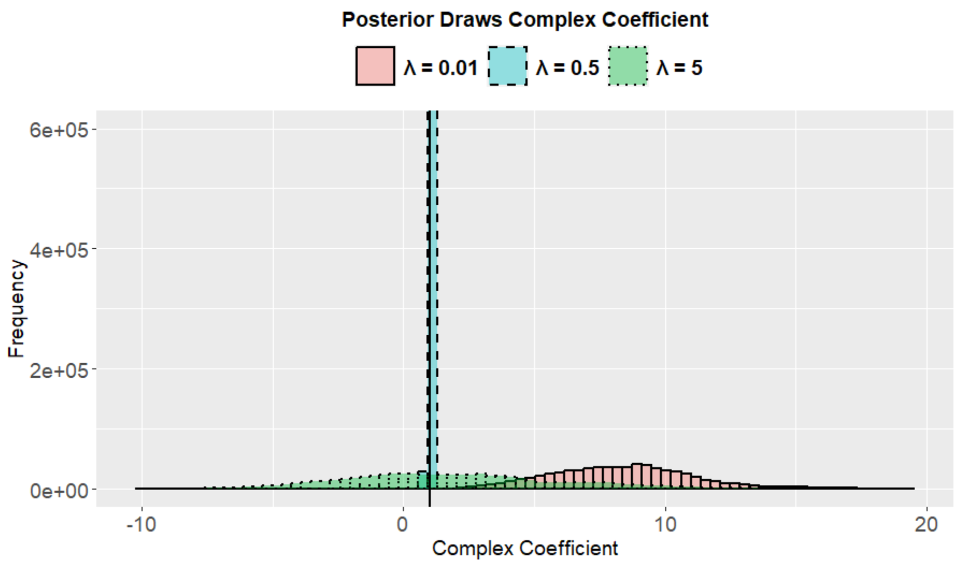

Figure 14: **Histogram of complex attribute MCMC estimates for different information processing costs λ.** This figure shows histograms of the posterior distribution of complex attribute coefficient estimates using 700,000 draws from the implied marginal posterior. The solid vertical line indicates data generating value. Note that we truncate the figure along the y-axis from above for illustration purposes.

5.2 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 experimental 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, 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.

5.3 Consumer Belief Extraction and their Role for Market Simulations

Prior beliefs play an important role for information search 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 (2021) 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 full information about prior beliefs as conveyed to DMs by the experimental design. However, this is no longer the case, even for the same DMs outside a particular experimental setting. For counterfactuals, e.g., market simulations, a key question is thus how to assess (likely) beliefs in the market setting to be simulated.[40]

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 empirically study potentially heterogeneous market beliefs. 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 and 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., 2021). For a recent overview in marketing see the literature section in Jindal and Aribarg (2021).

5.4 Alternative Information Cost Functions

There is an ongoing discussion in the economics literature on the question 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 experimental literature dealing with RI (implicitly) test 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.

For instance, experimental evidence suggest that non-linearities of information costs exist, where subjects pay too little attention to high rewards as compared to low rewards (Caplin and Dean, 2013; Dean and Neligh, 2019). Perhaps more important for the analysis of MAMA choice in DCEs, a linear Shannon entropy cost function precludes the concept of perceptual distance, where some states are harder to identify than others (Dean and Neligh, 2019; Hébert and Woodford, 2021). Relatedly, this feature of Shannon costs also rules out that the location on

[40] 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.

a screen where informative signals appear (for experimental evidence of this effect see Woodford, 2012) or the ordering and similarity of choice options matter (Fosgerau et al., 2020).

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 address the problem that some states are harder to distinguish than others.[41] A detailed discussion of this mostly theoretical literature is beyond the scope of this paper.[42]

From our perspective, 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 the distinction 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 in the context of DCEs. 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.

5.5 Conceptualization of Free Information

Somewhat related to both the previous discussions of cost functions and the empirical identification of what are simple and complex attributes (sources of free and costly information), we conclude with a more general perspective.

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 rich 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.[43]

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 find it easy to recognize

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

[42] 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. (2021).

[43] See also Maćkowiak et al. (2021) who discuss the idea that RI provides a model for the formation of heuristics.

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 gives rise to 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. The previous discussion regarding the empirical identification of simple vs. complex attributes should be viewed as a special case of this broader 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 maybe 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 applications to industry-grade DCEs with many attributes in particular, this is a crucial part of future research.

6 Bibliography

Ambuehl, S., A. Ockenfels, and C. Stewart (2020). Who opts in? *Rotman School of Management Working Paper* (3154197).

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

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

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

Bertoli, S., J. F.-H. Moraga, and L. Guichard (2020). Rational inattention and migration decisions. *Journal of International Economics 126*, 103364.

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

Blattberg, R. and S. Neslin (1989, 12). Sales promotion: The long and the short of it. *Marketing Letters 1*, 81–97.

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

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

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

Brown, Z. Y. and J. Jeon (2021). Product proliferation under rational inattention: Application to health insurance. *AEA Papers and Proceedings 111*, 554–59.

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

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

Caplin, A. and M. Dean (2015). Revealed preference, rational inattention, and costly information acquisition. *The American Economic Review 105*(7), 2183–2203.

Caplin, A., M. Dean, and J. Leahy (2019). Rational Inattention, Optimal Consideration Sets, and Stochastic Choice. *The Review of Economic Studies 86*(3), 1061–1094.

Caplin, A., J. Leahy, and F. Matějka (2016). Rational inattention and inference from market share data. *Unpublished manuscript, New York University and CERGE-EI*.

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

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

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

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

Cheremukhin, A., A. Popova, and A. Tutino (2015). A theory of discrete choice with information costs. *Journal of Economic Behavior & Organization 113*, 34–50.

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

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

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

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

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

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

Cover, T. M. and J. A. Thomas (1991). *Elements of Information Theory.* Jon Wiley and Sons, Inc.

Csaba, D. (2018). Attentional complements. *Unpublished manuscript, New York University.*

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

De Palma, A., G. M. Myers, and Y. Y. Papageorgiou (1994). Rational choice under an imperfect ability to choose. *The American Economic Review*, 419–440.

Dean, M. and N. Neligh (2019). Experimental tests of rational inattention. *Unpublished manuscript*.

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

DeShazo, J. and G. Fermo (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., R. Grewal, and J. Liechty (2005). Incentive-aligned conjoint analysis. *Journal of Marketing Research 42*(1), 67–82.

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

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

Dotson, J. P., J. R. Howell, J. D. Brazell, T. Otter, P. J. Lenk, S. MacEachern, and G. M. Allenby (2018). A probit model with structured covariance for similarity effects and source of volume calculations. *Journal of Marketing Research 55*(1), 35–47.

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

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

Fosgerau, M., E. Melo, A. De Palma, and M. Shum (2020). Discrete choice and rational inattention: A general equivalence result. *International Economic Review 61*(4), 1569–1589.

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

Gardete, P. and M. Hunter (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*.

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

Gilbride, T. J., G. M. Allenby, and J. D. Brazell (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. and J. D. Little (1983). A logit model of brand choice calibrated on scanner data. *Marketing Science 2*(3), 203–238.

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

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

Hébert, B. and M. Woodford (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. (2014). Quantifying search and switching costs in the us auto insurance industry. *The RAND Journal of Economics 45*(4), 847–884.

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

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

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

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

Janssen, A. and J. Kasinger (2021). Obfuscation and rational inattention in digitalized markets. *SAFE Working Paper No. 306*.

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

Joo, J. (2022). Rational inattention as an empirical framework for discrete choice and consumer-welfare evaluation. *Journal of Marketing Research*.

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

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

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

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

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

Leong, W. and D. A. Hensher (2012). Embedding decision heuristics in discrete choice models: A review. *Transport Reviews 32*(3), 313–331.

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

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

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

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

Lynch, J. G. J. (1985). Uniqueness issues in the decompositional modeling of multiattribute overall evaluations: An information integration perspective. *Journal of Marketing Research 22*(1), 1–19.

Maćkowiak, B., F. Matějka, and M. Wiederholt (2021). Rational inattention: A review. *ECB Working Paper* (2570).

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

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

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

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

McFadden, D., M. J. Machina, and J. Baron (1999). Rationality for economists? In *Elicitation of Preferences*, pp. 73–110. Springer.

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

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

Moore, W. L. and M. B. Holbrook (1990). Conjoint analysis on objects with environmentally correlated attributes: The questionable importance of representative design. *Journal of Consumer Research 16*(4), 490–497.

Moraga-González, J. L., Z. Sándor, and M. R. Wildenbeest (2021). Consumer search and prices in the automobile market. *CEPR Discussion Paper* (DP16574).

Musalem, A., R. Montoya, M. Meißner, and J. Huber (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*.

Novák, V., A. Matveenko, and S. Ravaioli (2021). The status quo and belief polarization of inattentive agents: Theory and experiment. *IGIER Working Paper*.

Ohler, T., A. Le, J. Louviere, and J. Swait (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. Technical report, Sawtooth Software.

Pinkovskiy, M. L. (2009). Rational inattention and choice under risk: Explaining violations of expected utility through a shannon entropy formulation of the costs of rationality. *Atlantic Economic Journal 37*(1), 99–112.

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., V. Watson, H. Burnett, and W. J. Ungar (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. and J. M. Lattin (1991). Development and testing of a model of consideration set composition. *Journal of Marketing Research 28*(4), 429–440.

Rooderkerk, R. P., H. J. Van Heerde, and T. H. Bijmolt (2011). Incorporating context effects into a choice model. *Journal of Marketing Research 48*(4), 767–780.

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

Russo, J. E. and F. Leclerc (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. and A. Newell (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.

Stan Development Team (2022). Stan modeling language users guide and reference manual, version 2.31.

Swait, J. and W. Adamowicz (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., M. Ban, and G. M. Allenby (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., Q. Zhang, and T. Erdem (2021). Prior information and consumer search: Evidence from eye-tracking. *Unpublished manuscript, New York University.*

Woodford, M. (2012). Inattentive valuation and reference-dependent choice. Technical report, Columbia University.

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

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

A Appendix

A.1 Identification of simple attributes: Log-likelihood traces based on MCMC estimation with linear coding

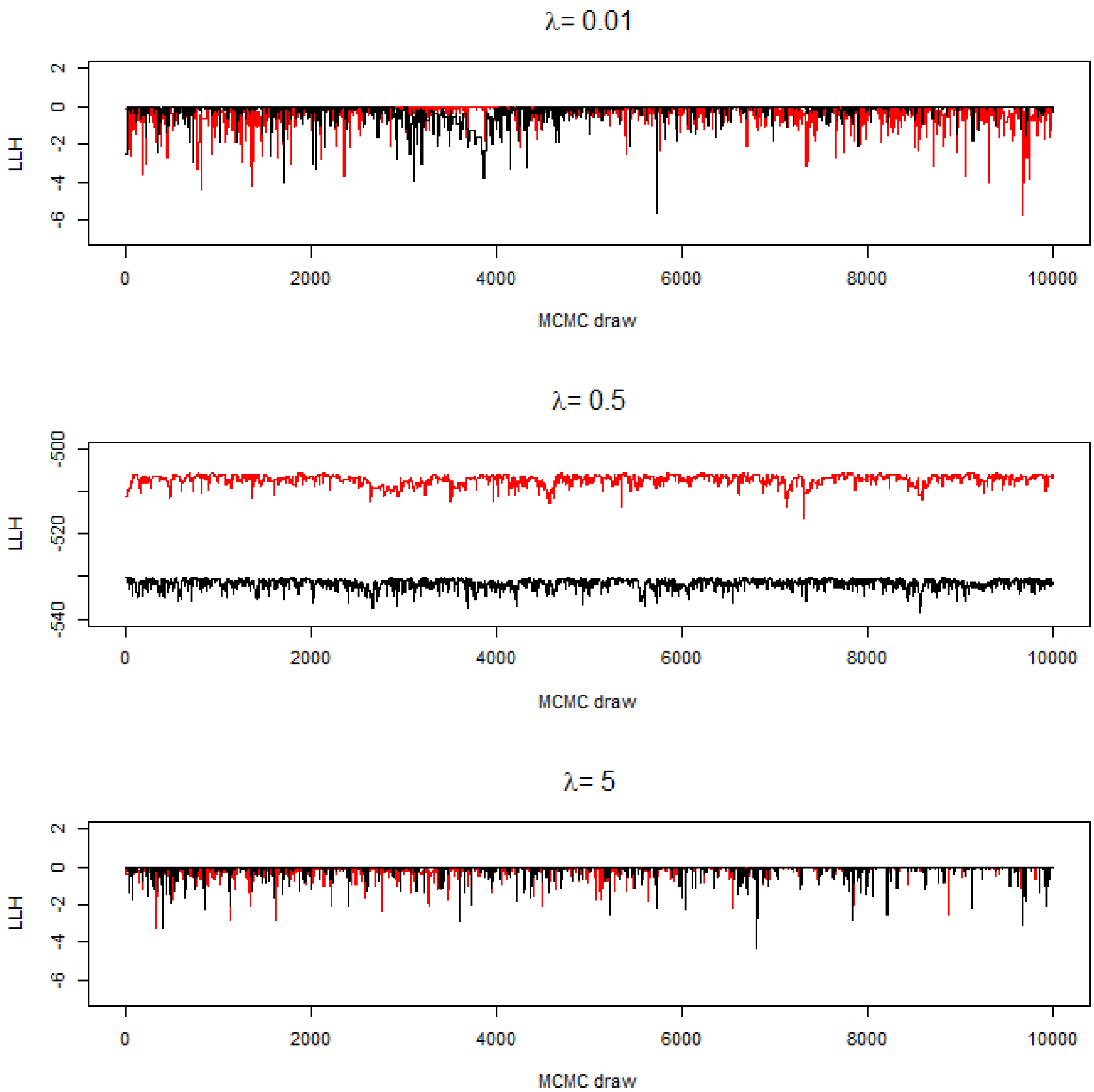

Figure 15: **Log-likelihood traces for both the correct (red) and false (black) specification of the simple attribute.**

A.2 Identification of simple attributes: Categorical coding

$\lambda = \mathbf{0.5}$					
Model	min	25%	50%	75%	max
Correct - Linear	-518.18	-510.46	-508.60	-504.05	-503.46
Wrong - Categorical	-536.14	-526.93	-525.01	-524.22	-521.02
Wrong - Linear	-541.94	-538.48	-536.71	-531.15	-530.52

Table 9: **Quartiles and min/max of Log-Likelihood MCMC draws with $\lambda = 0.5$ for the correct linear, wrong linear, and wrong categorical specification respectively.**