

Random utility and limited consideration

VICTOR H. AGUIAR

Department of Economics, University of Western Ontario

MARIA JOSE BOCCARDI

Amazon

NAIL KASHAEV

Department of Economics, University of Western Ontario

JEONGBIN KIM

Department of Economics, Florida State University

The random utility model (RUM, [McFadden and Richter \(1990\)](#)) has been the standard tool to describe the behavior of a population of decision makers. RUM assumes that decision makers behave as if they maximize a rational preference over a choice set. This assumption may fail when consideration of all alternatives is costly. We provide a theoretical and statistical framework that unifies well-known models of random (limited) consideration and generalizes them to allow for preference heterogeneity. We apply this methodology in a novel stochastic choice data set that we collected in a large-scale online experiment. Our data set is unique since it exhibits both choice set and (attention) frame variation. We run a statistical survival race between competing models of random consideration and RUM. We find that RUM cannot explain the population behavior. In contrast, we cannot reject the hypothesis that decision makers behave according to the logit attention model ([Brady and Rehbeck \(2016\)](#)).

KEYWORDS. Random utility, experimental discrete choice, random consideration sets, frames.

JEL CLASSIFICATION. C12, C90, D12, D81.

Victor H. Aguiar: vaguiar@uwo.ca

Maria Jose Boccardi: majoboccardi@gmail.com

Nail Kashaev: nkashaev@uwo.ca

Jeongbin Kim: jbkimecon@gmail.com

This paper was previously circulated as “Does Random Consideration Explain Behavior when Choice is Hard? Evidence from a Large-scale Experiment.” We are grateful to 4 anonymous referees for insightful comments and suggestions. We would like to thank Roy Allen, Jose Apesteguia, Miguel Ballester, Levon Barseghyan, Juan Dubra, Mikhail Freer, Yoram Halevy, Yuichi Kitamura, Paola Manzini, Marco Mariotti, John Rehbeck, and Jörg Stoye for useful comments and suggestions. We also thank the participants of the Barcelona GSE Summer Forum (Stochastic Choice), BRIC 2019, IWABE 2019, ASSA 2020 (Econometrics of Decision and Demand) for useful feedback. Mingshi Kang provided excellent research assistance. Aguiar and Kashaev gratefully acknowledge financial support from the Western Social Science Faculty grant (FRDF R5533A02) and Social Sciences and Humanities Research Council of Canada Insight Development Grant (430-2019-0426). This study is approved by the Caltech’s IRB No. 18-0812. The experiment was run prior Boccardi joined Amazon.

1. INTRODUCTION

A fundamental question in social science is how to describe the behavior of a population of decision makers (DMs). The random utility model (RUM, [McFadden and Richter \(1990\)](#)) is the standard tool to *describe* behavior.¹ RUM assumes that DMs behave *as if* they maximize their preferences over their choice set. However, RUM may fail at describing behavior if DMs do not consider all available alternatives. For instance, DMs may behave as if they use a two-stage procedure: first simplifying choice by using a consideration set, and only then choosing the best alternative among those considered. Thus, DMs may choose dominated alternatives.² A large literature, pioneered by [Masatlioglu, Nakajima, and Ozbay \(2012\)](#) and [Manzini and Mariotti \(2014\)](#), has proposed theories of consideration-mediated choice. These theories accommodate departures from RUM caused by inattention, feasibility, categorization, and search.³ In contrast to RUM,⁴ little is known about the empirical validity of these models. Our work aims to fill this important gap in the literature.

Methodologically, we provide a unifying theoretical framework that generalizes well-known theories of random consideration. We unify these theories with a new concept called attention-index. The attention-index is a net measure of how enticing a collection of alternatives is, or how costly it is to pay attention to it. We show how to test these theories statistically, and how to recover the preference distribution and consideration rules. Our framework extends many theories of consideration-mediated choice to allow for preference heterogeneity. This allows us to take these theories of individual behavior to the population level, thus permitting the use of cross-sectional data sets to test them. Following [McFadden and Richter \(1990\)](#) and [Kitamura and Stoye \(2018\)](#), we take seriously the fact that all theories of stochastic choice have as their primitive the unobserved distribution over choices that can only be estimated in finite samples by sample frequencies of choice. Thus, to test these models in finite samples, we need to account for sampling variability.

Empirically, we design a large experiment⁵ with two independent sources of exogenous variation: (i) full variation in choice sets (menus), and (ii) variation in frames. We conducted this experiment online in Amazon Mechanical Turk (MTurk), collecting 12,297 independent choice observations from 2135 individuals. A *frame* consists of

¹RUM was first proposed by [Block and Marschak \(1960\)](#) and [Falmagne \(1978\)](#) in an environment similar to ours.

²For evidence of choice of dominated alternatives, see [De Los Santos, Hortaçsu, and Wildenbeest \(2012\)](#), [Honka \(2014\)](#), [Ho, Hogan, and Morton \(2017\)](#), [Honka, Hortaçsu, and Vitorino \(2017\)](#), [Hortaçsu, Madanizadeh, and Puller \(2017\)](#), [Barseghyan, Coughlin, Molinari, and Teitelbaum \(2021\)](#), and [Heiss, McFadden, Winter, Wuppermann, and Zhou \(2021\)](#).

³See, for instance, [Aguiar, Boccardi, and Dean \(2016\)](#), [Brady and Rehbeck \(2016\)](#), [Aguiar \(2017\)](#), [Lleras, Masatlioglu, Nakajima, and Ozbay \(2017\)](#), [Caplin, Dean, and Leahy \(2019\)](#), [Horan \(2019\)](#), [Kovach and Ülkü \(2020\)](#), and [Cattaneo, Ma, Masatlioglu, and Suleymanov \(2020\)](#).

⁴For evidence for RUM, see [Kitamura and Stoye \(2018\)](#) and [McCausland, Marley, Davis-Stober, Brown, and Park \(2020\)](#).

⁵Other experiments that we are aware of that have collected stochastic choice data focusing on choice set variation are [Apesteguia and Ballester \(2021\)](#) (87 individuals) and [McCausland et al. \(2020\)](#) (141 participants). In contrast to our work, both focus mainly on binary choice sets and goodness-of-fit measures (including the computation of Bayes factors).

observable information that is irrelevant in the rational assessment of the alternatives (Salant and Rubinstein (2008)).⁶ Full variation in choice sets means that all possible choice sets are observed by the researcher. It allows us to test consideration-mediated choice theories in a large cross-section of heterogeneous individuals. Frame variation means that we vary the complexity of the description of alternatives without affecting the relevant payoffs. This is equivalent to varying the cost of consideration. In this sense, we induce an attention frame. This variation in frames allows us to differentiate between RUM and models of limited consideration since consideration could change with frames but preferences must remain stable.

In general, without frame variation, many models of consideration are empirically indistinguishable from RUM. For instance, if a data set, without frame variation, is consistent with the classical consideration model of Manzini and Mariotti (2014), then it is also consistent with RUM. However, this model and RUM will typically recover a distinct distribution of preferences. Varying frames, we can test whether the distribution of preferences remains the same across frames. This will imply that the model in Manzini and Mariotti (2014) and RUM may have different empirical implications in our experimental setup. The same reasoning applies to any model of limited consideration.

In our experimental design, we introduce three attention-cost/complexity frames for every choice set. Each DM faces all three frames. These treatments require the DM to solve a simple cognitive task to understand the alternatives. The consideration cost is progressively reduced across frames while we keep the choice set fixed. The fact that the choice remains the same is not explicitly stated in the experiment instructions. Within this design, we can understand how changing consideration costs across frames may affect choices while we keep (the distribution of) preferences fixed. Under our incentives protocol (pay-at-random across tasks), and under consequentialism, the distribution of preferences must remain constant regardless of the frame. By exploiting this feature of our design, we show that, in our sample, RUM fails to describe the population behavior, while the logit attention model of Brady and Rehbeck (2016) describes it well.

Our paper generalizes the methodology of Cattaneo et al. (2020) by allowing heterogeneity in preferences when the choice sets include a default alternative. We show that in our setting, if one assumes independence of preferences and consideration, at the population level, testing a consideration-mediated choice theory with heterogeneous preferences and a dominated default is equivalent to testing whether this hypothetical full-consideration distribution over choices is consistent with RUM. We, therefore, use this result to test these models using the framework of Kitamura and Stoye (2018).

To exploit all possible implications of the limited-consideration models of interest, we need full variation in choice sets. A limited consideration model may describe behavior well for a nonexhaustive data set, but it may fail to do so for an extended one. That is, one may have false positives when observing choices from a nonexhaustive set of menus, as discussed at length in Cattaneo et al. (2020) and De Clippel and Rozen (2021). Nonetheless, full exogenous variation in choice sets is an important data feature that

⁶We interpret the notion of rational assessment to be an assessment compatible with consequentialism and RUM.

is usually not satisfied in field data. Our testing procedure exploits rich experimental variation to increase the statistical power of our tests.

In addition, our experiment introduces a dominated default alternative that works as an opportunity cost of paying attention. For any choice set and frame, the default is always present and shown first. Moreover, it is preselected as the default choice—if the subject decides to skip the task, she is informed that the default alternative will be chosen for her. This design allows us to use the default alternative as the opportunity cost of incurring in the cost of consideration and understanding the other alternatives in the choice set. The set of alternatives in our experiment consists of lotteries. Hence, we use a degenerate lottery as the default due to its simplicity. In this sense, we believe the default alternative in our design has effectively zero cost of consideration. This dominated default is key to disentangle the distribution of preferences from random consideration.⁷

We use these theoretical and experimental innovations to test two well-known models of random consideration: (i) the logit attention model of [Brady and Rehbeck \(2016\)](#) (LA), and (ii) the version of elimination-by-aspects model of [Tversky \(1972\)](#) characterized by [Aguiar \(2017\)](#) (EBA).⁸ There is increasing interest in incorporating limited consideration in discrete choice.⁹ In particular, the influential and tractable model of [Manzini and Mariotti \(2014\)](#) (MM) has become an important tool for the analysis of limited attention in empirical work (e.g., [Dardanoni et al. \(2020\)](#), [Abaluck and Adams \(2021\)](#), and [Kashaev and Lazzati \(2021\)](#)). However, MM is highly stylized and assumes that consideration is driven by an item-dependent parameter (i.e., independence in consideration). We investigate from an experimental perspective whether this strong assumption is effective in explaining choice when consideration is hard. To do so, we consider two extensions of MM that allow for substitution and complementarity in consideration, the LA and EBA models. These two generalizations have the property that their intersection is exactly the MM model ([Kovach and Suleymanov \(2021\)](#)). This implies that MM explains the population behavior if and only if both the LA and EBA models explain it.

We test these models and the benchmark RUM conditioning on the frame. Crucially, we require the underlying preference relation to be stable among frames while allowing the consideration rules to vary with the frame. Our main findings are: (i) We reject the hypothesis that RUM provides a good description of population behavior. (ii) In contrast, the LA model with heterogeneous preferences cannot be rejected at the 5% significance level. (iii) However, we reject the hypothesis that EBA, and hence MM, describe the population behavior.

Our work contributes to the recent experimental literature on stochastic choice, limited consideration, and departures from RUM. Even though by now, limited attention in many environments is well documented, it is less clear what structural models of limited attention should the practitioner use. We see our contribution as providing answers

⁷We formulate a sensitivity analysis when the default is not dominated in Appendix B.

⁸We refer to EBA as the version of the original model in [Tversky \(1972\)](#) by [Cattaneo et al. \(2020\)](#) (Example 6). The model in [Aguiar \(2017\)](#) coincides with EBA in the special case where there is a dominated default with a restriction that any category that does not contain the default has zero mass. The EBA model characterization for the case without a default remains an open question.

⁹See, for instance, [Goeree \(2008\)](#), [Dardanoni, Manzini, Mariotti, and Tyson \(2020\)](#), [Abaluck and Adams \(2021\)](#), [Barseghyan et al. \(2021\)](#), and [Barseghyan, Molinari, and Thirkettle \(2021\)](#).

to this second issue.¹⁰ We hope that our findings about which models of limited consideration are successful empirically will inform future empirical work in the field. For instance, our findings have already been used to motivate the choice of the parametric specification of limited consideration in the recent work of [Abaluck and Adams \(2021\)](#).¹¹

Outline

The paper proceeds as follows. Section 2 presents our model. Section 3 details frame variation and our testing procedure. Section 4 presents our experiment. Section 5 presents the testing results. Finally, Section 6 concludes. All proofs and additional results are in the Appendix. The Online Supplementary Material ([Aguiar, Boccardi, Kashaev, and Kim \(2023\)](#)) include the replication files (data and code) and the instructions of our experiment.

2. ENVIRONMENT—MODEL

We consider a finite choice set X and we denote the outside alternative or default as $o \notin X$. We let the set of all possible choice sets be $\mathcal{A} = 2^X \setminus \{\emptyset\}$, where 2^X denotes the set of all subsets of X . A probabilistic choice rule is a mapping $p : X \cup \{o\} \times \mathcal{A} \mapsto [0, 1]$. The probabilistic choice rules for a given choice set add up to 1, $\sum_{a \in A} p(a, A) + p(o, A) = 1$. Moreover, $p(a, A) = 0$ if $a \notin A$. We fix $p(o, \emptyset) = 1$. A complete stochastic choice rule is a vector $P = (p(a, A))_{A \in \mathcal{A}, a \in A \cup \{o\}}$. For identification purposes, we treat P as a known object. In practice, we do not observe P , but can consistently estimate it by the collection of sample frequencies \hat{P} (see Section 3.2 for further details).

2.1 *Random behavioral model*

We consider an environment where DMs, faced with a choice set $A \in \mathcal{A}$, first pick $D \subseteq A$ (consideration set) and then choose the alternative in D that maximizes their preferences. With probability $\pi(\succ)$, DMs are endowed with preferences $\succ \in X \times X$ drawn from the set of all linear orders (strict preferences) on X , $R(X)$.¹² Note that since $o \notin X$, following [Manzini and Mariotti \(2014\)](#), we implicitly assume that the default is picked if and only if nothing else is considered. A typical interpretation of this situation is the sleeping agent behavior (see, for instance, [Abaluck and Adams \(2021\)](#)). When the agent is sleeping (i.e., she considers the empty set) she chooses the default alternative. Otherwise, the agent wakes up and considers some nonempty set and maximizes her preferences in her consideration set. In the second case, the default is assumed to be dominated by the rest of alternatives. We show how to relax this assumption in Appendix B.

¹⁰There is a vast literature documenting departures from fully rational behavior, but it is not focused on limited consideration. See [Rieskamp, Busemeyer, and Mellers \(2006\)](#) for a survey.

¹¹[Abaluck and Adams \(2021\)](#) structurally estimate a model of discrete choice with limited consideration under the LA and MM model.

¹²Linear orders are complete, reflexive, transitive, and antisymmetric orders.

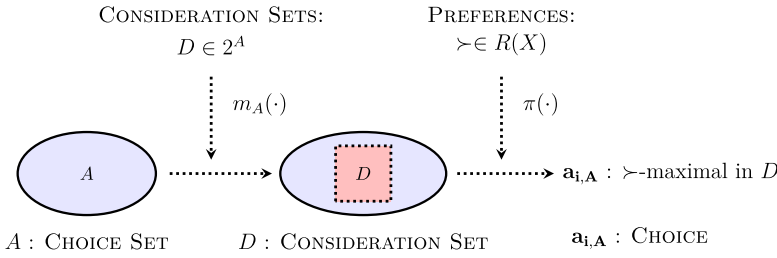


FIGURE 1. Consideration mediated choices. Choices are the result of a two stage process, first pick a consideration set and then pick the best alternative in that set. We observe choices (\mathbf{a}_i) and menus (A). We do not observe and we aim to identify the theory objects: distribution of preferences in the population (π) and stochastic choice rule (m_A).

The distribution $\pi \in \Delta(R(X))$ fully captures preference heterogeneity.¹³ The distribution over random consideration sets given the menu A is fully characterized by $m_A : 2^A \rightarrow [0, 1]$, $\sum_{D \subseteq A} m_A(D) = 1$. In other words, m_A is an element of the simplex $\Delta(2^A)$. Let m denote the complete collection of those distributions for all possible menus. That is, $m = (m_A(D))_{A \in \mathcal{A}, D \in 2^A}$. We assume that the random consideration sets and random preferences are *independent*.

DEFINITION 1 (Random behavioral model, B-rule). A complete stochastic choice rule P is a B-rule if there exists a pair (m, π) such that

$$p(a, A) = \sum_{> \in R(X)} \pi(>) \sum_{D \subseteq A} m_A(D) \mathbb{1}(a > b, \forall b \in D)$$

for all $a \in X$ and $A \in \mathcal{A}$.

This choice rule is illustrated in Figure 1. Definition 1 implicitly assumes that the random consideration set rule and the heterogeneous preferences are independent. Independence is a good starting assumption in the sterile environment of our experiment, as we want to achieve a decomposition of any observed probabilistic choice rule into its consideration (captured by m) and preference (captured by π) components. Independence has been assumed successfully in the structural work of [Abaluck and Adams \(2021\)](#). Also, we are interested in modeling decision-making in two-stages where DMS simplify a hard choice task using fast-and-frugal heuristics (consideration) that are independent of preferences, and then choose rationally from the simplified choice set. If the researcher observes additional information (e.g., age, gender, education, and income levels of individuals), then random consideration rule and random preferences need to be independent only conditionally on those observables.¹⁴

Independence holds trivially for the case of homogeneous preferences, such as all models covered by the Random Attention Model (RAM) of [Cattaneo et al. \(2020\)](#). More-

¹³For any set C , $\Delta(C)$ denotes the set of all probability distributions (simplex) on C . $\mathbb{1}(B)$ denotes indicator of the statement B and is equal to 1 if B is true and is equal to zero otherwise.

¹⁴See [Kashaev and Aguiar \(2022\)](#) for a study of the correlation between preferences and consideration.

over, as the following lemma demonstrates, the B-rule does not have empirical content even under the independence assumption.

LEMMA 1. *Every complete stochastic choice rule P is a B-rule.*

Without additional restriction on π and m , the model is not falsifiable. That is, any P can be generated by some independent π and m . We will impose constraints on m that will allow us to test an important class of random consideration sets models without restricting heterogeneity in preferences.

2.2 Attention-index consideration set rule

We restrict m by considering a family of consideration set rules that are governed by an *attention index*. The attention-index $\eta \in \Delta(2^X)$ is a distribution over the power set. The value $\eta(D)$ captures the unconditional attention that DMs pay to the set $D \in 2^X$. The attention index of a set is a net measure of its attractiveness with respect to how hard it is to consider it. The attention index measures how enticing a consideration set is, and how complex it is to understand. Therefore, $\eta(C) > \eta(D)$ means that C , in net terms, attracts more attention than D .

DEFINITION 2 (Attention-index representation). A consideration set rule m admits an attention-index representation if there exists and attention-index η , a link function ψ , and an index correspondence g such that $g(D, A) \subseteq 2^X \setminus D$ and

$$m_A(D) = \psi\left(\eta(D), \sum_{C \in g(D, A)} \eta(C)\right)$$

for all $A \in \mathcal{A}$ and $D \subseteq A$.

In what follows, we assume that the link function ψ and the correspondence g are known. A given link function captures the particular way in which the attention index shapes consideration given a choice set or menu. In other words, the link function transforms unconditional attention into conditional (on the choice set) attention. The index set $g(D, A)$ captures the collection of sets that are used in the attention aggregator $\sum_{C \in g(D, A)} \eta(C)$. However, we do not assume that η is known and do not impose any restrictions on it, so the setting is still semiparametric.

Next, we define several important models of limited consideration that admit the attention-index representation. They only differ in how they use the attention index to form the consideration set at a given menu.

DEFINITION 3. The logit attention (LA, Brady and Rehbeck (2016)), the choice set independent (MM, Manzini and Mariotti (2014)), the random consideration (EBA, Tversky (1972), Aguiar (2017)¹⁵), and the full consideration (FC) models admit an attention-index representation such that:

¹⁵The model in Aguiar (2017) is a special case of the model in Tversky (1972) and coincides with it when the attention index of only sets that contain the default is nonzero.

- $m \in \mathcal{M}^{\text{LA}}$ if and only if there exists $\eta \in \Delta(2^X)$ such that for

$$m_A(D) = \frac{\eta(D)}{\sum_{C \subseteq A} \eta(C)} > 0$$

for all $A \in \mathcal{A}$ and $D \in 2^A$;

- $m \in \mathcal{M}^{\text{MM}}$ if and only if $m \in \mathcal{M}^{\text{LA}}$ with

$$\eta(A) = \prod_{a \in X \setminus A} (1 - \gamma(a)) \prod_{b \in A} \gamma(b)$$

for a given $\gamma : X \rightarrow (0, 1)$ and for all $A \in 2^X$;

- $m \in \mathcal{M}^{\text{EBA}}$ if and only if there exists $\eta \in \Delta(2^X)$ such that

$$m_A(D) = \sum_{C: C \cap A = D} \eta(C)$$

for all $A \in \mathcal{A}$ and $D \in 2^A$;

- $m \in \mathcal{M}^{\text{FC}}$ if and only if $m \in \mathcal{M}^{\text{EBA}}$ with

$$\eta(A) = \mathbb{1}(A = X).$$

The MM-rule imposes independence in consideration across items, making the model highly stylized and tractable. Both the LA- and the EBA-rules generalize this item-independence to allow for substitution and complementarity of attention between different items. The LA-rule predicts that the probability of considering D is proportional to the attention-index value $\eta(D)$. The EBA-rule predicts that the probability of considering a subset D given menu A is equal to the probability that D is the intersection of the subset of alternatives (considered randomly using the attention index) and the choice set.¹⁶ In our application, we focus on these four particular models because they allow us to learn about the true data generating process governing our experimental application systematically.¹⁷ However, our approach extends to any model of consideration that admits an attention-index representation.¹⁸

Given the definition of the attention-index representation, we can define a restricted version of the B-rule (L-B-rule). Let \mathcal{M}^{L} be the set of all consideration rules induced by a given link function ψ^{L} , the correspondence g^{L} , and all attention-indices η ,

$$\mathcal{M}^{\text{L}} = \left\{ m : m_A(D) = \psi^{\text{L}} \left(\eta(D), \sum_{C \in g^{\text{L}}(D, A)} \eta(C) \right) \text{ for some } \eta \in \Delta(2^X) \right\}.$$

¹⁶Note that for the LA-model $\psi^{\text{LA}}(x, y) = x/(x + y)$ and $g^{\text{LA}}(D, A) = \{C \in 2^X \setminus D : C \cap A = C\}$, and for the EBA-model $\psi^{\text{EBA}}(x, y) = x + y$ and $g^{\text{EBA}}(D, A) = \{C \in 2^X \setminus D : C \cap A = D\}$.

¹⁷LA and EBA are completely distinct generalizations of MM. Thus, rejecting or accepting either of these models is very informative.

¹⁸For instance, one can generate a continuum of such models by taking all possible convex combinations of consideration rules m^{LA} and m^{EBA} that are generated by the same attention-index η .

DEFINITION 4 (L-B-rule). A complete stochastic choice rule P is L-B-rule if P is a B-rule with $m \in \mathcal{M}^L$.

We believe that random consideration rules that admit an attention-index representation have several theoretical advantages over other generalizations in the literature. First, they cover, as special cases, well-known models of consideration and allow us to introduce a new class of semiparametric models of limited consideration with heterogeneous preferences. In other words, the L-B-rule unifies existing models into a common structure. This structure helps us to understand the common traits of these models, to classify them, and to provide an identification and testing framework to possibly new models covered in this structure.

Second, they enable the unique identification of the consideration rule and the underlying stochastic choice full consideration probability from a cross-section of choices with menu variation and heterogeneous preference (see Theorem 3). We are not aware of any other work that achieves the same and is more general than ours. Allowing heterogeneous preferences is important in analysis of any data sets, and is unavoidable in cross-sections. An important alternative generalization is RAM, which imposes only a shape constraint on random consideration. However, it does not allow heterogeneous preferences, nor it obtains point identification of consideration or the underlying full consideration choice rule. Moreover, RAM is completely uninformative about preferences, when the stochastic choice rule is regular,¹⁹ as in the random utility models.²⁰ Also, the L-B-rule allows cycles of probabilities that can violate regularity while RAM cannot (see Appendix C.2).

Third, the concept of attention index is intuitive, simple, and of behavioral interest, as it provides a useful (unconditional) index of attention for any subset of alternatives. The link function and the attention index follow the tradition of classical stochastic choice theory of simple scalability where the probability of choice of an alternative in a menu is a nonlinear function (a link function) of some scale/index (Krantz (1965), Tversky (1972)). In the simple scalability tradition, the scale/index captures the intensity of the stimuli associated with a particular alternative. A classical example is the logit model of choice. Here, we apply the same intuition to consideration—the attention index captures the net attractiveness of a consideration set.

Fourth, consideration rules that admit an attention-index representation are compatible with optimal random consideration of a representative DM. Here, we show that a L-B-rule can be obtained as the result of allocating attention optimally (see Example 2).

In addition, consideration rules that admit an attention-index representation are compatible with a flexible interpretation of randomness. In our framework, the randomness due to limited consideration can arise both at the individual and population level. Indeed, consideration can be random at the individual level and independent and identically distributed (i.i.d.) at the population level; and consideration can be determinis-

¹⁹Regular random choice means that $p(a, A) \geq p(a, B)$ for any two menus $A \subseteq B$ and any $a \in A$.

²⁰Kashaev and Aguiar (2022) extend RAM to allow for heterogeneous preferences and show that when the stochastic choice rule is regular nothing can be said as well about random consideration and preferences even under independence among them.

tic at the individual level but heterogeneous at the population level. The next example shows how the latter case can be described by a L-B-rule.

EXAMPLE 1 (Heterogeneous categorization). Consider a population of DMs with two types of agents endowed with different deterministic attention rules. Assume that half of the DMs are fully attentive, while the other half follows a rule of thumb: they pay full attention to option b if it is present in a given choice set, else the consideration set is empty. The DMs pick the best alternative, according to the (heterogeneous) preference realization governed by some $\pi \in \Delta(R(X))$. If the consideration set is empty, then the outside option is selected. This population has heterogeneous (deterministic) consideration, however, the population behavior can be fully captured by a random consideration rule with the EBA restriction. Namely, let $\eta(X) = \frac{1}{2}$ and $\eta(\{b\}) = \frac{1}{2}$. Then the EBA-B-rule can describe this population behavior.

The next example demonstrates that the consideration rules that admit an attention-index representation can be derived as a solution to the problem where representative DMs optimally allocate their attention.

EXAMPLE 2 (Costly attention allocation). Consider a (representative) DM whose preferences are given by a (mean) utility $u : X \rightarrow \mathbb{R}$ and additive random vector $\xi = (\xi_x)_{x \in X}$ such that the random utility of a given item x is given by $u(x) + \xi_x$. The taste shocks ξ are distributed with respect to some continuous distribution.²¹ When the DM is faced with a menu A , she needs to allocate her attention, measured by $m_A \in \Delta(2^A)$, over all possible consideration sets in A (including the empty set). The attractiveness of a set D is captured by the McFadden's surplus of a given set defined by $\alpha(D) = \mathbb{E}[\max_{x \in D} (u(x) + \xi_x)]$ for all $D \subseteq 2^A \setminus \{\emptyset\}$, where the expectation is taken with respect to ξ . The attractiveness of the empty set is normalized to be 0, $\alpha(\emptyset) = 0$. The surplus $\alpha(D)$ is a measure of average attractiveness, capturing how enticing a consideration set is for the representative DM. The difficulty of picking a consideration set, or the cost of attention, is captured by a cognitive cost function $K : [0, 1] \rightarrow \mathbb{R} \cup \{\infty\}$. If D is considered with probability $m(D)$, then the cognitive cost is $K(m(D))$. The cost function is menu independent, but depends on the allocated attention $m(D)$. Following [Fudenberg, Iijima, and Strzalecki \(2015\)](#), we assume that K is convex. In this case, DM's problem is to find $m_A \in \Delta(2^A)$ that maximizes the expected attractiveness of the menu, given the cognitive cost of processing it. Formally,

$$m_A = \arg \max_{m \in \Delta(2^A)} \sum_{D \subseteq A} [m(D)\alpha(D) - K(m(D))].$$

When $K(t) = 0$ for all $t \in [0, 1]$, the solution is such that $m_A(A) = 1$. That is, the DM is consistent with full consideration (FC). When $K(t) = -t \log(t)/\theta$, where θ is the cost parameter, we get that

$$m_A(D) = \frac{\exp(\theta\alpha(D))}{\sum_{C \subseteq A} \exp(\theta\alpha(C))}.$$

²¹This guarantees that the implied random utility rule π exists because ties have zero probability.

That is, optimal consideration in this case is consistent with LA with the attention-index $\eta(D) = \exp(\theta\alpha(D))$. Note that in this definition of η , the independence assumption is still satisfied since $\eta(D)$ is an aggregate quantity and only depends on the mean utilities $(u(x))_{x \in X}$ and the distribution of ξ but not on the realizations of ξ (e.g., when ξ follows the Gumbel distribution, then $\alpha(D) = \log \sum_{x \in D} e^{u(x)}$). Of course, we can replace $\alpha(D)$ with any other measure of attractiveness and all our derivations will go through. This means we can generate any model consistent with LA this way. Finally, when $K(t) = \frac{1}{2}t^2$, we can get a special case of the EBA model.²²

Finally, the attention-index representation also has several econometric advantages. It allows for a significant reduction of the dimensionality of the consideration set rules, making them tractable. In general, the number of parameters controlling the random consideration is $\sum_{A \subseteq X} 2^{|A|} - 1$. The single-attention index and the link function reduce the number of unknown parameters to $2^{|X|} - 1$. It also leads to statistical testing (i.e., we can take into account sample variability) of known models of random consideration in a cross-section data set of choices (see Section 5). Thus, one can confront existing models to experimental data sets in a competitive fashion to guide the exploration of models of consideration sets and to inform which models are more successful.

2.3 Characterization and identification of the L-B-model

In this section, we answer the following questions: (i) When can we recover different consideration rules from the data? (ii) What are their observable implications? We answer these questions by *decomposing* the observed probabilities of choice P into an attention rule m and a distribution of preferences π . In other words, we recover from the data set P the primitives of the B-rule that generated it, and provide necessary and sufficient conditions that guarantee that a data set P can be generated by a L-B-rule.

Our starting point is to exploit the fact that if a consideration rule m admits an attention-index representation, then the probability of choosing the default alternative is completely determined by the attention-index η . In particular, the probability of choosing the outside option is independent of the distribution of preferences due to the independence assumption that we have imposed between preferences and consideration. In addition, recall that in our model the outside option is only chosen when nothing else in the menu is considered.²³ If we denote $p_o = (p(o, A))_{A \in \mathcal{A}}$ and $\psi_\emptyset(\eta) = (\psi(\eta(\emptyset), \sum_{C \in \mathcal{g}(\emptyset, A)} \eta(C)))_{A \in \mathcal{A}}$, abusing notation we can write the system of equations

$$p_o = \psi_\emptyset(\eta).$$

When ψ_\emptyset is *invertible*, we can uniquely recover the random consideration rule from the probability of choosing the outside alternative from all different menus. Since our objective is the identification of the consideration rule, we provide a natural restriction on

²²The EBA model then requires that the DM does not reoptimize in smaller menus $A \subset X$. Instead, she uses the heuristic that whatever category she draws at X it is intersected with the given menu to obtain the consideration set.

²³For an extension that relaxes this assumption, see Appendix B.

the attention-index rule that guarantees invertibility of ψ . This restriction is satisfied by the models of interest in this paper (but not restricted to them).

DEFINITION 5 (Totally monotone consideration). A consideration rule m admitting an attention-index representation is *totally monotone*, if we can write, for all $A \in \mathcal{A}$,

$$m_A(\emptyset) = \varphi\left(\eta(\emptyset), \sum_{C \subseteq A} \eta(C)\right),$$

where $\varphi : [0, 1] \times [0, 1] \rightarrow [0, 1]$ is a strictly monotone in each argument function.

The probability of not considering any object, conditional on a given choice set, is assumed to be a monotonic function of the cumulative probability of paying attention to at least some alternative in the menu (according to η), and of the probability of not considering anything (unconditionally). Crucially, the dependence on the correspondence g disappears in a totally monotone attention-index representation, with respect to the general attention-index representation. Note, this happens only for the case of not considering anything (i.e., the consideration set is \emptyset).

Totally monotone attention-index rules are such that the random consideration is monotone as in Cattaneo et al. (2020), namely $m_A(\emptyset) \leq m_B(\emptyset)$ if $B \subseteq A$, when φ is strictly increasing in the first entry. However, in this case, they imply more. Since the mapping $m_{(\cdot)}(\emptyset) : \mathcal{A} \rightarrow [0, 1]$ is a function of the cumulative probability of considering at least one item in any given menu (i.e., a function of $\sum_{C \subseteq A} \eta(C)$), the behavior of the probability of choosing the outside option will be restricted. For instance, for the case of EBA it will satisfy a form of marginal decreasing propensity of choice (Aguiar (2017)).

We highlight that total monotonicity is testable. Observe that the marginal probability of choosing the outside option, when a new set of alternatives is added to a menu, is weakly decreasing (e.g., $\Delta_C p(o, A) = p(o, A \cup C) - p(o, A) \leq 0$ and $\Delta_D(\Delta_C p(o, A) \leq 0)$).²⁴ Alternatively, φ can be strictly decreasing in the first entry, in which case it provides an antithetic behavior to that of RAM (yet testable). In this sense, this restriction on attention is neither weaker nor stronger than the monotonicity restriction in Cattaneo et al. (2020).

Strict monotonicity of φ in each of its entries implies the invertibility of ψ .²⁵ This sufficient condition for invertibility of the link function ψ is mild. It holds in all the examples of interest in this work. Importantly, it is a testable restriction. Note that since the inverse of ψ is known under the model of interest, we can compute a candidate η from the data P . If the computed η is not an element of the simplex $\Delta(2^X)$, then ψ is not invertible.

The next lemma shows that the models we consider admit a totally monotone attention-index representation.

²⁴Note, however, that this restriction does not mean that m satisfies the monotonicity property for other menus different from the empty set.

²⁵This is a consequence of Möbius invertibility of the mapping $v(\cdot) = \sum_{C \subseteq \cdot} \eta(C)$ (Chateauneuf and Jaffray (1989)).

LEMMA 2. Any $m \in \mathcal{M}^L$, $L \in \{LA, MM, EBA, FC\}$, admits a totally monotone attention-index representation with

- $\varphi^{LA}(\eta_o, t) = \frac{\eta_o}{t}$;
- $\varphi^{EBA}(\eta_o, t) = 1 - t + \eta_o$;
- $\varphi^{MM}(\eta_o, t) = \varphi^{LA}(\eta_o, t)$ and $\varphi^{MM}(\eta_o, t) = \varphi^{EBA}(\eta_o, t)$;
- $\varphi^{FC}(\eta_o, t) = \varphi^{EBA}(\eta_o, t)$.

The proof is omitted because of its simplicity for the cases of LA, EBA, and FC. For the case of MM, the statement follows from Brady and Rehbeck (2016) and Aguiar (2017).
26

The key assumption in this section has been that the default that is always present is always dominated. We formulate a sensitivity analysis when the default is not dominated in Appendix B.

2.4 Characterization of the L-B-model

As a preliminary step for characterizing the L-B-model, we construct a candidate calibrated attention-index η^L from the data P . Informally, this calibrated (revealed) attention index is the result of inverting the link function φ with respect to the probability of choosing the default alternative. The link function invertibility is a consequence of the monotonicity assumptions and the existence of a unique Mobius inverse of the cumulative attention-index $v(\cdot) = \sum_{C \subseteq \cdot} \eta(C)$ (Chateauneuf and Jaffray (1989)). We do this recursively. For a given φ^L , let $\varphi_1^{-1,L}$ and $\varphi_2^{-1,L}$ be the inverses of φ^L with respect to the first and the second argument, respectively. Let $|A|$ denote the cardinality of a finite set A .

DEFINITION 6 (Calibrated attention index). For given P , $\eta^L : 2^X \rightarrow \mathbb{R}$ is such that (i) $\eta^L(\emptyset) = \varphi_1^{-1,L}(p(o, X), 1)$, and (ii) for all $D \in 2^X \setminus X$,

$$\eta^L(D) = \sum_{B \subseteq D} (-1)^{|D \setminus B|} \varphi_2^{-1,L}(\eta^L(\emptyset), p(o, D)).$$

The calibrated attention index depends only on the data set P and the model L . If the calibrated attention index of a set is negative, then P could not have been generated by model L . This testable implication is analogous to the Block and Marschak (1960) inequalities.

Now, we construct an object, m^L , that is a distribution over consideration sets if the model L is correctly specified.

²⁶Note that since MM and FC are special cases of LA and EBA, respectively, they share the same link function. However, empirically we will be able to differentiate among them because of the additional restrictions they pose on η .

DEFINITION 7. For a given P , let $m^L = (m^L_A(D))_{A \in \mathcal{A}, D \in 2^A}$, where $m^L_A : 2^A \rightarrow \mathbb{R}$ is such that for all $A \in \mathcal{A}$ and $D \in 2^A$,

$$m^L_A(D) = \psi^L \left(\eta^L(D), \sum_{C \in \mathcal{g}^L(D, A)} \eta^L(C) \right).$$

We can apply this generic formula for totally monotone attention-index rules to the specific models of interest.

EXAMPLE 3. • $m^{LA}_A(D) = \frac{\eta^{LA}(D)}{\sum_{C \subseteq A} \eta^{LA}(C)}$, where $\eta^{LA}(D) = \sum_{B \subseteq D} (-1)^{|D \setminus B|} \frac{p(o, X)}{p(o, B)}$;

- $m^{MM}_A(D) = \frac{\eta^{MM}(D)}{\sum_{C \subseteq A} \eta^{MM}(C)}$, where $\eta^{MM}(D) = \prod_{a \in X \setminus D} (1 - \gamma^{MM}(a)) \prod_{b \in D} \gamma^{MM}(b)$, and $\gamma^{MM} : X \rightarrow \mathbb{R}$ such that $\gamma^{MM}(a) = 1 - \frac{p(o, A)}{p(o, A \setminus \{a\})}$ for some $A \in \mathcal{A}$ that contains a ;
- $m^{EBA}_A(D) = \sum_{C: C \cap A = D} \eta^{EBA}(C)$, where $\eta^{EBA}(D) = \sum_{A \subseteq D: D \in \mathcal{A}} (-1)^{|D \setminus A|} p(o, X \setminus A)$;
- $m^{FC}_A(D) = \mathbb{1}(A = D)$.

In general, m^L may not be a distribution (some components may be negative or greater than 1) since m^L is calibrated from observed frequencies. Moreover, m^{LA} or m^{MM} may not be well-defined if probabilities of choosing the outside option for some choice sets are zero.

To be able to estimate m^L from the data with probability approaching 1, we need the following definition that formalizes the above discussion.

DEFINITION 8 (Well-defined m^L). m^L is well-defined if $m^L_A \in \Delta(2^A)$ for all $A \in \mathcal{A}$.

We are ready to state our main result.

THEOREM 1. For every link function L , the following are equivalent:

- (i) P is a L-B-rule;
- (ii) m^L is a well-defined and P is a B-rule described by (m^L, π) .

Theorem 1 provides a full characterization of well-defined models with link functions. If P is a L-B-rule, then m^L has to be well-defined. Theorem 1 implies that to test a given model one does not need to consider all possible distributions over consideration sets. It suffices to check the unique distribution that is calibrated from observed P according to Definition 7.

Initially, we had to find two objects (the distribution over preferences π and the distribution over consideration sets m) to make the data consistent with the model. Now we just need to find π . In other words, we simplified the testing problem. Unfortunately, the testing problem is still not tractable since the set of all possible distributions over preferences $\Delta(R(X))$ is big. To solve this problem, we introduce another fictitious object.

DEFINITION 9. For given model L and P , let $P_\pi^L = (p_\pi^L(a, A))_{A \in \mathcal{A}, a \in X}$, where $p_\pi^L : X \times \mathcal{A} \rightarrow \mathbb{R}$ is such that for all $A \in \mathcal{A}$ and $a \in A$,

$$p_\pi^L(a, A) = \frac{p(a, A) - \sum_{C \subset A} m_A^L(C) p_\pi^L(a, C)}{m_A^L(A)}.$$

Note that when P has been generated by a L-B-rule, P_π^L corresponds to the underlying full-consideration random utility rule. In fact, we can write a L-B model equivalently as

$$p(a, A) = \sum_{D \subseteq A} m_A(D) p_\pi(a, D),$$

where $p_\pi(a, A) = \sum_{\succ \in R(X)} \pi(\succ) \mathbb{1}(a \succ b, \forall b \in A)$ is the underlying FC distribution over (nondefault) choices that is weighted by the random consideration rule m to produce the observed behavior. When P has been generated by this L-B-rule, it follows that $p_\pi^L = p_\pi$. That is why we call P_π^L the calibrated full consideration rule.

Similar to m^L , P_π^L has interpretation when the L-B-rule is consistent with the data. The next theorem provides the last missing piece of our characterization before testing.

THEOREM 2. Suppose that for given model L and stochastic choice rule P (i) m^L is well-defined, (ii) $m_A^L(A) > 0$ for all $A \in \mathcal{A}$. Then the following are equivalent:

- (i) P is a L-B-rule;
- (ii) P_π^L is a FC-B-rule.

Note that both $m^L \geq 0$ and P_π^L can be computed from P . Thus, Theorem 2 implies that to test a given model L it is necessary and sufficient to test whether m^L is well-defined, and whether calibrated P_π^L is a full consideration rule. Theorems 1 and 2 provide a generalization of the characterization results in Manzini and Mariotti (2014), Brady and Rehbeck (2016) and Aguiar (2017). Moreover, they provides a unified result for all models that admit a (totally monotone) attention-index representation.²⁷

In practice, we do not observe P , but can consistently estimate it by the collection of sample frequencies \hat{P} . In Section 3.2, we discuss how to test the L-B-rule accounting for sampling variability in \hat{P} .

2.5 Identification

Assuming independence between the distribution of preferences and the random consideration set rule, we uniquely identify the consideration set rule from P if it is a L-B-rule, for all models with totally monotone link functions. Moreover, if there is a positive mass of individuals that consider all alternatives in the choice set, the recoverability of preferences is as good as in the case of full consideration.

²⁷Note that Theorem 2 simplifies the testing problem because it avoids the problem of computing the distribution over choices for every consideration set in every menu. We only need to focus on computing the distribution over choices in each menu.

THEOREM 3 (Identification). *Suppose that for a given model L (i) P is a L-B-rule and (ii) $m_A^L(A) > 0$ for all $A \in \mathcal{A}$. If P is described by (m, π) and (m', π') , then $m = m'$ and $p_\pi = p_{\pi'}$.*

We underline that we achieve a unique decomposition of the data set P into its attention and preference components. Identification of preferences and consideration rules is not a trivial task. Even for simple data sets where stochastic behavior arises from only one channel (e.g., limited consideration), models that only allow for stochastic behavior because of preference heterogeneity (e.g., RUM), or because of random consideration/attention without additional assumptions (e.g., RAM) may fail to identify underlying preferences even when they perfectly describe observed choices. Our framework shows that the recoverability of preferences is as good as the RUM benchmark in stark contrast with RAM, where nothing can be learned about preferences for regular random choice.

3. FRAMES AND TESTING

3.1 Frames

In this section, we introduce another source of variation in the data-attention frames. This additional source of variation will allow us to differentiate between behavior consistent with RUM and behavior consistent with L-B. Following [Salant and Rubinstein \(2008\)](#), we define the extended choice set as a pair of a choice set $A \in \mathcal{A}$ and an attention frame $f \in F$. In this extended environment, for a given frame $f \in F$, we can define a probabilistic choice rule p_f . Similarly, a complete stochastic choice rule with frame $f \in F$ is $P_f = (p_f(a, A))_{A \in \mathcal{A}, a \in A \cup \{o\}}$.

The elements of F contain descriptions of physical items that only vary in presentation but not in the information they contain. Our interpretation of attention frames is the same as [Bhattacharya, Mukherjee, and Sonal \(2021\)](#). These descriptions are available to DMs and should not affect their preferences, but may influence their attention.

EXAMPLE 4 (Stochastic choice rule with frames). Let $X = \{a, b, c\}$ and consider two frames: (i) f that describes $a = 1$ token, $b = 2$ tokens, and $c = 3$ tokens; (ii) and f' that describes $a = 3 - 2$ tokens, $b = 10 - 8$ tokens, and $c = 2 + 1$ tokens. An example of an (incomplete) stochastic choice rule with frame f is $P_f = (p_f(a, \{a, b\}) = 0, p_f(b, \{a, b\}) = 1, p_f(b, \{b, c\}) = 0, p_f(c, \{b, c\}) = 1)$. An example of an (incomplete) stochastic choice rule with frame f' is $P_{f'} = (p_{f'}(a, \{a, b\}) = 1/2, p_{f'}(b, \{a, b\}) = 1/2, p_{f'}(b, \{b, c\}) = 1/2, p_{f'}(c, \{b, c\}) = 1/2)$.

In Example 4, f and f' present the same information about the same alternatives in two distinct ways. In particular, different frames correspond to different numbers of arithmetic operations used to describe the value of the option. The same value is expressed in each frame with a different sequence of arithmetic operations.

The B-rule with frame f can be defined analogously to the definition of the B-rule in Section 2. A complete stochastic choice rule with frame f , P_f , is a B-rule if there exists a

frame dependent distribution over random consideration sets, m_f and a frame dependent distribution over strict linear orders π_f such that

$$p_f(a, A) = \sum_{> \in R(X)} \pi_f(>) \sum_{D \subseteq A} m_{f,A}(D) \mathbb{1}(a > b, \forall b \in D).$$

We take the stand that preferences should not depend on the way the information is presented. Only attention may change due to frames. We formalize these ideas by the assumption of consequentialism, which is a common implicit assumption in the RUM framework of [McFadden and Richter \(1990\)](#).

DEFINITION 10 (Consequentialism). A collection of B-rules $(P_f)_{f \in F}$ described by $((m_f, \pi_f))_{f \in F}$ is said to satisfy consequentialism if there exists π such that $\pi_f = \pi$ for all $f \in F$.

Consequentialism means that an attention frame does not alter the payoffs that a DM obtains from choosing a given alternative. In the same way that the standard rational choice framework imposes frame independence ([Salant and Rubinstein \(2008\)](#)), the classical RUM imposes consequentialism ([McFadden and Richter \(1990\)](#)).

At this point, it is useful to formally define the Random Utility Model (RUM) over the whole choice set $X \cup \{o\}$. RUM treats the default alternative as just another item with no special status. That is, it is not assumed to be a dominated alternative. Let $R(X \cup \{o\})$ be a set of linear orders over the extended choice set $X \cup \{o\}$.

DEFINITION 11 (Random Utility Model, RUM). A collection of complete stochastic choice rules with frame $(P_f)_{f \in F}$ is consistent with random utility if there exists $\pi_o \in \Delta(R(X \cup \{o\}))$ such that

$$p_f(a, A) = \sum_{> \in R(X \cup \{o\})} \pi_o(>) \mathbb{1}(a > b, \forall b \in A),$$

for all $a \in A, A \in \mathcal{A}$, and all $f \in F$.

Note that RUM satisfies consequentialism and π_o does not depend on the frame. The L-B-rule extends the RUM framework to allow for frame dependence only through the random consideration rule.

By assuming that preferences satisfy consequentialism, we are essentially making two assumptions: (i) we assume that the distribution of preferences does not depend on how the alternatives are presented; (ii) consideration in turn can depend on payoff irrelevant information. These assumptions are important, since our experiment has three treatments where a subject must perform a different number of arithmetic operations to evaluate the monetary value of a prize.

These assumptions are auxiliary and are not necessary for either RUM or L-B-rules. However, these assumptions are a reasonable baseline to compare RUM and random consideration models. The reason that these assumptions are not immediately implied is because RUM is often viewed as descriptive. Taking this one step further, one could assume that the distribution of preferences varies with the difficulty of evaluating a task in

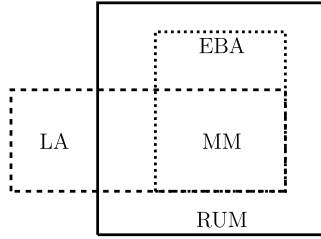


FIGURE 2. Relation among consideration set rules for a fixed frame: RUM, LA, MM, and EBA.

our experiment. This would violate our assumption of consequentialism. While frame-dependent preference could be a reasonable assumption in some settings, it would limit the ability to predict counterfactual choices. For example, one would need to identify preferences in each setting to get a prediction on choices.

Specifically, frame variation allows us to differentiate between L-B-rules and RUM as follows. Without it, LA is not nested by nor nests RUM. For example, LA allows for attraction effect, which violates regularity and, therefore, is inconsistent with RUM. Also, their intersection is nonempty because MM is both consistent with RUM and LA (Manzini and Mariotti (2014), Brady and Rehbeck (2016)). Moreover, for a fixed frame EBA is nested in RUM and nests MM, therefore, its intersection with LA is nonempty (Aguiar (2017)) (see Figure 2).²⁸ However, with frame variation this is no longer true: RUM and EBA intersect, but are not nested within each other.²⁹

3.2 Testing procedure

Theorem 2 allows us to test whether for a given frame f , a given stochastic choice rule P_f is a L-B-rule: it is necessary and sufficient to test whether m_f^L is well-defined (satisfies a set of linear inequalities) and $P_{f,\pi}^L$ is consistent with the full consideration model. Note that the full consideration model is equivalent to the random utility model without outside option. Testing for RUM is a well-understood problem and amounts to solving a quadratic optimization with cone constraints (see McFadden and Richter (1990) and Kitamura and Stoye (2018)). The approach proposed by Kitamura and Stoye (2018) allows us to test these conditions while accounting for sampling variability induced by using \hat{P}_f instead of unknown P_f .

We, however, need to slightly modify the testing procedure in Kitamura and Stoye (2018) to take into account the frame variation. First, we describe the testing procedure for the fixed frame and then extended it to environments with frame variation.

To introduce the testing procedure, we need to define several objects. Note that, for a fixed frame f , the calibrated full consideration rule, $P_{f,\pi}^L$, is a vector of length $d_p = \sum_{k=1}^{|X|} k \binom{|X|}{k}$.³⁰ The k th element of $P_{f,\pi}^L$ corresponds to some pair (a, A) such that $a \in A$.

²⁸Kovach and Suleymanov (2021) shows that MM is the intersection of LA and EBA.

²⁹All of these relationships are preserved when allowing for heterogeneous preferences under the independence assumption of preferences and attention. The reason is that the outside probability does not depend on the distribution of preferences. For more details, see Appendix C.

³⁰ $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ and $n! = 1 \cdot 2 \cdot \dots \cdot n$.

Let B_1 be the matrix of the size $d_p \times |X|!$ such that (k, l) element of it is equal to

$$B_{1,k,l} = \mathbb{1}(a \in A) \mathbb{1}(a \succ_l c, \forall c \in A),$$

where k corresponds to a pair (a, A) such that $a \in A$, and \succ_l is l th linear order on X . We define G_1 as the matrix of size $(d_p + d_m) \times d_1$, where $d_1 = |X|! + d_m$ and $d_m = \sum_{A \subset X} 2^{|A|}$ is the dimension of m_f^L , such that

$$G_1 = \begin{bmatrix} B_1 & 0_{d_p \times d_m} \\ 0_{d_m \times |X|!} & I_{d_m} \end{bmatrix},$$

where $0_{d_p \times d_m}$ denotes the zero matrix of size $d_p \times d_m$, and I_{d_m} denotes the identity matrix of size $d_m \times d_m$. The next result establishes an equivalent characterization of the L-B-rule via m_f^L and $P_{f,\pi}^L$. Let $\mathbb{R}_+^{d_1}$ denote componentwise nonnegative elements of the d_1 -dimensional Euclidean space \mathbb{R}^{d_1} .

THEOREM 4. *For a fixed frame f , the following are equivalent:*

- (i) $P_{f,\pi}^L$ is a FC-B-rule and m_f^L is well-defined;
- (ii) $\inf_{v \in \mathbb{R}_+^{d_1}} \|g_f^L - G_f v\| = 0$, where $g_f^L = (P_{f,\pi}^L, m_f^L)'$.

PROOF. See McFadden and Richter (1990) and Kitamura and Stoye (2018). □

Theorem 4 implies that we can test the null hypothesis that $\inf_{v \in \mathbb{R}_+^{d_1}} \|g_f^L - G_1 v\| = 0$. Fortunately, this testing problem can be directly cast to the testing problem in Kitamura and Stoye (2018).

To take into account the frame variation and consequentialism, we need to modify the matrix G_1 . Let $d_f = |F|$ and B be a matrix of size $d_f \cdot d_p \times |X|!$ that consists of d_f matrices B_1 stacked together. That is, $B = (B_1' B_1' \dots B_1')'$. Let $d = |X|! + d_f \cdot d_m$. Similar to G_1 , define

$$G = \begin{bmatrix} B & 0_{d_f \cdot d_p \times d_f \cdot d_m} \\ 0_{d_f \cdot d_m \times |X|!} & I_{d_f \cdot d_m} \end{bmatrix}.$$

Note that when $d_f = 1$ (i.e., the frame is fixed), then $G = G_1$.

COROLLARY 1. *The following are equivalent:*

- (i) $(P_f)_{f \in F}$ satisfies consequentialism; $P_{f,\pi}^L$ is a FC-B-rule and m_f^L is well-defined for all $f \in F$.
- (ii) $\inf_{v \in \mathbb{R}_+^d} \|g^L - G v\| = 0$, where $g^L = ((P_{f,\pi}^L)_{f \in F}', m_f^L)_{f \in F}'$.

4. THE EXPERIMENT

Our testing approach does not have requirements in terms of repeated individual choice data. Exploiting this feature, our experiment was designed to study the performance of

different theories of random consideration sets with few observations per individual. In particular, we conducted the experiment in Amazon MTurk for a large cross-section with at most two (disjoint) choice sets per individual (see Section 4.1). The large sample size of the data set generated by our experiment is fundamental for ensuring high statistical power using the tools in [Kitamura and Stoye \(2018\)](#).

All sessions were run between August 25, 2018, and September 17, 2018, on the MTurk platform with surveys designed in Qualtrics.³¹ We surveyed 2135 individuals. They were paid on average \$1.09 as a result of \$0.25 for participation fee and the outcome of a randomly selected task that pays a minimum of \$0 and a maximum of \$2. All payments were made in USD. The average duration of the session was 251.68 seconds (slightly over 4 minutes).³² This means that our average payment per hour is roughly \$15.

The payment in our experiment is comparable to other well-known experiments conducted in MTurk. To name a few, [Horton, Rand, and Zeckhauser \(2011\)](#) studied behavior in MTurk using games with the payment range between \$0.40 and \$1.60. They find that behavior in MTurk is consistent with behavior in the lab, where the stakes of games are ten times bigger. They also estimate the median minimum wage in MTurk as \$0.14 per hour.³³ [Dean and McNeill \(2014\)](#) conducted experiments of decision-making. The average payment for completing 15-min long tasks was between \$1.35 and \$1.55 including the show-up fee of \$0.25. [Kim \(2020\)](#) conducted an experiment in MTurk for several weeks with one 10-min task each week. The average earnings from each week's task were below \$1.00. [Rand, Greene, and Nowak \(2012\)](#) also conducted a public good game with MTurkers and the payment range was between \$0.90 and \$1.50 including the show-up fee of \$0.50.

4.1 *Experimental design*

In our experiment, we have two independent sources of exogenous variation: full variation in choice sets and variation in frames. Recall that full variation in choice sets means that all possible choice sets are observed, and variation in frames means that we vary observable information without affecting the rational assessment of the alternatives. These two sources of variation allow us to test consideration-mediated choice theories in a large cross-section of heterogeneous individuals, and differentiate these theories from RUM. The latter is possible since consideration is allowed to change with frames, but preferences must remain stable due to consequentialism. We vary the frame through changes in the cost of consideration.

³¹By clicking the link on the MTurk page, subjects were randomly directed to one of the treatments implemented by Qualtrics. After completing their task, subjects were also asked to complete a short survey regarding their demographic information. Subjects were not allowed to participate in the experiment more than once. Only subjects living in the USA were recruited.

³²The average duration of each task is about 23 seconds, and the duration is significantly correlated with the length of the choice set and the frame.

³³The minimum wage here refers to the reservation wage the MTurk subjects have for performing a given task.

TABLE 1. Lotteries measured in tokens, expected values, and variance.

Lottery	Expectation	Variance	Preference Rank $u(x) = \frac{x^{1-\sigma}}{1-\sigma}$ with σ						
			-2	0	0.25	0.30	0.50	0.75	
(1) $\frac{1}{5}50 + \frac{1}{5}0$	25.000	625.00	1	1	2	5	5	6	
(2) $\frac{1}{7}30 + \frac{1}{7}10$	20.000	100.00	5	5	5	2	1	1	
(3) $\frac{1}{4}50 + \frac{1}{4}30 + \frac{1}{4}10 + \frac{1}{4}0$	22.500	368.75	3	3	4	4	3	4	
(4) $\frac{1}{4}50 + \frac{1}{5}48 + \frac{3}{20}14 + \frac{2}{5}0$	24.125	511.73	2	2	1	3	4	5	
(5) $\frac{1}{5}48 + \frac{1}{4}30 + \frac{3}{20}14 + \frac{1}{4}10 + \frac{3}{20}0$	21.625	251.11	4	4	3	1	2	3	
(o) 12 with probability 1	12.000	0.00	6	6	6	6	6	2	

Choice set design

To induce preference heterogeneity, we consider lottery alternatives with different expected values and variances. Table 1 shows the alternatives and implied preference rankings if DMs are expected utility maximizers with CRRRA Bernoulli utility function. The outside option is dominated for moderate levels of risk aversion (e.g., Holt and Laury (2002)).³⁴

Let $X = \{l_1, l_2, l_3, l_4, l_5\}$ be the set of all nondefault alternatives, and let o be the default/outside option. All menus $A \in \mathcal{A}$ are observed in the sample. The outside option is always present and is shown first, while the order of other alternatives is randomized. Menus can be thought as different treatments.

Our primitive to test L-B is $\hat{P} = (\hat{p}(a, A))_{a \in A \cup \{o\}, A \in \mathcal{A}}$, therefore, we proceeded with stratified sampling, setting the minimal number of observations per choice set to be proportional to its cardinality, that is, $n_A = \lambda(|A| + 1)$ with $\lambda \geq 30$. This design requires a minimum of $\sum_{A \in \mathcal{A}} |A| = 3330$ tasks.

For each menu, the DM faced three consideration frames or cost treatments: High (H), Medium (M), and Low (L). These frames/cost treatments were induced by introducing a k -length two digit addition/subtraction to compute each prize in the lottery. The length k was set equal to 5, 3, and 1, for the high, the medium, and the low cost, respectively. Since in our experiment attention frames only change the complexity of the description of lotteries, we assume that preferences of DMs do not depend on the way alternatives are described, thus are consistent with consequentialism.

The numbers for the cognitive task were randomly generated. Examples can be seen in Figure 3. The default alternative o was presented as is, and there was no need to solve an arithmetic problem to understand it across the different levels of cost.

To prevent possible learning that could attenuate consideration costs, subjects were faced with disjoint choice sets. That is, subjects were either presented with the full choice set and the outside option ($X \cup \{o\}$); or a partition of X (presented at random order), that is, $A_j \cup \{o\}, A_k \cup \{o\}$ with $A_j \cup A_k = X$ and $A_j \cap A_k = \emptyset$. Our experimental design is summarized in Figure 4.

³⁴Recall that the assumption that the default is dominated is a testable assumption in our framework. In addition, without cost of consideration treatments, the outside alternative is easier to understand than the rest because of its simplicity. Hence, it works as a consideration-reference point in the sense of Kovach and Suleymanov (2021).

Select (only) one of the following alternatives. Click over the selected option and hit next [->].

Get 12 tokens for sure.

Prize	Probability
(3-8+22+4 +9) tokens	50%
(5-7 + 41 +10 -38) tokens	50%

Prize	Probability
(27-10 -2 +11 +24) tokens	25%
(25+6 -1+3+15) tokens	20%
(1-18 +3 +17 +11) tokens	15%
(9-6 -7 -4 + 8) tokens	40%

(a) High cost

Select (only) one of the following alternatives. Click over the selected option and hit next [->].

Get 12 tokens for sure.

Prize	Probability
(18+2+10) tokens	50%
(4+5+1) tokens	50%

Prize	Probability
(66+5-21) tokens	25%
(28-24+44) tokens	20%
(10-3+7) tokens	15%
(2-5+3) tokens	40%

(b) Medium cost

Select (only) one of the following alternatives. Click over the selected option and hit next [->].

Get 12 tokens for sure.

Prize	Probability
50 tokens	25%
48 tokens	20%
14 tokens	15%
0 tokens	40%

Prize	Probability
30 tokens	50%
10 tokens	50%

(c) Low cost

FIGURE 3. Consideration cost treatments. Different induced costs for choice set $\{o, l_2, l_4\}$.

The default alternative Our design allows us to use o as the opportunity cost of incurring in the cost of consideration and understanding the other lotteries in the choice set. We use a degenerate lottery as the default due to its simplicity. In this sense, we believe the alternative o in our design has effectively zero cost of consideration. Recall, that for any choice set/frame the outside option is always present and shown first. Moreover, it is preselected as the default alternative. If the subject decides to skip the task, she is informed that o will be chosen for her.

4.2 Sample

The sample consists of 2135 individuals that selected alternatives from one or two choice sets for all costs of attention, as shown in Figure 4, for a total of 12,297 observations. The number of observations per alternative/choice set are shown in Table 4 in Appendix D.

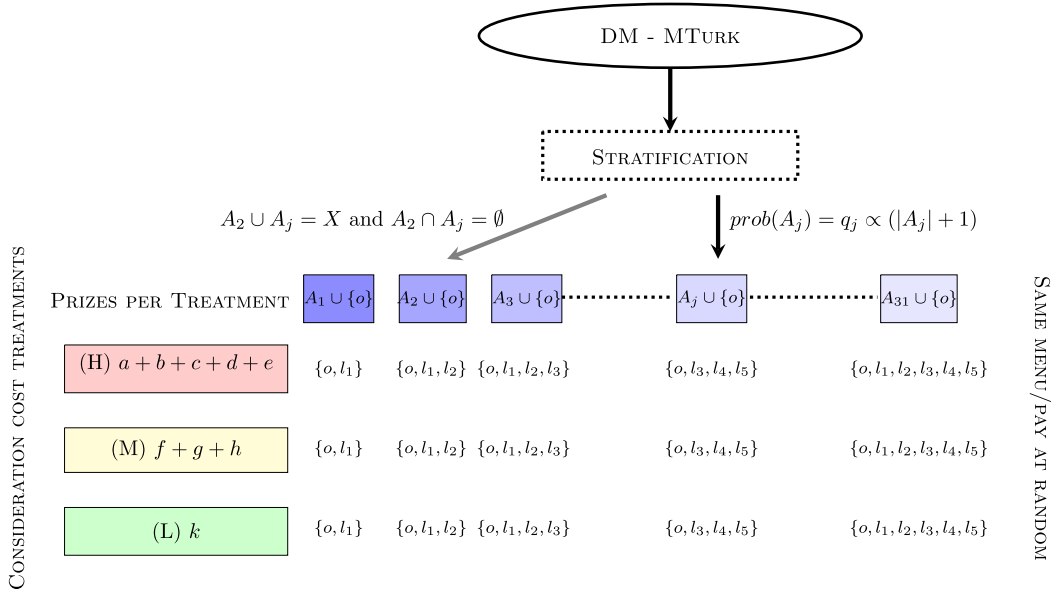


FIGURE 4. Experimental design: DM i draws with probability $p(A_j)$ menu A_j with $|A_j| \in \{3, 4, 5\}$. In the picture, $A_j = \{l_3, l_4, l_5\}$. Therefore, she is asked to choose from menus $A_j \cup \{o\}$ and $A_2 \cup \{o\}$, since $A_2 \cup A_j = X$ and $A_2 \cap A_j = \emptyset$. She is faced with one of these menus first (randomly selected) and asked to choose when the cost is H, M, and L. Then she faced the other menu for the three cost treatments.

Based on these observations the primitive for our analysis is the collection of observed frequencies $(\hat{p}(a, A))_{a \in A, A \in \mathcal{A}}$. We compute these frequencies for all costs. Unless otherwise stated, $\hat{p}(a, A)$ refers to observed frequency in the data pooled across attention costs.

Figure 5 summarizes the distribution of gender, age, education, ethnicity, labor, and income in our sample. Our subjects are a diverse sample of US individuals. By design, demographics are balanced across consideration cost treatments and choice sets (that can also be thought of as treatments).

4.3 Descriptive analysis: Evidence for costly consideration and frame effects

In this section, we describe the behavior of individuals in our sample and present suggestive evidence that our cost treatments effectively induce costly consideration or frame effects. In particular, we observe that the consideration cost treatments/frames: (i) affect the choice frequency of the outside option; (ii) have a heterogeneous effect on the choice frequencies of all other alternatives; (iii) and affect the patterns of choice with respect to the size of the menu. Moreover, all these effects depend monotonically on the level of difficulty of choice induced by each treatment.

Under the null hypothesis of full consideration and consequentialism, the observed frequency of choice of the outside option should remain the same across frames. The reason is that the choice menu remains the same across cost treatments, and payment is

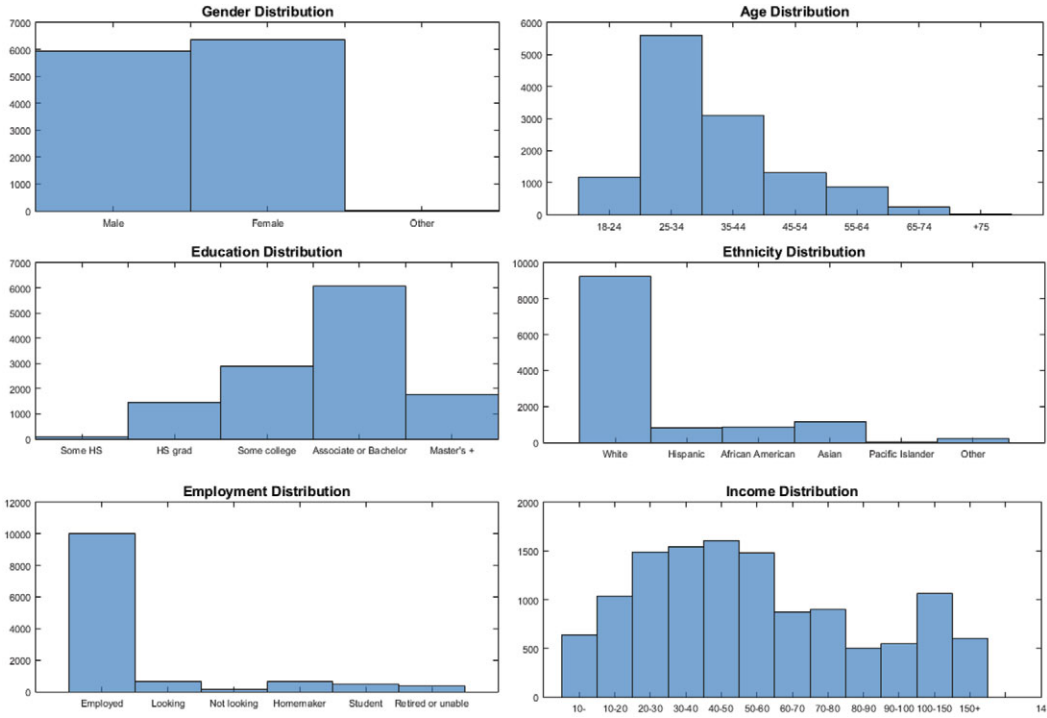


FIGURE 5. Distribution of demographics in sample.

at random. However, the outside option is chosen more often as the cost increases (see Figure 6). This is evidence against full consideration and in favor of frame-dependent choice.

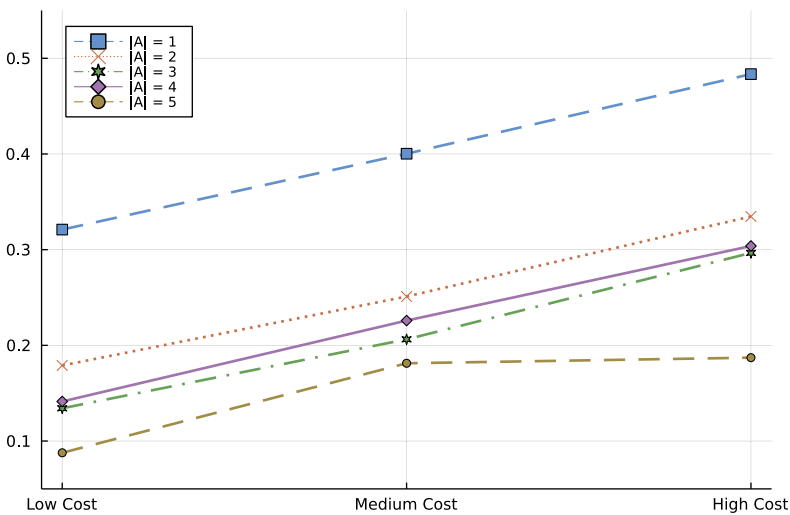


FIGURE 6. Frequency of the outside option conditional on the menu size and frame.

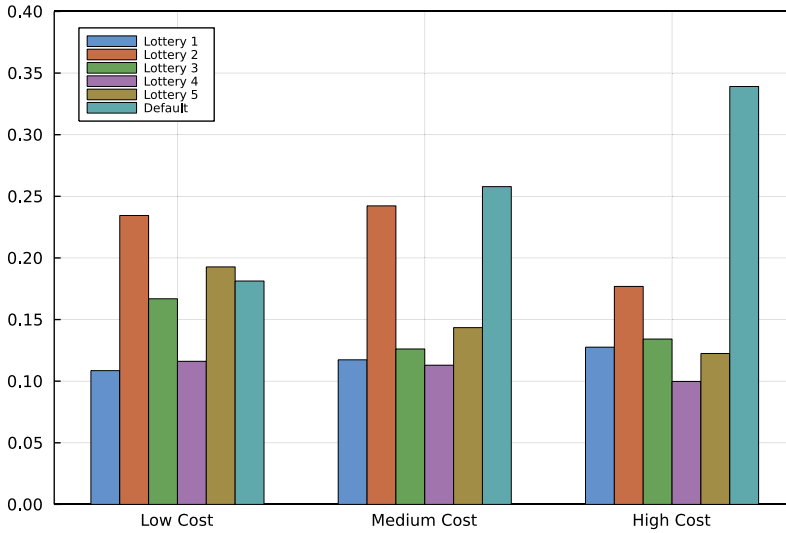


FIGURE 7. Frequencies of choices for different lotteries and frames.

We remind the reader here that low, medium, and high cost corresponds to 1, 3, and 5 arithmetic operations required to understand the monetary prize of each lottery, respectively. The monotone relation shows that our treatments were effective, and that the frequency of choice of the default is in fact ordered in the way it was expected.

Figure 7 shows the effect of the frames/cost-treatments on the choice frequencies. The harder it is to understand the lotteries,³⁵ the more likely subjects opt to not consider them and instead choose the outside option. These results support the effectiveness of the induced treatments. Figure 7 also shows that the effect of the cost treatment is not homogeneous across alternatives. The choice frequency of lottery 1 (the simplest to understand after the default) increases with the cost treatment; the cost does not have a significant impact on the probability of selecting lottery 4; while it has a negative impact on the other lotteries.

Overall, the probability of selecting any given lottery does not necessarily decrease with the size of the menu, suggesting that regularity is violated in our sample, as shown in Figure 8. Indeed, we test RUM later and confirm that these violations are significant, since we reject RUM at the 5% significance level.

Notice that, for a fixed frame/attention cost, some lotteries are harder to compute than others. For instance, lottery 2 can pay 30 or 10 tokens, while lottery 3 can pay 50, 30, 10, or 0. So, for every frame, lottery 3 is harder to compute than lottery 2. Our attention-index framework allows for alternatives or lotteries with heterogeneous complexity (e.g., the simplest model MM has a lottery-specific attention parameter). For instance, the simplest lottery 1 is picked more often as the cost rises, while the hardest to compute

³⁵Here, simplicity comes in the form of how easy (number of arithmetic operations) it is to compute expectation, variance, and expected utility of the lottery in terms of the number of prizes and whether the probabilities are uniform on the support of the lottery or not.

lottery 5 displays the opposite pattern. These observations confirm the presence of heterogeneous attention patterns.

4.4 Evidence for total monotonic attention

Recall that, under the null hypothesis, the data set is generated by an attention-index model with a link function that is totally monotone. Hence, the frequency of choosing the outside option is equal to a monotone transformation of cumulative attention associated with the attention-index $\sum_{C \subseteq A} \eta(C)$:

$$p(o, A) = m_A(\emptyset) = \varphi\left(\sum_{C \subseteq A} \eta(C), \eta(\emptyset)\right).$$

In Figure 8, we observe that the frequency of choice of the outside option *is not* a decreasing function of the cardinality of the choice set. The latter is inconsistent with total monotonicity, and thus, we may naively conclude that neither RUM, LA, nor EBA can explain our data set. However, as we confirm in the next section, this slight violation of monotonicity is an artifact of sample variability, and it is not statistically significant.

Choice overload refers to the case when the propensity of not choosing any alternative (i.e., the probability of picking the default alternative) increases with the size of the choice sets (Iyengar and Lepper (2000)). Our findings are informative on whether this effect, which may be present at the individual level, still matters at the population level. Note that neither RUM nor any model of limited consideration that we study (including Cattaneo et al. (2020)) can rationalize choice overload.³⁶ Existing models of limited consideration are fundamentally at odds with choice overload, since one of the important reasons to form a consideration set is to simplify choice.³⁷ We find no statistical support for choice overload in our data set. However, our choice set is of moderate size.

4.5 Differentiating between LA and EBA

In Figure 8, we also plot the frequencies of choice of the nondefault alternatives and find that there are violations of total monotonicity (e.g., lottery 3). This evidence suggests that EBA cannot describe this data set because total monotonicity should hold for all alternatives (Block and Marschak (1960), Aguiar (2017)). However, this is not enough to conclude that EBA cannot describe the data set because of sample variability. Nonetheless, we reject the null hypothesis that EBA can describe this data set. In addition, we must highlight that LA is the only candidate that can accommodate the observed non-monotonicity of the nondefault alternatives observed in this data set. We confirm this insight in our formal testing section by showing that LA does a good job at describing this data set.

³⁶Chernev, Böckenholt, and Goodman (2015) provides a meta-data analysis of the determinants of choice overload.

³⁷Other models of stochastic choice that are not models of limited consideration usually can accommodate choice overload; see, Fudenberg, Iijima, and Strzalecki (2015), Echenique, Saito, and Tserenjigmid (2018), Natenzon (2019), and Kovach and Tserenjigmid (2022).

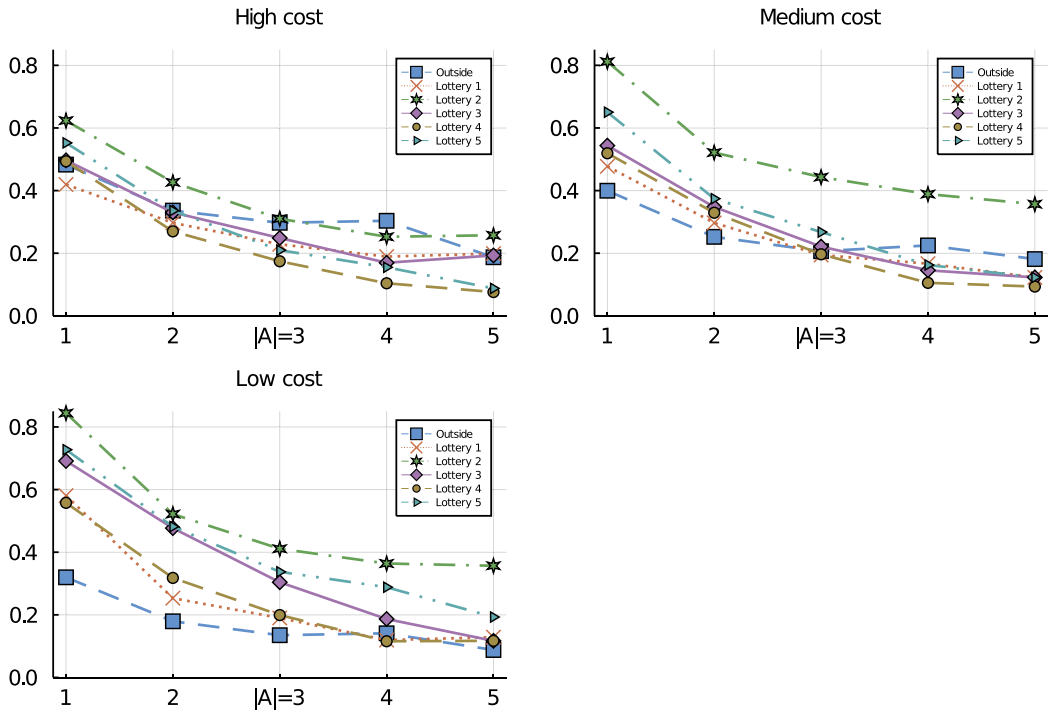


FIGURE 8. Frequency of choice for all frames as a function of the menu size.

The nonmonotonicity that we observe in Figure 8 is usually called *attraction effect* (Huber, Payne, and Puto (1982)). The attraction effect refers to a phenomenon when, as a new alternative is added to the choice set, the probability of the existing items is boosted. Our findings support the presence of the attraction effect in our experimental sample.

5. TESTING RANDOM CONSIDERATION MODELS

In this section, we report the results of testing the ability of RUM, LA, and EBA to describe our experimental data. We test these models without imposing any restrictions on preferences except stability over frames. Unless otherwise stated, the tested hypothesis is that, for a particular specification of our model (consideration set stochastic rule), there exists (m, π) that is a L-B representation for behavior.

5.1 Econometric testing

For every frame f , although P_f is not observed, the realized choice frequencies \hat{P}_f are. For every $A \in \mathcal{A}$ and $f \in F$, let $n_{f,A}$ denote the number of individuals in the sample that faced choice set A and frame f , and let $\mathbf{a}_{i,f,A}, i = 1, \dots, n_A$ be the observed choice of individual i from choice set $A \cup \{o\}$ and frame f . We assume that the researcher observes a cross-section of observations (i.e., i.i.d. observations) for every menu and frame.³⁸ Then

³⁸For a given frame, this is a standard stochastic choice data set in the literature on limited consideration.

we define the estimated stochastic choice rule as

$$\hat{P}_f = (\hat{p}_f(a, A))_{A \in \mathcal{A}, a \in A \cup \{o\}}$$

with $\hat{p}_f(a, A) = n_{f,A}^{-1} \sum_{i=1}^{n_{f,A}} \mathbb{1}(\mathbf{a}_{i,f,A} = a)$ for any $a \in A \cup \{o\}$.

Given the model of interest L and the estimator of P_f, \hat{P}_f , we can compute the estimators of m_f^L and $P_{f,\pi}^L, \hat{m}_f^L$ and $\hat{P}_{f,\pi}^L$, using Definitions 7 and 9.³⁹ Given the results of Corollary 1, a natural test statistic is

$$T_n = n \min_{v \in \mathbb{R}_+^d} \|\hat{g}^L - Gv\|^2,$$

where $n = \min_f(\sum_A n_{f,A})$ is the smallest sample size across frames and $\hat{g}^L = ((\hat{P}_{f,\pi}^L)'_{f \in F}, (\hat{m}_f^L)'_{f \in F})'$.

Let $\hat{g}_l^{L,*}, l = 1, \dots, L$, be bootstrap replications of \hat{g}^L . Let $\tau_n \geq 0$ be a tuning parameter and ι be a vector of ones of dimension d .⁴⁰ To compute the critical values of T_n , we follow the bootstrap procedure proposed in Kitamura and Stoye (2018):

- (i) Compute $\hat{\eta}_{\tau_n} = Gv_{\tau_n}$, where v_{τ_n} solves

$$n \min_{[v - \tau_n \iota / d] \in \mathbb{R}_+^d} \|\hat{g}^L - Gv\|^2;$$

- (ii) Compute the bootstrap test statistic

$$T_{n,l}^* = n \min_{[v - \tau_n \iota / d] \in \mathbb{R}_+^d} \|\hat{g}_l^{L,*} - \hat{g}^L + \hat{\eta}_{\tau_n} - Gv\|^2, \quad l = 1, \dots, L;$$

- (iii) Use the empirical distribution of the bootstrap statistic to compute critical values of T_n .

For a given significance level $\alpha \in (0, 1/2)$, the decision rule for the test is “reject the null hypothesis if $T_n > \hat{c}_{1-\alpha}$,” where $\hat{c}_{1-\alpha}$ is an $(1 - \alpha)$ -quantile of the empirical distribution of the bootstrap statistic.

We would like to conclude this section by observing that we can test the model conditional on additional observables (e.g., gender, income brackets, and education level). For discrete (or discretized) covariates, one just needs to perform the test for a subgroup of the population.

5.2 Survival race the L-B-rule versus RUM: Stability of preferences

Without frame variation, many models of consideration are empirically indistinguishable from RUM. For instance, if a data set is consistent with EBA or MM, for a fixed

³⁹To compute $\hat{P}_{f,\pi}^L$, in our empirical application, we minimized the Euclidean distance between \hat{P}_f and $(\sum_{D \subseteq A} \hat{m}_f^L p_{f,\pi}(a, D))_{a \in A, A \in \mathcal{A}}$ subject to $p_{f,\pi}(a, A) \geq 0, \sum_{a \in D} p_{f,\pi}(a, D) = 1$ for all a and A , and $p_{f,\pi}(a, D) = 0$ for all D and $a \notin D$.

⁴⁰In our empirical application, we conducted tests for different values of τ_n (e.g., $\tau_n = \sqrt{\frac{\log(\min_{f,A} n_{f,A})}{\min_{f,A} n_{f,A}}}$ following Kitamura and Stoye (2018), and $\tau_n = 0$). The results are qualitatively the same.

TABLE 2. Testing results under preference stability.

Model	T_n	p -value
RUM	3231.59	<0.001
LA	24,959.06	0.524
EBA	24,840.23	0.001

Note: Number of bootstrap replications = 1000.

frame, then it is also consistent with RUM. However, these models and RUM will typically recover a distinct distribution of preferences. Varying frames, we can test whether the distribution of preferences remains the same across frames.

L-B-model assumes that the distribution of preferences in the population is independent of the consideration rule. In our experiment, the choice sets faced by any subject are exactly the same for the three consideration cost treatments. Given our pay-at-random incentives scheme, choices from each choice set can be considered as i.i.d. draws from the underlying random utility distribution under the null hypothesis of stochastic rationality. Therefore, the independence assumption together with our experimental design imply that if one of the L-B theories describes the behavior of the high-cost treatment, it must also describe the behavior of the low-cost treatment. That is, if the independence assumption holds, then the distribution of preferences, π , should be invariant to changes in consideration costs for theories that we cannot reject. We check the validity of the different models of interest under this preference stability restriction.

We apply the procedure described in Section 5.1 to test whether the RUM, LA, and EBA models can explain the data with the restriction that the distribution of preferences remains frame-independent (i.e., consequentialism). The results of testing are presented in Table 2. In this table, we report the values of the test statistic and the corresponding p -values coming from the bootstrap distribution (1000 bootstrap replications for every test statistic were conducted) for different models.⁴¹ First, we strongly reject RUM at any reasonable significance level. In other words, for RUM we reject the hypothesis that the same distribution of preferences can rationalize behavior across consideration cost treatments. In contrast, we cannot reject the LA model at any standard significance level. In addition, we reject the hypothesis that EBA explains the population behavior under preference stability. Note that we can discriminate between RUM and EBA because of variation in frames—EBA is more general than RUM because it allows for flexible attention per frame. So, the rejection of EBA with stable preferences does not follow from the rejection of RUM. Taken together, our results show that our experimental subjects behave as if they are maximizing their preferences given a consideration set that follows the LA restriction.

⁴¹The p -value is interpreted as the probability of observing a realization of the test statistic that is above the one that is actually observed due to sample variability, if the null hypothesis is indeed correct. Then the smaller the p -value is, the more evidence the researcher has to reject the hypothesis of the validity of a given model.

5.3 Discussion

Our findings strongly support the hypothesis that the population behaves as if it is consistent with the LA model of limited consideration and has a stable distribution of preferences across frames.⁴² All frame effects observed in our descriptive analysis are fully captured by the variation in the random consideration rule that changes conditional on the frame. In contrast, the traditional RUM fails to describe the population behavior. To confirm that our testing procedure has power against LA, in Appendix E, we assess the performance of our procedure using the Monte Carlo simulation. In particular, we show that our test can reject the false null hypothesis of data being consistent with the LA model with high frequency in finite samples that are comparable to our experiment.

We highlight that our analysis cannot exclude the possibility that other models of behavior could also explain the population behavior in our experiment. We have only established that the population behavior is *as if* it is consistent with a L-B-rule.

Although we do not impose any restrictions on preferences, for example, expected utility, our results relate to the work of Freeman et al. (2019). They provide an alternative mechanism for the selection of a riskless lottery (default) over dominant risky choices from pairwise comparisons, when binary choice sets are presented as lists. They propose a theoretical explanation of the choice of the riskless choice with a model of reference dependence. The class of reference dependence models used by these authors is a special case of utility maximization. Recall that we find evidence against RUM in our experiment, thus ruling out Freeman et al. (2019) mechanism for our environment with costly consideration. In addition, in our experimental design, subjects are not required to choose from lists nor are restricted to pairwise comparisons. For an extended discussion of the role of misperception, see Appendix C.3.

We have maintained the assumption that the default alternative is also the worst alternative for both the LA and EBA models. However, RUM allows the default to be ranked arbitrarily. Hence, the main findings that LA explains the data set and RUM fails to do the same are robust to this assumption. We leave it as an open question whether EBA can explain this data set if this assumption is relaxed. Nevertheless, we believe that this assumption is reasonable in our experimental setup.

We have done our empirical analysis without conditioning on observable heterogeneity (e.g., age or gender). Attention and preferences may differ across different demographic groups. Methodologically, our tools can be applied after conditioning on observable heterogeneity, as explained in Kitamura and Stoye (2018). The study of consideration set rules and their relation to demographics is beyond the scope of this paper.

It is noteworthy that if individuals have convex risk preferences to mix between lotteries, repetition of discretized choice tasks may make it hard to estimate risk preferences. Feldman and Rehbeck (2022) show that subjects who mix between lotteries in convex budget sets are more likely to randomize choices in a repeated discretized task. Our experiment that uses different cost-treatments and randomizes the order of lotter-

⁴²We also rejected the hypothesis of whether the LA model with homogeneous preferences can explain the data. See Appendix F.

ies in presentation may alleviate such concerns compared to previous experiments in which the same set of lotteries in a fixed order was repeatedly presented to subjects.⁴³

We finish this section by discussing our model and our findings in relation to Rational Inattention (RI) models. [Caplin, Dean, and Leahy \(2019\)](#) shows that rational inattentive DMs form (deterministic) consideration sets. Generally, RI primitives cannot be point-identified with standard stochastic choice data sets. Nonetheless, RI models may still have testable implications in standard stochastic choice data sets. In [Appendix C.4](#), we show that a representative RI DM is compatible with deterministic consideration sets (i.e., the presence of zero probability of choice), which is not supported in our data. The case of a population of heterogeneous rational inattentive DMs and the aggregation of such behavior in the population is left for future research.

6. CONCLUSION

We have designed and implemented a novel experiment with a large sample that allowed us to statistically discern among competing models of population behavior. By exogenously varying choice sets and the frames induced by the cost of considering alternatives, we can disentangle two sources of stochastic behavior: limited consideration and preference heterogeneity. We use this novel data set to test RUM and two models of limited consideration, LA and EBA.

These models provide testable implications on choices that uniquely identify the stochastic consideration set rule from data. By calibrating consideration given the theory, we show that testing the L-B-model can be cast into [Kitamura and Stoye \(2018\)](#) framework for testing RUM. That is, we show that there exists a stochastic rule (computed from data) that is RUM if and only if observed choices are generated by a population of individuals consistent with the L-B-model.

We provide evidence against classical RUM, since consideration costs are binding for some individuals in the population. In contrast, we find support for the LA model with heterogeneous preferences. Crucially, we cannot reject that the distribution of preferences implied by LA is the same across all attention frames. This means that once we disentangled attention and preferences under LA, the recovered distribution of preferences does not change with the frame.

APPENDIX A: PROOFS

A.1 Proof of Lemma 1

We define $m_A(\{a\}) = p(a, A)$, and $m_A(D) = 0$ for all $D \subseteq A, D \neq \{a\}$. Let $\tilde{\pi} \in \Delta(R(X))$ be the uniform distribution. The pair $((m_A)_{A \in \mathcal{M}}, \tilde{\pi})$ is a B-rule. We now prove that it generates any data P . By definition, if P can be generated by a B-rule, then we have that

$$p(a, A) = \sum_{D \subseteq A} m_A(D) \sum_{\succ \in R(X)} \tilde{\pi}(\succ) \mathbb{1}(a \succ b, \forall b \in D)$$

⁴³We thank an anonymous referee for pointing out this issue.

for all A and $a \in A$. Rearranging and replacing the choice of $\tilde{\pi}$ in the above equation, we get that

$$\sum_{D \subseteq A} m_A(D) \sum_{\succ \in R} \tilde{\pi}(\succ) \mathbb{1}(a \succ b, \forall b \in D) = \frac{1}{|R(X)|} \sum_{\succ \in R(X)} \left[\sum_{D \subseteq A} m_A(D) \mathbb{1}(a \succ b, \forall b \in D) \right].$$

For given \succ and $m_A(\{a\}) = p(a, A)$, we have

$$\sum_{D \subseteq A} m_A(D) \mathbb{1}(a \succ b, \forall b \in D) = p(a, A) \mathbb{1}(a \succ a) = p(a, A)$$

because \succ includes the diagonal $a \succ a$ for all $a \in X$.

This implies that

$$\frac{1}{|R(X)|} \sum_{\succ \in R(X)} \left[\sum_{D \subseteq A} m_A(D) \mathbb{1}(a \succ b, \forall b \in D) \right] = p(a, A)$$

given that

$$\frac{1}{|R(X)|} \sum_{\succ \in R(X)} \left[\sum_{D \subseteq A} m_A(D) \mathbb{1}(a \succ b, \forall b \in D) \right] = \frac{1}{|R(X)|} \sum_{\succ \in R(X)} p(a, A) = p(a, A).$$

A.2 Proof of Theorem 1

(i) implies (ii). A complete stochastic choice rule P is a B-rule if there exists a pair (m, π) such that

$$p(a, A) = \sum_{D \subseteq A} m_A(D) \sum_{\succ \in R(X)} \pi(\succ) \mathbb{1}(a \succ b, \forall b \in D),$$

for all $a \in X$ and $A \in \mathcal{A}$, where we exchanged the summation operator with respect to the consideration sets and the linear orders exploiting independence.

Note that we can write the probability of the default alternative as $p(o, A) = 1 - \sum_{a \in A} p(a, A)$. This implies that

$$1 - p(o, A) = \sum_{D \subseteq A} m_A(D) \left[\sum_{a \in A} \sum_{\succ \in R(X)} \pi(\succ) \mathbb{1}(a \succ b, \forall b \in D) \right],$$

where the summation operator with respect to the items $a \in A$ can be exchanged with the summation over consideration sets. This is possible because the latter summation does not depend on the items $a \in A$.

Now, we notice that $\sum_{a \in A} \sum_{\succ \in R(X)} \pi(\succ) \mathbb{1}(a \succ b, \forall b \in D) = 1$ for all $D \subseteq A$. This implies that the default probability does not depend on the distribution of preferences and can be written in terms of the cumulative distribution of the consideration set distribution:

$$p(o, A) = 1 - \sum_{D \subseteq A, D \neq \emptyset} m_A(D).$$

We let the capacity $\varphi^* : 2^X \rightarrow [0, 1]$ be defined by $\varphi^*(A) = p(o, A)$.

The fact that $\eta = \eta^L$ under the correct specification of the link function follows from our monotonicity assumptions and the existence of a unique Mobius inverse of the mapping $v(\cdot) = \sum_{C \subseteq \cdot} \eta(C)$ (Shafer (1976), Chateauneuf and Jaffray (1989)). We provide specific derivations for each of the models of interest in this paper, to connect them with the existing literature, but they follow directly from the general η^L formula.

For given $L \in \{\text{LA, MM, EBA, FC}\}$ and P :

- If $m \in \mathcal{M}^{\text{LA}}$, then $m_A(D) = \frac{\eta(D)}{\sum_{C \subseteq A} \eta(C)}$ for some $\eta \in \Delta(2^X) \cap \mathbb{R}_{++}$.

This means that $\frac{\varphi^*(X)}{\varphi^*(A)} = \sum_{D \subseteq A} \eta(D)$. Then by Shafer (1976) it must be that

$$\eta(D) = \sum_{B \subseteq D} (-1)^{|D \setminus B|} \frac{\varphi^*(X)}{\varphi^*(B)} = \sum_{B \subseteq D} (-1)^{|D \setminus B|} \frac{p(o, X)}{p(o, B)};$$

- If $m \in \mathcal{M}^{\text{MM}}$, then $m_A(D) = \frac{\eta(D)}{\sum_{C \subseteq A} \eta(C)}$ for some $\eta \in \Delta(2^X) \cap \mathbb{R}_{++}$ with

$$\eta(D) = \prod_{a \in X \setminus D} (1 - \gamma(a)) \prod_{b \in D} \gamma(b),$$

and $\gamma : X \rightarrow (0, 1)$. This implies by simple computation that

$$\gamma(a) = 1 - \frac{\varphi^*(A)}{\varphi^*(A \setminus \{a\})} = 1 - \frac{p(o, A)}{p(o, A \setminus \{a\})}$$

for some $A \in \mathcal{A}$ that contains a ;

- If $m \in \mathcal{M}^{\text{EBA}}$, then $m_A(D) = \sum_{C: C \cap A = D} \eta(C)$ for some $\eta \in \Delta(2^X)$. Then

$$\varphi^*(A) = \sum_{D \cap A \neq \emptyset} \eta(D).$$

Using Shafer (1976) and Chateauneuf and Jaffray (1989), we conclude that

$$\eta(D) = \sum_{A \subseteq D: D \in \mathcal{A}} (-1)^{|D \setminus A|} (1 - \varphi^*(X \setminus A)) = \sum_{A \subseteq D: D \in \mathcal{A}} (-1)^{|D \setminus A|} (p(o, X \setminus A));$$

- If m is FC, then obviously $m_A(D) = \mathbb{1}(A = D)$.

To establish that $m = m^L$ for given $L \in \{\text{LA, MM, EBA, FC}\}$ and P , we exploit the uniqueness of m , which is a consequence of the invertibility of the Mobius transform and the completeness of P . In particular, if (m, π) and (m', π) represent the same P , then it must be that $m' = m$ for the cases of $L \in \{\text{LA, MM, EBA, FC}\}$. To see that this is true, recall that if P is a L-B-rule with (m, π) , then $1 - \sum_{D \subseteq A, D \neq \emptyset} m_A(D) = \varphi^*(A)$. This is exactly the same for the case where there is homogeneity in the preferences such that there is a linear order $\succ \in R(X)$ such that $\pi(\succ) = 1$. Since this equivalence does not depend on the distribution of preferences and due to the completeness of the data set, we can use this fact to apply known results from the consideration set literature regarding the uniqueness of m .

Now, by the Mobius inverse, it follows that

$$\eta^{LA}(D) = \sum_{B \subseteq D} (-1)^{|D \setminus B|} \frac{p(o, X)}{p(o, B)},$$

for all $D \in 2^X$. In particular,

- By Theorem 3.1 in Brady and Rehbeck (2016), it must be that m is uniquely identified by

$$m_A(D) = \frac{\eta(D)}{\sum_{C \subseteq D} \eta(C)},$$

where $\eta \in \Delta(2^X) \cap \mathbb{R}_{++}$ follows from the requirement that $\sum_{B \subseteq D} (-1)^{|D \setminus B|} \frac{p(o, X)}{p(o, B)} > 0$ for all $D \in 2^X$.

- Given $\gamma^{MM}(a) = 1 - p(o, a) \in (0, 1)$ for all $a \in X$ (which is well-defined by the completeness of P) and $\eta^{MM}(D) = \prod_{a \in X \setminus D} (1 - \gamma^{MM}(a)) \prod_{b \in D} \gamma^{MM}(b)$, it follows that m is uniquely identified by

$$m_A(D) = \prod_{a \in D} \gamma^{MM}(a) \prod_{b \in A \setminus D} (1 - \gamma^{MM}(b)),$$

for all $A \subseteq D$. Note that $\prod_{b \in \emptyset} \gamma^{MM}(b) = 1$ by convention. Also observe that uniqueness follows from Theorem 3.3 in Brady and Rehbeck (2016) since the MM restriction is a special case of the LA restriction.

- Given $\eta^{EBA}(D) = \sum_{A \subseteq D: D \in A} (-1)^{|D \setminus A|} (1 - p(o, X \setminus A)) \geq 0$ it follows by Theorem 1 in Aguiar (2017) that m is uniquely identified by

$$m_A^{EBA}(D) = \sum_{C: C \cap A = D} \eta^{EBA}(C),$$

for all $D \subseteq A$, where $D \neq \emptyset$ and $m_A(\emptyset) = 1 - \sum_{D \subseteq A, D \neq \emptyset} m_A(D)$.

- The case of FC is trivial.

A.3 Proof of Theorem 2

(i) implies (ii). If P is a L-B-rule then by Theorem 1, under conditions (i) and (ii), it must be that

$$p_{m, \pi}^L(a, A) = \frac{p_{m, \pi}(a, A) - \sum_{C \subset A} m_A^L(C) p_{\pi}^L(a, C)}{m_A^L(A)},$$

where $p_{m,\pi}(a, A) = \sum_{D \subseteq A} m_A^L(D) [\sum_{\succ \in R(X)} \pi(\succ) \mathbb{1}(a \succ b \forall b \in D)]$. Following the recursive formula, we can show that

$$p_{\pi}^L(a, A) = \frac{m_A^L(A) \left[\sum_{\succ \in R(X)} \pi(\succ) \mathbb{1}(a \succ b \forall b \in A) \right]}{m_A^L(A)} = \sum_{\succ \in R(X)} \pi(\succ) \mathbb{1}(a \succ b \forall b \in A).$$

This implies that P^L is a FC-B-rule.

(ii) implies (i). Under conditions (i) and (ii), the fact that

$$p_{\pi}^L(a, A) = \frac{p_{m,\pi}(a, A) - \sum_{C \subset A} m_A^L(C) p_{\pi}^L(a, C)}{m_A^L(A)}$$

implies that for all $A \in \mathcal{A}$ and all $a \in A$,

$$p(a, A) = \sum_{D \subseteq A} m_A^L(D) p_{\pi}^L(a, D).$$

If P^L is a FC-B-rule, then it implies that there exists $\pi \in \Delta(R(X))$ such that

$$p_{\pi}^L(a, A) = \sum_{\succ \in R(X)} \pi(\succ) \mathbb{1}(a \succ b \forall b \in A).$$

Hence, P is a L-B-rule. In fact, for all $A \in \mathcal{A}$ and all $a \in A$, it must be that the pair (m^L, π) generates the data set P :

$$p(a, A) = \sum_{D \subseteq A} m_A^L(D) \sum_{\succ \in R(X)} \pi(\succ) \mathbb{1}(a \succ b \forall b \in A), .$$

A.4 Proof of Theorem 3

We first prove that if P is described by (m, π) and (m', π') , then it must be that $m = m'$. This follows from [Chateauneuf and Jaffray \(1989\)](#). In particular, [Brady and Rehbeck \(2016\)](#) shows the identification results for $L = LA$, while [Aguiar \(2017\)](#) provides identification results for $L = EBA$. For $L = MM$, the result holds trivially.

Fixing m , if P is described by both (m, π) and (m, π') , then

$$p_{\pi}^L(a, A) = \frac{p(a, A) - \sum_{C \subset A} m_A^L(C) p_{\pi}^L(a, C)}{m_A^L(A)},$$

and

$$p_{\pi'}^L(a, A) = \frac{p(a, A) - \sum_{C \subset A} m_A^L(C) p_{\pi'}^L(a, C)}{m_A^L(A)},$$

for all $a \in A$ and nonempty $A \subseteq X$, which follows from [Definition 9](#). By condition (ii), $m_A^L(A) > 0$ and using the recursive definitions above for binary sets, we can see that

$p_{\pi}^L(a, \{a, b\}) = p_{\pi'}^L(a, \{a, b\})$ for any $a, b \in X$. For a fixed m , the recursive formula leads to the equivalence $p_{\pi'}^L = p_{\pi}^L$.

APPENDIX B: SENSITIVITY ANALYSIS FOR THE DEFAULT

One key assumption in our setup is that the outside option is only picked under full consideration if the choice set only contains the outside option. In our notation, we write this as $p_{\pi}(o, \emptyset) = 1$ and $p_{\pi}(o, A) = 0$ for all $A \neq \emptyset$. In this section, we relax this assumption to allow the probability of choosing the default to satisfy $p_{\pi}(o, \emptyset) = 1$ and $p_{\pi}(o, A) = \frac{e}{|A|}$ with $e \in [0, 1)$ for all $A \neq \emptyset$. The sensitivity parameter e is interpreted as the fraction of DMs that choose the dominated default even when there are other available alternatives. This is a violation of the assumption that the default is the worst alternative. Then this implies that under the null of consistency with the B-rule the probability of choosing the default is

$$p(o, A) = m_A(\emptyset) + \frac{e}{|A|}(1 - m_A(\emptyset)).$$

This assumption is compatible with RUM and implies that the probability of o being the best alternative in all linear orders is constant across menus under full consideration.⁴⁴ Notice that under this assumption we can calibrate $m_A(\emptyset)$ for all $A \in \mathcal{A}$:

$$m_A(\emptyset) = \frac{p(o, A) - \frac{e}{|A|}}{\left(1 - \frac{e}{|A|}\right)}.$$

Then for a given sensitivity parameter $e \in [0, 1)$, we can calibrate the empirical attention-index η^L without any changes. Next, we can compute p_{π} given the calibrated m^L , for $A \cup \{o\}$, and we can test RUM here using the tools we have developed. In our current empirical results, since the LA model passes for $e = 0$, it is unnecessary to do this sensitivity analysis.

APPENDIX C: COMPARISON WITH MODELS OF STOCHASTIC CHOICE

In this Appendix, we analyze the connection between the three consideration-mediated choice theories discussed in this paper and models that allow for stochastic behavior exclusively in preferences or in consideration.

C.1 Comparison with RUM

As explained previously, randomness arising from limited consideration as in EBA and MM can be rationalized under the umbrella of random utility for a fixed frame. However, LA allows for behavior that is inconsistent with regularity. Therefore, LA is not nested in RUM. By construction, our model L-B generalizes FC by allowing for independent

⁴⁴Indeed, this assumption allows the default to be chosen when compared to other alternatives. This is achieved by putting a mass (equal to $e/|A|$) on the event that o is the first among all alternatives in a given menu. This is a restriction on the distribution of preferences.

variation in choices due to limited consideration. In particular, L-B is RUM defined over X (what we call, equivalently, FC) when the stochastic choice rule is such that $m_A(D) = \mathbb{1}(D = A)$. We call this model FC.

Moreover, the L-B model is more general than RUM. This follows from the analysis in previous sections. In particular, fixing preferences, $\pi(\succ_i) = \mathbb{1}(\succ_i = \succ)$ for $\succ_i \in R(X)$, L-B with $L = LA$ reduces to original LA model by Brady and Rehbeck (2016) and, therefore, potentially inconsistent with RUM. Of course, when we add frame variation we can distinguish between EBA and RUM.

C.2 Comparison to RAM

Cattaneo et al. (2020) extends many theories of consideration by proposing the Random Attention Model (RAM). The authors allow for random consideration maps in the context of limited attention models. RAM abstracts away from the particular consideration-set-formation rule by considering a class of nonparametric random attention rules. The authors acknowledge that *RAM is best suited for eliciting information about preference ordering of a single decision-making unit when her choices are observed repeatedly*, which justifies the preference homogeneity assumption in their setting.

Many of the canonical models of limited attention proposed in the literature satisfy the monotonic attention property of Cattaneo et al. (2020). For instance, RAM nests LA, MM, and EBA without preference heterogeneity among other salient models of consideration sets. Additionally, RAM is a strict generalization of RUM. However, our L-B is not nested in RAM; see Cattaneo et al. (2020) for a complete description of its relationship to the literature.

Here, we show that in the presence of preference heterogeneity RAM may fail to rationalize behavior that can be explained by L-B. First, we formally define the restrictions imposed by RAM.

RAM imposes a *monotonic attention* restriction on consideration rules: the probability of paying attention to a particular subset does not decrease when the total number of possible consideration sets decreases. Formally,

DEFINITION 12 (Monotonic attention). For any $a \in A \setminus D$, $m_A(D) \leq m_{A \setminus \{a\}}(D)$.

Moreover, Cattaneo et al. (2020) provides a characterization of the model in terms of the revealed preference information inferred from data. Formally, we have the following.

DEFINITION 13 (Revealed preference, RAM). Let p be a RAM. Define P_R as the transitive closure of P defined as

$$aPb \text{ if there exists } A \in \mathcal{A} \text{ with } a, b \in A \text{ such that } p(a, A) > p(a, A \setminus \{b\}).$$

Then a is revealed preferred to b of and only if $a P_R b$.

Then a choice rule has a RAM representation if and only if P_R has no cycles. The following example of a L-B-rule, which results from a $m \in \mathcal{M}^{LA}$ for two linear orders \succ_1 and \succ_2 with $\pi(\succ_i) = 0.5$ with $i = 1, 2$, cannot be generated by RAM.

TABLE 3. Example 5 stochastic choice rule and random consideration set probability. p is consistent with LA-B but cannot be generated by RAM.

	$\{a, b, c\}$	$\{a, b\}$	$\{a, c\}$	$\{b, c\}$	$\{a\}$	$\{b\}$	$\{c\}$	\emptyset
a	0.305	0.339	0.157		0.208			
b	0.250	0.339		0.227		0.208		
c	0.255		0.300	0.341			0.345	
o	0.190	0.322	0.543	0.432	0.792	0.792	0.655	1
$\eta(D)$	0.20	0.30	0.01	0.10	0.05	0.05	0.10	0.19

EXAMPLE 5 (RAM violation). Let $X = \{a, b, c\}$ and consider a LA model for the random consideration set probability measure with $\eta(D)$ given as in Table 3. Moreover, consider two preference relations \succ_1 such that $a \succ_1 b \succ_1 c$, and \succ_2 such that $c \succ_2 b \succ_2 a$.⁴⁵ The resulting probabilistic choice rule is generated by a LA-B by construction. However, it cannot be rationalized by RAM since both aPb and bPa (i.e., $p(a, \{a, b, c\}) > p(a, \{a, c\})$ and $p(b, \{a, b, c\}) > p(b, \{b, c\})$). This means that a LA-B-rule allows cycles of the revealed preference relation P , which is ruled out by RAM.

C.3 Consideration cost and imperfect perception

One possible concern with our design is that DMs consider an alternative but misperceives the attributes (i.e., computes the wrong utility). We must point out that this concern applies broadly to any experimental design in which subjects have a nontrivial cognitive task. The following analysis assumes that the consideration cost is fixed. First, we need some preliminaries. For a given distribution of preferences $\pi \in \Delta(R(X))$, with perfect perception, there exists a random utility array $\mathbf{u} = (\mathbf{u}_a)_{a \in A}$ supported on $\mathbb{R}^{|A|}$ such that for a given menu of alternatives $A \in \mathcal{A}$:

$$\mathbb{P}(\mathbf{u}_a > \mathbf{u}_b, \forall b \in A \setminus \{a\}) = \pi(\succ: a \succ b, \forall b \in A \setminus \{a\}).$$

Now, misperception of any alternative $a \in A$ can be represented by another (possibly wrong) random utility variable \mathbf{w}_a supported on the reals. We let $\mathbf{w} = (\mathbf{w}_a)_{a \in A}$ be the array of such random variables. This random variable represents the subjective value that the DMs assigns to alternative a given her own perception of the item. Hence, \mathbf{w} and \mathbf{u} may be different (even when they may be correlated). Without loss of generality, we can define misperception as

$$\mathbf{e}_a = \mathbf{w}_a - \mathbf{u}_a,$$

⁴⁵In our experiment, this preference heterogeneity can be explained by heterogeneity in risk aversion. For example, let $a \equiv l_4$, $b \equiv l_3$, $c \equiv l_2$, and assume that DMs are expected-utility-maximizers with CRRA Bernoulli utilities. Then $a > b > c$ for individuals that are risk-neutrals ($\sigma = 0$), while $c > b > a$ for risk averse individuals ($\sigma > 0.5$). Holt and Laury (2002) finds that these types are common in their experiment across payment schemes.

for all $a \in A$ (and $\mathbf{e} = (\mathbf{e}_a)_{a \in A}$).⁴⁶

Under the assumption of independence of preference and attention. The only implication of misperception is that subjects' behavior will be governed not by π but rather by a different distribution of preferences π_e such that

$$\pi_e(\succ: a \succ b \forall b \in A \setminus \{a\}) = \mathbb{P}(\mathbf{u}_a + \mathbf{e}_a > \mathbf{u}_b + \mathbf{e}_b \forall b \in A \setminus \{a\}).⁴⁷$$

In other words, the population of DMs behavior captured by P will still be represented by a L-B model with (π_e, m) instead of the true (π, m) . This means that our design is robust to any arbitrary misperception error, in terms of the validity of our conclusions about how good are the different models to describe the population.

The only possible problem induced by misperception of alternatives is that we may lose the capacity to identify the true distribution of preferences. This possibility again is unavoidable in any experimental design that has a cognitive task. Nonetheless, this possibility is testable in our framework. In particular, if misperception exists, it must depend on the cognitive cost. Hence, we have the triple $(\mathbf{e}_H, \mathbf{e}_M, \mathbf{e}_L)$ that represents the misperception random array for the high, medium, and low cost, respectively. Then the distribution of preferences for any L-B model must not be stable across attention treatments with corresponding $(\pi_{e_H}, \pi_{e_M}, \pi_{e_L})$ distribution of preferences (that differ among costs).

However, we cannot reject the null hypothesis that LA has a stable distribution of preferences among the different cost distributions (i.e., $\pi_{e_H} = \pi_{e_M} = \pi_{e_L} = \pi$). In that sense, there is no evidence that misperception is important in our design.

C.4 Relation with rational inattention models

Rational inattention (RI) models have recently gained a lot of interest to model situations when choice is hard. However, RI models usually need very rich data sets to be identified/tested. That is, generally they cannot be identified with standard stochastic choice data sets. In that sense, we cannot do a full comparison between RI models and our approach, since the data set requirements are different. However, we can derive some implications of RI behavior for our data set.

RI is a model for individual behavior. To the best of our knowledge, the aggregate implications of this model are unknown. Hence, we will focus on comparing our approach to a case of a representative RI behavior. The problem of the representative RI DM is to choose the best possible alternative from a choice set. She has a prior μ over the true value of alternatives, $V = (v_k)_{k \in X \cup \{o\}}$, with $\mu \in \Delta(V)$. In response to the information structure, the RI DM chooses her optimal information to acquire an optimal action. We focus here on the subclass of RI problems with an additive cost of perception. The result of this problem is a true-value or state dependent stochastic choice rule

⁴⁶Note that in our experimental design, menus are randomly assigned to a DM. In addition, the presentation of each alternative remains the same across menu, conditional on the cost of consideration. Then it must be that the distribution of misperception (by-design) is the same across menus.

⁴⁷Where $\pi_e(\succ: \text{property})$ denotes the cumulative probability of preferences that have a certain property.

$p_v(\cdot, A) \in \Delta(A \cup \{o\})$, defined as

$$p_v(\cdot, A) = \arg \max_p \sum_{a \in A \cup \{o\}} p_v(a, A) v_a \mu(v_a) - \kappa(p_v(\cdot, A), \mu).$$

For the specification of κ , we focus on the generalized entropy proposed in Fosgerau, Melo, de Palma, and Shum (2017), which generalizes widely used entropic cost. Fosgerau et al. (2017) shows that this state-dependent stochastic choice is observationally equivalent to an additive random utility choice rule conditional on the support. That is, if $p_v(\cdot, A) \in \Delta(A \cup \{o\})$ (positive probability of choice), then $p_v(\cdot, A)$ is a random utility rule. Even when the underlying utility is fixed (and equal to v without loss of generality), there is randomness in choice due to a costly information acquisition. The state-dependent stochastic choice only differs from RUM when there are items in the choice set that are never chosen. Therefore, the RI DM is compatible with **deterministic** consideration sets. However, in our experiment we do not observe any element chosen with zero probability. In fact, the lowest probability of choice is 6% across all alternatives in $X \cup \{o\}$ and across all choice sets.

We have to aggregate across states to derive testable implications for the representative RI DM for our data set. This is because in our setup, the experimenter does not know ex ante the true value of alternatives. Preferences over lotteries (when there is not first-order stochastic dominance ordering among them) is unknown before choice. This is an important difference between our experiment and RI experimental literature, since they generally rely on enhanced data sets. We focus on collecting data sets that replicate standard stochastic choice data.

Using the fact that the sum of random utility rules is also a random utility rule, we notice that the marginal probability of choosing across different states is just the sum over the likelihood of this states (or the distribution of the true preferences). Then, if P admits a representative RI DM:

$$p(a, A) = \sum_{v \in V} p_v(a, A) \rho(v),$$

where $\rho \in \Delta(V)$ is the objective probability of the unobserved states.

LEMMA 3. *If $p_v(a, A) > 0$ for all $a \in A \cup \{o\}$, and all $v \in V$, it follows that if P admits a representative RI DM then P also admits a RUM representation.*

The proof of this lemma follows from Aguiar, Boccardi, and Dean (2016) and Fosgerau et al. (2017) that showed that the weighted sum of RUM is also RUM. The case in which one allows heterogeneity in discrete consideration sets, induced by RI, is difficult and left for future research.

APPENDIX D: ADDITIONAL DETAILS ON THE SAMPLE DESIGN

The primitive for the considered models is the estimated stochastic choice rule $\hat{p}_f(a, A)$ for $f \in \{H, M, L\}$. Therefore, for a fixed level of the cost f , the minimal required sample size was calculated to be proportional to the cardinality of the choice set. To maximize

TABLE 4. Average number of observations per alternative/choice set.

Choice set	N	$N/ A $	Choice set	N	$N/ A $
o12345	171	28.50	o12	131	43.67
o2345	155	31.00	o13	118	39.33
o1345	154	30.80	o14	125	41.67
o1245	149	29.80	o15	116	38.67
o1235	156	31.20	o23	112	37.33
o1234	143	28.60	o24	123	41.00
o345	131	32.75	o25	120	40.00
o245	118	29.50	o34	121	40.33
o235	125	31.25	o35	122	40.67
o234	116	29.00	o45	119	39.67
o145	112	28.00	o1	155	77.50
o135	123	30.75	o2	154	77.00
o134	120	30.00	o3	149	74.50
o125	121	30.25	o4	156	78.00
o124	122	30.50	o5	143	71.50
o123	119	29.75			

the number of observations for a given set of individuals, some individuals faced two decision tasks. To prevent possible learning, these subjects faced disjoint choice sets. That is, every subject faced either the full choice set $X \cup \{o\}$ or two choice sets that only had the outside option in common. Therefore, because of random assignment, in our experiment

- (i) 171 subjects faced only the whole choice set (the targeted number is 180);
- (ii) 757 subjects faced pairs of disjoint choice sets: the set of size 4 and the set of size 1 (the targeted number is 750);
- (iii) 1207 subjects faced pairs of disjoint choice sets: the set of size 3 and the set of size 2 (the targeted number is 1200).

This implies a total of 2135 subjects (the targeted number is 2130) for a total 4099 choices (the targeted number is 4080). Additionally, demographic data and preferences over binary comparison of lotteries were asked and incentivized. The effective number of observations per alternative/choice set/cost is summarized in Table 4.

APPENDIX E: PERFORMANCE OF THE TEST

In this section, we study the performance of our test in terms of statistical power. We are going to test the null hypothesis of LA-B when the true choice process presents choice overload. We consider behavior arising from a mixed population. A fraction $\lambda \in [0, 1]$ of the population is consistent with MM-B with $\gamma(x) = 1/2$ for all $x \in X$ and preferences consistent with expected utility maximization. The remaining fraction, λ , follows simple heuristics such the DM chooses outside option with probability proportional to the cardinality of the set. If she decides to pay attention to the menu, then she chooses uni-

TABLE 5. The table displays the proportion of rejections at the 10% and 5% significance levels for LA-B. Sample size = 4000. Number of MC replications = 1000. Number of bootstrap replications = 1000.

Process	Significance level	
	10%	5%
$\lambda = 0.25$	1	1
$\lambda = 0.50$	0.434	0.694

formly at random from it. The process is then consistent with the following stochastic choice rule:

$$p(a, A) = \lambda p^{\text{MM-B}}(a, A) + (1 - \lambda) p^{\text{CO}}(a, A), \tag{1}$$

where $p^{\text{MM-B}}(a, A)$ is consistent with MM-B with (m^{MM}, π) and $\pi(\succ) = 1/120$ for all $\succ \in R(X)$; and

$$p^{\text{CO}}(o, A) = \frac{|A| + 1}{|X| + 1} \quad \text{and} \quad p^{\text{CO}}(a, A) = \frac{1 - p^{\text{CO}}(o, A)}{|A|}.$$

The assumed process implies that a fraction $1 - \lambda$ of the population exhibits choice overload.⁴⁸

As the proportion of the population that exhibits choice overload increases so should increase the probability of rejecting the null that population behavior is generated by L-B. On the other extreme, when $\lambda = 1$ we should not reject the model. In particular, for any $\lambda < 32/39$ the process defined by equation (1) exhibits choice overload. However, for high values of λ the magnitude of this effect may not be significant to reject L-B.

Table 5 presents the results for power simulations for sample size 4000 and $\lambda \in \{0.25, 0.50\}$. For 1000 replications, the table displays the proportion of simulations that are rejected at the 10% and 5% significance levels. As expected, the fraction of rejections is bigger for smaller values of λ . For $\lambda = 0.25$, the rejection rate is 100%. We observe that at comparable sample size to our experiment the mixed process is rejected with power close to 1 when the choice overload fraction of DMs is moderate.

APPENDIX F: THE L-MODEL WITH HOMOGENEOUS PREFERENCES

In our setup, given an attention rule L (e.g., LA or MM) and a given frame, we can recover the underlying full consideration probabilistic choice rule p_π^L . Under the null hypothesis that our experimental data set can be generated by a L-rule with a strict preference relation over $X \cup \{o\}$ (such that o is the worst alternative), it must be that p_π^L

⁴⁸This process may intuitively arise when a DM that faced with a choice set only knows the size of the choice set and the alternatives in the grand set X . Knowing about the alternatives implies paying a cost c per alternative. Assume that preferences over information are modeled by a willingness to pay attention variable, w , that is distributed uniformly in $[0, 1]$. Then, given a choice set realization, after knowing $|A|$ DM i decides to pay attention to choice set A if $w_i > |A| \times c$. This implies that the DM pays attention and decide in the interior of the set with probability $\sum_{a \in A} p(a, A) = 1 - c|A|$; and $p(o, A) = c|A|$.

is degenerate (i.e., $p_{\pi}^L(a, A) \in \{0, 1\}$ for all A and $a \in A$). In particular, if we reject that $p_{\pi}^L(a_1, \{a_1, a_2\}) \in \{0, 1\}$ for some binary menu $\{a_1, a_2\}$ and for some attention cost, then we have to reject the L model with homogeneous preferences.

Given that the only model with heterogeneous preferences that was not rejected in our experiment is LA, to show the importance of preference heterogeneity, we tested whether the LA model with homogeneous preferences can explain the data. We took menu $\{1, 3\}$ and the high cost frame and computed the implied by the LA model the full consideration probability $\hat{p}_{\pi}^{LA}(1, \{1, 3\})$. If the data can be explained by the LA model with homogeneous preferences, then $\hat{p}_{\pi}^{LA}(1, \{1, 3\})$ should converge in probability to either 0 ($3 \succ 1$ with probability 1) or 1 ($1 \succ 3$ with probability 1). We tested two hypotheses: (i) $p_{\pi}(1, \{1, 3\}) = 0$ and (ii) $p_{\pi}(1, \{1, 3\}) = 1$. Both were rejected at the 5% significance level (p-value $< 10^{-3}$). As a result, we can conservatively claim that the null hypothesis that our experimental data set can be explained by the LA-rule with a single preference relation is rejected at least at the 5% significance level.

REFERENCES

- Abaluck, Jason and Abi Adams (2021), “What do consumers consider before they choose? Identification from asymmetric demand responses.” *The Quarterly Journal of Economics*, 136 (3), 1611–1663. [74, 75, 76]
- Aguiar, Victor H. (2017), “Random categorization and bounded rationality.” *Economics Letters*, 159, 46–52. [72, 74, 77, 82, 83, 85, 88, 96, 104, 105]
- Aguiar, Victor H., Maria Jose Boccardi, and Mark Dean (2016), “Satisficing and stochastic choice.” *Journal of Economic Theory*, 166, 445–482. [72, 110]
- Aguiar, Victor H., Maria Jose Boccardi, Nail Kashaev, and Jeongbin Kim (2023), “Supplement to ‘Random utility and limited consideration’.” *Quantitative Economics Supplemental Material*, 14, <https://doi.org/10.3982/QE1861>. [75]
- Apesteguia, Jose and Miguel Ángel Ballester (2021), “Separating predicted randomness from residual behavior.” *Journal of the European Economic Association*, 19 (2), 1041–1076. [72]
- Barseghyan, Levon, Maura Coughlin, Francesca Molinari, and Joshua C. Teitelbaum (2021), “Heterogeneous choice sets and preferences.” *Econometrica*, 89 (5), 2015–2048. [72, 74]
- Barseghyan, Levon, Francesca Molinari, and Matthew Thirkettle (2021), “Discrete choice under risk with limited consideration.” *American Economic Review*, 111 (6), 1972–2006. [74]
- Bhattacharya, Mihir, Saptarshi Mukherjee, and Ruhi Sonal (2021), “Frame-based stochastic choice rule.” *Journal of Mathematical Economics*, 97, 102553. [86]
- Block, Henry David and Jacob Marschak (1960), “Random orderings and stochastic theories of responses.” *Contributions to probability and statistics*, 2, 97–132. [72, 83, 96]

Brady, Richard L. and John Rehbeck (2016), “Menu-dependent stochastic feasibility.” *Econometrica*, 84 (3), 1203–1223. [71, 72, 73, 74, 77, 83, 85, 88, 104, 105, 107]

Caplin, Andrew, Mark Dean, and John Leahy (2019), “Rational inattention, optimal consideration sets, and stochastic choice.” *The Review of Economic Studies*, 86 (3), 1061–1094. [72, 101]

Cattaneo, Matias D., Xinwei Ma, Yusufcan Masatlioglu, and Elchin Suleymanov (2020), “A random attention model.” *Journal of Political Economy*, 128 (7), 2796–2836. [72, 73, 74, 76, 82, 96, 107]

Chateauneuf, Alain and Jean-Yves Jaffray (1989), “Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion.” *Mathematical social sciences*, 17 (3), 263–283. [82, 83, 103, 105]

Chernev, Alexander, Ulf Böckenholt, and Joseph Goodman (2015), “Choice overload: A conceptual review and meta-analysis.” *Journal of Consumer Psychology*, 25 (2), 333–358. [96]

Dardanoni, Valentino, Paola Manzini, Marco Mariotti, and Christopher J. Tyson (2020), “Inferring cognitive heterogeneity from aggregate choices.” *Econometrica*, 88 (3), 1269–1296. [74]

De Clippel, Geoffroy and Kareen Rozen (2021), “Bounded rationality and limited datasets.” *Theoretical Economics*, 16 (2), 359–380. [73]

Babur De Los Santos, Ali Hortaçsu, and Matthijs R. Wildenbeest (2012), “Testing models of consumer search using data on web browsing and purchasing behavior.” *American Economic Review*, 102 (6), 2955–2980. [72]

Dean, Mark and John McNeill (2014), *Preference for Flexibility and Random Choice: An Experimental Analysis*. Technical report, Working Paper. Brown University, Department of Economics. <http://hdl.handle.net/10419/145422>. [90]

Echenique, Federico, Kota Saito, and Gerelt Tserenjigmid (2018), “The perception-adjusted Luce model.” *Mathematical Social Sciences*, 93, 67–76. [96]

Falmagne, Jean Claude (1978), “A representation theorem for finite random scale systems.” *Journal of Mathematical Psychology*, 18 (1), 52–72. [72]

Feldman, Paul and John Rehbeck (2022), “Revealing a preference for mixtures: an experimental study of risk.” *Quantitative Economics*, 13 (2), 761–786. [100]

Fosgerau, Mogens, Emerson Melo, André de Palma, and Matthew Shum (2017), *Discrete Choice and Rational Inattention: A General Equivalence Result*. [110]

Freeman, David, Yoram Halevy, Terri Kneeland et al. (2019), “Eliciting risk preferences using choice lists.” *Quantitative Economics*, 13 (2), 217–237. [100]

Fudenberg, Drew, Ryota Iijima, and Tomasz Strzalecki (2015), “Stochastic choice and revealed perturbed utility.” *Econometrica*, 83 (6), 2371–2409. [80, 96]

Goeree, Michelle Sovinsky (2008), “Limited information and advertising in the us personal computer industry.” *Econometrica*, 76 (5), 1017–1074. [74]

Heiss, Florian, Daniel McFadden, Joachim Winter, Amelie Wuppermann, and Bo Zhou (2021), “Inattention and Switching Costs as Sources of Inertia in Medicare Part d.” *American Economic Review*, 111 (9), 2737–2781. [72]

Ho, Kate, Joseph Hogan, and Fiona Scott Morton (2017), “The impact of consumer inattention on insurer pricing in the medicare part d program.” *The RAND Journal of Economics*, 48 (4), 877–905. [72]

Holt, Charles A. and Susan K. Laury (2002), “Risk aversion and incentive effects.” *American Economic Review*, 92 (5), 1644–1655. [91, 108]

Honka, Elisabeth (2014), “Quantifying search and switching costs in the us auto insurance industry.” *The RAND Journal of Economics*, 45 (4), 847–884. [72]

Honka, Elisabeth, Ali Hortaçsu, and Maria Ana Vitorino (2017), “Advertising, consumer awareness, and choice: Evidence from the us banking industry.” *The RAND Journal of Economics*, 48 (3), 611–646. [72]

Horan, Sean (2019), “Random consideration and choice: A case study of default options.” *Mathematical Social Sciences*, 102, 73–84. [72]

Hortaçsu, Ali, Seyed Ali Madanizadeh, and Steven L. Puller (2017), “Power to choose? An analysis of consumer inertia in the residential electricity market.” *American Economic Journal: Economic Policy*, 9 (4), 192–226. [72]

Horton, John J., David G. Rand, and Richard J. Zeckhauser (2011), “The online laboratory: Conducting experiments in a real labor market.” *Experimental economics*, 14 (3), 399–425. [90]

Huber, Joel, John W. Payne, and Christopher Puto (1982), “Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis.” *Journal of consumer research*, 9 (1), 90–98. [97]

Iyengar, Sheena S. and Mark R. Lepper (2000), “When choice is demotivating: Can one desire too much of a good thing?” *Journal of personality and social psychology*, 79 (6), 995. [96]

Kim, Jeongbin (2020), *Discounting, Dynamic Consistency, and Cooperation in an Infinitely Repeated Game Experiment*. Technical report, Working paper. [90]

Kitamura, Yuichi and Jörg Stoye (2018), “Nonparametric analysis of random utility models.” *Econometrica*, 86 (6), 1883–1909. [72, 73, 88, 89, 90, 98, 100, 101]

Kovach, Matthew and Gerelt Tserenjigmid (2022), “The focal Luce model.” *American Economic Journal: Microeconomics*, 14 (3), 378–413. [96]

Kovach, Matthew and Levent Ülkü (2020), “Satisficing with a variable threshold.” *Journal of Mathematical Economics*, 87, 67–76. [72]

- Krantz, David H. (1965), *The Scaling of Small and Large Color Differences*. PhD thesis. [79]
- Lleras, Juan Sebastian, Yusufcan Masatlioglu, Daisuke Nakajima, and Erkut Y. Ozbay (2017), “When more is less: Limited consideration.” *Journal of Economic Theory*, 170, 70–85. [72]
- Manzini, Paola and Marco Mariotti (2014), “Stochastic choice and consideration sets.” *Econometrica*, 82 (3), 1153–1176. [72, 73, 74, 75, 77, 85, 88]
- Masatlioglu, Yusufcan, Daisuke Nakajima, and Erkut Y. Ozbay (2012), “Revealed attention.” *American Economic Review*, 102 (5), 2183–2205. [72]
- McCausland, William J., Anthony Arthur John Marley, Clinton Davis-Stober, Nicholas Brown, and Sanghyuk Park (2020), “Testing the random utility hypothesis directly.” *The Economic Journal*, 130 (625), 183–207. [72]
- McFadden, Daniel and Marcel K. Richter (1990), “Stochastic rationality and revealed stochastic preference.” In *Preferences, Uncertainty, and Optimality, Essays in Honor of Leo Hurwicz*, 161–186, Westview Press, Boulder, CO. [71, 72, 87, 88, 89]
- Kashaev, Nail and Victor H. Aguiar (2022), “A random attention and utility model.” *Journal of Economic Theory*, 105487. [76, 79]
- Kashaev, Nail and Natalia Lazzati (2021), “Peer effects in random consideration sets.” Preprint. arXiv:1904.06742. [74]
- Natenzon, Paulo (2019), “Random choice and learning.” *Journal of Political Economy*, 127 (1), 419–457. [96]
- Rand, David G., Joshua D. Greene, and Martin A. Nowak (2012), “Spontaneous giving and calculated greed.” *Nature*, 489 (7416), 427. [90]
- Rieskamp, Jörg, Jerome R. Busemeyer, and Barbara A. Mellers (2006), “Extending the bounds of rationality: Evidence and theories of preferential choice.” *Journal of Economic Literature*, 44 (3), 631–661. [75]
- Salant, Yuval and Ariel Rubinstein (2008), “(a, f): Choice with frames.” *The Review of Economic Studies*, 75 (4), 1287–1296. [73, 86, 87]
- Shafer, Glenn (1976), *A Mathematical Theory of Evidence*, Vol. 42. Princeton University Press. [103]
- Kovach, Matthew and Elchin Suleymanov (2011), “Reference Dependence and Random Attention.” Preprint. arXiv:2106.13350. [74, 88, 91]
- Tversky, Amos (1972), “Elimination by aspects: A theory of choice.” *Psychological review*, 79 (4), 281. [74, 77, 79]

Co-editor Christopher Taber handled this manuscript.

Manuscript received 28 February, 2021; final version accepted 2 July, 2022; available online 28 July, 2022.