

Random Utility and Limited Consideration^{*†}

Victor H. Aguiar[‡] Maria Jose Boccardi[§] Nail Kashaev[¶] Jeongbin Kim^{||}

This version: February 2021/ First version: November 2018

Abstract 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 dataset that we collected in a large-scale online experiment. Our dataset 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](#)).

JEL classification numbers: C90, C12, D81, D12.

Keywords: random utility, experimental discrete choice, random consideration sets, frames.

^{*}This paper was previously circulated as "Does Random Consideration Explain Behavior when Choice is Hard? Evidence from a Large-scale Experiment." 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 and Social Sciences and Humanities Research Council.

[†]This study is approved by the Caltech's IRB No. 18-0812.

[‡]Department of Economics, University of Western Ontario, vaguiar@uwo.ca

[§]Amazon, majoboccardi@gmail.com. The experiment was run prior Boccardi joined Amazon

[¶]Department of Economics, University of Western Ontario, nkashaev@uwo.ca

^{||}Department of Marketing, National University of Singapore, bizkj@nus.edu.sg

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 et al. (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 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 datasets 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

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 et al. (2012), Honka (2014), Heiss et al. (2016), Ho et al. (2017), Honka et al. (2017), Hortaçsu et al. (2017) and Barseghyan et al. (2019a).

³See, for instance, Aguiar et al. (2016), Brady and Rehbeck (2016), Aguiar (2017), Lleras et al. (2017), Caplin et al. (2019), Horan (2019), Kovach and Ülkü (2020), and Cattaneo et al. (2020).

⁴For evidence for RUM see Kitamura and Stoye (2018) and McCausland et al. (2018).

⁵Other experiments that we are aware of that have collected stochastic choice data focusing on choice set variation are Apesteguia et al. (2018) (87 individuals) and McCausland et al. (2018) (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).

online in Amazon Mechanical Turk (MTurk), collecting 12297 independent choice observations from 2135 individuals. A *frame* consists of 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 dataset, 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

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

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 dataset, 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 [De Clippel and Rozen \(2018\)](#) and [Cattaneo et al. \(2020\)](#). Nonetheless, full exogenous variation in choice sets is an important data feature that 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 pre-selected 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

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 \(2008\)](#), [Barseghyan et al. \(2019a,b\)](#), [Dardanoni et al. \(2020\)](#), and [Abaluck and Adams \(2021\)](#).

investigate from an experimental perspective whether this strong assumption is effective in explaining choice when it 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 (Suleymanov, 2018). 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 95 percent confidence 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 to this second issue.¹⁰ We hope that our findings of 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 which extends theories of consideration sets while allowing for preference heterogeneity. This section characterizes the restrictions under each theory considered and describes our testing procedure. Section 3 details our experimental design that exogenously varies choice sets and frames/consideration-cost, thereby allowing us to identify limited consideration in the sample. We describe the observed pattern of choices and discuss whether there is evidence of departures from rational behavior. Section 4 presents the testing results for the heterogeneous random-consideration theories in our sample and discusses their implications. Finally, Section 5 concludes. All proofs and additional results can be found in

¹⁰There is a vast literature documenting departures from fully rational behavior but it is not focused on limited consideration. See Rieskamp et al. (2006) for a survey.

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

the appendix.

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

2.1. Random Behavioral Model

We consider an environment where decision makers (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 $R(X)$ the set of all strict linear orders on X . The probability measure π , which is an element of the simplex on $R(X)$ denoted by $\Delta(R(X))$, fully captures preference heterogeneity. The distribution over random consideration sets is fully characterized by the probability measure $m_A : 2^A \rightarrow [0, 1]$, $\sum_{D \subseteq A} m_A(D) = 1$. In other words, m_A is an element of the simplex on 2^A , which we denote by $\Delta(2^A)$ for every $A \in \mathcal{A}$. Let m denote the complete collection of those probability measures. That is, $m = (m_A(D))_{A \in \mathcal{A}, D \in 2^A}$. Formally,

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

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

for all $a \in X$ and $A \in \mathcal{A}$, where $\mathbf{1}(\cdot)$ denotes the indicator function.¹²

¹² $\mathbf{1}(B)$ is equal to 1 if the statement B is true, otherwise it is equal to zero.

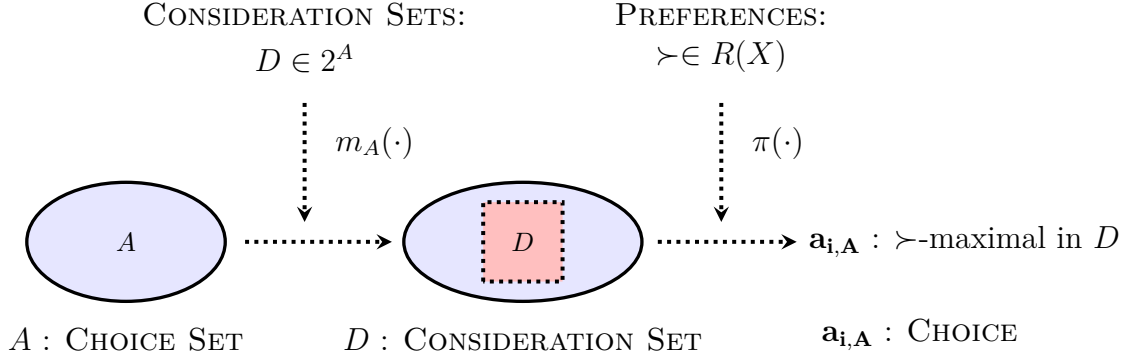


Figure 1 – CONSIDERATION MEDIATED CHOICES Choices are the result of a two stage process, first pick a consideration set and then pick the best alternative in that set. We observe choices (\mathbf{a}_i) and menus (A). We do not observe and we aim to identify the theory objects: distribution of preferences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

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). In addition, in our setup with exogenous frame variation, independence allows isolating the frame effects (attention effect). 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. Evidently, studying the correlation between preferences and consideration is an important problem but this is beyond the scope of this paper.

Our independence assumption holds trivially for the cases of homogeneous preferences such as all models covered by the Random Attention Model (RAM) of Cattaneo et al. (2020). Moreover, as the following lemma demonstrates, the B-rule does not have empirical content even under the independence assumption. This means that assuming independence between preferences and consideration rules, at this stage, can be done without loss of generality.

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

Without additional restriction on π and m the model is not falsifiable. 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 consider a family of consideration set rules that are governed by an *attention-index*. An attention-index is a probability measure over the power set, $\eta \in \Delta(2^X)$. The value $\eta(D)$ captures the unconditional attention that DMs pay to the set of alternatives $D \in 2^X$. The attention-index of a set is a net measure of its attractiveness with respect to how hard is to consider it. The attention-index measures how enticing a consideration set is, and how complex it is to understand. Therefore, if $\eta(C) > \eta(D)$, we say that C , in net terms, attracts more attention than D . We introduce here the family of attention-index consideration set rules.

Definition 2 (Attention-index representation). We say that a consideration set rule m admits an attention-index representation if there exists a link function and an index correspondence $(\psi(\cdot), g(\cdot, \cdot))$ 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 $D \subseteq A$ and some attention index η .

The attention-index representation is defined in an analogous way to the models of *simple scalability* in stochastic choice by [Krantz \(1965\)](#) and [Tversky \(1972\)](#). However, instead of depending on an arbitrary way in the entries of η , as in the simple scalability models, here we require that it depends on two forces: (i) the attention-index of D , and (ii) the cumulative attention to sets different from D captured by $g(D, A)$.¹³

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 $\sum_{C \in g(D, A)} \eta(C)$.

¹³However, scalar link function here is defined in terms of the attention-index instead of an average utility like in the traditional Luce model. In fact, our single-attention index is compatible with violations of regularity and other implications of the traditional stochastic choice models of simple scalability as discussed in [Tversky \(1972\)](#).

We focus on the random consideration rules that admit an attention-index representation because: (i) They cover, as special cases, well-known models of consideration (see below). (ii) They allow for unique identification of the consideration rule. (iii) They are intuitive, simple, and of behavioral interest, as they provide a useful (unconditional) index of attention for any subset of alternatives. (iv) They allow for a significant reduction of the dimensionality of the consideration set rules making them tractable.

The following models admit an attention-index representation: the full attention model (FC), the logit attention model (LA) of [Brady and Rehbeck \(2016\)](#), the choice set independent attention model (MM) of [Manzini and Mariotti \(2014\)](#), and the elimination-by-aspects model (EBA) of [Tversky \(1972\)](#) in its special version with a default by [Aguiar \(2017\)](#).¹⁴ Two distinct consideration set rules that admit an attention-index representation only differ in how they use the attention-index to form the consideration set at a given choice set.

Definition 3. The logit attention (LA), the choice set independent (MM), the random consideration (EBA), and the full consideration (FC) models admit an attention-index representation such that:

- $m \in \mathcal{M}^{\text{LA}}$ 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

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

¹⁴Indeed, the the model in [Aguiar \(2017\)](#) coincides with the elimination-by-aspects, when the former has an attention-index where there is positive mass only on menus that contain the default. In that case, the numerator of the elimination-by-aspects model, in the formula in [Cattaneo et al. \(2020\)](#), becomes 1, and both models are indistinguishable.

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

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

This list of attention-index consideration rules is nonexhaustive; however, we work with these models because they will allow us to learn about the true data generating process governing our experimental application in a systematic way. In particular, the well-known MM-rule is exactly the intersection of the LA and EBA models (Suleymanov, 2018). Thus, LA and EBA are completely distinct generalizations of MM. In that sense, rejecting or accepting either of these models is very informative.

The MM rule imposes independence in consideration across items. This restriction makes the model highly stylized and tractable. Both the LA and EBA rules generalize this independence assumption to allow for substitution and complementarity of attention between different items. The LA rule predicts that the probability of considering D when the choice set is A is proportional to the attention-index value $\eta(D)$. The EBA rule predicts that the probability of considering a subset D given choice set A is the probability that D is exactly the intersection of the category or subset of alternatives (considered randomly using the attention-index) and the choice set.¹⁵

In our framework, the randomness due to limited consideration can arise for two possible reasons: (i) consideration is deterministic at the individual level but heterogeneous at the population level; and (ii) consideration is random at the individual level and independent and identically distributed (i.i.d.) at the population level.

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

Definition 4 (L-B-rule). For a given link function ψ^L a complete stochastic choice rule P is L-B-rule if P is a B-rule with $m \in \mathcal{M}^L$.

Definition 4 summarizes all models that have the attention-index representation. In our

¹⁵Note that for the LA-model $\psi^{\text{LA}}(x, y) = \frac{x}{x+y}$ 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 $g^{\text{EBA}}(D, A) = \{C \in 2^X \setminus D : C \cap A = D\}$.

application we focus in four different models ($L \in \{LA, MM, EBA, FC\}$). However, our approach extends to any model of consideration that admits an attention-index representation.

The next example shows how a population of individuals that are heterogeneous in terms of their (deterministic) limited consideration can be described by a L-B-rule.

Example 1. 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 that 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 following example demonstrates a possible mechanism that allows an attention index representation when a representative DM exists.

Example 2. Consider a (representative) DM that, faced with a menu A , needs to allocate her attention, measured by $m_A \in \Delta(2^A)$, over all possible consideration sets in A (including the empty set). We assume that the DM is endowed with an attention-index $\eta \in \Delta(2^X)$. Choice difficulty is mediated by a cognitive cost function $K : [0, 1] \rightarrow \mathbb{R} \cup \{+\infty\}$ that is independent from menus. For a fixed $D \subseteq A$, $K(m_A(D))$ measures the cost of allocation attention to D using the consideration set rule m_A . Following [Fudenberg et al. \(2015\)](#) we assume that $K(\cdot)$ is strictly convex. The DM's problem is to find $m_A \in \Delta(2^A)$ that maximize expected attractiveness of the menu given the cognitive cost of processing it. Formally, m_A is a solution to

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

In this case, the strict convexity of K together with the first-order conditions of the above optimization problem would imply that the optimal consideration rule admits an attention-index representation. Thus, there exists a link function that connects m and η , making this behavior a special case of the attention-index family. In particular the LA rule is compatible with this interpretation (see [Appendix C.4](#) for details).

2.3. Characterization and Identification of the L-B-model

In this section, we provide an answer to 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 dataset P the primitives of the B rule that generated it, and provide necessary and sufficient conditions that guarantee that a dataset 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 the outside option is only chosen when nothing else in the choice set is considered. If we denote $p_o = (p(o, A))_{A \in \mathcal{A}}$ and $\psi_\emptyset(\eta) = \psi(\eta(\emptyset), \sum_{C \in g(\emptyset, A)} \eta(C))$, 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 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 $A \in \mathcal{A}$:

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

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).

Henceforth, we impose the total monotonicity restriction on the known link function, and we refer to the attention-index representation and totally monotone attention-index representation interchangeably.

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 $\Delta_D(\Delta_C p(o, A) \leq 0)$).¹⁶ 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).

Observe that strict monotonicity of φ in each of its entries implies the invertibility of ψ .¹⁷ We underline that this sufficient condition for invertibility of the link function ψ is mild. In fact, 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.

Lemma 2. *Any consideration rule with link function $L \in \{\text{LA}, \text{MM}, \text{EBA}, \text{FC}\}$ admits a totally monotone attention-index representation with:*

- $\varphi^{\text{LA}}(\eta_o, t) = \frac{\eta_o}{t}$;

¹⁶Note, however, that this restriction does not mean that the consideration set rule m satisfies the monotonicity property for other menus different from the empty set.

¹⁷This is a consequence of Mobius invertibility of the mapping $v(\cdot) = \sum_{C \subseteq \cdot} \eta(C)$ (Chateauneuf and Jaffray, 1989).

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

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 \(2017\)](#).¹⁸

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 .¹⁹ 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 $A \in 2^X \setminus X$

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

The calibrated attention-index depends only on the dataset P . We will show that P has been generated by an attention-index model if and only if the calibrated attention-index generates a well-defined consideration rule. In other words, the testable implication of a particular link function L on the probability of choosing the outside option is that the calibrated attention-index does not

¹⁸Note 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 .

¹⁹The calibrated index η^L will not be a well-defined distribution when the model is misspecified.

put negative attention on any set $A \in 2^X$. 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 under the correct specification of the model.

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

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

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

Definition 8. For given $L \in \{\text{LA}, \text{MM}, \text{EBA}, \text{FC}\}$ and P , let $m^L = (m_A^L(D))_{A \in \mathcal{A}, D \in 2^A}$, where $m_A^L : 2^A \rightarrow \mathbb{R}$ is such that, for all $A \in \mathcal{A}$ and $D \in 2^A$,

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

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. However, if m^L exists, then $\sum_{D \subseteq A} m_A^L(D) = 1$ for all $A \in \mathcal{A}$. Thus, m^L is a collection of distributions over consideration sets if and only if $m^L \geq 0$. To be able to estimate m^L from the data with probability approaching 1, we need the following definition.

Definition 9 (Well-defined m^L). m^L is a well-defined collection of distributions over consideration sets if $m_A^L \in \Delta(2^A)$ for all $A \in \mathcal{A}$.

For the examples of interest, m^L is a well-defined collection of distributions over consideration sets if

- $m_A^{\text{LA}}(D) > 0$ for all $D \in 2^A$;

- $\gamma^{\text{MM}}(a) \in (0, 1)$ for all $a \in X$;
- $m_A^{\text{EBA}}(D) \geq 0$ for all $A \in 2^A$;
- $m_A^{\text{FC}}(D) = \mathbb{1}(A = D)$.

The testable implications of the different attention-index link functions that we considered are summarized by the requirement that m^L is well-defined.

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 collection of distributions over consideration sets such that P is a B-rule described by (m, π) with $m = m^L$.

Theorem 1 provides a testable implication of the model. 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 considerations 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 achieved dimensionality reduction. 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 10. For given link function 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 $a \in A$,

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

We call P_π^L the calibrated full consideration rule.

Notice that when P has been generated by a L -B, 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 $p_\pi^L = p_\pi$.

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 link function 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. We do not observe P but can consistently estimate it by the collection of sample frequencies \hat{P} . In Section 2.7 we discuss how to test the L-B-rule accounting for sampling variability in \hat{P} .

Remarkably, Theorems 1 and 2 provide a generalization of the characterization results in Manzini and Mariotti (2014), Brady and Rehbeck (2016), and Aguiar (2017). Moreover, it provides a unified result for all models that admit a (totally monotone) attention-index representation.

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 or link functions that are totally monotone. 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.

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

We underline that the identification properties of the totally monotone attention-index consideration models are very sharp. Indeed, we achieve a decomposition of the dataset P into its attention

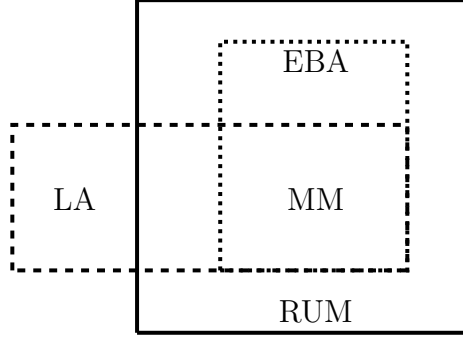


Figure 2 – RELATION AMONG CONSIDERATION SET RULES: RUM, LA, MM, AND EBA

and preference components in a unique way. Identification of preferences and consideration rules is not a trivial task. Even for simple datasets where stochastic behavior arises from only one channel (for example 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 identification of preferences for regular random choice is not possible.²⁰

2.6. Relation Among Different Attention-Index Models and Random Utility

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. This is just the standard model by [McFadden and Richter \(1990\)](#). Let $R(X \cup \{o\})$ be a set of linear orders over the extended choice set $X \cup \{o\}$. Also, let $\pi_o \in \Delta(R(X \cup \{o\}))$ be a probability measure over $R(X \cup \{o\})$.

Definition 11. (Random Utility Model, RUM) A complete stochastic choice rule P is consistent with random utility if there exists a $\pi_o \in \Delta(R(X \cup \{o\}))$ such that

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

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

²⁰Regular random choice means that for two menus $A \subseteq B$, for any $a \in A$ it follows that $p(a, A) \leq p(a, B)$.

The relationship between the models in Definition 3 and RUM are summarized in Figure 2. The relationship is valid for a fixed frame. 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, as discussed by Manzini and Mariotti (2014) and Brady and Rehbeck (2016), respectively. Moreover, EBA is nested in RUM and nests MM, therefore its intersection with LA is nonempty (Aguar, 2017). Crucially, for a fixed frame if a dataset is consistent with EBA it is consistent for RUM. However, when with frame variation this is no longer true and RUM and EBA intersect but are not nested within each other. Suleymanov (2018) shows that MM exactly describes the intersection between LA and EBA.²¹

2.7. Testing Procedure

Theorem 2 allows us to test whether a given stochastic choice rule P is a L-B-rule: it is necessary and sufficient to test whether m^L is well-defined (satisfies a set of linear inequalities) and P_π^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} instead of unknown P .

To introduce the testing procedure we need to define several objects. Note that the calibrated full consideration rule, P_π^L , is a vector of length $d_p = \sum_{k=1}^{|X|} k \binom{|X|}{k}$.²² The k -th element of P_π^L corresponds to some pair (a, A) such that $a \in A \cup \{o\}$.

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

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

²¹All 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.

²² $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ and $n! = 1 \cdot 2 \cdot \dots \cdot n$.

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 as the matrix of size $(d_p + d_m) \times d$, where $d = |X|! + d_m$ and $d_m = \sum_{A \subseteq X} 2^{|A|}$ is the dimension of m^L , such that

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

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^L and P_π^L . Let \mathbb{R}_+^d denote component wise nonnegative elements of the d -dimensional Euclidean space \mathbb{R}^d .

Theorem 4. *The following are equivalent.*

- (i) P_π^L is a FC-B-rule and m^L is well-defined;
- (ii) $\inf_{v \in \mathbb{R}_+^d} \|g^L - Gv\| = 0$, where $g^L = (P_\pi^{L'}, m^{L'})'$.

Proof. See McFadden and Richter (1990) and Kitamura and Stoye (2018). ■

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

Although P is not observed, the realized choice frequencies \hat{P} are. For every $A \in \mathcal{A}$ let n_A denote the number of individuals in the sample that faced choice set A , and let $\mathbf{a}_{i,A}$, $i = 1, \dots, n_A$ be the observed choice of individual i from choice set $A \cup \{o\}$. We assume that the researcher observes a cross-section of observations for every choice set. Then we define the estimated stochastic choice rule as

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

with $\hat{p}(a, A) = n_A^{-1} \sum_{i=1}^{n_A} \mathbf{1}(\mathbf{a}_i = a)$ for any $a \in A \cup \{o\}$.

Given the model of interest L and the estimator of P , \hat{P} , we can compute the estimators of m^L and P_π^L , \hat{m}^L and \hat{P}_π^L , using Definitions 7 and 10.²³ Given the results of Theorem 4, a natural test statistic is

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

²³To compute \hat{P}_π^L , in our empirical application, we minimized the Euclidean distance between \hat{P} and $(\sum_{D \subseteq A} \hat{m}^L p_\pi(a, D))_{a \in A, A \in \mathcal{A}}$ subject to $p_\pi(a, A) \geq 0$, $\sum_{a \in D} p_\pi(a, D) = 1$ for all a and A , and $p_\pi(a, D) = 0$ for all D and $a \notin D$.

where $n = \sum_A n_A$ is the sample size and $\hat{g}^L = (\hat{P}_\pi^{L'}, \hat{m}^{L'})'$.

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 .²⁴ 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

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

(ii) Compute the bootstrap test statistic

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

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

For a given confidence 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.

3. 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

²⁴In our empirical application we conducted tests for different values of τ_n (e.g., $\tau_n = \sqrt{\frac{\log(\min_A n_A)}{\min_A n_A}}$ following Kitamura and Stoye (2018), and $\tau_n = 0$). The results are qualitatively the same.

individual (see Section 3.1). The large sample size of the dataset 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.²⁵ 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 \$2.²⁶ The average duration of the session was 251.68 seconds (slightly over 4 minutes).²⁷ 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 et al. (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 minimum wage in MTurk as \$0.14 per hour. 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 (2016) 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 et al. (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.

3.1. 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. We vary the frame through changes in

²⁵By 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.

²⁶All payments were made in USD.

²⁷The 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.

the cost of consideration.

To induce (potentially) 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 CRRA Bernoulli utility function

$$u(x) = \begin{cases} \frac{x^{1-\sigma}}{1-\sigma}, & \sigma \neq 1, \\ \ln(x), & \sigma = 1. \end{cases}$$

Additional details about implied rankings are presented in Appendix D. The outside option is dominated for moderate levels of risk aversion (e.g. Holt and Laury, 2002).²⁸

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
(1) $\frac{1}{2}50 + \frac{1}{2}0$	25.000	625.00	1	1	2	5	5	6	6
(2) $\frac{1}{2}30 + \frac{1}{2}10$	20.000	100.00	5	5	5	2	1	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
(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
(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	3
(o) 12 with probability 1	12.000	0.00	6	6	6	6	6	2	2

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 choice sets $A \in \mathcal{A}$ are observed in the sample. The outside option is always present and is shown first, while the order of other alternatives is randomized. Choice sets can be thought as different treatments. In fact, many of the behavioral implications of stochastic models of choice are testable in terms of the change in behavior when adding (removing) an alternative from the menu.

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, i.e. $n_A = \lambda(|A| + 1)$ 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,

²⁸In 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 Suleymanov (2018).

for the high, the medium, and the low cost, respectively. 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), i.e. $A_j \cup \{o\}, A_k \cup \{o\}$ with $A_j \cup A_k = X$ and $A_j \cap A_k = \emptyset$. Our experimental design is summarized in Figure 4.²⁹

The default alternative For any choice set/frame the outside option is always present and shown first. Moreover, it is pre-selected as the default alternative. If the subject decides to skip the task, she is informed that o will be chosen for her. This 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.

3.2. 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 12297 observations. The number of observations per alternative/choice set are shown in Table 4. 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).

²⁹ This is done to increase the number of observations given the fixed cost of the participation fee.

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+10-39) 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 $\{o, l_2, l_4\}$

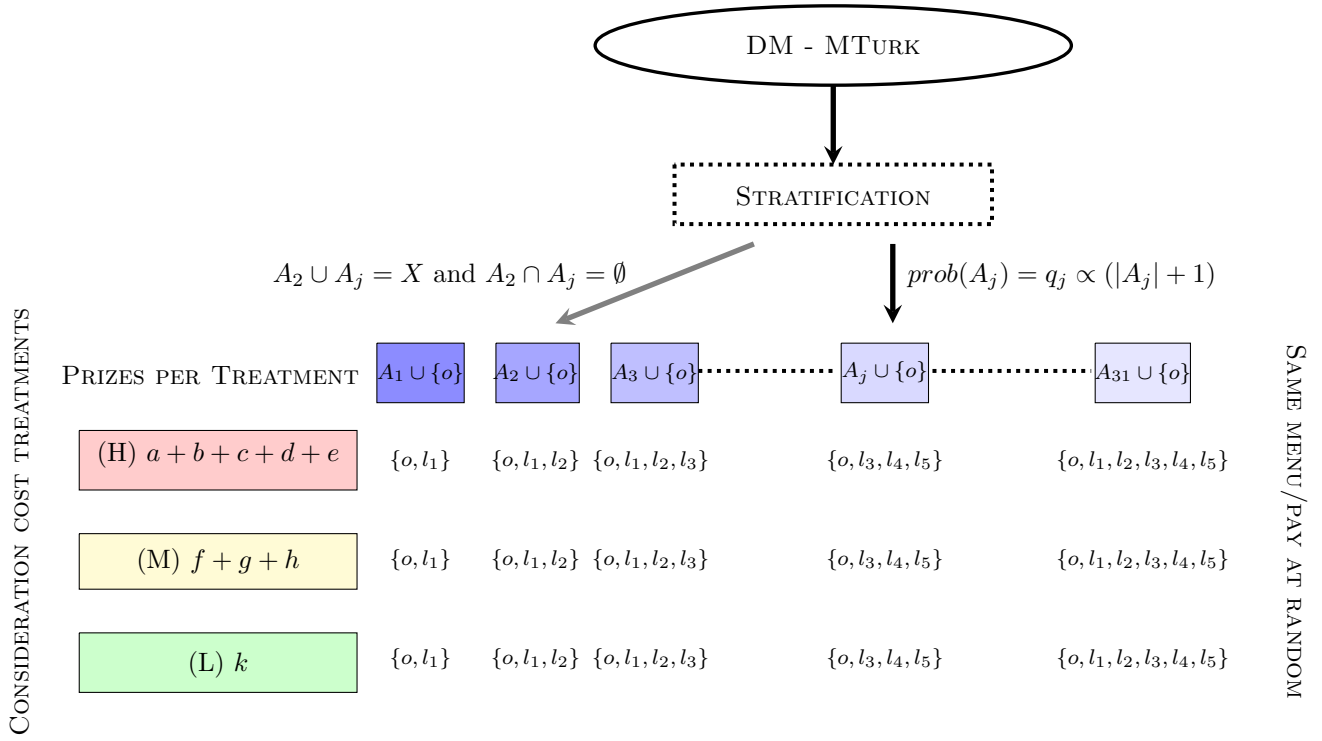


Figure 4 – EXPERIMENTAL DESIGN: DM i draws with probability $p(A_j)$ menu A_j with $|A_j| \in \{3, 4, 5\}$. In the picture, $A_j = \{l_3, l_4, l_5\}$. Therefore she is asked to choose from menus $A_j \cup \{o\}$ and $A_2 \cup \{o\}$, since $A_2 \cup A_j = X$ and $A_2 \cap A_j = \emptyset$. She is faced 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.

3.3. 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 (i.e., frame independence or ineffective consideration cost), 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 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.

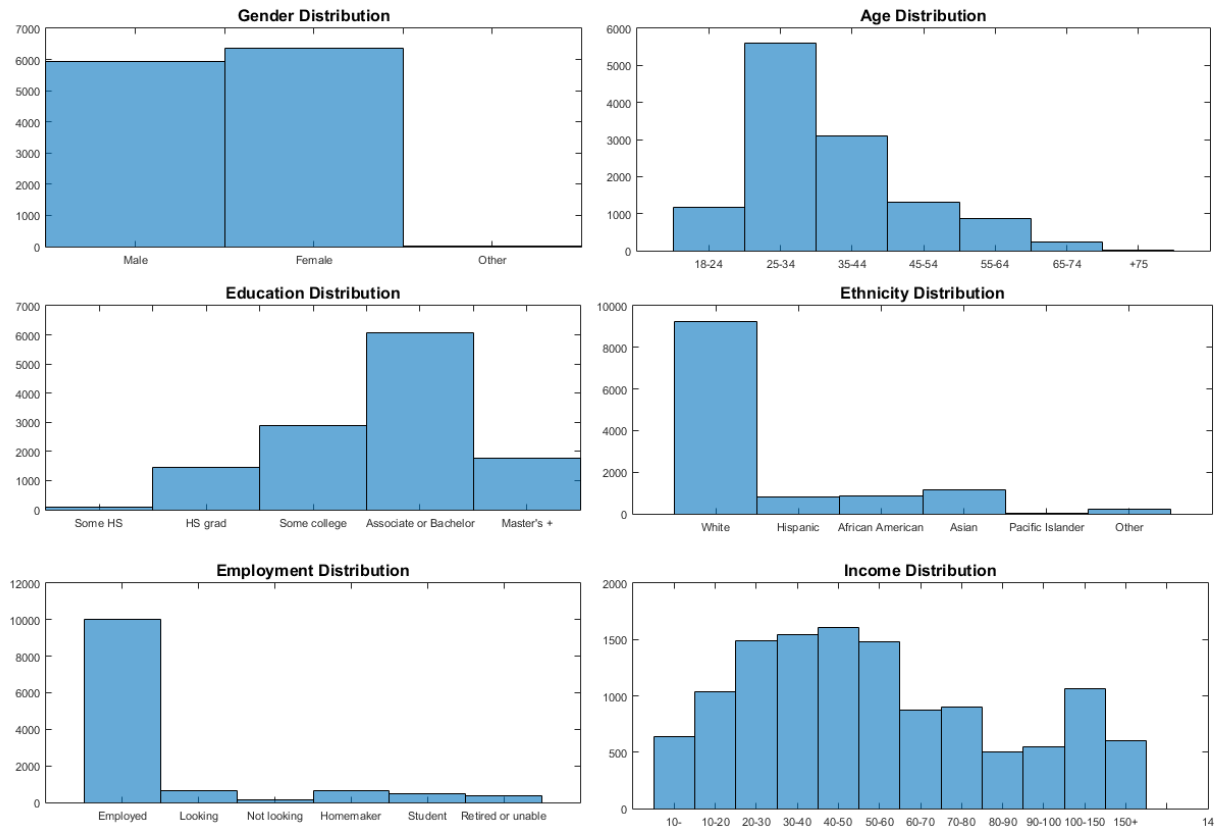


Figure 5 – DISTRIBUTION OF DEMOGRAPHICS IN SAMPLE

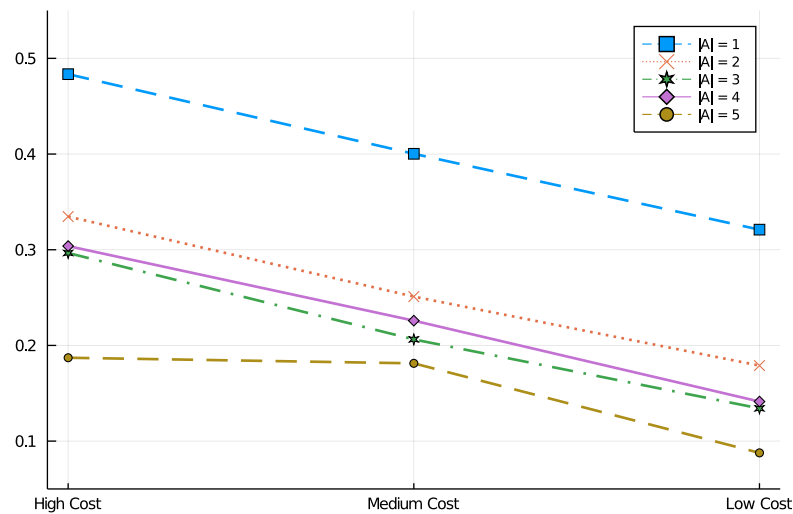


Figure 6 – FREQUENCY OF THE OUTSIDE OPTION CONDITIONAL ON THE MENU SIZE AND FRAME.

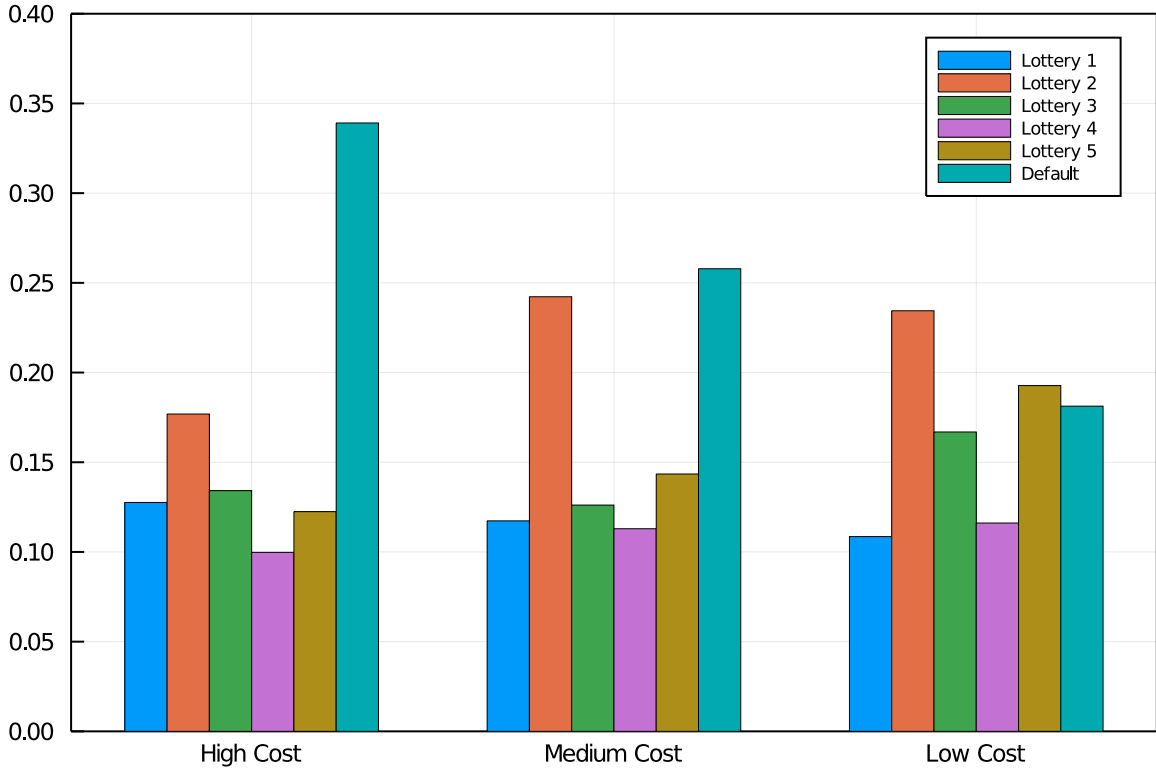


Figure 7 – FREQUENCIES OF CHOICES FOR DIFFERENT MENUS AND FRAMES

In passing, we highlight the remarkable linear relationship between the outside option frequency of choice and the number of operations (consideration cost) for almost all menu sizes. We remind the reader here that high, medium, and low cost corresponds to 5, 3, and 1 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. Nonetheless, the linear relation was unexpected. To our knowledge no theory of (random) consideration predicts it. This is a new stylized fact and we leave its further analysis for future work.

Figure 7 shows the effect of the frames/cost-treatments on the choice frequencies. The harder it is to understand the lotteries,³⁰ 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

³⁰Here 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.

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 95 percent confidence level.

3.4. Evidence for Total Monotonic Attention

Recall that, under the null hypothesis, the dataset 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 dataset. 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.

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 RUM cannot describe this dataset because total monotonicity should hold for all alternatives (Block and Marschak, 1960).³¹ However, this is not enough to conclude that RUM cannot describe the dataset, because of sample variability. Nonetheless, we reject the null hypothesis that RUM can describe this dataset. In addition, the nonmonotonicity of the nondefault alternatives is evidence in favor of LA as a good candidate to describe this dataset. We confirm this insight in our formal testing section by showing that LA does a good job at describing this dataset.

³¹Consistency with RUM is equivalent to satisfying the Block-Marschack conditions.

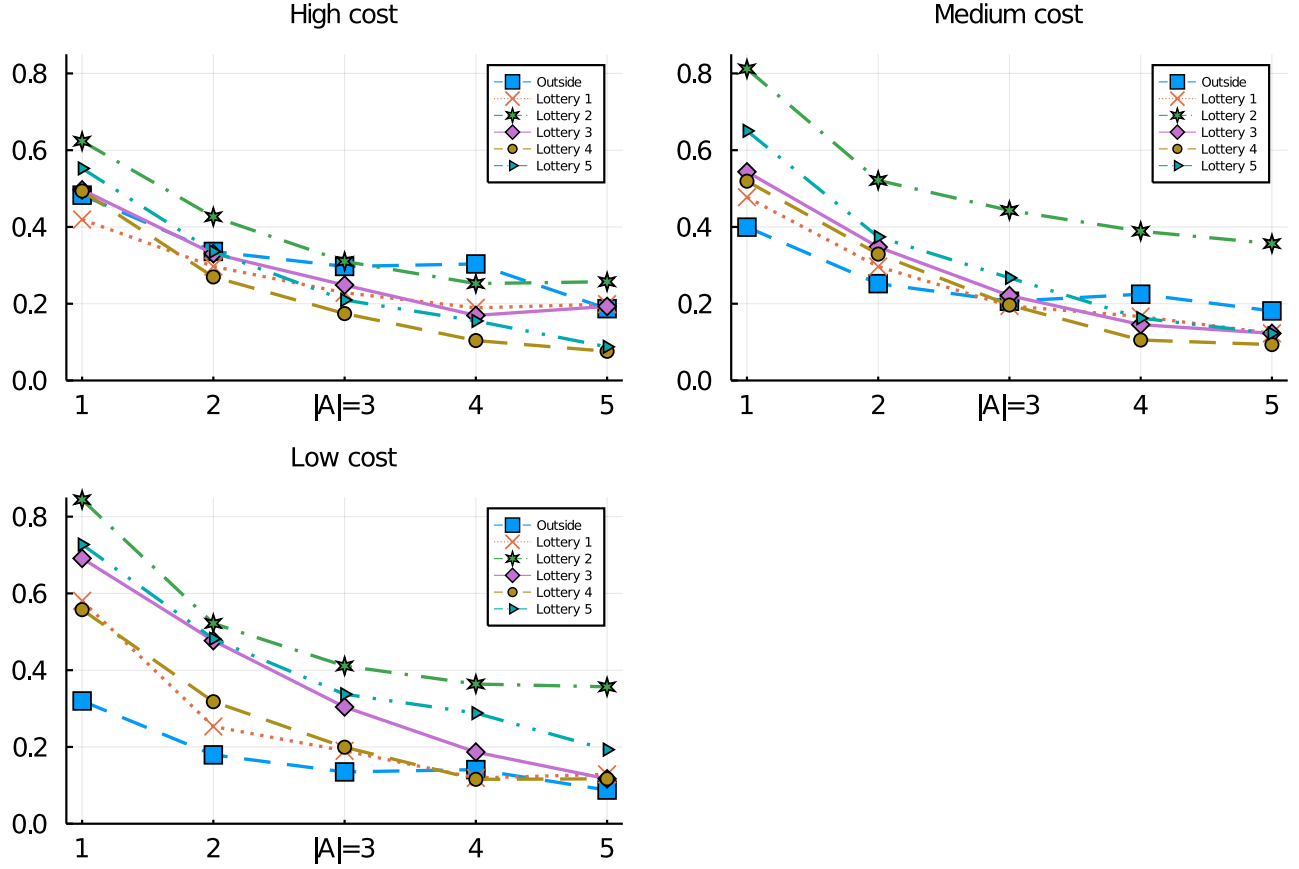


Figure 8 – FREQUENCY OF CHOICE FOR ALL FRAMES AS A FUNCTION OF THE MENU SIZE

4. 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.

4.1. Survival race the L-B-rule v.s. RUM: Stability of Preferences

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

First, we test whether the RUM model can explain the data with the restriction that the distribution of preferences remains frame-independent:

$$\pi_{\text{Low}}^{\text{RUM}} = \pi_{\text{Medium}}^{\text{RUM}} = \pi_{\text{High}}^{\text{RUM}}.$$

Second, we test whether the underlying distribution of preferences at different cost levels is the same for the LA and EBA models by imposing

$$\pi_{\text{Low}}^{\text{LA}} = \pi_{\text{Medium}}^{\text{LA}} = \pi_{\text{High}}^{\text{LA}}$$

and

$$\pi_{\text{Low}}^{\text{EBA}} = \pi_{\text{Medium}}^{\text{EBA}} = \pi_{\text{High}}^{\text{EBA}},$$

respectively. The distributions $\pi_{\text{cost}}^{\text{LA}}$ and $\pi_{\text{cost}}^{\text{EBA}}$, $\text{cost} \in \{\text{Low}, \text{Medium}, \text{High}\}$, are the distribution over preferences implied by the LA and the EBA models.

Adding the preference stability restriction does not require any modification to our testing framework. One just needs to compute \hat{g}^L for every frame and then stuck them together. The matrix G is replaced by the stacked matrix, $[G'G'G']'$; the columns remain the same as in the test for a fixed cost because the preference support is unchanged.

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

Table 2 – TESTING RESULTS UNDER PREFERENCE STABILITY

MODEL	T_n	P-VALUE
RUM	3231.59	<0.001
LA	32635.85	0.767
EBA	25304.92	0.001

Notes: Number of bootstrap replications=1000.

replications for every test statistic were conducted) for different models.³² First, we strongly reject RUM at any reasonable confidence 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 confidence 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.

4.2. 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. 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 Monte Carlo simulation. In particular, we show that our test can reject the false null hypothesis of data being consistent with

³²The 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.

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, e.g. expected utility, our results relate to the work of [Freeman et al. \(2018\)](#). 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. \(2018\)](#) 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 dataset and RUM fails to do the same are robust to this assumption. In contrast, we leave it as an open question whether EBA can explain this dataset 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 \(2018\)](#). However, we may lose statistical power. The study of consideration set rules and their relation to demographics is beyond the scope of this paper.

We finish this section by discussing our model and our findings in relation to Rational Inattention (RI) models. [Caplin et al. \(2019\)](#) shows that rational inattentive DMs form (deterministic) consideration sets. Generally, RI primitives cannot be point-identified with standard stochastic choice datasets. Nonetheless, RI models may still have testable implications in standard stochastic choice datasets. 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.

Choice Overload and Attraction Effect in the Aggregate Regarding, the traditional evidence against RUM, two important empirical anomalies contradict it: the attraction effect (Huber et al., 1982) and choice overload (Iyengar and Lepper, 2000).³³ Our findings are informative on whether these effects, which may be present at the individual level, still matter at the population level. We find evidence for the attraction effect driving the rejection of RUM. In contrast to RUM, LA can accommodate the attraction effect (Brady and Rehbeck, 2016). This and the presence of the attraction effect in our sample allow us to differentiate between RUM and LA. However, neither RUM nor any model of limited consideration that we study (including Cattaneo et al., 2020) can rationalize choice overload.³⁴ 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.³⁵ We find no statistical support for choice overload in our dataset. However, our choice set is of moderate size. We leave the question of whether choice overload matters at the aggregate population level as an open question.

5. 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 dataset to test RUM and two models of limited consideration, LA and EBA.

These models provide testable implications on choices that uniquely identify the stochastic

³³The attraction effect refers to a phenomenon when, as a new alternative is added to the choice set, the probability of some of the existing items is boosted. 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.

³⁴Chernev et al. (2015) provides a meta-data analysis of the determinants of choice overload.

³⁵ Other models of stochastic choice that are not models of limited consideration usually can accommodate choice overload. See, Fudenberg et al. (2015), Echenique et al. (2018), Kovach and Tserenjigmid (2018), and Natenzon (2018).

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 on 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.

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 $\pi \in \Delta(R(X))$ be a uniform probability. The pair $((m_A)_{A \in \mathcal{M}}, \pi)$ 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

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

Rearranging and replacing the choice of π in the above equation we get:

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

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

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

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

- If $m \in \mathcal{M}^{\text{LA}}$, then $m_A(D) = \sum_{C \subseteq A} \frac{\eta(D)}{\eta(C)}$ 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 $m_A(D) = \sum_{C \subseteq A} \frac{\eta(D)}{\eta(C)}$ 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 $L \in \{\text{LA}, \text{MM}, \text{EBA}, \text{FC}\}$ 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 $L \in \{LA, MM, EBA, FC\}$. 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 dataset, 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

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

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 $\eta^{EBA}(D) = \sum_{A \subseteq D: D \in \mathcal{A}} (-1)^{|D \setminus A|} (1 - p(o, X \setminus A)) \geq 0$ 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

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

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

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

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

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

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

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 dataset 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 10. 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 $p_{\pi}^L = p_{\pi'}^L$.

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. 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.

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 Random Utility Model

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 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 the Random Attention Model

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,

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 if 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 $m \in \mathcal{M}^{\text{LA}}$ 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 3 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 3 (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 $c \succ_2 b \succ_2 a$.³⁶ 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 $p(b, \{a, b, c\}) > p(b, \{b, c\})$).

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, miss-perception 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

³⁶In 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 $a \succ b \succ c$ for individuals that are risk-neutrals ($\sigma = 0$), while $c \succ b \succ a$ for risk averse individuals ($\sigma > 0.5$). Holt and Laury (2002) finds that these types are common in their experiment across payment schemes.

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 miss-perception as:

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

for all $a \in A$ (and $\mathbf{e} = (\mathbf{e}_a)_{a \in A}$).³⁷

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

$$\pi_e(\succ: a \succ b \ \forall b \in A \setminus \{a\}) = \mathbb{P}(\mathbf{u}_a + \mathbf{e}_a > \mathbf{u}_b + \mathbf{e}_b \ \forall b \in A \setminus \{a\}).³⁸$$

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 miss-perception 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 miss-perception of alternatives is that we may loose 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 miss-perception 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 miss-perception 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 miss-perception is important in our design.

³⁷Note 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 miss-perception (by-design) is the same across menus.

³⁸Where $\pi_e(\succ: \text{property})$ denotes the cumulative probability of preferences that have a certain property.

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 datasets to be indentified/tested. That is, generally they cannot be identified with standard stochastic choice datasets. In that sense, we cannot do a full comparison between RI models and our approach since the dataset requirements are different. However, we can derive some implications of RI behavior for our dataset.

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 and 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 $p_v(\cdot, A) \in \Delta(A \cup \{o\})$, 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 et al. \(2017\)](#), 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 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 percent 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 dataset. 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 datasets. We focus on collecting datasets 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 [Fosgerau et al. \(2017\)](#) and from [Aguiar et al. \(2016\)](#) that showed the that 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.

Optimal Random Consideration for LA It is still an open question whether the (reduced-form) LA random consideration model can be generated from a behavioral optimization problem. Here we show that LA can be obtained as the result of allocating attention optimally. Consider a DM that, faced with a menu A , needs to allocate her attention, measured by $m_A \in \Delta(2^A)$, over all possible consideration sets in A , including the empty set. We assume that each set D has a (deterministic) attractiveness index $\alpha(D)$, with $\alpha : 2^X \rightarrow [0, 1]$, that measures both, how enticing a consideration set is, but also how complex it is to understand. That is, is a net measure of attractiveness with respect to how hard is to consider it. Therefore, if $\alpha(C) > \alpha(F)$ we say that C is in net terms more attractive than F . The key assumption is that α is deterministic and does not depend on the distribution of preferences $\pi \in \Delta(R(X))$ in the population. This implies independence between random consideration consideration sets captured by m_A and random preferences captured by π .

Choice difficulty is captured by a cognitive cost function K that is menu independent but depends on the allocated attention, measured by $m_A(D)$, for a fixed $D \subseteq A$: $K : [0, 1] \rightarrow \mathbb{R} \cup \infty$. Following [Fudenberg et al. \(2015\)](#) we assume that K is strictly convex. Then, DM's problem is to optimally find $m_A \in \Delta(2^A)$ to maximize expected attractiveness of the menu given the cognitive

cost of processing it. Formally:

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

where $\alpha_A : 2^A \rightarrow \mathbb{R}$ is a menu-dependent attractiveness that depends on α (the menu independent attractiveness defined before).

When $K(m(C)) = 0$, and $\alpha_C(D) = \alpha(D)$ for any $C \subseteq D$, under the assumption that $\alpha(A) > \alpha(D)$ for all $D \subset A$ we get $m_A^{\text{FC}}(A) = 1$. That is, the DM is consistent with FC.

More importantly, under entropy cost, $K(m(D)) = \theta m(D) \log m(D)$, from [Fudenberg et al. \(2015\)](#) (and $\alpha_A(D) = \alpha(D)$) we get that:

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

That is, optimal consideration is consistent with LA. Note that the attention-index can be defined as:

$$\eta(D) = \exp(\theta \alpha(D)),$$

for all $D \in 2^X$.

Turns out that EBA can also come from a different optimization problem with a quadratic cost $K(m) = \frac{\nu}{2} m^2$, where the attractiveness of the menu is given by the cumulative attractiveness that is menu dependent:

$$\alpha_A(D) = \sum_{C \in 2^X: C \cap A = D} \alpha(C).$$

With this in hand we get:

$$m_A^{\text{EBA}}(D) = \frac{1}{\nu} \sum_{C \in 2^X: C \cap A = D} \alpha(C),$$

where ν ensure that this is well-behaved probability distribution. In this case, the attention-index is $\eta = \alpha$.

D. Experiment

D.1. Experimental Design

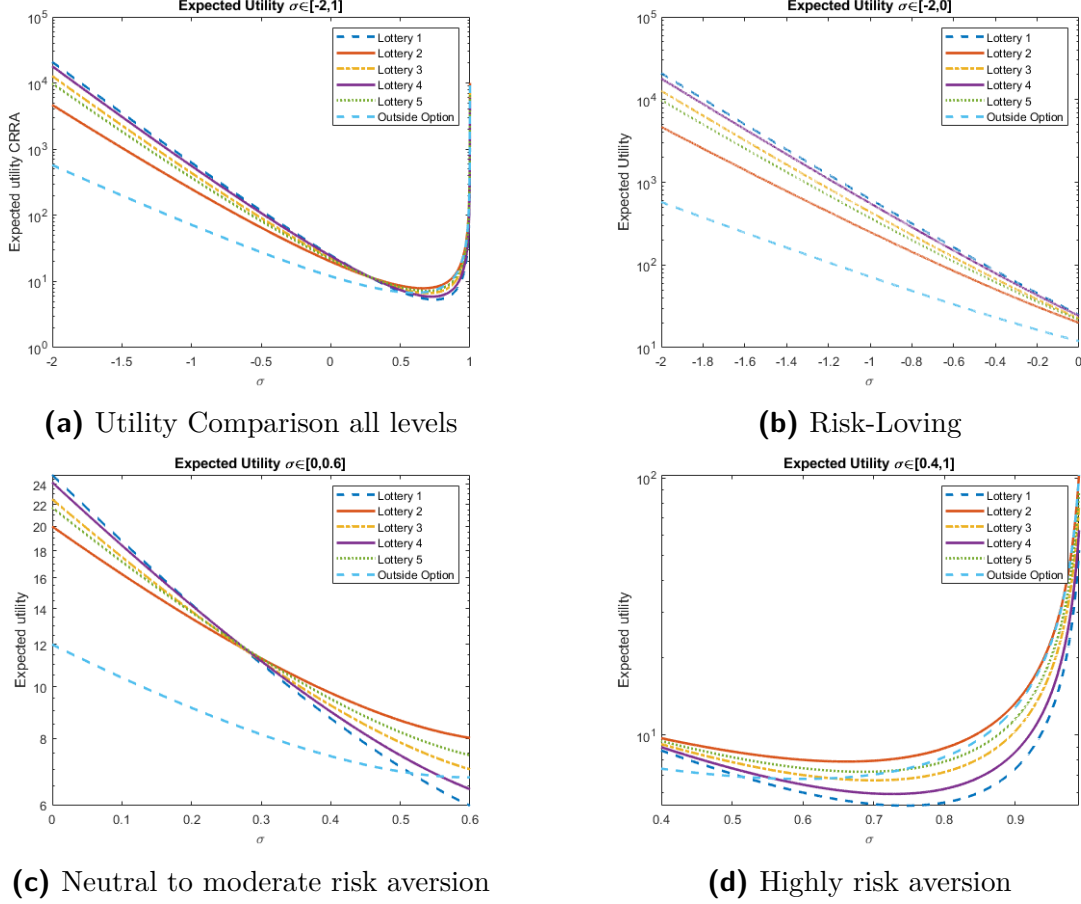


Figure 9 – COMPARISON BETWEEN LOTTERIES IN TERMS OF RISK AVERSION. Simulations for EU maximizers individuals, with CRRA utility function $u(x) = \frac{x^{1-\sigma}}{1-\sigma}$.

D.2. Sample

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

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			

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 number of 2135 (the targeted number is 2130) observations for 4099 tasks (the targeted number of tasks 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.

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

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

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

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

The assumed process implies that a fraction $1 - \lambda$ of the population exhibits choice overload.³⁹

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 500 replication the table displays the proportion of simulations that are rejected at the 90 percent and 95 percent confidence levels. As expected, the fraction of rejections is bigger for smaller values of λ . For $\lambda = 0.25$ the rejection rate is 100 percent. We observe that at comparable sample size to our experiment the mixed process is rejected with power close to 1 when the choice overload

³⁹This 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 modelled 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 $p(o, A) = c|A|$.

Table 5 – The table displays the proportion of rejections at the 90 percent and 95 percent confidence levels for LA-B. Sample size=4000. Number of MC replications=500. Number of bootstrap replications=500

PROCESS	CONFIDENCE LEVEL	
	90%	95%
$\lambda = 0.25$	1	1
$\lambda = 0.50$	0.464	0.73

fraction of DMs is moderate.

F. Experiment Instructions

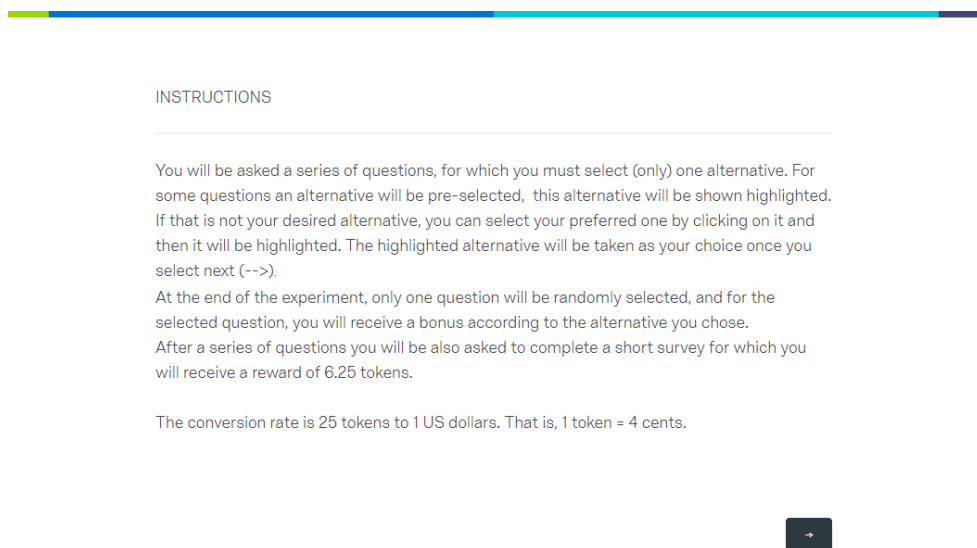


Figure 10 – Instructions page 1

For example, if you face the following question

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

Get 5 tokens for sure.

Prize	Probability
8 tokens	50%
2 tokens	50%

Prize	Probability
28 tokens	25%
0 tokens	75%

Figure 11 – Instructions page 2

For example, if you face the following question, you must select an answer. Imagine you have selected alternative 3.

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

Get 5 tokens for sure.

Prize	Probability
8 tokens	50%
2 tokens	50%

Prize	Probability
28 tokens	25%
0 tokens	75%

Figure 12 – Instructions page 3

For example, if you face the following question, you must select an answer. Imagine you have selected alternative 3. Then you must click on it, and it will be highlighted.

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

Get 5 tokens for sure.

Prize	Probability
8 tokens	50%
2 tokens	50%

Prize	Probability
28 tokens	25%
0 tokens	75%

Figure 13 – Instructions page 4

For example, if you face the following question, you must select an answer. Imagine you have selected alternative 3. Then you must click on it, and it will be highlighted. Once you have chosen your desired alternative, hit next (-->)

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

Get 5 tokens for sure.

Prize	Probability
8 tokens	50%
2 tokens	50%

Prize	Probability
28 tokens	25%
0 tokens	75%

Figure 14 – Instructions page 5

You may also get questions in a slightly modified format. You must choose your preferred alternative among the ones displayed on the screen. For example:

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

Get 5 tokens for sure.

Prize	Probability
$(8+4-3-5-2)$ tokens	25%
$(11-9+5+3-2)$ tokens	25%
$(2+3+1-5+11)$ tokens	50%

Prize	Probability
$(2+5+9+7-3)$ tokens	25%
$(8+9-7-5-3)$ tokens	75%

→

Figure 15 – Instructions page 6

INSTRUCTIONS

You will be paid your show up fee of 10 tokens and a variable amount of tokens that depends on your choices. The variable amount will be the outcome of a randomly selected question among the answered ones. Given the selected question, your chosen alternative will be executed and you will be paid accordingly.

For example, assume that you were only asked the two questions above. Namely:

QUESTION 1

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

Get 5 tokens for sure.

Prize	Probability
8 tokens	50%
2 tokens	50%

Prize	Probability
28 tokens	25%
0 tokens	75%

Figure 16 – Instructions page 7

INSTRUCTIONS

You will be paid a bonus of tokens that depends on your choices in addition to the reward of 25 cents. The bonus will be the outcome of a randomly selected question among the answered ones. Given the selected question, your chosen alternative will be executed and you will be paid accordingly.

For example, assume that you were only asked the two questions above. Namely:

QUESTION 2

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

Get 5 tokens for sure.

Prize	Probability
$(8+4-3-5-2)$ tokens	25%
$(11-9+5+3-2)$ tokens	25%
$(2+3+1-5+1)$ tokens	50%

Prize	Probability
$(2+5+9+7-3)$ tokens	25%
$(8+9-7-5-3)$ tokens	75%

Figure 17 – Instructions page 8

INSTRUCTIONS

You will be paid a bonus of tokens that depends on your choices in addition to the reward of 25 cents. The bonus will be the outcome of a randomly selected question among the answered ones. Given the selected question, your chosen alternative will be executed and you will be paid accordingly.

For example, assume that you were only asked the two questions above. Assume that these were your choices.

CHOICE IN QUESTION 1

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

Get 5 tokens for sure.

Prize	Probability
8 tokens	50%
2 tokens	50%

Prize	Probability
28 tokens	25%
0 tokens	75%

Figure 18 – Instructions page 9

INSTRUCTIONS

You will be paid a bonus of tokens that depends on your choices in addition to the reward of 25 cents. The bonus will be the outcome of a randomly selected question among the answered ones. Given the selected question, your chosen alternative will be executed and you will be paid accordingly.

For example, assume that you were only asked the two questions above. Assume that these were your choices.

CHOICE IN QUESTION 2

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

Get 5 tokens for sure.

Prize	Probability
$(8+4-3-5-2)$ tokens	25%
$(11-9+5+3-2)$ tokens	25%
$(2+3+1-5+11)$ tokens	50%

Prize	Probability
$(2+5+9+7-3)$ tokens	25%
$(8+9-7-5-3)$ tokens	75%

Figure 19 – Instructions page 10

INSTRUCTIONS

You will be paid a bonus of tokens that depends on your choices in addition to the reward of 25 cents. The bonus will be the outcome of a random selected question among the answered ones. Given the selected question, your chosen alternative will be executed and you will be paid accordingly.

Since you answered two questions, each will be selected with probability $1/2$. You can think that a coin is being flipped and if it shows heads then you will be paid according to your choice in Question 1; while if it shows tails you will be paid according to your choice in Question 2. If you were to answer 4 questions, then each of these will be selected with probability $1/4$.

For the example this implies that:

With probability $1/2$ you will get the outcome of Question 1. That is, with probability $1/4$ you get 28 tokens, with probability $3/4$ you get 0.

With probability $1/2$ you will get the outcome of Question 2. That is you would get 5 tokens for sure.



Figure 20 – Instructions page 11

Before you start making your choices here is an example.

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

Get 5 tokens for sure.

Prize	Probability
8 tokens	50%
2 tokens	50%

Prize	Probability
28 tokens	25%
0 tokens	75%



Figure 21 – Instructions page 12

With probability 1/2 you get 8 tokens, with probability 1/2 you get 2 tokens.



Figure 22 – Instructions page 13



Figure 23 – Instructions page 14

References

- Abaluck, J. and Adams, A. (2021). What do consumers consider before they choose? identification from asymmetric demand responses. *The Quarterly Journal of Economics*, Accepted.
- Aguiar, V. H. (2017). Random categorization and bounded rationality. *Economics Letters*, 159:46–52.
- Aguiar, V. H., Boccardi, M. J., and Dean, M. (2016). Satisficing and stochastic choice. *Journal of Economic Theory*, 166:445 – 482.
- Apesteguia, J., Ballester, M. Á., et al. (2018). Separating predicted randomness from noise. Technical report.
- Barseghyan, L., Coughlin, M., Molinari, F., and Teitelbaum, J. C. (2019a). Heterogeneous choice sets and preferences. *arXiv preprint arXiv:1907.02337*.
- Barseghyan, L., Molinari, F., and Thirkettle, M. (2019b). Discrete choice under risk with limited consideration. *arXiv preprint arXiv:1902.06629*.
- Block, H. D. and Marschak, J. (1960). Random orderings and stochastic theories of responses. *Contributions to probability and statistics*, 2:97–132.
- Brady, R. L. and Rehbeck, J. (2016). Menu-dependent stochastic feasibility. *Econometrica*, 84(3):1203–1223.
- Caplin, A., Dean, M., and Leahy, J. (2019). Rational inattention, optimal consideration sets, and stochastic choice. *The Review of Economic Studies*, 86(3):1061–1094.

- Cattaneo, M. D., Ma, X., Masatlioglu, Y., and Suleymanov, E. (2020). A random attention model. *Journal of Political Economy*, 128(7):2796–2836.
- Chateauneuf, A. and Jaffray, J.-Y. (1989). Some characterizations of lower probabilities and other monotone capacities through the use of möbius inversion. *Mathematical social sciences*, 17(3):263–283.
- Chernev, A., Böckenholt, U., and Goodman, J. (2015). Choice overload: A conceptual review and meta-analysis. *Journal of Consumer Psychology*, 25(2):333–358.
- Dardanoni, V., Manzini, P., Mariotti, M., and Tyson, C. J. (2020). Inferring cognitive heterogeneity from aggregate choices. *Econometrica*, 88(3):1269–1296.
- De Clippel, G. and Rozen, K. (2018). Bounded rationality and limited datasets.
- De Los Santos, B., Hortaçsu, A., and Wildenbeest, M. R. (2012). Testing models of consumer search using data on web browsing and purchasing behavior. *American Economic Review*, 102(6):2955–80.
- Dean, M. and McNeill, J. (2014). Preference for flexibility and random choice: an experimental analysis. Technical report, Working Paper, Brown University, Department of Economics.
- Echenique, F., Saito, K., and Tserenjigmid, G. (2018). The perception-adjusted luce model. *Mathematical Social Sciences*, 93:67–76.
- Falmagne, J. C. (1978). A representation theorem for finite random scale systems. *Journal of Mathematical Psychology*, 18(1):52–72.
- Fosgerau, M., Melo, E., Palma, A. d., and Shum, M. (2017). Discrete choice and rational inattention: A general equivalence result.
- Freeman, D., Halevy, Y., Kneeland, T., et al. (2018). Eliciting risk preferences using choice lists. Technical report, Vancouver School of Economics.
- Fudenberg, D., Iijima, R., and Strzalecki, T. (2015). Stochastic choice and revealed perturbed utility. *Econometrica*, 83(6):2371–2409.
- Goeree, M. S. (2008). Limited information and advertising in the us personal computer industry. *Econometrica*, 76(5):1017–1074.

- Heiss, F., McFadden, D., Winter, J., Wuppermann, A., and Zhou, B. (2016). Inattention and switching costs as sources of inertia in medicare part d. Technical report, National Bureau of Economic Research.
- Ho, K., Hogan, J., and Scott Morton, F. (2017). The impact of consumer inattention on insurer pricing in the medicare part d program. *The RAND Journal of Economics*, 48(4):877–905.
- Holt, C. A. and Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review*, 92(5):1644–1655.
- Honka, E. (2014). Quantifying search and switching costs in the us auto insurance industry. *The RAND Journal of Economics*, 45(4):847–884.
- Honka, E., Hortaçsu, A., and Vitorino, M. A. (2017). Advertising, consumer awareness, and choice: Evidence from the us banking industry. *The RAND Journal of Economics*, 48(3):611–646.
- Horan, S. (2019). Random consideration and choice: A case study of default options. *Mathematical Social Sciences*, 102:73–84.
- Hortaçsu, A., Madanizadeh, S. A., and Puller, S. L. (2017). Power to choose? an analysis of consumer inertia in the residential electricity market. *American Economic Journal: Economic Policy*, 9(4):192–226.
- Horton, J. J., Rand, D. G., and Zeckhauser, R. J. (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental economics*, 14(3):399–425.
- Huber, J., Payne, J. W., and Puto, C. (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of consumer research*, 9(1):90–98.
- Iyengar, S. S. and Lepper, M. R. (2000). When choice is demotivating: Can one desire too much of a good thing? *Journal of personality and social psychology*, 79(6):995.
- Kashaev, N. and Lazzati, N. (2021). Peer effects in random consideration sets. *arXiv preprint arXiv:1904.06742*.
- Kim, J. (2016). Discounting, dynamic consistency, and cooperation in an infinitely repeated game experiment. Technical report, Working paper.

- Kitamura, Y. and Stoye, J. (2018). Nonparametric analysis of random utility models. *Econometrica*, 86(6):1883–1909.
- Kovach, M. and Tserenjigmid, G. (2018). The imbalanced luce model.
- Kovach, M. and Ülkü, L. (2020). Satisficing with a variable threshold. *Journal of Mathematical Economics*, 87:67–76.
- Krantz, D. H. (1965). *The scaling of small and large color differences*. PhD thesis.
- Lleras, J. S., Masatlioglu, Y., Nakajima, D., and Ozbay, E. Y. (2017). When more is less: Limited consideration. *Journal of Economic Theory*, 170:70–85.
- Manzini, P. and Mariotti, M. (2014). Stochastic choice and consideration sets. *Econometrica*, 82(3):1153–1176.
- Masatlioglu, Y., Nakajima, D., and Ozbay, E. Y. (2012). Revealed attention. *American Economic Review*, 102(5):2183–2205.
- McCausland, W. J., Marley, A., Davis-Stober, C., Brown, N., and Park, S. (2018). Testing the random utility hypothesis directly. *The Economic Journal*.
- McFadden, D. and Richter, M. K. (1990). Stochastic rationality and revealed stochastic preference. *Preferences, Uncertainty, and Optimality, Essays in Honor of Leo Hurwicz*, Westview Press: Boulder, CO, pages 161–186.
- Natenzon, P. (2018). Random choice and learning.
- Rand, D. G., Greene, J. D., and Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature*, 489(7416):427.
- Rieskamp, J., Busemeyer, J. R., and Mellers, B. A. (2006). Extending the bounds of rationality: Evidence and theories of preferential choice. *Journal of Economic Literature*, 44(3):631–661.
- Salant, Y. and Rubinstein, A. (2008). (a, f): choice with frames. *The Review of Economic Studies*, 75(4):1287–1296.
- Shafer, G. (1976). *A mathematical theory of evidence*, volume 42. Princeton university press.

Suleymanov, E. (2018). Stochastic attention and search.

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