Does Random Consideration Explain Behavior when Choice is Hard? Evidence from a Large-scale Experiment*†

Victor H. Aguiar‡ Maria Jose Boccardi§

Nail Kashaev¶ Jeongbin Kim‖

This version: November 2019/ First version: November 2018

Abstract We study population behavior when choice is hard because considering alternatives is costly. To simplify their choice problem, individuals may pay attention to only a subset of available alternatives. We design and implement a novel online experiment that exogenously varies choice sets and consideration costs for a large sample of individuals. We provide a theoretical and statistical framework that allows us to test random consideration at the population level. Within this framework, we compare competing models of random consideration. We find that the standard random utility model fails to explain the population behavior. However, our results suggest that a model of random consideration with logit attention and heterogeneous preferences provides a good explanation for the population behavior. Finally, we find that the random consideration rule that subjects use is different for different consideration costs while preferences are not. We observe that the higher the consideration cost the further behavior is from the full-consideration benchmark, which supports the hypothesis that hard choices have a substantial negative impact on welfare via limited consideration.

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

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

*We would like to thank Roy Allen, Jose Apesteguia, Miguel Ballester, Juan Dubra, Mikhail Freer, Yoram Halevy, Paola Manzini, Marco Mariotti, John Rehbeck, and Jörg Stoye for useful comments and suggestions. Aguiar and Kashaev gratefully acknowledge financial support from the Western Social Science Faculty grant (FRDF R5533A02).

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

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

§Social Sciences Division, New York University Abu Dhabi, maria.jose.boccardi@nyu.edu

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

‖Division of the Humanities and Social Sciences, California Institute of Technology, jbkim@caltech.edu

1. Introduction

A fundamental question in social science is how to describe the behavior of a population of decision makers (DMs). This question is tightly related to identifying the distribution of preferences that generates behavior. The random utility model (RUM) (McFadden and Richter (1990)) is the standard tool to *describe* behavior and to *identify* preferences.[1] RUM assumes that DMs maximize their preferences over their choice set. However, RUM may fail at describing behavior and identifying preferences if DMs do not consider all available alternatives when choice is hard. We say choice is hard when there is a cost to understanding the decision task. In this situation, DMs may use two-stage procedures: first simplifying choice by using a consideration set, and only then choosing the best alternative among those considered. Thus, DMs may choose dominated alternatives when facing a cost of consideration.[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 and their ability to identify the distribution of preferences. By focusing on the inattention channel, 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, in order to test these models in finite samples, we need to account for sampling variability.

Empirically, we design a novel experiment with two independent sources of exogenous variation: (i) full variation in choice sets (menus), and (ii) three levels of consideration cost. Full variation in choice sets means that all possible choice sets are observed by the researcher. Full variation in

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

[2]Ho et al. (2017) and Heiss et al. (2016) show the effect of limited consideration in the health insurance market and its effects in overspending. Hortaçsu et al. (2017) shows that for the residential electricity market in Texas, low-cost information interventions could increase consumer surplus by $100 or more per year. Honka (2014) and Honka et al. (2017) examine the effects of limited consideration in the US auto insurance and banking industries, respectively. De Los Santos et al. (2012) uses data on web browsing to show that consumers consider a relatively small number of websites when shopping online. Barseghyan et al. (2018) uses administrative data on insurance purchases to show that DMs choose dominated options due to limited consideration, in an environment without menu variation.

[3]See, for instance, Aguiar et al. (2016), Brady and Rehbeck (2016), Caplin et al. (2016), Aguiar (2017), Kovach and Ulku (2017), Lleras et al. (2017), Cattaneo et al. (2017), and Horan (2018).

[4]For evidence from the field for RUM using surveys, see Kitamura and Stoye (2018). For experimental evidence, see McCausland et al. (2018).

choice sets allows us to test consideration-mediated choice theories, and to identify preferences in a large cross-section of heterogeneous individuals. Variation in the consideration cost allows us to study the relevance of different consideration set models. We also introduce a dominated default alternative that works like an opportunity cost of paying attention. We conducted this experiment online in Amazon Mechanical Turk (MTurk), collecting 12297 independent choice observations from 2135 individuals. To the best of our knowledge, this is the largest experiment to date that permits the testing of RUM and its limited consideration extensions.[5] We reject the null hypothesis of the validity of RUM and fail to reject the hypothesis that DMs have limited consideration.

We model behavior as if individuals may not consider all alternatives in a given choice set. DMs first pick a subset of alternatives, which we refer to as their *consideration set*. Choices are then given by the best alternative (as determined by the DMs' preferences) among the ones considered. Thus, the choices observed are the result of a combination of the unobserved preferences and the consideration set rules. Since unobserved preferences and consideration sets may vary across individuals, it is difficult to test consideration-mediated choice theories. We show that full variation in choice sets, and the presence of a default alternative, are sufficient to uniquely recover distribution over consideration sets given a particular consideration set rule. Moreover, recoverability of preferences under limited consideration is as good as in RUM if, in addition to full variation in choice sets, there is a positive mass of DMs who consider the full choice set.[6]

Once the underlying preference distribution is identified, we can construct the hypothetical distribution over choices generated by the identified preference distribution under full consideration. We show that, at the population level, testing a consideration-mediated choice theory with heterogeneous preferences 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).

In order 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 dataset. 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). Nonetheless, full exogenous variation in choice sets is an important data feature that is usually not satisfied in field data. In fact, one must observe, at least some individuals facing each possible choice set.

Additionally, Kitamura and Stoye (2018) argue that lack of exogenous variation in choice sets invalidates the standard RUM framework provided by McFadden and Richter (1990). In field studies the choice set variation may not be exogenous because the constraints that a DM faces may

[5] The only other 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). Both focus mainly on binary choice sets that are not sufficient for our goals, nor do they test limited consideration. In addition, we focus on statistical testing instead of goodness-of-fit measures or the computation of Bayes factors (the two approaches they use).

[6] Cattaneo et al. (2017) shows that many random consideration rules are equivalent to heterogeneity in deterministic attention filters (Masatlioglu et al. (2012)). In this sense, at the population level, we allow heterogeneity in both preferences and in consideration.

be correlated with her preferences.

Exogenous variation in the consideration cost allows us to study conditions under which different consideration models provide a better description of the behavior of the population. We introduce three cost treatments for every choice set. Each DM faces all three costs. These treatments require the DM to solve a simple cognitive task to understand the alternatives. The consideration cost is progressively reduced while we keep the choice set fixed.[7] Within this design we are able to understand how changing consideration costs may affect choices while we keep (the distribution of) preferences fixed.

We use these theoretical and experimental innovations to test well-known models of random consideration:[8] (i) the choice-set independent-consideration model of Manzini and Mariotti (2014) (MM), (ii) the logit attention model of Brady and Rehbeck (2016) (LA), and (iii) the random categorization model of Aguiar (2017) (RCG). There is increasing interest in incorporating limited consideration in discrete choice.[9] In particular, the influential and tractable MM model has become important tool for the analysis of limited attention in empirical work (e.g., Barseghyan et al. (2019), Dardanoni et al. (2019), and Kashaev and Lazzati (2019)). However, MM is highly stylized and assumes that consideration is driven by an item-dependent parameter (i.e., independence in consideration). We investigate from an experimental perspective whether this strong assumption is effective in explaining choice when it is hard. To do so we consider two extensions of MM that allow for substitution and complementarity in consideration, the LA and RCG models. These two generalizations have the property that their intersection is exactly the MM model (Suleymanov (2018)). We exploit the rich relationship among these models to learn about the true data generating process. We hope that our findings inform future empirical work in the field. Note that in the context of our experiment, full choice set variation allows us to test the assumption of independence in consideration.

First, we test these models and the benchmark RUM with and without conditioning on the cost level. Later we require the underlying preference relation to be stable among cost treatments while allowing the consideration rules to vary with the cost. Our main findings are: (i) We reject the hypothesis that RUM provides a good description of population behavior, when we pool observations across consideration costs. In contrast, the LA model with heterogeneous preferences cannot be rejected at the 95 percent confidence level. (ii) RUM provides a good description of behavior for high and low consideration costs while failing for the case of intermediate cost. LA with heterogeneity also provides a good description of the behavior of the population across all consideration cost levels. (iii) Perhaps surprisingly, the highly stylized MM model with preference heterogeneity cannot be rejected as a good description of the data in the high-cost treatment (it is rejected at the medium and low costs). The same is true for RCG (which nests MM).

By varying consideration costs while fixing the choice sets faced by DMs, our experimental design facilitates detecting failures of a given model at the task of identifying stable preferences

[7]The fact that the choice remains the same is not explicitly stated in the experiment instructions.

[8]Note that our methodology can be applied to other stochastic consideration rules.

[9]See, for instance, Goeree (2008), Abaluck and Adams (2017), Barseghyan et al. (2018), Barseghyan et al. (2019), and Dardanoni et al. (2019).

that are independent of the way alternatives are presented. Across our three cost treatments, the choice set remains the same. Under our incentives protocol (pay-at-random across tasks), the distribution of preferences must remain constant regardless of the consideration cost. By exploiting this feature of our design, we show that, in our sample, RUM fails in its identification task even when it describes behavior well. We provide evidence that population behavior is different between the high and low consideration cost treatments. This implies that we reject the null hypothesis that the same distribution of preferences, under RUM, governs the choice of the population in the different cost treatments. In sharp contrast, we fail to reject the null hypothesis that the underlying distribution of preferences is the same across cost treatments for LA.

Our work contributes to the recent experimental literature on stochastic choice, limited consideration, and departures from stochastic rationality (RUM).[10] Two important empirical anomalies contradict RUM: the attraction effect and choice overload. We are interested in whether these effects, which may be present at the individual level, still matter at the population level. The attraction effect refers to the case in which, as a new alternative is added to the choice set, the probability of some of the existing items is boosted. The attraction effect cannot be rationalized by RUM. In contrast, the LA model can explain it. The presence of the attraction effect in our sample allows us to differentiate between these two models. We find evidence for the attraction effect only for the intermediate consideration cost; for this consideration cost RUM fails but, at the 95 percent confidence level, we cannot reject LA.

Choice overload refers to the case in which the propensity of not choosing (or the probability of picking a default alternative) increases in larger choice sets. Neither RUM nor any model of limited consideration that we studied can rationalize choice overload.[11] Limited consideration is fundamentally at odds with choice overload since one of the important reasons to form a consideration set is to simplify choice. In fact, not even the most general model of random consideration, which so far, is the Random Attention Model (RAM) by Cattaneo et al. (2017), can explain choice overload.[12] We find no statistical support for choice overload. In sum, we find experimental evidence that random consideration with heterogeneous preferences can explain behavior when choice is hard, while random utility cannot.

This paper generalizes the methodology of Cattaneo et al. (2017) by allowing heterogeneity in preferences, when the choice sets include a default alternative. Cattaneo et al. (2017) provides a general framework to test different models of stochastic consideration when preferences are fixed (with and without a default). Therefore, their work is better suited for individual stochastic choice data.[13] Our contribution is designed to work at the (large) cross-section level where each individual,

[10]There is a vast literature documenting departures from fully rational behavior. Rieskamp et al. (2006) reviews many of these violations and the theories that have been proposed to rationalize them.

[11]Chernev et al. (2015) provides a meta-data analysis of the determinants of choice overload, reviewing previous literature on this subject. In this context, the presence of a large choice set of difficult complexity can increase the likelihood of both effects.

[12] 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).

[13]To the best of our knowledge, Allen and Rehbeck (2018) is the only other paper that proposes a (nonstatistical) test of stochastic choice from a nonparametric revealed-preferences perspective. Another way in which our approach

in principle, can face only one choice set. Moreover, our dataset allows us to do asymptotic inference accounting for sampling variability. In sharp contrast with the RAM framework, we show that identification of preferences is possible when choice is regular. We show that the default alternative is key in achieving point identification of the stochastic consideration rule, even under heterogeneity in preferences.

We also contribute to the empirical literature focused on identifying the distribution of preferences in the presence of limited consideration. With limited consideration, the classical RUM fails at identifying preferences. In addition, we show that, even under preference homogeneity, the recent approach of Cattaneo et al. (2017) may also fail at the identification task. In contrast to these approaches, we achieve exact identification of the distribution of consideration rules and the same level of preference identification as does RUM with full consideration. We do this by introducing a default alternative that has, plausibly, zero consideration cost and that captures the opportunity cost of considering any item in a choice set. Our approach differs from previous efforts that have used enhanced datasets to test for the presence of consideration in that we use only a standard stochastic choice dataset widely used in the discrete-choice literature. For example, employing eye-tracking data to identify attention, Reutskaja et al. (2011) provides evidence of search and satisficing.[14] Goeree (2008) and Van Nierop et al. (2010) approach the identification task by assuming that advertisements affect the formation of a consideration set, but not consumer preference.[15] Recently, Abaluck and Adams (2017) use structural restrictions on the elasticity behavior of demand to identify consideration sets and preferences. They implement an experiment in which they validate their model with choice set variation. (However, they deliberately ignore this information and instead try to recover it using their methodology). We differ from that work because we are focused on a large population (whereas their sample consists of only 150 subjects) (allowing us to do statistical inference) and we do not observe attributes (e.g., prices).

Finally, we estimate the distribution of preferences and consideration sets in our dataset, assuming that (i) the consideration rule is LA (logit attention), and (ii) preferences satisfy the expected utility restrictions (the independence axiom) and exhibit constant relative risk aversion (CRRA). We are able to do this because, when we use our testing procedure, we cannot reject these restrictions. We also cannot reject the fact that preferences are stable across consideration cost treatments while consideration is not. Imposing the CRRA restriction on preferences allows us to uniquely recover the distribution of preference types in our population (Apesteguia et al. (2017)).

We find that the higher the consideration cost, the higher the proportion of individuals who suboptimize. Our results indicate that when the consideration cost is high, 24.07 percent of the population may be suboptimal because of lack of full consideration. The fractions of suboptimizing

---

differs from theirs is that they focus on the case of stochastic choice with attributes.

[14] Other authors such as Honka et al. (2017), and Draganska and Klapper (2011) use additional surveys to identify limited consideration. In order to identify consideration set models Kawaguchi et al. (2016), and Conlon and Mortimer (2013) exploit variation in product availability, while Dehmamy and Otter (2014) and Huang and Bronnenberg (2018) exploit variations in quantity purchased and products purchased. Gabaix et al. (2006) exploits mouse-tracking data to identify the pieces of information or attributes that a DM considers that will make her consider a particular alternative.

[15] Roberts and Lattin (1997) and Van Nierop et al. (2010) provide summaries of the marketing literature on consideration set models.

individuals for the medium and low costs are 17.34 percent and 3.76 percent, respectively. These findings reveal substantial welfare implications of hard choices. Our estimates of the distribution of preferences imply that there are two major preference types in our population if we condition on the recovered stochastic consideration rule. This finding supports the presence of heterogeneity in preferences. In addition, we cannot reject the possibility that individuals in our sample are risk-averse expected utility maximizers conditional on limited consideration; if this were true, then that would support the model by Gul et al. (2014), as long as that model is augmented with limited consideration.[16]

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 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. Section 5 presents the structural estimation of the random consideration rules as well as the distribution of preferences and the main implications of hard choice for welfare. Finally, Section 6 concludes.

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 Consideration Sets and Preference Heterogeneity

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

[16]This observation supports the findings in List and Haigh (2005) that violations of expected utility are less prevalent at the population level.

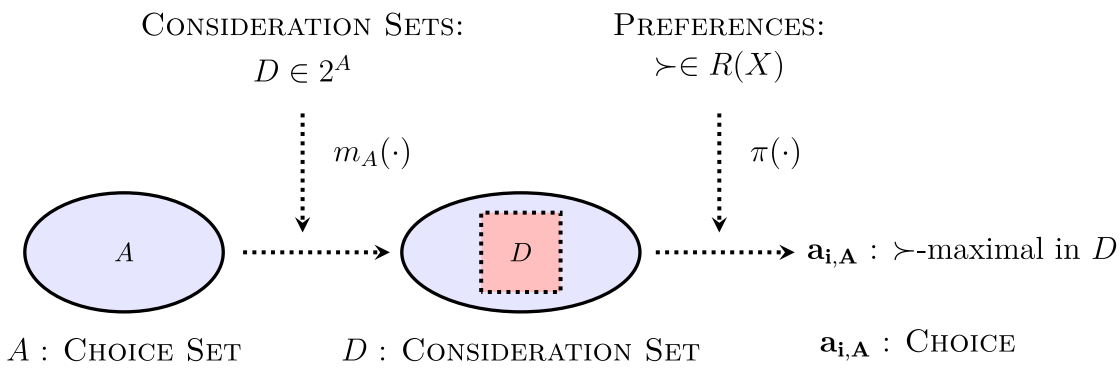


Figure 1 – Consideration mediated choices Choices are the result of a two stage process, first pick a consideration set and then pick the best alternative in that set. We observe choices ($\mathbf{a_i}$) and menus (A). We do not observe and we aim to identify the theory objects: distribution of preferences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

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 $m_A : 2^A \to [0,1]$, $\sum_{D \subseteq A} m_A(D) = 1$. 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. (Heterogeneous Preferences Random Consideration Set Rule, HRC-rule) A complete stochastic choice rule P is a HRC-rule if there exists a pair (m, π) such that

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

for all $a \in X$ and $A \in \mathcal{A}$, where $\mathbb{1}(\,\cdot\,)$ denotes the indicator function.[17]

This choice rule is illustrated in Figure 1. We can specialize the HRC-rule for different restrictions on the heterogeneity of preferences by replacing the set of all linear orders $R(X)$ in Definition 1 by $R^c(X) \subseteq R(X)$ that satisfies property c.[18]

Definition 1 implicitly assumes that the random consideration set rule and the heterogeneous preferences are independent. Independence is a very natural restriction in this environment 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. Also, we are interested in modelling decision making in two-stages where DMs simplify a hard choice task by means of fast-and-frugal

[17] $\mathbb{1}(\,B\,)$ is equal to 1 if the statement B is true, otherwise it is equal to zero.

[18] In what follows, we will allow for several restrictions on c: (i) linear orders must satisfy the independence axiom over lotteries. We call this restriction Expected Utility (EU) ($c = \text{EU}$). For short the notation for these restricted models is c-L-HRC. (ii) We fix a single linear order or homogeneous preferences ($c = \text{HM}$). The EU restricted set of preference orders $R^{\text{EU}}(X)$ only contains linear orders that can be represented by an expected utility, namely there exists a collection $\{u_z\}_{z \in Z}$, which represents utilities over the set of prizes Z, such that each lottery $l \in \Delta(Z)$ has a utility $\text{EU}(l) = \sum_{z \in Z} l(z) u_z$. See Appendix B.

heuristics (consideration) that are independent from 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. Moreover, as the following lemma demonstrates, the HRC-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* HRC*-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. We say that a consideration set rule m admits an attention-index representation if there exists a *link* function $\psi(\cdot) = (\psi_{A,D}(\cdot))_{A\in\mathcal{A},D\subseteq A}$ such that

$$m_A(D) = \psi_{A,D}(\eta),$$

for all $A \in \mathcal{A}$ and $D \subseteq A$. The collection of all consideration rules induced by a given link function ψ^{L} (indexed by L) is denoted by $\mathcal{M}^{\mathrm{L}}$.

In what follows we assume that the link function $\psi(\cdot)$ is 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. 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 random categorization model (RCG) of Aguiar (2017). 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 (RCG), and the full consideration (FC) models admit an attention-index representation by the link function ψ^{L}, $\mathrm{L} \in \{\mathrm{LA}, \mathrm{MM}, \mathrm{RCG}, \mathrm{FC}\}$ such that

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

$$m_A(D) = \psi^{\mathrm{LA}}_{A,D}(\eta) = \frac{\eta(D)}{\sum_{C \subseteq A} \eta(C)} > 0,$$

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

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

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

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

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

$$m_A(D) = \psi^{\mathrm{RCG}}_{A,D}(\eta) = \sum_{C : C \cap A = D} \eta(C),$$

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

- $m \in \mathcal{M}^{\mathrm{FC}}$ if and only if $m \in \mathcal{M}^{\mathrm{RCG}}$ 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. (An example of a consideration rule that admits an attention-index representation is the *Elimination by Aspects* (Tversky (1972)). Our general result covers this case that so far has not been fully characterized in the literature.) In particular, the well-known MM-rule is exactly the intersection of the LA and RCG models (Suleymanov (2018)). Thus, LA and RCG 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 RCG rules generalize this independence assumption in order 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 RCG 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.

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. 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-HRC-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 RCG restriction. Namely, let $\eta(X) = \frac{1}{2}$ and $\eta(\{b\}) = \frac{1}{2}$. Then the RCG-HRC-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 $K : [0, 1] \to \mathbb{R} \cup \{+\infty\}$ 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.5 for details).

Given the definition of the attention-index representation we can define a restricted version of the HRC-rule (L-HRC-rule).

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

Definition 4 summarizes all models that have the attention-index representation. In our application we focus in four different models ($\mathrm{L} \in \{\mathrm{LA}, \mathrm{MM}, \mathrm{RCG}, \mathrm{FC}\}$). However, our approach extends to any model of consideration that admits an attention-index representation.

2.3. Characterization and Identification of the L-HRC-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 HRC rule that generated it, and provide necessary and sufficient conditions that guarantee that a dataset P can be generated by a L-HRC 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 $\psi_{\emptyset}(\eta) = (\psi_{A,\emptyset}(\eta))_{A\in\mathcal{A}}$ we can write the equation:

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

When $\psi_{\emptyset}$ 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. A consideration rule m admitting an attention-index representation is *totally monotone* if, we can write, for all $A \in \mathcal{A}$:

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

where $\varphi : [0, 1] \times [0, 1] \to [0, 1]$ is a strictly monotone link function in each entry or coordinate.

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).

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. (2017), 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 $m_{(\cdot)}(\emptyset) : \mathcal{A} \to [0,1]$ 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 RCG it will satisfy a form of marginal decreasing propensity of choice (Aguiar (2017)). Informally, 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 $\Delta_D(\Delta_C p(o, A) \leq 0)$).[19] Alternatively, φ can be strictly decreasing in the first entry in which case it provides an antithetic behavior to that of the RAM. In this sense, this restriction on attention is neither weaker not stronger than the monotonicity restriction in Cattaneo et al. (2017).

Observe that strict monotonicity of φ in each of its entries implies the invertibility of ψ.[20] 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* $\mathrm{L} \in \{\mathrm{LA}, \mathrm{MM}, \mathrm{RCG}, \mathrm{FC}\}$ *admits a totally monotone attention-index representation with:*

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

The proof is omitted because of its simplicity for the cases of LA, RCG and FC. For the case of MM the statement follows from Brady and Rehbeck (2016) and Aguiar (2017).[21]

2.4. Characterization of the L-HRC-model

As a preliminary step for characterizing the L-HRC-model, we construct a candidate calibrated attention-index η^{L} from the data P.[22] Informally, this calibrated (revealed) attention-index is

[19] Note, 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.

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

[21] Note that since MM and FC are special cases of LA and RCG 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.

[22] The calibrated index η^{L} will not be a well-defined distribution when the model is misspecified.

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 $\varphi_1^{-1,\mathrm{L}}$ and and $\varphi_2^{-1,\mathrm{L}}$ 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, $\eta^{\mathrm{L}} : 2^X \to \mathbb{R}$ is such that (i) $\eta^{\mathrm{L}}(\emptyset) = \varphi_2^{-1,\mathrm{L}}(1, p(o, X))$, and (ii) for all $A \in 2^X \setminus X$

$$\eta^{\mathrm{L}}(A) = \sum_{B \subseteq D} (-1)^{|D \setminus B|} \varphi_1^{-1,\mathrm{L}}(p(o, D), \eta^{\mathrm{L}}(\emptyset)).$$

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 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 correct specification of the model.

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

$$m_A^{\mathrm{L}}(D) = \psi_{A,D}^{\mathrm{L}}(\eta^{\mathrm{L}}).$$

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

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

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

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 $\sum_{D \subseteq A} m_A^{\mathrm{L}}(D) = 1$ for all $A \in \mathcal{A}$. Thus, m^{L} is a collection of distributions over consideration sets if and only if $m^{\mathrm{L}} \geq 0$. 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 $m^{\mathrm{L}}_A(D) \in \Delta(2^A)$ for all $A \in \mathcal{A}$ and all $D \in 2^A$.

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

- $m^{\mathrm{LA}}_A(D) > 0$ for all $D \in 2^A$;
- $\gamma^{\mathrm{MM}}(a) \in (0,1)$ for all $a \in X$;
- $m^{\mathrm{RCG}}_A(D) \geq 0$ for all $A \in 2^A$;
- $m^{\mathrm{FC}}_A(D) = \mathbb{1}\left(A = D \right)$.

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-HRC-*rule;*

(ii) m^{L} *is a well-defined collection of distributions over consideration sets such that* P *is a* HRC-*rule described by* (m, π) *with* $m = m^{\mathrm{L}}$.

Theorem 1 provides a testable implication of the model. If P is a L-HRC-rule, then m^{L} has to be well-defined. Theorem 1 implies that in order 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 $P^{\mathrm{L}}_\pi = (p^{\mathrm{L}}_\pi(a, A))_{A \in \mathcal{A}, a \in X}$, where $p^{\mathrm{L}}_\pi : X \times \mathcal{A} \to \mathbb{R}$ is such that, for all $A \in \mathcal{A}$ and $a \in A$,

$$p^{\mathrm{L}}_\pi(a, A) = \frac{p(a, A) - \sum_{C \subset A} m^{\mathrm{L}}_A(C) p^{\mathrm{L}}_\pi(a, C)}{m^{\mathrm{L}}_A(A)}.$$

We call P^{L}_π *the calibrated full consideration rule.*

Notice that when P has been generated by a LHRC, P^{L}_{π} corresponds to the underlying full-consideration random utility rule. In fact, we can write a LHRC model equivalently as:

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

where $p_\pi(a, A) = \sum_{\succ \in R(X)} \pi(\succ) \mathbb{1}\left(a \succ b,\ \forall b \in A \right)$ 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 LHRC rule, it follows that $p^{\mathrm{L}}_{\pi} = p_\pi$.

Similar to m^{L}, P^{L}_{π} has interpretation when the L-HRC-rule is consistent with the data. 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) $m^{\mathrm{L}}_A(A) > 0$ for all $A \in \mathcal{A}$. Then the following are equivalent:*

(i) P is a L-HRC-*rule;*

(ii) P^{L}_{π} is a FC-HRC-*rule.*

Note that both $m^{\mathrm{L}} \geq 0$ and P^{L}_{π} can be computed from P. Thus, Theorem 2 implies that in order to test a given model L it is necessary and sufficient to test whether m^{L} is well defined, and whether calibrated P^{L}_{π} 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-HRC-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-HRC-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-HRC-*rule, and (ii) $m^{\mathrm{L}}_A(A) > 0$ 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 and preference components in a unique way. Appendix C.3 shows that 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.[23]

We can add restrictions on the set of linear orders $R(X)$ to uniquely identify the distribution of preferences. In particular, we can follow Apesteguia et al. (2017) and replace $R(X)$ in Definition 1 by the set of linear orders that satisfy a single crossing condition.[24] In our experiment we consider lotteries, therefore it is natural to impose restrictions for Expected Utility rationalization. If we restrict preferences to be governed by a constant absolute risk aversion (CARA) or constant relative risk aversion (CRRA) Expected Utility model with full heterogeneity in the risk parameter, then the restricted set of linear orders satisfies single crossing (Apesteguia et al. (2017)). This is an important special case that we study in Section 5.

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

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

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

The relationship between the models in Definition 3 and RUM are summarized in Figure 2. 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, RCG is nested in RUM and nests MM, therefore its intersection with LA is nonempty (Aguiar (2017)). Suleymanov (2018) shows that MM exactly describes the intersection between LA and RCG.[25] Table 1 summarizes some well documented

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

[24]The single crossing condition is a restriction on the support of (strict) preferences that requires them to be ordered, such that any pairwise comparison only switches once (Apesteguia et al. (2017)).

[25]All these relationships are preserved when allowing for heterogeneous preferences under the independence

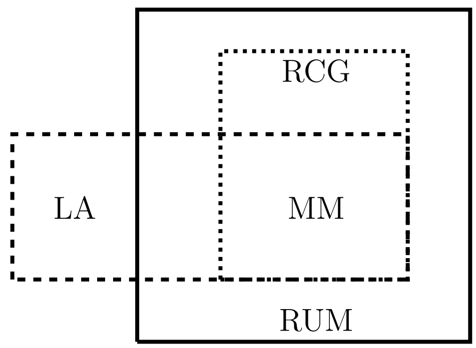


Figure 2 – Relation among Consideration Set Rules: RUM, LA, MM, and RCG

Table 1 – Behavior that cannot be rationalized by LA, RCG, MM and RUM

Allows for:	RUM	MM	LA	RCG
Attraction Effect	✗	✗	✓	✗
Choice overload	✗	✗	✗	✗

features of stochastic choice and whether the consideration rules in Definition 3 can rationalize them.[26]

2.7. Testing Procedure

Theorem 2 allows us to test whether a given stochastic choice rule P is a L-HRC-rule: it is necessary and sufficient to test whether m^{L} is well-defined (satisfies a set of linear inequalities) and P^{L}_{π} 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) allow 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^{L}_{π}, is a vector of length $d_p = \sum_{k=1}^{|X|} k\binom{|X|}{k}$.[27] The k-th element of P^{L}_{π} corresponds to some pair (a, A) such that $a \in A \cup \{o\}$.

assumption of preferences and attention. The reason is that the outside probability does not depend on the distribution of preferences.

[26] For more details on the comparison among stochastic consideration set models, some other models in the literature, and their empirical implications see Appendix C.

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

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

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

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-HRC-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-HRC-*rule and* m^{L} *is well-defined;*

(ii) $\inf_{v \in \mathbb{R}_+^d} \left\| g^{\mathrm{L}} - Gv \right\| = 0$, *where* $g^{\mathrm{L}} = (P_\pi^{\mathrm{L}\prime}, m^{\mathrm{L}\prime})'$.

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

Theorem 4 implies that we can test the null hypothesis that $\inf_{v \in \mathbb{R}_+^d} \left\| g^{\mathrm{L}} - Gv \right\| = 0$. 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 $\boldsymbol{a}_{i,A}$, $i = 1, \ldots, n_A$ 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 $\hat{p}(a, A) = n_A^{-1} \sum_{i=1}^{n_A} \mathbb{1}\left(\boldsymbol{a}_i = a \right)$ 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}, $\hat{m}^{\mathrm{L}}$ and $\hat{P}_\pi^{\mathrm{L}}$, using Definitions 7 and 10. Given the results of Theorem 4, a natural test statistic is

$$\mathrm{T}_n = n \min_{v \in \mathbb{R}_+^d} (\hat{g}^{\mathrm{L}} - Gv)' \hat{\Omega}^{-} (\hat{g}^{\mathrm{L}} - Gv),$$

where $n = \sum_A n_A$ is the sample size; $\hat{g}^{\mathrm{L}} = (\hat{P}_\pi^{\mathrm{L}\prime}, \hat{m}^{\mathrm{L}\prime})'$; $\hat{\Omega}^{-}$ is a generalized inverse of a diagonal matrix $\hat{\Omega}$ such that the i-th diagonal element $\hat{\Omega}_{i,i}$ is a consistent estimator of the asymptotic variance of the i-th component of $\hat{g}^{\mathrm{L}}$.

Let $\hat{g}_l^{\mathrm{L},*}$, $l = 1, \ldots, L$, be bootstrap replications of $\hat{g}^{\mathrm{L}}$. Let $\tau_n \geq 0$ be a tuning parameter and ι

be a vector of ones of dimension d.[28] To compute the critical values of T_n we follow the bootstrap procedure proposed in Kitamura and Stoye (2018):

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

$$n \min_{[v-\tau_n \iota/d]\in\mathbb{R}^d_+} (\hat{g}^{\mathrm{L}} - Gv)'\hat{\Omega}^{-}(\hat{g}^{\mathrm{L}} - Gv);$$

(ii) Compute

$$\hat{g}_l^{\mathrm{L},*} = \hat{g}_l^{\mathrm{L},*} - \hat{g}^{\mathrm{L}} + \hat{\eta}_{\tau_n},$$

for every $l = 1, \dots, L$ and Ω^*;

(iii) Compute the bootstrap test statistic

$$\mathrm{T}^*_{n,l} = n \min_{[v-\tau_n \iota/d]\in\mathbb{R}^d_+} (\hat{g}_l^{\mathrm{L},*} - Gv)'\hat{\Omega}^{*-}(\hat{g}_l^{\mathrm{L},*} - Gv), \quad l = 1, \dots, L;$$

(iv) 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 of the L-HRC-rule if $\mathrm{T}_n > \hat{c}_{1-\alpha}$", where $\hat{c}_{1-\alpha}$ is an $(1-\alpha)$-quantile of the empirical distribution of the bootstrap statistic. Note that if the hull hypothesis is wrong and the asymptotic variance of $\hat{g}^{\mathrm{L}}$ is bounded from above and is bounded away from zero, then the test statistic diverges to infinity as the sample size grows. So asymptotically we will reject the wrongly specified null hypothesis with probability approaching 1.

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 need to perform the test for a subgroup of population.

3. The Experiment

Our testing approach requires full choice set variation but does not have requirements in terms of repeated individual choice data. Exploiting this feature, our experiment was designed to study the performance of different theories of random consideration sets with few observations per individual. In particular, we conducted the experiment in Amazon MTurk for a large cross-section with at most two (disjoint) choice sets per individual (see Section 3.1).

[28]In 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.

All sessions were run between August 25, 2018 and September 17, 2018 on the MTurk platform with surveys designed in Qualtrics.[29] 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.[30] The average duration of the session was 251.68 seconds (slightly over 4 minutes).[31] 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 was 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

The experiment is designed to obtain a dataset with exogenous variations in choice sets and in the cost of consideration. In addition, our design produces a complete stochastic choice dataset with full choice set variation. To induce (potentially) preference heterogeneity we consider lottery alternatives with different expected value and variance. Table 2 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)).[32]

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

[29] By 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 USA were recruited.

[30] All payment were made in USD.

[31] The average duration of each task is 23 seconds, and it is significantly correlated with the length of the choice set (correlation=0.1677, p-value$< 10^{-4}$) and the cost (correlation=0.2581, p-value$< 10^{-4}$).

[32] We provide evidence that in our experimental sample DMs have risk aversion levels such that the default is dominated by all other alternatives. Additionally, it can be argued that, without cost of consideration treatments, the outside alternative is easier to understand than the rest because of its simplicity. Hence, it works as a reference point (Suleymanov (2018)).

Table 2 – 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

is shown first, while the order of other alternatives is randomized.[33] 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-HRC 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 cost treatments: High (H), Medium (M), and Low (L). These costs of treatment were induced by introducing a k-length two digit addition/subtraction to compute each prize in the lottery. The length k was set equal to 5, 3 and 1, for the high, the medium, and the low cost, respectively. 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.

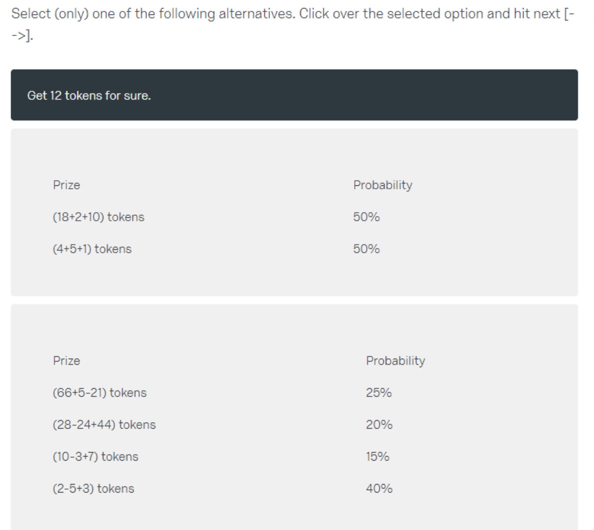
To prevent possible learning that could attenuate consideration costs, subjects were faced with disjoint choice sets. That is, subjects were either presented with the full choice set and the outside option ($X\cup\{o\}$); or a partition of X (presented at random order), i.e. $A_j\cup\{o\}, A_k\cup\{o\}$ with $A_j\cup A_k=X$ and $A_j\cap A_k=\emptyset$. Our experimental design is summarized in Figure 4.[34]

The default alternative For any choice set/treatment cost 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.

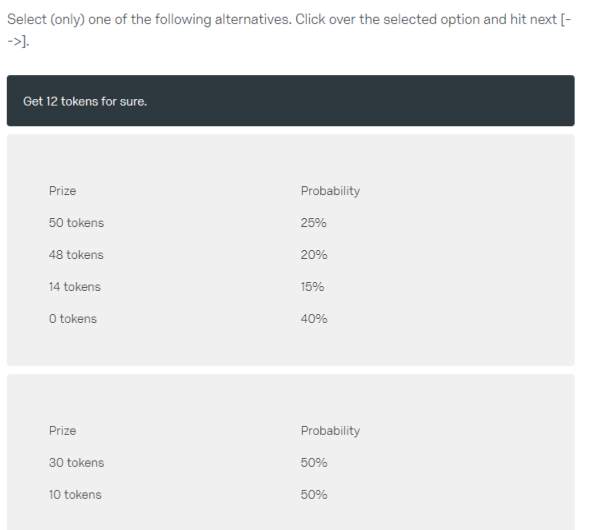
Structure of the Lotteries. Since alternatives in our experiment are lotteries, a natural theory to describe behavior under full consideration is expected utility. Here we show the special structure

[33] Without this randomization the order of the choices in menus may have effect on agents decisions.
[34] This is done to increase the number of observations given the fixed cost of the participation fee.

(a) High Cost



(b) Medium Cost



(c) Low Cost

Figure 3 – Consideration Cost Treatments Different induced costs for Choice set $\{o, l_2, l_4\}$

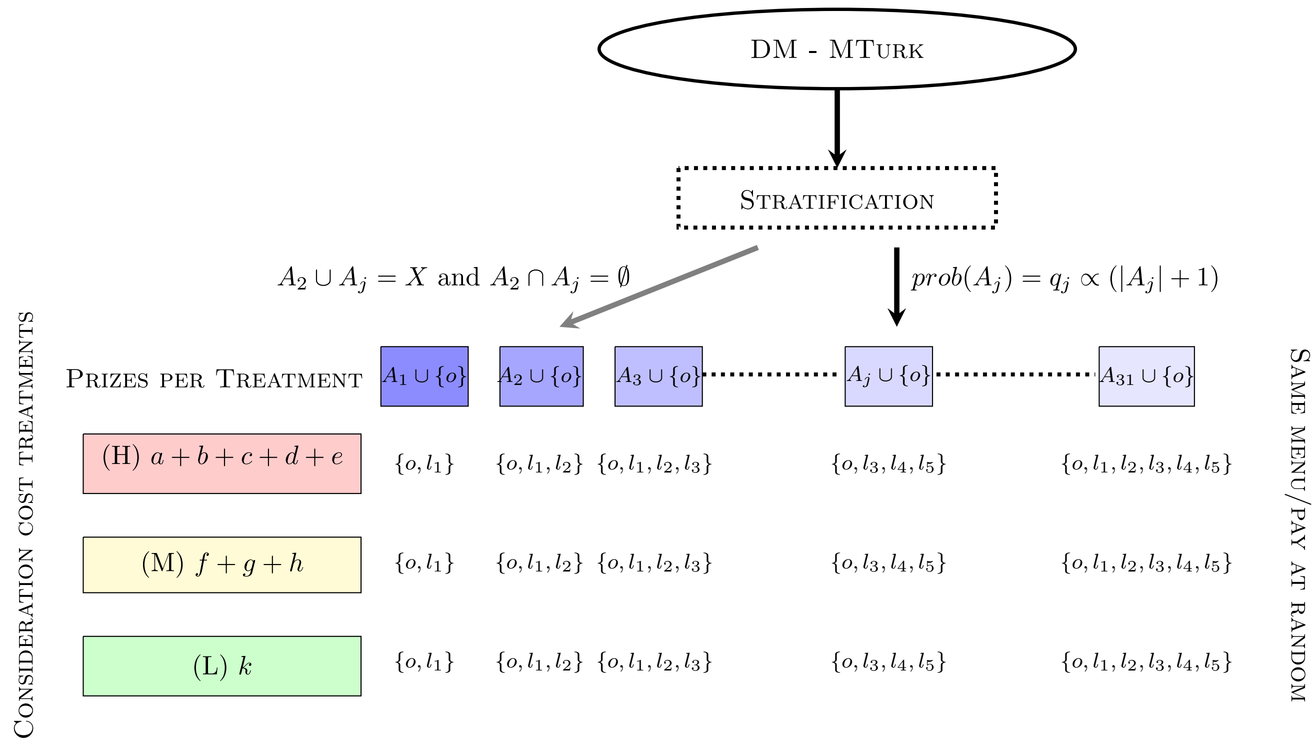


Figure 4 – Experimental Design: DM i draws with probability $p(A_j)$ menu A_j with $|A_j| \in \{3, 4, 5\}$. In the picture, $A_j = \{l_3, l_4, l_5\}$. Therefore she is asked to choose from menus $A_j \cup \{o\}$ and $A_2 \cup \{o\}$, since $A_2 \cup A_j = X$ and $A_2 \cap A_j = \emptyset$. She is faced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

of our alternatives that allows us to test expected utility (and the independence axiom) as a restriction on the set of linear orders (i.e, we use $R^c(X)$ with $c = \text{EU}$ in Definition 1).

Lotteries are defined on $\Delta(Z)$ with Z being the set of prizes. In our experiment $Z = (50, 48, 30, 14, 12, 10, 0)$.[35] Independence implies that, for any $a, b, c \in \Delta(Z)$, and any $\alpha \in (0, 1)$

$$a \succ b \iff \alpha a + (1-\alpha)c \succ \alpha b + (1-\alpha)c\,.$$

To understand additional restrictions that are imposed by independence note that lotteries in our experiment can be written as $l_1 = (1/2, 0, 0, 0, 0, 0, 1/2)$ and $l_2 = (0, 0, 1/2, 0, 0, 1/2, 0)$. By defining the auxiliary lottery $a = (0, 2/5, 0, 3/10, 0, 0, 3/10)$, the following relations can be established:

$$l_3 = \frac{1}{2}l_1 + \frac{1}{2}l_2, \quad l_4 = \frac{1}{2}l_1 + \frac{1}{2}a, \quad l_5 = \frac{1}{2}l_2 + \frac{1}{2}a\,.$$

This structure restricts the possible orders that are compatible with expected utility: (i) if $l_1 \succ l_2$ then $l_1 \succ l_3$, $l_3 \succ l_2$ and $l_4 \succ l_5$; or (ii) if $l_2 \succ l_1$ then $l_2 \succ l_3$, $l_3 \succ l_1$ and $l_5 \succ l_4$. That is, this assumption restricts the support for $\pi \in \Delta(R(X))$, therefore reducing the dimension of the testing

[35] For L-HRC we do not require preferences to be defined over the outside option, however we also consider RUM where o is considered as one alternative more and therefore we define Z accordingly.

problem. As a result, by imposing independence we decrease the set of preferences that can generate the data.

However, the previous restrictions are only implications of the expected utility assumption. The necessary and sufficient condition is the existence of a Bernoulli utility vector $u = (u_z)_{z \in Z} \in \mathbb{R}^{|Z|}$ that represents the preferences over the restriction $\succ \in R^{\mathrm{EU}}(X)$ (i.e., $l_1 \succ l_2 \iff \sum_{z\in Z} l_{1,z} u_z > \sum_{z \in Z} l_{2,z} u_z$). The necessary and sufficient conditions for deciding whether a linear order is compatible with EU (i.e., it can be extended to the simplex $\Delta(Z)$ as an expected utility order) with a finite set of prizes and lotteries are provided (as a linear programming problem) in Appendix B. Therefore, imposing the necessary and sufficient conditions restricts the total number of linear orders on X consistent with expected utility from 120 to 10.

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 11. 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 as treatments).

3.3. Descriptive Analysis: Evidence for Costly Consideration

In this section we describe behavior of individuals in our sample and show that our cost treatments effectively induce costly consideration in a reduced form analysis.[36] In particular, we observe that the consideration cost treatments: (i) have a significant effect on the choice frequency of the outside option; (ii) have a heterogeneous effect on the choice frequencies of all other alternatives; (iii) affect the dynamics of choice with respect to the size of the menu; and (iv) have a significant effect on the time dedicated to each task. Moreover, all these effects depend monotonically on the level of difficulty of choice induced by each treatment.

Under the null hypothesis of full consideration (ineffective consideration cost), the observed frequency of choice of the outside option should remain the same. 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 and Table 14). This is evidence in favor of limited consideration. In addition, we observe that the choice frequency of the outside

[36] We use here a linear probability regression framework that delivers a simple correlation analysis.

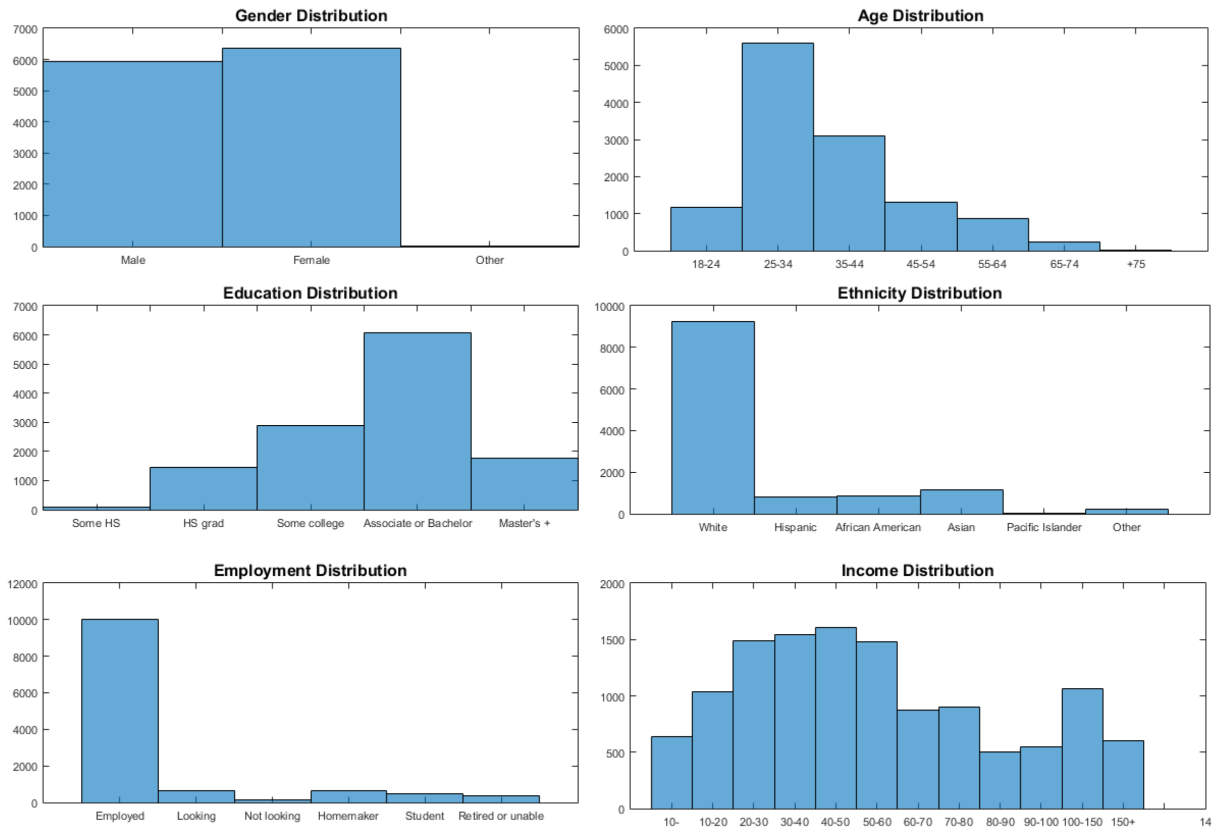


Figure 5 – Distribution of demographics in sample

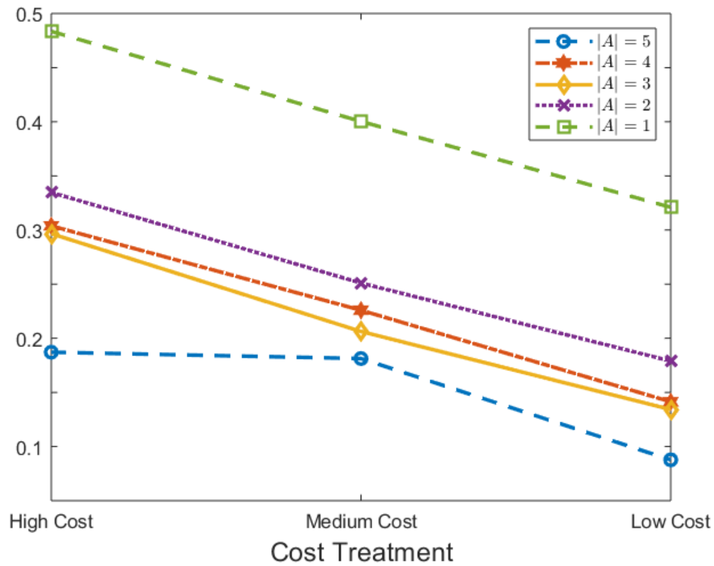


Figure 6 – Estimated frequency of the Outside Option: Average by menu size.

option decreases with the menu size, providing evidence against choice overload. We remind the reader that choice overload can not be generated by L-HRC.

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. 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 seems to be a new stylized fact that we document here, but leave its further analysis for future work.

Table 12 shows that the net effect of the menu size on the choice frequency of the outside option is positive for the medium and high costs; when it is very costly to compute the lotteries is more likely that DMs decide not to consider any alternative and choose the outside option.

Figure 7 shows the effect of the treatments on the choice frequencies while Tables 13-14 in Appendix D show the correlation between choices, consideration cost treatments, size of the menu and the order of the alternatives within the menu.

The harder it is to understand the lotteries,[37] 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 alternative 1 increases with the cost treatment; the cost does not have a significant impact on the probability of selecting lottery 4; while it has a negative impact on the other alternatives.

[37]Here 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.

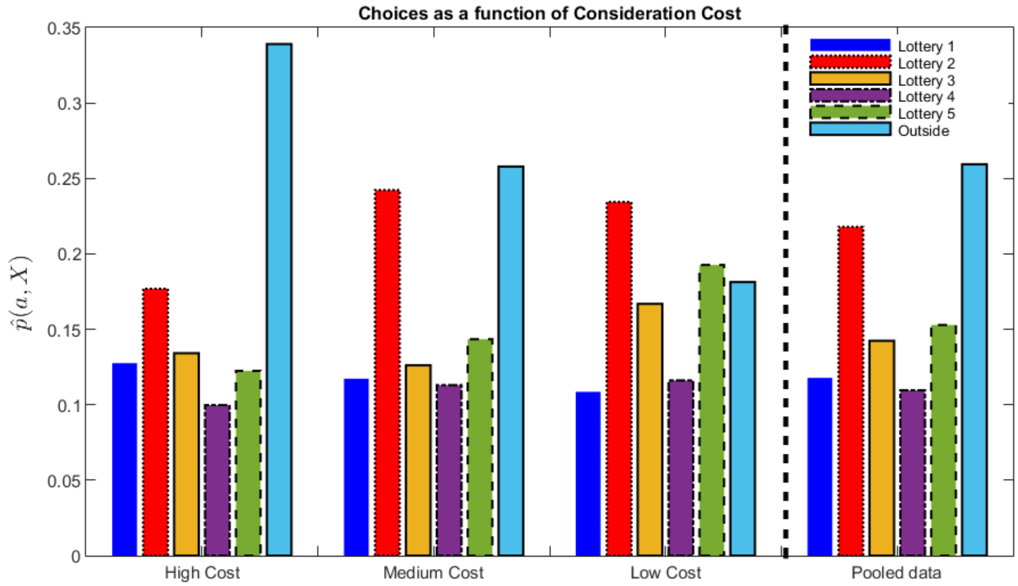


Figure 7 – Choices and Consideration Cost

Overall, the probability of selecting any given lottery decreases with the size of the menu, suggesting that regularity is satisfied in most of our sample, as shown in Figure 8. However, this relation is nonmonotonic for the empirical choice frequency for the pooled data, in particular around menus of size 4 suggesting a violation of regularity. This nonmonotonicity appears also for the medium and high cost treatments, as shown in Figure 12. In Appendix E we provide an extended analysis of the evidence of choice overload and attraction effect.[38]

Finally, we find a significant effect of the cost treatment on the time dedicated to complete tasks. The time it takes to complete a task (choice set/cost) may be related to the size of the choice set and cost of consideration. These correlations are shown in Table 15. The sensitivity of the individual decision times to our treatments is evidence of their effectiveness.[39]

4. Testing Random Consideration Models

In this section we report the results of testing the validity of RUM and L-HRC to describe our experimental data. In particular, we focus on the EU restriction on the preference support

[38] We find evidence that regularity fails only for the medium case using our formal testing procedure. We also find that regularity fails in the pooled data.

[39] We also find that the frequency of choice of an alternative that is different from the outside option increases when the alternative is seen immediately after the outside option. The higher the consideration cost is, the stronger the order effect becomes. This result is consistent with a story of limited consideration where the DM may just consider the first few alternatives she sees in the menu. Search and satisficing can be seen as a special case of random consideration when the threshold is random (Aguiar et al. (2016)). The study of order effects is left for future work.

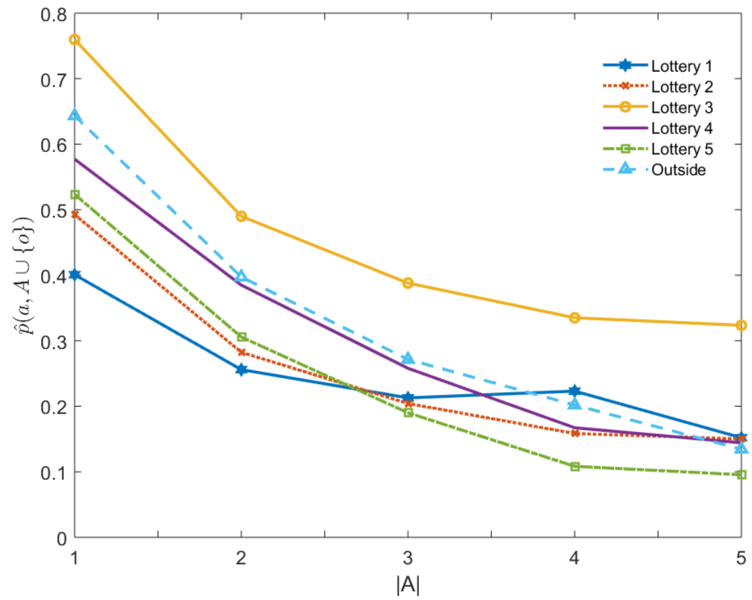


Figure 8 – Frequencies of choice in pooled data as a function of menu size

$(R^{\mathrm{EU}}(X))$. Unless otherwise stated the tested hypothesis is that, for a particular specification of our model (consideration set stochastic rule/cost treatment), there exists (m, π) that is a L-HRC representation for behavior. We report the value of the test statistic and the corresponding p-value coming from the bootstrap distribution for the test statistic (500 bootstrap replications for every test statistic were conducted). The 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 L-HRC representation.

4.1. Results per Consideration Cost Treatment

We have run a (statistical) survival race among competing models of behavior. We test for the validity of RUM, EU-RUM, EU-LA, EU-RCG, and EU-MM. We test the special case of RUM with the EU restriction because our choice sets are composed by lotteries. As we explained before, this restriction is a test of Random Expected Utility Model by Gul et al. (2014) in the case with full consideration. In the rest of this section we will refer to the L-HRC models without EU for short.

We use the relation between the different models (see Figure 2) to learn about their empirical ability to describe the behavior of our experimental population. First, we use the fact that RUM and LA are the biggest models that are not nested by any other model of interest. Our first finding is that LA cannot be rejected at all standard confidence levels for the three costs of consideration (see Table 3). In contrast, RUM is rejected by the intermediate cost of consideration at the 90 percent confidence level (see Table 4). Our main conclusion is that the standard RUM cannot describe the behavior of the population for all costs, but LA is successful at this task. We underline

Table 3 – RESULTS FOR EU-L-HRC - LIMITED CONSIDERATION Summary of the testing results for L-HRC with $L \in \{LA, MM, RCG\}$ when DMs are assumed to be expected utility maximizers within their consideration sets.

		CONSIDERATION SET RULE					
		LA		MM		RCG	
SAMPLE	N	TEST-STAT.	P-VALUE	TEST-STAT.	P-VALUE	TEST-STAT.	P-VALUE
LOW COST	4099	105305	0.4117	63302	0.0220	45412	0.0235
MEDIUM COST	4099	301213	0.4813	-	-	-	-
HIGH COST	4099	78699	0.7006	59175	0.2740	24090	0.4186

Table 4 – RESULTS FOR RUM: Summary of the testing results for RUM and EU-RUM

		RUM		EU-RUM	
SAMPLE	N	TEST-STATISTIC	P-VALUE	TEST-STATISTIC	P-VALUE
LOW COST	4099	271.07	0.5326	2211.95	<0.002
MEDIUM COST	4099	401.74	0.0919	1292.72	<0.002
HIGH COST	4099	225.97	0.8966	1567.60	<0.002

the fact that a population of DMs with full consideration (and arbitrary distribution of preferences) would have been described at all cost of consideration by RUM. This is not the case. Hence, this is evidence in favor of the presence of limited consideration at the population level. Moreover, when we restrict our attention to EU-RUM then all three costs of consideration cannot be described by the model proposed in Gul et al. (2014). Nonetheless, we also find that both RUM and LA cannot be rejected for the high and low costs of consideration.

Given these results, we turn to testing the two additional models of interest RCG and MM, which are subsets of the models above. The main purpose of this exercise is to further understand the population behavior for these cases.[40] (For the intermediate case, given that RCG and MM are subsets of RUM, we must also reject them.) For the high and low costs of consideration, we have strikingly different results. Whereas the RCG and MM models are rejected at all standard

[40]If the population of DMs had full consideration in the high and low cost, given that the choice sets faced by DMs across different cost levels are the same, and payment is at random, it must be that the same distribution of preferences must describe the two cost levels. This is not the case, as we will see next, providing further evidence against full consideration happening for both the high and low cost.

confidence levels for the low cost of consideration, the same two models cannot be rejected for the high cost case. The latter result implies that the highly stylized MM model can describe well the behavior of the population for the high cost, and perhaps surprisingly, we cannot reject the hypothesis that MM is the only rule of consideration that describes the population behavior within the class of LA and RCG rules. This follows from the fact that the MM model is exactly the intersection of the LA and RCG models. This implies that if the true data generating process for this population behavior is consistent with MM, then no other LA or RCG model can describe the same data.

For the low cost case, the behavior of the population is described well by both LA and RUM, but RCG (and hence MM) fail to do so.[41]

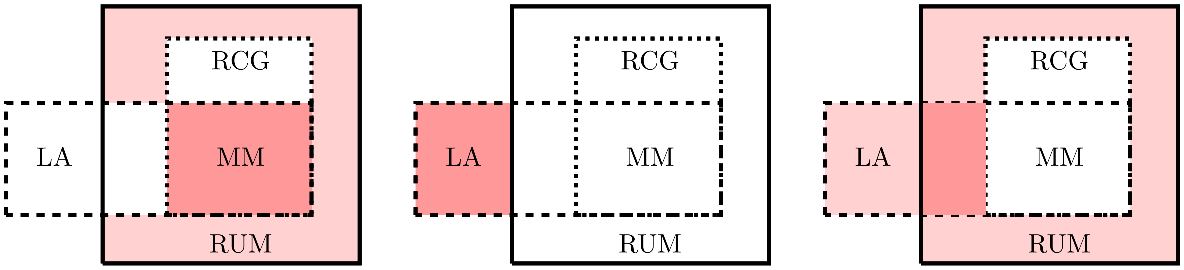
Taken together, these findings imply that even when RUM can describe well behavior in the high and low costs, it is not identifying a stable distribution of preferences that are consistent with full consideration. Moreover, we find that this instability in the distribution of preferences is compatible with limited consideration with a LA specification.

Our results provide the first empirical description of how different models of limited consideration explain the population behavior for different levels of the cost of consideration. These results are summarized in Figure 9.

Evidence for Limited Consideration First, the finding that full consideration and (stable across costs) RUM fail to describe the population behavior suggests that DMs may be suboptimizing due to limited consideration. Nevertheless, even when there is evidence in favor of limited consideration, DMs may look consistent with RUM for fixed cost level. In fact, many fast and frugal heuristics and limited consideration behaviors are consistent with this finding. A prominent example is MM, a special case of RUM that successfully describes the high cost case. In other words, without the presence of varying consideration costs with a fixed choice set and payment at random, it would be impossible to differentiate between RUM and L-HRC.

Second, we test the LA, MM, and RCG models with the EU restriction. Table 3 presents the formal results. We find that LA with heterogeneous preferences of the EU type provides a good description for each of the consideration cost treatments, since we cannot reject the hypothesis that behavior is generated by LA-HRC. This is the first evidence in favor of LA as a feasible extension of EU-RUM that describes well the lottery choice behavior of a large population of DMs. The EU-RUM is strongly rejected for all cost cases as seen in Table 4. In this sense, we find that EU is a reasonable description of behavior once we have taken into account limited consideration. This finding has implications for the literature interested in violations of the independence axiom in the context of choice over lotteries. Here we have shown that a limited consideration explanation for observed departures from the EU maximizing benchmark is empirically supported.

[41]This finding suggests that there is a special LA consideration rule that is also RUM that is not one of the other models. This model, is to our knowledge, unknown in the literature and it remains an open question what this model looks like.



(a) High Cost **(b)** Medium Cost **(c)** Low Cost

Figure 9 – Summary of the Testing Results for L-HRC. Light shaded areas correspond to the set of behaviors that are not rejected for each of the tested models. Dark shaded areas indicate the set of behaviors that is consistent with two or more models.

Choice Overload and Attraction Effect in the Aggregate Third, our findings provide evidence against choice overload in our population. The key common feature of LA, RCG, and MM is that they are not compatible with choice overload. Even the most general model of limited consideration yet, the RAM (Cattaneo et al. (2017)), does not allow for choice overload. The reason is simply that limited consideration rules allow the DMs to simplify large choice sets and, hence, they are immune to the choice overload phenomenon. Since we cannot reject the LA model in all cost levels we find no evidence in favor of choice overload.[42] We may see this phenomenon spuriously as an artifact of sample variability. Note that this is a population result and we cannot say anything about the importance of choice overload for individual datasets. However, this finding puts in perspective the relevance of choice overload at the population level.[43]

Fourth, our results are consistent with attraction effect. In particular, we reject RUM for the intermediate cost case. We remind the reader that LA is consistent with attraction effect while RUM, MM, and RCG are not. The only way RUM is rejected and LA is not is that there are violations of regularity, while the probability of choosing the outside option is well-behaved or consistent with regularity. In our setup this is evidence in favor of the attraction effect and against choice overload. We do not find support for the attraction effect at the high and low costs of consideration. This result is new in the literature and leaves as an open question what is the role of the consideration cost in generating anomalies of behavior.

Revealed Consideration Set Rules and Cost of Consideration Finally, our evidence suggests that DMs actively change their consideration rules according to the cost of consideration in a way that is always compatible with LA. However, there is a lot of heterogeneity within costs of consideration. For example we cannot reject that MM (therefore RCG) describes behavior in the high cost treatment. Intuitively, as the cost reduces, DMs may exert more effort and abandon fast and frugal heuristics but still do not fully consider all alternatives in their choice set. Instead, they may follow more reflective consideration rules that can be captured by our L-HRC model. Indeed, at the low cost and the medium cost the RCG and MM models are rejected. This suggests that,

[42] In addition, in our regression analysis of the experimental data we show that the relative frequency of choosing any of the lotteries increases with the size of the menu (Tables 13-14).

[43] See Appendix E.1 for more details on choice overload and attraction effect.

when DMs are able to better understand choices, they stop using simple consideration rules such as categorization or choice set independent consideration. Recall that LA describes well behavior for the low cost case.

A limitation of our experimental design is that at the low cost of consideration we cannot be sure that DMs have full consideration. Even when they do not have to face a cognitive task when facing an alternative, consideration may still be costly. Given our experimental design it is expected that at the low cost the population behaves as if it is consistent with RUM. In this case, we cannot differentiate whether DMs have limited consideration of the LA type or if they behave as if they are maximizing their own preferences with full consideration over the extended choice set (i.e., $X \cup \{o\}$).[44] Both the LA model and RUM do a fine job in describing the low cost behavior. The main difference between these two models is how they rationalize the positive mass of DMs that choose the outside option. RUM places positive probability on the preference types that set o first in their preference order. In contrast, LA places o last, while explaining the positive mass of choice on the outside option by the probability that DMs did not consider anything else.

As a final consideration, our findings support the hypothesis that limited consideration in high cost of consideration environments may be in fact compatible with the i-Independence condition in Manzini and Mariotti (2014). This condition requires that the probability of considering any given item is independent from the choice set. This finding is novel and we believe it informs about the conditions under which it may be better to use the highly tractable MM model.

4.2. Validity of the L-HRC-rule - Stability of Preferences

L-HRC assumes that the distribution of preferences in the population is independent of the consideration rule. We remind the reader that 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-HRC 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 this hypothesis in this section.

Given that LA is the only model we cannot reject as a description of population behavior at every cost level, we first test whether the underlying distribution of preferences at different cost levels is the same: $\pi^{\mathrm{LA}}_{\mathrm{Low}} = \pi^{\mathrm{LA}}_{\mathrm{Medium}} = \pi^{\mathrm{LA}}_{\mathrm{High}}$, were $\pi^{\mathrm{LA}}_{\mathrm{cost}}$ is the distribution over preference implied by m^{LA} at cost $\in \{\mathrm{Low}, \mathrm{Medium}, \mathrm{High}\}$. The results are presented in Table 5. We cannot reject this hypothesis of joint stability of preferences inferred from the LA specification, and its ability to

[44]Future extensions to our experimental design may address this question by studying an outside option that is dominated by all other lotteries in the first-order stochastic sense.

Table 5 – TEST FOR THE CONSISTENCY OF PREFERENCE IDENTIFICATION Joint test for the stability of identified preference distribution π across consideration costs.

MODEL	HYPOTHESIS	P-VALUE
RUM	$\pi^{\text{RUM}}_{\text{Low}} = \pi^{\text{RUM}}_{\text{Medium}} = \pi^{\text{RUM}}_{\text{High}}$	<0.002
LA	$\pi^{\text{LA}}_{\text{Low}} = \pi^{\text{LA}}_{\text{Medium}} = \pi^{\text{LA}}_{\text{High}}$	0.6623

describe the population behavior.[45]

Next, we test a similar hypothesis for RUM. In stark contrast to LA, even though RUM does a good job describing the data for the high and the low cost treatments, it fails to identify a stable underlying preference distribution. For RUM we reject the hypothesis that the same distribution of preferences can rationalize behavior across consideration cost treatments (See Table 5). This is an intuitive result since behavior at the high cost is consistent with MM but behavior at the low cost is not.

Summarizing, LA is not rejected in any of the consideration cost treatments. Moreover, we cannot reject the hypothesis that the underlying distribution of preferences is the same across consideration costs under LA. This finding supports our independence assumption between consideration and preferences and supports the empirical validity of our identification results.

4.3. Evidence from the pooled data: Do results stand if we aggregate across cost treatments?

As an important robustness check, we analyze whether our main findings stand after aggregating behavior across consideration costs. We test whether pooled behavior admits a representation by a L-HRC model or by RUM. This is important since: (i) The evidence against RUM and in favor of LA may not stand after aggregating behavior across cost treatments. After all, in many applications we cannot condition on the consideration cost. (ii) The nonrejection of LA may be an artifact of our sample being finite. The probability of making type-II error may be relatively big even for our sample size. We remind the reader that our experiment is by far the largest of its kind. Testing in the pooled data directly addresses the first concern and partially address the concerns about the sample size (after pooling the data the sample size increases from 4099 observations to 12297 observations).

[45]The test of stable preferences does not require any modification to the Kitamura and Stoye (2018) framework. The data is the stacked vector of all calibrated $\hat{P}^{\text{LA}}_{\pi_{\text{cost}}}$ and the stacked vector of calibrated $m^{\text{LA}}_{\text{cost}}$, both ordered according to the consideration cost. The matrix B is replaced by the horizontally stacked matrix, $[B'B'B']'$; the columns remain the same as in the test for a fixed cost because the preference support is unchanged. The diagonal matrix I_{d_m} of dimension d_m is multiplied by 3.

Table 6 – Results for Pooled Data - 12297 observations

Full Consideration			Limited Consideration		
Model	Test-Statistic	p-value	Model	Test-Statistic	p-value
RUM	468.06	0.0037	LA	681260	0.2590
EU-RUM	4170.17	<0.002	MM	322469	0.0087
			RCG	202249	0.0084

We find that in the pooled data we reject RUM at the 95 percent confidence level while we cannot reject LA. These findings are summarized in Table 6. This confirms our previous results and provides evidence against the concerns described above.

Note that for LA-HRC, rationalization for each consideration cost by the model does not imply that there is a representative agent that is consistent with it in the pooled data. In contrast, if RUM describes well every single cost treatment it must also describe the pooled behavior (see Apesteguia et al. (2016), and Aguiar et al. (2016).). In this sense, the pooled test has high statistical power against LA and less so against RUM. Nevertheless, we find that we reject RUM and we cannot reject LA. This should inform us that even with the total sample of 12297 independent choices there is evidence against the standard model of stochastic rationality and in favor of limited consideration.

4.4. Discussion

Our testing results are robust to DMs mistakes when understanding the lotteries once they decide to pay attention to them, as discussed in Appendix C.4. However, as in any design with a cognitive task, we may not identify the true distribution of preferences. If this is the case, then preference distribution will depend on the cognitive cost. However, we find that we cannot reject the null hypothesis that LA has a stable distribution of preferences across consideration cost treatments. Therefore we find no evidence for systematic misperception.

We show that limited consideration behavior changes across consideration costs providing evidence that limited consideration may be the mechanism behind choosing the outside option. The consideration models we study are also robust to the outside option playing the role of a reference point (see the work of Suleymanov (2018)).

Freeman et al. (2018) provides an alternative mechanism for the selection of a riskless lottery over dominant risky choices from pairwise comparisons, when binary choice sets are presented as lists. Freeman et al. (2018) proposes 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 are 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, our experimental design subjects are not required to choose from lists nor are restricted to pairwise comparisons, they face one choice set at a time.

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

We finish this section by discussing our model and our findings in relation to Rational Inattention (RI) models. Caplin et al. (2016) 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.5 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.

5. Structural Estimation of Preferences and Attention: Implications for Welfare

In this section we use Theorem 3 to estimate the severity of limited consideration across consideration costs, its effect on choices and therefore on welfare. We also identify preferences.

We start with considering the LA restriction on random consideration because it is the only model that is not rejected across all costs of consideration. We find that the estimated distribution over consideration sets exhibits substantial heterogeneity. For example, the probability of considering the grand set X is decreasing with the cost of consideration.

Next we restrict the analysis from the previous section to a class of CRRA Bernoulli utility functions. That is, we consider a parametric utility function

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

The CRRA restriction further reduces the cardinality of $R^{\mathrm{EU}}(X)$ from 10 orders to 6 orders. We consider this restriction because it exhibits single-crossing which guarantees uniqueness of the distribution over preference orders π (Apesteguia et al. (2017)), and it has been extensively studied

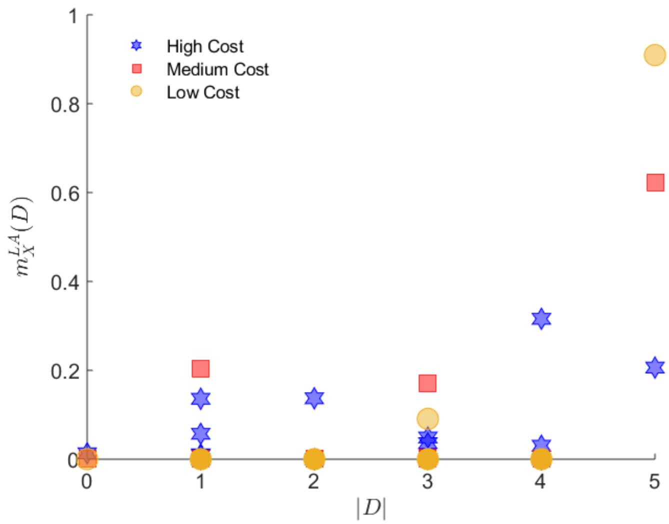


Figure 10 – Estimated Stochastic Consideration Rule across Cost Treatments

in the literature (e.g., Holt and Laury (2002)). We find that the CRRA restriction does not change our previous findings. That is, we do not reject our hypothesis of independence of preferences (consistent with the CRRA restrictions) and random consideration (consistent with the LA-rule); while the random consideration rule changes across consideration costs.[46]

5.1. Estimated Stochastic Consideration Rule

We estimate $\hat{m}^{\text{LA}}$ from $(\hat{p}(o, A))_{A\in\mathcal{A}}$.[47] We now focus on the distribution over consideration sets when DMs face the grand choice set X (the analysis can be conducted for any other menu $A \subseteq X$). Focusing on X allows us to estimate the distribution of consideration over all possible menus in our sample.[48] The estimated stochastic rule for X across cost treatments as a function of the size of the consideration set is presented on Figure 10. The figure shows that when DMs face the high consideration cost, their attention is stochastic across different subsets of the menu. In particular, the estimated probability of considering the entire menu is $\hat{m}^{\text{LA}}_{X,\text{High}}(X) = 0.2061$. As the cost decreases consideration shifts to bigger subsets. In particular, when there is no additional costs to compute lotteries (low cost), the estimated probability of considering the entire menu is close to 1 ($\hat{m}^{\text{LA}}_{X,\text{Low}}(X) = 0.9092$). Table 7 summarizes these results.

The results in Table 7 can be interpreted as the proportion of individuals that, when faced the grand set X, consider sets of size 0,1,2,3,4, and 5. Attention shifts towards bigger sets as the consideration cost decreases. This is consistent with our expectations as the fraction of fully

[46]In particular, we cannot reject stability of preferences across consideration cost treatments. The p-value of the hypothesis that $\pi^{\text{CRRA-LA}}_{\text{Low}} = \pi^{\text{CRRA-LA}}_{\text{Medium}} = \pi^{\text{CRRA-LA}}_{\text{High}}$ is 0.49.

[47]The estimation has two steps, first we use Definition 7 to construct the first stage estimates. Next we project them to the simplex. The second step estimator is consistent when LA is the true model, which we cannot reject.

[48]Note that for LA it must be that $m^{\text{LA}}_X(D) = \eta^{\text{LA}}(D)$ for all $D \in 2^X$. Therefore, η^{LA} is the stochastic consideration rule over 2^X.

Table 7 – AGGREGATE ESTIMATED STOCHASTIC CONSIDERATION RULE PER CONSIDERATION TREATMENT The table displays $\sum_{D\in 2^X:|D|=k} \hat{m}^{\mathrm{LA}}_X(D)$ with $D \in 2^X$.

	CONSIDERATION SET SIZE					
COST	$\lvert D\rvert = 0$	$\lvert D\rvert = 1$	$\lvert D\rvert = 2$	$\lvert D\rvert = 3$	$\lvert D\rvert = 4$	$\lvert D\rvert = 5$
HIGH COST	0.0133	0.2125	0.1376	0.0855	0.3464	0.2061
MEDIUM COST	0.0001	0.2047	0.0006	0.1743	0.0003	0.6218
LOW COST	0.0001	0.0003	0.0006	0.0913	0.0003	0.9092

Estimates are bounded away from zero for attention to be well-defined for all menus, therefore $\min_A \hat{m}_X(A) = 0.0001$.

rational or full consideration DMs grows as the cost of consideration decreases.

Figure 10 also shows that the attention is not uniformly distributed across (strict) subsets of X. For example, for the high cost, the distribution of attention for subsets of size 4 is

$$\begin{aligned}
&\hat{m}^{\mathrm{LA}}_{X,\mathrm{High}}(X \setminus \{l_1\}) = \hat{m}^{\mathrm{LA}}_{X,\mathrm{High}}(X \setminus \{l_2\}) = \hat{m}^{\mathrm{LA}}_{X,\mathrm{High}}(X \setminus \{l_4\}) = 0.0001,\\
&\hat{m}^{\mathrm{LA}}_{X,\mathrm{High}}(X \setminus \{l_3\}) = 0.0303,\\
&\hat{m}^{\mathrm{LA}}_{X,\mathrm{High}}(X \setminus \{l_5\}) = 0.3159.
\end{aligned}$$

This suggests that different lotteries attract attention differently. To simplify the exposition, we compute the following average attention index for each lottery:

$$I_A(a) = \sum_{D\in 2^A:\, a\in D,} \frac{1}{|D|} m_A(D).$$

This index is an average of the consideration paid to subsets where the lottery is present weighted by the inverse of the size of the menu. It measures the average contribution to consideration of alternative a in X.[49] Table 8 shows $I_X(a)$ for each consideration cost treatment.

The average consideration index can also be understood as the weighted probability of choice frequencies if DMs were to have the counterfactual uniform distribution over all preference orders. Under this hypothesis, full consideration would imply that each alternative would be chosen, in the aggregate, with probability 1/5. Table 8 shows that, when there is no cost of consideration (i.e, low cost treatment) the implied indices per lottery are not very different from full consideration. This is not the case for the medium and high cost.

For the medium and high costs lottery 2 has the highest average attention. On the other hand, lottery 5 seems to grab proportionally less attention when there is a cost of considering alternatives. We remind the reader that the consideration cost is induced on prizes, while probabilities are seen without distortion. Arguably lottery 2 is, ex-ante, one of the simplest ones to compute since it

[49] Note that this index can easily by extended to compute the contribution of pairs, triples, or any other strict subset; to compute for example how any given two lotteries compete for attention.

Table 8 – Average Consideration index per lottery / consideration treatment, $I_X(a)$

	Consideration Cost - Data			Theory
Lottery	High	Medium	Low	FC
Lottery 1	0.1966	0.1247	0.1822	0.2000
Lottery 2	0.3607	0.3870	0.2125	0.2000
Lottery 3	0.1475	0.1826	0.2125	0.2000
Lottery 4	0.1486	0.1814	0.1822	0.2000
Lottery 5	0.1346	0.1260	0.2125	0.2000

is uniform in two prizes. With the same argument, lottery 5 seems to be the more complex to compute (recall that the probabilities that are needed to understand the lottery are 1/5, 1/4, 3/20, 1/4, and 3/20). Note that we have not imposed any structure on m but our estimates correspond naturally with the reasonable idea that simpler lotteries will be in average considered more often than complex ones when choice is hard.

5.2. Estimated Preference Distribution

We remind the reader that we cannot reject the hypothesis that stochastic choices are generated by CRRA-LA DMs whose preferences are stable across consideration cost treatments (p-value=0.49). Imposing the restriction that expected utility maximizers can be represented by a CRRA Bernoulli utility function we gain the uniqueness of the estimated π as discussed above.[50] Therefore we can estimate the implied distribution over preference types, map it to our parametric assumption, and set-estimate the implied CRRA parameter.

Our results show that $\hat{\pi}$ places substantial positive probability on the two out of six preference orders consistent with CRRA-LA. The fractions of individuals with preferences corresponding to other orders are below 5 percent, as reported in Table 9. Almost 50 percent of DMs prefer lottery 2 to the rest with $\sigma \in (0.3002, 1]$. Roughly 30 percent of DMs prefer lottery 1 the most, and around 20 percent of DMs have lottery 4 or 5 as the most preferred item.[51] In particular, this

[50] In order to implement this in practice, we simulate all possible orders for a CRRA parameter ranging from $\sigma \in [-1, 1]$ with grid with step=10^{-4}. Since the number of lotteries is finite and CRRA is special case of EU, we know that we can have at most 10 orders. We compute 6 orders in total, each one corresponding to a bracket of values of σ. Due to the single crossing property these brackets are in fact connected intervals. The single crossing property is easily verified with respect to the enumeration of orderings given by σ in which $\succ_1$ corresponds to $\sigma \in [-1, 0.2287)$ and $\succ_6$ corresponds to $\sigma \in (0.3001, 1]$. For example the pair l_1 is preferred to l_3 from order 1 to order 3, and l_3 is preferred to l_1 from order 4 to order 6. All pairs satisfy this property for different breaking points.

[51] This results are very much in line with the estimation for risk aversion in Holt and Laury (2002). The comparison cannot be directly performed since the intervals for σ cannot be directly mapped to ours due to the structure of the

Table 9 – Estimated distribution across preference types given CRRA Utility and LA-rule

Pref. Order	$\hat{\pi}$	Implied σ
$l_1 \succ l_4 \succ l_3 \succ l_5 \succ l_2$	0.30500	[-1,0.2287)
$l_4 \succ l_1 \succ l_5 \succ l_3 \succ l_2$	0.04905	(0.2287,0.2606)
$l_4 \succ l_5 \succ l_1 \succ l_3 \succ l_2$	0.04905	(0.2606,0.2728)
$l_5 \succ l_4 \succ l_2 \succ l_3 \succ l_1$	0.04905	(0.2728,0.2832)
$l_5 \succ l_2 \succ l_4 \succ l_3 \succ l_1$	0.04905	(0.2832,0.3001)
$l_2 \succ l_5 \succ l_3 \succ l_4 \succ l_1$	0.49880	(0.3001,1]

preference are consistent with the fact that the default alternative is the worst according to the DMs preferences and will only be picked if no lottery is considered. This implies that the observed probability of choice over the outside option is a revealed measure of welfare loss.

Note that we cannot reject that the majority our DMs are risk-averse. Also, we cannot reject that for the two major preference types, the default alternative o is ranked last according to the CRRA parameter implied rank over $X \cup \{o\}$. We underline that this preferences distribution is the same for all three costs of consideration. Also, we remind the reader that we strongly rejected the EU-RUM model, but under limited consideration of the LA type we can recover a stable preference distribution under the CRRA restriction.

5.3. Welfare consequences of Limited Consideration

Our model allows to disentangle the effects of random consideration and heterogeneity in preferences in stochastic choice data. In our data, both sources of stochastic behavior are present. We show that, in particular for the high and medium consideration cost, the extent of limited consideration not only changes with the consideration cost treatments, but also depends on the alternatives and the size of the menu. Moreover, our population is heterogeneous in terms of preferences.

Our identification results also permit us to examine the welfare effects of limited consideration in our population. *The structural estimate of welfare loss* is the fraction of DMs that are suboptimizing. Suboptimizing means that a DM does not consider the best possible lottery according to her preference type. We are able to compute this welfare loss exactly in a cross-section of individuals

lotteries. For example, for the low and real payments -comparable payments in our design, Holt and Laury (2002) estimates that 34 percent of the population can be represented by $\sigma < 0.15$, while 40 percent correspond to $\sigma > 0.41$. They find similar results for larger real and hypothetical payments.

by using the independence assumption between preferences and consideration.[52] The structural estimation of LA-HRC with CRRA preferences makes sure that the preference distribution and the random consideration rules are well-defined (this is not needed for testing).

To better understand suboptimization consider the probability of choosing the outside option. In our model the outside option is only chosen if nothing in the menu is considered. Therefore, $(\hat{m}_A^{\mathrm{LA}}(\{\emptyset\}))_{A\in\mathcal{A}}$ provides an estimate of the fraction of individual who would be better off if they considered anything but the default option only.

On top of individuals that are suboptimizing because they do not pay attention to anything but the default option, there are also DMs that pay attention to some alternatives beyond the default, but these alternatives were not the best according to their preference orders. Thus, these DMs would also be better off if they considered all available alternatives. Our measure of welfare loss takes into account these DMs as well.

The results for the grand choice set X are: (i) in the high cost case 24.07 percent of individuals are not choosing their first option; (ii) in the medium cost case 17.34 percent of individuals are suboptimizing; and (iii) in the low cost case only 3.76 percent of individuals fail to optimize. These findings support the null hypothesis that hard choices impact welfare negatively via limited consideration. In the high cost case almost 25 percent of the population chooses a dominated option. The cognitive task we introduce is arguably quite modest, however it has a measurable impact on the welfare of DMs. We also observe that the effect of the consideration cost on welfare is strictly increasing.

6. Conclusion

We have designed and implemented a novel experiment with a large sample that allowed us to statistically discern among competing models of population behavior. By exogenously varying choice sets and the 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 the extensions of the models of limited consideration(LA (Brady and Rehbeck (2016)), RCG (Aguiar (2017), and MM (Manzini and Mariotti (2014)) with or without restrictions on preference heterogeneity. We call this extension L-HRC.

These models provide testable implications on choices that uniquely identify the stochastic consideration set rule from data. By calibrating consideration given the theory, we show that testing for L-HRC can be casted 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 L-HRC.

[52]The fraction of DMs suboptimizing for a fixed preference type is going to be given by the fraction of DMs that are not considering the best alternative according to the preference type multiplied by the probability of the preference type.

We provide evidence against classical RUM, since consideration costs are binding for some individuals in the population. On the other hand, we find support for the LA model with heterogeneous preferences. We cannot reject it neither in the pooled data nor in the data that corresponds for different levels of the consideration cost. Moreover, we cannot reject that the distribution on preferences, implied by LA, is the same across all attention costs. This means that once we disentangled attention and preferences under LA, the recovered distribution does not change with the consideration cost.

Under the assumption that DMs have CRRA preferences and use LA-rule, we uniquely recover the distribution over consideration costs and the distribution of preferences. Our findings indicate that (i) the distribution over consideration costs is heterogeneous; (ii) there are two major preference types in our sample. Using the estimated distributions over consideration costs and preferences we quantify the fraction of individuals that are suboptimizing because of limited consideration for different consideration costs. This fraction can be as high as 25 percent.

References

Abaluck, J. and Adams, A. (2017). What do consumers consider before they choose? identification from asymmetric demand responses. Technical report, National Bureau of Economic Research.

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.

Allen, R. and Rehbeck, J. (2018). Stochastic choice with observable characteristics.

Apesteguia, J., Ballester, M. A., et al. (2016). Stochastic representative agent. Technical report.

Apesteguia, J., Ballester, M. Á., et al. (2018). Separating predicted randomness from noise. Technical report.

Apesteguia, J., Ballester, M. A., and Lu, J. (2017). Single-crossing random utility models. *Econometrica*, 85(2):661–674.

Barseghyan, L., Coughlin, M., Molinari, F., and Teitelbaum, J. (2018). Heterogeneous choice sets and preferences. Presented at Oxford 2018, SARP 2.

Barseghyan, L., Molinari, F., and Thirkettle, M. (2019). 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. (2016). Rational inattention, optimal consideration sets and stochastic choice. Technical report, Working paper.

Cattaneo, M. D., Ma, X., Masatlioglu, Y., and Suleymanov, E. (2017). A random attention model. *arXiv preprint arXiv:1712.03448*.

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.

Conlon, C. T. and Mortimer, J. H. (2013). Demand estimation under incomplete product availability. *American Economic Journal: Microeconomics*, 5(4):1–30.

Dardanoni, V., Manzini, P., Mariotti, M., and Tyson, C. (2019). Inferring cognitive heterogeneity from aggregate choices. Technical report, Department of Economics, University of Sussex.

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.

Dehmamy, K. and Otter, T. (2014). Utility and attention-a structural model of consideration. *Available at SSRN*.

Draganska, M. and Klapper, D. (2011). Choice set heterogeneity and the role of advertising: An analysis with micro and macro data. *Journal of Marketing Research*, 48(4):653–669.

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.

Gabaix, X., Laibson, D., Moloche, G., and Weinberg, S. (2006). Costly information acquisition: Experimental analysis of a boundedly rational model. *American Economic Review*, 96(4):1043–1068.

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

Gul, F., Natenzon, P., and Pesendorfer, W. (2014). Random choice as behavioral optimization. *Econometrica*, 82(5):1873–1912.

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. (2018). Random consideration and choice: a case study of default options. Technical report.

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.

Huang, Y. and Bronnenberg, B. J. (2018). Pennies for your thoughts: Costly product consideration and purchase quantity thresholds. *Marketing Science*.

Kashaev, N. and Lazzati, N. (2019). Peer effects in random consideration sets. *arXiv preprint arXiv:1904.06742*.

Kawaguchi, K., Uetake, K., and Watanabe, Y. (2016). Identifying consumer attention: A product-availability approach.

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 Ulku, L. (2017). Satisficing with a variable threshold.

List, J. A. and Haigh, M. S. (2005). A simple test of expected utility theory using professional traders. *Proceedings of the National Academy of Sciences*, 102(3):945–948.

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.

Reutskaja, E., Nagel, R., Camerer, C. F., and Rangel, A. (2011). Search dynamics in consumer choice under time pressure: An eye-tracking study. *American Economic Review*, 101(2):900–926.

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.

Roberts, J. H. and Lattin, J. M. (1997). Consideration: Review of research and prospects for future insights. *Journal of Marketing Research*, pages 406–410.

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.

Van Nierop, E., Bronnenberg, B., Paap, R., Wedel, M., and Franses, P. H. (2010). Retrieving unobserved consideration sets from household panel data. *Journal of Marketing Research*, 47(1):63–74.

A. Proofs

A.1. Proof of Lemma 1

Proof. 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 HRC. We now prove that it generates any data P. By definition if P can be generated by a HRC, we have that

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

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

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

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

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

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

This implies that

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

given that

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

∎

A.2. Proof of Theorem 1

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

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

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

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

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 $\sum_{a\in A}\sum_{\succ\in R(X)}\pi(\succ)\mathbb{1}\left(\, a\succ b,\;\forall\, b\in D\,\right)=1$ 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 $\varphi^*:2^X\to[0,1]$ be defined by $\varphi^*(A)=p(o,A)$.

The fact that $\eta=\eta^{\mathrm{L}}$ 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 $\mathrm{L}\in\{\mathrm{LA},\mathrm{MM},\mathrm{RCG},\mathrm{FC}\}$ and P:

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

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

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

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

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

and $\gamma : X \to (0,1)$. 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 $m \in \mathcal{M}^{\text{RCG}}$, 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 $m = m^{\text{L}}$ for given $\text{L} \in \{\text{LA}, \text{MM}, \text{RCG}, \text{FC}\}$ 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 $\text{L} \in \{\text{LA}, \text{MM}, \text{RCG}, \text{FC}\}$. To see that this is true recall that if P is a L-HRC-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 Shafer (1976) it follows that

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

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

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

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

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

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

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

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

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

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

This implies that P^{L} is a FC-HRC-rule.

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

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

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

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

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

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

Hence, P is a L-HRC-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:

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

■

A.4. Proof of Theorem 3

Proof. 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 $\mathrm{L} = \mathrm{LA}$, while Aguiar (2017) provides identification results for $\mathrm{L} = \mathrm{RCG}$. For $\mathrm{L} = \mathrm{MM}$ the result holds trivially.

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

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

and

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

for all $a \in A$ and nonempty $A \subseteq X$, which follows from Definition 10. By condition (ii), $m_A^{\mathrm{L}}(A) > 0$ and using the recursive definitions above for binary sets, we can see that $p_{\pi}^{\mathrm{L}}(a, \{a, b\}) = p_{\pi'}^{\mathrm{L}}(a, \{a, b\})$ for any $a, b \in X$. For a fixed m the recursive formula leads to the equivalence $p_{\pi'}^{\mathrm{L}} = p_{\pi}^{\mathrm{L}}$. ■

B. Testing Random Expected Utility with a Finite Choice Set and Finite Prizes

In our experiment we have a finite set of prizes Z and a finite set of lotteries $X \subset \Delta(Z)$. Here we derive the necessary and sufficient conditions for testing the null hypothesis that the population behavior captured by P is described by a distribution of preferences $\pi \in \Delta(R^{\mathrm{EU}}(X))$ that is defined over the EU restriction. Namely, the necessary and sufficient condition is that there exist a Bernoulli utility vector $u = (u_z)_{z \in X}$ such that for any pair of lotteries $x, y \in X$ and for a preference

$\succ \in R^{\text{EU}}(X)$ in the restricted support (i.e., $\pi(\succ) > 0$), $x \succ y$ if and only if $x'u > y'u$. Evidently, the order $\succ$ defined over X can be extended to the whole simplex $\Delta(Z)$, with the following expected utility for any lottery $w \in \Delta(X)$ by: $U(w) = w'u = \sum_{z \in Z} u_z w_z$, where u is the same Bernoulli utility vector as before (since prizes have not changed).

In practice, this means that we have to verify which elements of the total set of linear orders $R(X)$ are compatible with the expected utility restriction. We solve this problem by proposing a simple linear programming approach. Enumerate the lotteries in X such that $X = \{x_1, x_2, \cdots, x_n\}$ following a given candidate linear order $\succ$, such that $x_1 \succ x_2 \succ x_3 \cdots \succ x_n$. Define the row vector $a_i = (x_i - x_{i+1})'$ for all $i = 1, \ldots, n-1$. Stack the rows into a matrix $A = (a_i)_{i=1}^{n-1}$ (the size is $(n-1) \times |Z|$). Then, $\succ$ is compatible with the EU restriction if and only if there exists a vector $u \in \mathbb{R}^{|Z|}$ such that:

$$Au > 0,$$

where $0 \in \mathbb{R}^{n-1}$. This is easily checked using a linear programming approach where we minimize u subject to $Au > 0$.

The linear programming algorithm establishes if the problem is feasible and finds a solution if and only if the order $\succ$ is compatible with EU. Else, it declares that the program not feasible and we eliminate $\succ$ from our list of candidates.

C. Comparison with Models of Stochastic Choice

In this section 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 in the previous section, randomness arising from limited consideration as in RCG and MM can be rationalized under the umbrella of random utility. However, LA allows for behavior that is inconsistent with regularity. Therefore LA is not nested in RUM. By construction our model L-HRC generalizes FC by allowing for independent variation in choices due to limited consideration. In particular, L-HRC is RUM defined over X (what we call, equivalently, FC) when the stochastic choice rule is such that $m_A(D) = \mathbb{1}\left(D = A\right)$. We call this model FC.

Moreover, L-HRC is more general than RUM. This follows from the analysis in previous section. In particular, fixing preferences, $\pi(\succ_i) = \mathbb{1}\left(\succ_i = \succ\right)$ for $\succ_i \in R(X)$, L-HRC with L = LA reduces to original LA model by Brady and Rehbeck (2016), and therefore potentially inconsistent with RUM.

C.2. Comparison to the Random Attention Model

Cattaneo et al. (2017) 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. (2017). For instance, RAM nests LA, MM and RCG without preference heterogeneity among other salient models of consideration sets. Additionally, RAM is a strict generalization of RUM. However, our L-HRC is not nested in RAM, see Cattaneo et al. (2017) 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-HRC. 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. (2017) 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 of and only if $a\ P_R\ b$.

Then, a choice rule has a RAM representation if and only if P_R has no cycles. The following example of a L-HRC-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.

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 10. Moreover, consider two preference relations $\succ_1$ such that $a \succ_1 b \succ_1 c$, and $\succ_2$ such that $c \succ_2 b \succ_2 a$.[53] The resulting probabilistic

[53] In 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 EU-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.

Table 10 – EXAMPLE 3 Stochastic choice rule and random consideration set probability. p is consistent with LA-HRC 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

choice rule is generated by a LA-HRC 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. Comparison across stochastic choice rules for identification

The following simple example shows that identification of preferences and attention is not a trivial matter. In particular, many models may describe behavior accurately and cannot be rejected given observed data. However, identification may not be unique. This is important when predicting behavior and designing policy interventions.

Example 4 (Identification of m and π)**.** Consider a population of individuals that are utility maximizers but may not consider all alternatives available to them. Their behavior can be described by an independent consideration rule as in Manzini and Mariotti (2014). For simplicity we assume that all individuals share the same preferences ($a \succ b$) and that the probability of considering any alternative is $\frac{1}{2}$. Let $X = \{a,b\}$ and all menus $A \cup \{o\}$ with $A \in \mathcal{A}$ are observed. The table below shows the resultant stochastic choice rule.

	A		
	$\{a,b\}$	$\{a\}$	$\{b\}$
a	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
b	$\frac{1}{4}$		$\frac{1}{2}$
o	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{2}$

Identification under MM-HRC From Definition 7 we recover uniquely m from observed data. Given m, we recover preferences uniquely for this dataset.

Identification under RUM We can perfectly described the data with a RUM on $X \cup \{o\}$. However we do not recover the true preferences. For example, $\pi(\succ_i) = 1/4$ for all $i \in \{1, 2, 3, 4\}$ with $\succ_1$: $b \succ_1 o \succ_1 a$; $\succ_2$: $a \succ_2 o \succ_2 b$; $\succ_3$: $o \succ_3 a \succ_3 b$ and $\succ_4$: $a \succ_4 b \succ_4 o$

Identification under RAM (Cattaneo et al. (2017)) The random attention model (RAM) proposed by Cattaneo et al. (2017) fixes preferences and consider a random attention rule that must satisfy monotonic attention.[54] Even without preference heterogeneity RAM may fail to identify preferences for datasets consistent with it. By construction, the stochastic choice rule described in Example 4 is consistent with RAM. However, preferences are not uniquely identified. In particular, $b \succ a \succ o$ describes p under RAM with the attention rule $m_A(D)$ for $D \subseteq A$ as described below.

	Consideration set D						
Menu A	$\{a, b, o\}$	$\{a, o\}$	$\{b, o\}$	$\{a, b\}$	$\{a\}$	$\{b\}$	$\{o\}$
$\{a, b, o\}$	0	$\frac{1}{4}$	$\frac{1}{8}$	0	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{4}$
$\{a, o\}$		$\frac{1}{4}$	$\frac{1}{4}$		$\frac{1}{4}$		$\frac{1}{4}$
$\{b, o\}$			$\frac{1}{4}$			$\frac{1}{4}$	$\frac{1}{2}$

C.4. 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 $A \in \mathcal{A}$:

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

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 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:

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

[54] A stochastic consideration rule satisfies monotonic attention if for any $x \in A \setminus D$, $m_A(D) \leq m_{A \setminus \{x\}}(D)$.

for all $a \in A$ (and $\mathbf{e} = (\mathbf{e}_a)_{a\in A}$).[55]

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:

$$\pi_e(\succ\colon a \succ b\ \forall b \in A \setminus \{a\}) = \mathbb{P}\left(\mathbf{u}_a + \mathbf{e}_a > \mathbf{u}_b + \mathbf{e}_b\ \forall b \in A \setminus \{a\}\right).^{56}$$

In other words, the population of DMs behavior captured by P will still be represented by a L-HRC 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-HRC 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.

C.5. 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 adquire 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

[55] Note 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.

[56] Where $\pi_e(\succ\colon \text{property})$ denotes the cumulative probability of preferences that have a certain property.

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

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

For the specification of κ, we focus on the generalized entropy proposed in Fosgerau 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 data sets that replicate standard stochastic choice data.

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

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

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

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

The proof of this lemma follows from 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 $\alpha : 2^X \to [0, 1]$, 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 $D \subseteq A$: $K : [0,1] \to \mathbb{R} \cup \infty$. 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:

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

where $\alpha_A : 2^A \to \mathbb{R}$ 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 $m_A^{\mathrm{FC}}(A) = 1$. 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:

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

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 RCG 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:

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

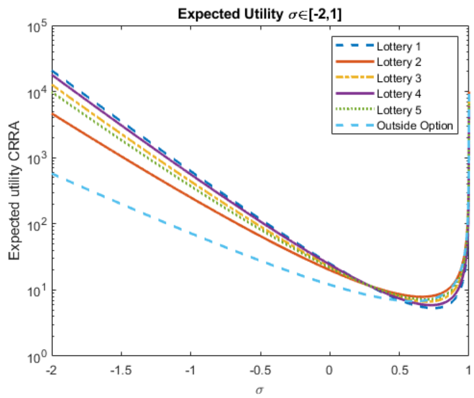
With this in hand we get:

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

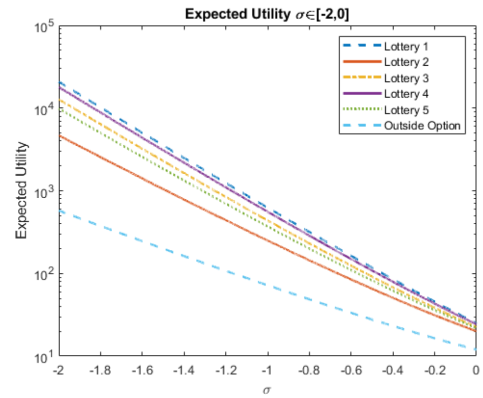
where ν ensure that this is well-behaved probability distribution. In this case, the attention-index is $\eta = \alpha$.

D. Experiment

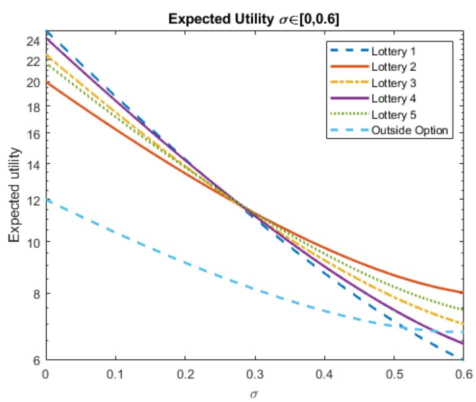
D.1. Experimental Design



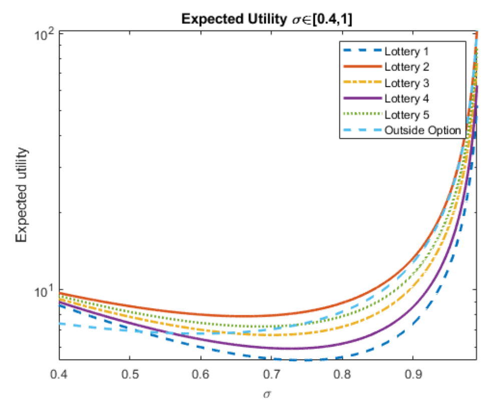
(a) Utility Comparison all levels



(b) Risk-Loving



(c) Neutral to moderate risk aversion



(d) Highly risk aversion

Figure 11 – 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 $k = \{\text{H, M, L, pooled}\}$. 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. In order to prevent possible learning, these subjects faced disjoint choice sets. That is, every subject faced either the full choice set $X \cup \{o\}$ or two choice sets that only had the outside option in common. Therefore, because of random assignment, in our experiment

(i) 171 subjects faced only the whole choice set (the targeted number is 180);

Table 11 – Average number of observations per alternative/choice set

Choice set	N	$N/\lvert A\rvert$	Choice set	N	$N/\lvert A\rvert$
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			

(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 11.

D.3. Comparative Statistics

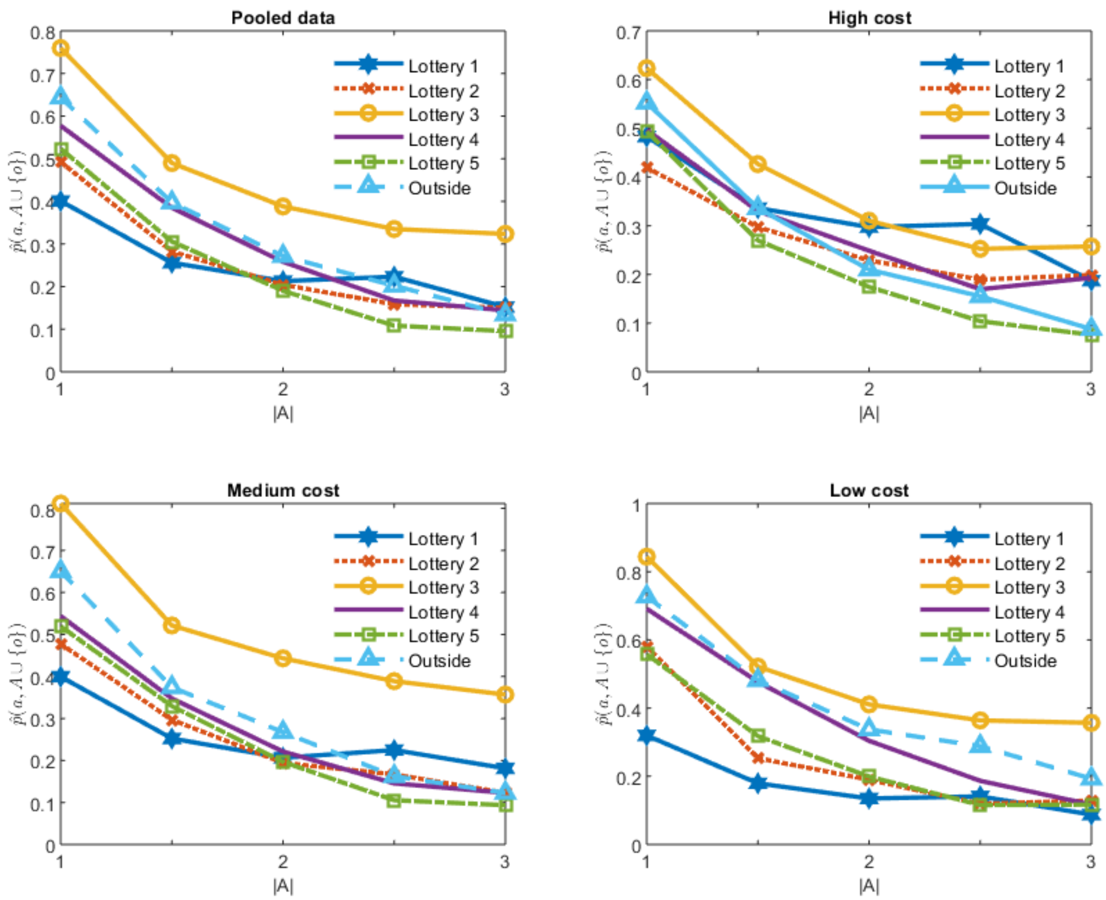


Figure 12 – Dynamics of Choice for all treatments

Table 12 – Determinants of Selecting the outside option Likelihood of choosing the outside option as a function of size of the menu, induced cost of attention and whether it is the first or second menu. Standard errors are in ().

	P(Choice=Outside Option)- Linear Probability Model				
	(1)	(2)	(3)	(4)	(5)
Size Menu 2	-0.1467***	-0.1472***	-0.2229***	-0.1420***	-0.1423***
	(0.0116)	(0.01157)	(0.0153)	(0.0198)	(0.0198)
Size Menu 3	-0.1892***	-0.1855***	-0.2637***	-0.1868***	-0.1831***
	(0.0116)	(0.0116)	(0.0153)	(0.0198)	(0.0198)
Size Menu 4	-0.1779***	-0.1748***	-0.2571***	-0.1797***	-0.1766***
	(0.0128)	(0.0129)	(0.0180)	(0.0220)	(0.0220)
Size Menu 5	-0.2495***	-0.2319***	-0.2962***	-0.2333***	-0.2156***
	(0.0211)	(0.0216)	(0.0343)	(0.0362)	(0.0364)
High Cost				0.1625***	0.1625***
				(0.0220)	(0.0220)
Medium Cost				0.0793***	0.0793***
				(0.0220)	(0.220)
High Cost x Size Menu 2			0.1559***	-0.0067	-0.0066
			(0.0174)	(0.0281)	(0.0280)
High Cost x Size Menu 3			0.1625***	-0.0001	0.0000
			(0.0174)	(0.0281)	(0.0280)
High Cost x Size Menu 4			0.1625***	0.0000	0.0000
			(0.0220)	(0.0311)	(0.0311)
High Cost x Size Menu 5			0.0994**	-0.0631	-0.0631
			(0.0463)	(0.0512)	(0.0512)
Medium Cost x Size Menu 2			0.0713***	-0.0072	-0.0080
			(0.0174)	(0.0281)	(0.0280)
Medium Cost x Size Menu 3			0.0721***	-0.0072	-0.0071
			(0.0174)	(0.0281)	(0.0280)
Medium Cost x Size Menu 4			0.0854***	0.0053	0.0053
			(0.0220)	(0.0311)	(0.0311)
Medium Cost x Size Menu 5			0.0936**	0.0143	0.0143
			(0.0463)	(0.0512)	(0.0512)
First Menu		-0.0322***	-0.0322***		-0.0322***
		(0.0079)	(0.0080)		(0.0079)
Constant	0.4016	0.4161	0.4161	0.3210	0.3356
	(0.0091)	(0.0098)	(0.0097)	(0.0155)	(0.0159)
Observations	12297	12297	12297	12297	12297
Adjusted R^2	0.0263	0.0274	0.0445	0.0474	0.0486
p-value F	0.000	0.000	0.000	0.000	0.000

Table 13 – DETERMINANTS OF CHOICE Linear probability model coefficients are reported with standard errors (), and p-values are displayed.

	LOTTERY 1		LOTTERY 2		LOTTERY 3	
	(1)	(2)	(1)	(2)	(1)	(2)
Cost High	0.0256***	-0.0505**	-0.0482***	-0.0179	-0.0258***	-0.0811***
	(0.0071)	(0.0205)	(0.0090)	(0.0260)	(0.0077)	(0.0222)
	0.000	0.014	0.000	0.491	0.001	0.000
Cost Medium	0.0126*	-0.0200	0.01330	-0.0146	-0.0367***	-0.0423*
	(0.0071)	(0.0208)	(0.0090)	(0.0264)	(0.0077)	(0.0226)
	0.074	0.344	0.138	0.580	0.000	0.061
Seen First	0.0848***	0.1411***	0.1209***	0.2168***	0.0880***	0.1463***
	(0.0063)	(0.0175)	(0.0079)	(0.022)	(0.0068)	(0.0190)
	0.0000	0.0000	0.000	0.000	0.000	0.000
Size Menu	0.0185***	0.0182***	0.0516***	0.0730***	0.0150***	0.0196***
	(0.0027)	(0.0052)	(0.0034)	(0.0066)	(0.0029)	(0.0057)
	0.000	0.001	0.000	0.000	0.000	0.001
Size × Seen First		-0.0309***		-0.0490***		-0.0297***
		(0.0056)		(0.0071)		(0.0061)
		0.000		0.000		0.000
Cost High × Seen First		0.0419***		0.0530***		0.0393**
		(0.0153)		(0.0194)		(0.0166)
		0.006		0.006		0.018
Cost Medium × Seen First		0.0147		0.0228		0.0028
		(0.0152)		(0.0193)		(0.0165)
		0.333		0.237		0.865
Cost High × Size		0.0239***		-0.0182**		0.0164**
		(0.0066)		(0.0083)		(0.0071)
		0.000		0.030		0.022
Cost Medium × Size		0.0101		0.0073		0.0015
		(0.0066)		(0.0084)		(0.0072)
		0.126		0.387		0.832
Constant	0.0269	0.0327	0.0525	0.0017	0.0932	0.0841
	(0.0034)	(0.0162)	(0.0206)	(0.1383)	(0.0102)	(0.0176)
	0.004	0.044	0.934	0.000	0.000	0.000
Observations	12297	12297	12297	12297	12297	12297
Adj R^2	0.0160	0.0196	0.0335	0.0385	0.0160	0.0185
p-value F	0.0000	0.0000	0.0000	0.0000	0.0000	0.000
Frequency of Choice	0.1178		0.2179		0.1424	

Table 14 – DETERMINANTS OF CHOICE, CONTINUATION TABLE 13. Linear probability model coefficients are reported with standard errors (), and p-values are displayed.

	LOTTERY 4		LOTTERY 5		OUTSIDE OPTION	
	(1)	(2)	(1)	(2)	(1)	(2)
Cost High	-0.0104	-0.0118	-0.0637***	-0.0033	0.1578***	0.1687***
	(0.0069)	(0.0199)	(0.0079)	(0.0229)	(0.0095)	(0.0242)
	0.132	0.554	0.000	0.886	0.000	0.000
Cost Medium	0.0003	-0.0137	-0.0454***	0.0061	0.0766***	0.0697***
	(0.0069)	(0.0202)	(0.0079)	(0.0232)	(0.0095)	(0.0242)
	0.963	0.499	0.000	0.793	0.000	0.004
Seen First	0.0770***	0.1247***	0.0844***	01386***		
	(0.0061)	(0.0170)	(0.0070)	(0.0195)		
	0.0000	0.0000	0.000	0.000		
Size Menu	0.0009	0.0119**	0.0182***	0.0435***	-0.0543***	-0.0538***
	(0.0026)	(0.0051)	(0.0030)	(0.0058)	(0.0035)	(0.0061)
	0.731	0.019	0.000	0.000	0.000	0.000
Size × Seen First		-0.02747***		-0.262***		
		(0.0054)		(0.0062)		
		0.000		0.000		
Cost High × Seen First		0.0247*		0.0190		
		(0.0148)		(0.0170)		
		0.096		0.264		
Cost Medium × Seen First		0.0367**		0.0131		
		(0.0148)		(0.0170)		
		0.013		0.440		
Cost High × Size		-0.0027		-0.0255***		-0.0042
		(0.0064)		(0.0073)		(0.0086)
		0.673		0.001		0.626
Cost Medium × Size		0.0001		-0.0217***		0.0026
		(0.0064)		(0.0074)		(0.0086)
		0.984		0.003		0.758
Constant	0.0835	0.0603	0.1122	0.0477	0.3228	0.3215
	(0.0091)	(0.0158)	(0.0105)	(0.0181)	(0.0113)	(0.0171)
	0.000	0.0000	0.000	0.008	0.000	0.000
Observations	12297	12297	12297	12297	12297	12297
Adj. R^2	0.0138	0.0161	0.0188	0.0212	0.0403	0402
p-value F	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Frequency of Choice	0.1096		01529		0.2594	

Table 15 – Time per choice set as a function of the cost and size of the choice set

	Dependent Variable: Time to Submit Choice (secs.)					
	(1)	(2)	(3)	(4)	(5)	(6)
Cost		12.5601***	3.8275***	3.8275**	4.4134***	4.3847***
		(0.4175)	(1.0629)	(1.0616)	(1.0560)	(1.0550)
		0.000	0.000	0.000	0.000	0.000
Size Menu		6.0122***	-0.6941	-1.1898	-1.2745	-1.8242**
		(0.3076)	(0.8113)	(0.9704)	(0.8094)	(0.8074)
		0.000	0.392	0.220	0.024	
Cost × Size			3.3532***	3.3532***	3.4316***	3.4316***
			(0.3755)	(0.3751)	(0.3740)	(0.3723)
			0.000	0.000	0.001	0.000
First Question						7.2823***
						(0.6847)
						0.000
Constant	23.2605	-17.5174	-0.0521	3.2407	-5.9433	-8.0540
	(0.3583)	(1.2064)	(2.2961)	(3.1367)	(2.3641)	(2.3619)
	0.000	0.0000	0.982	0.302	0.012	0.001
Menu				✓		
Choice					✓	✓
Observations	12297	12297	12297	12297	12297	12297
F-statistic	0.00	643.48	458.31	44.96	190.90	183.46
p-value		0.00	0.00	0.00	0.00	0.00
R^2 adjusted		0.0946	0.1004	0.1027	0.1100	0.1179

E. Choice Overload, Attraction Effect, and Simulation Results

E.1. Choice Overload and Attraction Effect

In this section we examine whether there is evidence of choice overload and attraction effect in the sample. We focus on these two effects because: (i) they have been observed at the individual level, see Rieskamp et al. (2006); (ii) choice overload cannot be rationalized by any model that we are considering; while (iii) attraction effect can be accommodated by LA, it cannot be explained by other consideration set rules examined in this paper. Because of sampling variability, any evidence of choice overload and attraction effect in finite sample may not imply that the underlying stochastic choice rule is also consistent with choice overload and attraction effect. Next section shows that, in finite samples, datasets generated by utility maximizing individuals may still exhibit choice overload. Nonetheless, the magnitude of the finite sample choice overload can be informative.

Mediating choice with consideration sets allows the DM to simplify her choice problem. Therefore, if DMs mediate choice with consideration sets, it is not expected that the problem becomes overwhelming as the choice set size increases. That is, if DMs are indeed simplifying their choice problem, then it is not expected that the probability of considering something in the menu decreases monotonically with the size of the menu. Formally, we say that there is evidence of choice overload if the probability of choosing the outside option increases when new alternatives are added to the choice set.

Definition 14 (Choice Overload (CO))**.** We say there is evidence of choice overload if for some $A \in A$ and some $a \in A$ it is the case that $p(o, A) > p(o, A \setminus \{a\})$, where o denotes the outside option.

The results for the sample relative frequencies $\hat{P}$ are summarized in Table 16. Intuitively, there is evidence of choice overload if bigger choice sets drive decision makers to not consider any element in the choice set and choose the outside option. There is finite sample evidence of choice overload at all attention costs. Evidence for choice overload seems stronger for choice sets of 3 or 4 alternatives. However, it is possible that the evidence of choice overload is an artifact of sample variability. In fact, the mean intensity of the CO violations is below 0.03 for all cost levels which is arguably small. Consistent with this observation, in Section 4 we find statistical evidence against choice overload on the population level.

Attraction effect instead refers to the effect of an added alternative on the probability of choosing another alternative. Attraction effect is incompatible with RUM, MM, and RCG, but it can be explained by LA. Formally,

Definition 15 (Attraction Effect (AE))**.** We say that there is evidence for the attraction effect if for some $A \in A$, $a \in A$, and $x \notin A$

$$p(a, A \cup \{x\}) > p(a, A)\,.$$

Table 16 – EVIDENCE FOR CHOICE OVERLOAD IN SAMPLE For any pair of menus A, $A \setminus \{a\}$ the magnitude of deviation is defined as $(\hat{p}(o, A) - \hat{p}(o, A \setminus \{a\}))$ if positive, 0 otherwise. Mean refers to the average magnitude of deviations and Std. Dev as the standard deviation of the magnitudes of deviations across all menu pairs of the form A, $A \setminus \{a\}$. Proportion is the percentage of CO inequalities that are violated in sample.

		All A	$\|A\| = 5$	$\|A\| = 4$	$\|A\| = 3$	$\|A\| = 2$
POOL	Proportion	0.2933		0.6500	0.3000	
	Mean	0.0098		0.0261	0.0071	
	Std. Dev	0.0189		0.0241	0.0160	
LOW	Proportion	0.2267		0.6000	0.1667	
	Mean	0.0081		0.0238	0.0043	
	Std. Dev	0.0196		0.0288	0.0140	
MEDIUM	Proportion	0.3867	0.4000	0.7500	0.3333	0.1000
	Mean	0.0175	0.0040	0.0406	0.0138	0.0036
	Std. Dev	0.0276	0.0060	0.0344	0.0232	0.0110
HIGH	Proportion	0.2133		0.5000	0.2000	
	Mean	0.0098		0.0214	0.0103	
	Std. Dev	0.0244		0.0276	0.0292	

Table 17 – ATTRACTION EFFECT IN SAMPLE Deviations from regularity supporting AE are given by $\hat{p}(a, A \cup \{x\}) - \hat{p}(a, A)$ if positive, 0 otherwise. Mean and Std Dev are the mean and standard deviations of the magnitude of deviations across the 140 possible comparisons $(a, A \cup \{x\}), (a, A)$. Proportion is the proportion of inequalities that evidence AE.

		DECOY				
		$x = 1$	$x = 2$	$x = 3$	$x = 4$	$x = 5$
POOL	Proportion	0.1071		0.0357	0.1071	0.0741
	Mean	0.0121		0.0035	0.0251	0.0155
	Std. Dev	0.0065			0.0239	0.0090
LOW	Proportion	0.2500	0.0357	0.1429	0.1786	0.0714
	Mean	0.0239	0.0325	0.0215	0.0362	0.0176
	Std. Dev	0.0166		0.0187	0.0196	0.0012
MEDIUM	Proportion	0.1429		0.1786	0.1429	0.0714
	Mean	0.0114		0.0419	0.0497	0.0324
	Std. Dev	0.0100		0.0350	0.0365	0.0007
HIGH	Proportion	0.0714	0.0714	0.1786	0.1429	0.1786
	Mean	0.0259	0.0164	0.0174	0.0232	0.0306
	Std. Dev	0.0264	0.0176	0.0152	0.0152	0.0236

Table 17 summarizes the evidence for attraction effect on the relative choice frequencies $\hat{P}$. Results show that the evidence for the attraction effect is weaker, and almost insignificant when the alternative introduced as a potential decoy is lottery 2. The attraction effect seems stronger, for the medium cost. However, the evidence is not conclusive at this point without a proper statistic analysis. We discuss the evidence for attraction effect in Section 4, finding statistical support for it only for intermediate costs of consideration.

E.2. Small Samples and Sampling Variability

The data exhibits evidence of choice overload, however it might be just the result of sampling variability in finite samples. Next we present a simple example where the realization of a probability rule consistent with RUM exhibits choice overload in small samples.

Example 5 (Choice overload for RUM data)**.** Let P be a probabilistic choice rule that is consistent with RUM over $\{a, b\} \cup \{o\}$ and is such that $p(o, \{a\}) = \frac{1}{2}$ and $p(o, \{a, b\}) = \frac{1}{4}$. Assume that the dataset consists of two observations: a is chosen from $\{a, o\}$, and o is chosen from $\{a, b, o\}$. Based on this data the estimated probability rule is given by $\hat{p}(o, \{a\}) = 0$ and $\hat{p}(o, \{a, b\}) = 1$. These choice frequencies exhibit choice overload (adding alternative b to the choice set boosts the probability of the outside option), and may lead to rejection of RUM if one does not take into account sample variability.

This simple example demonstrates the importance of statistical testing. In the limit, if the data is generated by a model that satisfies regularity, all evidence of choice overload should dissipate. However, considering two observations is not restrictive. To provide some Monte Carlo evidence about spurious occurrences of choice overload or attraction effect in finite samples, we simulate behavior for N individuals choosing from choice sets as in our experiment.

In particular, DMs are assumed to be utility maximizers over $X \cup \{o\}$. We simulated behavior for different numbers of observations and different levels of heterogeneity. A total of 1000 repetitions for 2, 5, 10, 50, 100 and 200 different preference orders (uniformly picked at random) were considered.[57] Table 18 presents the results of these simulations that are summarized in Figure 13.

The results show that, as the number of observations increases, the evidence for choice overload dissipates for all levels of heterogeneity. It is important to note than our pooled sample is of 12297; and the sample for each attention cost is of 4099, which implies that choice overload should dissipate if the underlying population rule is consistent with the L-HRC model.

[57] We also conducted this exercise with nonuniform distributions over preference orders. The results a qualitatively the same.

Table 18 – Simulations under RUM and evidence of choice overload with $|X| = 5$

	Number preference relations					
	2	5	10	50	100	200
N	Proportion of CO					
100	0.814	0.972	0.996	1.000	1.000	1.000
200	0.802	0.981	0.999	1.000	1.000	1.000
500	0.815	0.976	1.000	1.000	1.000	1.000
1000	0.810	0.980	1.000	1.000	1.000	1.000
5000	0.806	0.974	0.997	0.990	0.991	0.985
10000	0.818	0.975	0.997	0.966	0.923	0.901
15000	0.808	0.976	0.997	0.918	0.837	0.751
N	Total Magnitude					
100	4.794	7.185	7.855	8.614	8.524	8.520
200	3.043	4.677	4.905	5.362	5.304	5.230
500	1.829	2.446	2.518	2.392	2.326	2.300
1000	1.234	1.660	1.544	1.249	1.212	1.180
5000	0.553	0.716	0.553	0.226	0.177	0.148
10000	0.388	0.505	0.368	0.105	0.065	0.046
15000	0.315	0.411	0.306	0.068	0.034	0.022
N	Avg. Marginal Magnitude					
100	0.056	0.087	0.095	0.106	0.105	0.105
200	0.032	0.053	0.057	0.063	0.063	0.062
500	0.018	0.026	0.028	0.027	0.026	0.026
1000	0.012	0.018	0.017	0.014	0.013	0.013
5000	0.005	0.007	0.006	0.003	0.002	0.002
10000	0.004	0.005	0.004	0.001	0.001	0.001
15000	0.003	0.004	0.003	0.001	0.000	0.000

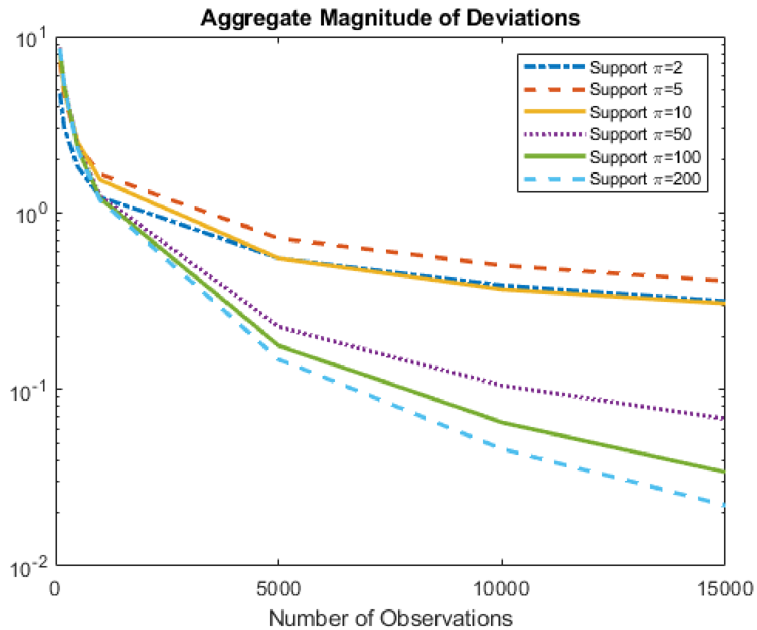


Figure 13 – Choice overload for RUM DMs - LogScale

F. 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-HRC 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-HRC 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

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

where $p^{\text{MM-HRC}}(a, A)$ is consistent with MM-HRC with (m^{MM}, π) and $\pi(\succ) = 1/10$ for all $\succ \in R^{\text{EU}}(X)$; and

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

The assumed process implies that a fraction $1-\lambda$ of the population exhibits choice overload.[58]

[58] This 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 19 – The table displays the proportion of rejections at the 90 percent and 95 percent confidence levels for LA-HRC with EU preferences. Sample size=4000. Number of MC replications=200.

Process	Confidence level 90%	Confidence level 95%
$\lambda\ = 0.25$	0.990	0.940
$\lambda\ = 0.50$	0.870	0.705
$\lambda\ = 0.75$	0.680	0.465

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-HRC. 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-HRC.

Table 19 presents the results for power simulations for sample size 4000 and $\lambda \in \{0.25, 0.50, 0.75\}$. For 200 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 99 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 fraction of DMs is moderate.

G. Experiment Instructions

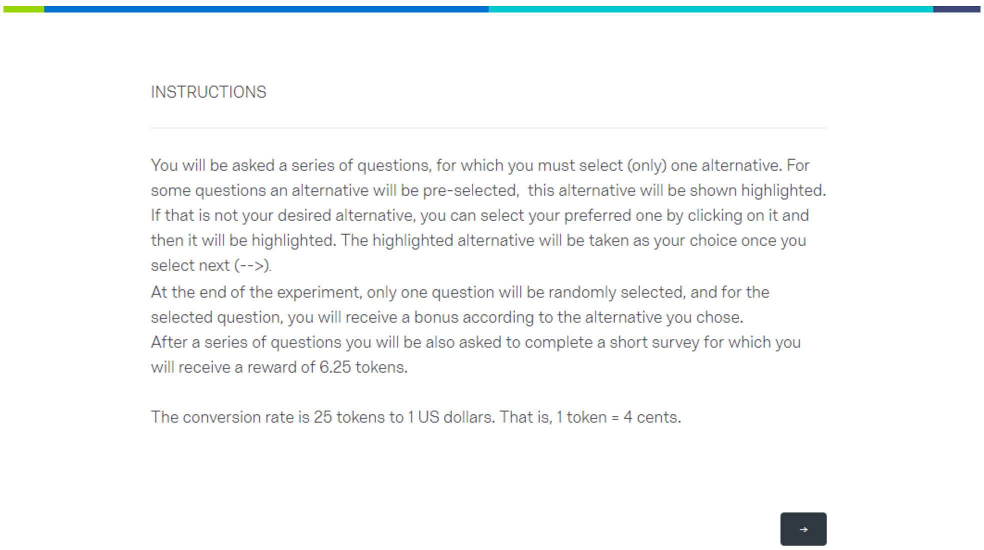


Figure 14 – Instructions page 1

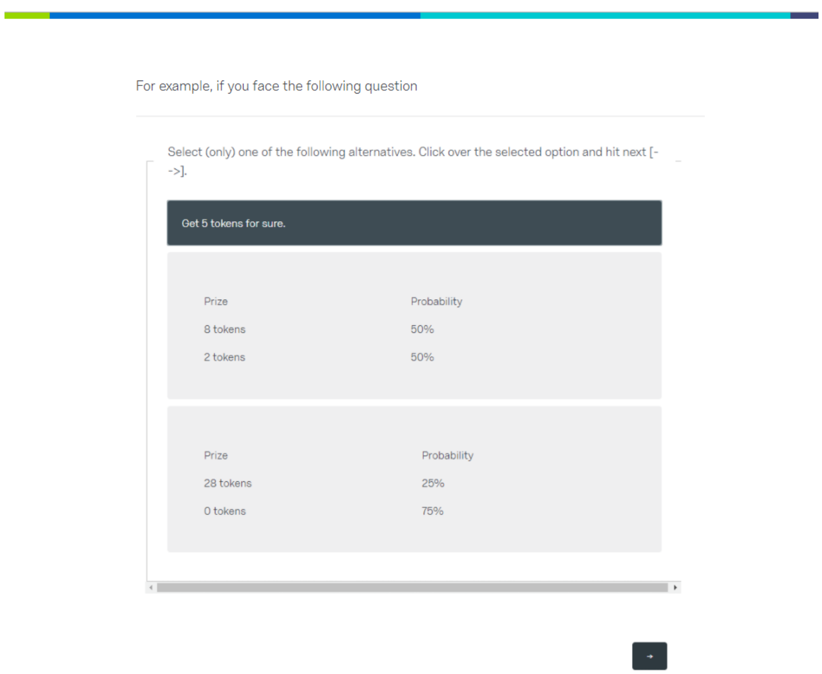


Figure 15 – Instructions page 2

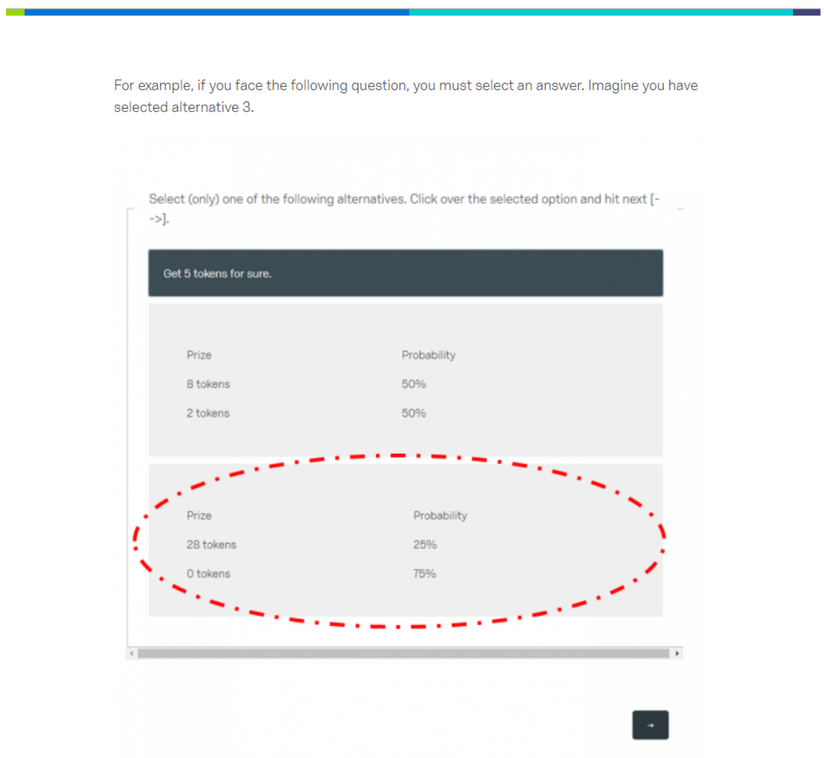


Figure 16 – Instructions page 3

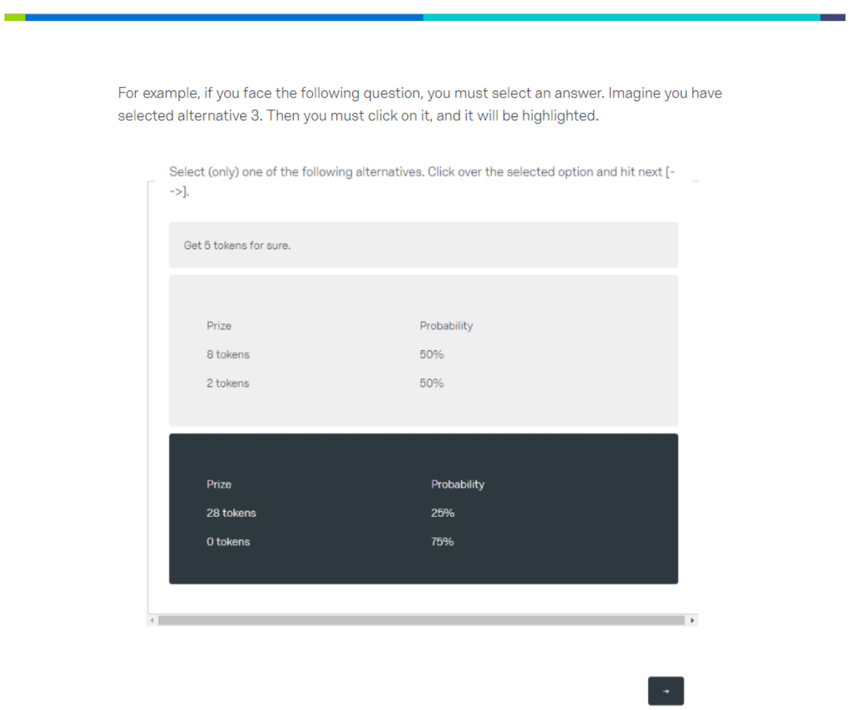


Figure 17 – Instructions page 4

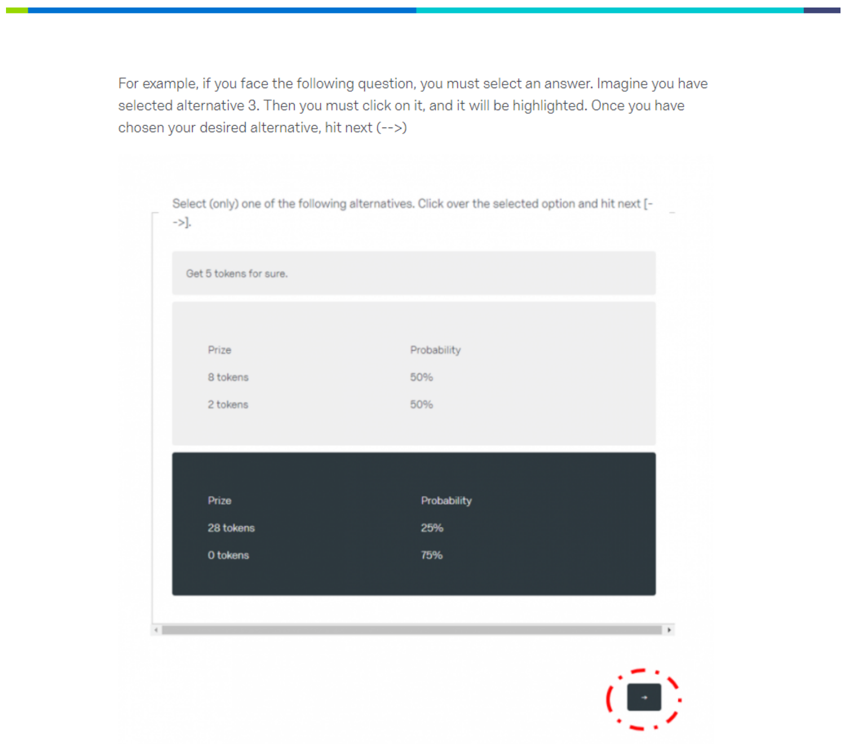


Figure 18 – Instructions page 5

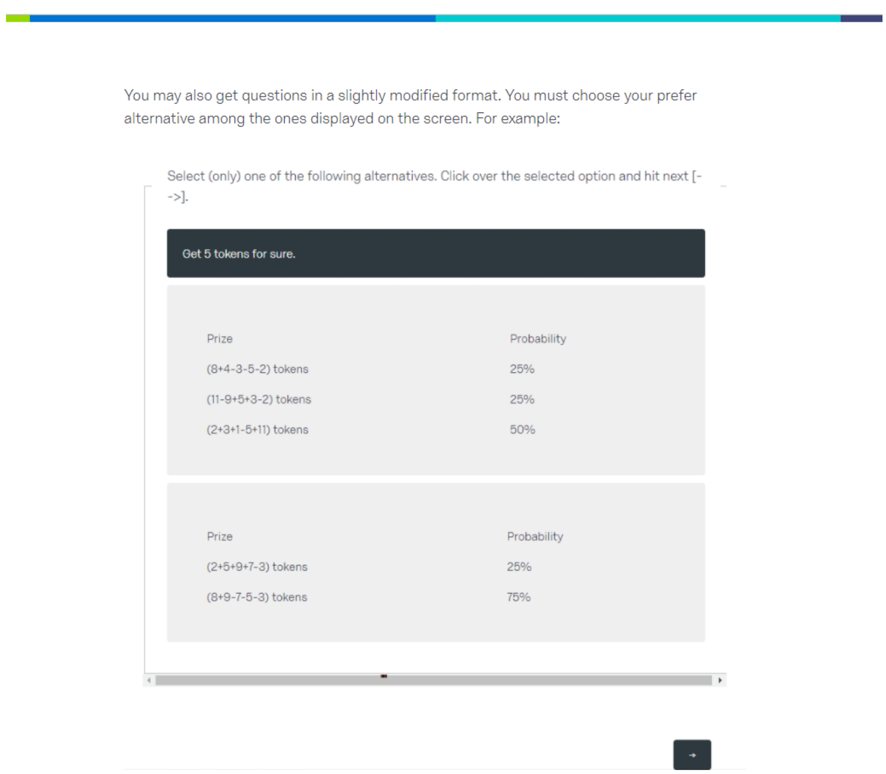
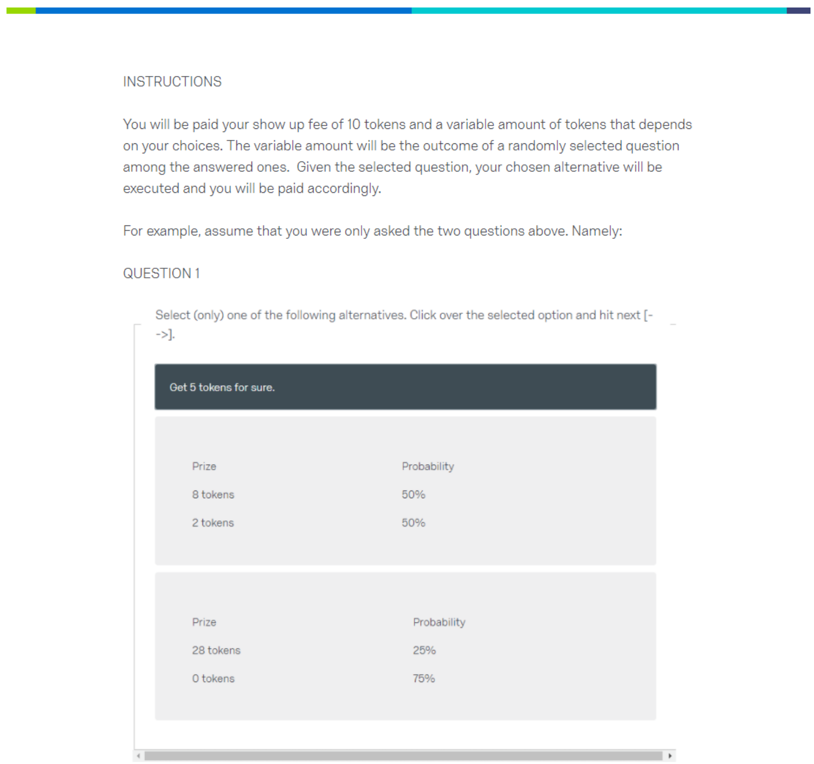


Figure 19 – Instructions page 6



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

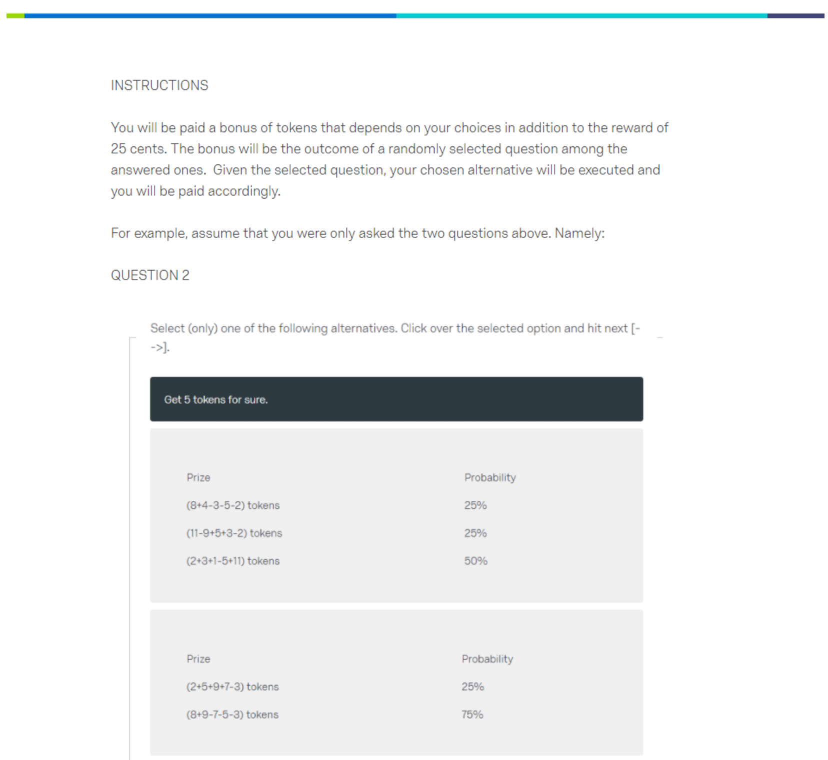
Select (only) one of the following alternatives. Click over the selected option and hit next [-->].

Get 5 tokens for sure.

Prize	Probability
8 tokens	50%
2 tokens	50%

Prize	Probability
28 tokens	25%
0 tokens	75%

Figure 20 – 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

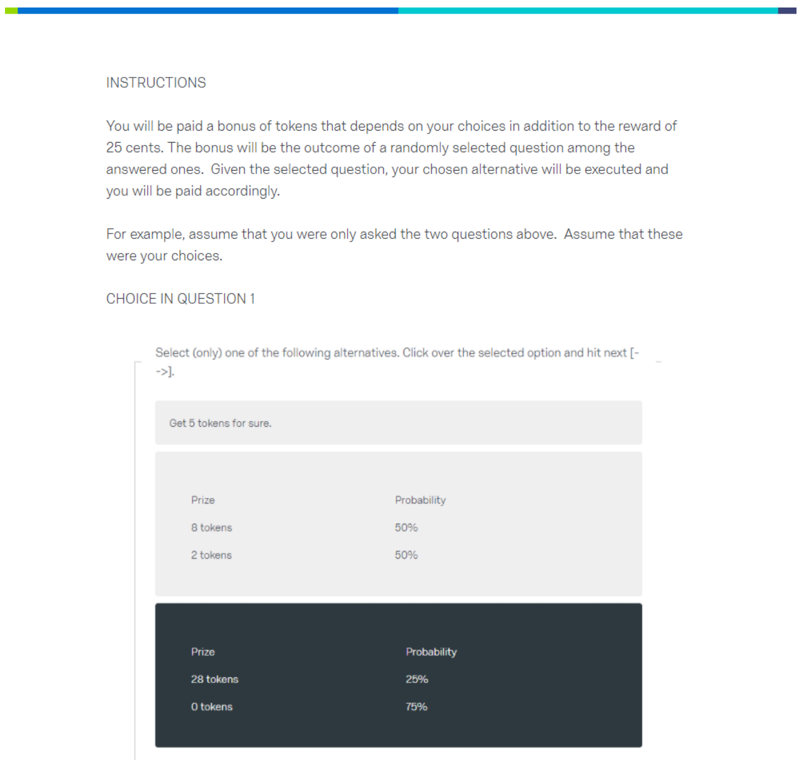
Select (only) one of the following alternatives. Click over the selected option and hit next [-->].

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+11) tokens	50%

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

Figure 21 – 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

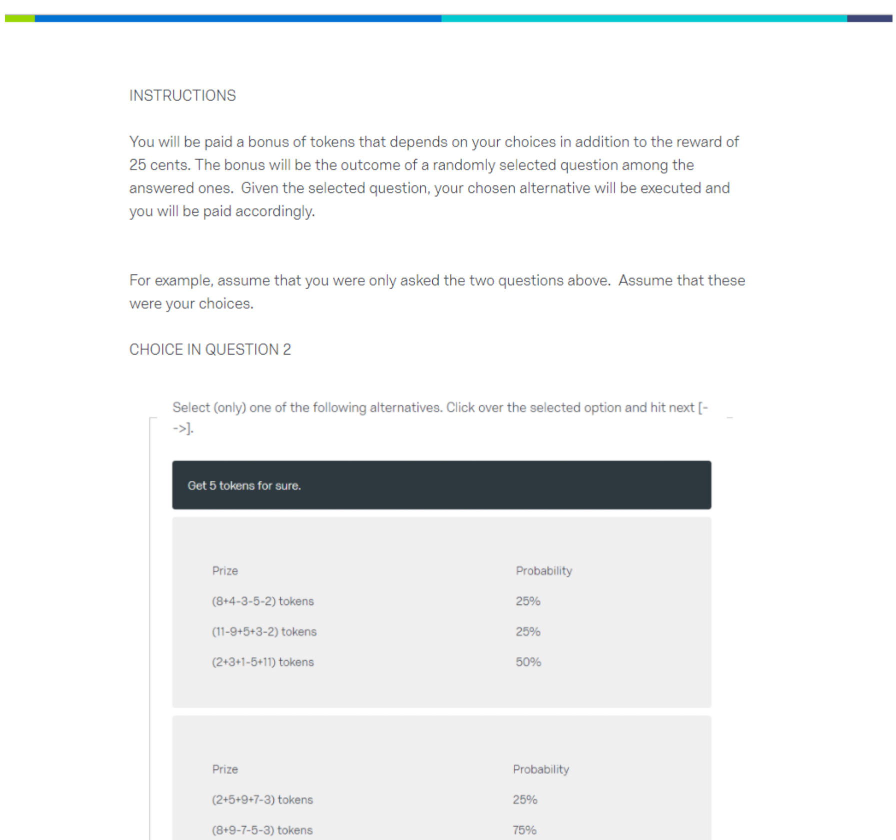
Select (only) one of the following alternatives. Click over the selected option and hit next [-->].

Get 5 tokens for sure.

Prize	Probability
8 tokens	50%
2 tokens	50%

Prize	Probability
28 tokens	25%
0 tokens	75%

Figure 22 – 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

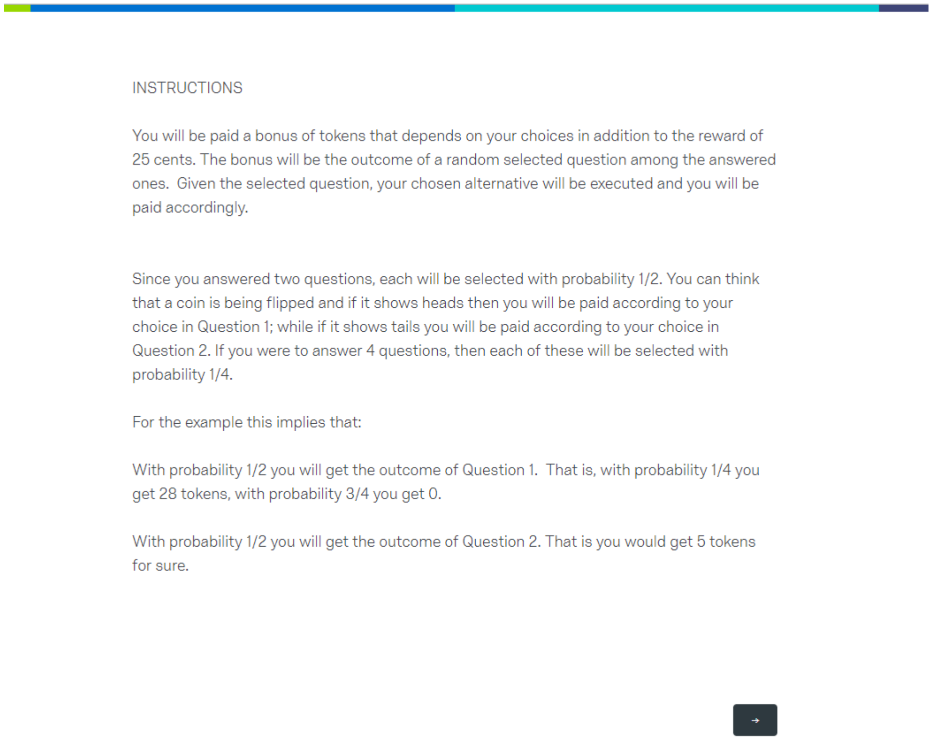
Select (only) one of the following alternatives. Click over the selected option and hit next [-->].

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 23 – 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.

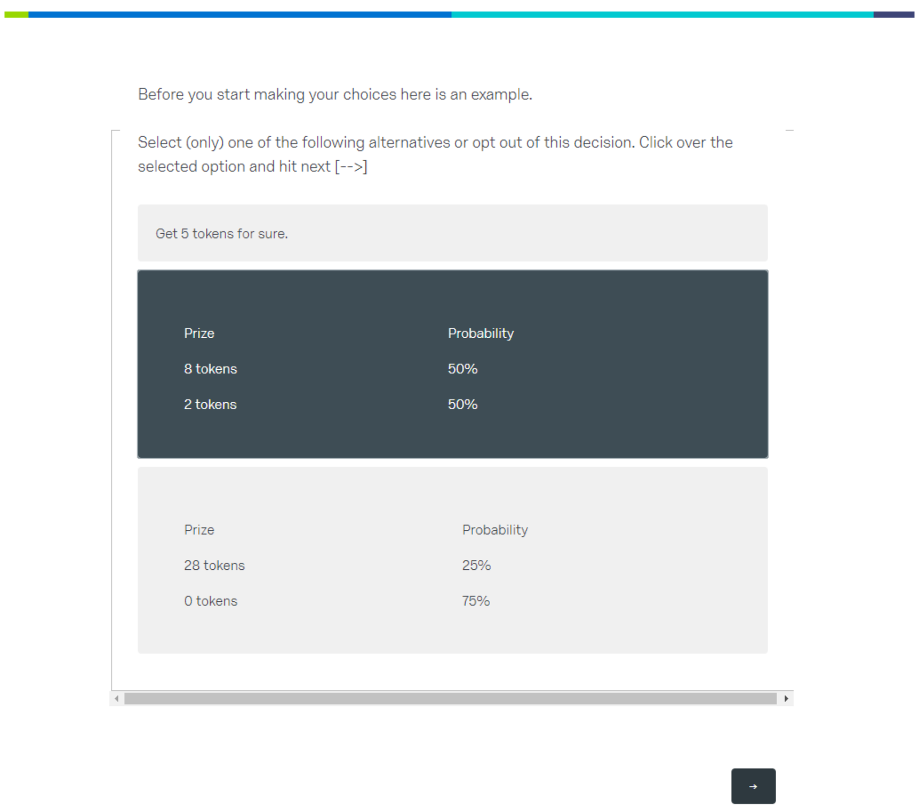
Since you answered two questions, each will be selected with probability 1/2. You can think that a coin is being flipped and if it shows heads then you will be paid according to your choice in Question 1; while if it shows tails you will be paid according to your choice in Question 2. If you were to 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 24 – 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 25 – Instructions page 12

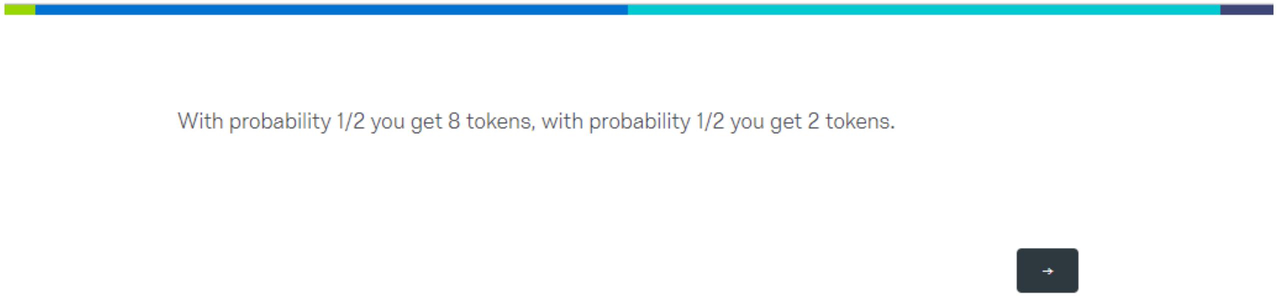


Figure 26 – Instructions page 13

Figure 27 – Instructions page 14