A Random Attention and Utility Model[*]

Nail Kashaev Ⓡ Victor H. Aguiar[†]

This version: October, 2021 / First Version: January, 2021

Abstract We generalize the stochastic revealed preference methodology of McFadden and Richter (1990) for finite choice sets to settings with limited consideration. Our approach is nonparametric and requires partial choice set variation. We impose a monotonicity condition on attention first proposed by Cattaneo et al. (2020) and a stability condition on the marginal distribution of preferences. Our framework is amenable to statistical testing. These new restrictions extend widely known parametric models of consideration with heterogeneous preferences.

JEL classification numbers: C50, C51, C52, C91.

Keywords: random utility, random consideration sets.

1. Introduction

The stochastic revealed preference methodology of McFadden and Richter (1990) is a cornerstone of economic analysis. This research agenda aims at explaining the behavior of a population of decision makers (DMs) as if each DM maximizes her utility, which is an independent identically distributed draw from a distribution of preferences, over their choice set (i.e., menus). This theory is usually referred to as random utility model (RUM).[1] If RUM is successful at describing behavior, then the analyst can use it to recover the distribution of heterogeneous preferences solely from observing the probability of choice of a finite set of alternatives from different menus justifying the name of revealed preference.[2] This distribution of preferences is an important input for many social sciences and can play a key role in policymaking. However, RUM may fail at describing behavior if DMs do not consider all available alternatives. This may happen, for instance, if there is a cost to understanding the decision task. In this situation, DMs may use a two-stage procedure:[3] first selecting a subset of the given menu (consideration set), and only then choosing the best alternative from that set. Given that there may be latent heterogeneity in DMs preferences, and in how DMs form consideration sets, from the analyst's standpoint, both the consideration sets and choices from these sets are random. As a result of this two-stage procedure, if the consideration set does not contain the most preferred alternative of a DM, the DM will choose a dominated alternative, failing to be consistent with RUM.[4]

---

[1] RUM was originally formulated by Block and Marschak (1960) and Falmagne (1978).

[2] Since Fishburn (1998), it is known that only partial identification of the distribution of preferences is possible under RUM. The probability of an item being ranked first is uniquely identified (Aguiar et al., 2016).

[3] Since Manzini and Mariotti (2007), there has been a renewed interest in studying models of sequential choice in economics. The main aim of this research is to accommodate context effects that produce violations of the standard utility maximization framework. In a sequential procedure of choice or a two-stage choice procedure, the DMs first simplify the original choice problem using some heuristic and then choose rationality from the simplified choice problem. Limited consideration is just one example of a factor that affects the first stage, which determines the effective choice set used in the second stage. Other factors may be willpower and status-quo bias (Horan, 2016).

[4] For examples of the distortions created by limited consideration, see Ho et al. (2017) and Heiss et al. (2016) (the health insurance market), Hortaçsu et al. (2017) (the residential electricity market in Texas), Honka (2014) and Honka et al. (2017) (the US auto insurance and banking industries), De Los Santos et al.

This paper proposes a generalization of the stochastic revealed preference methodology that is robust to limited consideration, allows for heterogeneous preferences that can be correlated with consideration, and is amenable to statistical testing. In doing so, we provide nonparametric restrictions on limited consideration and preferences that make partial recoverability of the distribution of preferences possible in a large class of stochastic choice data sets. Similar to Kitamura and Stoye (2019), our framework also permits robust counterfactual welfare analysis and out-of-sample predictions.

A large literature, pioneered by Masatlioglu et al. (2012) and Manzini and Mariotti (2014), has proposed theories of consideration-mediated choice. These theories accommodate some departures from RUM caused by inattention, feasibility, categorization, and search.[5] However, in contrast to our work, most existing theories of random consideration have assumed that preferences are *homogeneous* (Cattaneo et al., 2020). This restriction implies that these models are not well suited to describe behavior at the population level. Our framework allows for homogeneous and heterogeneous preferences that may be correlated with consideration, making it suitable for both experimental and field data sets.

The closest paper to our work is Cattaneo et al. (2020). They provide a general framework, Random Attention Model (RAM), to test different models of stochastic consideration when preferences are homogeneous. Therefore, their work is better suited for individual stochastic choice data. Cattaneo et al. (2020) impose a set-monotonicity restriction on the probability of considering a set of alternatives given the menu. Namely, they assume that the probability of considering a given set cannot increase if the menu is getting larger. We study the implications of imposing this set-monotonicity constraint as well, but we allow for heterogeneous preferences.

The second assumption we impose is *preference stability*. This condition requires that the marginal distribution of preferences does not depend on the menu. The same stability

---

(2012) (web browsing behavior of consumers when shopping online), Barseghyan et al. (2021) (insurance purchases).

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

assumption is used in McFadden and Richter (1990) and it means that the variation in menus does not change the preferences of DMs. The stability assumption is satisfied in many empirical and theoretical settings (see Sections 3 and 4 for further details). Together with set-monotonicity, it extends the RAM framework to settings with preference heterogeneity and the RUM framework to settings with limited consideration. Importantly, stability puts no restrictions on the dependence structure between menus and consideration, is consistent with statistical dependence between random consideration and random preferences, and together with set-monotonicity makes our framework testable. We show that none of these assumptions alone has empirical content, but together they restrict behavior meaningfully and allow for welfare and counterfactual analysis.

Set-monotonicity and stability are compatible with the behavior of a mixture of DMs, where each DM's behavior is consistent with stochastic limited consideration governed by a RAM (i.e., requiring set-monotonicity). Under this interpretation, stability requires that the heterogeneity of preferences of the population of DMs is independent of the choice set, as in the McFadden and Richter's (1990) stochastic revealed preference framework.

We also demonstrate that set-monotocity and stability are satisfied by several important models of limited consideration with heterogeneous preferences, such as a variant of Manzini and Mariotti's (2007) model of alternative specific consideration with heterogeneous preferences, Brady and Rehbecks's (2016) model of logit attention with random utility, Tversky's (1972) model of elimination by aspects, search and satisficing with random search and random utility (Aguiar et al., 2016), and a version of rational inattention with Shannon's cost of information Caplin et al. (2018).[6]

Our approach differs from previous works that have used enhanced data sets to test for the presence of consideration. In particular, we only need a standard stochastic choice data

---

[6]Aguiar et al. (2021) considers the same primitives but imposes full independence between random preferences and attention. In contrast to this paper, Aguiar et al. (2021) requires the presence of a default alternative and imposes parametric restrictions on the random attention.

set widely used in the discrete-choice literature.[7] Recently, Abaluck and Adams (2021) use structural restrictions on the elasticity behavior of demand to identify consideration sets and preferences. We differ from that work because we do not observe attributes (e.g., prices). Barseghyan et al. (2021) obtain information about parametric distribution of preferences in a domain with attributes variation by introducing a support restriction on possible consideration sets. Our framework does not impose any parametric restrictions on the distribution of preferences, and allows both shape and support restrictions on consideration probabilities. Kashaev and Lazzati (2021) develop a dynamic model of discrete choice that incorporates peer effects into random consideration sets. They identify preferences and consideration probabilities in a fixed menu settings by using variation in choices of peers. We assume menu variation and do not have access to panel data.

Aguiar ⓡ Kashaev (2021) study nonparametric identification and estimation of the distribution of consideration sets and preferences without menu variation in panel data settings. Dardanoni et al. (2020b) provide identification of the consideration probabilities given a known distribution of preferences in a fixed menu. They also consider grouped data sets where three instances of choice of the same consumers is observed to enhance identification. We assume menu variation, do not need to know the distribution of preferences, and do not use enriched stochastic choice data sets. More recently, Dardanoni et al. (2020a) provide identification arguments for both preferences and cognition heterogeneity (including consideration probabilities) in mixture data sets. In contrast to our work, their method requires observing the joint distribution of choice over different menus. Also, their results are focused on parametric heterogeneity.

The paper is organized as follows. Section 2 introduces our general framework. In Section 3, we provide several justifications of the stability assumption. Section 4 shows that our framework generalizes several important models of limited consideration. Section 5 provides

---

[7]For examples of enriched data sets that identify limited consideration, see Reutskaja et al. (2011) (eye-tracking data); Honka et al. (2017) and Draganska and Klapper (2011) (additional surveys); Kawaguchi et al. (2016) and Conlon and Mortimer (2013) (variation in product availability); Dehmamy and Otter (2014) and Huang and Bronnenberg (2018) (variations in quantity purchased and products purchased); and Gabaix et al. (2006) (mouse-tracking data).

the characterization of our model which is amenable for statistical testing, and discusses the computational aspects of our model. Section 6 studies the implications of our model for preference revelation and counterfactual welfare analysis. Section 7 concludes our paper. All proofs can be found in Appendix A.

2. Model

Let X be a finite choice set. The collection of choice sets (menus) is denoted by a nonempty subset of the power set $\mathcal{A} \subseteq 2^X \setminus \{\emptyset\}$. We define the stochastic choice function $\rho_A \in \Delta(A)$, where $\Delta(A)$ denotes the set of all probability distributions on A, for $A \in \mathcal{A}$ such that $\rho_A(a)$ denotes the probability of choosing $a \in A$. The stochastic choice data set is the vector $\rho = (\rho_A)_{A \in \mathcal{A}}$. We call a stochastic choice data set complete if $\mathcal{A} = 2^X \setminus \{\emptyset\}$ and incomplete otherwise.

We let $U \subseteq X \times X$ be the set of linear orders (strict preference relations) defined on X. The typical element will be denoted by $\succ \in U$. Note that U can be restricted to any subset of linear orders exhibiting some property such as single-crossing (Apesteguia et al., 2017) or expected utility. Our theory applies to these restrictions without changes. However, we present our results for the unrestricted U to maximize generality.

Within our framework, DMs may exhibit limited consideration. DMs exhibit limited consideration when they maximize their preferences in a strict subset of the menu. This strict subset is called a consideration set. We model limited consideration using the notion of consideration filters.

Definition 1 (Consideration Filter)**.** We say that $\phi : 2^X \setminus \{\emptyset\} \to 2^X$ is a feasible (consideration)

filter if there exists $D \in 2^X \setminus \{\emptyset\}$ such that

$$\phi(A) = \begin{cases} D, & D \subseteq A, \\ \emptyset, & \text{otherwise} \end{cases}$$

for all $A \in 2^X \setminus \{\emptyset\}$.

Let Φ be a finite collection of all feasible filters. The typical element of it will be denoted by $\phi \in \Phi$.

We consider a random attention and utility model (RAUM). A *behavioral type* of this model is determined by a pair of preferences and filter $(\succ, \phi) \in U \times \Phi$. A RAUM rule $\pi = (\pi_A)_{A \in 2^X \setminus \{\emptyset\}}$ is a collection of probability distribution over preferences and filters $\pi_A \in \Delta(U \times \Phi)$ such that $\pi_A(\succ, \phi) = 0$ whenever $\phi(A) = \emptyset$ for all choice sets $A \in 2^X \setminus \{\emptyset\}$.

Definition 2. A stochastic choice data set ρ admits a RAUM representation π if

$$\rho_A(a) = \sum_{(\succ,\phi)\in U\times\Phi} \pi_A(\succ, \phi)\mathbb{1}\left(a \in \phi(A),\ a \succ b,\ \forall b \in \phi(A) \setminus \{a\} \right),$$

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

RAUM is so general that it does not have any empirical content. That is, feasible filters together with unrestricted (possibly menu dependent) distribution of preferences are permissive enough to explain any behavior. Hence, without further constraints, it is impossible to falsify RAUM or to recover the (marginal) distribution of preferences of a population of DMs (i.e., $\pi_A^*(\succ) = \sum_\phi \pi_A(\succ, \phi)$). We impose the following stability constraint on the RAUM representation π.

Assumption 1 (Stability). There exists $\pi^* \in \Delta(U)$ such that $\pi_A^*(\succ) = \pi^*(\succ)$ for any $A \in 2^X \setminus \{\emptyset\}$ and $\succ \in U$.

Note that stability is equivalent to requiring that $\pi_A^*(\succ) = \pi_B^*(\succ)$ for any $A, B \in 2^X \setminus \{\emptyset\}$ and $\succ \in U$, thus, justifying its name.

One interpretation of stability is that it restricts limited consideration such that the marginal distribution of preferences of the general RAUM is equivalent to the *true* distribution of heterogeneous preferences in the population. The true distribution of preferences is the distribution on U that controls behavior in the counterfactual situation of absence of limited consideration. Stability does not require the knowledge of such distribution. Moreover, our stability assumption is a natural analogue, within our more general framework, of the assumption of preference stability in the stochastic rationality model of McFadden and Richter (1990).

Importantly, stability is consistent with stochastic dependence between consideration filters and random preferences. We only require that the (marginal) distribution of preferences remains the same across exogenously given menus of alternatives. We further explore limitations of the stability assumption in the next section (Section 3).

Even under stability, limited consideration has to be further restricted to have empirical bite as we will show in Proposition 1. Here, we follow Cattaneo et al. (2020) and impose the following restriction.

Assumption 2 (Set-monotonicity). For any $\succ$, ϕ, A, and B such that $A \subseteq B$ and $\phi(A) \neq \emptyset$, it must be that $\pi_A(\phi| \succ) \geq \pi_B(\phi| \succ)$.

Set-monotonicity means that the conditional probability of a given filter, ϕ, conditional on a preference type, $\succ$, cannot increase as the menu expands. That is, DMs will pay more attention to a set when the menu of alternatives is smaller. Intuitively, larger menus have a higher opportunity cost of consideration. Cattaneo et al. (2020) show that many models of random consideration satisfy set-monotonicity.

The interaction of set-monotonicity and stability does not imply independence of consideration and preferences, as the following example demonstrates.

Example 1. Let $X = \{a, b\}$ and $\mathcal{A} = \{\{a\}, \{b\}, \{a, b\}\}$. Let $a \succ_1 b$ and $b \succ_2 a$, and assume that only two filters below realize with nonzero probability:

$$\phi_1(A) = \begin{cases} \{a\}, & \text{if } A \in \{\{a\}, \{a, b\}\}, \\ \emptyset, & \text{otherwise,} \end{cases} \qquad \phi_2(A) = \begin{cases} \{b\}, & \text{if } A \in \{\{b\}, \{a, b\}\}, \\ \emptyset, & \text{otherwise.} \end{cases}$$

Consider the following distributions $\pi_A(\succ, \phi)$ over the above two preference orders and two filters for different menus:

$\pi_A(\succ, \phi)$	$(\succ_1, \phi_1)$	$(\succ_1, \phi_2)$	$(\succ_2, \phi_1)$	$(\succ_2, \phi_2)$
$\{a, b\}$	1/3	1/6	1/6	1/3
$\{a\}$	1/2	0	1/2	0
$\{b\}$	0	1/2	0	1/2

Note that $\pi_A(\succ_1) = \pi_A(\succ_2) = 1/2$ for all A. However,

$$\frac{1}{3} = \pi_{\{a,b\}}(\succ_1, \phi_1) \neq \pi_{\{a,b\}}(\succ_1) \sum_{i=1}^{2} \pi_{\{a,b\}}(\succ_i, \phi_1) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

That is, preferences and filters are not independent. Moreover, set-monotonicity is also satisfied. For instance, $\pi_{\{a\}}(\phi_1| \succ_1) = 1 \geq 2/3 = \pi_{\{a,b\}}(\phi_1| \succ_1)$.

The next proposition qualifies the importance of stability and set-monotonicity working together. Neither of these restrictions alone are enough for empirical relevance of the model. However, when they are combined together, the model becomes falsifiable even with limited menu variation. We see the combination of these two restrictions as a baseline of empirical content that makes our study empirically meaningful.

Proposition 1. The following statements are true:

(i) Any ρ admits a stable RAUM representation.

(ii) Any ρ admits a set-monotone RAUM representation.

(iii) There exists an incomplete ρ that does not admit a set-monotone and stable RAUM representation.

Here we provide a sketch of the proof of (iii). We construct an incomplete data set (i.e., $\mathcal{A} \neq 2^X \setminus \{\emptyset\}$) that does not admit a set-monotone and stable RAUM. Let $X = \{a, b, c, d\}$ and

$$\mathcal{A} = \{\{a,b\}, \{a,c\}, \{b,d\}, \{a,b,d\}, \{a,c,d\}, \{b,c,d\}\}.$$

Suppose the observed ρ is as follows

	$\{a,b\}$	$\{a,c\}$	$\{b,d\}$	$\{a,b,d\}$	$\{a,c,d\}$	$\{b,c,d\}$
$\rho_A(a)$	1	1	-	0	0	-
$\rho_A(b)$	0	-	1	1	-	α_b
$\rho_A(c)$	-	0	-	-	1	α_c
$\rho_A(d)$	-	-	0	0	0	α_d

where $\alpha_d > 0$. Consider the pair $\{a, b\}$ and $\{a, b, d\}$. From observing $\rho_{\{a,b\}}(b) = 0$ and $\rho_{\{a,b,d\}}(b) = 1$, we can conclude that $b \succ d$ with probability 1 or b and d are never considered together. Similarly, from observing $\rho_{\{a,c\}}(c) = 0$ and $\rho_{\{a,c,d\}}(c) = 1$, we can make analogous conclusion about c and d. As a result, the fact that $\alpha_d > 0$ then implies that the probability of considering the singleton consideration set $\{d\}$ must be nonzero (otherwise d is either never considered or dominated by b or c). But the latter is impossible because in menu $\{b, d\}$ option d is never chosen. The formal details of the sketch above can be found in Appendix A.1.

3. Stability of Preferences as Structured Heterogeneity in a Population of Inattentive DMs

The set-monotonicity property is well-understood and justified due to Cattaneo et al. (2020). Here, we provide several justifications of stability as a reasonable assumption for the RAUM representation.

One possible interpretation of the set-monotone and stable RAUM representation is that it represents a mixture of the behavior of a population of DMs, where the behavior of each DM is consistent with RAM.

Definition 3 (Random Attention Model, RAM, Cattaneo et al., 2020)**.** A stochastic choice data set ρ admits a RAM representation if there exist a preference order $\succ$ and a collection of distributions over consideration filters (an attention rule) $\lambda_\succ = (\lambda_{A,\succ})_{A \in \mathcal{A}} \in \Delta(\Phi)^{|\mathcal{A}|}$, where $|\mathcal{A}|$ is the cardinality of $\mathcal{A}$, such that $\lambda_{A,\succ}(\phi) = 0$ whenever $\phi(A) = \emptyset$; $\lambda_{A,\succ}(\phi) \geq \lambda_{B,\succ}(\phi)$ for all $A, B \in \mathcal{A}$ such that $A \subseteq B$ and $\phi(A) \neq \emptyset$; and

$$\rho_A(a) = \sum_\phi \lambda_{A,\succ}(\phi) \mathbb{1}\left(a \in \phi(A),\ a \succ b,\ \forall b \in \phi(A) \setminus \{a\} \right)$$

for each $A \in \mathcal{A}$.

To better understand the relation between RAM and RAUM, consider the following data generating process for ρ. Fix some distribution over preference orders $\pi^* \in \Delta(U)$ and some collection of attention rules for all possible preference orders $\{\lambda_\succ\}_{\succ \in U}$. Every DM draws a preference order $\succ$ from π^* independently of other DMs, as in the McFadden and Richter (1990)'s framework. Given the preference order $\succ$ and attention rule $\lambda_\succ$, the DM chooses alternatives from A according to the RAM rule (that induces a probability of choice)

$$\rho_{A,\succ}(a) = \sum_\phi \lambda_{A,\succ}(\phi) \mathbb{1}\left(a \in \phi(A),\ a \succ b,\ \forall b \in \phi(A) \setminus \{a\} \right).$$

Let ρ be a mixture of the above RAM rules weighted by π^*. That is,

$$\rho_A(a) = \sum_\succ \pi^*(\succ)\rho_{A,\succ}(a)$$

for each $A \in \mathcal{A}$. Note that this data set admits a set-monotone and stable RAUM representation π with $\pi_A(\succ, \phi) = \lambda_{A,\succ}(\phi)\pi^*(\succ)$ for all $\succ$, λ, and A. Moreover, by construction, any ρ that admits a set-monotone and stable RAUM π is also a mixture of RAM rules induced by π^* and attention rules $\{\pi_A(\cdot \mid \succ)\}_{\succ \in U}$. In other words, the set-monotone and stable RAUM framework is an extension of the RAM framework to heterogeneous preferences structured as in the McFadden and Richter (1990)'s framework.

Alternatively, since ρ admits a RUM representation if there exists a distribution over preference orders $\pi^* \in \Delta(U)$ such that

$$\rho_A(a) = \sum_\succ \pi^*(\succ)\mathbb{1}\left(a \succ b, \, \forall b \in A \setminus \{a\} \right),$$

we can think of RUM rules as mixtures of RAM rules obtained from particular degenerate attention rules (i.e., $\lambda_{A,\succ}(\phi) = \mathbb{1}\left(\phi(A) = A\right)$). Hence, the set-monotone and stable RAUM framework is an extension of the RUM framework to heterogeneous consideration filters, as in the Cattaneo et al. (2020)'s framework.

It is important to point out that in this generalization of the RAM framework, which allows for heterogeneous preferences, we use the same stability assumption as in the classical stochastic revealed preference framework of McFadden and Richter (1990). Kitamura and Stoye (2018) have given a modern interpretation to this assumption as an exogeneity restriction requiring preferences and menus to be independent. There are many settings where this assumption is satisfied: (i) experiments, where the experimenter exogenously varies the menus (Aguiar et al., 2021); (ii) environments with predetermined choice sets in time (e.g., such as modes of transport or elections McFadden, 1986); (iii) choice problems with frequently-purchased and

inexpensive products (Lu, 2021).[8] Moreover, the stability assumption has been the standard in decision theory with menu variation since the work of Falmagne (1978).

An important reason why the distribution of preferences is assumed to be independent of menus is that this restriction allows for well-defined and informative welfare and counterfactual choice analysis out-of-sample (i.e., in menus that are not part of the data set). This stable distribution over preferences or types can be interpreted as a **true preference distribution**. This follows the tradition of the seminal consideration set papers in assuming that each DM is endowed with a menu independent strict preference relation (Masatlioglu et al., 2012, Manzini and Mariotti, 2014, Cattaneo et al., 2020).

The stability assumption may not be suitable for all choice situations. A classical example arises in survey data, such as the application studied in Kitamura and Stoye (2018). Stability may fail because income, which determines the budget faced by a DM, may be correlated with preferences. In these situations, Kitamura and Stoye (2018) suggest using the control function approach (Blundell and Powell, 2001, Imbens and Newey, 2009) to suitably modify the stochastic revealed preference framework in McFadden and Richter (1990). A study of the connection between their solution to the problem of endogeneity and our framework is left for future work.

4. Examples of RAUM

Notwithstanding the apparent restrictiveness of the stability and set-monotonicity assumptions, RAUM generalizes a wide variety of models of limited consideration and random utility, allowing for correlation between attention and preferences, as showcased in the next examples.

---

[8]Note that if there are observable covariates (e.g., product characteristics), then our analysis goes through after conditioning on such covariates. In that case, stability would require preferences to be independent of menus conditional on observed covariates.

Attention-index Models

Given $\succ$, an attention-index $\eta_\succ : 2^X \to \mathbb{R}_+$ such that $\eta_\succ(\emptyset) = 0$ is a capacity over subsets of X capturing how enticing they are conditional on a given preference type. The following models of consideration are examples of rules that are governed by attention-indexes. These type of models are studied in Aguiar et al. (2021) with an independence assumption between preferences and consideration. Abaluck and Adams (2021) also uses these type of models in a different domain with the same independence assumption. Here, we show that models with an attention-index consideration that depends on heterogeneous preferences admit a set-monotone and stable RAUM representation.

Example 2 (Logit Attention, LA, Brady and Rehbeck (2016)). Consider $X = \{a, b, c\}$. Assume that there is a vector of random utilities $\mathbf{u} = (\mathbf{u}_y)_{y \in X} \in \mathcal{U}$ associated with X that has a continuous cumulative distribution function (c.d.f.) $F_\mathbf{u}$. The random vector $\mathbf{u}$ induces a distribution over the preference orders π^* such that

$$\pi_A^*(\succ) = \int_\mathcal{U} \mathbb{1}\left(u_{y_1} > u_{y_2} > u_{y_3} \right) dF_\mathbf{u}(u),$$

for all A, and all y_1, y_2, and y_3 in X such that $y_1 \succ y_2 \succ y_3$.[9] For every $\succ$ such that $y_1 \succ y_2 \succ y_3$ (and y_1, y_2, y_3 are distinct elements in X) define an attention index $\eta_\succ : 2^X \to \mathbb{R}$ such that

$$\eta_\succ(D) = \int_{u_{a_1} > u_{a_2} > u_{a_3}} (|\{y \in D \ : \ u_y \geq \kappa\}|) dF(u),$$

for some $\kappa \in \mathbb{R}$ such that $\mathbf{u}_y \geq \kappa$, $y \in X$, with positive probability (i.e., $\eta_\succ(D) > 0$). The attention index $\eta_\succ$ captures the attractiveness of D by computing the cardinality of the set of all alternatives in D that are above a threshold κ.[10] Given $\eta_\succ$ and π^* we construct a rule π

---

[9]Continuity of $F_\mathbf{u}$ implies that $\mathbf{u}_{y_1} = \mathbf{u}_{y_2}$, $y_1 \neq y_2$ with probability zero.

[10]When $\kappa \to -\infty$ the attention-index converges to $|D|$ and there is no dependence on preferences. This special case has been explored in Cattaneo et al. (2020) with homogeneous preferences.

such that $\pi(\succ, \phi) = \pi_A(\phi| \succ)\pi_A^*(\succ)$, where

$$\pi_A(\phi| \succ) = \frac{\eta_\succ(\phi(A))}{\sum_{\phi' \in \Phi} \eta_\succ(\phi'(A))}$$

is a conditional-on-preferences version of the logit attention model of Brady and Rehbeck (2016).

The RAUM rule π constructed in Example 2 is set-monotone and stable. Set-monotonicity follows from the fact that $\pi_A(\cdot| \succ)$ is of the LA form (Brady and Rehbeck, 2016), which is set-monotone (Cattaneo et al., 2020). In addition, stability holds because the c.d.f. $F_\mathbf{u}$ is menu independent. In the absence of limited consideration, the distribution over preferences π^* constitute a random utility rule as in McFadden and Richter (1990). For instance, if $\mathbf{u}_y = w(y) + \boldsymbol{\epsilon}_y$ for some utility function $w : X \rightarrow \mathbb{R}$ and ϵ distributed according to the Gumbel distribution, then the random utility (marginal) takes the logit form and

$$\pi_A^*(\succ) = \frac{e^{w(y_1)}}{\sum_{y \in X} e^{w(y)}} \frac{e^{w(y_2)}}{e^{w(y_2)} + e^{w(y_3)}}.$$

We highlight again that in this example random consideration is correlated with random preferences, which strictly generalizes the logit attention model with heterogeneous preferences studied in Aguiar et al. (2021).

Example 3 (Item Specific Attention, MM,Manzini and Mariotti, 2014, Horan, 2019)**.** Let $X = \{a, b, c\}$ and as in Example 2 let $\mathbf{u} = (\mathbf{u}_y)_{y \in X} \in \mathcal{U}$ be a random utility vector that has a continuous c.d.f. $F_\mathbf{u}$. The random vector $\mathbf{u}$ induces π^* such that

$$\pi_A^*(\succ) = \int_\mathcal{U} \mathbb{1}\left(u_{y_1} > u_{y_2} > u_{y_3} \right) dF_\mathbf{u}(u),$$

for all A, and all y_1, y_2, and y_3 in X such that $y_1 \succ y_2 \succ y_3$. For all distinct $y_1, y_2, y_3 \in X$

and $\succ$ such that $y_1 \succ y_2 \succ y_3$, define $\gamma_\succ : X \to (0,1)$ as

$$\gamma_\succ(y) = \int_{\mathbb{R}, u_{y_1} > u_{y_2} > u_{y_3}} \mathbb{1}\left(u_y \geq \xi \right) dF_{(\boldsymbol{\xi}, \mathbf{u})}(\xi, u),$$

where $\boldsymbol{\xi} \in \mathbb{R}$ is a random threshold that is used to assess how salient the alternative y is to the DM; $F_{(\boldsymbol{\xi}, \mathbf{u})}$ is the joint c.d.f. of $(\boldsymbol{\xi}, \mathbf{u}')'$ and is such that $\gamma_\succ(y) \in (0,1)$ for all $y \in X$ (e.g., $(\boldsymbol{\xi}, \mathbf{u}')'$ is normally distributed random vector). The MM is an LA model, where $\eta_\succ$ satisfies

$$\eta_\succ(D) = \frac{\prod_{a \in X \setminus D} (1 - \gamma_\succ(a)) \prod_{b \in D} \gamma_\succ(b)}{1 - \prod_{c \in X} (1 - \gamma_\succ(c))}$$

for all $A \in 2^X$ and $\succ \in U$ with the convention that $\prod_{b \in \emptyset} \gamma_\succ(b) = 0$.

Example 3 satisfies stability because the distribution of $\mathbf{u}$ does not depend on the menu. In addition, the conditional MM model of attention satisfies set-monotonicity (Cattaneo et al., 2020). Similar to Example 2, we allow for unrestricted correlation between preferences driven by $\mathbf{u}$ and attention driven by $\boldsymbol{\xi}$.

Example 4 (Elimination by Aspects, EBA, Tversky, 1972)**.** Let $X = \{a, b, c\}$ and as in Examples 2-3 and let $\mathbf{u} = (\mathbf{u}_y)_{y \in X} \in \mathcal{U}$ be a random utility vector that has a continuous c.d.f. $F_\mathbf{u}$. The random vector $\mathbf{u}$ induces π^* such that

$$\pi_A^*(\succ) = \int_{\mathcal{U}} \mathbb{1}\left(u_{y_1} > u_{y_2} > u_{y_3} \right) dF_\mathbf{u}(u),$$

for all A, and all y_1, y_2, and y_3 in X such that $y_1 \succ y_2 \succ y_3$. For every $\succ$ such that $y_1 \succ y_2 \succ y_3$ (and y_1, y_2, y_3 are distinct elements in X) define an attention index $\eta_\succ$ such that

$$\eta_\succ(D) = \int_{\xi, u_{y_1} > u_{y_2} > u_{y_3}} \mathbb{1}\left(\sum_{y \in D} (u_y + \xi_{y,D}) \geq 0 \right) dF_{(\boldsymbol{\xi}, \mathbf{u})}(\xi, u),$$

where $\boldsymbol{\xi} = (\boldsymbol{\xi}_{a,D})_{D \in 2^X \setminus \{\emptyset\}, a \in D}$ is a vector of set-choice-specific random distortions of item attractiveness conditional on the consideration set D; $F_{(\boldsymbol{\xi}, \mathbf{u})}$ is the joint c.d.f. of $(\boldsymbol{\xi}', \mathbf{u}')'$. The

16

attention-index in this case is the probability that the consideration set perceived value is nonnegative. The perceived value is computed as the sum of all utilities of the items in the consideration set plus a distortion term. Given $\eta_\succ$ and π^*, we construct a rule π such that $\pi(\succ, \phi) = \pi_A(\phi|\succ)\pi_A^*(\succ)$, where

$$\pi_A(\phi|\succ) = \sum_{C \subseteq X \,:\, C \cap A = \phi(A)} \frac{\eta_\succ(C)}{\sum_{B \subseteq X \,:\, B \cap A \neq \emptyset} \eta_\succ(B)},$$

for all ϕ such that $\phi(A) \neq \emptyset$ will form an elimination by aspects model.[11]

Stability is satisfied in this example because $F_{\mathbf{u}}$ does not depend on the menu. Set-monotonicity is also satisfied because of the EBA restriction on conditional attention (Cattaneo et al., 2020). Correlation between attention and preferences is modeled thought the joint distribution $F_{(\boldsymbol{\xi}, \mathbf{u})}$. Note that in this example attention and preference correlation depends on the consideration set, but not on the menu.

Search and Satisficing

The search and satisficing behavioral model (Simon, 1955) is an important example of a model with dependence between preferences and consideration driven by random utility directly affecting the way choice sets are formed.

Example 5 (Search and Satisficing). For $X = \{a, b, c\}$, let $\mathbf{s} = (\mathbf{s}_y)_{y \in X} \in \mathcal{U}$ be a random search index and $\mathbf{u} = (\mathbf{u}_y)_{y \in X} \in \mathcal{U}$ be a random utility with a joint continuous c.d.f. $F_{(\mathbf{s}, \mathbf{u})}$. The random vector $\mathbf{u}$ induces π^* such that

$$\pi_A^*(\succ) = \int_{\mathcal{U}} \mathbb{1}\left(u_{y_1} > u_{y_2} > u_{y_3}\right) dF_{\mathbf{u}}(u),$$

---

[11]Sometimes we have an alternative that is present in every menu. That is, there exists o such that $o \in A$ for all $A \in \mathcal{A}$. We call this the default alternative. Item specific attention model with default is an MM model with a default such that $\gamma_\succ(o) = 1$ for all $\succ$ and o is the worst item with probability 1. The EBA is also called Random Categorization Rule (RCG) (Aguiar, 2017) when there is a default alternative that always attracts attention. Suleymanov (2018) shows that MM with a default is the intersection of LA and EBA/RCG with a default.

for all A, and all y_1, y_2, and y_3 in X such that $y_1 \succ y_2 \succ y_3$. An alternative y_1 is searched earlier than y_2 if and only if its search index is bigger (i.e., $s_{y_1} \geq s_{y_2}$). All DMs face a common threshold $\tau \in \mathbb{R}$, and their preferences are captured by random utility $\boldsymbol{u}$. For every $\succ$ such that $y_1 \succ y_2 \succ y_3$ (and y_1, y_2, y_3 are distinct elements in X) define $\pi_A(\cdot \mid \succ)$ as the conditional distribution over filters such that $\pi_A(\phi \mid \succ) = 0$ if $\phi(A) = \emptyset$ and for $\phi(A) \neq \emptyset$

$$\pi_A(\phi \mid \succ) = \int_{s, u_{y_1} > u_{y_2} > u_{y_3}} \mathbb{1}\left(\phi(A) = \{y \in A \mid \ \nexists y' \in A : s_y < s_{y'}, u_{y'} \geq \tau\} \right) dF_{(\mathbf{s},\mathbf{u})}(s, u),$$

where $\tau \in \mathbb{R}$ is a threshold that is common for all DMs. We assume $F_{(\mathbf{s},\mathbf{u})}$ is continuous to avoid indifference in preferences and to avoid the case that many items are searched simultaneously. In words, the probability of the consideration set being equal to $\phi(A) = D$ is given by the probability of items in D being searched before the items whose utility value are above the threshold in a menu. If no item is satisficing, then the whole menu A is searched.

The rule $\pi(\succ, \phi) = \pi_A(\phi \mid \succ)\pi_A^*(\succ)$ from Example 5 is set-monotone because the threshold is constant and the distribution of the search index does not depend on the menu.[12] In addition, stability is satisfied because the joint distribution of the random utility and the search index does not depend on the menu. This version of search and satisficing was studied and characterized by Aguiar et al. (2016). The interpretation of this particular RAUM is compatible with ρ being generated by the behavior of a population of DMs. Each of these DMs chooses according to random satisficing behavior (with a fixed distribution of search) as in Aguiar et al. (2016). The preference heterogeneity is governed by π^*.

Correlation between preferences and consideration arises endogenously due to the use of the utility to stop the search process once a satisficing item is found. Caplin et al. (2011) showed that search and satisficing behavior, as in our example, can be written as an optimization problem that produces optimal consideration sets given search costs and expected value of each alternative. In that sense, correlation between preferences and consideration is the result

---

[12]This example fits Example 9 in the supplement of Cattaneo et al. (2020). But it is easy to see that a bigger menu (i.e., $B \supseteq A$) will only decrease the probability of filter ϕ (i.e., $\pi_B(\phi \mid \succ) \leq \pi_A(\phi \mid \succ)$).

of optimizing behavior of DMs.

Set-monotonicity breaks if we allow menu dependent thresholds (Aguiar and Kimya, 2019) or menu dependent search indexes (Aguiar et al., 2016). However, if the threshold is random[13], as in Kovach and Ülkü (2020), then set-monotonicity condition is satisfied if the distribution of the random threshold is menu independent.

Rational Inattention

Caplin et al. (2016) pointed out the relationship between the theory of rational inattention in discrete choice and the theory of consideration sets. Even though the domain of the theory of rational inattention as presented in Caplin et al. (2016) is described in a different domain than the one over which our theory is developed, we show next that some special cases of the rational inattention model admit a set-monotone and stable RAUM representation. The first example of this section follows the consumer problem 1 in Caplin et al. (2016), but in contrast to them, (i) we allow for menu variation, and (ii) we assume that priors change with the menu in a way consistent with Bayes' rule.

Example 6 (Rational Inattention). Let $X = \{a, b, c\}$ be the choice set and the (unobserved) state space be equal to the choice set $\Omega = X$. Consider an individual DM. Let $\mu(y)$, $y \in X$, be the prior that y is of high quality and the rest of alternatives are of low quality. Without loss of generality, assume that $\mu(a) \geq \mu(b) \geq \mu(c)$. Assume that, given menu $A \subseteq X$, the prior belief about the state of the world is updated according to Bayes' rule. That is, the prior that y in menu A is of the high quality is $\mu_A(y) = \mu(y)/\sum_{y' \in A} \mu(y')$.

There is a state dependent utility $u : X \times \Omega \to \{0, 1\}$ such that, $u(a, y) = u_G$ if $a = y$ and otherwise $u(a, y) = u_B$ with $u_G > u_B$. The DM expends attention effort following Shannon's model of rational inattention with a fixed parameter $\lambda > 0$.[14] For tractability, define a net

---

[13]τ becomes random and there is a joint c.d.f. $F_{(\tau, \mathbf{s}, \mathbf{u})}$ governing the model.

[14]We consider several values of λ as a comparative statics analysis, but we assume that the DM has a fixed cost of attention effort.

payoff of identifying the high-quality item that takes into account the attention effort as

$$\delta = \exp\left(\frac{u_G - u_B}{\lambda}\right) - 1.$$

Define also,

$$\delta_1^* = -3 + \frac{1}{\mu(c)}, \quad \delta_2^* = -1 + \frac{\mu(a)}{\mu(c)}, \quad \delta_3^* = -1 + \frac{\mu(b)}{\mu(c)}, \quad \delta_4^* = -1 + \frac{\mu(a)}{\mu(b)}.$$

Note that since $1 - 2\mu(c) = \mu(a) + \mu(b) - \mu(c) \geq \mu(a)$ and $\mu(a) \geq \mu(b) \geq \mu(c)$, it follows that $\delta_1^* \geq \delta_2^* \geq \delta_3^*$ and $\delta_2^* \geq \delta_4^*$. Caplin et al. (2016) show that rational inattention produces a deterministic consideration set that remains the same across states (i.e., $\pi_A(\phi^* | \succ) = 1$ for all $\succ \in U$ and fixed ϕ^*). Tables 1 and 2 display the deterministic considerations set (i.e., $\phi(A)$ with probability 1) for different values of δ and menu A. Table 1 corresponds to the case when $\delta_3^* \geq \delta_4^*$ (i.e., $\mu(b)/\mu(c) \geq \mu(a)/\mu(b)$). Table 2 considers the case when $\delta_4^* \geq \delta_3^*$.

$A/\delta \in$	$(0, \delta_4^*]$	$(\delta_4^*, \delta_3^*]$	$(\delta_3^*, \delta_2^*]$	$(\delta_2^*, \delta_1^*]$	$(\delta_1^*, +\infty)$
$\{a,b,c\}$	$\{a\}$	$\{a,b\}$	$\{a,b\}$	$\{a,b\}$	$\{a,b,c\}$
$\{a,b\}$	$\{a\}$	$\{a,b\}$	$\{a,b\}$	$\{a,b\}$	$\{a,b\}$
$\{b,c\}$	$\{b\}$	$\{b\}$	$\{b,c\}$	$\{b,c\}$	$\{b,c\}$
$\{a,c\}$	$\{a\}$	$\{a\}$	$\{a\}$	$\{a,c\}$	$\{a,c\}$

Table 1 – Consideration sets for different menus A and δ when $\delta_3^* \geq \delta_4^*$.

$A/\delta \in$	$(0, \delta_3^*]$	$(\delta_3^*, \delta_4^*]$	$(\delta_4^*, \delta_2^*]$	$(\delta_2^*, \delta_1^*]$	$(\delta_1^*, +\infty)$
$\{a,b,c\}$	$\{a\}$	$\{a\}$	$\{a,b\}$	$\{a,b\}$	$\{a,b,c\}$
$\{a,b\}$	$\{a\}$	$\{a\}$	$\{a,b\}$	$\{a,b\}$	$\{a,b\}$
$\{b,c\}$	$\{b\}$	$\{b,c\}$	$\{b,c\}$	$\{b,c\}$	$\{b,c\}$
$\{a,c\}$	$\{a\}$	$\{a\}$	$\{a\}$	$\{a,c\}$	$\{a,c\}$

Table 2 – Consideration sets for different menus A and δ when $\delta_4^* \geq \delta_3^*$

Tables 1-2 imply that set-monotonicity is satisfied. Indeed, the deterministic consideration set described in each column of these tables satisfy the attention filter property described in Masatlioglu et al. (2012).[15]

---

[15]The attention filter property states that if we remove an item from a menu that is not in the deterministic

Following Caplin et al. (2016), it can be shown that given a consideration set D, the distribution over choices for this rationally inattentive DM in D is such that

$$P_D(y) = \frac{\mu_D(y)(|D| + \delta) - 1}{\delta}.$$

For instance, if $\delta \in (\delta_2^*, \delta_1^*]$ (i.e, $D = \{a, b\}$), then $\rho_{\{a,b,c\}}(a) = P_{\{a,b\}}(a)$, $\rho_{\{a,b,c\}}(b) = P_{\{a,b\}}(b)$, and $\rho_{\{a,b,c\}}(c) = 0$. Stability also holds. Since ρ_A only depends on A via consideration set D, for $\delta \leq \delta_1^*$, it is trivial to construct a menu independent distribution over preference orders over X. Hence, ρ generated by this rational inattention model admits a set-monotone and stable RAUM representation.

When $\delta > \delta_1^*$, DMs consider all alternatives. Hence, in this case, since there are only 3 alternatives, ρ admits a RAUM representation if and only if ρ is regular (Block and Marschak, 1960, Falmagne, 1978). That is, $\rho_A(y) \geq \rho_X(y)$ for all $A \subseteq X$ and $y \in A$. Note that $\rho_{\{a,b\}}(a) = \frac{\mu_{\{a,b\}}(a)(2 + \delta) - 1}{\delta}$ and $\rho_{\{a,b,c\}}(a) = \frac{\mu(a)(3 + \delta) - 1}{\delta}$. Hence,

$$\rho_{\{a,b\}}(a) - \rho_{\{a,b,c\}}(a) = \frac{\mu(a)\mu(c)(\delta - \delta_1^*)}{\delta(\mu(a) + \mu(b))} > 0,$$

where the last inequality is implied by $\delta > \delta_1^*$. Similar inequalities can be derived for all other menus and options. For the formal construction of $\pi^*(\succ)$ see Appendix B.

Caplin et al. (2016) remark informally that the consideration sets produced by the Shannon's model of rational inattention satisfy the attention filter property first proposed in Masatlioglu et al. (2012). Deterministic consideration sets that are attention filters satisfy set-monotonicity (Cattaneo et al., 2020). However, this statement is only true under certain assumptions about how the prior is defined and how it changes when menus change. In Example 6, the deterministic consideration set is indeed an attention filter. However, if we replace μ with another prior that arbitrarily depends on the menu, this property will not hold in general.

---

consideration set, then the consideration set in the new menu is the same as the consideration set in the original menu.

Stability also holds in Example 6. Note that randomness in choice for the rational inattentive DM is driven by mistakes due to costly attentional effort. If $\lambda = 0$, then choice will be deterministic and consistent with utility maximization. The fact that stability holds in this setup is entirely due to the fact that Shannon's model of rational inattention admits an additive random utility equivalent representation, as stated in Fosgerau et al. (2017). Similarly, this equivalence will hold only under certain assumptions on the priors and how they change across menus. If priors are fixed, this equivalence holds. In addition, we showed in our example another way to change priors that also makes this equivalence hold. In other words, stability holds here because we restrict the dependence of the priors on the menu, hence making it possible for the implied distribution over preferences π^* to be menu independent as stability requires.

Not all probabilities of choice generated by Shannon's rational inattention model admit a set-monotone and stable RAUM representation. A well-known example in Matějka and McKay (2015) (see problem 4) exhibits Shannon's rational inattention model where priors change across menus such that stability is broken even with full consideration.

We conclude by noting that the example in this section admits neither a RUM representation nor a RAM representation. However, it admits a set-monotone and stable RAUM representation.

5. Characterization of Set-monotone and Stable RAUM

In this section we characterize the set-monotone and stable RAUM in a form amenable to (statistical) testing. In particular, we show that to conclude whether a given data set admits a set-monotone and stable RAUM representation, it suffices to check whether a particular linear program has a solution. This problem is similar to the one in McFadden and Richter

([1990](#)) that characterizes RUM.

Let d_ρ, d_m, and d_r denote the number of entries in ρ, the cardinality of $2^X \setminus \{\emptyset\}$, and the total number of linear restrictions imposed by feasibility, stability, and set-monotonicity on π, respectively. Also, define $g = (\rho', 1'_{d_m}, 0'_{d_r})' \in \mathbb{R}^{d_g}$, where 1_{d_m} is the vector of ones of length d_m and 0_{d_r} is the vector of zeros of length d_r.

Theorem 1. *Given a stochastic choice data set ρ the following are equivalent:*

   (i) *ρ admits a set-monotone and stable RAUM.*

   (ii) *There exists $v \in \mathbb{R}^d_+$, $d < \infty$, such that*

$$g = Gv$$

   where G is a known matrix that consists of -1, 0, 1.

Theorem 1 provides a linear characterization of a set-monotone and stable RAUM. It is important to note that without stability the problem is quadratic since set-monotonicity is imposed on the conditional distribution over filters $\pi_A(\phi| \succ)$. The linearity of our problem makes it amenable to statistical testing using tools in Deb et al. (2018) as we discuss below.

An Illustrative Example

To better understand how the matrix G is constructed, note that G consists of 5 types of linear restrictions on π. First, the observed $\rho_A(a)$ is a mixture of preference-filter-types such that a is considered and is the best among those considered:

$$\rho_A(a) = \sum_{(\succ,\phi)\in U\times\Phi} \pi_A(\succ, \phi)\mathbb{1}\left(a \in \phi(A),\ a \succ b,\ \forall b \in \phi(A) \setminus \{a\} \right).$$

If the choice set is $X = \{a, b\}$, preferences U are such that $a \succ_1 b$ and $b \succ_2 a$, and there are only 3 feasible filters in Φ: ϕ_1, ϕ_2 and ϕ_3 such that $\phi_1(\{a\}) = \phi_1(\{a, b\}) = \{a\}$, $\phi_2(\{b\}) = \phi_2(\{a, b\}) = \{b\}$, $\phi_3(\{a, b\}) = \{a, b\}$, then

$$\pi_{\{a,b\}}(\succ_1, \phi_1) + \pi_{\{a,b\}}(\succ_1, \phi_3) + \pi_{\{a,b\}}(\succ_2, \phi_1) = \rho_{\{a,b\}}(a).$$

Second, the definition of RAUM requires $\pi_A(\succ, \phi) = 0$ whenever $\phi(A) = \emptyset$. For instance, $\pi_{\{b\}}(\succ_1, \phi_1) = 0$.

Third, each π_A sums up to 1. For instance,

$$\pi_{\{a,b\}}(\succ_1, \phi_1) + \pi_{\{a,b\}}(\succ_1, \phi_2) + \pi_{\{a,b\}}(\succ_1, \phi_3) + \pi_{\{a,b\}}(\succ_2, \phi_1) + \pi_{\{a,b\}}(\succ_2, \phi_2) + \pi_{\{a,b\}}(\succ_2, \phi_3) = 1.$$

Fourth, π has to be stable. That is, $\sum_\phi (\pi_A(\succ, \phi) - \pi_B(\succ, \phi)) = 0$ for every $A, B \in \mathcal{A}$. For instance,

$$\pi_{\{a\}}(\succ_1, \phi_1) + \pi_{\{a\}}(\succ_1, \phi_2) - \pi_{\{a,b\}}(\succ_1, \phi_1) - \pi_{\{a,b\}}(\succ_1, \phi_2) = 0.$$

Finally, to impose set-monotonicity note that for $A \subseteq B$, under stability, $\pi_A(\phi \mid \succ) \geq \pi_B(\phi \mid \succ)$ is equivalent to

$$\pi_A(\phi \mid \succ)\pi^*(\succ) = \pi_A(\succ, \phi) \geq \pi_B(\succ, \phi) = \pi_B(\phi \mid \succ)\pi^*(\succ).$$

To turn the inequality into equality, we introduce a slack variable $\bar{v}_{A,B,\succ,\phi} \geq 0$. As a result, we get that

$$\pi_A(\succ, \phi) - \pi_B(\succ, \phi) - \bar{v}_{A,B,\succ,\phi} = 0.$$

For example,

$$\pi_{\{a\}}(\succ_1, \phi_1) - \pi_{\{a,b\}}(\succ_1, \phi_1) - \bar{v}_{\{a\},\{a,b\},\succ_1,\phi_1} = 0.$$

These are 5 types of linear equality restrictions on $v = (\pi', \bar{v}')'$, where $\bar{v} = (\bar{v}_{A,B,\succ,\phi})_{A \subseteq B \in \mathcal{A}, \succ \in U, \phi \in \Phi}$

24

($\bar{v}$ enters the first 4 restrictions with zero coefficients). Combining equations for all menus, choices, filters, and preference orders, we obtain the system of equations in Theorem 1.

Computational Aspects of Testing

We have assumed that we observe ρ. In reality, we have to estimate ρ from a sample of choices. To do this, in our preferred interpretation of RAUM, we need a cross-section of choices of a population of DMs with choice set variation.[16] Once we have the estimator of ρ, $\hat{\rho}$, we can use the testing procedure delineated in Deb et al. (2018), which amounts to solve a (convex)-quadratic problem with linear constraints. Such problems are well-known in the optimization literature and typically easy to solve (Kitamura and Stoye, 2018, Deb et al., 2018). The main computational cost can arise in the computation of the matrix G. The matrix G does not depend on $\hat{\rho}_n$, but its size grows exponentially with the size of the choice set X. For instance, for $|X| = 6$, G has about 1.8 million rows and 4.4 million columns. Fortunately, G is sparse – for $|X| = 6$ less than 0.00012 percent of entries of G are nonzero.[17]

In the extreme case where the choice set has a continuum of alternatives, such as in the framework of Kitamura and Stoye (2018), the construction of the analogue matrix for the special case of RUM can become computational prohibitive. However, this level of complexity does not typically arise in our setup, as we focus on discrete choice with a moderate choice set size.

In addition, recent methodological advances in random column generation in Smeulders et al. (2021) could allow us to handle the potentially large dimensionality in our problem arising from a large choice set. The framework in Smeulders et al. (2021) has been used successfully to speed and scale up the framework in Kitamura and Stoye (2018). It is left for future

---

[16]One example of such data set is the one collected in Aguiar et al. (2021). In that particular sample, there are 4099 independent choices from a choice set with $|X| = 6$ alternatives and $|\mathcal{A}| = 32$ randomly and exogenously assigned menus.

[17]In our simulations, the computation time grows exponentially as well. While for $|X| = 3$ matrix G is computed in approximately $3 \cdot 10^{-4}$ seconds, for $|X| = 6$ it takes about 74 minutes.

research to figure out how to use the random column generation technique in our problem.

Another key difference from Cattaneo et al. (2020) is that we do not require the data set to be complete. We show that verifying the conditions on Theorem 1 is necessary and sufficient to guarantee that there is a set-monotone and stable RAUM representation of the data set. This means that our methodology can be applied to data sets that do not contain full variation in menus, such as those collected in Apesteguia and Ballester (2021).[18]

Another approach to testing is to use the fact that any RAUM rule is a mixture of RAM rules

$$\rho_A(a) = \sum_{\succ} \pi^*(\succ)\rho_{A,\succ}(a)$$

for each $A \in \mathcal{A}$, where $\rho_\succ$ is the RAM rule that corresponds to a given preference order $\succ$ and π^* is the stable distribution over preferences.

Proposition 2. Given a stochastic choice data set ρ the following are equivalent:

(i) ρ admits a set-monotone and stable RAUM.

(ii) There exist $\pi^* \in \Delta(U)$ and a collection of rules $\{\rho_\succ\}_{\succ \in U}$ such that for each $A \in \mathcal{A}$ and $a \in A$

$$\rho_A(a) = \sum_{\succ} \pi^*(\succ)\rho_{A,\succ}(a),$$

and $R_\succ \rho_\succ \leq 0$ for all $\succ \in U$, where $R_\succ$ is the matrix from Theorem 3 in Cattaneo et al. (2020).

In contrast to the linear characterization in Theorem 1, the objective function in the optimization problem implied by Proposition 2 is not convex since ρ depends on unknown π^* and $\rho_\succ$ nonlinearly. However, the dimensionality of this problem is much smaller.

The proof of Proposition 2 is omitted because it follows directly from Theorem 3 in Cattaneo

---

[18]Incompleteness of the data set also leads to substantial decrease in the size of G since it is determined by $|\mathcal{A}|$.

et al. (2020) and the interpretation of our set-monotone and stable RAUM as a mixture of DMs whose behavior is consistent with RAM.

6. Partial Identification of Preferences and Welfare Analysis

Partial Identification of Preferences

Although RAUM is falsifiable, given that preferences are not homogeneous, it is important to learn whether RAUM reveals anything about preferences. In this section, we show that RAUM reveals information about the distribution of preferences in population in some data sets.

We say that ρ is regular if $\rho_A(a) \geq \rho_B(a)$ for all $A \subseteq B$ and $a \in A$. Otherwise, we call ρ irregular.

To formalize the notion of revelation of preferences, let R_ρ be a set of all set-monotone and stable RAUM representations of ρ. That is,

$$R_\rho = \{\pi \in \Delta(U \times \Phi) \ : \ \pi \text{ is set-monotone and stable RAUM rule and } \rho \text{ admits } \pi\}.$$

Next define the identified set for preference distributions implied by R_ρ as

$$\Pi(R_\rho) = \left\{\pi^* \in \Delta(U) \ : \ \text{exists } \pi \in R_\rho \text{ such that } \pi^*(\succ) = \sum_\phi \pi_A(\succ, \phi) \text{ for all } A, \succ\right\}.$$

Proposition 3. $\Pi(R_\rho)$ is a strict subset of $\Delta(U)$ for any irregular ρ. In particular, $\pi^*(\succ) < 1$ for any $\succ \in U$ such that $b \succ a$ for all $b \in X$ and some a in

$$\{a \in X \ : \ \rho_B(a) > \rho_A(a), \ a \in A \subseteq B, \ A, B \in \mathcal{A}\}.$$

Proposition 3 states that irregular data is always informative about preferences. For example, if $\rho_{\{a,b,c\}}(a) > \rho_{\{a,b\}}(a)$, then we can conclude that a cannot be the worst alternative with probability 1. There must exist a DM who ranks a above something else. Since set-monotone and stable RAUM is a generalization of RAM, the conclusion of Proposition 3 is a generalization of the results in Cattaneo et al. (2020) for heterogeneous preferences.[19]

We conclude this section by noting that, since RAUM is a strict generalization of RAM, in general, the distribution over preferences can not be pined down uniquely without imposing more restrictions. Moreover, if there are several different preference orders that can explain the observed data set under RAM, then any distribution over these orders can explain the data set under RAUM. In particular, for regular data sets, nothing can be learned about preferences under RAUM, since there is no revelation of information about preferences under RAM (Cattaneo et al., 2020).

Out-of-Sample Predictions and Counterfactual Analysis

Similar to Kitamura and Stoye (2019) who analyzed RUM, we can use our framework to conduct out-of-sample predictions and counterfactual analysis within the RAUM framework. In particular, we are interested in (i) predicting the choices of DMs in menus that are not observed in the data; and (ii) measuring welfare losses due to inattention as the fraction of individuals that do not achieve their first best due to limited consideration (i.e. the first best is the counterfactual situation where DMs consider the whole menu instead of its subsets).

By Theorem 1, we know that a given data set ρ, defined on the collection of menus $\mathcal{A}$, admits a set-monotone and stable RAUM representation π if and only if there exists $v \in \mathbb{R}_+^d$ such that

$$g = Gv. \tag{1}$$

---

[19]Note that $\Pi(\mathrm{R}_\rho)$ and $\Delta(U)$ are closed sets, hence, the difference between $\Delta(U)$ and $\Pi(\mathrm{R}_\rho)$ has a positive Lebesgue measure.

Recall that if v is a solution to Equation (1), then the first $|U| \cdot (2^{|X|} - 1) \cdot |\Phi|$ components of v correspond to the set-monotone and stable RAUM rule. We will abuse notation and use π^v to denote this rule. This π^v includes rules for all menus (even those that are not observed in the data). Thus, the set of solutions to Equation (1) characterizes all possible set-monotone and stable RAUM rules that can be admitted by given ρ.

To make out-of-sample predictions, note that given v (hence, π^v) that solves Equation (1), $\pi_A^v(\succ, \phi)$ will be a fraction of DMs in the population with preferences $\succ$ and filter ϕ for any A that is consistent with ρ. So we can compute the maximal out-of-sample probability of observing a from $A \notin \mathcal{A}$ (and $A \subseteq X$) that was not observed in ρ as

$$\max_{\pi^v} \sum_{(\succ, \phi) \in U \times \Phi} \pi_A^v(\succ, \phi) \mathbb{1}\left(a \in \phi(A),\ a \succ b,\ \forall b \in \phi(A) \setminus \{a\} \right),$$

$$\text{s.t. } Gv = g.$$

The minimal out-of-sample probability of observing a from $A \notin \mathcal{A}$ (and $A \subseteq X$) can be computed in the same way by replacing the max by the min operator.

Next, to measure welfare losses due to inattention as the fraction of individuals that do not achieve their first best due to limited consideration, for any $A \in \mathcal{A}$ such that $\sum_{\succ'} \pi_A(\succ', \phi) \neq 0$ define

$$\Delta P_A^{\mathrm{c}}(\pi^v) = \sum_{a \in A, \phi \in \Phi, \succ \in U} \pi_A^v(\succ, \phi)\Big[\mathbb{1}\left(a \in A,\ a \succ b,\ \forall b \in A \setminus \{a\} \right)$$

$$- \mathbb{1}\left(a \in \phi(A),\ a \succ b,\ \forall b \in \phi(A) \setminus \{a\} \right)\Big].$$

$\Delta P_A(\pi^v)$ measures that counterfactual fraction of DMs endowed with menu A who are strictly better off from considering all available alternatives in menu A. Hence,

$$\max_{v \in \mathbb{R}_+^d} \Delta P_A(\pi^v)$$

$$\text{s.t. } Gv = g$$

will be the maximal fraction of DMs, who faced menu A, that would be better off if they consider all alternatives in A. If one is interested in the total effect it suffices to replace $\Delta P_A(v)$ by $\sum_{A \in \mathcal{A}} \Delta P_A(v)$ in the last optimization. Similar to the out-of-sample predictions, the lower bound of $\Delta P_A(\pi^v)$ can be computed by replacing the max by the min operator.

Note that the counterfactual fully attentive behavior, assuming that each DM is fully attentive or considers the whole menu, is equivalent to the DMs behaving consistently with RUM governed by the true preference distribution (under our preferred interpretation of RAUM) $\pi^{*,v}$, where $\pi^{*,v}(\succ) = \sum_{\phi} \pi^v(\succ, \phi)$.

We finish this section by remarking that our ordinal approach puts no restriction on the random utility distribution. We do not need to integrate over (unknown) distributions of parameters of high order polynomial approximations of the utility function or the consideration probability. Instead, by taking a purely revealed preference approach, our out-of-sample predictions and counterfactual welfare analysis require solving a linear program that delivers sharp bounds (i.e., a point is within bounds if and only if there exists a data generating process that is consistent with observed data and the point).

7. Conclusions

We have extended the classical stochastic revealed preference methodology in McFadden and Richter (1990) for finite sets to allow for limited consideration. Our model allows for heterogeneous preferences that are correlated with consideration sets. We assume that consideration satisfies the set-monotonicity assumption of Cattaneo et al. (2020). We also introduce a new condition, called stability, that requires the marginal distribution of preferences

to be independent of menus. We show that this new restriction is satisfied in many theoretical and empirical settings. The proposed model and conditions are amenable to statistical testing using procedures proposed in Deb et al. (2018) and Kalouptsidi et al. (2020).

References

Abaluck, Jason and Abi Adams (2021) "What do consumers consider before they choose? Identification from asymmetric demand responses," *The Quarterly Journal of Economics*, Accepted. 1, 4

Aguiar, Victor H (2017) "Random categorization and bounded rationality," *Economics Letters*, 159, 46–52. 5, 11

Aguiar, Victor H., Maria Jose Boccardi, and Mark Dean (2016) "Satisficing and stochastic choice," *Journal of Economic Theory*, 166, 445 – 482, https://doi.org/10.1016/j.jet.2016.08.008. 2, 5, 1, 4

Aguiar, Victor H, Maria Jose Boccardi, Nail Kashaev, and Jeongbin Kim (2021) "Random Utility and Limited Consideration." 6, 3, 4, 4, 16

Aguiar, Victor H Ⓡ Nail Kashaev (2021) "Identification and Estimation of Discrete Choice Models with Unobserved Choice Sets." 1

Aguiar, Victor H and Mert Kimya (2019) "Adaptive stochastic search," *Journal of Mathematical Economics*, 81, 74–83. 4

Apesteguia, Jose, Miguel A Ballester, and Jay Lu (2017) "Single-Crossing Random Utility Models," *Econometrica*, 85 (2), 661–674. 2

Apesteguia, Jose and Miguel Ángel Ballester (2021) "Separating Predicted Randomness from Noise," *Journal of The European Economic Association*, forthcoming. 5

Barseghyan, Levon, Maura Coughlin, Francesca Molinari, and Joshua C Teitelbaum (2021) "Heterogeneous choice sets and preferences," *Econometrica*, forthcoming. 4, 1

Block, Henry David and Jacob Marschak (1960) "Random orderings and stochastic theories of responses," *Contributions to probability and statistics*, 2, 97–132. 1, 6

Blundell, Richard and James L Powell (2001) "Endogeneity in nonparametric and semiparametric regression models." 3

Brady, Richard L and John Rehbeck (2016) "Menu-Dependent Stochastic Feasibility," *Econometrica*, 84 (3), 1203–1223. 5, 1, 2, 4

Caplin, Andrew, Mark Dean, and John Leahy (2016) "Rational inattention, optimal consideration sets and stochastic choice,"Technical report, Working paper. 5, 4, 6, 6, 4, B

———— (2018) "Rational inattention, optimal consideration sets and stochastic choice,"Technical report. 1

Caplin, Andrew, Mark Dean, and Daniel Martin (2011) "Search and satisficing," *American Economic Review*, 101 (7), 2899–2922. 4

Cattaneo, Matias D, Xinwei Ma, Yusufcan Masatlioglu, and Elchin Suleymanov (2020) "A random attention model," *Journal of Political Economy*, 128 (7), 2796–2836. (document), 1, 2, 2, 3, 3, 10, 4, 4, 4, 12, 4, 5, (ii), 5, 6, 7

Conlon, Christopher T and Julie Holland Mortimer (2013) "Demand estimation under incomplete product availability," *American Economic Journal: Microeconomics*, 5 (4), 1–30. 7

Dardanoni, Valentino, Paola Manzini, Marco Mariotti, Henrik Petri, and Christopher J Tyson (2020a) "Mixture Choice Data: Revealing Preferences and Cognition." 1

Dardanoni, Valentino, Paola Manzini, Marco Mariotti, and Christopher J Tyson (2020b) "Inferring cognitive heterogeneity from aggregate choices," *Econometrica*, 88 (3), 1269–1296. 1

De Los Santos, Babur, Ali Hortaçsu, and Matthijs R. Wildenbeest (2012) "Testing Models of Consumer Search Using Data on Web Browsing and Purchasing Behavior," *American Economic Review*, 102 (6), 2955–80, 10.1257/aer.102.6.2955. 4

Deb, Rahul, Yuichi Kitamura, John K-H Quah, and Jörg Stoye (2018) "Revealed price preference: theory and empirical analysis," *arXiv preprint arXiv:1801.02702.* 5, 7

Dehmamy, Keyvan and Thomas Otter (2014) "Utility and attention-a structural model of consideration," *Available at SSRN.* 7

Draganska, Michaela and Daniel Klapper (2011) "Choice set heterogeneity and the role of advertising: An analysis with micro and macro data," *Journal of Marketing Research*, 48 (4), 653–669. 7

Falmagne, J. C. (1978) "A representation theorem for finite random scale systems," *Journal of Mathematical Psychology*, 18 (1), 52–72, 10.1016/0022-2496(78)90048-2. 1, 3, 6

Fishburn, Peter C (1998) "Stochastic utility," *Handbook of utility theory*, 1, 273–318. 2

Fosgerau, Mogens, Emerson Melo, André de Palma, and Matthew Shum (2017) "Discrete choice and rational inattention: A general equivalence result." 4

Gabaix, Xavier, David Laibson, Guillermo Moloche, and Stephen Weinberg (2006) "Costly information acquisition: Experimental analysis of a boundedly rational model," *American Economic Review*, 96 (4), 1043–1068. 7

Heiss, Florian, Daniel McFadden, Joachim Winter, Amelie Wuppermann, and Bo Zhou (2016) "Inattention and switching costs as sources of inertia in medicare part d,"Technical report, National Bureau of Economic Research. 4

Ho, Kate, Joseph Hogan, and Fiona Scott Morton (2017) "The impact of consumer inattention on insurer pricing in the Medicare Part D program," *The RAND Journal of Economics*, 48 (4), 877–905. 4

Honka, Elisabeth (2014) "Quantifying search and switching costs in the US auto insurance industry," *The RAND Journal of Economics*, 45 (4), 847–884, 10.1111/1756-2171.12073. 4

Honka, Elisabeth, Ali Hortaçsu, and Maria Ana Vitorino (2017) "Advertising, consumer awareness, and choice: Evidence from the us banking industry," *The RAND Journal of Economics*, 48 (3), 611–646. 4, 7

Horan, Sean (2016) "A simple model of two-stage choice," *Journal of Economic Theory*, 162, 372–406. 3

———— (2019) "Random consideration and choice: A case study of "default" options," *Mathematical Social Sciences*, 102, 73–84. 5, 3

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

Huang, Yufeng and Bart J Bronnenberg (2018) "Pennies for your thoughts: Costly product consideration and purchase quantity thresholds," *Marketing Science*. 7

Imbens, Guido W and Whitney K Newey (2009) "Identification and estimation of triangular simultaneous equations models without additivity," *Econometrica*, 77 (5), 1481–1512. 3

Kalouptsidi, Myrto, Yuichi Kitamura, Lucas Lima, and Eduardo A Souza-Rodrigues (2020)

"Partial identification and inference for dynamic models and counterfactuals,"Technical report, National Bureau of Economic Research. 7

Kashaev, Nail and Natalia Lazzati (2021) "Peer Effects in Random Consideration Sets," *arXiv preprint arXiv:1904.06742.* 1

Kawaguchi, Kohei, Kosuke Uetake, and Yasutora Watanabe (2016) "Identifying Consumer Attention: A Product-Availability Approach." 7

Kitamura, Yuichi and Jörg Stoye (2018) "Nonparametric analysis of random utility models," *Econometrica*, 86 (6), 1883–1909. 3, 5

——— (2019) "Nonparametric counterfactuals in random utility models," *arXiv preprint arXiv:1902.08350.* 1, 6

Kovach, Matthew and Levent Ülkü (2020) "Satisficing with a variable threshold," *Journal of Mathematical Economics*, 87, 67–76. 5, 4

Lleras, Juan Sebastian, Yusufcan Masatlioglu, Daisuke Nakajima, and Erkut Y Ozbay (2017) "When more is less: Limited consideration," *Journal of Economic Theory*, 170, 70–85. 5

Lu, Zhentong (2021) "Estimating multinomial choice models with unobserved choice sets," *Journal of Econometrics.* 3

Manzini, Paola and Marco Mariotti (2007) "Sequentially rationalizable choice," *American Economic Review*, 97 (5), 1824–1839. 3, 1

——— (2014) "Stochastic choice and consideration sets," *Econometrica*, 82 (3), 1153–1176. 1, 3, 3

Masatlioglu, Yusufcan, Daisuke Nakajima, and Erkut Y. Ozbay (2012) "Revealed Attention," *American Economic Review*, 102 (5), 2183–2205, 10.1257/aer.102.5.2183. 1, 3, 6, 4

Matějka, Filip and Alisdair McKay (2015) "Rational inattention to discrete choices: A new foundation for the multinomial logit model," *American Economic Review*, 105 (1), 272–98. 4

McFadden, Daniel (1986) "The choice theory approach to market research," *Marketing science*, 5 (4), 275–297. 3

McFadden, Daniel and Marcel K Richter (1990) "Stochastic rationality and revealed stochastic preference," *Preferences, Uncertainty, and Optimality, Essays in Honor of Leo Hurwicz, Westview Press: Boulder, CO*, 161–186. (document), 1, 2, 3, 4, 5, 7

Ray, Debraj ⓡ Arthur Robson (2018) "Certified random: A new order for coauthorship," *American Economic Review*, 108 (2), 489–520. ∗

Reutskaja, Elena, Rosemarie Nagel, Colin F Camerer, and Antonio Rangel (2011) "Search dynamics in consumer choice under time pressure: An eye-tracking study," *American Economic Review*, 101 (2), 900–926. 7

Simon, Herbert A (1955) "A behavioral model of rational choice," *The quarterly journal of economics*, 69 (1), 99–118. 4

Smeulders, Bart, Laurens Cherchye, and Bram De Rock (2021) "Nonparametric analysis of random utility models: computational tools for statistical testing," *Econometrica*, 89 (1), 437–455. 5

Suleymanov, Elchin (2018) "Stochastic attention and search." 11

Tversky, Amos (1972) "Elimination by aspects: A theory of choice.," *Psychological review*, 79 (4), 281. 1, 4

A. Proofs

A.1. Proof of Proposition 1

Since any ρ can be completed, it is sufficient to establish validity of statements (i) and (ii) for complete stochastic data sets (i.e., $\mathcal{A} = 2^X \setminus \{\emptyset\}$).

Proof of (i). Fix any complete ρ and let $\pi_A(\phi| \succ) = \rho_A(a)\mathbb{1}\left(\phi(A) = \{a\}\right)$ for all $a, A, \succ$. Then ρ admits a stable RAUM representation $\pi_A(\phi| \succ)\pi^*(\succ)$, where π^* is any element in $\Delta(U)$.

Proof of (ii). Fix any complete ρ. For any $\succ, A$ let $a_{\succ,A}$ be the best element in A according to $\succ$ and $\kappa_{\succ,A} = \sum_{\succ' \in U} \mathbb{1}\left(a_{\succ,A} = a_{\succ',A}\right)$ be the number of preference orders for which $a_{\succ,A}$ is also the best. Take $\pi_A(\succ) = \rho_A(a_{\succ,A})/\kappa_{\succ,A}$. Then ρ admits a set-monotone RAUM representation $\pi_A(\phi| \succ)\pi_A(\succ)$, where $\pi_A(\phi| \succ) = \mathbb{1}\left(\phi(A) = A\right)$ for all $A, \succ$.

Proof of (iii). To prove (iii) we will construct an incomplete data set (i.e., $\mathcal{A} \neq 2^X \setminus \{\emptyset\}$) that does not admit a set-monotone and stable RAUM. Let $X = \{a, b, c, d\}$ and

$$\mathcal{A} = \{\{a, b\}, \{a, c\}, \{b, d\}\{a, b, d\}, \{a, c, d\}, \{b, c, d\}\}.$$

Suppose the observed ρ is as follows

Menu	$\{a, b\}$	$\{a, c\}$	$\{b, d\}$	$\{a, b, d\}$	$\{a, c, d\}$	$\{b, c, d\}$
$\rho_A(a)$	1	1	-	0	0	-
$\rho_A(b)$	0	-	1	1	-	α_b
$\rho_A(c)$	-	0	-	-	1	α_c
$\rho_A(d)$	-	-	0	0	0	α_d

where $\alpha_d > 0$. (Columns in the above matrix correspond to different menus. For instance,

the third element of the second row is $\rho_{\{b,d\}}(b)$.) Consider $\{a, b\}$ and $\{a, b, d\}$. We will abuse notation and associate filters with consideration sets they imply. For example, if ϕ_1 is such that $\phi_1(A) = B$ we will write $\pi_A(\succ, B)$ instead of $\pi_A(\succ, \phi)$. Note that

$$0 = \rho_{\{a,b\}}(b) = \sum_{\succ} \pi_{\{a,b\}}(\succ, \{b\}) + \pi_{\{a,b\}}(\succ, \{a, b\})\mathbb{1}\left(b \succ a\right),$$

$$1 = \rho_{\{a,b,d\}}(b) = \sum_{\succ} \pi_{\{a,b,d\}}(\succ, \{b\}) + \pi_{\{a,b,d\}}(\succ, \{a, b\})\mathbb{1}\left(b \succ a\right)$$

$$+ \pi_{\{a,b,d\}}(\succ, \{b, d\})\mathbb{1}\left(b \succ d\right) + \pi_{\{a,b,d\}}(\succ, \{a, b, d\})\mathbb{1}\left(b \succ a, d\right).$$

Subtracting the first equation from the second one, we get that

$$1 = \sum_{\succ}[\pi_{\{a,b,d\}}(\succ, \{b\}) - \pi_{\{a,b\}}(\succ, \{b\})] + [\pi_{\{a,b,d\}}(\succ, \{a, b\}) - \pi_{\{a,b\}}(\succ, \{a, b\})]\mathbb{1}\left(b \succ a\right)$$

$$+ \pi_{\{a,b,d\}}(\succ, \{b, d\})\mathbb{1}\left(b \succ d\right) + \pi_{\{a,b,d\}}(\succ, \{a, b, d\})\mathbb{1}\left(b \succ a, d\right).$$

Set-monotonicity of π_A and stability of preferences then imply that

$$1 \leq \sum_{\succ} \pi_{\{a,b,d\}}(\succ, \{b, d\})\mathbb{1}\left(b \succ d\right) + \pi_{\{a,b,d\}}(\succ, \{a, b, d\})\mathbb{1}\left(b \succ a, d\right).$$

Since $\sum_{\succ} \sum_{D \subseteq \{a,b,d\}} \pi_{\{a,b,d\}}(\succ, D) = 1$, we can conclude that $(\pi(\succ)_{\succ \in U}$ is such that $b \succ d$ with probability 1. If we apply the above arguments to $\rho_{\{a,c\}}(c)$ and $\rho_{\{a,c,d\}}(c)$, we can deduce that $c \succ d$ with probability 1. Thus, with probability 1, d is never picked if it is considered together with b or c. Hence, in menu $\{b, c, d\}$ it can be picked with positive probability (i.e. $\alpha_d > 0$) if and only if set $\{d\}$ is considered with positive probability. The later is not possible since $\pi_{\{b,c,d\}}(\succ, \{d\}) \leq \pi_{\{b,d\}}(\succ, \{d\}) \leq \rho_{\{b,d\}}(d) = 0$ (d is never picked in menu $\{b, d\}$).

A.2. Proof of Theorem 1

Assume that π is a set-monotone and stable RAUM representation of possibly incomplete ρ.
Let $d_m = \left| 2^X \setminus \{\emptyset\} \times U \times \Phi \right|$ and $\mathcal{A}_a = \{(a, A) \in X \times \mathcal{A} : a \in A\}$. Fix one-to-one mapping
$i_1 : \mathcal{A}_a \rightarrow \{1, 2, \ldots, |\mathcal{A}_a|\}$ that maps a pair (a, A) to a corresponding element of vector ρ.
Also fix any one-to-one $i_2 : 2^X \setminus \{\emptyset\} \times U \times \Phi \rightarrow \{1, 2, \ldots, d_m\}$. Let B be a matrix of size
$|\mathcal{A}_a| \times d_m$ such that the (k, l)-element of it, $B_{k,l}$, is defined as follows

$$
B_{k,l} = \begin{cases} \mathbb{1} \left(a \succ b, \forall b \in \phi(A) \setminus \{a\} \right), & \text{if } k = i_1((a, A)), l = i_2(A, \succ, \phi) \text{ for some } \succ, \phi, (a, A) \in \mathcal{A}_a \\ 0, & \text{otherwise} \end{cases}.
$$

Hence, in matrix notation, if ρ admits a RAUM representation, then

$$
\rho = B\pi,
$$

where $\pi = (\pi_A(\succ, \phi))_{i_2(A, \succ, \phi)}$.

The rest of the restrictions will be imposed on *all* menus (including the ones that are not
present in $\mathcal{A}$). These restrictions do not use any data. First, we want to capture the fact that
$\pi_A(\cdot, \cdot)$ is a probability distribution and needs to sum up to 1. For any $A \in 2^X \setminus \{\emptyset\}$, let an
$i_2(A, \succ, \phi)$ element of a row of matrix O to be 1 for all $\succ$ and ϕ and to be zero otherwise.
Hence, the constraint can be written as

$$
O\pi = 1_{d_m},
$$

where O is the matrix of size $d_m \times d_1$.

The next set of restrictions captures feasibility: $\pi_A(\succ, \phi) = 0$ whenever $\phi(A) = \emptyset$. Let

$d_2 = \sum_{A,\succ,\phi} \mathbb{1}\left(\phi(A) = \emptyset\right)$. Then the feasibility constraint can be written as

$$F\pi = 0,$$

where F is a matrix of 0/1 that picks $i_2(A, \succ, \phi)$ elements of π that should be set to zero because of feasibility.

Next we want to rewrite the definition of stability in the matrix form. Note that stability can be written as $\sum_{\phi} \pi_A(\succ, \phi) = \sum_{\phi} \pi_B(\succ, \phi)$ for all A, B. Fix any A, B, and $\succ$. Let $\iota^{A,B,\succ}$ be a vector of length d such that

$$\iota_k^{A,B,\succ} = \mathbb{1}\left(\exists \phi\ :\ k = i_2(A, \succ, \phi)\right) - \mathbb{1}\left(\exists \phi\ :\ k = i_2(B, \succ, \phi)\right).$$

Take a collection of vectors $\left\{\iota_k^{A,B,\succ}\right\}_{A,B,\succ}$ and remove all linearly dependent or zero vectors. Let every element of what is left to be a row of a matrix S. Then, stability is equivalent to

$$S\pi = 0.$$

Finally, we want to build a matrix representation of set-monotonicity. Note that, under stability, $\pi_A(\phi \mid \succ) \geq \pi_B(\phi \mid \succ)$ is equivalent to $\pi_A(\succ, \phi) \geq \pi_B(\succ, \phi)$. Hence, similarly to stability, fix any $A, B, \succ, \phi$ such that $A \subseteq B$, $A \neq B$, and let $\iota^{A,B,\succ,\phi}$ be a vector of length d such that

$$\iota_k^{A,B,\succ,\phi} = \mathbb{1}\left(k = i_2(A, \succ, \phi)\right) - \mathbb{1}\left(k = i_2(B, \succ, \phi)\right).$$

Similarly to matrix S we can use vectors $\left\{\iota_k^{A,B,\succ,\phi}\right\}$ to build matrix M such that set-monotonicity is equivalent to

$$M\pi = \bar{v},$$

where $\bar{v}$ is a component-wise nonnegative vector. Define G as

$$
G = \begin{bmatrix} B & 0 \\ O & 0 \\ F & 0 \\ S & 0 \\ M & -I \end{bmatrix}.
$$

As a result, if ρ admits a set-monotone and stable RAUM representation, then the system $g = Gv$ has a component-wise nonnegative solution $(\pi', \bar{v}')'$.

Now suppose $g = Gv$ has a component-wise nonnegative solution $(\pi', \bar{v}')'$, we want to show that this π is a set-monotone and stable RAUM representation of ρ. By the definition of G, π is a complete (i.e., includes all possible menus) collection of distributions over $U \times \Phi$. Moreover, the constructed π is set-monotone and stable and can generate the observed ρ.

A.3. Proof of Proposition 3

Towards a contradiction assume $\Pi_\rho = \Delta(U)$. If ρ is irregular, then there exist $A, B \in \mathcal{A}$, $A \subseteq B$, and $a \in A$ such that $\epsilon \equiv \rho_B(a) - \rho_A(a) > 0$. Since by assumption $\Pi_\rho = \Delta(U)$, take any π such that a is the worst with probability 1. If ϕ^* is such that $\phi^*(A) = \phi^*(B) = \{a\}$, then

$$
\rho_B(a) = \sum_{\succ} \sum_{\phi} \pi_B(\phi| \succ)\pi(\succ)\mathbb{1}\left(a \succ b,\ \forall b \in \phi(B),\ a \in \phi(B) \right) = \sum_{\succ} \pi_B(\phi^*| \succ)\pi(\succ).
$$

Similarly,

$$
\rho_A(a) = \sum_{\succ} \pi_A(\phi^*| \succ)\pi(\succ).
$$

Taking the difference between these two equations we get that

$$0 < \epsilon = \sum_{\succ} [\pi_B(\phi^* | \succ) - \pi_A(\phi^* | \succ)] \pi(\succ) \le 0,$$

where the last inequality follows from set-monotonicity. This contradiction completes the proof.

B. Omitted Details from Example 6

There are 4 possible non-singleton menus: $\{a, b, c\}$, $\{a, b\}$, $\{b, c\}$, and $\{a, c\}$. Given δ and A, let $D_{\delta, A}$ be the deterministic consideration set.

Case 1, $A = \{a, b, c\}$. Theorem 1 in Caplin et al. (2016) implies that if $\mu(c) > 1/(3 + \delta)$ (i.e., $\delta > -3 + 1/\mu(c) = \delta_1^*$), then $D_{\delta, A} = A$. If $\delta < \delta_1^*$ and

$$\mu(b) \ge \frac{\mu(a) + \mu(b)}{2 + \delta} > \mu(c)$$

(i.e., $\delta_4^* \le \delta < \delta_1^*$), then $D_{\delta, A} = \{a, b\}$. Finally, if

$$\mu(a) \ge \frac{\mu(a)}{1 + \delta} > \mu(b)$$

(i.e., $\delta < \delta_4^*$), then the consideration set is $\{a\}$.

Case 2, $A = \{a, b\}$. Applying the same Theorem 1 in Caplin et al. (2016) we can obtain that if

$$\mu_A(b) = \frac{\mu(b)}{\mu(a) + \mu(b)} > 1/(2 + \delta)$$

(i.e., $\delta > -1 + \mu(a)/\mu(b) = \delta_4^*$), then $D_{\delta, A} = \{a, b\}$, and $D_{\delta, A} = \{a\}$ otherwise.

Case 3, $A = \{b, c\}$. Similarly to Case 2, if

$$\mu_A(b) = \frac{\mu(c)}{\mu(b) + \mu(c)} > 1/(2 + \delta)$$

(i.e., $\delta > -1 + \mu(b)/\mu(c) = \delta_e^*$), then $D_{\delta,A} = \{b, c\}$, and $D_{\delta,A} = \{b\}$ otherwise.

Case 4, $A = \{a, c\}$. If

$$\mu_A(c) = \frac{\mu(c)}{\mu(a) + \mu(c)} > 1/(2 + \delta)$$

(i.e., $\delta > -1 + \mu(a)/\mu(c) = \delta_2^*$), then $D_{\delta,A} = \{a, c\}$, and $D_{\delta,A} = \{a\}$ otherwise.

Given that $1 - 2\mu(c) = \mu(a) + (\mu(b) - \mu(c) \geq \mu(a)$ and $\mu(a) \geq \mu(b) \geq \mu(c)$, it follows that $\delta_1^* \geq \delta_2^* \geq \delta_3^*$ and $\delta_2^* \geq \delta_4^*$. Tables 1 and 2 summarize the above derivations for different values of δ and menu A. Table 1 corresponds to the case when $\delta_3^* \geq \delta_4^*$. Table 2 considers the case when $\delta_4^* < \delta_3^*$.

Next, we compute the implied by the model probabilities of choosing different options. Theorem 1 in Caplin et al. (2016) implies that the probability that $y \in D_{\delta,A}$ is chosen from $y \in D_{\delta,A}$ satisfies

$$P_{D_{\delta,A}}(y) = \frac{\mu_{D_{\delta,A}}(y(|D_{\delta,A}| + \delta) - 1)}{\delta}.$$

Assume that $\delta_3^* > \delta_4^*$ (the opposite case leads to the same conclusion). The following table displays $P_{D_{\delta,A}}(a)$.

$A/\delta \in$	$(0, \delta_4^*]$	$(\delta_4^*, \delta_3^*]$	$(\delta_3^*, \delta_2^*]$	$(\delta_2^*, \delta_1^*]$	$(\delta_1^*, +\infty)$
$\{a, b, c\}$	1	$P_{\{a,b\}}(a)$	$P_{\{a,b\}}(a)$	$P_{\{a,b\}}(a)$	$P_{\{a,b,c\}}(a)$
$\{a, b\}$	1	$P_{\{a,b\}}(a)$	$P_{\{a,b\}}(a)$	$P_{\{a,b\}}(a)$	$P_{\{a,b\}}(a)$
$\{a, c\}$	1	1	1	$P_{\{a,c\}}(a)$	$P_{\{a,c\}}(a)$

For options b and c the tables are

$A/\delta \in$	$(0, \delta_4^*]$	$(\delta_4^*, \delta_3^*]$	$(\delta_3^*, \delta_2^*]$	$(\delta_2^*, \delta_1^*]$	$(\delta_1^*, +\infty)$
$\{a, b, c\}$	0	$P_{\{a,b\}}(b)$	$P_{\{a,b\}}(b)$	$P_{\{a,b\}}(b)$	$P_{\{a,b,c\}}(b)$
$\{a, b\}$	0	$P_{\{a,b\}}(b)$	$P_{\{a,b\}}(b)$	$P_{\{a,b\}}(b)$	$P_{\{a,b\}}(b)$
$\{b, c\}$	1	1	$P_{\{b,c\}}(b)$	$P_{\{b,c\}}(b)$	$P_{\{b,c\}}(b)$

and

$A/\delta \in$	$(0, \delta_4^*]$	$(\delta_4^*, \delta_3^*]$	$(\delta_3^*, \delta_2^*]$	$(\delta_2^*, \delta_1^*]$	$(\delta_1^*, +\infty)$
$\{a, b, c\}$	0	0	0	0	$P_{\{a,b,c\}}(c)$
$\{b, c\}$	0	0	$P_{\{b,c\}}(c)$	$P_{\{b,c\}}(c)$	$P_{\{b,c\}}(c)$
$\{a, c\}$	0	0	0	$P_{\{a,c\}}(c)$	$P_{\{a,c\}}(c)$

Note that since $P_{\{a,b,c\}}(a) < P_{\{a,b\}}(a) \leq P_{\{a,c\}}(a)$, $P_{\{a,b,c\}}(b) < P_{\{a,b\}}(b) \leq P_{\{b,c\}}(b)$, and $P_{\{a,b,c\}}(c) < P_{\{a,c\}}(c) \leq P_{\{b,c\}}(c)$, the computes distributions do not violate regularity for all values of δ and all menus A.

If $\succ_{y_1 y_2 y_3}$ is such that $y_1 \succ_{y_1 y_2 y_3} y_2 \succ_{y_1 y_2 y_3} y_3$, then one can verify that the following π^* is a stable distribution that together with the deterministic consideration set is consistent with observed choices:

$$\pi^*(\succ_{abc}) = \frac{(\delta - \delta_1^* + \kappa_{ac})}{(\mu(b) + \mu(c))} \frac{\mu(a)\mu(b)}{\delta} \geq 0,$$

$$\pi^*(\succ_{acb}) = \frac{(\delta - \delta_1^* + \kappa_{ac})}{(\mu(b) + \mu(c))} \frac{\mu(a)\mu(c)}{\delta} \geq 0,$$

$$\pi^*(\succ_{bac}) = \frac{(\delta - \delta_1^* + \kappa_{bc})}{(\mu(a) + \mu(c))} \frac{\mu(a)\mu(b)}{\delta} \geq 0,$$

$$\pi^*(\succ_{bca}) = \frac{(\delta - \delta_1^* + \kappa_{bc})}{(\mu(a) + \mu(c))} \frac{\mu(b)\mu(c)}{\delta} \geq 0,$$

$$\pi^*(\succ_{cab}) = \frac{(\delta - \delta_1^*)}{(\mu(a) + \mu(b))} \frac{\mu(a)\mu(c)}{\delta} \geq 0,$$

$$\pi^*(\succ_{cba}) = 1 - (\pi_1 + \pi_2 + \pi_3 + \pi_4 + \pi_5) = \frac{(\delta - \delta_1^*)}{(\mu(a) + \mu(b))} \frac{\mu(b)\mu(c)}{\delta} \geq 0,$$

where $\kappa_{ac} = 1/\mu(c) - 1/\mu(a) \geq 0$ and $\kappa_{bc} = 1/\mu(c) - 1/\mu(b) \geq 0$.