

PROGRESSIVE RANDOM CHOICE*

EMEL FILIZ-OZBAY[†] AND YUSUFCAN MASATLIOGLU[§]

ABSTRACT. We introduce a flexible framework to study probabilistic choice that accommodates heterogeneous types and bounded rationality. We provide a novel progressive structure for the heterogeneous types to capture heterogeneity due to varying levels of a behavioral trait such as limited attention, willpower stock, shortlisting constraint, degree of loss aversion, or being pro-social. Given an order of alternatives, our progressive structure sorts the types by the extent to which they align with this order. Unlike the Random Utility Model, our model uniquely identifies the heterogeneity, allowing policymakers to perform an improved welfare analysis. As a showcase, we provide characterization of a well-studied type of bounded rationality: “less-is-more.” In addition, we provide conditions for unique identification of the underlying order for the less-is-more structure.

Date: May, 2022.

* We thank David Ahn, Jose Apesteguia, Miguel Ballester, Paul Cheung, David Dillenberger, Keaton Ellis, Jay Lu, Paola Manzini, Marco Mariotti, Collin Raymond, Erkut Ozbay, Ece Yegane, and Kemal Yildiz for helpful comments and discussions. This version is immensely benefited from the inputs of the editor and four careful reviewers.

[†] University of Maryland, 3114 Tydings Hall, 7343 Preinkert Dr., College Park, MD 20742. Email: efozbay@umd.edu.

[§] University of Maryland, 3114 Tydings Hall, 7343 Preinkert Dr., College Park, MD 20742. E-mail: yusufcan@umd.edu.

1. INTRODUCTION

In the empirical analysis of human behavior, the data often comes in a probabilistic form indicating the choice frequencies, such as how often a specific restaurant is visited in a city or how frequently a particular insurance policy is purchased. The Random Utility Model (RUM) is the standard tool for analyzing probabilistic choice in which the randomness in choices is attributed to the variation in tastes or types. This approach makes a strong behavioral assumption that each type is a utility maximizer. However, there is abundant evidence against utility maximization in various fields such as law, economics, psychology, and marketing.¹ In this paper, we introduce a random choice model, which offers a flexible framework with which to study probabilistic choice and accommodates heterogeneous types and bounded rationality. Our framework can be applied to both the choices of a single individual in different situations (intrapersonal) and the choices of different individuals in the same environment (interpersonal).

A random choice model is defined by having a collection of choice functions and a probability distribution over them.² This collection contains heterogeneous types which may or may not be consistent with a utility maximization. Without further restrictions, this model is too permissive to make any prediction in terms of the observed behavior. In our framework, each choice function corresponds to a particular behavioral type in the collection. In many circumstances, the types are naturally sorted according to some ingrained characteristic. For example, consider a situation in which the agents in a population have varying levels of environmental cautiousness. The greener types choose more environmentally friendly policies than the less green types. Therefore, the types are naturally sorted according to their environmental cautiousness. To capture the idea of this example, we focus on domains where alternatives are ranked according to a reference order (such as environmental friendliness). In many other economically relevant contexts, such an order naturally arises. For example, the tax policies can be ordered by the revenue they generate (Roberts [1977]), public goods can

¹See Huber et al. [1982], Ratneshwar et al. [1987], Tversky and Simonson [1993], Kelman et al. [1996], Prelec et al. [1997], Echenique et al. [2011], and Trueblood et al. [2013].

²McFadden and Richter [1990] briefly mentions the possibility of working with choice functions outside of preference maximization. However, they do not investigate the idea further. Recently, in an independently developed paper, Dardanoni et al. [2020a] works with a model of randomization over choice functions. Their focus is mostly on the identification of preferences and cognitive distributions in specific models of choice by assuming an observable mixture of choice functions. Instead, our focus is on the representation of probabilistic choice by a random choice model. Hence, the two papers complement each other.

be ordered by their provision levels (Epple et al. [2001]), insurance offers can be ordered by their deductibles (Barseghyan et al. [2019]), and scheduled payments can be ordered by their present value based on an interest rate (Manzini and Mariotti [2006a]). Given a reference order, a collection of types is called “progressive” if the types can be sorted according to their alignment with this order. The random choice model in which the collection of types is progressive is called *Progressive Random Choice* (PRC).

Our progressive structure is a generalization of a well-known condition called the single crossing property. Indeed, the two conditions are equivalent if each type is generated by a utility maximization. The single crossing property appears in various economic models, including Mirrlees [1971], Roberts [1977], Grandmont [1978], Rothstein [1990], Milgrom and Shannon [1994], and Gans and Smart [1996].³ This property enables economists to perform comparative statics comparing the heterogeneous types in terms of their alignment with the underlying order such as one type choosing more prosocial policies or avoiding loss more than the other type. The empirical analysis of type distributions often relies on ordered heterogeneous types. For example, Chiappori et al. [2019] identifies the distribution of individual risk attitudes from aggregate data, and Barseghyan et al. [2019] show how the heterogeneity in consideration sets and risk aversion can be identified. Our model enables researchers to extend this type of analysis to boundedly rational agents.

In this framework, we first answer questions about characterization and identification. When do progressive heterogeneous types exist for a given reference order? Can one identify the heterogeneous types and their distribution from the observed aggregate choice? Our main result, Theorem 1, answers both questions and enables us to provide comparative statics between any two probabilistic choices. We show that one can uniquely identify the PRC representation—both the collection of choice types and the probability weight assigned to each choice type in the collection. This feature contrasts with the undesirable non-uniqueness of the RUM. The unique identification allows policy makers to craft policies while knowing the unique weights of each behavioral type, which in turn allows an improved welfare analysis. While the model is uniquely identifiable, this structure does not entail any restrictions on the observed choices—any probabilistic choice has a *unique* PRC representation.

³Apesteguia et al. [2017] applied this property to random utility models.

The empirical validity of our framework can be investigated by designing careful experiments. Manzini and Mariotti [2006b] provides a rare opportunity to test our model. Our analysis of their data confirms that a progressive structure exists in the observed collection types. One may check the progressive structure on other rich data sets when types are observed in contexts such as decision making under risk (McCausland et al. [2020]), time preferences, and portfolio allocation.

The main advantage of our framework over the RUM is its flexibility to study phenomena that are outside the utility maximization paradigm. A researcher with a specific type of bounded rationality in mind can answer the following questions: What are the behavioral implications of this type of bounded rationality for the probabilistic data? What properties does the probabilistic data need to satisfy for the existence of progressive types exhibiting such bounded rationality? Theorem 1 becomes invaluable in answering the second question since it already provides the unique PRC representation for the observed data. Because the unique heterogeneous types are already identified by Theorem 1, the researcher needs only to verify the property of interest to be satisfied for those identified types. Therefore, Theorem 1 is a crucial step to study any bounded rationality and to derive such characterization results.

As a showcase, we perform the aforementioned characterization for the “less-is-more type” of bounded rationality. “Less-is-more” refers to a behavior of making fewer mistakes on smaller sets. It aims to capture one of the most studied behavioral phenomena, namely choice overload. Chernev et al. [2015] argues that choice overload can have negative welfare consequences; hence, having fewer options can lead to an increase in consumer welfare, or “less-is-more”.⁴ For example, the complexity caused by the abundance of alternatives may lead to choosing lower-ranked alternatives (see, e.g., Iyengar and Lepper [2000], Chernev [2003], Iyengar et al. [2004], and Caplin et al. [2009]). Kamenica [2008] argues that consumers might be better off with smaller varieties if they are uninformed and choose randomly from the available options. According to Gul and Pesendorfer [2001], a diner may choose unhealthy but tempting dishes on a large restaurant menu while choosing the healthy dishes in smaller menus because she controls herself better with smaller menus.

⁴The “less-is-more” property is rich enough to accommodate models such as shortlisting (Manzini and Mariotti [2007]), rationalization (Cherepanov et al. [2013]), preferred personal equilibrium (Kőszegi and Rabin [2006]), limited attention (Lleras et al. [2017]) and categorization (Manzini and Mariotti [2012]).

We uncover that the types with less-is-more property have a simple and intuitive implication for the probabilistic data. We call this behavioral postulate *U-Monotonicity*. It simply states that the choice frequency of the upper contour set of an alternative is monotonic with respect to set inclusion. We show that U-Monotonicity is not only necessary but also sufficient for the existence of a progressive representation with less-is-more. Thanks to Theorem 1, this representation is unique.

All the results above assume that the reference order is exogenously given. In some applications, the order is conceivably observable to the researcher (for example, it may correspond to a social norm of a society) or the researcher may be the one designing the menu of options in a controlled experiment so that an objective order (such as tax revenue, the level of public good provision, or the time schedule of payments) is imposed. However, in some other contexts, especially when the order corresponds to the underlying preferences, deriving it from the probabilistic choice might be needed. Due to Theorem 1, without further restrictions, any data has a PRC representation with respect to any reference order. However, once we commit to a specific type of bounded rationality, we restrict the behavior as well, as we do in the less-is-more showcase. Focusing on this specific behaviour allows us to identify the reference order endogenously. We provide a procedure to derive a binary relation which must be part of the unobserved reference order (Proposition 1). Finally, we provide sufficient conditions so that this derived binary relation becomes complete. Together with Theorem 1, we therefore establish that the reference order, the collection of types, and their weights in the PRC representation are uniquely constructed for less-is-more. Endogenizing the order improves the applicability of the model. The less-is-more showcase should be viewed as a guideline for future researchers to characterize other types of bounded rationality of interest. The contribution of our paper is to offer this novel framework of PRC and enable researchers study bounded rationality in the probabilistic domain.

Our theoretical contribution complements the models of Apesteguia and Ballester [2020] and Dardanoni et al. [2020a], who utilize models of randomization over choice functions for empirical applications to identify the heterogeneity in the data. Dardanoni et al. [2020a] prove the usefulness of the random choice model in the identification of preferences and cognitive distributions in specific models of choice. As in Apesteguia et al. [2017], Apesteguia and Ballester [2020] work with preference types, but they apply the progressive structure only locally and allow for limited data. Our paper is also related to the recent literature

that combines decision theory and econometric analysis. The most closely related papers in this literature are Abaluck and Adams [2017]; Barseghyan, Coughlin, Molinari, and Teitelbaum [2018]; and Dardanoni, Manzini, Mariotti, and Tyson [2020b]. In a general setup, Abaluck and Adams [2017] show that, by exploiting asymmetries in cross-partial derivatives, consideration set probabilities and utilities can be separately identified from observed choices when rich exogenous variation exists in the observed covariates. Barseghyan et al. [2018] provide partial identification results when exogenous variation in observed covariates is more restricted. Lastly, similar to previous papers, Dardanoni et al. [2020b] study choices from a fixed menu of alternatives. They consider aggregate choices in which individuals might differ in terms of both their consideration capacities and preferences. Finally, the PRC with less-is-more offers predictions distinct from those in several well-known probabilistic choice models such as Manzini and Mariotti [2014], Brady and Rehbeck [2016], and Cattaneo et al. [2019] (see Section 5).

The rest of the paper is organized as follows. Section 2 introduces the random choice model and the progressiveness notion. We provide three distinct classes of models and general conditions that guarantee the progressive structure for each class. Section 3 provides comparative statics between any two models within our framework. Section 4 characterizes *less-is-more* and shows that, for the less-is-more structure, the unique reference order is derived under mild assumptions. Section 5 summarizes how our models relate to other well-known probabilistic choice models. Section 6 concludes.

2. MODEL

Let X be a non-empty and finite set of alternatives and let \mathcal{X} denote all non-empty subsets of X where $|X| \geq 3$. A probabilistic choice function is a mapping $\pi : X \times \mathcal{X} \rightarrow [0, 1]$ such that for any $S \subseteq X$, (i) $\pi(x|S) > 0$ only if $x \in S$; (ii) $\sum_{x \in S} \pi(x|S) = 1$. We interpret $\pi(x|S)$ as the probability of choosing x from alternative set S . We denote the sum of all choice probabilities in $T \subset S$ with $\pi(T|S)$, i.e. $\pi(T|S) = \sum_{x \in T} \pi(x|S)$. A deterministic choice function on X is a mapping $c : \mathcal{X} \rightarrow X$ such that $c(S) \in S$ for any $S \subseteq X$. Let \mathbb{C} denote the set of all choice functions on X .

We now introduce a Random Choice Model (RCM). Let μ be a probability distribution on \mathbb{C} , so $\mu(c)$ is the probability of c being the choice function. The probability distribution

μ constitutes a probabilistic choice function π_μ such that

$$(RCM) \quad \pi_\mu(x|S) = \sum_{c(S)=x} \mu(c)$$

The probabilistic choice function induced by μ sets the probability of an alternative x being chosen from an alternative set S as the sum of probabilities of choice functions that select x from S . We call the choices in the support of μ (the choice functions with strictly positive weights) the choice types. Hence, the probability of x being chosen from S is the frequency of those *choice types* who choose x from S .

We say that a probabilistic choice function π has a Random Choice representation if there exists μ such that $\pi = \pi_\mu$. If the support of μ consists of only distinct choice functions resulting from utility maximization, then π_μ corresponds to the well-known Random Utility Model (RUM). Hence, the RUM is a special case of the RCM.

The RCM enjoys high explanatory power: Any probabilistic choice function can be represented within the random choice framework. Moreover, the well-known non-uniqueness result of the RUM implies that the RCM representation is not unique in general. We address this issue by studying the RCM on more structured domains. We show that this approach generates not only an economically-meaningful relationship between the choice types but also deliver uniqueness.

We consider the set of alternatives that are naturally ordered by a linear order \succ .⁵ Examples of such domains are in abundance as discussed in the introduction. We need further notation for the rest of the paper. We define $x \succeq y$ if $x \succ y$ or $x = y$. Given a relation R , we denote the upper and lower contour sets of an alternative x with respect to R by $U_R(x) = \{y \in X \mid yRx\}$ and $L_R(x) = \{y \in X \mid xRy\}$, respectively.⁶

We now propose a condition to relate the choice types with the order on the domain. We do this by sorting the types in terms of how much the behavior is in line with the reference order. In the public good provision problem, the level of public good picked by more prosocial types is above the level chosen by the less prosocial types. In the restaurant example, if an individual with less self-control resists temptation, another individual with more self-control resists that too. We now state our main concept formally.

⁵Throughout the paper, we simply say an “order” instead of a “linear order”.

⁶We also define $U_R(S) = \bigcup_{x \in S} U_R(x)$ and $L_R(S) = \bigcup_{x \in S} L_R(x)$.

Definition 1. A collection of distinct choice types $\mathcal{C} \subseteq \mathbb{C}$ is *progressive* with respect to \triangleright if \mathcal{C} can be sorted $\{c_1, c_2, \dots, c_T\}$ such that $c_t(S) \succeq c_s(S)$ for all S and for any $t > s$.⁷

The progressiveness imposes an ordered structure on the collection of choices types such that a higher indexed type cannot choose an alternative that is dominated by the choice of a lower indexed type from the same set. In other words, the progressiveness requires type t to be more aligned with \triangleright than type s for $t > s$. The idea of progressive types reduces the heterogeneity of types in RCM into one dimension. That dimension of heterogeneity can be caused by the varying levels of willpower, prosocial behavior, attention capacities, or loss aversion coefficients. Hence, the progressive structure allows to study the heterogeneity within a given phenomena of interest. We now define progressive random choice formally.

Definition 2. π has a **progressive** random choice representation with respect to \triangleright , (PRC $_{\triangleright}$), if there exists μ on \mathbb{C} such that $\pi = \pi_{\mu}$ and the support of μ is progressive with respect to \triangleright .

We view our novel progressive structure as a strength of the model because it provides a meaningful interpretation for the support of the RCM. Recall that the support of the RCM (or its special case, RUM) consists of several independent types and there is no immediate comparison between them. In contrast, the PRC orders the choice types with respect to a natural order \triangleright on the domain. This is interpreted as choice types gradually becoming more aligned with the reference type induced by \triangleright .

2.1. Progressiveness in Different Classes of Models. One might wonder when the progressive structure holds within a class of models. In the case of utility maximization, this question has been extensively studied (see e.g. Mirrlees [1971], Roberts [1977], Grandmont [1978], Rothstein [1990], Milgrom and Shannon [1994], Gans and Smart [1996]). In this class, each choice type operates as a utility maximizer, hence it is denoted by $\{u_t\}$.⁸ For every type t there exists a utility function, u_t , such that for all S ,

$$\text{(Class 1)} \quad c_t(S) = \underset{x \in S}{\operatorname{argmax}} u_t(x)$$

⁷The betweenness property defined by Albayrak and Aleskerov [2000], Horan and Sprumont [2016] in a different context is a closely related concept.

⁸We assume that u_t is strict so that the choice is unique.

The well-known single crossing property of utility functions guarantees the progressive structure which is studied by Apesteguia et al. [2017] within the probabilistic choice context. Below, we formally define the single crossing condition for a collection of $\{u_t\}$.

Condition 1: Let \triangleright be an order. For any $x \triangleright y$ and $t > s$,

$$u_s(x) > u_s(y) \Rightarrow u_t(x) > u_t(y)$$

Remark 1. Within Class 1, a collection of choice functions defined by $\{u_t\}$ satisfies Condition 1 with respect to \triangleright if and only if it is progressive with respect to \triangleright .

The proof of Remark 1 is in the Appendix. Finding a necessary and sufficient condition for progressiveness is not straightforward when we leave the utility maximization paradigm. Next, we perform this task for two other classes of models covering different bounded rationality models.

The next one is the class of two-stage procedures which contain several well-studied decision making models as listed below. In this class, the first stage determines a constrained set induced by a particular behavioral limitation. In the second stage, the decision maker optimizes over this constraint. An RCM in which each type performs a two-stage procedure and the heterogeneity is due to the variation in the behavioral limitation of different types while the utilities are type independent is denoted by $(u, \{\Gamma_t\})$.⁹ Formally, for every type t and for all S ,

$$\text{(Class 2)} \quad c_t(S) = \operatorname{argmax}_{x \in \Gamma_t(S)} u(x)$$

The models of Limited Attention by Masatlioglu et al. [2012], Lleras et al. [2017], Short-listing by Manzini and Mariotti [2007], Rationalization by Cherepanov et al. [2013], Categorization by Mariotti and Manzini [2012], Willpower by Masatlioglu et al. [2020], and Preferred Personal Equilibrium by Kőszegi and Rabin [2006] can be written as in Class 2 formulation. Here, all types maximize the same utility function. In an intrapersonal decision setting, this can be thought as a person with fixed preferences but facing an idiosyncratic behavioral constraint such as having idiosyncratic levels of attention, willpower, loss-aversion etc. In a

⁹We assume that u is strict, i.e., $u(x) \neq u(y)$ for all $x \neq y$ to be in line with our earlier assumption of the reference order being linear order.

population setting, this can be thought as the existence of a common attribute that ranks the alternatives such as the lowest price in shopping or the shortest distance in route choice and each type has a different constraint.

Next, we provide a sufficient condition on the class of two stage procedures for having a progressive structure in the RCM. Our condition is a monotonicity requirement on how the constraints of each type evolve. Particularly, the feasible set of the higher types contains higher utility options than that of lower types.

Condition 2: Let \triangleright be an order. For any $t > s$, $L_{\triangleright}(\Gamma_s(S)) \subseteq L_{\triangleright}(\Gamma_t(S))$.

Note that Condition 2 is weaker than requiring the higher types to consider more options. Hence, a limited attention model in which consideration sets grow with types also satisfies this condition.

Remark 2. Within Class 2, a collection of choice functions defined by $(u, \{\Gamma_t\})$ satisfies Condition 2 with respect to \triangleright that is ordinally equivalent to u if and only if it is progressive with respect to \triangleright .

The proof is provided in the Appendix. Note that when a type with a lower index chooses an alternative, it has to be in the constraint set of that type. Then by Condition 2, a higher type must consider some alternatives that weakly dominate the choice of the lower type. Therefore, the choice of a higher type can only be weakly better as well. This provides an ordered structure on the choice types as required by the progressiveness.

Interpretation of Condition 2 on well-known models of Class 2 is intuitive:

- *Shortlisting of Manzini and Mariotti [2007].* In this model, a decision maker first eliminates dominated alternatives with respect to a binary relation to form a shortlist (constraint) and then she maximizes her preferences on the shortlist. If the first stage binary relation gets more incomplete, the shortlists get gradually richer as Condition 2 requires.
- *Shortlisting of Tyson [2013].* As opposed to the above model, here, the alternatives are naturally ordered according to their salience- the property of standing out from the rest. A decision maker's preferences are imperfectly perceived due to cognitive or information-processing constraints. This determines the alternatives she dislikes

for sure. Among the ones that survive, she chooses the most salient option. If the information processing gets gradually more costly for higher types, then the shortlists get larger and Condition 2 is satisfied.

- *Preferred Personal Equilibrium of Kőszegi and Rabin [2006]*. Alternatives are ordered with respect to the consumption utility of a person. The constraint in this two-stage interpretation is the set of personal equilibria for each type and the utility is the consumption utility. Types have different loss aversion coefficients and their personal equilibrium contains the alternatives that are optimal conditional on them being the reference point (rational expectations). As the individual becomes more loss averse, the set of personal equilibria enlarges, hence, Condition 2 is satisfied.
- *Willpower of Masatlioglu et al. [2020]*. Consider a decision maker with limited willpower facing visceral urges. The alternatives are ordered with respect to the commitment utility. The constraint of each type contains the alternatives she can overcome her visceral urges with her willpower. As her willpower increases, she is able to overcome visceral urges more successfully and able to choose from a richer constraint set. Hence, this collection of choices also satisfies Condition 2.
- *Limited Attention of Masatlioglu et al. [2012]*. In this example, an individual maximizes her preferences on what she pays attention. Different types are able to attend to different set of alternatives. If the awareness of types extend gradually then Condition 2 is satisfied since the consideration sets become nested.
- *Rationalization of Cherepanov et al. [2013]*. The decision maker who is endowed with a set of rationales maximizes her preferences among alternatives that she can rationalize. A rationale can be intuitively understood as a story that states that some options are better than others. The choice types differ in terms of the set of rationales they use for that choice. As the set of rationales gradually gets larger, the corresponding collection of choices satisfies Condition 2.

The final class is the class of models with a menu-dependent behavioral cost. In this class, each type maximizes a utility function minus a menu-dependent cost function. The

heterogeneity is due to the variation in the behavioral cost but the utilities are type independent. Hence, we are interested in the collection of choices described by $(u, \{\kappa_t\})$. Formally, for every type t and for all S ,

$$\text{(Class 3)} \quad c_t(S) = \operatorname{argmax}_{x \in S} u(x) - \kappa_t(x, S)$$

where κ_t is the menu-dependent cost function.¹⁰

The models of Temptation and Self Control by Gul and Pesendorfer [2001], Fudenberg and Levine [2006], Dekel et al. [2009], Noor and Takeoka [2010] and Social Norms and Shame by Dillenberger and Sadowski [2012] fall into this class. For example, an individual with a fixed commitment utility but idiosyncratic costs of self control generating the probabilistic choice data falls in this class. Alternatively, a population of individuals choosing a public policy can have the same selfish utility (for example, their objective can be minimizing costs) but they may differ in how much they care about following a norm (such as being prosocial or minimizing inequality). This is also within the Class 3.

Let \triangleright be the order defined by a behaviorally motivated norm such as a social norm or temptation. Here, \triangleright is *possibly* different than the common utility function. In order to get a progressive structure, we impose the assumption that the marginal net utility of switching to an alternative ranked higher in the social norm is decreasing in types. In other words, higher types act more in line with the norm. Formally,

Condition 3: Let \triangleright be an order. For any $x \triangleright y$ and $t > s$,

$$u(x) - \kappa_s(x, S) > u(y) - \kappa_s(y, S) \Rightarrow u(x) - \kappa_t(x, S) > u(y) - \kappa_t(y, S)$$

Remark 3. Within Class 3, a collection of choice functions defined by $(u, \{\kappa_t\})$ satisfies Condition 3 with respect to \triangleright if and only if it is progressive with respect to \triangleright .

The proof of Remark 3 is in the Appendix and it is based on the intuition that when a lower indexed type chooses an alternative, say y , it means that y provides the highest net utility. Then by Condition 3, it is not optimal for a higher type to switch to a lower ranked alternative according to \triangleright . Hence, the higher type chooses an alternative that is ranked weakly higher than y . This provides an ordered structure on the choice types as required by the $\text{PRC}_{\triangleright}$.

¹⁰We assume that this maximization problem has a unique solution, hence, the choice is unique.

We also provide a stronger sufficient condition for progressive structure based only on the menu dependent cost function. Formally, if the menu dependent cost function satisfies

$$\kappa_s(x, S) - \kappa_s(y, S) \geq \kappa_t(x, S) - \kappa_t(y, S)$$

for any $x \triangleright y$ and $t > s$, then Condition 3 holds and the corresponding model of Class 3 has a progressive structure with respect to \triangleright . This stronger condition is intuitively satisfied by the two well-known models below.

- *Costly Self-Control*. In this model, the decision maker chooses from a menu with tempting alternatives according to the following maximization:

$$c_t(S) = \operatorname{argmax}_{x \in S} \{u(x) - \alpha_t f(\max_{y \in S} v(y) - v(x))\}$$

When f is a linear function, this is equivalent to the model of Gul and Pesendorfer [2001] and when it is an increasing and convex function, it becomes the model of Fudenberg and Levine [2006], Noor and Takeoka [2010]. These models of Class 3 can be thought as a decision maker with a temptation order such that her cost of temptation is randomly determined. Note that here the heterogeneity of types is captured by the temptation cost parameter α_t . If α_t increases by the type, then Condition 3 is satisfied with respect to the temptation order, v .

- *Ashamed to be Selfish of Dillenberger and Sadowski [2012]*. Consider a decision maker facing a trade-off between choosing her best allocation and minimizing shame caused by not choosing the best allocation according to a social norm. Assume each type differs only in terms of how much they are influenced by the social norm. Formally,

$$c_t(S) = \operatorname{argmax}_{x \in S} \{u(x) - (\max_{y \in S} \psi(y) - \psi(x))^{\beta_t}\}$$

where u is a utility function over allocations, ψ represents the norm, and β_t is the shame parameter of type t . The amount, $(\max_{y \in A} \psi(y) - \psi(x))^{\beta_t}$, is interpreted as the shame from choosing x in comparison to the alternative that maximizes the norm. The stronger condition restricts how the cost of deviation from a social norm varies by the type.

Note that in all the examples of Class 3 models above, the reference order is not determined by the consumption utility, u , of an alternative but instead determined by the

temptation level or the social norm utility. Thus, Condition 3 requires the heterogeneity parameters in these examples to have a certain structure.

The aforementioned examples illustrate how the progressive structure can be interpreted on archetypical models of two-stage procedures and behavioral cost models. In each example, we demonstrated that the PRC structure allows for a substantial degree of heterogeneity of choice behavior when committed to a specific model. In each case, the alternatives are naturally ordered either by an attribute, common utility, a temptation ranking, or a shared norm. The types are sorted according to how closely their behavioral concern or bounded rationality allow them to follow this order. Moreover, we provided sufficient conditions for these classes of models to have the progressive structure.

One may further show that the behavior induced by models in Classes 1-3 are nested and the corresponding conditions gradually get stronger.

Remark 4. Class 1 \subset Class 2 \subset Class 3.¹¹

2.2. Existence of a Representation. $\text{PRC}_{\triangleright}$ imposes some compatibility among all the choice functions in a collection because the choices in the support gradually become more and more aligned with the choice induced by \triangleright . It also allows a substantial degree of heterogeneity of choice behavior as we will see in Theorem 1. Our first result below states that PRC is capable of explaining all probabilistic choices for a given reference order. In other words, PRC enjoys high explanatory power.

Theorem 1. *Let \triangleright be a reference order. Every probabilistic choice π has a unique $\text{PRC}_{\triangleright}$ representation.*

The power of Theorem 1 comes from the fact that it is applicable to any data. For any specific behavior outside of utility maximization, Theorem 1 provides both existence and uniqueness of PRC representation. For example, let's say that a researcher wants to characterize a behavior, say γ , generated by a set of heterogeneous types outside of utility maximization. Certainly not every data set will be in that class. A research agenda studying heterogeneous γ types should address the following questions: (i) When do progressive heterogeneous types exist for a given reference order? (ii) Is the type distribution unique?

¹¹Given this result, it is routine to show that Condition 3 \Rightarrow Condition 2 \Rightarrow Condition 1 for the appropriately defined dual models.

(iii) Are these types in line with γ ? (iv) Finally, is the reference order identified uniquely? Our Theorem 1 answers (i) and (ii) for any type of bounded rationality one wants to study. Hence, it can be directly applied in any future research. The researcher with a specific model in mind needs to answer only (iii) and (iv) for that model. To illustrate this point, we will perform these exercises in Section 4 for the choice overload phenomenon.

The proof of Theorem 1 is constructive and hence provides an algorithm generating the heterogeneous types and their weights *uniquely* from a given probabilistic choice data set. The construction is based on the choice probabilities of lower contour sets with respect to \triangleright .¹² We first calculate all cumulative probabilities on lower contour sets derived from the probabilistic choice. Next we sort these cumulative probabilities from the lowest to the highest, $0 < k_1 < k_2 < \dots < k_T$. Finally, we construct the collection of choices, \mathcal{C} , step by step. The first choice function assigns each alternative set its worst element with respect to \triangleright .¹³ The probability mass of this first choice, c_1 , is the lowest positive cumulative probability given by the aforementioned order, k_1 . In the second step, for each alternative set, we check if the cumulative probability of the lower contour set of the chosen alternative of c_1 equals to k_1 or it is strictly larger than k_1 . For the former case, we assign the second worst alternative as the choice by c_2 ; for the latter case, we keep c_2 equal to c_1 . Note that such a construction assigns the same or better alternative to each alternative set in c_2 than c_1 . The probability assigned to c_2 is $k_2 - k_1$. This procedure continues and defines each c_i based on c_{i-1} while respecting progressiveness, as the choices in each step gradually choose weakly better alternatives on any given set.

Note that Theorem 1 also states a uniqueness result; hence, the construction of the representation not only identifies the exact nature of heterogeneity but also provides a unique weight for each choice type in this heterogeneity. This feature allows a regulator/firm to calculate the effect of a policy on each type of agent in a heterogeneous population and aggregate those effects with uniquely determined weights. In addition to that, the uniqueness result of Theorem 1 can also be used for parameter estimations for the examples mentioned in Classes 2 and 3 above. Once the choice types are uniquely defined within a behavioral class committed by the researcher, she can estimate the parameters of the model. The uniqueness

¹²A similar construction could be done using both lower and upper counter sets and that would lead to the same collection of choice functions. We thank an anonymous reviewer for pointing this.

¹³This worst element needs to be chosen from among the ones which are chosen with positive probability.

of PRC is in sharp contrast to both the general RCM and RUM, which are well-known to admit multiple representations (see Fishburn [1998] for the RUM).

To appreciate the uniqueness result, consider an example with three alternatives. While there are six possible preference orderings, there are twenty four possible choice functions. Even after fixing an exogenous reference order, one can generate a large number of possible collections with the progressive structure with respect to this reference order. The uniqueness comes from the fact that each collection can only have a maximum of six elements due to the progressive structure.¹⁴ Hence, the probabilistic choice data has enough information to uniquely identify the right collection of types. When the alternative set is larger than three elements, even though the number of possible choice functions grows extensively, the maximum number of choice functions in a progressive collection cannot surpass the information given by the probabilistic choice data. On the contrary, in the RUM model, the maximum number of preference orderings exceeds the number of choices in the choice data, which causes undesirable non-uniqueness properties of the RUM.

2.3. An Illustration. We next illustrate how to apply Theorem 1 to actual data. Manzini and Mariotti [2006b] provides a rare data set which includes individuals' choices on each feasible choice set. So we can put our construction to a test. We first aggregate their data to report only the probabilistic choice in Table 1. Then we apply our construction above to this data to reveal the choice types and corresponding frequencies. This allows us to test whether our derived choice profiles match the observed ones.

Manzini and Mariotti [2006b] elicits the time preferences of subjects on three payment schedules, namely an increasing (I), a decreasing (D), a constant (K) schedule.¹⁵ Each plan has three installments (three months, six months, nine months) of a fixed amount of €48. Plan I pays (€8,€16,€24), Plan D pays (€24,€16,€8), and Plan K pays (€16,€16,€16). We provide their aggregate data among 68 subjects.¹⁶

¹⁴The progressive structure is the main driver behind the uniqueness result. A weaker condition, called *non-reversing* property, is introduced by Dardanoni et al. [2020a]. However, the non-reversing property is not sufficient for a unique random choice representation.

¹⁵They also use a fourth alternative. Without loss of generality, we do not include it in our analysis.

¹⁶In order to reduce the noise in the data, we eliminate rare choice profiles and report those that are followed by at least three subjects.

| π | $\{D, K, I\}$ | $\{D, K\}$ | $\{D, I\}$ | $\{K, I\}$ |
|-------|---------------|------------|------------|------------|
| D | 0.71 | 0.75 | 0.96 | – |
| K | 0.29 | 0.25 | – | 1 |
| I | 0 | – | 0.04 | 0 |

TABLE 1. The probabilistic choice induced by the data provided in Manzini and Mariotti [2006b] where $I=(\text{€}8, \text{€}16, \text{€}24)$, $D=(\text{€}24, \text{€}16, \text{€}8)$, and $K=(\text{€}16, \text{€}16, \text{€}16)$, and each vector correspondences to payments in three, six, and nine months, respectively.

First, our construction needs a reference order. These three plans are naturally sorted according to the payment in the first period: $D \triangleright K \triangleright I$. We first order the cumulative choice probabilities on lower contour sets according to \triangleright using Table 1: $0 = \pi(I|\{D, K, I\}) = \pi(I|\{K, I\}) < \pi(I|\{D, I\}) = 0.04 < \pi(K|\{D, K\}) = 0.25 < \pi(\{K, I\}|\{D, K, I\}) = 0.29$. We now construct the choice types with a progressive structure in Table 2. While I is the worst alternative in $\{D, K, I\}$ and $\{K, I\}$, it is chosen with zero probability in those sets. Since the first choice profile must select the worst alternative chosen with the positive probability, that alternative cannot be I on these sets. Hence, $c_1(\{D, K, I\}) = K = c_1(\{K, I\})$. Otherwise, the worst alternative is chosen, i.e., $c_1(\{D, K\}) = K$ and $c_1(\{D, I\}) = I$. The probability mass of c_1 , is the lowest positive cumulative probability observed, 0.04 (see the first row in Table 2). Note that the first choice exhausts all choice probabilities for I on $\{D, I\}$. Hence, all other choice profiles must chose D from $\{D, I\}$ implying that the second choice profile must switch to D on $\{D, I\}$. Since the rest of choice probabilities on other alternative sets are not exhausted yet, we keep the rest the same. The frequency of c_2 determined by $\pi(K|\{D, K\})$ since the choice from $\{D, K\}$ must switch. Hence, the weight of c_2 is $0.21 = \pi(K|\{D, K\}) - \pi(I|\{D, I\})$ (see the second row in Table 2). The rest is constructed similarly.

| | $\{D, K, I\}$ | $\{D, K\}$ | $\{D, I\}$ | $\{K, I\}$ | type frequencies | |
|-------|---------------|------------|------------|------------|--|----------|
| | | | | | derived | observed |
| c_1 | K | K | I | K | $0.04 = \pi(I \{D, I\})$ | 0.04 |
| c_2 | K | K | D | K | $0.21 = \pi(K \{D, K\}) - \pi(I \{D, I\})$ | 0.21 |
| c_3 | K | D | D | K | $0.04 = \pi(\{K, I\} \{D, K, I\}) - \pi(K \{D, K\})$ | 0.04 |
| c_4 | D | D | D | K | $0.71 = 1 - \pi(\{K, I\} \{D, K, I\})$ | 0.71 |

TABLE 2. The collection of choice types driven from Table 1 by using the construction algorithm given in the proof of Theorem 1.

Note that the data in Table 1 is not consistent with a RUM since it violates the regularity assumption¹⁷ $\pi(K|\{D, K\}) < \pi(K|\{D, K, I\})$. The magnitude of this violation is 0.04. It is worth noting that this is the weight assigned to the *only* “irrational type” (c_3) in our construction. Indeed, the derived choice types and their frequencies exactly match the observed data in Manzini and Mariotti [2006b] (see the last two columns in Table 2).

3. COMPARATIVE STATICS

Next we discuss how the comparative statics exercise can be performed to order any two PRC representations for a fixed reference order. Note that this discussion requires only progressiveness on the random choice model; hence, it automatically applies to any PRC model with additional behavioral conditions, such as the one considered in Section 4. To do this, we first introduce an order between distributions of choices in Definition 3. Before defining the order, we define, for all $\alpha \in (0, 1]$, $\mu_\alpha^{-1} := c_i \in \mathcal{C}$ such that $\mu(c_1) + \dots + \mu(c_{i-1}) < \alpha \leq \mu(c_1) + \dots + \mu(c_i)$ for a given $\mathcal{C} = \{c_1, \dots, c_T\}$ and μ . Hence, μ_α^{-1} identifies the choice function in the collection at which the cumulative distribution weakly exceeds α .

Definition 3. Let \triangleright be a reference order. Probability distribution μ defined on \mathcal{C} is **higher** than probability distribution η defined on \mathcal{C}' if $\forall \alpha \in (0, 1]$ and $\forall S \subset X$, $\mu_\alpha^{-1}(S) \triangleright \eta_\alpha^{-1}(S)$.

Definition 3 compares two probability distributions and identifies the one which is more in line with the underlying order, \triangleright , as the higher distribution. Note that the compared distributions do not need to have the same support. This allows us to order two PRCs, π_μ and π_η , with different choice collections as their supports or having the same support with different weights on choices in the support. If it is the latter case, then a distribution being higher simply means it first order stochastically dominates the other distribution. Note that the comparison is based on \triangleright ; hence, the compared models should have the same underlying \triangleright .

We order two probabilistic choices in the standard first order stochastic domination sense, i.e. one dominates the other if it assigns a higher probability of choice to all the upper contour sets defined by \triangleright when choosing from a set. This is formally stated below.

Definition 4. Let \triangleright be a reference order. Probabilistic choice π first order stochastically dominates probabilistic choice π' if for any set S and any $x \in S$, $\pi(U_{\triangleright}(x)|S) \geq \pi'(U_{\triangleright}(x)|S)$.

¹⁷The regularity axiom requires that for all $x \in T \subset S \subset X$, $\pi(x|S) \leq \pi(x|T)$. This is called strict regularity when the inequality is strict.

Now we can state our result on comparative statics between any two PRC representations for \triangleright .

Theorem 2. *Let both π_μ and π_η have a PRC_{\triangleright} representation. π_μ first order stochastic dominates π_η if and only if μ is higher than η .*

Recall that if the choices in the support of PRC are rational and represented by a collection of preferences, our model is equivalent to SCRUM (as stated by Remark 1.) For such models, Definition 3 is equivalent to Definition of “a SCRUM being higher” in Apestegua et al. [2017] (see page 667). Hence, their Proposition 2 is a special case of our Theorem 2.

Also note that if two decision makers (or two populations) have PRCs with the same underlying \triangleright and the same collection of choices in the support, the probabilistic choice of decision maker 1 first order stochastic dominates that of decision maker 2 if and only if the cumulative weighting function of the first decision maker first order stochastic dominates that of the second decision maker. This means that the second decision maker more often engages with choices that are less aligned with the choice rationalized by \triangleright . In other words, she makes worse mistakes (in the sense of not being aligned with \triangleright) more often.

As previously mentioned, two decision makers’ PRC_{\triangleright} may have different supports. For example, say two decision makers use limited attention models of Masatlioglu et al. [2012]. Assume that the first decision maker considers the worst element of a set in her first choice function in the support, then considers the worst two elements in her second choice function and so on. So this person’s consideration sets gradually extend and her choice becomes more aligned with \triangleright . The second person’s support has a single choice which relies on the full consideration set (she is not boundedly rational) and chooses according to the underlying \triangleright (so her choice is degenerate, she is fully attentive and her choice satisfies WARP). Then the probabilistic choice of the more attentive person (the second person) first order stochastic dominates the probabilistic choice of the less attentive one (the first person).

4. AN APPLICATIONS OF PRC FOR BOUNDED RATIONALITY

Our model allows us to address behavior that is possibly inconsistent with utility maximization. Behavioral Economics literature provides abundant evidence outside of the utility

maximization framework. The examples listed under Class 2 and Class 3 in Section 2 illustrate several types of bounded rationality with the PRC structure. A natural question to ask is how one can characterize a certain behavioral phenomena demonstrated by each type in our PRC framework. If each choice type acts according to a behavioral bias at a differentiated degree, what kind of probabilistic choice would they generate in the aggregate data?

To illustrate usefulness of PRC, we consider one of the most studied deviations from the utility maximization: the choice overload phenomenon. This phenomenon is also called “less-is-more.” The idea of less-is-more is based on the evidence that the decision makers may not benefit from having too many choices in a situation due to asymmetric information, limited attention, cognitive capacities, or reference-dependent evaluations. We first commit to less-is-more type of bounded rationality and find the corresponding necessary and sufficient conditions on probabilistic choice in order to have less-is-more type of bounded rationality. Thanks to Theorem 1, we already have a uniqueness result for this application, so we will only focus on the characterization.

As in Section 2, we first assume that the reference order, \triangleright , is observable. This assumption is reasonable in situations in which there is a single common attribute to rank all alternatives, such as the lowest price, shortest distance, the amount of carbon footprint, etc. Later, we will relax this assumption and endogenize the reference order. We next define the less-is-more property for a collection of choice functions.

Definition 5. We say that a collection of choice functions, \mathcal{C} , satisfies *less-is-more* with respect to \triangleright if for all t and for all $T \subset S$, $c_t(S) \in T \Rightarrow c_t(T) \succeq c_t(S)$.

In other words, $c_t(T)$ is more aligned with \triangleright than $c_t(S)$ when $T \subset S$, because the choice from a larger set is dominated by the choice from a smaller set. Note that if the choice functions in the support of randomization are rationalizable by preferences, then the less-is-more property trivially holds.¹⁸ This new concept restricts each possible choice function to be either rational or boundedly rational in the sense of less-is-more.

All the examples discussed under Classes 1-3 in Section 2 can be modified to accommodate the less-is-more structure. For the shortlisting example, in which the shortlists get gradually longer, imagine that the initial shortlist orders the alternatives based on a linear

¹⁸One should note that this observation makes $\text{SCRUM}_{\triangleright}$ a special case of $\text{PRC}_{\triangleright}$ with the less-is-more property.

order that is completely opposite of \triangleright , say $\tilde{\triangleright}$. Such a shortlist would report only the worst alternatives as undominated. Clearly, the choice implied by this shortlist would satisfy “less-is-more” since only a weakly better alternative can be shortlisted and chosen on a smaller set than on a larger set. When the shortlists in that example get gradually longer, due to reverse order implied by $\tilde{\triangleright}$, each choice satisfies less-is-more.

We should note that there are some well known examples that do not satisfy the less-is-more structure. For example, if the attention correspondences of the model described within Class 2 are attention filters (see Masatlioglu et al. [2012]), then the choice functions that are used in the PRC would not satisfy the less-is-more property. Due to the existence of such examples, this more demanding structure improves the prediction power of our model.

4.1. Characterization. Next we state our only axiom for the characterization of boundedly rational types in the sense of less-is-more. Our first axiom requires a monotonicity condition to hold for upper counter sets with respect to a reference order.

Axiom 1. (U-Monotonicity) For all $x \in T \subset S \subseteq X$ such that $\pi(x|S) \neq 0$

$$\pi(U_{\triangleright}(x)|S) \leq \pi(U_{\triangleright}(x)|T)$$

U-Monotonicity intuitively captures less-is-more because the better alternatives are chosen more frequently on smaller subsets. U-Monotonicity resembles the standard regularity condition. In the deterministic case, the regularity condition is equivalent to WARP, but U-Monotonicity is weaker than WARP. Note that if we have a deterministic choice function which satisfies WARP, then it can be represented by a preference relation. In such a case, U-Monotonicity holds with respect to that preference relation. On the other hand, if a deterministic choice function does not satisfy WARP, it may still satisfy U-Monotonicity with respect to an order. For example, consider the choices summarized by $\pi(z|\{x, y, z\}) = 1$, $\pi(x|\{x, y\}) = 1$, $\pi(y|\{y, z\}) = 1$, and $\pi(x|\{x, z\}) = 1$. This choice behavior does not satisfy WARP but it satisfies U-Monotonicity with respect to $x \triangleright y \triangleright z$.

While U-Monotonicity is a generalization of regularity for the deterministic choice, these two conditions are independent in the probabilistic choice environment. For example, a probabilistic choice satisfying strict regularity must violate U-Monotonicity for any order. To see this, take a set S and denote the worst and second worst alternatives in S for a given order

by z and y , respectively. By strict regularity, $\pi(z|S) < \pi(z|\{y, z\})$. This implies $\pi(U_{\succeq}(y)|S) > \pi(U_{\succeq}(y)|\{y, z\})$, a violation of U-Monotonicity. Finally, we must highlight that the inequality in U-Monotonicity might not hold for alternatives chosen with zero probability.¹⁹ We now state our characterization result for the less-is-more type of bounded rationality.

Theorem 3. *Let \triangleright be a reference order. A probabilistic choice π satisfies U-Monotonicity with respect to \triangleright if and only if there exists a unique PRC_{\triangleright} representation of π in which each choice satisfies the less-is-more condition.*

Note that Theorem 3 not only provides a necessary and sufficient condition for the less-is-more representation but also concludes that the representation is unique. The algorithm generating the unique representation is the one provided in the proof of Theorem 1. The proof provided in the Appendix shows that the random choice model generated by this algorithm not only satisfies progressiveness (as shown by Theorem 1) but also satisfies less-is-more given U-Monotonicity.

One might wonder whether the data in Table 1 has a PRC representation with less-is-more structure. Theorem 3 provides a sufficient condition, so we can test it. It is routine to verify that the data satisfies U-Monotonicity for the reference order $D \triangleright K \triangleright I$. Hence, Theorem 3 implies that the collection of choice types in PRC_{\triangleright} has less-is-more structure. Indeed, one can verify the uniquely constructed choice types in Table 2 have that structure.

4.2. Endogenous Reference Order. Up to now, we have taken the reference order as given. This is reasonable in some applications in which the true reference order (such as a social norm or a common attribute such as price or carbon footprint) is observable to the researcher. Our axioms are stated in terms of the reference order, a component of the model. Hence, they should be seen as a test that inputs *both* a probabilistic choice function and an order. For example, U-Monotonicity tests whether the data has a PRC representation with less-is-more condition for a given reference order. The order can convey much (but not all) of the psychology that the PRC captures. For example, say an outside observer believes that

¹⁹The next example shows that a stronger version of U-Monotonicity where we drop the non-zero requirement is not a necessary condition for PRC with less-is-more. Suppose $X = \{x, y, z\}$ where $x \triangleright y \triangleright z$. Consider a random utility that puts 50-50 weights on u_1 and u_2 where $u_1(x) < u_1(y) < u_1(z)$ and $u_2(y) < u_2(z) < u_2(x)$. Note that (u_1, u_2) satisfies single-crossing. Since each choice type is a utility maximizer, then the less-is-more condition is trivially hold. Therefore, this is a PRC satisfying the less-is-more structure. However $\pi(U_{\succeq}(y)|\{x, y, z\}) = 1/2 > 0 = \pi(U_{\succeq}(y)|\{y, z\})$ even though $\{y, z\} \subset \{x, y, z\}$. Hence, the non-zero requirement in the statement of U-Monotonicity is crucial. We thank to an anonymous referee for providing this example.

the heterogeneity is due to being environmental friendly at differentiated levels. However, she cannot decide which attribute (carbon emission or sustainability) sorts environmental friendliness. U-Monotonicity makes it straightforward to determine whether π constitutes a PRC representation with less-is-more under either of these two orders. This helps the outside observer endogenously pick the order which passes the test.

It is also possible that the outside observer does not have any prior knowledge about the reference order. We now describe to what extent one can identify the reference order of a given probabilistic choice function for a PRC representation. Note that this exercise only makes sense for a subclass of probabilistic choice functions, because from Theorem 1 every data has a PRC representation with respect to an arbitrary order. For that reason, we perform this task for the less-is-more structure.

For the revelation of the reference order, we assume that the model is correct and we ask when we can infer the underlying reference order. First, observe that if removing an alternative z from a tripleton causes a regularity violation, $\pi(y|\{x, y, z\}) > \pi(y|\{x, y\})$, we infer that x is ranked above y . To see this, assume for a contradiction that there exists an order \triangleright with $y \triangleright x$ and $\text{PRC}_{\triangleright}$ represents π with a support of types satisfying less-is-more. Since U-Monotonicity must hold for \triangleright , we have

$$\pi(U_{\triangleright}(y)|\{x, y, z\}) \leq \pi(U_{\triangleright}(y)|\{x, y\})$$

This yields $\pi(y|\{x, y, z\}) \leq \pi(y|\{x, y\})$, which contradicts with the assumption. Hence, π has no PRC representation for any reference order that ranks y above x . Similarly, we can show that the observation $\pi(y|S) > \pi(y|\{x, y\})$ where $x \in S$ also reveals that x is ranked above y , which is denoted by $x \triangleright_{\pi} y$.

There are two other choice patterns revealing $x \triangleright_{\pi} y$. Specifically, if there exists a z revealing $z \triangleright_{\pi} y$ (i.e., $\pi(y|\{x, y, z\}) > \pi(y|\{y, z\})$) and $\pi(x|\{x, y, z\}) < \pi(x|\{x, y\})$, we must have x is revealed to be better than y . To see this, assume x is not ranked above y , then there exists an order \triangleright with $y \triangleright x$ and $\text{PRC}_{\triangleright}$ represents π with a support satisfying less-is-more. Since we must have $z \triangleright_{\pi} y$, we must have $z \triangleright y \triangleright x$. Since $\pi(y|\{x, y, z\}) \neq 0$, U-Monotonicity for y with respect to \triangleright implies that $\pi(x|\{x, y, z\}) \geq \pi(x|\{x, y\})$, a contradiction. Hence x must be ranked above y in every PRC representation, i.e., $x \triangleright_{\pi} y$.

The second choice pattern is more involved. We argue that if there exists z such that $\pi(z|\{x, y, z\}) > \pi(z|\{y, z\})$, $\pi(x|\{x, y, z\}) < \pi(x|\{x, y\})$, $\pi(z|\{x, y, z\}) < \pi(z|\{x, z\})$, and $\pi(x|\{x, y, z\}) \neq 0$, then x must be ranked above y . To see this, assume that x is not ranked above y . The former inequality implies that $y \triangleright_\pi z$. Hence, there are two possible reference orderings: $y \triangleright_1 z \triangleright_1 x$ or $y \triangleright_2 x \triangleright_2 z$. Since $\pi(z|\{x, y, z\}) \neq 0$, U-Monotonicity for z with respect to \triangleright_1 implies that $\pi(x|\{x, y, z\}) \geq \pi(x|\{x, y\})$. Similarly, assuming $\pi(x|\{x, y, z\}) \neq 0$, U-Monotonicity for x with respect to \triangleright_2 implies that $\pi(z|\{x, y, z\}) \geq \pi(z|\{x, z\})$. Both cases imply a contradiction. Hence x must be ranked above y in every PRC representation satisfying less-is-more, i.e., $x \triangleright_\pi y$.

We now formally state the above observations: For any distinct x and y , define the following binary relation,

$$\begin{aligned}
 x \triangleright_\pi y \quad \text{if} \quad & (i) \pi(y|S) > \pi(y|\{x, y\}) \text{ for some } S \ni x, \\
 & (ii) \exists z \text{ s.t. } \pi(y|\{x, y, z\}) > \pi(y|\{y, z\}) \text{ and } \pi(x|\{x, y, z\}) < \pi(x|\{x, y\}) \\
 & (iii) \exists z \text{ s.t. } \pi(z|\{x, y, z\}) > \pi(z|\{y, z\}), \pi(x|\{x, y, z\}) < \pi(x|\{x, y\}), \\
 & \pi(z|\{x, y, z\}) < \pi(z|\{x, z\}), \text{ and } \pi(x|\{x, y, z\}) \neq 0
 \end{aligned}$$

If we have $x \triangleright_\pi z$ and $z \triangleright_\pi y$ revealed, we must have x to be ranked above y by transitivity even though $x \triangleright_\pi y$ is not revealed. The transitive closure of \triangleright_π , denoted by \triangleright_π^T , includes these additional revelations as well as \triangleright_π itself. The next proposition summarizes this observation.²⁰

Proposition 1. *[Revealed Preference] If π has a PRC $_{\triangleright}$ representation satisfying the less-is-more property, then \triangleright_π^T constructed above must be included in \triangleright (i.e., $\triangleright_\pi^T \subseteq \triangleright$).*

Given Proposition 1, \triangleright_π^T can help test endogenously whether the probabilistic choice function satisfies Axiom 1 (i.e., it has a PRC with less-is-more property with respect to the order.) Given π , we can first derive \triangleright_π^T as described above. If there is a cycle, then this means that π cannot be represented by a PRC satisfying less-is-more. If it does not have any cycle, then the implied \triangleright_π^T restricts the set of all reference orders which may be compatible

²⁰Since \triangleright_π is a subset of the reference order, it must be asymmetric as well.

with π . Indeed, if \triangleright_{π}^T is complete (see Theorem 4 for conditions to have a complete \triangleright_{π}^T), it must be the underlying reference order. Hence, we not only establish that the data has a PRC representation satisfying less-is-more condition, but also infer the underlying reference order endogenously. If \triangleright_{π}^T is not complete, then one may check Axiom 1 on this restricted set of orders including \triangleright_{π}^T . If this restricted set contains an order satisfying Axiom 1, then this order provides a PRC representation for the choice data. If none of them satisfies Axiom 1, then π cannot be represented by a PRC representation satisfying less-is-more condition.

To utilize Proposition 1, we revisit the data in Table 1. Note that $D \triangleright_{\pi} K$. This is a direct consequence of definition of \triangleright_{π} (part (i)). Proposition 1 implies that any PRC representation satisfying less-is-more condition must rank D over K . However, the data in Table 1 does not reveal any other reference order. In this example, \triangleright_{π} is incomplete. The next remark highlights that if the revealed order is complete and U-Monotonicity holds with respect to it, then the data has the endogenous PRC representation satisfying less-is-more.

Remark 5. If \triangleright_{π}^T is a linear order and π satisfies U-Monotonicity with respect to \triangleright_{π}^T , then Theorem 3 and Proposition 1 yield that π has a PRC representation satisfying less-is-more. More importantly, \triangleright_{π}^T is the unique reference order, which is derived endogenously from the data.

We next provide sufficient conditions for the completeness of the revealed reference order. Both of these condition can be directly checked from the data. The first one is Weak Binary Regularity, which is substantially weaker than regularity.

Axiom 2. (Weak Binary Regularity) For all $x \in S \subset X$ with $|S| \geq 2$,

$$\pi(x|S) \leq \max\{\pi(x|\{x, y\}) \mid x \neq y \in S\}$$

The standard regularity requires the choice probability of an alternative in a larger set to be weakly less than all the choice probabilities on every smaller sets (including both binary and non-binary sets). However, Axiom 2 requires the inequality to hold only for the maximum of the binary choice probabilities. So, our postulate allows binary regularity violations but it restricts the severity of them. Unlike U-Monotonicity, it does not require the knowledge of the reference order, hence it can be checked directly using the probabilistic choice data.

Axiom 3. For all $S \subset X$ with $|S| = 3$, $\pi(x|S) > \pi(x|\{x, y\})$ for some $x, y \in S$.

Note that if strict probabilistic choice data²¹ satisfies U-Monotonicity with respect to some order, it must satisfy Axiom 3. Hence, it needs to be satisfied in order to have a less-is-more representation. Again unlike U-Monotonicity, it can be directly verified using the stochastic choice data.

The next theorem shows that these two axioms are sufficient for the revealed reference order being complete.

Theorem 4. *If a strict probabilistic choice function π satisfies Axioms 2 and 3, then \triangleright_π^T is complete.*

Theorem 4 provides sufficient conditions for completeness of the revealed reference order. Then by Remark 5, we have a unique candidate for the endogenous reference order. Therefore, Theorem 4 is an identification result revealing endogenous reference order uniquely.

The next example illustrates the importance of strictness and Axiom 2 for Theorem 4.

Example 1. The following table provides a set of parametric probabilistic choice described by π_λ where $\lambda \in [0, 0.15)$.

| π_λ | $\{x, y, z\}$ | $\{x, y\}$ | $\{x, z\}$ | $\{y, z\}$ |
|---------------|-------------------|------------|------------|------------|
| x | $0.10 + \lambda$ | 0.30 | 0.40 | – |
| y | $0.30 - 2\lambda$ | 0.70 | – | 0.55 |
| z | $0.60 + \lambda$ | – | 0.60 | 0.45 |

When λ is positive, we have $\pi_\lambda(z|\{x, y, z\}) > \pi_\lambda(z|\{y, z\})$ and $\pi_\lambda(z|\{x, y, z\}) > \pi_\lambda(z|\{x, z\})$, which imply $y \triangleright_\pi z$ and $x \triangleright_\pi z$. However, we cannot conclude the order between x and y . It is routine to check that π_λ has multiple PRC representations satisfying less-is-more property with respect to both $x \triangleright_1 y \triangleright_1 z$ and $y \triangleright_2 x \triangleright_2 z$, when $\lambda \geq 0$. When $\lambda = 0$, the data violates strictness but Axiom 2 is satisfied. On the other hand, when $\lambda > 0$, the strictness holds but not Axiom 2.

5. RELATED LITERATURE

In this section, we compare our model with other well-known models of probabilistic choice from the literature. First, note that in terms of the explanatory power, PRC includes

²¹We call a probabilistic choice function strict if for all x, S, S' , $\pi(x|S) \neq \pi(x|S') > 0$.

all other models (see Theorem 1). Since the PRC with less-is-more condition imposes testable restrictions, we now compare this special sub-class to other probabilistic choice models. As we mentioned before, the SCRUM of Apesteguia et al. [2017] is a special case of this sub-class. Moreover, the PRC satisfying less-is-more includes some other RUM choices other than SCRUM. Hence, the PRC satisfying less-is-more has more explanatory power than SCRUM. It is also true that the PRC with less-is-more is distinct from RUM.

The Attribute Model of Gul, Natenzon, and Pesendorfer [2014] and the Fixed Distribution Satisficing Model of Aguiar, Boccardi, and Dean [2016] are both subsets of RUM. Hence, they are distinct from our model in terms of observed choices. Similarly, the Additive Perturbed Utility model of Fudenberg, Iijima, and Strzalecki [2015] satisfies regularity. Since our model allows for violations of regularity, they are distinct models.

The General Luce Model of Echenique and Saito [2019]²² and the Perception-Adjusted Luce Model of Echenique, Saito, and Tserenjigmid [2018] are generalizations of the Luce Model. When choice probabilities are strict and the outside option is never chosen, these two models reduce to the Luce rule. However, the PRC with less-is-more allows for violations of Luce’s independence of irrelevant alternatives condition under these assumptions.

Manzini and Mariotti [2014], Brady and Rehbeck [2016], and Cattaneo et al. [2019] provide probabilistic choice models in which randomness comes from random consideration rather than random preferences. While the first two provide parametric random attention models, the last offers a non-parametric restriction on the random attention rule. The first two models require the existence of a default option for their models. To provide an accurate comparison, we consider versions of those without an outside/default option.²³ The random attention model (RAM) of Cattaneo et al. [2019] covers the model of Brady and Rehbeck [2016] (BR), which in turn contains the model of Manzini and Mariotti [2014] (MM). All these models include preferences as one of the components of their models. First, we state the differences of these models for a given reference order. RAM includes RUM, BR, SCRUM and MM. However, RAM and PRC satisfying less-is-more are independent models because neither one is a subset of the other. When we consider endogenous reference order, RAM still includes RUM, BR, SCRUM and MM. In addition, it is still true that PRC satisfying

²²See Ahumada and Ulku [2018] and Horan [2018b] for related models.

²³See Horan [2018a] for an axiomatic characterization of the Manzini and Mariotti [2014] model when there is no default option.

less-is-more is different from RAM. For example, consider the following probabilistic choice with three alternatives, π : $\pi(z|\{x, y, z\}) = \pi(y|\{x, y, z\}) = \pi(z|\{y, z\}) = 0.3$, and $\pi(y|\{x, y\}) = \pi(z|\{x, z\}) = 0.2$. π is a PRC satisfying less-is-more condition but not RAM. Indeed, this example is outside of any model discussed above.

Finally, Cattaneo et al. [2021] proposes another random attention model in which alternatives compete for the decision maker's attention. When there are more alternatives, the decision maker pays less attention to each alternative (Attention Overload). Their axiom is weaker than U-Monotonicity, hence the PRC satisfying less-is-more is a special case of that model. We should note that our revealed preference relation is always richer than theirs.

6. CONCLUSION

We have introduced a novel PRC model that uniquely identifies the heterogeneous types that are possibly boundedly rational. Our examples illustrate that this model can prove useful in economic contexts in which one wishes to investigate interpersonal or intrapersonal variation in choices caused by a sorted behavioral trait such as willpower, loss aversion, attention, or limited cognitive ability.

The applications of the single-crossing property have been fruitful in mechanism design settings. Since the progressive structure allows us to extend the same logic to settings with bounded rationality, a natural avenue is to explore the implications of progressiveness in behavioral mechanism design.

In this paper, we investigated one specific application of our framework to study the less-is-more structure. That application highlights that a researcher with a specific type of bounded rationality in mind may investigate the properties of the data in order to have heterogeneous progressive types to behave according to that bounded rationality. In our application, U-Monotonicity characterizes the less-is-more types. However, there are some other well-studied bounded rationality models such as the limited attention model of (Masatlioglu et al. [2012]) that do not satisfy the less-is-more structure. Hence, another avenue for exploration is to study other such bounded rationality structures for the choices in the collection and their behavioral implications.

Finally, several empirical queries arise from the present work. Conducting experimental tests comparing the explanatory power of competing models that we reviewed in this paper would certainly be useful.

REFERENCES

- ABALUCK, J. AND A. ADAMS (2017): “What Do Consumers Consider Before They Choose? Identification from Asymmetric Demand Responses,” *NBER Working Paper No. 23566*.
- AGUIAR, V. H., M. J. BOCCARDI, AND M. DEAN (2016): “Satisficing and Stochastic Choice,” *Journal of Economic Theory*, 166, 445–482.
- AHUMADA, A. AND L. ULKU (2018): “Luce Rule with Limited Consideration,” *Mathematical Social Sciences*, 93, 52–56.
- ALBAYRAK, S. AND F. ALESKEROV (2000): “Convexity of choice function sets,” *Bogazici University Research Paper, ISS/EC-2000-01*.
- APESTEGUIA, J. AND M. BALLESTER (2020): “Random Utility Models with Ordered Types and Domains,” *Working Paper, University of Oxford*.
- APESTEGUIA, J., M. A. BALLESTER, AND J. LU (2017): “Single-Crossing Random Utility Models,” *Econometrica*, 85, 661–674.
- BARSEGHYAN, L., M. COUGHLIN, F. MOLINARI, AND J. C. TEITELBAUM (2018): “Heterogeneous Consideration Sets and Preferences,” work in progress, Cornell University.
- BARSEGHYAN, L., F. MOLINARI, AND M. THIRKETTLE (2019): “Discrete choice under risk with limited consideration,” *arXiv preprint arXiv:1902.06629*.
- BRADY, R. L. AND J. REHBECK (2016): “Menu-Dependent Stochastic Feasibility,” *Econometrica*, 84, 1203–1223.
- CAPLIN, A., M. DEAN, AND D. MARTIN (2009): “Search and Satisficing,” Working Paper.
- CATTANEO, M. D., P. CHEUNG, X. MA, AND Y. MASATLIOGLU (2021): “Attention Overload,” .
- CATTANEO, M. D., X. MA, Y. MASATLIOGLU, AND E. SULEYMANOV (2019): “A Random Attention Model,” *Journal of Political Economy*, Forthcoming.
- CHEREPANOV, V., T. FEDDERSEN, AND A. SANDRONI (2013): “Rationalization,” *Theoretical Economics*, 8, 775–800.
- CHERNEV, A. (2003): “When More Is Less and Less Is More: The Role of Ideal Point Availability and Assortment in Consumer Choice,” *Journal of Consumer Research*, 30, 170–183.
- CHERNEV, A., U. BÖCKENHOLT, AND J. GOODMAN (2015): “Choice overload: A conceptual review and meta-analysis,” *Journal of Consumer Psychology*, 25, 333–358.

- CHIAPPORI, P.-A., B. SALANIÉ, F. SALANIÉ, AND A. GANDHI (2019): “From Aggregate Betting Data to Individual Risk Preferences,” *Econometrica*, 87, 1–36.
- DARDANONI, V., P. MANZINI, M. MARIOTTI, P. HENRIK, AND C. J. TYSON (2020a): “Mixture Choice Functions: A Tool to Identify Preferences and Cognition,” *Working Paper*.
- DARDANONI, V., P. MANZINI, M. MARIOTTI, AND C. J. TYSON (2020b): “Inferring Cognitive Heterogeneity from Aggregate Choices,” *Econometrica*, *forthcoming*.
- DEKEL, E., B. L. LIPMAN, AND A. RUSTICHINI (2009): “Temptation-Driven Preferences,” *The Review of Economic Studies*, 76, 937–971.
- DILLENBERGER, D. AND P. SADOWSKI (2012): “Ashamed to be selfish,” *Theoretical Economics*, 7, 99–124.
- ECHENIQUE, F., S. LEE, AND M. SHUM (2011): “The money pump as a measure of revealed preference violations,” *Journal of Political Economy*, 119, 1201–1223.
- ECHENIQUE, F. AND K. SAITO (2019): “General Luce model,” *Economic Theory*, 68, 811–826.
- ECHENIQUE, F., K. SAITO, AND G. TSERENJIGMID (2018): “The Perception-Adjusted Luce Model,” *Mathematical Social Sciences*, 93, 67–76.
- EPPLE, D., T. ROMER, AND H. SIEG (2001): “Interjurisdictional sorting and majority rule: an empirical analysis,” *Econometrica*, 69, 1437–1465.
- FISHBURN, P. C. (1998): “Stochastic utility,” *Handbook of utility theory*, 1, 273–319.
- FUDENBERG, D., R. IJIMA, AND T. STRZALECKI (2015): “Stochastic Choice and Revealed Perturbed Utility,” *Econometrica*, 83, 2371–2409.
- FUDENBERG, D. AND D. LEVINE (2006): “A dual-self model of impulse control,” *The American Economic Review*, 1449–1476.
- GANS, J. S. AND M. SMART (1996): “Majority voting with single-crossing preferences,” *Journal of public Economics*, 59, 219–237.
- GRANDMONT, J.-M. (1978): “Intermediate preferences and the majority rule,” *Econometrica: Journal of the Econometric Society*, 317–330.
- GUL, F., P. NATENZON, AND W. PESENDORFER (2014): “Random Choice as Behavioral Optimization,” *Econometrica*, 82, 1873–1912.
- GUL, F. AND W. PESENDORFER (2001): “Temptation and Self-Control,” *Econometrica*, 69, 1403–1435.
- HORAN, S. (2018a): “Random Consideration and Choice: A Case Study of “Default” Options,” *Working Paper, Université de Montréal and CIREQ*.

- (2018b): “Threshold Luce Rules,” .
- HORAN, S. AND Y. SPRUMONT (2016): “Welfare criteria from choice: An axiomatic analysis,” *Games and Economic Behavior*, 99, 56–70.
- HUBER, J., J. W. PAYNE, AND C. PUTO (1982): “Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis,” *Journal of Consumer Research*, 9, 90–98.
- IYENGAR, S., G. HUBERMAN, AND W. JIANG (2004): “How much choice is too much? Contributions to 401(k) retirement plans,” in *Pension Design and Structure*, Oxford University Press, chap. 5, 83–97.
- IYENGAR, S. S. AND M. R. LEPPER (2000): “When choice is demotivating: can one desire too much of a good thing?” *Journal Personality and Social Psychology*, 79, 995–1006.
- KAMENICA, E. (2008): “Contextual inference in markets: On the informational content of product lines,” *American Economic Review*, 98, 2127–49.
- KELMAN, M., Y. ROTTENSTREICH, AND A. TVERSKY (1996): “Context-dependence in legal decision making,” *The Journal of Legal Studies*, 25, 287–318.
- KŐSZEGI, B. AND M. RABIN (2006): “A Model of Reference-Dependent Preferences,” *Quarterly Journal of Economics*, 121, 1133–1165.
- LLERAS, J. S., Y. MASATLIOGLU, D. NAKAJIMA, AND E. Y. OZBAY (2017): “When More Is Less: Limited Consideration,” *Journal of Economic Theory*, 170, 70–85.
- MANZINI, P. AND M. MARIOTTI (2006a): “Consumer Choice and Revealed Bounded Rationality,” Working Papers 571, Queen Mary, University of London, Department of Economics.
- (2006b): “Two-Stage Boundedly Rational Choice Procedures: Theory and Experimental Evidence,” *Working Paper*.
- (2007): “Sequentially Rationalizable Choice,” *American Economic Review*, 97, 1824–1839.
- (2012): “Categorize Then Choose: Boundedly Rational Choice and Welfare,” *Journal of the European Economic Association*, 10, 1141–1165.
- (2014): “Stochastic Choice and Consideration Sets,” *Econometrica*, 82, 1153–1176.
- MARIOTTI, M. AND P. MANZINI (2012): “Choice by lexicographic semiorders,” *Theoretical Economics*, 7.
- MASATLIOGLU, Y., D. NAKAJIMA, AND E. Y. OZBAY (2012): “Revealed Attention,” *American Economic Review*, 102, 2183–2205.

- MASATLIOGLU, Y., D. NAKAJIMA, AND E. OZDENOREN (2020): “Willpower and compromise effect,” *Theoretical Economics*, 15, 279–317.
- MCCAUSLAND, W. J., C. DAVIS-STOBER, A. MARLEY, S. PARK, AND N. BROWN (2020): “Testing the random utility hypothesis directly,” *The Economic Journal*, 130, 183–207.
- McFADDEN, D. AND M. 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.
- MILGROM, P. AND C. SHANNON (1994): “Monotone comparative statics,” *Econometrica: Journal of the Econometric Society*, 157–180.
- MIRRELES, J. A. (1971): “An exploration in the theory of optimum income taxation,” *The review of economic studies*, 38, 175–208.
- NOOR, J. AND N. TAKEOKA (2010): “Uphill self-control,” *Theoretical Economics*, 5, 127–158.
- PRELEC, D., B. WERNERFELT, AND F. ZETTELMEYER (1997): “The role of inference in context effects: Inferring what you want from what is available,” *Journal of Consumer research*, 24, 118–125.
- RATNESHWAR, S., A. D. SHOCKER, AND D. W. STEWART (1987): “Toward Understanding the Attraction Effect: The Implications of Product Stimulus Meaningfulness and Familiarity,” *Journal of Consumer Research*, 13, 520–533.
- ROBERTS, K. W. (1977): “Voting over income tax schedules,” *Journal of public Economics*, 8, 329–340.
- ROTHSTEIN, P. (1990): “Order restricted preferences and majority rule,” *Social choice and Welfare*, 7, 331–342.
- TRUEBLOOD, J. S., S. D. BROWN, A. HEATHCOTE, AND J. R. BUSEMEYER (2013): “Not just for consumers: Context effects are fundamental to decision making,” *Psychological science*, 24, 901–908.
- TVERSKY, A. AND I. SIMONSON (1993): “Context-dependent preferences,” *Management science*, 39, 1179–1189.
- TYSON, C. J. (2013): “Behavioral Implications of Shortlisting Procedures,” *Social Choice and Welfare*, 41, 941–963.

APPENDIX

Proof of Remark 1. Assume that $\{u_1, \dots, u_T\}$ satisfies Condition 1 with respect to \triangleright . For a contradiction, suppose the corresponding choice collection does not satisfy the progressiveness property. Then there exist $s, t \in \{1, \dots, T\}$ such that $t > s$ and $S \subset X$ where $c_s(S) \triangleright c_t(S)$. Since each u_i rationalizes the corresponding c_i , we have $u_s(c_s(S)) > u_s(c_t(S))$ and $u_t(c_t(S)) > u_t(c_s(S))$. Note that since $c_s(S) \triangleright c_t(S)$ by Condition 1 we must have $u_s(c_s(S)) > u_s(c_t(S)) \Rightarrow u_t(c_s(S)) > u_t(c_t(S))$, which is a contradiction.

For the other direction of the proof, assume that the collection of choices satisfies the progressiveness property with respect to \triangleright . For a contradiction, suppose the corresponding set of preferences does not satisfy Condition 1. Then there exists $x, y \in X$ such that $x \triangleright y$, $s, t \in \{1, \dots, T\}$ with $s > t$ and while $u_t(x) > u_t(y)$ we have $u_s(y) > u_s(x)$. Then $c_t(\{x, y\}) = x$ and $c_s(\{x, y\}) = y$. By progressiveness, we should have $c_s(\{x, y\}) \succeq c_t(\{x, y\})$ or equivalently, $y \succeq x$. This is a contradiction. \square

Proof of Remark 2. Assume Condition 2 with respect to \triangleright on a collection of $(u, \{\Gamma_t\})$ within Class 2 where u and \triangleright are ordinally equivalent. Let $t > s$ and $S \subseteq X$ then by definition of this class, $c_s(S) \in \Gamma_s(S)$. Then observe that $c_s(S) \in L_{\succeq}(\Gamma_s(S)) \subseteq L_{\succeq}(\Gamma_t(S))$. This implies existence of $y \in \Gamma_t(S)$ s.t. $y \succeq c_s(S)$. Since $c_t(S) \succeq x$ for any $x \in \Gamma_t(S)$, in particular we must have $c_t(S) \succeq y \succeq c_s(S)$. Hence, $\{c_t\}$ is a progressive collection with respect to \triangleright .

For the other direction of the proof assume that the collection of choices satisfies the progressive property with respect to \triangleright which is ordinally equivalent to u . Let $t > s$, S be a set of alternatives and $x \in L_{\succeq}(\Gamma_s(S))$. Then $\exists y \in \Gamma_s(S)$ s.t. $y \succeq x$. Then by definition $c_s(S) \succeq x$ which together with the progressiveness implies $c_t(S) \succeq c_s(S) \succeq x$. Then $x \in L_{\succeq}(\Gamma_t(S))$. This shows that $L_{\succeq}(\Gamma_s(S)) \subseteq L_{\succeq}(\Gamma_t(S))$. \square

Proof of Remark 3. Assume that a collection $(u, \{\kappa_t\})$ within Class 3 satisfies Condition 3 with respect to some \triangleright . For contradiction assume there are two types $t > s$ and $S \subseteq X$ such that $c_s(S) \triangleright c_t(S)$. Then by Condition 3, $u(c_s(S)) - \kappa_s(c_s(S), S) > u(c_t(S)) - \kappa_s(c_t(S), S)$ imply $u(c_s(S)) - \kappa_t(c_s(S), S) > u(c_t(S)) - \kappa_t(c_t(S), S)$. Since $c_s(S)$ is chosen by type s , the left hand side of the above statement is true. Then the left hand side must be true too but that contradicts with $c_t(S)$ being chosen by type t .

For the other direction of the proof, assume that the collection of choices satisfies the progressive property with respect to \triangleright . Let S be a set of alternatives containing two elements, $x \triangleright y$ and $t > s$. For contradiction, let $u(x) - \kappa_s(x, S) > u(y) - \kappa_s(y, S)$ but $u(x) - \kappa_t(x, S) < u(y) - \kappa_t(y, S)$ then $c_s(\{x, y\}) = x$ and $c_t(\{x, y\}) = y$. Then $c_s(\{x, y\}) \triangleright c_t(\{x, y\})$ which contradicts with the progressiveness. \square

Proof of Remark 4. Class 1 \subset Class 2: Let a collection $\{u_t\}_{t=1}^T$ with induced choice types of $\{c_t\}$ be in Class 1. We will define an equivalent collection of choice functions which are generated by a model in Class 2. Define $u := u_T$ and for any t and S , $\Gamma_t(S) = \{x \in S \mid u(c_t(S)) \geq u(x)\}$. Hence, we construct the constraint sets from the lower contour sets with respect to

u of the chosen element of type t . Then define $\tilde{c}_t(S) := \operatorname{argmax}_{x \in \Gamma_t(S)} u(x)$. By definition, $\{\tilde{c}_t(S)\}$ is in Class 2. Note that $c_t(S) \in \Gamma_t(S)$ by definition and $c_t(S) = \tilde{c}_t(S)$ for any S .

Class 2 c Class 3: Let a collection $\{c_t\}$ be generated by (u, Γ_t) in Class 2. We will define an equivalent collection of choice functions which are generated by a model in Class 3. Define \triangleright as the order induced by u , $v := -u$ and define the menu dependent behavioral cost functions as:

$$\kappa_t(x, S) = \begin{cases} -2u(x) & \text{if } x \in L_{\triangleright}(\Gamma_t(S)) \\ M & \text{if } x \notin L_{\triangleright}(\Gamma_t(S)) \end{cases}$$

where $M > \max_{x \in X} \{-2u(x)\}$. Consider $\tilde{c}_t(S) := \operatorname{argmax}_{x \in S} v(x) - \kappa_t(x, S)$. By definition, the collection $\{\tilde{c}_t\}$ is in Class 3. We now show $c_t(S) = \tilde{c}_t(S)$ for all S . For any $y \in L_{\triangleright}(\Gamma_t(S))$, $v(c_t(S)) - \kappa_t(x, S) = -u(c_t(S)) + 2u(c_t(S)) = u(c_t(S)) \geq u(y) = -u(y) + 2u(y) = v(y) - \kappa_t(y, S)$. And for any $y \notin L_{\triangleright}(\Gamma_t(S))$, $v(c_t(S)) - \kappa_t(x, S) = u(c_t(S)) > -u(y) - M = v(y) - \kappa_t(y, S)$. Hence, $c_t(S) = \tilde{c}_t(S)$ for any S .

Proof of Theorem 1. Let \triangleright be an order and a probabilistic choice function π be given. We will construct a collection of choice functions, \mathcal{C} , with the desired structure with respect to \triangleright and a probability distribution μ on \mathcal{C} such that $\pi_{\mu} = \pi$.

Define

$$K = \{\pi(L_{\triangleright}(x)|S) \mid S \subseteq X \text{ and } x \in S\}$$

This defines a collection of all cumulative probabilities on lower contour sets derived from the probabilistic choice. K is a finite subset of $[0, 1]$. Next we sort the strictly positive elements of K from the lowest to the highest, i.e., $0 < k_1 < k_2 < \dots < k_m = 1$.²⁴ Note that since X is finite, m is finite.

Next we will construct the set of choice functions, \mathcal{C} , recursively. Before that, we define a minimizing operator $\min_{\pi^+}(\triangleright, S)$, which selects the worst alternative in S according to \triangleright with strictly positive choice probability. That is,

$$\min_{\pi^+}(\triangleright, S) = \{x \in S \mid \pi(x|S) > 0 \text{ and } y \triangleright x \text{ whenever } \pi(y|S) > 0 \text{ and } y \neq x\}$$

For any set S , follow the steps below:

Step 1: Define

$$c_1(S) = \min_{\pi^+}(\triangleright, S) \text{ and } \mu(c_1) = k_1$$

Note that $\mu(c_1)$ is positive and for any S , $\pi(c_1(S)|S) = \pi(L_{\triangleright}(c_1(S))|S) \geq k_1$ as $\pi(c_1(S)|S)$ is an element of K and by definition k_1 is the smallest of those probabilities. Moreover, there exists a subset S such that $\pi(L_{\triangleright}(c_1(S))|S) = k_1$ since $k_1 \in K$.

Step 2: Define the second choice type as

$$c_2(S) = \begin{cases} c_1(S) & \text{if } \pi(L_{\triangleright}(c_1(S))|S) > k_1 \\ \min_{\pi^+}(\triangleright, S \setminus c_1(S)) & \text{if } \pi(L_{\triangleright}(c_1(S))|S) = k_1 \end{cases} \text{ and } \mu(c_2) = k_2 - k_1$$

²⁴ k_m is always equal to 1 since $\pi(L_{\triangleright}(x)|\{x\}) = 1$.

This is well-defined because by the construction in the first step: $\pi(L_{\succeq}(c_1(S))|S) \geq k_1$. Note that $\mu(c_2)$ is strictly positive as $k_1 < k_2$, and by step 1, c_1 is different from c_2 . Observe that for any S , $c_2(S) \succeq c_1(S)$ by definition of c_2 and hence, $\{c_1, c_2\}$ satisfies progressiveness with respect to \triangleright . Note that $\mu(c_1) + \mu(c_2) = k_2$. Moreover, there exists a subset S such that $\pi(L_{\succeq}(c_2(S))|S) = k_2$ since $k_2 \in K$.

Step i : Define the i^{th} choice as

$$c_i(S) = \begin{cases} c_{i-1}(S) & \text{if } \pi(L_{\succeq}(c_{i-1}(S))|S) > k_{i-1} \\ \min_{\pi^+}(\triangleright, S \setminus \bigcup_{k=1}^{i-1} c_{i-k}(S)) & \text{if } \pi(L_{\succeq}(c_{i-1}(S))|S) = k_{i-1} \end{cases}$$

and $\mu(c_i) = k_i - k_{i-1}$

This is well-defined because by construction in first $i - 1$ steps

$$\pi(L_{\succeq}(c_{i-1}(S))|S) = \sum_{\substack{y \preceq c_{i-1}(S) \\ y \in S}} \pi(y|S) \geq k_{i-1}$$

Note that by step $i - 1$, $c_{i-1} \neq c_i$, and by construction $c_i(S) \succeq c_{i-1}(S) \succeq c_{i-2}(S) \succeq \dots \succeq c_1(S)$ for all S . Hence, $\{c_1, c_2, \dots, c_i\}$ consists of distinct elements and satisfies progressiveness with respect to \triangleright . Note that $\sum_{t=1}^i \mu(c_t) = k_i$. This construction stops when we reach m^{th} step.

Define $\mathcal{C} = \{c_1, \dots, c_m\}$ where each c_i is defined in Step i above. Since \mathcal{C} satisfies progressiveness with respect to \triangleright , and $\sum_{t=1}^m \mu(c_t) = k_1 + \sum_{t=2}^m (k_t - k_{t-1}) = k_m = 1$, (μ, \mathcal{C}) constitutes a PRC, denoted by π_μ . That is,

$$\pi_\mu(x|S) = \sum_{\substack{x=c_k(S) \\ c_k \in \mathcal{C}}} \mu(c_k)$$

We need to show that the representation holds, i.e., $\pi_\mu = \pi$. Note that by construction $\pi_\mu(x|S) = 0$ for any $x \in S$ such that $\pi(x|S) = 0$.

Let $x \in S$ be an element with $\pi(x|S) \neq 0$. Let $\pi(L_{\succeq}(x)|S) = k_i$ and $\pi(L_{\triangleright}(x)|S) = k_j$. Since $L_{\triangleright}(x) \subset L_{\succeq}(x)$ and $\pi(x|S) \neq 0$, k_i is strictly greater than k_j . Then by construction, we have $c_{j+1}(S) = \dots = c_i(S) = x$. In addition, for all $k \leq j$, $x \triangleright c_k(S)$ and $x \triangleleft c_k(S)$ for all $k \geq i + 1$. Then we have

$$\begin{aligned} \pi_\mu(x|S) &= \sum_{t=j+1}^i \mu(c_t) = \sum_{t=j+1}^i (k_t - k_{t-1}) = k_i - k_j \\ &= \pi(L_{\succeq}(x)|S) - \pi(L_{\triangleright}(x)|S) \\ &= \pi(x|S) \end{aligned}$$

Hence, π_μ and π are the same.

(Uniqueness): Let μ_1 with support $\mathcal{C}^1 = \{c_1^1, \dots, c_{n_1}^1\}$ and μ_2 with support $\mathcal{C}^2 = \{c_1^2, \dots, c_{n_2}^2\}$ be two PRC representations of the same stochastic data described by π such that \mathcal{C}^1 and \mathcal{C}^2 satisfy progressiveness. We want to show that $\mathcal{C}^1 = \mathcal{C}^2$ and $\mu_1 = \mu_2$.

For contradiction, suppose $\mu_1 \neq \mu_2$. Define the c.d.f. implied by μ_i as $M_i(c_t^i) = \sum_{s \leq t} \mu_i(c_s^i)$ for $i = 1, 2$. Let M_i^{-1} be the inverse choice defined from the c.d.f such that

$$M_i^{-1}(\alpha) = \{c_t^i | M_i(c_{t-1}^i) < \alpha \leq M_i(c_t^i)\}$$

for $i = 1, 2$. Since $\mu_1 \neq \mu_2$, then there must be an $\alpha \in (0, 1)$ such that $M_1^{-1}(\alpha) \neq M_2^{-1}(\alpha)$. Let $M_1^{-1}(\alpha) = \{c_t^1\}$ and $M_2^{-1}(\alpha) = \{c_s^2\}$. These two choice functions should disagree on some sets, i.e. there must be $S \subset X$ such that $y = c_t^1(S)$ and $x = c_s^2(S)$. Without loss of generality assume $x \triangleright y$. By progressiveness, for any $k \leq t$, $c_k^1(S) \sqsubseteq y$ and for any $l \geq s$, $c_l^2(S) \sqsupseteq x$. Then $\pi_{\mu_2}(L_{\sqsupseteq}(y)|S) < \alpha \leq \pi_{\mu_1}(L_{\sqsupseteq}(y)|S)$ which is a contradiction because $\pi_{\mu_1} = \pi_{\mu_2}$ as both represent the original probabilistic choice described by π . \square

Proof of Theorem 2. Let π_μ and π_η be two $\text{PRC}_{\triangleright}$ with supports \mathcal{C} and \mathcal{C}' , respectively.

First we show the sufficiency. Let μ be higher than η ; and for contradiction assume that π_μ does not first order stochastically dominates π_η , i.e. there exists a set S and for some x

$$1 - \alpha := \pi_\mu(U_{\sqsupseteq}(x)|S) < \pi_\eta(U_{\sqsupseteq}(x)|S) := 1 - \beta$$

Note that α and β are the probability of choosing the strict lower contour set of x in S by using π_μ and π_η , respectively, hence $\alpha > \beta$.

Since \mathcal{C} and \mathcal{C}' are ordered choice collections satisfying progressiveness, there exists t and t' such that $\mu(c_1) + \dots + \mu(c_t) = \alpha$ and $x \triangleright c_t(S)$; $\eta(c'_1) + \dots + \eta(c'_{t'}) = \beta$ and $x \triangleright c'_{t'}(S)$. Let $c'_k = \eta_\alpha^{-1}$, then $k > t'$ since $\alpha > \beta$. Note that by the assumption of μ being higher than η , we must have $\mu_\alpha^{-1}(S) = c_t(S) \sqsupseteq c'_k(S) = \eta_\alpha^{-1}(S)$. Then we have $x \triangleright c_t(S) \sqsupseteq c'_k(S) \sqsupseteq x$. The last relation follows from the fact that t' is the highest index choice in \mathcal{C}' which chooses an element from the lower contour set of x and any choice with higher index chooses an element weakly better than x . This gives us the contradiction that needed for the proof.

Next we show the necessity. Let π_μ first order stochastically dominate π_η but μ not be higher than η . Then $\exists S \subset X$ and $\alpha \in (0, 1]$ such that $\eta_\alpha^{-1}(S) \triangleright \mu_\alpha^{-1}(S)$. Define x and y as $x = \eta_\alpha^{-1}(S)$ and $y = \mu_\alpha^{-1}(S)$, then $x \triangleright y$. Then we have

$$\pi_\eta(U_{\sqsupseteq}(x)|S) \geq \alpha > \pi_\mu(U_{\sqsupseteq}(x)|S)$$

which contradicts with the assumption that π_μ first order stochastically dominates π_η . \square

Proof of Theorem 3. (Necessity of U-Monotonicity): Let \triangleright be the reference order and (μ, \mathcal{C}) represent π such that \mathcal{C} satisfies less-is-more condition. Let $x \in T \subset S \subseteq X$ and $\pi(x|S) \neq 0$. First, we will show that for any $c_i \in \mathcal{C}$, $x \triangleright c_i(T) \Rightarrow x \triangleright c_i(S)$. Assume not, there exists i such that $c_i(S) \sqsupseteq x \triangleright c_i(T)$. If $c_i(S) \in T$ then the less-is-more property immediately yields a contradiction. Now consider $c_i(S) \notin T$. Then, since $\pi(x|S) \neq 0$, there must be an index $j \leq i$ such that $c_j(S) = x$. Then $c_j(S) = x \in T \subset S$. By the less-is-more property we have $c_j(T) \sqsupseteq c_j(S)$. Since $j \leq i$, by progressiveness, $c_i(T) \sqsupseteq c_j(T) \sqsupseteq c_j(S) = x$ which contradicts with $x \triangleright c_i(T)$. Therefore, we prove our claim that $x \triangleright c_i(T) \Rightarrow x \triangleright c_i(S)$. This implies the

following relation:

$$\sum_{x \triangleright_{c_i}(T)} \mu(c_i) \leq \sum_{x \triangleright_{c_i}(S)} \mu(c_i)$$

Hence,

$$\pi_\mu(U_{\geq}(x)|S) \leq \pi_\mu(U_{\geq}(x)|T)$$

(Sufficiency of U-Monotonicity): We assume that the probabilistic choice function π satisfies U-Monotonicity with respect to \triangleright and will show that the construction of \mathcal{C} given in the proof of Theorem 1 satisfies the less-is-more property with respect to this order.

Before we proceed, we note that U-Monotonicity can be expressed by strict lower counter sets. That is, for all $x \in T \subset S$,

$$\pi(U_{\geq}(x)|T) \geq \pi(U_{\geq}(x)|S) \iff \pi(L_{\triangleright}(x)|T) \leq \pi(L_{\triangleright}(x)|S)$$

We first show c_1 satisfies the less-is-more property. Let $c_1(S) \in T \subset S$. By construction, $\pi(c_1(S)|S) \neq 0$ and for all x such that $c_1(S) \triangleright x$ we have $\pi(x|S) = 0$. Hence, $\pi(L_{\triangleright}(c_1(S))|S) = 0$. By U-Monotonicity, $\pi(L_{\triangleright}(c_1(S))|T) = 0$. Hence, for all x such that $c_1(S) \triangleright x$, we have $\pi(x|T) = 0$. Since $\pi(c_1(T)|T) \neq 0$ by construction, we must have $c_1(T) \geq c_1(S)$.

Assume that all c_t satisfy the less-is-more property for all $t < i$. We now show that c_i also satisfies it. Let $c_i(S) \in T \subset S$. For contradiction, assume $c_i(S) \triangleright c_i(T)$. We consider two possible changes that may happen from Step $i-1$ to Step i .

Case 1) $c_{i-1}(S) = c_i(S)$. By the progressiveness property, $c_{i-1}(S) = c_i(S) \triangleright c_i(T) \geq c_{i-1}(T)$. Then transitivity implies $c_{i-1}(S) \triangleright c_{i-1}(T)$, which contradicts the fact that c_{i-1} satisfies the less-is-more property.

Case 2) $c_{i-1}(S) \neq c_i(S)$. Then the following relations hold according to the proof of Theorem 1;

$$\begin{aligned} k_i &\leq \pi(L_{\geq}(c_{i-1}(T))|T) \\ &\leq \pi(L_{\geq}(c_i(T))|T) \quad \text{since } c_i(T) \geq c_{i-1}(T) \\ &\leq \pi(L_{\triangleright}(c_i(S))|T) \quad \text{since } c_i(S) \triangleright c_i(T) \\ &\leq \pi(L_{\triangleright}(c_i(S))|S) \quad \text{since U-Monotonicity} \\ &= k_{i-1} \quad \text{since the choice on } S \text{ changed in step } i \end{aligned}$$

This observation contradicts with k_i being strictly increasing in i . Hence, we have $c_i(T) \geq c_i(S)$. This shows the less-is-more condition holds for c_i . \square

Proof of Theorem 4. For completeness of \triangleright_π^T , take two arbitrary alternatives x and y . If there exists z such that $\pi(y|\{x, y, z\}) > \pi(y|\{x, y\})$ or $\pi(x|\{x, y, z\}) > \pi(x|\{x, y\})$, then we must have either $x \triangleright_\pi y$ or $y \triangleright_\pi x$ by part (i) of definition of \triangleright_π^T .

Next assume that for any alternative z ,

$$(1) \quad \pi(y|\{x, y, z\}) \leq \pi(y|\{x, y\}) \text{ and } \pi(x|\{x, y, z\}) \leq \pi(x|\{x, y\})$$

Note also that the inequalities in Equation 1 are actually strict because π is strict. We assume $\pi(z|\{x, z\}) < \pi(z|\{y, z\})$. This assumption is without loss of generality since the same argument applies if the inequality is reversed.

We analyze four cases depending on whether $A_x := \pi(x|\{x, y\}) - \pi(x|\{x, z\})$ and $A_y := \pi(y|\{x, y\}) - \pi(y|\{y, z\})$ are positive or negative.

Case 1: $A_x < 0$ and $A_y < 0$.

By Equation (1) and $A_x, A_y < 0$, both x and y satisfy strict regularity in $\{x, y, z\}$. Then by Axiom 3, z must violate strict regularity. Since we assumed $\pi(z|\{x, z\}) < \pi(z|\{y, z\})$, together with Axiom 2, we must have $\pi(z|\{x, z\}) < \pi(z|\{x, y, z\}) < \pi(z|\{y, z\})$. By part (iii) of the construction of \triangleright_π , we get $y \triangleright_\pi x$. (Indeed, the entire order must be $y \triangleright_\pi x \triangleright_\pi z$ among these three alternatives.)

Case 2: $A_x > 0$ and $A_y > 0$.

Case 2(a): $\pi(x|\{x, z\}) < \pi(x|\{x, y, z\})$.

If $\pi(x|\{x, z\}) < \pi(x|\{x, y, z\})$ then $z \triangleright_\pi x$. Moreover, this case also implies that $\pi(z|\{x, z\}) > \pi(y|\{x, y, z\}) + \pi(z|\{x, y, z\})$. Hence, $\pi(z|\{y, z\}) > \pi(z|\{x, z\}) > \pi(z|\{x, y, z\})$ implying that z satisfies strict regularity. This gives us two possibilities: (i) $\pi(y|\{y, z\}) < \pi(y|\{x, y, z\})$, or (ii) $\pi(y|\{x, y, z\}) < \pi(y|\{y, z\})$.

From (i), Axiom 2 and strictness imply that $\pi(y|\{x, y, z\}) < \pi(y|\{x, y\})$. Given $\pi(x|\{x, z\}) < \pi(x|\{x, y, z\})$, part (ii) of the construction of \triangleright_π yields $y \triangleright_\pi x$. If (ii) holds, then $\pi(y|\{x, y, z\}) < \pi(y|\{y, z\}) < \pi(y|\{x, y\})$ (by $A_y > 0$) and $\pi(x|\{x, y, z\}) > \pi(x|\{x, z\})$ together implying that $y \triangleright_\pi x$ from part (ii) of the construction of \triangleright_π .

Case 2(b): $\pi(x|\{x, z\}) > \pi(x|\{x, y, z\})$.

Since $A_x > 0$, x satisfies the strict regularity in $\{x, y, z\}$. By Axiom 3, either y or z must violate regularity. If y violates it, then since $A_y > 0$ we have $\pi(y|\{y, z\}) < \pi(y|\{x, y, z\})$. Then by applying part (ii) of the construction of \triangleright_π , we get $x \triangleright_\pi y$.

Assume z violates regularity but not y . Since we assumed that $\pi(z|\{x, z\}) < \pi(z|\{y, z\})$, together with Axiom 2, we have $\pi(z|\{x, z\}) < \pi(z|\{x, y, z\}) < \pi(z|\{y, z\})$. Then, together with the assumption that y satisfies regularity and Axiom 2, we can apply part (iii) of the construction of \triangleright_π and conclude $y \triangleright_\pi x$.

Case 3: $A_x < 0 < A_y$. By Equation (1) and $A_x < 0$, x satisfies regularity in $\{x, y, z\}$. If $\pi(y|\{y, z\}) > \pi(y|\{x, y, z\})$, then y also satisfies regularity in $\{x, y, z\}$ by $A_y > 0$. Then by Axiom 3, z must violate the regularity. By applying Axiom 2 together with the assumption of $\pi(z|\{x, z\}) < \pi(z|\{y, z\})$, we get $\pi(z|\{x, z\}) < \pi(z|\{x, y, z\}) < \pi(z|\{y, z\})$. Then by applying part (iii) of the construction of \triangleright_π , we must have $y \triangleright_\pi x$.

If $\pi(y|\{y, z\}) < \pi(y|\{x, y, z\})$, then $z \triangleright_\pi y$. By part (iii) of definition of \triangleright_π , we have $x \triangleright_\pi z$. Then, we must have $x \triangleright_\pi^T y$.

Case 4: $A_x > 0 > A_y$. By Equation (1) and $A_y < 0$, y satisfies regularity in $\{x, y, z\}$. If $\pi(x|\{x, z\}) > \pi(x|\{x, y, z\})$, then x also satisfies regularity in $\{x, y, z\}$ by $A_x > 0$. Then by

Axiom 3, z must violate regularity. By applying Axiom 2 together with the assumption of $\pi(z|\{x, z\}) < \pi(z|\{y, z\})$, we get $\pi(z|\{x, z\}) < \pi(z|\{x, y, z\}) < \pi(z|\{y, z\})$. Then by applying part (iii) of the construction of \triangleright_π , we get $y \triangleright_\pi x$.

If $\pi(x|\{x, z\}) < \pi(x|\{x, y, z\})$, then $z \triangleright_\pi x$. By part (iii) of definition of \triangleright_π , we have $y \triangleright_\pi z$. Then, $y \triangleright_\pi^T x$.

□