PROGRESSIVE RANDOM CHOICE*

EMEL FILIZ-OZBAY† AND YUSUFCAN MASATLIOGLU§

Abstract. We introduce a Random Choice Model which randomizes over choice types that are possibly boundedly rational. This model can explain a rich set of stochastic choices. We further impose a progressive structure on the collection of choice types in order to capture heterogeneity due to varying level of a behavioral trait such as limited attention, willpower stock, shortlisting constraint, degree of loss aversion, or being pro-social. The progressiveness sorts the types based on how much the behavior of a type aligns with a reference ordering. Unlike the Random Utility Model, our Progressive model uniquely identifies the heterogeneity in the data, allowing policy makers to craft policies knowing the unique weights of each behavioral type and to perform an improved welfare analysis. As a showcase, we provide the characterizations of two useful types of bounded rationality: "less-is-more" and "no-simple-mistake". We provide a set of axioms for unique identification of the reference ordering for a progressive representation when it is not exogenously given.

Date: Sep, 2021.
* We thank David Ahn, Jose Apesteguia, Miguel Ballester, Paul Cheung, David Dillenberger, Keaton Ellis, Jay Lu, Paola Manzini, Marco Mariotti, Collin Raymond, and Erkut Ozbay for helpful comments and discussions.
† 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 empirical analysis of human behavior, the data often comes in stochastic form indicating the frequencies a certain choice is observed either in a population (interpersonal) or from an agent choosing in multiple occasions (intrapersonal). While a Random Utility Model (RUM) is mostly utilized for such analysis by embedding heterogeneous preference types in the model, it assumes each type to be rational and to act in accordance with a utility maximization. Since this assumption is often violated,[1] one may want the model to allow for boundedly rational types. In this paper we introduce a model that allows individual choice types to exhibit choice patterns outside of rational preference maximization framework. Our model has a collection of choice types, rather than preference types, endowed with a probability distribution. We call this the *Random Choice Model* (RCM).[2]

When the stochastic choice is generated by a collection of behavioral types, in many circumstances the types are naturally sorted according to how closely their behavior is aligned with an ordering induced by an ingrained characteristic. For example, in policy settings the tax policies are ordered by the revenue they generate (Roberts [1977]), public goods are ordered by the provision levels (Epple et al. [2001]), insurance offers are ordered by deductibles (Barseghyan et al. [2019]), policies are ordered by surplus they generated (Epple and Platt [1998]), and scheduled payments are ordered by the present value based on an interest rate (Manzini and Mariotti [2006a]); some well studied decision theoretic settings include lotteries ordered according to the second stochastic dominance, two-color Ellsberg urns ordered by the minimum number of balls in certain color (Chew et al. [2017]), options ordered according to their

[1]There is an abundance of evidence against utility maximization across several fields, including law, economics, psychology, and marketing, e.g. see Huber et al. [1982], Ratneshwar et al. [1987], Tversky and Simonson [1993], Kelman et al. [1996], Prelec et al. [1997], Echenique et al. [2011], Trueblood et al. [2013].

[2]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 stochastic choice by a random choice model. Hence, our works are complementary.

temptation levels (Shiv and Fedorikhin [1999]), and uncertain payments ordered by a maximum loss amount. Economists often perform comparative statics comparing the heterogeneous types in terms of the alignment with respect to the underlying ordering such as one type choosing more pro-social policy or avoiding loss more than the other type. The empirical analysis of type distributions also relies of 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.

We apprehend the idea of sorted types by focusing on a special case of RCM where the alternatives are sorted with respect to a reference ordering and the choice types are ordered in terms of how much the behavior of a type is aligned with this ordering. For example, consider a situation where the policies can be ranked by a social norm such as being environmental friendly. If the heterogeneity in the data is due to the agents' environmental cautiousness, then the choice of more "green" types will be more aligned with the reference ranking of the policies. In other words, the choice of a higher indexed type will be at least as green as the policy chosen by a lower indexed type. We call this structure *progressive* with respect to a reference ordering.[3] The random choice model where the collection of choice types is progressive is called *Progressive Random Choice* (PRC).

Our progressive structure is a generalization of a well-known concept called the single crossing property.[4] Indeed, the two concepts are the same if each choice type is generated by a preference/utility maximization. The single crossing property has been recently applied to RUM by Apesteguia et al. [2017].[5] In comparison to RUM, one

[3]The reference ordering in this example is the greenness of a policy but in other examples it may also correspond to the decision maker's preferences, a temptation ranking, or a measure of salience (see examples in the main text). While, initially, we take the reference ordering exogenous, later we make a revealed ordering analysis to endogenize it and provide conditions for unique identification of this ordering.

[4]The single crossing property plays important roles in economics: see Mirrlees [1971], Roberts [1977], Grandmont [1978], Rothstein [1990], Milgrom and Shannon [1994], Gans and Smart [1996].

[5]Apesteguia and Ballester [2020] builds on Apesteguia et al. [2017] within RUM, by applying the single crossing idea locally and only on alternatives that exist in the same choice problem.

advantage of having progressiveness structure in RCM framework is that we no longer need to commit to a particular functional form to order types.

Our first result identifies the type distribution given a stochastic choice data. We show with the minimal restriction on the distribution of choice types (progressiveness), one can uniquely identify the PRC representation –both the collection of choice functions and the probability weight assigned to each choice function in the collection. This result also shows the richness of the PRC representation because any stochastic choice has a PRC representation. We also provide a notion of comparative statics which allow us to rank any two PRCs within our framework.

Our framework enables us to study phenomena that are outside of the utility maximization paradigm. Since our PRC algorithm fully and uniquely identifies the choice types that generates the data, a researcher who wants to study a certain sorted behavioral trait can use it in identification and then analyze the implications of the behavioral property of interest on the stochastic choice data. For example, one can study the properties of data that is generated by behavioral types who choose more green policies on smaller set of policy options. We illustrate how to do this on two types of bounded rationality applications: (i) "no-simple-mistakes" (i.e., not choosing an option that has never been chosen in binary comparisons on triplets) and (ii) "less-is-more" (i.e., being more aligned with the ordering on smaller sets). For each of cases (i) and (ii), we provide the necessary and sufficient conditions on the stochastic choice in order to have a PRC representation with heterogeneity in the form of interest.

The empirical validity of our axioms 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 progressive structure exists in these observed collection types, and moreover the types in one of the data sets in that study satisfy the less-is-more property. One may check the progressive structure on other rich data sets when types are observed in contexts such as decision making under risk, time preferences, and portfolio allocation. With such fine data, we can also observe the types of bounded rationality of the choice types and question the implications of those types on the stochastic choice data.

Besides showing the usefulness of our PRC model, the two applications we have capture some well studied behavioral phenomena in decision science. No-simple-mistake addresses choice behavior where if x is chosen from $\{x, y, z\}$ then it must be chosen from either $\{x, y\}$ or $\{x, z\}$. Behavior violating this is called difficult choice in decision theory. We show that weakening both the standard regularity axiom and the centrality axiom of Apesteguia et al. [2017] are necessary and sufficient conditions for having a PRC with types making no-simple-mistake.

One of the most studied behavioral phenomena is so-called choice overload which may include situations where the complexity caused by abundance of alternatives may lead to choosing alternatives that are ranked lower by the reference ordering (see e.g. Iyengar and Lepper [2000], Chernev [2003], Iyengar et al. [2004], and Caplin et al. [2009]). Such choice overload is inevitable given the trend of abundance of options offered to consumers.[6] Chernev et al. [2015] argue that choice overload might have negative welfare consequences, hence, having less options can lead to an increase in consumer welfare: "less-is-more". One may choose unhealthy but tempting dishes on a large restaurant menu while she controls herself better on smaller menus. The 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]). Moreover, the less-is-more property is satisfied by the types observed in the choice data of Manzini and Mariotti [2006b] eliciting the time preferences of subjects on four payment schedules covering three periods.[7] We show that one simple axiom, *U-regularity*, is necessary and sufficient for having a PRC representation with types satisfying less-is-more property.

---

[6]For example, colgate.com offers fifty-three different kinds of toothpaste, and the typical supermarket in the United States carries 40,000 to 50,000 items nowadays, where as this number was 7,000 items as late as the 1990s.

[7]We considered only the choice types appeared at least 3% of the data set.

In the first part of the paper, the reference ordering defining the progressive structure is assumed to be exogenously given. In some applications, the reference ordering is conceivably observable to the researcher (for example 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 ordering (such as first order domination, riskiness of the options, or the time schedule of payment) is imposed. However, in some other contexts, especially when the ordering corresponds to the underlying preferences, it is crucial to derive it from the stochastic choice. This would improve the applicability of the model. We show to what extend the reference ordering can be revealed from the data. We also show that the axioms used in the two aforementioned applications are sufficient for unique identification of the reference ordering under a mild restriction.

Our theoretical contribution complements Apesteguia and Ballester [2020] and Dardanoni et al. [2020a] which utilize models of randomization over choice functions for empirical applications to identify heterogeneity in the data. Dardanoni et al. [2020a] prove usefulness of random choice model in 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 progressive structure only locally and allow for limited data. Our paper is also related to the recent literature which 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 utility can be separately identified from observed choices when there is rich exogenous variation in 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 choice where individuals might differ both in terms of their consideration capacities and preferences.

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 commits to one of two types of bounded rationality: *less-is-more* and *no-small-mistake*. This section provides a characterization theorem for each condition. Section 5 shows that, under less-is-more, the unique identification of the ordering is possible with mild assumptions. Hence, the reference is order endogenously inferred. Section 6 summarizes how our models relates to other well-known stochastic choice models. Section 7 concludes.

2. Model

Let X be a non-empty set of alternatives where $\mathcal{X}$ denotes all of its non-empty subsets. A stochastic choice function is a mapping $\pi : X \times \mathcal{X} \to [0,1]$ such that for any $S \subseteq X$, (i) $\pi(x|S) > 0$ only if $x \in S$; (ii) $\sum_{x\in S} \pi(x|S) = 1$. $\pi(x|S)$ is interpreted as the probability of choosing x from alternative set S. $\pi(T|S)$ is the sum of all choice probabilities in T, i.e. $\pi(T|S) = \sum_{x\in T} \pi(x|S)$. A choice function on X is a mapping $c : \mathcal{X} \to X$ such that $c(S) \in S$ for any $S \subseteq X$. $\mathbb{C}$ is the set of all choice functions on X.

We now introduce a Random Choice Model (RCM) where an individual stochastically engages with a choice function, c, from the collection of all choice functions, $\mathbb{C}$. Let μ be a probability distribution on $\mathbb{C}$. $\mu(c)$ represents the probability of c being realized as the choice function. μ constitutes a stochastic choice function π_μ such that

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

The stochastic choice 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 which 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 stochastic 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 generated by some utility maximization, then π_μ becomes the well-known Random Utility Model (RUM). Hence RUM is a special case of RCM. RCM enjoys a high explanatory power: any stochastic choice function can be represented within random choice framework. However, the representation is not unique in general as it is implied by the well-known non-uniqueness result of RUM. We will address this issue by studying RCM on more structured domains and show that this generates not only an interpretable relation between choice types but also deliver uniqueness.

We consider the set of alternatives that are naturally ordered by a linear order $\rhd$. Examples of such domains are in abundance as discussed in the introduction. In the problem of public good provision, $\rhd$ may order the alternatives based on the level of public good, in the problem of choosing a dish at a restaurant, $\rhd$ may sort the dishes based on the temptation levels.

We now propose a condition to relate the heterogeneity among choice types allowed by RCM with the order on the domain. We do this by sorting the heterogeneous types in terms of how much the behavior is in line with the reference ordering. In the public good provision problem, the level of public good picked by a more pro-social types will be above the level chosen by the less pro-social types. In the restaurant example, if an individual with less self-control resists a temptation, someone with more self-control will resist that too. Formally, we say a collection of distinct choice types $\mathcal{C} \subseteq \mathbb{C}$ is *progressive* with respect to $\rhd$ if $\mathcal{C}$ can be sorted $\{c_1, c_2, \ldots, c_T\}$ such that $c_t(S) \unrhd c_s(S)$ for all S and for any $t \geq s$.[8] 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, progressiveness requires type t to be more aligned with $\rhd$ than type s for $t \geq s$. The idea of progressive types reduces the heterogeneity of types in RCM into one dimension. That dimension of heterogeneity could be caused by the magnitudes of

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

willpower, levels of being pro-social, attention capacities, or loss aversion coefficients. Hence, the progressive structure will allow to study the heterogeneity within a given phenomena of interest. We now define progressive random choice formally.

Definition 1. π has a **progressive** random choice representation with respect to $\rhd$, ($\text{PRC}_{\rhd}$), if there exists μ on $\mathbb{C}$ such that $\pi = \pi_\mu$ and the support of μ is progressive with respect to $\rhd$.

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

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 and hence it is denoted by $\{u_t\}$.[9] Hence, for every type t there exists a utility function, u_t, such that for all S,

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

The well-known single crossing property of utility functions guarantees the progressive structure (see Lemma 1 in Appendix) which is studied by Apesteguia et al. [2017] in the stochastic choice context. A collection of $\{u_t\}$ satisfies *Single Crossing Property* with respect to $\rhd$ if for any $x \rhd y$ and $t > s$,

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

[9]We assume that u_t is strict so that the choice is unique.

Answering the same question is not straightforward when we step out the utility maximization paradigm. Next, we perform this task for two other classes of models covering different bounded rationality models. We then provide sufficient conditions for these classes to have the progressive structure.

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 and then the decision maker optimizes over this constraint. An RCM where 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 $(u, \{\Gamma_t\})$.[10] Formally, for every type t and for all S,

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

The models of Limited Attention by Masatlioglu et al. [2012], Lleras et al. [2017], Shortlisting 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 multi-individual setting, this can be thought as (i) the existence of a common attribute that ranks the alternatives such as the lowest price in shopping or the shortest distance in route choice, or (ii) having an objective ranking among the alternatives such as first-order stochastic dominance.

Next, we provide a sufficient condition on the class of two stage procedures for having a progressive structure in RCM. Let $L_u(T)$ denote the collection of alternatives

[10] We 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 ordering being linear order.

which have weakly lower utility than some element of T, i.e., $L_u(T) = \{x \in X | u(x) \leq u(y) \text{ for some } y \in T\}$. Our condition is a monotonicity requirement on how the constraints of each type evolve. Particularly, higher types' feasible set contains higher utility options than that of lower types.

Condition 1: For any types $t < s$, $\quad L_u(\Gamma_t(S)) \subseteq L_u(\Gamma_s(S))$.

Note that Condition 1 is weaker than requiring the higher types to consider more options. In other words, attention sets growing by types also satisfies this condition.

Remark 1. For a collection of choice functions defined by $(u, \{\Gamma_t\})$, Condition 1 implies progressive structure with respect to the order induced by u.[11]

The proof is easy to see and provided in 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 1, a higher type must consider some alternatives that weakly dominate the choice of the lower type. Therefore, the choice of 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 1 on well-known models 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 a preference ordering on the shortlist. If the first stage binary relation gets more incomplete, the shortlists get gradually richer as Condition 1 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

[11]One can easily show that Condition 1 is not only sufficient but also a necessary condition.

most salient option. If the information processing gets gradually more costly for higher types, then the shortlists get larger and Condition 1 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 equilibrium 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 expectation). As the individual becomes more loss averse, the set of personal equilibrium enlarges, hence, Condition 1 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 those she overcomes 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 1.

- *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 1 is satisfied since the attention 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 1.

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 collection of choices described by $(u, \{k_t\})$. Formally, for every type t and for all S,

$$\text{(Class 3)} \qquad c_t(S) = \operatorname*{argmax}_{x \in S} \; u(x) - k_t(x, S)$$

where k_t is the menu-dependent cost function.[12]

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. An individual with a fixed commitment utility but idiosyncratic cost of self control generating the stochastic choice data is in this class. Alternatively, a population of individuals who are choosing a public policy and while they all have the same selfish utility (for example minimizing cost), they differ how much they care about following a norm (such as being pro-social or minimizing inequality) also within Class 3.

Let $\rhd$ be the order defined by a behaviorally motivated norm such as a social norm or temptation. Here, $\rhd$ is *possibly* different than the common utility function. In order to get a progressive structure, we impose that the marginal costs of switching to an alternative ranked higher in the social norm is decreasing. In other words, it is easier for higher types to follow the norm. Formally,

Condition 2: Let $\rhd$ be an order. For any types $s > t$ and alternatives $x \rhd y$,

$$k_t(x, S) - k_t(y, S) \geq k_s(x, S) - k_s(y, S)$$

Remark 2. For a collection of choice functions defined by $(u, \{k_t\})$, Condition 2 with respect to $\rhd$ implies a progressive structure for the choice types with respect to $\rhd$.

The proof of Remark 2 is in 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

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

net utility. Then by Condition 2, it is not optimal for a higher type to switch to a lower ranked alternative according to $\triangleright$. Hence, the higher type will choose an alternative that is ranked weakly higher than y. This provides an ordered structure on the choice types as required by $\text{PRC}_{\triangleright}$.

The interpretation of Condition 2 on some well-known models is intuitive, as discussed below.

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

  $$c_t(S) = \underset{x \in S}{\text{argmax}} \quad \{u(x) - \alpha_t f(\max_{y \in S} v(y) - v(x))\}$$

  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 ordering and 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 2 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 it is influenced by the social norm. Formally,

  $$c_t(S) = \underset{x \in S}{\text{argmax}} \quad \{u(x) - (\max_{y \in S} \psi(y) - \psi(x))^{\beta_t}\}$$

  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. Condition 2 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 temptation level or social norm utility. Condition 2 requires the heterogeneity parameters to have 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, committing to a specific model, we demonstrated that the PRC structure allows for a substantial degree of heterogeneity of choice behavior. In each case, the alternatives are naturally ordered either by an attribute, common utility, temptation ranking, or 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.

PRC 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 $\rhd$. 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 stochastic choices for a given preference ordering. In other words, PRC enjoys high explanatory power.

Theorem 1. *Let $\rhd$ be a reference ordering. Every stochastic choice π has a PRC$_\rhd$ representation. Moreover, the representation is unique.*

The proof of Theorem 1 is constructive and hence provides an algorithm generating the heterogeneous types and their weights *uniquely* from a given stochastic choice data set. The construction is based on the choice probabilities of lower contour sets with respect to $\rhd$. We first calculate all cumulative probabilities on lower contour sets derived from the stochastic choice. Next we define an ordering function which sorts these cumulative probabilities from the lowest to the highest, $0 < k_1 < k_2 < \cdots < k_T$. 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 cumulative probability driven by the aforementioned ordering, 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, and the construction of the representation identifies the exact nature of heterogeneity and 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. Uniqueness 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, note that, for 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 ordering. The uniqueness comes from the fact that each collection can only have

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

a maximum of six elements due to the progressive structure.[14] Hence, the stochastic 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 grow extensively, the maximum number of choice functions in a progressive collection cannot surpass the information given by the stochastic choice data. On the contrary, in RUM model, the maximum number of preference orderings exceeds the number of choice data, which causes undesirable non-uniqueness of RUM.

3. Comparative Statics

Next we discuss how the comparative statics exercise can be performed to order any two PRCs. 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 those considered in Section 4. To do this, first we introduce an ordering relation between distributions of choices in Definition 2. Before defining the order, we define, for all $\alpha \in (0,1]$, $\mu_\alpha^{-1} := c_i \in \mathcal{C}$ such that $\mu(c_1) + .. + \mu(c_{i-1}) < \alpha \leq \mu(c_1) + .. + \mu(c_i)$ for given $\mathcal{C} = \{c_1, ..., c_T\}$ and μ. Hence, μ_α^{-1} identifies the choice function in the collection at which the cumulative distribution weakly exceeds α.

Definition 2. Probability distribution μ defined on $\mathcal{C}$ is **higher** than probability distribution η defined on $\mathcal{C}'$ if $\forall \alpha \in (0,1]$ and $\forall S \subset X$, $\mu_\alpha^{-1}(S) \gtrsim \eta_\alpha^{-1}(S)$.

Definition 2 compares two probability distributions and identifies the one which is more in line with the underlying ordering, $\rhd$, 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 stochastic dominates

[14]The 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.

the other distribution. Note that the comparison is based on $\triangleright$; hence, the compared models should have the same underlying $\triangleright$.

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

Definition 3. Stochastic choice π first order stochastic dominates stochastic choice π' if for any set S and any $x \in S$,

$$\pi(U(x)|S) \geq \pi'(U(x)|S)$$

Now we can state our result on comparative statics between any two PRCs.

Theorem 2. *Let π_μ and π_η be two $PRC_\triangleright$. π_μ first order stochastic dominates π_η if and only if μ is higher than η.*

Note that if the choices in the support of PRC are rational and represented by a collection of preferences, our model becomes equivalent to SCRUM (as stated by Lemma 1 in Appendix.) For such models Definition 2 is equivalent to Definition of "a SCRUM being higher" in Apesteguia 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 stochastic 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' $\text{PRC}_\triangleright$ 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 $\rhd$. 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 $\rhd$ (so her choice is degenerate, she is fully attentive and her choice satisfies WARP). Then the stochastic choice of the more attentive person (the second person) will first order stochastic dominate the stochastic choice of the less attentive one (the first person).

4. Applications of PRC for Certain Types of Bounded Rationality

The RCM allows us to address behavior that is possibly inconsistent with utility maximization because the choice types do not have to be in line with a preference maximization. Behavioral Economics literature provides abundance of 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 PRC structure. A natural question to ask is how one may 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 stochastic choice they would generate in the aggregate data. We show how to address this question on two examples in Subsection 4.1 and 4.2. In each subsection, we commit to one type of bounded rationality and find the corresponding necessary and sufficient conditions on stochastic choice in order to have that kind of bounded rationality. Thanks to Theorem 1, we already have a uniqueness result for these applications, so we will only focus on the characterization.

4.1. **Less is More.** One of the most studied deviations from the rational model is the choice overload phenomenon, i.e. the welfare improving effect of having less options (see Schwartz [2005], Iyengar and Lepper [2000], Chernev [2003], Iyengar et al. [2004], Caplin et al. [2009]). This phenomenon is called "less-is-more." While the classical rational choice theory concludes that the exuberance of choice has positive welfare implications, the idea of less-is-more is based on the evidence that the decision makers may not benefit from having too much choice in many situations. Due to their limited attention spans, cognitive capacities, or reference dependent evaluations, they may under-perform and deviate from their underlying preferences when they choose from very large set of options.

In this section we interpret the reference ordering of a PRC model as the common underlying preferences. Then each choice type can be viewed as a type of bounded rationality with a certain level of choice overload. Less options might be better for these types since they may choose sub-optimally on larger alternative sets and act more in line with the preferences on subsets. Then each choice type in the progressive collection of choice functions represents how severely that type of decision maker is affected from having too much choice.

As in the previous section, we first assume that the reference ordering, $\rhd$, is observable. This assumption is reasonable in situations where there is a single common attribute to rank all alternatives, such as the lowest price, shortest distance, the amount of carbon footprint, etc. We say that a collection of choice functions, $\mathcal{C}$, satisfies *less-is-more* with respect to $\rhd$ if for all t and for all $T \subset S$, $c_t(S) \in T \Rightarrow c_t(T) \unrhd c_t(S)$. In other words, $c_t(T)$ is more aligned with $\rhd$ 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 a preference ordering, then the less-is-more property trivially holds.[15] 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, where shortlists get gradually longer, imagine that the initial shortlist orders the alternatives based on a linear order that is completely opposite of $\rhd$, say $\tilde{\rhd}$. Such a shortlist would

[15]One should note that this observation makes SCRUM$_\rhd$ a special case of PRC$_\rhd$ with the less-is-more property.

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 ordering 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 will improve the prediction power of our model.

Next we state our only axiom for the characterization of boundedly rational types in the sense of less-is-more. This axiom is closely related to the well-known *regularity axiom*: For all $x \in T \subset S \subseteq X$, $\pi(x|S) \leq \pi(x|T)$. If the inequalities are strict, this is called strict regularity. This standard regularity axiom states that the choice frequency of an alternative is higher on smaller sets. Our first axiom requires the regularity condition to hold on upper counter sets. Let $U(x) = \{y \in X \mid y \trianglerighteq x\}$ denotes the upper contour set of an alternative, x.

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

$$\pi(U(x)|S) \leq \pi(U(x)|T)$$

Regularity and U-regularity coincide for the best alternative in any set. However, U-regularity allows for other regularity violations.

In deterministic case, the regularity condition is equivalent to WARP. But U-regularity is weaker than WARP. Note that if we have a deterministic choice which satisfies WARP than it can be represented by a preference relation. In such a case, U-regularity holds with respect to that preference relation. On the other hand, if a deterministic choice does not satisfy WARP, it may still satisfy U-regularity with respect to a linear ordering. 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-regularity with respect to $x \rhd y \rhd z$.

We now state our characterization result for the less-is-more type of bounded rationality.

Theorem 3. *Let $\rhd$ be a reference ordering. A stochastic choice π satisfies U-regularity with respect to $\rhd$ if and only if there exists a unique PRC$_{\rhd}$ representation of π where each choice satisfies less-is-more condition.*

Note that Theorem 3 not only provides a necessary and sufficient condition for 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-regularity.

4.2. **No Simple Mistakes.** One of the very basic rationality requirements is that if an alternative is never chosen in binary comparisons against some alternatives, then it should not be chosen from the set either. This requirement is quite demanding since the decision makers can make mistakes when they face with many options. Indeed, one may want to allow for such deviations from rationality when the choice environment is cognitively too demanding but require it for smaller set. The minimal of such requirements is to make this assumption for sets of three alternatives. In this subsection, we characterize the stochastic choice generated by types who may be boundedly rational but capable of avoiding mistakes on small choice sets. Formally, we say a collection of choice types, $\mathcal{C}$, satisfies *no-simple-mistake* if for all t, $c_t(\{x,y\}) \neq x$ and $c_t(\{x,z\}) \neq x \Rightarrow c_t(\{x,y,z\}) \neq x$. Rejecting an alternative over two different alternatives separately in binary comparisons indicates that this alternative is not desirable. Then choosing this alternative when all alternatives are simultaneously available is identified as a mistake. In this subsection, we eliminate such simple mistakes.[16]

[16]A more stronger version of this property could be that if an alternative is never chosen in binary comparisons, then it cannot be chosen in a grand set of any size. Note that such a stronger property

The definition of no-simple-mistake allows for bounded rationality without having extreme irrationality. Note that if the choice functions in the support of randomization are rationalizable by a preference ordering, then this property trivially holds. This new concept restricts each possible choice function to be either rational or boundedly rational in the sense of no-simple-mistake.

Next, we provide two axioms on a stochastic choice function, π, with respect to an ordering $\triangleright$.

Axiom 2. (Binary Weak Regularity) If $x \triangleright y \triangleright z$, then

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

and

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

This axiom is a weaker version of regularity written only for the best and the worst alternatives of a tripleton. By replacing the max operator with the min operator, one could get the classical regularity condition: $\pi(x|\{x,y,z\})$ is smaller than both $\pi(x|\{x,y\})$ and $\pi(x|\{x,z\})$. Our axiom is weakening that by only requiring it to be smaller than at least one of them.

Axiom 3. (Weak Centrality) Let $x \triangleright y \triangleright z$ and $\pi(y|\{x,y,z\}) > 0$. If $\pi(y|\{x,y\}) + \pi(y|\{y,z\}) < 1$, then $\pi(x|\{x,y\}) \leq \pi(x|\{x,y,z\})$ or $\pi(z|\{y,z\}) \leq \pi(z|\{x,y,z\})$.

This axiom states that in a tripleton, when the intermediate alternative is chosen with positive probability but relatively unattractive to the decision-maker in the sense that its total choice probability in binary comparisons is less than 1, then removing one of the extreme alternatives will weakly lower the other extreme alternative's choice probability.

could be in conflict with the idea of less-is-more where people could make mistakes when they face a large choice sets.

Theorem 4. *Let $\rhd$ be a reference ordering. A stochastic choice π satisfies Binary Weak Regularity and Weak Centrality with respect to $\rhd$ if and only if there exists a unique PRC$_{\rhd}$ representation of π where each type in the support makes no-simple-mistake.*

5. Endogenous Reference Order

Up to now, we have taken the reference ordering, $\rhd$, as given. This was reasonable in some applications where the true reference ordering (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 $\rhd$, a component of the model. Hence, they should be seen as a test that inputs *both* a stochastic choice and an order. For example, U-regularity tests whether the data has a PRC representation with less-is-more condition for a given reference order. The order might convey much (but not all) of the psychology that PRC captures. For example, say an outside observer believes that 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-regularity 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. The next remark makes this point for both bounded rationality conditions we studied earlier.

Remark 3. [Characterization] A stochastic choice π has an endogenous PRC representation satisfying less-is-more (no-simple-mistake) condition if and only if there exist preferences $\rhd$ such that $(\pi, \rhd)$ satisfies Axiom 1 (Axioms 2-3).

It is also possible that the outside observer does not have a prior knowledge about the reference ordering. We now describe to what extend one can identify the reference ordering of a given stochastic choice for a PRC representation. Note that this exercise only makes sense for a subclass of stochastic choice because otherwise from Theorem 1 every data has a PRC representation with respect to an arbitrary ordering. We perform this task for the bounded rationality restrictions described in the previous section.

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

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

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 ordering 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 y is ranked above x, which is denoted by xPy.

There are two other choice patterns revealing xPy. Specifically, if there exists a z revealing zPy (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 ordering $\rhd$ with $y \rhd x$ and $\text{PRC}_{\rhd}$ represents π with a support satisfying less-is-more. Since we must have zPy, we must have $z \rhd y \rhd x$. Since $\pi(y|\{x,y,z\}) \neq 0$, U-Regularity for y with respect to $\rhd$ 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., xPy.

The second choice pattern is more involved. We assume that 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 $\pi(x|\{x,y,z\}) \neq 0$. Assume that x is not ranked above y. The former inequality implies that yPz. Hence, there are two possible reference orderings: $y \rhd_1 z \rhd_1 x$ or $y \rhd_2 x \rhd_2 z$. Since $\pi(z|\{x,y,z\}) \neq 0$, U-Regularity for z with respect to $\rhd_1$ implies that $\pi(x|\{x,y,z\}) \geq \pi(x|\{x,y\})$. Similarly, since $\pi(x|\{x,y,z\}) \neq 0$, U-Regularity for x with respect to $\rhd_2$ 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, i.e., xPy.

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

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

If we have xPz and zPy revealed, we must have x to be ranked above y by transitivity even though xPy is not revealed. The transitive closure of P, denoted by P_T, includes these additional revelations as well as P itself. The next proposition summarizes this observation.

Proposition 1. *If π has a PRC$_\triangleright$ representation satisfying the less-is-more property, then P_T must be included in $\triangleright$.*

Next example illustrates that by having only the less-is-more structure in the PRC, there might be multiple reference orderings representing the same stochastic choice. Nevertheless, P_T is still in the intersection of those orders as stated by Proposition 1.

Example 1. The following table provides a set of of parametric stochastic choice described by π_λ where $\lambda \in [-0.10, 0.10]$.

When $\lambda > 0$, 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 yPz and xPz. However, we cannot conclude the ordering between x and y. Notice that we only employ the first part of definition of P since the rest of the conditions does not apply in this case. Therefore, 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

$\lambda > 0$.[17]

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 $-0.10 < \lambda < 0$, we have $\pi_\lambda(z|\{x,y,z\}) > \pi_\lambda(z|\{y,z\})$ revealing yPz, but $\pi_\lambda(z|\{x,y,z\}) < \pi_\lambda(z|\{x,z\})$. However, part (iii) of the definition of P implies that we must have $xPy(Pz)$. We have a unique revelation.[18]

When $\lambda = -0.10$, part (iii) of the definition of P no longer applies since $\pi_\lambda(x|\{x,y,z\}) = 0$. In that case, we only reveal yPz. Multiple orders containing yPz would give a PRC representation satisfying less-is-more property.

As the above Example 1 illustrates, the less-is-more condition may not be enough for unique identification of the reference ordering. Note that in all the cases without uniquely defined ordering in Example 1, we have either a violation of Axiom 2 (when $\lambda > 0$) or the stochastic choice has zero probability in it (when $\lambda = -0.10$). The next theorem shows that when we impose more restrictions, then the unique identification of the ordering is possible for the less-is-more type of bounded rationality. We call a stochastic choice strict if for all x, S, S', $p(x|S) \neq p(x|S') > 0$.[19]

Theorem 5. *If a strict stochastic choice π has PRC representation satisfying less-is-more with respect to $\rhd_1$ and $\rhd_2$ and Axiom 2 holds for both orders, $\rhd_1$ must be equivalent to $\rhd_2$, hence the reference order is unique.*

Proposition 1 is an identification result for the possible reference ordering. Theorem 5 is on the other hand a uniqueness result, stating that there is a unique completion of P_T acting as the reference order. One can use these results to provide a procedure

[17]It is straight forward to check these are the only reference orderings representing π even when $\lambda = 0$.
[18]This is surprising since there is only one regularity violation in this case. In other words, while there is less direct revelation compared to the case $\lambda > 0$, the revelation is unique.
[19]This assumption cannot be rejected by any finite data set. In addition, it is usually made for estimation purposes.

to test endogenously whether the stochastic choice satisfies Axioms 1-3 (i.e., it has a PRC with less-is-more and no-simple mistake properties with respect to the orderings.)

Given π, we can first derive P as described above. If there is a cycle, then this means that π cannot be represented by PRC satisfying less-is-more. If it does not have any cycle, then the implied P_T restricts the set of all reference orders which might be compatible with π. Then one may check Axioms 1-3 on this restricted set of orders. If there is an order satisfying axioms, then this order must be unique by Theorem 5. Hence there is no further need to check others. If none of them satisfies all the axioms, this implies that π cannot be represented by any PRC satisfying less-is-more and no-simple mistake conditions.

6. Related Literature

In this section, we compare our model with other well-known models of stochastic choice from the literature. First, note that in terms of explanatory power, PRC includes all the other models (see Theorem 1). Since PRC with less-is-more and/or no-simple-mistake condition impose testable restrictions, we now compare these special sub-classes to other stochastic choice models. As we mentioned before, SCRUM of Apesteguia et al. [2017] is a special case of this class. Moreover, the PRC satisfying less-is-more and no-simple mistake conditions includes other RUM choices other than SCRUM.

Manzini and Mariotti [2014], Brady and Rehbeck [2016], and Cattaneo et al. [2019] provide stochastic models where 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.[20] The random attention model (RAM) of Cattaneo et al. [2019] covers the model of Brady

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

and Rehbeck [2016] (BR), which in turn contains the model of Manzini and Mariotti [2014] (MM). All these models include an preference ordering as one of the components of their models. First, we state the differences of these models for a given reference ordering. In that case, RAM includes RUM, BR, SCRUM and MM. However, RAM, PRC satisfying less-is-more, and PRC satisfying no-simple-mistake are independent models because neither one is a subset of the other. When we consider endogenous reference ordering, RAM still includes RUM, BR, SCRUM and MM. In addition, it is still true that PRC satisfying less-is-more and no-simple-mistake is different from RAM. For example, consider the following stochastic 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 and no-simple-mistake conditions but not RAM. Indeed, this example is outside of any models discussed above. Moreover, it is routine to show PRC satisfying less-is-more is independent of PRC satisfying no-simple-mistake.

In the model of Gul, Natenzon, and Pesendorfer [2014] the decision maker first randomly picks an attribute using the Luce rule given the weights of all attributes. Then she picks an alternative using the Luce rule given the intensities of all alternatives in that attribute. Gul, Natenzon, and Pesendorfer [2014] show that any attribute rule is a random utility model. Hence, their model is distinct from the PRC model satisfying less-is-more and no-simple mistake conditions.

Echenique and Saito [2019] consider a general Luce model (GLM) where the decision maker uses the Luce rule to choose from among alternatives in her (deterministic) consideration set instead of the whole choice set.[21] GLM reduces to the Luce rule when all alternatives are chosen with positive probability in all menus. Hence, GLM is also distinct from our model in terms of observed choices.

Echenique, Saito, and Tserenjigmid [2018] propose a model (PALM) which uses violations of Luce's IIA to reveal perception priority of alternatives. For an example of stochastic choice data which can be explained by our model but not PALM, consider any data where the outside option is never chosen. When the outside option is never

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

chosen, PALM reduces to the Luce rule. However, PRC with less-is-more and no-simple mistake properties allows for violations of Luce's IIA in the absence of an outside option.

Fudenberg, Iijima, and Strzalecki [2015] consider a model of Additive Perturbed Utility (APU) where agents randomize, as making deterministic choices can be costly. In their model, choices satisfy regularity. Since Ł-PRC allows for violations of regularity, they are distinct models.

Aguiar, Boccardi, and Dean [2016] consider a satisficing model where the decision maker searches until she finds an alternative above a satisficing utility level. If there is no alternative above the satisficing utility level, the decision maker picks the best available alternative. They focus on two special cases of this model: (i) the Full Support Satisficing Model, where in any menu each alternative has a positive probability of being searched first, and (ii) the Fixed Distribution Satisficing Model. They show that the second model is a subset of RUM. On the other hand, the first model has no restrictions on observed choices if all alternatives are always chosen with positive probability. Hence, ours is distinct from these, too.

7. Conclusion

We have introduced a novel PRC model which not only uniquely identifies the random choices utilized but also has an intuitive progressive structure. As the examples we have provided throughout the article suggest, this model may prove useful in economic contexts where one wishes to investigate interpersonal or intrapersonal variation in choice and the variation is based on a sorted behavioral trait such as willpower, loss aversion, attention, or limited cognitive ability.

We see several directions in which the present work can be extended. We investigated two extensions in this paper imposing the less-is-more and no-simple mistake structures on the choices of a PRC model. These models have stronger prediction power and still apply to many examples covered by the literature. However, there are still some well-studied bounded rationality models such as the limited attention model with attention filter (Masatlioglu et al. [2012]) that can be explained in our framework.

Hence, one obvious avenue for exploration is to study other bounded rationality structures on the choices in the collection and their behavioral implications. Additionally, one might gradually impose more structure on PRC. In our Theorem 5, we perform this exercise for adding additional mild restrictions on less-is-more and get uniqueness of the endogenous reference ordering. Finally, several empirical queries arise from the present work. It would certainly be useful to conduct experimental tests comparing the explanatory power of competing models that we reviewed in this paper.

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., 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.

Chew, S. H., B. Miao, and S. Zhong (2017): "Partial ambiguity," *Econometrica*, 85, 1239–1260.

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. and G. J. Platt (1998): "Equilibrium and Local Redistribution in an Urban Economy when Households Differ in both Preferences and Incomes," *Journal of Urban Economics*, 43, 23–51.

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. Iijima, 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.

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.

Milgrom, P. and C. Shannon (1994): "Monotone comparative statics," *Econometrica: Journal of the Econometric Society*, 157–180.

Mirrlees, 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.

Schwartz, B. (2005): *The Paradox of Choice : Why More Is Less*, Harper Perennial.

Shiv, B. and A. Fedorikhin (1999): "Heart and mind in conflict: The interplay of affect and cognition in consumer decision making," *Journal of consumer Research*, 26, 278–292.

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

Single Crossing and Progressiveness. As mentioned in the main text, progressiveness generalizes the single-crossing idea recently studied by Apesteguia et al. [2017] within the RUM framework. A collection of preferences, $\{P_1, ..., P_T\}$, satisfies the single-crossing property with respect to $\rhd$ if for every $x \rhd y$ and every $s > t$, xP_ty implies xP_sy. As the next Lemma shows, if the choices in the support of a random choice model are rational and generated by maximization of preferences, then progressiveness is equivalent to the support of the corresponding RUM satisfying the single crossing property.

Lemma 1. *Let $\{c_1, ..., c_T\}$ be a collection of choices where each c_i is derived from maximization of a complete and transitive ordering P_i. Then $\{P_1, ..., P_T\}$ satisfies the single-crossing property with respect to $\rhd$ if and only if $\{c_1, ..., c_T\}$ satisfies the progressiveness property with respect to $\rhd$.*

Proof. Assume that $\{P_1, ..., P_T\}$ satisfies the single-crossing property with respect to $\rhd$. For contradiction, suppose the corresponding choice collection does not satisfy the progressiveness property. Then there exist $s, t \in \{1, ..., T\}$ such that $s > t$ and $S \subset X$ where $c_t(S) \rhd c_s(S)$. Since each P_i represents the corresponding c_i, we have $c_s(S)P_sc_t(S)$ and $c_t(S)P_tc_s(S)$. Note that since $c_t(S) \rhd c_s(S)$ by single crossing property we must have $c_t(S)P_tc_s(S) \Rightarrow c_t(S)P_sc_s(S)$, which is a contradiction.

For the other direction of the proof, assume that the collection of choices satisfies the progressiveness property with respect to $\rhd$. For contradiction, suppose the corresponding set of preferences does not satisfy the single-crossing property. Then there exists $x, y \in X$ such that $x \rhd y$, $s, t \in \{1, ..., T\}$ with $s > t$ and while xP_ty we have yP_sx. Then $c_t(\{x, y\}) = x$ and $c_s(\{x, y\}) = y$. By progressiveness, we should have $c_s(\{x, y\}) \unrhd c_t(\{x, y\})$ or equivalently, $y \unrhd x$. This is a contradiction. □

Proof of Remark 1. Assume Condition 1 on a collection of $(u, \{\Gamma_t\})$ within Class 2. Let $s > t$ and $S \subseteq X$ then by definition of this class, $c_t(S) \in \Gamma_t(S)$. Then observe that $c_t(S) \in L_u(\Gamma_t(S)) \subseteq L_u(\Gamma_s(S))$. This implies existence of $y \in \Gamma_s(S)$ s.t. $u(y) \geq u(c_t(S))$. Since $u(c_s(S)) \geq u(x)$ for any $x \in \Gamma_s(S)$, in particular we must have $u(c_s(S)) \geq u(y) \geq u(c_t(S))$. Hence, $\{c_t\}$ is a progressive collection with respect to u. □

Proof of Remark 2. Assume Condition 2 on a collection $(u, \{k_t\})$ within Class 3. For contradiction assume there are two types $s > t$ and $S \subseteq X$ such that $c_t(S) \rhd c_s(S)$. then by Condition 2, we have

$$k_t(c_t(S), S) - k_t(c_s(S), S) \geq k_s(c_t(S), S) - k_s(c_s(S), S)$$

Since $c_t(S)$ is chosen by type t, we have the followings:

$u(c_t(S)) - k_t(c_t(S), S) \geq u(c_s(S)) - k_t(c_s(S), S)$

$u(c_t(S)) - u(c_s(S)) \geq k_t(c_t(S), S) - k_t(c_s(S), S) \geq k_s(c_t(S), S) - k_s(c_s(S), S)$

Since we assumed that the choice of each type is unique, the inequality above is strict. Then type s would choose $c_t(S)$ as well, giving us a contradiction. Hence, the collection $\{c_t\}$ must be progressive with respect to $\rhd$. □

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

Define

$$K = \{\pi(\bar{L}(x) \cup x|S) \mid S \subseteq X \text{ and } x \in S\}^{22}$$

where $\bar{L}(x) = \{y \in X \mid x \rhd y\}$ (strict lower counter set). This defines a collection of all cumulative probabilities on lower contour sets derived from the stochastic 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., $0 < k_1 < k_2 < \cdots < k_m = 1$.[23] 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 $\min_{\pi^+}(\rhd, S)$, which selects the worst alternative in S according to $\rhd$ with strictly positive choice probability. That is,

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

For any set §, follow the steps below:

Step 1: Define

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

Note that $\mu(c_1)$ is positive and for any S, $\pi(c_1(S)|S) = \pi(\bar{L}(c_1(S)) \cup c_1(S)|S) \geq k_1$ 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 $\pi(\bar{L}(c_1(S)) \cup c_1(S)|S) = k_1$ since $k_1 \in K$.

Step 2: Define the second choice type as

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

This is well-defined because by the construction in the first step: $\pi(\bar{L}(c_1(S)) \cup c_1(S)|S) \geq k_1$. 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, $c_2(S) \trianglerighteq c_1(S)$ by definition of c_2 and hence, $\{c_1, c_2\}$

[22]We abuse the notation and write $A \cup x$ instead of $A \cup \{x\}$.

[23]k_m is always equal to 1 since $\pi(\bar{L}(x) \cup x|\{x\}) = 1$.

satisfies progressiveness with respect to $\triangleright$. Note that $\mu(c_1) + \mu(c_2) = k_2$. Moreover, there exists a subset S such that $\pi(\bar{L}(c_2(S)) \cup c_2(S)|S) = k_2$ since $k_2 \in K$.

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

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

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

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

$$\pi(\bar{L}(c_{i-1}(S)) \cup c_{i-1}(S)|S) = \sum_{\substack{y \trianglelefteq c_{i-1}(S) \\ y \in S}} \pi(y|S) \geq k_{i-1}$$

Note that by step $i-1$, $c_{i-1} \neq c_i$, and by construction $c_i(S) \trianglerighteq c_{i-1}(S) \trianglerighteq c_{i-2}(S) \trianglerighteq \ldots \trianglerighteq c_1(S)$ for all S. Hence, $\{c_1, c_2, \cdots, c_i\}$ 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 $\mathcal{C} = \{c_1, \ldots, c_m\}$ where each c_i is defined in Step i above. Since $\mathcal{C}$ satisfies progressiveness with respect to $\triangleright$, and $\sum_{t=1}^{m} \mu(c_t) = k_1 + \sum_{t=2}^{m}(k_t - k_{t-1}) = k_m = 1$, $(\mu, \mathcal{C})$ 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 $\pi(\bar{L}(x) \cup x|S) = k_i$ and $\pi(\bar{L}(x)|S) = k_j$. Since $\bar{L}(x) \subset \bar{L}(x) \cup x$ and $\pi(x|S) \neq 0$, k_i is strictly greater than k_j. Then by construction, we have $c_{j+1}(S) = \cdots = c_i(S) = x$. 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

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

Hence, π_μ and π are the same.

(Uniqueness): Let μ_1 with support $\mathcal{C}^1 = \{c_1^1, \ldots c_{n_1}^1\}$ and μ_2 with support $\mathcal{C}^2 = \{c_1^2, \ldots c_{n_2}^2\}$ 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 $x \rhd y$. By progressiveness, for any $k \leq t$, $c_k^1(S) \unlhd y$ and for any $l \geq s$, $c_l^2(S) \unrhd x$. Then $\pi_{\mu_2}(\bar{L}(y) \cup y|S) < \alpha \leq \pi_{\mu_1}(\bar{L}(y) \cup y|S)$ which is a contradiction because $\pi_{\mu_1} = \pi_{\mu_2}$ as both represent the original stochastic choice described by π. $\square$

Proof of Theorem 2. Let π_μ and π_η be two L-PRC$_\succ$ 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 $S = \{a_1, ..., a_n\}$ and for some $1 \leq i \leq n$

$$\pi_\mu(\{a_i, a_{i+1}, ..., a_n\}, S) < \pi_\eta(\{a_i, a_{i+1}, ..., a_n\}, S)$$

.

Define α and β as the probability of choosing the strict lower contour set of a_i in S by using π_μ and π_η, respectively, i.e. $\alpha = \pi_\mu(\bar{L}(a_i), S)$ and $\beta = \pi_\eta(\bar{L}(a_i), S)$; then $\alpha > \beta$.

Since $\mathcal{C}$ and $\mathcal{C}'$ are ordered choice collections satisfying progressiveness, there exists t and t' such that $\mu(c_1)+...+\mu(c_t) = \alpha$ and $c_t(S) \lhd a_i$; $\eta(c_1')+...+\eta(c_{t'}') = \beta$ and $a_i \rhd c_{t'}'(S)$. 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 $\mu_\alpha^{-1}(S) = c_t(S) \unrhd c_k'(S) = \eta_\alpha^{-1}(S)$. Then we have $a_i \rhd c_t(S) \unrhd c_k'(S) \unrhd a_i$. 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 a_i and any choice with higher index chooses an element weakly better than a_i. This gives us the contradiction that needed for the proof.

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

$$\pi_\mu(L(y) \cup \{y\}, S) \geq \alpha > \pi_\eta(L(y) \cup \{y\}, S)$$

Then we have

$$\pi_\mu(U(y), S) < \pi_\eta(U(y), S)$$

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

Proof of Theorem 3. (Necessity of U-Regularity): Let $\rhd$ be the reference ordering and $(\mu, \mathcal{C})$ 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 $c_i \in \mathcal{C}$, $x \rhd c_i(T) \Rightarrow x \rhd c_i(S)$. Assume not, there exists i such that $c_i(S) \trianglerighteq x \rhd c_i(T)$. 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 $c_j(T) \trianglerighteq c_j(S)$. Since $j \leq i$, by progressiveness, $c_i(T) \trianglerighteq c_j(T) \trianglerighteq c_j(S) = x$ which contradicts with $x \rhd c_i(T)$. Therefore, we prove our claim that $x \rhd c_i(T) \Rightarrow x \rhd c_i(S)$. This implies the following relation:

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

Hence,

$$\pi_\mu(U(x)|T) \geq \pi_\mu(U(x)|S)$$

(Sufficiency of U-Regularity): We assume that the stochastic choice function π satisfies U-regularity with respect to $\rhd$ 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 ordering.

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

$$\pi(U(x)|T) \geq \pi(U(x)|S) \Longleftrightarrow \pi(\bar{L}(x)|T) \leq \pi(\bar{L}(x)|S)$$

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 $c_1(S) \rhd x$ we have $\pi(x|S) = 0$. Hence, $\pi(\bar{L}(c_1(S))|S) = 0$. By U-regularity, $\pi(\bar{L}(c_1(S))|T) = 0$. Hence, for all x such that $c_1(S) \rhd x$, we have $\pi(x|T) = 0$. Since $\pi(c_1(T)|T) \neq 0$ by construction, we must have $c_1(T) \trianglerighteq c_1(S)$.

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 $c_i(S) \rhd c_i(T)$. 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, $c_{i-1}(S) = c_i(S) \rhd c_i(T) \trianglerighteq c_{i-1}(T)$. Then transitivity implies $c_{i-1}(S) \rhd c_{i-1}(T)$, 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;

$$
\begin{aligned}
k_i &\le \pi(\bar{L}(c_{i-1}(T)) \cup c_{i-1}(T)|T) \\
&\le \pi(\bar{L}(c_i(T)) \cup c_i(T)|T) \quad \text{since } c_i(T) \trianglerighteq c_{i-1}(T) \\
&\le \pi(\bar{L}(c_i(S))|T) \quad \text{since } c_i(S) \triangleright c_i(T) \\
&\le \pi(\bar{L}(c_i(S))|S) \quad \text{since U-regularity} \\
&= k_{i-1} \quad \text{since the choice on } S \text{ changed in step } i
\end{aligned}
$$

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

Proof of Theorem 4. (Necessity of Axioms 2-3): Let $\triangleright$ be the reference ordering and $(\mu, \mathcal{C})$ represent π such that $\mathcal{C}$ satisfies the no-simple-mistake condition. We show that π satisfies Axiom 2 and Axiom 3 with respect to $\triangleright$.

We first show that Axiom 2 is necessary. For a contradiction, assume that there are three distinct alternatives $x \triangleright y \triangleright z$ s.t. $\pi(x|\{x,y,z\}) > \max\{\pi(x|\{x,y\}), \pi(x|\{x,z\})\}$. Then, $\pi(x|\{x,y,z\}) > 0$ and hence, there exists c_i such that $c_i(x|\{x,y,z\}) = x$ and, for any $k < i$, $c_k(x|\{x,y,z\})$ is different from x. In other words, i is the smallest index type choosing x from $\{x,y,z\}$. Since $\pi(x|\{x,y,z\}) > \pi(x|\{x,y\})$, we must have $c_i(\{x,y\}) = y$. Similarly, $\pi(x|\{x,y,z\}) > \pi(x|\{x,z\})$ implies $c_i(\{x,z\}) = z$. These imply that c_i violates the no-simple-mistake condition. This establishes the first part of Axiom 2.

We now show the second part of Axiom 2. For a contradiction, we assume that $\pi(z|\{x,y,z\}) > \max\{\pi(z|\{x,z\}), \pi(z|\{y,z\})\}$. Then, $\pi(z|\{x,y,z\}) > 0$, and hence, there exists c_j such that $c_j(z|\{x,y,z\}) = z$ and, for any $k > j$, $c_k(z|\{x,y,z\})$ is different from z. In other words, j is the largest indexed type choosing z from $\{x,y,z\}$. Since $\pi(z|\{x,y,z\}) > \pi(z|\{x,z\})$, we must have $c_j(\{x,z\}) = x$. Similarly, $\pi(z|\{x,y,z\}) > \pi(z|\{y,z\})$ implies $c_j(\{y,z\}) = y$, which implies c_j violates the no-simple-mistake condition. Hence, Axiom 2 is satisfied for the best and worst alternatives.

We now show that Axiom 3 is necessary. Let $x \triangleright y \triangleright z$, $\pi(y|\{x,y,z\}) > 0$ and $\pi(y|\{x,y\}) + \pi(y|\{y,z\}) < 1$, but $\pi(x|\{x,y\}) > \pi(x|\{x,y,z\})$ and $\pi(z|\{y,z\}) > \pi(z|\{x,y,z\})$. Since $\pi(y|\{x,y,z\}) > 0$, there exist two types $i \le j$ such that $c_i(\{x,y,z\}) = c_j(\{x,y,z\}) = y$ and, for any $k < i$ or $j < k$, $c_k(\{x,y,z\})$ is different from y. In other words, i and j are the smallest and largest indexed types choosing y from $\{x,y,z\}$, respectively. $\pi(x|\{x,y\}) > \pi(x|\{x,y,z\})$ implies that $c_j(\{x,y\}) = x$. Similarly, $\pi(z|\{y,z\}) > \pi(z|\{x,y,z\})$ implies that $c_i(\{y,z\}) = z$. Then $\pi(y|\{x,y\}) + \pi(y|\{y,z\}) < 1$ guarantees that there exists t such that $c_t(\{x,y\}) = x$ and $c_t(\{y,z\}) = z$. If t is between i and j, then by the assumption $c_t(\{x,y,z\}) = y$, which yields a contradiction for the no-simple-mistake condition. If $t < i$, then the progressive structure implies $c_i(\{x,y\}) = x$ since $c_j(\{x,y\}) = x = c_t(\{x,y\})$, and $t < i \le j$. Then c_i violates the no-simple-mistake condition. Similarly, if $t > j$, c_j violates it too. This completes the proof of necessity.

(Sufficiency of Axioms 2-3): We assume the stochastic choice function π satisfies Axiom 2 and Axiom 3 with respect to $\rhd$ and will show that the construction of $\mathcal{C}$ given in the proof of Theorem 1 satisfies the no-simple-mistake condition.

If $|S| = 2$, the model has no restriction. Hence assume $|S| \geq 3$. For a contradiction, assume that there exists $c_i \in \mathcal{C}$ such that $c_i(\{x,y,z\}) = x$, $c_i(\{x,y\}) = y$, and $c_i(\{x,z\}) = z$ for three distinct alternatives.

If x is the worst alternative among $\{x,y,z\}$, then by the representation, we must have $\pi(x|\{x,y,z\}) \geq \mu(c_i) + \sum_{t=1}^{i-1} \mu(c_t) > \sum_{t=1}^{i-1} \mu(c_t) \geq \pi(x|T)$ for all $T \subset S$ since $\mu(c_i) > 0$ and μ is a PRC representation of π. This contradicts to Axiom 2.

If x is the best alternative among $\{x,y,z\}$, the representation implies that $\pi(x|\{x,y,z\}) \geq \mu(c_i) + \sum_{t=i+1}^{n} \mu(c_t) > \sum_{t=i+1}^{n} \mu(c_t) \geq \pi(x|T)$ for all $T \subset S$ since $\mu(c_i) > 0$ and μ is a PRC representation of π. Axiom 2 is violated.

If x is the middle alternative among $\{x,y,z\}$, w.l.o.g. assume $y \rhd x \rhd z$. Since $\mu(c_i) > 0$, we have $\pi(x|\{x,y,z\}) \neq 0$. By the PRC structure, we must have $\sum_{t=1}^{i} \mu(c_t) > \pi(x|\{x,y\})$ and $\sum_{t=i+1}^{n} \mu(c_t) \geq \pi(x|\{x,z\})$. Hence $1 > \pi(x|\{x,y\}) + \pi(x|\{x,z\})$. Axiom 3 implies $\pi(y|\{x,y\}) \leq \pi(y|\{x,y,z\})$ or $\pi(z|\{x,z\}) \leq \pi(z|\{x,y,z\})$. The former inequality does not hold since $\pi(y|\{x,y,z\}) < \sum_{t=i}^{n} \mu(c_t) \leq \pi(y|\{x,y\})$. On the other hand, since $c_i(\{x,z\}) = z$ and $c_i(\{x,y,z\}) = x$, we must have $\pi(z|\{x,z\}) > \pi(z|\{x,y,z\})$. This means the second inequality does not hold either, giving us a contradiction.

This completes the proof of sufficiency. □

Proof of Theorem 5. Before we proceed with the proof, we prove one lemma.

Lemma 2. *If xPy and xPz for the revealed preferences corresponding to the stochastic choice function assumed in Theorem 5, then either $\pi(y|\{x,y,z\}) > \pi(y|\{y,z\})$ or $\pi(z|\{x,y,z\}) > \pi(z|\{y,z\})$.*

Proof. Since π has a PRC representation and xPy and xPz are assumed, there are two possible reference orderings between y and z: $x \rhd_1 y \rhd_1 z$ or $x \rhd_2 z \rhd_2 y$. If $\text{PRC}_{\rhd_1}$ represents π then π satisfies U-Regularity with respect to $\rhd_1$ implying that $\pi(\{x,y\}|\{x,y,z\}) \leq \pi(y|\{y,z\})$. Then, $\pi(z|\{x,y,z\}) \geq \pi(z|\{y,z\})$. Similarly, if $\text{PRC}_{\rhd_2}$ represents π then π satisfies U-Regularity with respect to $\rhd_2$ implying $\pi(y|\{x,y,z\}) \geq \pi(y|\{y,z\})$. Since π is strict, we have the desired result. □

Next we will prove Theorem 5, i.e., the uniqueness of the reference ordering for PRC satisfying less-is-more. For a contradiction, assume that π has two such PRC representations: $\text{PRC}_{\rhd_1}$ and $\text{PRC}_{\rhd_2}$, both satisfying Axioms 1-2 with respect to the corresponding orderings, $\rhd_1 \neq \rhd_2$. Then there exist two alternatives x and y such that x is ranked above y in one representation ($x \rhd_1 y$) and y is ranked above x in another representation ($y \rhd_2 x$). This implies that we cannot have xPy or yPx because $P \subseteq \rhd_1 \cap \rhd_2$. Hence, for any alternative z, we must have

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

This is true because otherwise the construction of P in part (i) would imply xPy or yPx and that would give a contradiction. Note also that the inequalities in Equation 1 are actually strict because π is strict.

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. For the first two cases where $A_x, A_y > 0$ and $A_x, A_y < 0$, 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. For the last two cases where $A_x < 0 < A_y$ and $A_x > 0 > A_y$ are symmetric so the proofs of these cases follow by symmetric arguments, hence, we will only illustrate it for the case of $A_x < 0 < A_y$.

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 $\{x,y,z\}$. Since π satisfies U-regularity, there exists at least one regularity violations (the worst alternative among x,y,z in the corresponding ordering will violate regularity.) Since x or y do not violate it, z must be the worst alternative. 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 P, we get yPx. by Proposition 1, $P \subseteq \rhd_1$, giving us a contradiction.

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 zPx. 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), we must have zPy. Since we also have zPx, either $\pi(y|\{x,y,z\}) > \pi(y|\{x,y\})$ or $\pi(x|\{x,y,z\}) > \pi(x|\{x,y\})$ by Lemma 2 contradicting with Equation 1. 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 yPx

from part (ii) of the construction of P. By Proposition 1, $P \subseteq \rhd_1$, giving us a contradiction.

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

Since $A_x > 0$, x satisfies the strict regularity in $\{x,y,z\}$. y must also satisfies it because otherwise, if y violates it, then we have $\pi(y|\{y,z\}) < \pi(y|\{x,y,z\})$. Then by applying part (ii) of the construction of P, we get xPy. Then by Proposition 1, $P \subseteq \rhd_2$, giving us a contradiction. Then z must violate the regularity and be the worst alternative among x, y and z in any ordering used for PRC. 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 observation above that y satisfies strict regularity, we can apply part (iii) of the construction of P and conclude yPx. By Proposition 1, this implies $P \subseteq \rhd_1$, giving us a contradiction.

Case 3: $A_x < 0 < A_y$. By Equation (1) and $A_x < 0$, x satisfies the strict regularity in $\{x,y,z\}$. We consider two sub-cases: $\pi(z|\{x,z\}) < \pi(z|\{y,z\})$, and $\pi(z|\{x,z\}) > \pi(z|\{y,z\})$.

Case 3(a): Assume $\pi(z|\{x,z\}) < \pi(z|\{y,z\})$.

If $\pi(y|\{y,z\}) > \pi(y|\{x,y,z\})$, then y also satisfies the strict regularity in $\{x,y,z\}$ by $A_y > 0$. Since π satisfies U-regularity, z must violate the regularity and be the worst alternative in any ordering used for PRC. 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 P, we get yPx. By Proposition 1, $P \subseteq \rhd_1$, giving us a contradiction.

If $\pi(y|\{y,z\}) < \pi(y|\{x,y,z\})$, then zPy. Given that, we have two possibilities: (i) $\pi(z|\{x,z\}) < \pi(z|\{x,y,z\})$, or (ii) $\pi(z|\{x,y,z\}) < \pi(z|\{x,z\})$. From (i), we must have xPz. Then, transitivity implies xP_Ty. By Proposition 1, it must be $P_T \subseteq \rhd_2$, giving us a contradiction. If (ii) holds, then both x and z satisfy the strict regularity in $\{x,y,z\}$. Then y must violate it due to U-regularity and hence, y must be the worst alternative in every ordering representing the stochastic choice. This gives us a contradiction since $y \rhd_2 x$.

Case 3(b): Assume $\pi(z|\{x,z\}) > \pi(z|\{y,z\})$. Since x satisfies strict regularity then either y or z must violate it.

If z violates the strict regularity, then by the assumption of Case 3(b), we must have $\pi(z|\{x,y,z\}) > \pi(z|\{y,z\})$ or equivalently, $\pi(\{x,y\}|\{x,y,z\}) < \pi(y|\{y,z\})$. Then, $\pi(y|\{x,y,z\}) \leq \pi(\{x,y\}|\{x,y,z\}) < \pi(y|\{y,z\}) < \pi(y|\{x,y\})$. This means that y satisfies strict regularity. Then by U-regularity, z must be the worst alternative among x, y and z in any ordering used for a PRC representation. By applying Axiom 2, we get $\pi(z|\{x,y,z\}) < \pi(z|\{x,z\})$. Together with x satisfying strict regularity, part (iii) of the construction of P implies xPy. By Proposition 1, $P \subseteq \rhd_2$, giving us a contradiction.

If y violates strict regularity, then $\pi(y|\{x,y,z\}) > \pi(y|\{y,z\})$ since $A_y > 0$. This is equivalent to $\pi(\{x,z\}|\{x,y,z\}) < \pi(z|\{y,z\})$ implying that $\pi(z|\{x,y,z\}) \leq \pi(\{x,z\}|\{x,y,z\}) < \pi(z|\{y,z\}) < \pi(z|\{x,z\})$. This means, z satisfies strict regularity which makes y the worst alternative among x, y and z in any ordering used for a PRC representation due to U-regularity. This contradicts with the assumption that $y \rhd_2 x$.

Case 4: $A_x > 0 > A_y$. The proof of this case follows the same argument in the proof of Case 3, hence it is omitted. $\square$