PROGRESSIVE RANDOM CHOICE*

EMEL FILIZ-OZBAY† AND YUSUFCAN MASATLIOGLU§

Abstract. We introduce a new model of stochastic choice, the "Progressive Random Choice (PRC) model." In a PRC model, the decision maker randomizes over a collection of choice functions which are ordered with respect to some reference order. A PRC representation identifies the collection of choices and associated probabilities uniquely. Moreover, it can explain a rich set of stochastic choices. We are particularly interested in PRC where each choice function satisfies a well-known bounded rationality structure, namely "less-is-more." The characterization of less-is-more-PRC relies on two simple axioms: U-regularity and weak-regularity. We further show that the reference ordering can be endogenously derived in this class.

Date: November, 2020.

* We thank David Ahn, Jose Apesteguia, Miguel Ballester, Paul Cheng, 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

Sorted type spaces are ubiquitous in economics. In many circumstances, behavioral types are naturally sorted according to an ingrained characteristic (such as productivity, taste for public goods, political ideology, etc.). Sorted behavioral type structures have important applications, such as performing comparative statics or empirical analysis of type distributions. 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.

The random utility model (RUM) is often utilized to study stochastic data in understanding underlying heterogeneity in the population as it has heterogeneous preference types naturally embedded into the model. RUM requires that each type behaves in accordance with the utility maximization paradigm. However, this assumption could be violated especially when the behavioral types are due to loss aversion, preference for commitment, level of attentiveness, strength of willpower, or social preferences of decision makers. Since people systematically violate the requirements of utility maximization,[1] one may not want to force the model to allow only for rational types. Instead, 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 utility functions, endowed with a probability distribution. We call this model the *random choice* model.[2] Our framework can be interpreted as both the choices of a single individual in different situations (intrapersonal) and the choices of different individuals in the same environment (interpersonal).

[1]There is an abundance of evidence across several fields, including law, economics, psychology, and marketing that individual behavior is not in accord with this assumption, 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.

In order to apprehend the idea of sorted types, we focus on collections of choice functions that can be sorted. For this purpose, we use a reference ordering on the alternative set, and types that act more aligned with this ordering are indexed higher in the collection. We call this structure *progressive* with respect to a reference ordering.[3] One advantage of this approach is that we eliminate the need to commit to a particular functional form to order types. For example, consider a situation where the alternatives can be ranked by a social norm.[4] If the heterogeneity is due to how closely the decision makers follow the social norm, then the social norm corresponds to the reference ordering in our model and ordered types gradually become more motivated by the social norm.[5] In other words, if a low indexed type chose according to the norm then a higher indexed one should follow the norm too. The random choice model where the collection of choice types is progressive is called *Progressive Random Choice* (PRC).

Examples of PRC are abundant. For example, consider a decision maker who uses a collection of satisficing models in the sense of Simon [1955] and chooses the first alternative in a fixed list that provides a minimum level of satisfaction.[6] The choice functions that she randomly employs are represented by different satisficing threshold levels. Then as the minimum level of satisfaction increases, the choices become more inline with the underlying preferences. Another example of progressiveness can be generated within the limited attention framework (see Lleras et al. [2017] and Masatlioglu et al. [2012]). If the decision maker has limited attention in each choice used in a PRC, then progressiveness ordering is interpreted as the decision maker becoming gradually more attentive. A third example could be built within a social norm model (see Dillenberger and Sadowski [2012]) where each choice function corresponds to a different type of agent who is affected by the norm at a different level characterized by their

[3]In the first part of the paper, the reference ordering is taken exogenously. We later endogenize it in Section 4.
[4]This example is formalized in Example 1.
[5]The reference ordering may also correspond to the decision maker's preferences, temptation ranking, or measure of salience (see examples in the main text).
[6]When there is no alternative above the threshold, the decision maker searches the entire set and chooses the best one.

shame parameters. Ordering the types according to the shame parameters imposes a progressive structure to the choice functions collection. The first two examples are capturing intrapersonal randomization of a decision maker and the reference ordering corresponds to her preferences. On the other hand, the third example captures an interpersonal heterogeneity and the reference ordering is the social norm.

The heterogeneous choice types may not be observable in general and they need to be derived from the stochastic choice data. However, in some exceptions such as controlled experiments, the researcher can observe the choice functions of each subject and study the choice types. Manzini and Mariotti [2006] provides a unique opportunity to test the progressive structure. They collected two data sets generated by 102 subjects and document choice types generating the stochastic choices. In the first data set, only three choice functions are used most frequently by the subjects (any other choice behavior is performed by only one or two subjects.) In the second data set, four-choice functions are used frequently. Both of these collections have the progressive structure. Moreover, half of the subjects do not exhibit utility maximizing behavior, which implies that RUM is not a right model for those individuals.

Our progressive structure is a generalization of a well-known concept called the single crossing property. Indeed, it is equivalent to the single crossing property if each choice function in the collection is generated by a preference/utility maximization. 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]. It has been recently applied to RUM by Apesteguia et al. [2017]. Again, in the RUM framework, Apesteguia and Ballester [2020] builds on Apesteguia et al. [2017] by applying the single crossing idea locally and only on alternatives that exist in the same choice problem.

Our first result is about identification of type distribution. We ask under which conditions it is possible to recover the type distribution from stochastic 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.

The main purpose of our model is to model heterogeneity in the choice data through sorted types. Since the PRC model fully serves this purpose and uniquely identifies the types, it can be used to model any particular class of choice types or behavioral traits of interest. Specifically, our framework enables us to study phenomena that are outside of the utility maximization paradigm. To illustrate this, we focus on choice types that are likely to make mistakes on larger choice sets. We then provide the complete set of behavioral implications of these types and identify these types from the observable stochastic choice data.

One of the most studied behavioral phenomena is so-called choice overload which is defined as decision makers being worse off by the complexity of choice problems. Here, we consider a special type of choice overload where complexity is measured by the number of alternatives (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.[7]

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". This concept of bounded rationality captures situations where a decision maker is more likely to choose suboptimally on larger alternative sets than smaller ones. Note that the idea of less-is-more potentially allows extreme mistakes: a decision maker may pick the worst alternative on the larger sets even though she chooses optimally in binary comparisons. Our less-is-more property rules out such extreme mistakes. Despite this additional restriction, our property is still rich enough to accommodate models such as shortlisting (Manzini and Mariotti [2007]), rationalization (Cherepanov et al.

[7]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.

[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 [2006] eliciting the time preferences of subjects on four payment schedules covering three periods.[8]

We focus on a subclass of PRC models satisfying the less-is-more property: L-PRC. Our first result guarantees identification for L-PRC. Next we turn to testability: does L-PRC generate testable restrictions on stochastic data? The answer is that two simple axioms, *U-regularity* and *Weak-regularity*, summarize all the testable implications of L-PRC. U-regularity requires the stochastic choice to assign higher probability to the upper contour set of an alternative on a smaller choice set than a larger one. Weak-regularity on the other hand is a weaker version of the standard regularity condition. While regularity requires the choice probabilities to be monotonic for *any* subset, the Weak-regularity demands such monotonicity to hold at least for *some* subset.

In the first part of the paper, the reference ordering defining the progressive structure is assumed to be 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 that one can uniquely identify the reference ordering of an L-PRC model under a mild restriction. We also show that endogenous L-PRC offers distinct predictions from several well-known stochastic choice models.[9]

[8]We considered only the choice types appeared at least 3% of the data set.
[9]The models provided by Manzini and Mariotti [2014], Brady and Rehbeck [2016], and Cattaneo et al. [2019] are distinct from L-PRC. We also show that while SCRUM (Apesteguia et al. [2017]) is a special case of L-PRC, the intersection of RUM and L-PRC is a strict superset of SCRUM.

The L-PRC model and its characterization illustrate the link between a behavioral trait (namely, less-is-more) and the stochastic choice data generated by types with different levels of that behavioral motive. Our algorithm constructs the collection of choice types in a PRC uniquely, hence, a researcher who wants to study a certain sorted behavioral types can use our algorithm to identify the types and then see the implications of the behavioral property of interest on the stochastic choice. We present how to do this for a particular type of behavioral trait capturing choice overload but a similar idea can be carried out for other type of behavioral traits under consideration.

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. It also presents the first representation result for the PRC model. Section 3 provides the necessary and sufficient conditions for existence of a PRC model with the *less-is-more* structure. Section 4 states our identification result for deriving the unique underlying reference ordering of an L-PRC model. Section 5 provides comparative statics between any two models within our framework. Section 6 summarizes how L-PRC relates to other well-known stochastic choice models. Section 7 concludes.

2. Model

Let X denote a finite set of alternatives and $\succ$ be a linear order on X where $\succsim$ is the weak order derived from $\succ$ in the usual sense.[10] $\succ$-best denotes the best alternative of a set with respect to relation $\succsim$. A stochastic choice function is a mapping $\pi : X \times 2^X \setminus \emptyset \to [0,1]$ 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 $c : 2^X \setminus \emptyset \to X$ such that $c(S) \in S$ for any $S \subseteq X$. $\mathbb{C}$ is the set of all choice functions on X.

In a random choice model, 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. Given a set of available alternatives S, the probability of an alternative x being chosen is determined by the sum of probabilities of choice functions which select x. Therefore, μ constitutes a stochastic choice function π_μ such that

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

We say that a stochastic choice function π has a random choice representation if there exists μ such that $\pi = \pi_\mu$. Let the support of μ be denoted by $\{c \in \mathbb{C} \mid \mu(c) > 0\}$. If the support of μ consists of only distinct choice functions generated by some linear order, then π_μ becomes the well-known RUM. Hence RUM is a special case of the random choice model. Any stochastic choice function can be represented within random

[10] While we assume an exogenous reference ordering, $\succ$, here, in Section 4 we will provide conditions for deriving $\succ$ endogenously.

choice framework. That is, for every π, there exists μ such that $\pi = \pi_\mu$. However, the representation is not unique in general.[11]

We impose a structure on the support of the probability distribution. Our condition, called progressiveness, is inspired by single-crossing preferences, which play important roles in economics (see e.g. Mirrlees [1971], Roberts [1977], Grandmont [1978], Rothstein [1990], Milgrom and Shannon [1994], Gans and Smart [1996].) Recently, this property was used by Apesteguia et al. [2017] in the context of a random utility model. We say a collection of distinct choice functions $\mathcal{C} \subseteq \mathbb{C}$ is *progressive* with respect to $\succ$ if $\mathcal{C}$ can be sorted $\{c_1, c_2, \ldots, c_T\}$ such that $c_t(S) \succsim c_s(S)$ for all S and for any $t \geq s$.[12] Progressiveness imposes an ordered structure on the collection of choices such that a higher indexed choice function cannot choose an alternative that is dominated by the choice of a lower indexed choice function on the same choice set. In other words, c_t is more aligned with $\succ$ than c_s when $t > s$. Note that any subset of a progressive set is also progressive. We now define progressive random choice formally.

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

We view our novel progressive structure as a strength of the model because it provides a meaningful interpretation for the support of random choice. Recall that the support of RUM consists of several independent utilities and there is no immediate comparison between them. In contrast, PRC orders the choice functions with respect to a reference choice function implied by $\succ$ and these functions gradually become more aligned with the reference. The use of each choice function can be interpreted as lapses into bounded rationality and the more distant the used choice function is from the

[11] Consider a stochastic choice function generated by $\{c_1, c_2\}$ with $\mu(c_1) = \mu(c_2) = 0.5$ where $c_1(x,y,z) = x, c_1(x,y) = x, c_1(x,z) = x, c_1(y,z) = y$, and $c_2(x,y,z) = y, c_2(x,y) = y, c_2(x,z) = z$, and $c_2(y,z) = y$. The same stochastic choice can be generated by $\{c'_1, c'_2\}$ with $\mu'(c'_1) = \mu'(c'_2) = 0.5$ where $c'_1(x,y,z) = y, c'_1(x,y) = x, c'_1(x,z) = z, c'_1(y,z) = y$ and $c'_2(x,y,z) = x, c'_2(x,y) = y, c'_2(x,z) = x$, and $c'_2(y,z) = y$ as well.

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

reference, the more severe the lapse is. The examples below illustrate on archetypical models how progressiveness can be interpreted as the random choices becoming less and less boundedly rational in the paradigm of that specific model. They also demonstrate the fact that PRC allows a substantial degree of heterogeneity of choice behavior. In each example, the choice types are sorted according to one behavioral trait. As our examples will clarify, the reference ordering may correspond to the decision maker's preferences as well as her social norm, measure of salience, and temptation ranking, see e.g. Examples 1, 2, and 4.

Example 1. (Ashamed to be Selfish) *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 (Dillenberger and Sadowski [2012]). Assume each type differs only in terms of how much it is influenced by the social norm. Each type cares more and more about the shame component, and hence, the collection of choices is progressive with respect to the social norm. Formally, we assume that the choice can be written as*

$$c_s(T) = \operatorname*{argmax}_{x \in T} \quad \{u(x) - (\max_{y \in T} \psi(y) - \psi(x))^s\}$$

where u is a utility function over allocations, ψ represents the norm, and s is the shame parameter. The amount, $(\max_{y\in A} \psi(y) - \psi(x))^s$, is interpreted as the shame from choosing x in comparison to the alternative that maximizes the norm. To provide a numerical example, we assume there are three possible allocations: x, y and z. While utilities are $u(x) = 4, u(y) = 3$ and $u(z) = 1$, the norm values are $\psi(x) = 1, \psi(y) = 4$ and $\psi(z) = 6$. We consider five different shame parameter values $s = 0, 0.3, 0.6, 0.9,$ and 1.2. Table 1 illustrates the corresponding collection of ordered types on $\{x, y, z\}$.

This collection is progressive with respect to $z \succ y \succ x$, which also represents the social norm. The ordered types are becoming gradually more inline with the social norm. While c_5 is completely aligned with the social norm and represented by the reference ordering, c_2 cannot be represented by any preferences.

	c_1	c_2	c_3	c_4	c_5
$\{x,y,z\}$	x	x	y	y	z
$\{x,y\}$	x	y	y	y	y
$\{x,z\}$	x	x	x	z	z
$\{y,z\}$	y	y	y	y	z
s	0	0.3	0.6	0.9	1.2

TABLE 1. Five choice function according to shame parameter.

Example 2. (Shortlisting) *Let each choice type in the PRC be determined by first eliminating dominated alternatives with respect to a binary relation to form a shortlist and then by maximizing a preference ordering on the shortlist (Manzini and Mariotti [2007]). Assume each type shares the same preferences. If the first stage binary relation gets more incomplete, the shortlist gets gradually richer and the choice becomes more aligned with the underlying order of alternatives. Hence, this collection of choices is progressive with respect to the preference ordering.*

The model of Tyson [2013] is another shortlisting model example with a different interpretation. The first criterion is the decision maker's preferences that are imperfectly perceived due to cognitive or information-processing constraints. The second criterion is interpreted as a measure of salience—the property of standing out from the rest. If the information processing gets gradually more costly, then the shortlists get larger and hence the choice becomes more aligned with the salience order rather than a preference order.

Example 3. (Preferred Personal Equilibrium) *Let each choice type in the PRC be determined by the preferred personal equilibrium concept introduced by Kőszegi and Rabin [2006]. According to this model, each type is endowed with a consumption utility as well as a gain-loss utility. As the individual becomes more loss averse, the set of personal equilibrium enlarges. Hence, the preferred personal equilibrium becomes more aligned with the consumption utility. This collection of choices is progressive with respect to the consumption utility.*

Example 4. (Temptation) *Consider a decision maker facing temptation with limited willpower (Masatlioglu et al. [2020]). The decision maker picks the alternative that maximizes her commitment utility from the set of alternatives where she overcomes temptation with her willpower. Each choice type in the PRC differs in terms of willpower stock. As her willpower stock increases, she is able to overcome temptation more successfully and able to choose an alternative more aligned with her commitment preferences. Hence, this collection of choices also satisfies progressiveness with respect to the commitment utility.*[13]

Example 5. (Limited Attention) *In this example, suppose that each choice type in the PRC pays attention to a consideration set and chooses the most preferred alternative in the consideration set with respect to the decision maker's underlying preferences (Masatlioglu et al. [2012], Lleras et al. [2017]). If each type's awareness extends gradually, then her choices become closer to the rational choice implied by her preferences. For this example, the progressiveness structure is equivalent to increased attentiveness.*

Example 6. (Rationalization) *Consider a decision maker providing a line of reasoning (a rationale) to justify her choice behavior (Cherepanov et al. [2013]) for each choice type in the PRC. A rationale can be intuitively understood as a story that states that some options are better than others. The decision maker maximizes her preferences among alternatives she can rationalize. Each choice type differs only in terms of the set of rationales she uses for that choice. As the set of rationales gradually gets larger, the corresponding collection of choices satisfies progressiveness with respect to her preferences.*

As mentioned above, 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 $\succ$ if for every $x \succ y$ 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

[13]A similar example of temptation can be created based on the convex self-control model of Noor and Takeoka [2010].

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 $\succ$ if and only if $\{c_1, ..., c_T\}$ satisfies the progressiveness property with respect to $\succ$.*

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 rational choice induced by $\succ$. It also allows a substantial degree of heterogeneity of choice behavior as we will see in Theorem 1. We will also show that the PRC structure leads to uniquely defined weights on the collection of choices.

We now state our first result. Theorem 1 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 $\succ$ be a reference ordering. Every stochastic choice π has a PRC$_\succ$ representation. Moreover, the representation is unique.*

The proof of Theorem 1 is constructive. The construction is based on the choice probabilities of lower contour sets with respect to $\succ$. We 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 $\succ$.[14] 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 is strictly larger than k_1. For the former case,

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

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 better alternatives on any given set.

Note that Theorem 1 is also a uniqueness result, and the construction of the representation provides the exact weights for each choice function in the support.[15] This is in sharp contrast to both the general random choice model and RUM, which are well-known to admit multiple representations (see Fishburn [1998] for the RUM and footnote 11 for the random choice model). Hence, one obvious strength of the PRC is its uniqueness.

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. Hence, the progressive structure by itself does not identify the right collection of choices for the representation immediately. However, the uniqueness comes from the fact that each collection can only have a maximum of six elements due to the progressive structure. 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.

[15]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.

3. Less is More

The random choice model allows us to address behavior that is inconsistent with utility maximization because the choice functions in the randomization support do not have to be in line with a preference maximization. Behavioral Economics literature provides abundance of evidence outside of the utility maximization framework. One of the most studied deviations from the rational model is the choice overload phenomenon, i.e. 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 represent 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 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 etc., or where alternatives are objectively ranked independent of personal (idiosyncratic) preferences such as lotteries being ordered according to first-order stochastic dominance. Having said that, in the next section, we will drop this assumption and show that one can identify the underlying preferences from the observed choices.

Definition 2. We say a collection of choice functions $\mathcal{C}$ satisfies *less-is-more* with respect to $\succ$ if for all t,
(i) for all $T \subset S$, $c_t(S) \in T \Rightarrow c_t(T) \succsim c_t(S)$,
(ii) for all $x \in S$ and for all $T \subset S$ such that $x \in T$ and $|T| \neq 1$ $c_t(T) \neq x \Rightarrow c_t(S) \neq x$.

The first condition means that $c_t(T)$ is more aligned with $\succ$ than $c_t(S)$ when $T \subset S$, because the choice from a larger set is dominated by the choice from a smaller set. The second condition says if an alternative is never chosen in any subset, then it cannot be chosen in the grand set. While the first condition requires the better alternatives to be chosen on smaller sets, it does not rule out the worst alternative to be chosen only in the larger set. The second condition restricts such extreme mistakes. This definition allows bounded rationality without having extreme irrationality. 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. This new concept restricts each possible choice function to be either rational or boundedly rational in the sense of less-is-more.[16]

Definition 3. π has a **less-is-more** PRC representation with respect to $\succ$, (L-PRC$_\succ$), if there exists μ such that the support of μ satisfies progressiveness and less-is-more with respect to $\succ$ and $\pi = \pi_\mu$.

All the models discussed in Examples 1 - 6 can be modified to accommodate the less-is-more structure. For the shortlisting example, Example 2, where shortlists get gradually longer, imagine that the initial shortlist orders the alternatives based on a linear order that is completely opposite of $\succ$, say $\tilde{\succ}$. Such a shortlist would report only the worst alternatives as undominated. Clearly, the choice implied by this shortlist would satisfy "less-is-more" since on a smaller set only a weakly better alternative can be shortlisted and chosen than on a larger set. When the shortlists in that example get gradually longer, due to reverse ordering implied by $\tilde{\succ}$, each choice satisfies less-is-more.

[16]One should note that this observation makes SCRUM$_\succ$ a special case of L-PRC$_\succ$.

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 in Example 5 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 axioms for the stochastic choice that will characterize L-PRC$_{\succ}$. These axioms are closely related to the well-known regularity axiom: For all $x \in T \subset S \subseteq X$,

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

The 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 $U(x) = \{y \in X \mid x \succsim y\}$ 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 regularity violations. Specifically, if we apply U-regularity to the second worst alternative, we can show that the worst alternative violates the regularity condition weakly. To show this, let y_S and z_S be the second-worst and the worst alternatives of $S \subseteq X$, respectively. Then, for T such that $y_S, z_S \in T \subset S$ and $\pi(y_S|S) \neq 0$, U-regularity implies that,

$$\begin{aligned} \pi(U(y_S)|S) &\leq \pi(U(y_S)|T) \\ 1 - \pi(z_S|S) &\leq 1 - \pi(z_S|T) \\ \pi(z_S|S) &\geq \pi(z_S|T) \end{aligned}$$

Hence, the regularity might be violated for z_S. By this argument, if the stochastic choice is a strict RUM,[17] then it violates U-regularity.

[17]We say π is a strict RUM if $\pi(x|S) < \pi(x|T)$ for all $x \in T \subset S \subseteq X$.

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 $\pi(z|xyz) = 1, \pi(x|xy) = 1, \pi(y|yz) = 1$, and $\pi(x|xz) = 1$. This choice behavior does not satisfy WARP but it satisfies U-regularity with respect to $x \succ y \succ z$.

The axiom requires the better elements to be more likely to be chosen on smaller sets. Therefore, we are interested in behavior such that the decision maker makes more mistakes and chooses a dominated alternative with respect to the underlying linear order more often on larger sets than smaller ones. Having said that, one might want to limit the amount of mistakes. In the above example, the worst alternative is chosen in the grand set with probability one. While this individual makes zero mistakes in all smaller sets, she makes the mistake with extreme probability in the larger set. The next axiom limits such examples. Hence, the deviations from RUM is minimal.

Axiom 2. (Weak-regularity) For all $S \subset X$ with $|S| > 2$,

$$\pi(x|S) \leq \max\{\pi(x|T) \mid x \in T \subset S \text{ and } |T| > 1\}$$

This axiom is a weaker version of regularity. While the regularity requires that $\pi(x|S)$ is smaller than all $\pi(x|T)$ for all subsets of S, the weak-regularity only requires it to be smaller than at least one of them.

We now state our characterization result for L-PRC$_\succ$.

Theorem 2. *Let $\succ$ be a reference ordering. A stochastic choice π satisfies U-regularity with respect to $\succ$ and weak-regularity if and only if there exists a unique L-PRC$_\succ$ representation of π.*

Note that Theorem 2 not only provides the necessary and sufficient conditions for L-PRC$_\succ$ 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 and weak-regularity.

Less-is-more property (ii) does not allow choice functions that presents difficult choice type of behavior in the support of randomization. However, some well known bounded rationality models such as limited attention allows for this behavior. Proof of Theorem 2 shows that dropping weak-regularity axiom on the stochastic choice is equivalent to dropping property (ii) from the less-is-more definition.

Remark 1. Let $\succ$ be a reference ordering. A stochastic choice π satisfies U-regularity with respect to $\succ$ if and only if there exists unique μ such that the support of μ satisfies progressiveness and property (i) of less-is-more with respect to $\succ$ and $\pi = \pi_{\mu}$.

4. Endogenous L-PRC

Up to now, we have taken the reference ordering as given. In some applications, the true reference ordering (such as a social norm) is observable to the researcher. However, when the reference ordering corresponds to the underlying preference ordering, it might not be exogenously given. In such cases, it must be inferred from choice. This section explores how one can identify the reference ordering of a stochastic choice which has some L-PRC representation. We first define an endogenous L-PRC representation: A stochastic choice π has an *endogenous* L-PRC representation if there exist an ordering $\succ$ such that L-PRC$_{\succ}$ represents π.

Next we provide a sufficient condition for the revealed ordering between a pair of alternatives.

Proposition 1. *Assume that π has an endogenous L-PRC representation. If $\pi(y|S) > \pi(y|\{x,y\})$ and $x \in S$, then x must be ranked above y for any reference ordering representing π.*

To see this, assume for contradiction there exists an ordering $\succ$ with $y \succ x$ and L-PRC$_\succ$ represents π. Since U-regularity must hold for $\succ$, for S such that $x, y \in S$ we have

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

This yields $\pi(y|S) \leq \pi(y|\{x,y\})$, which contradicts with the assumption of Proposition 1.

For unique identification of the reference ordering, next we assume strict stochastic choice: for all x, S, S', $p(x|S) \neq p(x|S') > 0$.[18]

Theorem 3. *If a strict stochastic choice π has an endogenous L-PRC representation, then the reference ordering is uniquely identified.*

The two preceding results, Proposition 1 and Theorem 3, were about revealing reference ordering. They are not applicable, however, unless the observed choice data has an endogenous L-PRC representation. Therefore, a question to ponder is: how can we test whether a stochastic choice data has a L-PRC representation? Surprisingly, it turns out that it can be characterized by using U-regularity and weak-regularity.

Note that U-regularity is defined for a specific reference ordering. In other words, it is applied on the pair $(\pi, \succ)$. When we do not know the underlying reference ordering, we need to check whether U-regularity holds for some $\succ$. If it does hold, we can represent the stochastic choice by endogenous L-PRC, otherwise the stochastic choice is outside of L-PRC framework. Therefore, we can define the axiom over the choice data only. In particular, we say that π satisfies U-regularity if there exist preferences $\succ$ such that $(\pi, \succ)$ satisfies U-regularity.

Remark 2. [Characterization] A stochastic choice π has an endogenous L-PRC representation if and only if π satisfies weak-regularity and there exist preferences $\succ$ such that $(\pi, \succ)$ satisfies U-regularity.

[18] This assumption cannot be rejected by any finite data set. In addition, it is usually made for estimation purposes.

This remark immediately follows from Theorem 2 because if there exists an endogenously defined $\succ$ making π satisfy U-regularity, then we can apply Theorem 2 to that $\succ$.

5. 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 L-PRC cases. To do this, first we introduce an ordering relation between distributions of choices in Definition 4. 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 4. Probability distribution μ defined on $\mathcal{C}$ is **higher** than probability distribution η defined on $\mathcal{C}'$ if $\forall \alpha \in (0, 1]$ and $\forall S \subset X$, $\mu_\alpha^{-1}(S) \succsim \eta_\alpha^{-1}(S)$.

Definition 4 compares two probability distributions and identifies the one which is more in line with the underlying preference, $\succ$, 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 the other distribution. Note that the comparison is based on $\succ$; hence, the compared models should have the same underlying $\succ$.

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 when choosing from a set. This is formally stated below.

Definition 5. 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 4. *Let π_μ and π_η be two $PRC_\succ$. 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.) For such models Definition 4 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 4.

Also note that if two decision makers (or two populations) have PRCs with the same underlying $\succ$ 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 $\succ$. In other words, she makes worse mistakes (in the sense of not being aligned with $\succ$) more often.

As previously mentioned, two decision makers' $\text{PRC}_\succ$ may have different supports. For example, say two decision makers use limited attention models similar to Example 5. 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 $\succ$. 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 $\succ$ (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).

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 the endogenous L-PRC model imposes testable restrictions, we now compare this special subclass to other stochastic choice models. As we mentioned before, SCRUM of Apesteguia et al. [2017] is a special case of L-PRC. Moreover, L-PRC 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.[19] The random attention model (RAM) of Cattaneo et al. [2019] covers the model of Brady and Rehbeck [2016] (BR), which in turn contains the model of Manzini and Mariotti [2014] (MM). Indeed, RAM includes RUM, BR, SCRUM and MM. However, RAM and L-PRC are independent models because there are choice data represented by L-PRC but not RAM and vice versa. 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 belongs to L-PRC but not RAM. Indeed, this example is outside of any models discussed above. Moreover, it is routine to show that L-PRC is independent of RAM, RUM, BR, and MM.

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 L-PRC.

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

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.[20] GLM reduces to the Luce rule when all alternatives are chosen with positive probability in all menus. Hence, L-PRC and GLM are distinct 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 L-PRC but not PALM, consider any data where the outside option is never chosen. When the outside option is never chosen, PALM reduces to the Luce rule. However, L-PRC 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 L-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, L-PRC is distinct from these, too.

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

7. Conclusion

We have introduced a novel PRC model which not only identifies uniquely 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. The first extension we investigated in this paper imposes the less-is-more structure on the choices of a PRC model, and introduces L-PRC. This model has stronger prediction power and still applies to most of the examples we discussed. However, it leaves out some well-studied bounded rationality models such as the limited attention model with attention filter. 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 L-PRC. Finally, several empirical queries arise from the present work. It would certainly be useful to conduct tests comparing the explanatory power of competing models that we reviewed in this paper.

The empirical validity of our axioms can be investigated. Our initial analysis of the data sets provided by Manzini and Mariotti [2006] on time preferences, confirm 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.

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.

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*.

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.

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.

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.

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 (2006): "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.

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.

Simon, H. A. (1955): "A Behavioral Model of Rational Choice," *Quarterly Journal of Economics*, 69, 99–118.

Trueblood, J. S., S. D. Brown, A. Heathcote, and J. R. Busemeyer (2013): "Not just for consumers: Context effects are fundamental to decision making," *Psychological science*, 24, 901–908.

Tversky, A. and I. Simonson (1993): "Context-dependent preferences," *Management science*, 39, 1179–1189.

Tyson, C. J. (2013): "Behavioral Implications of Shortlisting Procedures," *Social Choice and Welfare*, 41, 941–963.

Appendix

Proof of Lemma 1. Assume that $\{P_1, ..., P_T\}$ satisfies the single-crossing property with respect to $\succ$. 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 $c_s(S) \prec c_t(S)$. By definition of c, we have $c_s(S) P_s c_t(S)$ and $c_t(S) P_t c_s(S)$. Note that since $c_t(S) \succ c_s(S)$ by single crossing property we must have $c_t(S) P_t c_s(S) \Rightarrow c_t(S) P_s c_s(S)$, which is a contradiction.

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

Proof of Theorem 1. We assume that the stochastic choice function π is given and will construct a collection of choice functions $\mathcal{C}$ satisfying the progressiveness conditions and a probability distribution μ on it where the corresponding PRC with μ is equivalent to π.

Define
$$K = \{\pi(L(x) \cup x | S) \mid S \subseteq X \text{ and } x \in S\}^{21}$$
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., $0 < k_1 < k_2 < \cdots < k_m = 1$.[22] 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 $\min_{\pi^+}(\succ, S)$, which selects the worst alternative in S according to $\succ$ with strictly positive choice probability. That is,
$$\min_{\pi^+}(\succ, S) = \{x \in S \mid \pi(x|S) > 0 \text{ and } y \succ x \text{ whenever } \pi(y|S) > 0 \text{ and } y \neq x\}$$

Step 1: Define
$$c_1(S) = \min_{\pi^+}(\succ, S) \text{ and } \mu(c_1) = k_1$$
Note that $\mu(c_1)$ is positive and for any S, $\pi(c_1(S)|S) = \pi(L(c_1(S)) \cup c_1(S)|S) \geq k_1$ 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 $\pi(L(c_1(S)) \cup c_1(S)|S) = k_1$ since $k_1 \in K$.

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

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

Step 2: Define the second choice as

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

This is well-defined because by construction in first step: $\pi(L(c_1(S)) \cup c_1(S)|S) \geq k_1$. 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, $c_2(S) \succsim c_1(S)$ by definition of c_2 and hence, $\{c_1, c_2\}$ satisfies progressiveness with respect to $\succ$. Note that $\mu(c_1) + \mu(c_2) = k_2$. Moreover, there exists a subset S such that $\pi(L(c_2(S)) \cup c_2(S)|S) = k_2$ since $k_2 \in K$.

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

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

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

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

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

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

Define $\mathcal{C} = \{c_1, ..., c_m\}$ where each c_i is defined in Step i above. Since $\mathcal{C}$ satisfies progressiveness with respect to $\succ$, 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 $\pi(L(x) \cup x|S) = k_i$ and $\pi(L(x)|S) = k_j$. Since $L(x) \subset L(x) \cup x$ 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 $k \leq j$, $x \succ c_k(S)$ and $x \prec c_k(S)$ for all $k \geq i+1$. Then we have

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

Hence, π_μ and π are the same.

(Uniqueness): Let μ_1 with support $\mathcal{C}^1 = \{c_1^1, ...c_{n_1}^1\}$ and μ_2 with support $\mathcal{C}^2 = \{c_1^2, ...c_{n_2}^2\}$ 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 $M_i(c_t^i) = \sum\limits_{s\leq t}\mu_i(c_s^i)$ 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 $x \succ y$. By progressiveness, for any $k \leq t$, $c_k^1(S) \precsim y$ and for any $l \geq s$, $c_l^2(S) \succsim x$. Then $\pi_{\mu_2}(L(y)\cup y|S) < \alpha \leq \pi_{\mu_1}(L(y)\cup y|S)$ 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. **(Necessity):** Let $(\mu, \mathcal{C})$ represent π such that $\mathcal{C}$ satisfies less-is-more condition. Let $x \in T \subseteq S \subseteq X$ and $\pi(x|S) \neq 0$. First, we will show that for any $c_i \in \mathcal{C}$, $c_i(T) \prec x \Rightarrow c_i(S) \prec x$. Assume not, there exists i such that $c_i(T) \prec x \precsim c_i(S)$. 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 $c_j(T) \succsim c_j(S)$. Since $j \leq i$, by the betweenness property, $c_i(T) \succsim c_j(T) \succsim c_j(S) = x$ which contradicts with $c_i(T) \prec x$. Therefore, we prove our claim, which gives the following relations:

$$\pi_\mu(L(x)|T) = \sum_{c_i(T)\prec x}\mu(c_i) \leq \sum_{c_i(S)\prec x}\mu(c_i) = \pi_\mu(L(x)|S)$$

Hence,

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

We will next show the necessity of the weak-regularity condition. For contradiction assume there exists S and $x \in S$ such that for every $T \subset S$, $\pi_\mu(x|S) > \pi_\mu(x|T)$. Since x is chosen from S with positive probability, there must be some choice functions in $\mathcal{C}$ that chooses x on S. Let c_i be the choice function with the highest index and choosing x, i.e., $c_i(S) = x$ and for any $j > i$ $c_j(S) \neq x$. By less-is-more property (ii), there must be $T' \subset S$ such that $c_i(T') = x$. Denote the strict upper contour set of x as $\bar{U}(x) = \{y \mid y \succ x\}$. Then by construction

$$\pi_\mu(\bar{U}(x)|S) = \mu(c_{i+1}) + ... + \mu(c_N) \geq \pi_\mu(\bar{U}(x)|T')$$

since i was the highest indexed choice function choosing x on S and $\mathcal{C}$ has progressive structure. Then

$$\pi_\mu(U(x)|S) = \pi_\mu(x|S) + \pi_\mu(\bar{U}(x)|S) > \pi_\mu(x|T') + \pi_\mu(\bar{U}(x)|T') = \pi_\mu(U(x)|T')$$

This contradicts with U-regularity that we proved above. Hence, weak-regularity must be satisfied by the stochastic choice.

(Sufficiency): We assume the stochastic choice function π satisfies U-regularity and weak-regularity and will show that the construction of $\mathcal{C}$ given in the proof of Theorem 1 satisfies the less-is-more property.

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

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

where $L(x) = \{y \in X \mid x \succ y\}$ (strict lower counter set).

We first show c_1 satisfies the less-is-more property (i). Let $c_1(S) \in T \subseteq S$. By construction, $\pi(c_1(S)|S) \neq 0$ and for all $x \prec c_1(S)$ we have $\pi(x|S) = 0$. Hence, $\pi(L(c_1(S))|S) = 0$. By Axiom 1, $\pi(L(c_1(S))|T) = 0$. Hence, for all $x \prec c_1(S)$ we have $\pi(x|T) = 0$. Since $\pi(c_1(T)|T) \neq 0$ by construction, we must have $c_1(T) \succsim c_1(S)$.

Assume that all c_t satisfy the less-is-more property (i) when $t < i$. We now show that c_i also satisfy it. Let $c_i(S) \in T \subseteq S$. For contradiction, assume $c_i(S) \succ c_i(T)$. 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, $c_{i-1}(T) \precsim c_i(T) \prec c_i(S) = c_{i-1}(S)$. Then transitivity implies $c_{i-1}(S) \succ c_{i-1}(T)$, which contradicts the fact that c_{i-1} satisfies the less-is-more property (i).

Case 2) $c_{i-1}(S) \neq c_i(S)$. Then the following relations hold

$$\begin{aligned}
k_i &\leq \pi(L(c_{i-1}(T)) \cup c_{i-1}(T)|T) \\
&\leq \pi(L(c_i(T)) \cup c_i(T)|T) \quad \text{since } c_{i-1}(T) \precsim c_i(T) \\
&\leq \pi(L(c_i(S))|T) \quad \text{since } c_i(S) \succ c_i(T) \\
&\leq \pi(L(c_i(S))|S) \quad \text{since Axiom 1} \\
&= k_{i-1} \quad \text{since the choice on } S \text{ changed in step } i
\end{aligned}$$

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

We now show c_i satisfies the less-is-more property (ii). For a contradiction, let $S \subset X$ such that $c_i(S) = x$ and for all $\{x\} \neq T \subset S$, we have $c_i(T) \neq x$. By the less-is-more property (i), it must yield $c_i(T) \succ x$. If x is the worst alternative in S according to $\succ$, then $\pi(x|S) \geq \mu(c_i) + \sum_{t=1}^{i-1} \mu(c_t) > \pi(x|T)$ since $\mu(c_i) > 0$. This contradicts to the weak-regularity. If x is not the worst alternative in S, then there exists $y \in S$ such that $x \succ y$. Then $c_i(\{x, y\})$ must be x by the less-is-more property (i). This contradicts with $c_i(T) \neq x$ for every $T \subset S$. This shows the less-is-more condition (ii) for c_i.

This completes the proof of sufficiency. $\square$

Proof of Theorem 3. Before we proceed with the proof, we prove two lemmata.

Lemma 2. *The reference ordering among x, y and z is $xPyPz$ if one of the following conditions holds:*

(1) $\pi(y|S) > \pi(y|\{x, y\})$ *and* $\pi(z|T) > \pi(z|\{y, z\})$ *where* $x \in S$ *and* $y \in T$,
(2) *Among x, y and z, choices satisfy strict regularity except* $\pi(z|\{x, y, z\}) > \pi(z|\{y, z\})$.

Proof. The first part of Lemma 2 follows from the application of Proposition 1 (xPy and yPz) and transitivity ($xPyPz$). Hence, two regularity violations could deliver unique identification with three alternatives.

To prove the second part, assume $\pi(z|\{x, y, z\}) > \pi(z|\{y, z\})$, which implies yPz by Proposition 1. We now argue that x must ranked above y, hence the reference ordering is uniquely identified. Otherwise, there are two possible reference orderings: $y \succ_1 z \succ_1 x$ or $y \succ_2 x \succ_2 z$. Since the worst alternative with respect to each ordering used for a representation has to weakly violate the regularity condition, we must have either $\pi(x|\{x, y, z\} \geq \pi(x|\{x, z\})$ since $\succ_1$ represents or $\pi(z|\{x, y, z\}) \geq \pi(z|\{x, z\})$ since $\succ_2$ represents. Both cases imply a strict violation of regularity. Hence $xPy(Pz)$. $\square$

Lemma 3. *If xPy and xPz, 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 an L-PRC representation and xPy and xPz are assumed, there are two possible reference orderings between y and z: $x \succ_1 y \succ_1 z$ or $x \succ_2 z \succ_2 y$. Since the worst alternative with respect to each ordering used for a representation has to weakly violate the regularity condition (Axiom 2), we must have either $\pi(z|\{x, y, z\} \geq \pi(z|\{y, z\})$ or $\pi(y|\{x, y, z\}) \geq \pi(y|\{y, z\})$. $\square$

Next we will prove Theorem 3, i.e., the uniqueness of the reference ordering for an L-PRC. For contradiction, assume that π has two L-PRC representations: L-PRC$_{\succ_1}$

and L-PRC$_{\succ_2}$. Then there exist two alternatives x and y such that x is ranked above y in one representation ($x \succ_1 y$) and y is ranked above x in another representation ($y \succ_2 x$). This implies that we cannot have xPy or yPx. 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 Proposition 1 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 $A_xA_y > 0$, 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 $A_xA_y < 0$, $A_x < 0 < A_y$ and $A_x > 0 > A_y$ are symmetric so the proofs or 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. By Axiom 2, we must have $\pi(z|\{x,z\}) < \pi(z|\{x,y,z\}) < \pi(z|\{y,z\})$. Since the second part of Lemma 2 is applicable, the revealed reference ordering is unique, a contradiction.

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

Since we assumed above without loss of generality $\pi(z|\{x,z\}) < \pi(z|\{y,z\})$, Axiom 2 implies $\pi(z|\{x,y,z\}) < \pi(z|\{y,z\})$.

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, 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 3 contradicting with Equation 1. If (ii) holds, then both y and z satisfy strict regularity in $\{x,y,z\}$. By the second part of Lemma 2, the revealed reference ordering is unique, a contradiction.

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

Since $A_x > 0$, x satisfies the strict regularity in $\{x,y,z\}$. Therefore, either y or z (or both) violates the strict regularity in $\{x,y,z\}$. If only y violates it, then we must have $\pi(y|\{y,z\}) < \pi(y|\{x,y,z\}) < \pi(y|\{x,y\})$ by having $A_y > 0$ and weak-regularity.

Again, by the second part of Lemma 2, the revealed reference ordering is unique, a contradiction. Similarly, we get a contradiction if only z violates it. Hence, both y and z must violate it. Then, by Equation 1, we have $\pi(y|\{y,z\}) < \pi(y|\{x,y,z\})$ and by the assumption of $\pi(z|\{x,z\} < \pi\{z|\{y,z\}$, we have $\pi(z|\{x,z\}) < \pi(z|\{x,y,z\})$. By the first part of Lemma 2, the revealed reference ordering is unique, 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: 3(a) $\pi(z|\{x,z\}) < \pi(z|\{y,z\})$, and 3(b) $\pi(z|\{x,z\}) > \pi(z|\{y,z\})$.

Case 3(a): Assume $\pi(z|\{x,z\}) < \pi(z|\{y,z\})$. If $\pi(y|\{y,z\}) > \pi(y|\{x,y,z\})$, y also satisfies the strict regularity in $\{x,y,z\}$ by Equation (1). Since π satisfies U-regularity, z must violate the regularity. By Axiom 2, we must have $\pi(z|\{x,z\}) < \pi(z|\{x,y,z\}$. Since the second part of Lemma 2 is applicable, the revealed reference ordering is unique, 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. By the first part of Lemma 2, the revealed reference ordering is unique, a contradiction. If (ii) holds, then both x and z satisfy the strict regularity in $\{x,y,z\}$. By the second part of Lemma 2, the revealed reference ordering is unique, a contradiction.

Case 3(b): Assume $\pi(z|\{x,z\}) > \pi(z|\{y,z\})$. Since x satisfies strict regularity then either y or z must violate the strict regularity. We next show that both y and z cannot violate the strict regularity at the same time. If y violates it, then we must have $\pi(y|\{y,z\}) < \pi(y|\{x,y,z\})$. This implies $\pi(z|\{y,z\}) > \pi(z|\{x,y,z\})$ since $\pi(x|\{x,y,z\}) > 0$. Then z must satisfy it. Similarly, if z violates it, then we must have $\pi(z|\{y,z\}) < \pi(z|\{x,y,z\})$. This implies $\pi(y|\{y,z\}) > \pi(y|\{x,y,z\})$ since $\pi(x|\{x,y,z\}) > 0$, hence y must satisfy it. By Lemma 2, the revealed reference ordering is unique, a contradiction.

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. □

Proof of Theorem 4. 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 $S = \{x_1, ..., x_n\}$ 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 from the probability of choosing lower contour set of a_i in S by using π_μ and π_η, respectively, i.e. $\alpha = \pi_\mu(L(a_i), S)$ and $\beta = \pi_\eta(L(a_i), S)$; 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 $c_t(S) \prec a_i$; $\eta(c'_1) + ... + \eta(c'_{t'}) = \beta$ and $c'_{t'}(S) \prec a_i$. 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 $\mu_\alpha^{-1}(S) = c_t(S) \succsim c'_k(S) = \eta_\alpha^{-1}(S)$. Then we have $a_i \succ c_t(S) \succsim c'_k(S) \succsim a_i$. 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 $\eta_\alpha^{-1}(S) \succ \mu_\alpha^{-1}(S)$. Define x and y as $x = \eta_\alpha^{-1}(S)$ and $y = \mu_\alpha^{-1}(S)$, then $x \succ y$. 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. $\square$