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 an exogenous order and we call this structure progressiveness. Progressiveness imposes an intuitive structure on well known models. A PRC representation identifies the collection of choices and associated probabilities uniquely. We further impose a "less-is-more property" on the choice functions in the collection and call this an L-PRC model. The characterization of L-PRC relies on one simple axiom on stochastic choice: Lower Contour Set Monotonicity.

Date: April, 2020.
* We thank David Ahn, Paul Cheng, David Dillenberger, Keaton Ellis, and Collin Raymond 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

Understanding human behavior is at the core of economics as well as other disciplines such as psychology and marketing. This task presents challenges when the observed choice behavior is obtained in the form of aggregate data generated by a population or individual data collected under different situations where the frequencies of certain choices (rather than degenerate choices) are available. Such data sets present stochastic choice behavior. This fundamental problem was first studied by Thurstone [1927] and later developed by Marschak [1959], Block et al. [1959], Luce [1959], Suppes and Luce [1965]. The Thurstonian Random Utility Maximization (RUM) model is a collection of utility functions endowed with a probability distribution. In RUM, the choice frequency of an alternative is determined by the sum of the probability masses of utility functions where the alternative is the utility maximizer. RUM has been proven to be an important tool for the analysis of both "interpersonal" and "intrapersonal" variation in choices.

The key assumption of RUM is that each variation in choices is modeled as a preference maximization. However, 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 (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]). People systematically violate the requirements of utility maximization. Motivated by these findings, we construct a model that allows individual choices to exhibit desired choice patterns which are outside of the utility maximization paradigm. In this model, there is a collection of choice functions rather than utility functions endowed with a probability distribution. We call this model the *random choice* model. As in RUM, the collection of choice functions in a random choice 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).

While the random choice model is a useful framework, it has two shortcomings: First, it is too general to make any prediction (i.e. any choice behavior can be generated as an outcome of random choice.) Second, the representation is not unique (i.e., the same choice behavior can be explained by two completely different collections of choices). To address the second issue, we propose a useful restriction on the collection of choice functions which has practical and theoretical relevance. In particular, we focus on cases where each collection can be ordered in such a way that the choice functions become more and more aligned with a choice induced by a reference ordering. We call this property progressiveness with respect to an exogenous reference ordering and name such models Progressive Random Choice (PRC) models. The reference ordering can correspond to the decision maker's preferences, temptation ranking, measure of salience, or social norms (see examples in the main text.)

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.[1] 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 the willpower model (see Strotz [1955], Masatlioglu et al. [2020]). If each choice used in a PRC follows a willpower model with different levels of willpower, the monotone relationship in the willpower will lead to progressive choice functions.

The progressiveness structure in the PRC is the novelty of our model, as this assumption provides a meaningful interpretation for our representation. Recall that the

[1]When there is no alternative above the threshold, the decision maker searches the entire set and chooses the best one.

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 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 reference, the more severe the lapse is. In addition, PRC allows a substantial degree of heterogeneity of choice behavior. To illustrate this, we provide other examples of the random choice satisfying progressiveness property within the class of the shortlisting models of Manzini and Mariotti [2007] and Tyson [2013], the rationalization model of Cherepanov et al. [2013], the preferred personal equilibrium model of Kőszegi and Rabin [2006], and the other-regarding preference model of Dillenberger and Sadowski [2012].

The progressiveness 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 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].

Contrary to RUM, our representation is unique. We show that we can uniquely identify both the collection of choice functions and the exact probability weight of each choice in the collection satisfying progressiveness. The unique weight attached to each choice function is the likelihood of having that sort of deviation from the reference (rational) choice. We also provide a notion of comparative statics which allow us to rank any two PRCs within our framework.

The unique PRC representation of a stochastic choice behavior enjoys high explanatory power. To improve the prediction power of the model, we impose a bounded rationality condition, namely "less-is-more." That is, choices in the smaller sets cannot be dominated by the choices in the larger sets with respect to the underlying preference. This concept of bounded rationality is used to capture situations where it is more likely to choose suboptimally on larger alternative sets than smaller ones.

We call stochastic choices that can be represented by a random choice model satisfying the progressiveness and less-is-more properties L-PRC. The characterization of L-PRC requires a single axiom we call *Lower Contour Set Monotonicty*. The axiom restricts the stochastic choice to choose any lower contour set with higher probability on a smaller alternative set than a larger one. The axiom collapses to the standard weak axiom of revealed preference if the data is degenerate.

L-PRC is distinct from several well-known stochastic choice models. We show that while SCRUM (Apesteguia et al. [2017]) is a special case of L-PRC, the intersection of RUM and L-PRC contains more than SCRUM. The models provided by Manzini and Mariotti [2014], Brady and Rehbeck [2016], and Cattaneo et al. [2019] are also distinct from L-PRC.

Our paper is 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 [2018]. 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. [2018] 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 imposes an additional structure, *less-is-more*, on the PRC and states the single necessary and sufficient condition for the existence of a representation with such structure, called L-PRC. Section 4 provides comparative statics between any two models within our framework. Section 5 summarizes how L-PRC relates to other well-known stochastic choice models. Section 6 discusses how the underlying preference used for the representation can be endogenized and how one can perform revealed preference analysis in this setting. Section 7 concludes.

2. Model

Let X denote a finite set of alternatives and $\succ$ be a linear order on X where $\succeq$ is the weak order derived from $\succ$ in the usual sense.[2] $\succ$-best denotes the best alternative of a set with respect to relation $\succeq$. 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

[2]While we assume an exogenous $\succ$, in Section 6 we will discuss when $\succ$ can be endogenously derived.

order, then π_μ becomes the well-known RUM. Hence RUM is a special case of random choice. If we do not impose any restriction on the support of μ, any stochastic choice function can be represented within random choice framework. That is, for every π, there exists μ such that $\pi = \pi_\mu$. Moreover, the representation is not unique in general.[3]

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 $t \geq s$.[4] Progressiveness imposes a structure on the sorted collection of choices such that a higher indexed choice function cannot choose an alternative that is dominated by the lower indexed choice function on the same choice set. In other words, c_t is more aligned with $\succ$ than c_s when $s < t$. Note that any subset of a progressive set is also progressive. Throughout this paper we will refer to this exogenous reference ordering, $\succ$, as preference. However, as some examples will clarify, the ordering may correspond to the decision maker's temptation ranking, measure of salience, or social norms, see e.g. Examples 1, 3, and 5.

To illustrate the concept of progressiveness better, we provide some examples. Table 1 illustrates two collections of choice functions, $\mathcal{C}$ and $\mathcal{C}'$, on $\{x, y, z\}$. Given three alternatives, the cardinality of the progressive set cannot be more than 6. Hence, any collection with more than six choice functions cannot be progressive. $\mathcal{C}$ is progressive with respect to relation $x \succ y \succ z$ since, for each row, the choices improves in the

[3]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.

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

direction of $\succ$.[5] Four out of six types of individual choices (c_1, c_3, c_4, and c_6) are rational (with different preferences). Hence, this concept allows q high-degree of heterogeneity in terms of preferences. While c_6 is completely aligned with $\succ$, c_1 is completely opposite of it. $\mathcal{C}'$ is also progressive with respect to $x \succ y \succ z$. Contrary to $\mathcal{C}$, $\mathcal{C}'$ contains only four types of individual choices. All of these violate WARP; hence, there is no choice type completely aligned with $\succ$. The collection of the six choice functions generated by the six possible preference relations for three objects is not progressive for any ordering.

	$\mathcal{C}$						$\mathcal{C}'$			
	c_1	c_2	c_3	c_4	c_5	c_6	c'_1	c'_2	c'_3	c'_4
$\{x, y, z\}$	z	z	y	y	y	x	z	z	y	x
$\{x, y\}$	y	y	y	y	x	x	y	x	x	x
$\{x, z\}$	z	z	z	x	x	x	z	z	z	z
$\{y, z\}$	z	y	y	y	y	y	y	y	y	y

TABLE 1. Two progressive collections of choice functions.

We view the progressiveness structure as a strength of our model, as this assumption provides a meaningful interpretation for the support of our model. Recall that the support of RUM consists of several independent preferences. In contrast, progressiveness imposes a consistency between the different types of choices in the support with respect to an exogenous reference relation, $\succ$. These types can be ordered in a way such that choices of each type become more and more rational with respect to this relation. 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. In each example, there is one parameter varying across choice types.

Example 1. (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 created by the first binary*

[5] Note that $\mathcal{C}$ is also progressive with respect to $z \succ y \succ x$. Indeed, our definition implies that if $\mathcal{C}$ is progressive with respect to $\succ$, it is also progressive with respect to the opposite of $\succ$.

relation (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 2. (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 3. (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.*[6]

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

Example 4. (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 5. (Other-regarding Preferences) *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. As each type cares more and more about the shame component, the collection of choices is progressive with respect to the social norm.*[7]

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.[8] As the next Lemma shows, if the choices in the support of a random choice model are rational and generated

[7]Formally, we assume that the choice can be written as $c_s(T) = \text{argmax}_{x\in T} \quad \{u(x) - (\max_{y\in T}\psi(y) - \psi(x))^s\}$ where u is a utility function over allocations, ψ represents the norm, $(\max_{y\in A}\psi(y) - \psi(x))^s$ is interpreted as the shame from choosing x in comparison to the alternative that maximizes the norm, and s is the shame parameter. Assume that we have a unique maximizer $x(s)$ so that choice is a function. In addition, assume that the shame function is in the following form: $g(a, b) = (b - a)^s$ where s is the shame parameter. Then if each choice type in the PRC has a different shame parameter, s, then the larger s brings the choice closer to the social norm.

[8]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.

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

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

When there is no need to specify, we will abuse the notation and denote such representations by PRC without indicating the corresponding preference.

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. *Every stochastic choice π has a PRC representation with respect to $\succ$. 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 $\succ$.[9] 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, 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. 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 3 for the random choice model). Hence, one obvious strength of the PRC is its uniqueness. Moreover, it is easy to construct an example of stochastic choice that has a unique RUM representation, such as the Luce model, whose support of preference relations is strictly larger than the support of the stochastic choice's PRC representation. As can be seen in the next example, the RUM representation of the Luce model makes use of all the preference relations and assigns positive weights to each of them. In contrast, the PRC representation of the same stochastic choice has a smaller support and the relation between the elements of the support can be easily interpreted by using the progressiveness property.

Example 7. Let π be given by

π	$\{x,y,z\}$	$\{x,y\}$	$\{x,z\}$	$\{y,z\}$
x	1/3	1/2	1/2	–
y	1/3	1/2	–	1/2
z	1/3	–	1/2	1/2

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

This choice data has a unique RUM (or Luce) representation that assigns equal weight to all six possible linear orders on $\{x, y, z\}$ as shown in the left hand side of Table 2. This choice data also has a unique PRC representation with $x \succ y \succ z$ with only four choice functions in the support as shown in the right hand side of Table 2. Moreover, there is an intuitive interpretation of the behavior under the PRC representation. If we assume that $\succ$ is the decision makers rational choice and her stochastic choice is the result of her mistakes, then we interpret c_4 as her most rational choice and the other lower indexed choices she randomly uses are ordered with respect to $\succ$. Hence, every time she uses a choice function different than c_4 we can measure how far her choice is from the rational one and how often she makes that kind of mistake.

	RUM						PRC			
	$c_{\succ_1}$	$c_{\succ_2}$	$c_{\succ_3}$	$c_{\succ_4}$	$c_{\succ_5}$	$c_{\succ_6}$	c_1	c_2	c_3	c_4
$\{x, y, z\}$	z	z	y	y	x	x	z	y	y	x
$\{x, y\}$	y	x	y	y	x	x	y	y	x	x
$\{x, z\}$	z	z	z	x	x	x	z	z	x	x
$\{y, z\}$	z	z	y	y	z	y	z	z	y	y
μ	1/6	1/6	1/6	1/6	1/6	1/6	1/3	1/6	1/6	1/3

TABLE 2. RUM and PRC representations for the choice data provided in Example 7

3. Less is More

In this section, we impose an additional structure on the elements of the choice collection used in a PRC and increase the prediction power of the model. Since our aim is to allow stochastic choice of boundedly rational individual(s), we utilize one condition proposed in the literature for bounded rationality, namely "less-is-more". This property requires choices on smaller sets to be preferred to the choices on larger sets. We say a collection of choice functions $\mathcal{C}$ satisfies *less-is-more* with respect to $\succ$ if for all t, $c_t(T) \succsim c_t(S)$ whenever $c_t(S) \in T \subset S$. That also 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. This restricts each possible choice in the support of randomization to be boundedly rational in the sense of "less-is-more."

Definition 2. 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 $\pi = \pi_{\mu}$.

When there is no need to specify, we will abuse the notation and denote such representations by L-PRC without indicating the corresponding preference.

All the models discussed in Examples 1 - 6 can be modified to accommodate the less-is-more structure. For the shortlisting example, Example 1, 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. Similarly, for the limited attention example, Example 4, if the attention correspondence in each choice in the RCM is based on a competition filter (i.e. each consideration operator, Γ_i, considers weakly better alternatives on smaller sets, see Lleras et al. [2017]), then the RCM 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 in Example 4 are attention filters (i.e. each consideration operator, Γ_i, considers weakly dominated alternatives on smaller sets, see Masatlioglu et al. [2012]), then the choices 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.

The characterization of L-PRC relies on a single axiom, defined below. Let $\pi(x|S)$ be a stochastic choice function and $\succ$ be an exogenous linear order. $L(x) = \{y \in X \mid x \succ y\}$ denotes the lower contour set of an alternative, x.

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

$$\pi(L(x)|S) \geq \pi(L(x)|T)$$

To grasp the intuition behind this axiom, consider the deterministic choices, i.e., $\pi(x|S) \in \{0,1\}$. First, if this choice satisfies the weak axiom of revealed preference (WARP) and is represented by $\succ$ then $\pi(x_S|S) = 1$ if and only if x_S is the $\succ$-best alternative in S. By definition, $\pi(L(x_S)|S) = 0$. In this case, the axiom implies that whenever x_S is present in any $T \subset S$, x_S is going to be chosen with probability 1. Hence, WARP and Axiom 1 imply the same restrictions on choice for this situation.

Axiom 1 is more general than WARP even in the deterministic world. To see this, assume $\pi(x_S|S) = 0$. In other worlds, the best alternative according to $\succ$ is not chosen (i.e. the choice cannot be represented by $\succ$.) Since the choices are deterministic, there exists an alternative x^* in S such that $\pi(x^*|S) = 1$. Hence, $\pi(L(x^*)|S) = 0$. Our axiom implies that $\pi(L(x^*)|T) = 0$ on the smaller set, T. Hence, the choice on T must be weakly better than x^*, as anything worse than x^* is chosen with probability zero. Therefore, Axiom 1 restricts the choice in deterministic case even when WARP is not satisfied.

The axiom requires the best elements to be more likely to be chosen on smaller sets. In other words, more mistakes are made on larger 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.

Axiom 1 is also related to the regularity property: $\pi(x|S) \leq \pi(x|T)$ for all $x \in T \subset S \subseteq X$. While Axiom 1 allows regularity violations, for the $\succ$-best alternative,

x_S of S, regularity and Axiom 1 coincide. To see this, if $\pi(x_S|S) = 0$, the claim is true trivially. Assume not, then for T such that $x_S \in T \subset S \subseteq X$ Axiom 1 implies that

$$\begin{aligned} \pi(L(x_S)|S) &\geq \pi(L(x_S)|T) \\ 1 - \pi(x_S|S) &\geq 1 - \pi(x_S|T) \\ \pi(x_S|S) &\leq \pi(x_S|T) \end{aligned}$$

Hence, the regularity holds for x_S for all S. Moreover, if we apply Axiom 1 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, respectively. Then, for T such that $y_S, z_S \in T \subset S \subseteq X$ and $\pi(y_S|S) \neq 0$, Axiom 1 implies that,

$$\begin{aligned} \pi(L(y_S)|S) &\geq \pi(L(y_S)|T) \\ \pi(z_S|S) &\geq \pi(z_S|T) \end{aligned}$$

Hence, the regularity is violated weakly for z_S for all S. Next is our main theorem.

Theorem 2. *Let π be a stochastic choice. π satisfies Axiom 1 with respect to $\succ$ if and only if there exists an L-PRC$_\succ$ representation of π. Moreover, the representation is unique.*

Note that Theorem 2 not only provides the necessary and sufficient condition for L-PRC representation but also concludes that the representation is unique. The algorithm generating the 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 Axiom 1.

4. 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 need to introduce an ordering relation between distributions of choices as in Definition 3. Before defining the order, we define, for all $\alpha \in (0,1]$, $\mu_\alpha^{-1} := c_i \in \mathcal{C}$ such that $\mu(c_1) + .. + \mu(c_{i-1}) < \alpha \leq \mu(c_1) + .. + \mu(c_i)$ 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 3. Probability distribution μ defined on $\mathcal{C}$ is **higher** than probability distribution η defined on $\mathcal{C}'$ if $\forall \alpha \in (0,1]$ and $\forall S \subset X$, $\mu_\alpha^{-1}(S) \succeq \eta_\alpha^{-1}(S)$.

Definition 3 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 4. Stochastic choice π first order stochastic dominates stochastic choice π' if for any set S and any $x \in S$,

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

where $U(x)$[10] is the upper contour set of x.

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

Theorem 3. *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.*

[10] $U(x) = \{y \in X \mid y \succeq x\}$

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 3 is equivalent to Definition of "a SCRUM being higher" in Apesteguia and Ballester [2013] (see page 667). Hence, their Proposition 2 is a special case of our Theorem 3.

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 4. 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).

5. L-PRC versus Other Stochastic Choice Models

In this section, we compare L-PRC with other well-known models of stochastic choice from the literature. Figure 1 illustrates the relationship between L-PRC and five well known stochastic choice models on the Marschak-Machina triangle in a domain of three alternatives, $X = \{x, y, z\}$. The alternatives are at the corners of the triangle. For each model, we assume that the underlying preference is $x \succ y \succ z$. We then fix stochastic choice for binary sets: $\pi(y|\{x, y\}) > \pi(z|\{x, z\}) > \pi(z|\{y, z\})$ and denote these by the red dots on the edges of the triangle. The figure identifies all possibilities for $\pi(\cdot|\{x, y, z\})$ for each model given these binary choices. The green area in Figure 1 denotes all the stochastic choices that satisfy Axiom 1. Hence, this region captures the explanatory power of L-PRC for these binary choices. As we mentioned before, SCRUM is a special case of our model, which is represented by the black dot in the interior of the Marschak-Machina triangle. RUM predictions are highlighted by the small upside down triangle with blue edges in the figure. Since the boundaries of RUM and L-PRC overlap, L-PRC includes many RUM choices other than SCRUM.

Manzini and Mariotti [2014], Brady and Rehbeck [2016], and Cattaneo et al. [2019] are the papers related to our work. Similar to ours, each of these models has a preference relation as one of their parameters. While the first two provide parametric random attention models, the last offers a non-parametric restriction on the 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.[11] Similar to SCRUM, Manzini and Mariotti [2014] can accommodate only one choice probability, which is marked as MM in Figure 1. Brady and Rehbeck [2016] can be in line with more choices, which are indicated by the black line in Figure 1. It is apparent that these models are not only different from each other but also distinct from L-PRC since their predictions do not overlap with the green region. While the MM point is included in the RUM triangle (the blue triangle), BR might be outside the RUM paradigm (so is L-PRC).

Finally, the random attention model (RAM) of Cattaneo et al. [2019] is highlighted by the orange area. The RAM model includes RUM, BR, SCRUM and MM. However,

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

RAM and L-PRC are independent models because there are choice data represented by L-PRC but not RAM and vice versa, as can be seen in the figure.

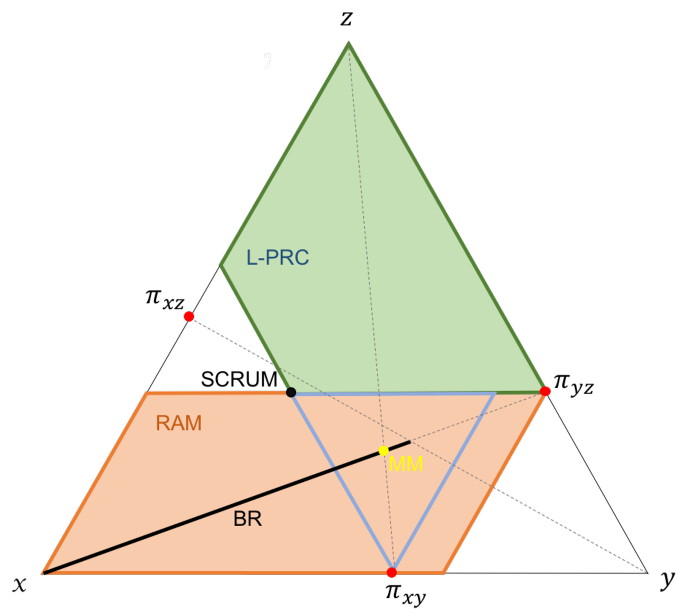

FIGURE 1. Given $x \succ y \succ z$, and choices on binary sets, this figure identifies all possibilities for $\pi(\cdot|xyz)$ for each model of interest. The predictions of L-PRC are highlighted by the green shaded area.

OTHER RELATED MODELS: Gul, Natenzon, and Pesendorfer [2014] consider the following model. The decision maker first randomly picks an attribute using the Luce rule given the weights of all attributes. Then the decision maker 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.

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.[12] GLM reduces to the Luce rule when all alternatives are chosen with positive probability in all menus. Since the choices on doubleton sets in Figure 1 do not satisfy Luce's Independence of Irrelevant Alternatives

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

condition (IIA)[13], GLM does not exist in Figure 1. If we pick binary data satisfying Luce's IIA, GLM either coincides with the Luce model or at least one alternative must be chosen with zero probability from the tripleton set. 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.

[13]Luce's Independence of Irrelevant Alternatives condition is violated since $\frac{\pi(x|\{x,y,z\})}{\pi(x|\{x,z\})} \neq \frac{\pi(z|\{x,y,z\})}{\pi(z|\{x,z\})}$.

6. Endogenous Reference Ordering

We have developed our model assuming an exogenously given reference ordering $\succ$. Indeed, in most applications, the outside observer has knowledge about the underlying ordering. For example, the prices of alternatives can be observed and used as a one dimensional attribute, sorting the alternatives exogenously. The exogenous ordering used in our construction of random choice model can be interpreted as such observable sorting. In this section, we discuss how we can extend our model to endogenous ordering, which can be recovered from the choice data.

Note that Axiom 1 is defined for a specific ordering. In other words, it is applied on the pair $(\pi, \succ)$. When we do not know the underlying preference, we need to check whether Axiom 1 holds for any preferences. If it does, we can represent the stochastic choice by L-PRC; otherwise, the stochastic choice is outside of the L-PRC framework. Therefore, we can define the axiom over the choice data only. In particular, we say that π satisfies Axiom 1 if there exist preferences $\succ$ such that $(\pi, \succ)$ satisfies Axiom 1.

Corollary 1. *A stochastic choice π has an L-PRC representation if and only if there exist preferences $\succ$ such that $(\pi, \succ)$ satisfies Axiom 1.*

This corollary immediately follows from Theorem 2 because if there exists an endogenously defined $\succ$ making π satisfy Axiom 1, then we can apply Theorem 2 to that $\succ$.

Since the choice of preferences brings an extra degree of freedom to our model, a natural question is whether the L-PRC model can explain any stochastic choice. The answer is no. Specifically, if the stochastic choice is a strict RUM,[14] then it cannot have an L-PRC representation. This follows directly from the fact that the worst alternative according to preference must violate the regularity condition weakly.[15] The following

[14]We say π is a strict RUM if $\pi(x|S) < \pi(x|T)$ for all $x \in T \subset S \subseteq X$, which is denoted by the interior of the blue triangle in the figure.

[15]Note that in Figure 1 when the preference is $x \succ y \succ z$, L-PRC intersects with RUM on one edge of the blue triangle. Similarly, if we change the preference, the intersection of L-PRC and RUM can only

example illustrates that L-PRC makes prediction even outside of a strict RUM as it shows a situation that cannot be explained by either a RUM or L-PRC with respect to some preference.

Example 8. Let π_λ be given by

π_λ	$\{x, y, z\}$	$\{x, y\}$	$\{x, z\}$	$\{y, z\}$
x	λ_x	0.70	0.65	–
y	λ_y	0.30	–	0.50
z	λ_z	–	0.35	0.50

where $\lambda_i \geq 0$ be the probability of choosing alternative i from $\{x, y, z\}$ and $\lambda_x + \lambda_y + \lambda_z = 1$. If $\lambda_x < 0.65$, $\lambda_y < 0.30$, and $\lambda_z < 0.35$, then π_λ is a strict RUM, which is a non-empty set. Therefore, such data does not have any L-PRC representation while it has a RUM representation. On the other hand, if $0.30 < \lambda_y < 0.50$ and $0.35 < \lambda_z < 0.50$, π_λ does not satisfy Axiom 1. Hence, in this case, the stochastic choice has neither RUM nor L-PRC representation. Therefore, L-PRC makes a prediction even outside of strict RUM.

Next, we introduce an identification for revealed preference. For this purpose, we will utilize Axiom 1 and focus on those linear orders that make a stochastic choice satisfy the axiom. Then, we will take the intersection of those orderings. Such a conservative specification of revealed preference has been employed also in Cattaneo et al. [2019] and other deterministic limited consideration models such as Masatlioglu et al. [2012].

Definition 5. Let $\{\succ_j\}_{j=1,\ldots,J}$ be all the preference relations such that $(\pi, \succ_j)$ satisfies Axiom 1. Then x is *revealed preferred* to y if $x \succ_j y$ for all j. Denote such a revealed preference by xPy.

occur on the other edges of the blue triangle. Hence, the interior of the blue triangle cannot intersect with L-PRC.

This definition provides an algorithm to define the revealed preference. Remember that Axiom 1 is the same as the regularity condition for the best alternative. Hence, the algorithm must start with such an alternative. Then we can check whether a specific binary relation satisfies Axiom 1. In Example 8, if $0 < \lambda_x < 0.70$, $0 < \lambda_y < 0.50$, and $0 < \lambda_z < 0.50$ and the representation exists, the preference is completely identified. On the other hand, if $\lambda_z = 1$, then we have x, yPz. Since the regularity condition is satisfied for both x and y, the model cannot reveal whether x is better than y, or the other way around.

To see how this definition aids revealed preference, let us assume that $\pi(z|S) > \pi(z|T)$ and $T \subset S$. By Axiom 1, we know that there must be something that is better than z in the set T. Also, if the smaller set contains only two elements, we can even claim revealed preference from this property.

Remark 1. If $(\pi, \succ)$ satisfies Axiom 1 and $\pi(z|S) > \pi(z|T)$, then there exists $x \in T$ such that $x \succ z$. In particular, if $T = \{x, z\}$, then xPz.

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 inter-personal or intra-personal variation in choice and the variation is solely based on one underlying exogenous parameter such as willpower, loss aversion, attention, limited cognitive ability, etc.

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 or alternative structures on PRC rather than less-is-more to investigate their implications on the choice behaviour. 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. Similarly, the empirical validity of our axiom can be investigated.

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. A. Ballester (2013): "Choice by Sequential Procedures," *Games and Economic Behavior*, 77, 90–99.

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.

Block, H. D., J. Marschak, et al. (1959): "Random orderings and stochastic theories of response," Tech. rep., Cowles Foundation, Yale University.

Brady, R. L. and J. Rehbeck (2016): "Menu-Dependent Stochastic Feasibility," *Econometrica*, 84, 1203–1223.

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.

Dardanoni, V., P. Manzini, M. Mariotti, and C. J. Tyson (2018): "Inferring Cognitive Heterogeneity from Aggregate Choices," *Working Paper Series 1018, Department of Economics, University of Sussex.*

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.

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.

Luce, R. D. (1959): *Individual choice behavior*, Wiley, New York.

Manzini, P. and M. Mariotti (2007): "Sequentially Rationalizable Choice," *American Economic Review*, 97, 1824–1839.

——— (2014): "Stochastic Choice and Consideration Sets," *Econometrica*, 82, 1153–1176.

Marschak, J. (1959): "Binary choice constraints and random utility indicators," Tech. rep., Cowles Foundation, Yale University.

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.

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

Strotz, R. H. (1955): "Myopia and Inconsistency in Dynamic Utility Maximization," *The Review of Economic Studies*, 23, 165.

Suppes, P. and R. D. Luce (1965): "Preference, Utility, and Subjective Probability," *Handbook of Mathematical Psychology*, 3, 249–410.

Thurstone, L. L. (1927): "A law of comparative judgment." *Psychological review*, 34, 273.

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 $c_s(S)P_sc_t(S)$ and $c_t(S)P_tc_s(S)$. Note that since $c_t(S) \succ c_s(S)$ 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 $\succ$. 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 xP_ty we have yP_sx. Then note that $c_t(\{x, y\}) = x$ and $c_s(\{x, y\}) = y$. By progressiveness, we should haves $c_s(\{x, y\}) \succeq c_t(\{x, y\})$ or equivalently, $y \succeq x$. This is a contradiction. □

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

$$K = \{\pi(L(x) \cup x|S) \mid S \subseteq X \text{ and } x \in S\}^{16}$$

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., $0 < k_1 < k_2 < \cdots < k_m = 1$.[17] 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

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

[16]We abuse the notation and write $A \cup x$ instead of $A \cup \{x\}$.
[17]k_m is always equal to 1 since $\pi(L(x) \cup x|\{x\}) = 1$.

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 $k_1 \in K$.

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, $c_2(S) \succeq c_1(S)$ 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

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

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

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

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

Note that by step $i-1$, $c_{i-1} \neq c_i$, and by construction $c_i(S) \succeq c_{i-1}(S) \succeq c_{i-2}(S) \succeq ... \succeq c_1(S)$ $\forall S$. 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 $M_i(c_t^i) = \sum_{s \leq t} \mu_i(c_s^i)$ 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 $k \leq t$, $c_k^1(S) \preceq y$ and for any $l \geq s$, $c_l^2(S) \succeq x$. 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 $c_i(T) \prec x \preceq c_i(S)$. 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 $c_j(T) \succeq c_j(S)$. Since $j \leq i$, by the betweenness property, $c_i(T) \succeq c_j(T) \succeq c_j(S) = x$ which contradicts with $c_i(T) \prec x$. Therefore, we prove our claim, which gives the following relations:

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

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

We first show c_1 satisfies the less-is-more property. Let $c_i(S) \in T \subseteq S$. 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 $c_1(T) \succeq c_1(S)$.

Assume that all c_t satisfy the less-is-more property 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, $c_{i-1}(T) \preceq c_i(T) \prec c_i(S) = c_{i-1}(S)$. 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.

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

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

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

This shows less-is-more condition for c_n. □

Proof of Theorem 3. Let π_μ and π_η be two L-PRC 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 $\mu_\alpha^{-1}(S) = c_t(S) \succeq c'_k(S) = \eta_\alpha^{-1}(S)$. Then we have $a_i \succ c_t(S) \succeq c'_k(S) \succeq a_i$. 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. $\square$