2016

2016-3 Random Categorization and Bounded Rationality

Victor H. Aguiar

Random Categorization and Bounded Rationality

by

Victor H. Aguiar

Research Report # 2016-3 **August 2016**

Department of Economics Research Report Series

Department of Economics
Social Science Centre
Western University
London, Ontario, N6A 5C2
Canada

Random Categorization and Bounded Rationality.*

Victor H. Aguiar†

This version: August 2016

Abstract

In this study we introduce a new stochastic choice rule that categorizes objects in order to simplify the choice procedure. At any given trial, the decision maker deliberately randomizes over mental categories and chooses the best item according to her utility function within the realized consideration set formed by the intersection of the mental category and the menu of alternatives. If no alternative is present both within the considered mental category and within the menu the decision maker picks the default option. We provide the necessary and sufficient conditions that characterize this model in a complete stochastic choice dataset in the form of an acyclicicity restriction on a stochastic choice revealed preference and other regularity conditions. We recover the utility function uniquely up to a monotone transformation and the probability distribution over mental categories uniquely. This model is able to accommodate violations of IIA (independence of irrelevant alternatives), of stochastic transitivity, and of the Manzini-Mariotti menu independence notion (i-Independence). A generalization of the categorizing procedure accommodates violations of regularity and thus provides an alternative model to random utility.
JEL classification numbers: C60, D10.
Keywords: decision theory, random choice, bounded rationality, categorization, consideration sets.

1 Introduction

Categorization has been recognized as an important part of the decision making process. Decision makers (DMs) categorize in order to simplify complex decision situations (Manzini & Mariotti, 2012). At the same time new evidence suggests that decision makers deliberately randomize when choosing (Agranov & Ortoleva, 2014). Here we provide a model that connects categorization, bounded rationality, and randomness in choice. The DM has access to a fixed set of categories defined as bundles of alternatives and, at any given trial, she considers with fixed probability one of those categories. Then the DM chooses according to her preferences the best item at the intersection of the considered category that is available in the menu. The probability of choosing a particular item in a menu is the sum of the probabilities of all mental categories that have a

*An earlier version of this paper was circulated with the name of "Random consideration sets: Inferring preferences from psychological biases" (2015). This paper also subsumes Aguiar (2015a).

†Department of Economics, University of Western Ontario, vaguiar@uwo.edu. An important part of this paper was written as part of my dissertation in Brown University. I am grateful to Roberto Serrano, Geoffroy de Clippel, Susanne Schennach, Drew Fudenberg, Tomasz Strzalecki, Itzhak Gilboa, Federico Echenique, David Jacho-Chavez, Anthony Marley, Miguel Castro, Luis Castro and John Rehbeck for useful comments and encouragement.

non-empty intersection with the menu and that, even more importantly, are such that there are no better alternatives within it than the fixed item.

This probabilistic categorization rule allows for menu dependence, and thus is more general than the popular model of limited attention and random choice put forward by Manzini and Mariotti(2014) (hereinafter MM), which is characterized among other conditions by a menu independence axiom on the propensity of considering any given item (i-Independence). It also allows for degenerate probabilities and, in fact, the proposed model nests the standard rational model with strict preferences with a categorization rule that consists of considering the category of all alternatives with probability 1. Stochastic intransitivity is also accommodated, as well as the similarity effect which represents a violation of the IIA axioms (independence of irrelevant alternatives).

The random categorization (RC) rule is characterized by the acyclicity of a stochastic revealed preference relation that consists of declaring a stochastically revealed preferred (strictly) to b if and only if the probability of choosing b in a menu is changed (either negatively or positively) by introducing a into such menu (i.e., if a has a non-zero impact in the probability of choosing b in a menu). The second condition is a total monotonicity requirement that is equivalent to the Block-Marschak conditions and thus makes our model a subcase of the random-utility model (RUM). A direct generalization of the main categorizing procedure that allows for category avoidance (modeled as a signed measure over categories instead of as a probability distribution) constitutes a new choice rule that is different from RUM. This generalization is able to accommodate violations of the regularity condition (i.e., introducing new alternatives to a menu can only decrease the probability of choosing the existing items).

The random categorization rule is distinct from other efforts to generalize MM and, in particular, it is neither nested nor does it nest the random feasibility rule proposed by Brady & Rehbeck (2015). Recently, Zhang (2016) proposes an empirically different categorization procedure that ours, that produces stochastic choice (e.g., the categories in his work are a partition while here they are arbitrary, among other differences). The author establishes that his rule is different from previous versions of the categorization rule presented in this paper (Aguiar, 2015a,b). The reader ought to compare the similarities.

Section 2 presents the environment and the dataset. Section 3 describes the representation of the RC rule and provides illustrations. Section 4 presents the characterization of the RC rule. Section 5 presents a generalization of this rule that allows for violations of the regularity condition that is necessary for RUM. Section 6 describes its relation with the literature. Section 7 concludes.

2 The Environment and Dataset

Formally, consider a finite choice set X. There is an always-available option $\{o\}$ (i.e., not choosing or a default that is always possible to obtain). A stochastic choice dataset is a set of menus $\mathcal{M} \subseteq 2^X$ together with a probabilistic choice map for any given menu and an item inside the menu: $p : X \cup \{o\} \times \mathcal{M} \to [0,1]$. That is the sequence $\{A, p(a,A)\}_{A\in\mathcal{M}, a\in X\cup\{o\}}$. The probability of choice is such that $\sum_{a\in A} p(a,A) + p(o,A) = 1$, where $p(o,A)$ denotes the probability of not choosing anything from A, and thus picking the outside option or default.[1] We fix $p(o,\emptyset) = 1$.

[1]The agent may or may not be forced to choose. When forced to choose $\sum_{a\in A} p(a,A) = 1$.

We say that the stochastic choice dataset is complete if $\mathcal{M}$ is the power set.

3 The Model: Random Categorization (RC) rule.

Having been given a menu, a DM who follows the random-categorization rule selects a mental category with a fixed probability and then chooses the item that maximizes its utility from those alternatives that belong to the considered category and to the given menu. In the event that no item in the considered category is in the menu, the DM picks the default alternative $\{o\}$.

Formally, a DM is endowed with a collection of categories over the choice set X. We take the categories as given but we do not observe them. Categories are a collection of subsets of X, formally $\mathcal{D} \subseteq 2^X$. The DM has a probability measure defined over the categories that represents her propensity to consider a given category at any given trial. A probability of consideration is a mapping $m : \mathcal{D} \mapsto [0,1]$ such that $\sum_{D\in\mathcal{D}} m(D) = 1$ and $m(D) \in [0,1]$. Finally, the DM also is endowed with a fixed utility function $u : X \mapsto \mathbb{R}$ that represents her tastes, we assume that it is injective or equivalently we rule out the possibility of indifference.

When facing a menu, the DM draws a mental category $D \in \mathcal{D}$ with probability $m(D)$ and then forms a consideration set $\mathbf{\Gamma}(D, A) = D \cap A$. Then she picks $a = argmax_{b\in\mathbf{\Gamma}(D,A)}u(b)$, the item that maximizes her utility in the consideration set. Thus under the RC rule the probability of choosing $a \in A$ is given by $p_{RC}(a, A) = \sum_{D\cap A\neq\emptyset:D\in\mathcal{D}} \mathbb{I}(u(a) > u(b)\forall b \in (D \cap A)\backslash\{a\})m(D)$, where $\mathbb{I}(u(a) > u(b)\forall b \in (D \cap A)\backslash\{a\})$ is equal to 1 if the condition is true and is equal to zero if not. Alternatively, we can write $p_{RC}(a, A) = \sum\limits_{\{a\}\cap D\neq\emptyset;\mathbf{B}_A(a)\cap D=\emptyset:D\in\mathcal{D}} m(D)$ where $\mathbf{B}_A(a) = \{b \in A : u(b) > u(a)\}$ is the set of better than a elements in the menu A.

Definition 1. (Random Categorization rule -RC) A stochastic choice dataset has a Random Categorization rule representation if there is a triple u, m and $\mathcal{D}$ that are the injective utility function, the probability of consideration map, and the mental categories respectively, such that the probability of choosing $a \in X$ in a menu $A \in \mathcal{M}$, is the cumulative probability of all categories that produce a consideration set where $a \in A$ is the best element available:

$$p_{RC}(a, A) = \sum_{\{a\}\cap D\neq\emptyset;\mathbf{B}_A(a)\cap D=\emptyset:D\in\mathcal{D}} m(D).$$

Finally, the probability to choose the default is $p_{RC}(o, A) = \sum_{A\cap D=\emptyset} m(D)$.

Summarizing, the probability to choose $a \in A$ under the RC rule $p_{RC}(a, A)$ is equivalent to the probability that the DM considers $a \in A$ but does not considers any alternative that is better than it. By definition $\sum_{a\in A} p_{RC}(a, A) + p_{RC}(o, A) = 1$.

There are two important special cases of the RC rule, namely the standard rational model (without indifference) and the MM model of consideration sets with menu independence.

Example 1. (Standard Rational DM) A standard rational DM has probability of choice $p_{SR}(a, A) = \mathbb{I}(u(a) > u(b)\forall b \in A)$, for an injective utility function $u : X \mapsto \mathbb{R}$ and the indicator function $\mathbb{I}(\cdot)$. Clearly, this is a RC rule with categories $\mathcal{D} = \{X\}$ with probability $m(X) = 1$ and the with same utility function u.

The MM model is also a special case of the RC rule.

Example 2. (MM stochastic consideration with menu independence) The MM model of consideration set consists of an attention parameter $\gamma : X \mapsto (0,1)$ and a utility function $u : X \mapsto \mathbb{R}$ such that $p_{MM}(a, A) = \gamma(a) \prod_{b \in A: u(b) > u(a)} (1 - \gamma(b))$. In this case, the categories are comprised of all possible subsets of X including the empty set (which has positive probability), $\mathcal{D} = 2^X$ and the probability of consideration of the categories is $m(D) = \sum_{A \subseteq D} (-1)^{|D \setminus A|} (\prod_{a \in X \setminus A} (1 - \gamma(a)))$, which the reader can verify generates $p_{RC}(a, A) = \sum_{\{a\} \cap D \neq \emptyset; \mathbf{B}_A(a) \cap D = \emptyset : D \in \mathcal{D}} m(D) = \gamma(a) \prod_{b \in A: u(b) > u(a)} (1 - \gamma(b))$. The fact that γ is non-degenerate implies that the support of m is the whole power set 2^X, or alternatively that the categories include all elements of the power set. Further discussion of the relation of the RC rule and MM is provided in the sequel.

Of course, the RC rule, allows for probabilistic datasets that cannot be accommodated by neither the rational model or the MM model.

Example 3. (Categorization) Consider the choice set $X = \{a, b, c, d\}$, such that $u(c) > u(a) > u(b) > u(d)$ with categories$\mathcal{D} = \{\{a, c, b\}, \{a\}, \{b, d\}\}$. We let the map $m : \mathcal{D} \mapsto (0,1)$ be any non-degenerate probability over the categories, for all $D \in \mathcal{D}$. Then we give the DM the menus $A = \{a, c, b\}$, $B = \{a, d, b\}$, we have: (i) $p_{RC}(a, A \setminus \{b\}) / p_{RC}(a, A) = m(\{a\}) / m(\{a\})$, (ii) $p_{RC}(a, B \setminus \{b\}) = m(\{a, c, b\}) + m(\{a\})$, and (iii) $p_{RC}(a, B) = m(\{a\})$. Finally, (i), (ii) and (iii) imply that $\frac{p_{RC}(a, A \setminus \{b\})}{p_{RC}(a, A)} < \frac{p_{RC}(a, B \setminus \{b\})}{p_{RC}(a, B)}$, when $m(D) > 0$ for all $D \in \mathcal{D}$, so i-Independence, that is a necessary condition for MM, that requires that $\frac{p_{RC}(a, A \setminus \{b\})}{p_{RC}(a, A)} = \frac{p_{RC}(a, B \setminus \{b\})}{p_{RC}(a, B)}$ is violated. Clearly, this model cannot be generated by a standard rational model (without indifference) either.

Finally, we provide an example where the RC rule exhibits the *similarity effect* that is a violation of IIA. The similarity effect, seem to be a well-established empirical fact (Trueblood et al., 2013). This effect is usually called a “context effect” in the psychological literature and it has inspired an important literature on random choice with attributes (for a nice discussion see (Trueblood et al., 2014; Swait et al., 2014)). The similarity effect is the observation that the probability of choosing an alternative decreases when another dominating object is introduced to the menu (this is Tversky’s (1972) similarity hypothesis). Formally:

Definition 2. (Similarity effect) For any $a, b, c \in X$ and any $B \in \mathcal{M}$ we observe $\frac{p(b, B \cup \{a\})}{p(c, B \cup \{a\})} < \frac{p(b, B)}{p(c, B)}$.

Example 4. (Similarity Effect) Consider a set of political candidates $X = \{a, b, c\}$ with $\mathcal{D} = \{\{a, b\}, \{b, c\}, \{c\}\}$ where $\{a, b\}$ is the category of candidates who are from region I, the group of candidates who belong to the ethnic group II $\{b, c\}$, and$\{c\}$ is the group of candidates who are of religion III. All members of the electorate have the following preferences over candidates (quality) $u(a) > u(b) > u(c)$; however the are are restricted in their voting by their regional, ethnic and religious affiliations. The proportion of voters who belong to ethnic group II is $m(\{b, c\}) = \frac{4}{6}$, while the proportion of voters who belong to region I and religion III are $m(\{a, b\}) = m(\{c\}) = \frac{1}{6}$ respectively. This is an RC rule because, even though the voters care about the candidates quality, they have subjective restrictions. The probability of choosing candidates of groups I and III when the group’s II candidate is not participating in the race is $p_{RC}(a, \{a, c\}) = \frac{1}{6}$ and $p_{RC}(c, \{a, c\}) = \frac{5}{6}$. The probability of choosing candidate a when all three candidates are running for office remains the same $p_{RC}(a, X) = \frac{1}{6}$ while the probability of choosing group's III candidate decreases significantly as a result of the presence of the dominating candidate from group II $p_{RC}(c, X) = \frac{1}{6}$, recall that $u(b) > u(c)$. This example exhibits the similarity effect because $\frac{p_{RC}(c,X)}{p_{RC}(a,X)} = 1$ while $\frac{p_{RC}(c,\{a,c\})}{p_{RC}(a,\{a,c\})} = 5$.

4 Characterization and Recoverability

4.1 Characterization

Let us begin by defining a revealed preference relation in the context of stochastic choice that will serve as the basis of our characterization.

Definition 3. (Stochastic revealed preference $\succ$) $\succ \subset X \times X$ is defined as $a \succ b$ if and only if $p(b, A \cup \{a\}) \neq p(b, A)$.

In words, an item a is **stochastically revealed preferred to** b if and only if a has a non-zero impact on the probability to choose b in any menu. The relation $\succ$ captures the idea that introducing a new alternative a in a menu has some form of psychological higher status for her than b.

Now, we impose a condition on the revealed stochastic preference relation $\succ$, in the same spirit as the Strong Axiom of Revealed Preference (SARP).

Axiom 1. *(Acyclicity) $\succ$ on X is acyclic. i.e. there exists no $a_1, \cdots, a_n \in X$ such that $a_i \succ a_{i+1}$ for $i = 1, \cdots n-1$ and $a_n \succ a_1$.*

The relation $\succ$ is asymmetric but may not be total. This is a standard internal consistency requirement in revealed preference relations.

The second axiom that characterizes the RC rule is **Weakly Decreasing Marginal Propensity (WDMP) of Choice** Axiom.

Before stating this axiom we first need to define the successive differences of any probability of choice.

Definition 4. (Successive differences) Successive differences for $p(a, A)$ for a fixed $a \in A \cup \{o\}$ and $A \in \mathcal{M}$. For the probability $p(a, A)$, define recursively:

$\Delta_{A_1} p(a, A) = p(a, A) - p(a, A \cup A_1)$ for $A, A_1 \in \mathcal{M}$,

$\Delta_{A_n} \cdots \Delta_{A_1} p(a, A) = \Delta_{A_{n-1}} \cdots \Delta_{A_1} p(a, A) - \Delta_{A_{n-1}} \cdots \Delta_{A_1} p(a, A \cup A_n)$ for all $n \geq 2$ and for all $A, A_1 \cdots A_n \in \mathcal{M}$.

We are ready to state the WDMP axiom.

Axiom 2. *(Weakly decreasing marginal propensity of choice -WDMP-.) For all $A \in \mathcal{M}$, and any $\{A_i\}_{i=1}^n \in \mathcal{M}^n$ the successive differences are non-negative $\Delta_{A_n} \cdots \Delta_{A_1} p(o, A) \geq 0$ for all $n \geq 1$.*

Now, we describe the WDMP condition. The first difference $-\Delta_{A_1} p(o, A) = p(o, A \cup A_1) - p(o, A)$ measures the direct impact of adding A_1 to a menu A on the probability of choosing the default o; this impact is required to be non-positive. This condition is usually known as **regularity** or monotonicity because this is equivalent to saying that $p(o, A) \geq p(o, B)$ for $A \subseteq B$. Moreover, the marginal impact of adding A_1 to a menu that contains A_2 (i.e. $-\Delta_{A_2} \Delta_{A_1} p(o, A) =$

$\Delta_{A_1} p(o, A \cup A_1) - \Delta_{A_1} p(o, A))$ is also non-positive and not weaker in magnitude than the direct impact.

This axiom is equivalent to Block-Marschak (1960) polynomials non-negativity condition (equivalently Total Monotonicity) only for the case of the default alternative probability of choice.[2] We are ready to state our main result.

Theorem 1. *A complete stochastic choice dataset admits a Random Categorization rule (RC) representation if and only if it satisfies Axiom (1) -Acyclicity- and Axiom (2) -WDMP-.*

It turns out, that the WDMP and the Acyclicity condition imply the Total Monotonicity condition/Block-Marschak regularity for all alternatives $a \in X \cup \{o\}$.[3] This fact, makes the RC rule a special case of RUM.

Corollary 1. *A complete stochastic choice dataset admits that admits a RC representation also admits a RUM representation.*

Remark. We believe that the RC rule has one important advantage with respect to the more general RUM, that is its improved recoverability properties that will be apparent in the sequel and its compatibility with a fundamental cognitive process in decision making such as categorization.

4.2 Recoverability

The results set out hereunder are derived trivially from theorem (1).

Corollary 2. *In a complete stochastic choice dataset generated by the RC rule, the categories $\mathcal{D} \subseteq 2^X$ and the probability of consideration $m : \mathcal{D} \mapsto [0, 1]$ are recovered uniquely.*

The uniqueness is a consequence of defining m as a Mobius inverse that is always unique, as is detailed in the last part of the proof of theorem (1).

Corollary 3. *In a complete stochastic choice dataset generated by the RC rule, we can recover a set of injective utility functions $\mathcal{U}$ that represent the each element of the transitive closure of the relation $\succ$.*

Now we establish the sufficient condition for the recoverability of the injective utility up to a monotone transformation.

Axiom 4. *$\succ$-Totality. $\succ$ is total if and only if for any $a, b \in X$ $a \succ b$ or $b \succ a$.*

Recall that $\succ$ is a stochastic revealed preference relation on the finite set X, thus $\succ$-totality is a testable restriction on the stochastic choice. Totality means that all items are related among them by the psychological precedence captured by $\succ$.

Corollary 4. *In a complete stochastic choice dataset generated by the RC rule, that satisfies Axiom (4) we can recover an injective utility function u up to a monotone transformation (i.e., if the data admits an RC $(u, m, \mathcal{D})$ representation it also admits an RC $(f \circ u, m, \mathcal{D})$ representation where $f : \mathbb{R} \mapsto \mathbb{R}$ is a monotone function).*

[2] The axiom is stated using the quantifier $\forall n \geq 1$, for the sake of generality because the representation can be readily extended to countable choice sets. Note that in contrast with Block-Marschak (1960) polynomials, these differences are well defined for some cases where X is infinite (e.g. when it is a locally compact Hausdorff second countable space like $\mathbb{R}^d$).

[3]

Axiom 3. *(Total Monotonicity) For all $a \in X \cup \{o\}, A \in \mathcal{M}$, such that $a \in A$ and any $\{A_i\}_{i=1}^n \in \mathcal{M}^n$ the successive differences are non-negative $\Delta_{A_n} \cdots \Delta_{A_1} p(a, A) \geq 0$ for all $n \geq 1$.*

5 Beyond Random Utility: Generalized Random Categorization

Here we study a generalization of the RC rule that is characterized uniquely by the acyclicity of the stochastic revealed preference relation $\succ$ (Axiom 1) and is a different model from RUM. This generalization is of interest because it allows for violations of regularity. The main change is that we allow the DM to have a signed measure defined over her mental categories instead of a probability distribution m discussed above. We say that $\mu : \mathcal{D} \mapsto \mathbb{R}$ is a signed measure defined over mental categories if and only if $\sum_{D\in\mathcal{D}} \mu(D) = 1$. We interpret $\mu(D) > 0$ as a propensity to consider a category while $\mu(D) \leq 0$ measures the avoidance of a category.

Definition 5. (General Random Categorization rule -GRC-) A stochastic choice dataset has a General Random Categorization rule representation if there is a triple u, μ and $\mathcal{D}$ that are the injective utility function, the signed measure of consideration map, and the mental categories respectively, such that the probability of choosing $a \in X$ in a menu $A \in \mathcal{M}$, is the sum of the propensities of all categories that produce a consideration set where $a \in A$ is the best element available:

$$p_{GRC}(a, A) = \sum_{\{a\}\cap D\neq\emptyset; \mathbf{B}_A(a)\cap D=\emptyset: D\in\mathcal{D}} \mu(D).$$

Finally, the probability to choose the default is $p_{GRC}(o, A) = \sum_{A\cap D=\emptyset} \mu(D)$.

Summarizing, under the GRC rule the probability of choosing an item is sum of propensities of mental categories where the fixed item is the best available alternative.

Theorem 2. *A stochastic random choice dataset can be generated by a General Random Categorization rule (GRC) if and only if it satisfies Axiom (1).*

Surprisingly the recoverability properties of the GRC model are the same as the RC model, so we refer to the reader to the previous section in regard to this matter.

The model is able to accommodate violations of the regularity condition, or the restriction that $p(a, A \cup \{a, b\}) \leq p(a, A \cup \{a\})$ for all $a, b \in X$ and all $A \in \mathcal{M}$. This is a necessary condition for the RUM.

Example 5. (Regularity Violations) The following DM faces the following set of alternatives: three cars (models) $X = \{a, b, c\}$. The first car, a, has two attributes: it has the positive attribute of a low price and the negative attribute of high fuel consumption. The second car, b, has high performance but also shares with a the high fuel consumption. Finally, c is consider low fuel consumption. The corresponding categories associated with this attributes are $\mathcal{D} = \{\{a\}, \{b\}, \{a, b\}, \{c\}\}$ with the signed-measure $\mu(\{a\}) = \frac{2}{3}$, $\mu(\{a, b\}) = -\frac{1}{3}$, $\mu(\{b\}) = \frac{1}{3}$, $\mu(\{c\}) = \frac{1}{3}$ which correspond to the propensities both positive and negative of each mental category. We assume that $u(a) > u(b) > u(c)$; this reflects preferences over the car models themselves. We first observe the probability of choosing something different from the default: $p_{GRC}(a, \{a\}) = \frac{1}{3}$, $p_{GRC}(b, \{b\}) = 0$, $p_{GRC}(c, \{c\}) = \frac{1}{3}$. The case of the second alternative has a probability of zero because the DM prefers not to buy a car or to buy the default alternative when faced with only b. When the DM faces $\{a, b\}$ as alternatives, then the probability of choosing b increases. This happens in the GRC model because of the interaction of the utility and the categories weights, since $u(a) > u(b)$ it receives in the mind of the DM the weight of being the high fuel consumption car when making the choice. Thus a distracts from b the bad attribute because it is in the top of DM's mind. Formally, $p_{GRC}(b, \{a, b\}) = \frac{1}{3}$ while $p_{GRC}(a, \{a, b\}) = \frac{1}{3}$, and so we have $p_{GRC}(b, \{a, b\}) > p_{GRC}(b, \{b\})$ which is a violation of regularity. This implies that this model cannot be accommodated by a RUM.

6 Relation with Models of Stochastic Consideration Sets

The present work has explored a new model of random choice generated by a randomization process over categories. This produces a random consideration set that is then combined with a preference relation to produce a choice; in that sense, it generalizes the seminal work of Manzini and Mariotti (2014). In fact, the RC model nests the random consideration set of MM. Also, this model is related to the recent work of Brady and Rehbeck (2014); however the RC is not nested in their contribution. The RC model is nested in the Random Utility representation because it implies Total Monotonicity. The RC rule shares some similarities with the "Elimination by Aspects" choice rule described in Tversky (1972) but it is not nested in such model because RC can be shown to fail weak stochastic transitivity. This feature makes the RC rule able to accommodate the match-up effect.

The work of Fudenberg et al. (2013) proposes a tractable model of random utility based on optimization of expected utility that is perturbed by a non-linear additive term; again the RC model is not nested in their model because of the failure of stochastic transitivity. The work (Perception Adjusted Luce Model PALM) by Echenique, Saito and Tserenjigmid (2013) differs from the RC framework in that PALM concerns itself with intensities of random utility rather than random consideration sets. They also have a form of acyclicity condition on a revealed priority of perception relation (a hazard rate) however the RC model is not nested in PALM and vice-verse because RC allows for degenerate probabilistic choice rule (beyond the outside option). Moreover, RC does not nest the Luce model but PALM does. Recently, Zhang (2016) proposes an empirically different categorization procedure that produces stochastic choice. The author establishes that his rule is different from the previous versions of the categorization rule presented in this paper (Aguiar, 2015a,b). The reader ought to compare the similarities.

Manzini and Mariotti (2014) Stochastic Consideration Sets

The RC rule nests the MM model. The reader can quickly check that MM axioms imply our conditions but ours do not imply their axioms. The main differences is the possibility of accommodating violations of the i-Independence axiom, that is a form of menu independence consideration.

Definition 6. (i-Independence) For all $a, b \in X$ and all $A, B \in \mathcal{M}$: $\frac{p(a,A\setminus\{b\})}{p(a,A)} = \frac{p(a,B\setminus\{b\})}{p(a,B)}$ and $\frac{p(o,A\setminus\{b\})}{p(o,A)} = \frac{p(o,B\setminus\{b\})}{p(o,B)}$.

Notice that i-Asymmetry and i-Indepence in MM imply Axiom (2) and Axiom (1) but the converse is not true, so MM model is nested in RC.

Menu Dependence and Random Feasibility

Interestingly, another model that nests MM, provided by Brady and Rehbeck (2014), the Menu Dependent Stochastic Consideration -MDSC-, (also renamed as Random Feasibility) (Brady & Rehbeck, 2014, 2015) is not nested in RC nor is RC nested in their model. Evidently, the intersection of their contribution and the RC rule is non-empty. The axiom of MDSC that can be shown to fail in general in our set-up is as follows:

Definition 7. Axiom ASI (Asymmetric Sequential Independence): For all distinct $a, b \in X$, exactly one of the following holds.

$$p(a, \{a, b\}) = p(a, \{a\})p(\{a, o\}, \{a, b\}) \text{ or } p(b, \{a, b\}) = p(b, \{b\})p(\{b, o\}, \{a, b\}).$$

Claim 1. RC is not nested in the MDSC.

7 Conclusion

We have presented and characterized a new model of stochastic choice that deals with a DM with bounded rationality. Otherwise rational and deterministic choices are mediated by a random categorizing procedure. We have provided (necessary and sufficient) conditions under which an observer can infer standard rational preferences from a standard stochastic choice dataset and also recover uniquely the random categorization devices. This new model of random choice nests the deterministic rational choice model and the MM random consideration set model.

References

Agranov, M. & Ortoleva, P. (2014). *Stochastic Choice and Preferences for Randomization*. Technical report, mimeo.

Aguiar, V. (2015a). Stochastic Choice and Attention Capacities: Inferring Preferences from Psychological Biases. *Available at SSRN 2607602*.

Aguiar, V. (2015b). Stochastic choice and attention capacities: Inferring preferences from psychological biases. *Available at SSRN 2607602*.

Block, H. D. & Marschak, J. (1960). Random orderings and stochastic theories of responses. *Contributions to probability and statistics*, 2, 97–132.

Brady, R. & Rehbeck, J. (2014). Menu-dependent stochastic consideration.

Brady, R. & Rehbeck, J. (2015). Menu-Dependent Stochastic Feasibility. *Mimeo*.

Chateauneuf, A. & Jaffray, J.-Y. (1989). Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion. *Mathematical social sciences*, 17(3), 263–283.

Echenique, F., Saito, K., & Tserenjigmid, G. (2013). *The Perception-Adjusted Luce Model*. Technical report, Discussion paper.

Falmagne, J. C. (1978). A representation theorem for finite random scale systems. *Journal of Mathematical Psychology*, 18(1), 52–72.

Fudenberg, D., Iijima, R., & Strzalecki, T. (2013). Stochastic choice and revealed perturbed utility.

Manzini, P. & Mariotti, M. (2012). Categorize then choose: Boundedly rational choice and welfare. *Journal of the European Economic Association*, 10(5), 1141–1165.

Manzini, P. & Mariotti, M. (2014). Stochastic choice and consideration sets. *Econometrica*, 82(3), 1153–1176.

Swait, J., Brigden, N., & Johnson, R. D. (2014). Categories shape preferences: A model of taste heterogeneity arising from categorization of alternatives. *Journal of Choice Modelling*.

Trueblood, J. S., Brown, S. D., & Heathcote, A. (2014). The multiattribute linear ballistic accumulator model of context effects in multialternative choice. *Psychological review*, 121(2), 179.

Trueblood, J. S., Brown, S. D., Heathcote, A., & Busemeyer, J. R. (2013). Not just for consumers context effects are fundamental to decision making. *Psychological science*, 24(6), 901–908.

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

Zhang, J. (2016). Stochastic Choice with Subjective Categorization. *Available at SSRN*.

8 Appendix

8.1 Proof of Theorem (1)

Before stating the proof of Theorem (1) we need the following lemmata and some preliminaries.

Definition 8. (A Random Consideration Set and Hitting Functional) A Random Set $\mathbf{\Gamma}(X)$ is a random mapping that takes values on the set of categories $\mathcal{D}$ defined on X with probability of having a non-empty intersection with any given menu $A \in \mathcal{M}$ is given by $T(A) = \mathbf{P}(\mathbf{\Gamma}(X) \cap A \neq \emptyset) = \sum_{D \cap A \neq \emptyset : D \in \mathcal{D}} m(D)$ (we call $T : 2^X \mapsto [0,1]$ the hitting functional).

Lemma 1. *A RC rule can be equivalently represented by Random Set Hitting Functional, where* $p_{RC}(a, A) = \mathbf{P}(\mathbf{\Gamma}(X) \cap \{a\} \neq \emptyset; \mathbf{\Gamma}(X) \cap \mathbf{B}_A(a) = \emptyset)$*, where* $\mathbf{P}(\mathbf{\Gamma}(X) \cap \{a\} \neq \emptyset; \mathbf{\Gamma}(X) \cap \mathbf{B}_A(a) = \emptyset) = T(\mathbf{B}_A(a) \cup \{a\}) - T(\mathbf{B}_A(a))$ *and* $p_{RC}(o, A) = \mathbf{P}(\mathbf{\Gamma}(X) \cap A = \emptyset) = 1 - T(A)$.

Proof. This is done by direct computation, notice that:

$T(\mathbf{B}_A(a) \cup \{a\}) - T(\mathbf{B}_A(a)) = \sum_{(\mathbf{B}_A(a) \cup \{a\}) \cap D \neq \emptyset} m(D) - \sum_{\mathbf{B}_A(a) \cap D \neq \emptyset} m(D)$, thus

$T(\mathbf{B}_A(a) \cup \{a\}) - T(\mathbf{B}_A(a)) = \sum_{\{a\} \cap D \neq \emptyset; \mathbf{B}_A(a) \cap D = \emptyset : D \in \mathcal{D}} m(D)$, because $T(A) = \sum_{D \cap A \neq \emptyset : D \in \mathcal{D}} m(D)$.

□

Lemma 2. *The Hitting Functional T is totally monotone, such that for any $A \in \mathcal{M}$ and any $\{A_i\}_{i=1}^n \in \mathcal{M}^n$ the successive differences are non-positive $\Delta_{A_n} \cdots \Delta_{A_1} T(A) \leq 0$ for all $n \geq 1$. Moreover, its first difference $\Delta_B T(A)$ for any $B \in \mathcal{M}$ is such that for any $\{B_i\}_{i=1}^n \in \mathcal{M}^n$ the successive differences $\Delta_{B_n} \cdots \Delta_{B_1} \Delta_B T(A) \leq 0$ for all $n \geq 1$.*

Proof. Notice that $\Delta_{A_n} \cdots \Delta_{A_1} T(A) = -\mathbf{P}(\mathbf{\Gamma}(X) \cap A \neq \emptyset; \mathbf{\Gamma}(X) \cap A_i = \emptyset \forall i = 1, \cdots, n) \leq 0$, and $\Delta_{B_n} \cdots \Delta_{B_1} \Delta_B T(A) = -\mathbf{P}(\mathbf{\Gamma}(X) \cap A \neq \emptyset; \mathbf{\Gamma}(X) \cap B = \emptyset; \mathbf{\Gamma}(X) \cap B_i = \emptyset \forall i = 1, \cdots, n) \leq 0$. □

Now we are ready to prove Theorem (1).

Proof. If a stochastic choice dataset is generated by a RC rule then by lemma (1) we know that there is a random set $\mathbf{\Gamma}(X)$ defined on 2^X with hitting functional such that $p_{RC}(a, A) = T(\mathbf{B}_A(a) \cup \{a\}) - T(\mathbf{B}_A(a)) = -\Delta_{\{a\}} T(\mathbf{B}_A(a))$ and $p_{RC}(o, A) = 1 - T(A)$. Then by lemma (2) it follows that for a fixed $a \in X$, and any $A \in \mathcal{M}$ such that $a \in A$, $\Delta_{A_{n-1}} \cdots \Delta_{A_1} \Delta_{\{a\}} p_{RC}(a, A) = \Delta_{A_{n-1}} \cdots \Delta_{A_1} \Delta_{\{a\}} T(\mathbf{B}_A(a)) \geq 0$ for all $\{A_i\}_{i=1}^n \in \mathcal{M}^n$ for all $n \geq 1$, also $\Delta_{A_{n-1}} \cdots \Delta_{A_1} p_{RC}(o, A) = \Delta_{A_{n-1}} \cdots \Delta_{A_1} T(A) \geq 0$ for all $\{A_i\}_{i=1}^n \in \mathcal{M}^n$ for all $n \geq 1$. This is exactly Axiom (2) or WDMP. Now, we deal with $\succ$ Acyclicity or Axiom (1) necessity. Notice that if we observe $b \succ a$ that is $p_{RC}(a, A \cup \{b\}) < p_{RC}(a, A)$ it means that $T(\mathbf{B}_{A \cup \{b\}}(a) \cup \{a\}) - T(\mathbf{B}_{A \cup \{b\}}(a)) < T(\mathbf{B}_A(a) \cup \{a\}) - T(\mathbf{B}_A(a))$ this can only be true if $b \in \mathbf{B}_{A \cup \{b\}}(a)$ or equivalently $u(b) > u(a)$. This means that if we observe $b \succ a$ this implies that $u(b) > u(a)$, thus further implying that $\succ$ is acyclic, because u is an injective utility (no indifference).

If a stochastic choice dataset satisfies Axiom (1), namely acyclicity of $\succ$ and Axiom (2) or the WDMP condition then we can build a RC rule that generates the dataset. The main part of the rest of the proof is to use Axiom (2) to set the candidate hitting functional $\varphi(A) = 1 - p(o, A)$ and to notice that by this axiom $\varphi : 2^X \mapsto [0,1]$ is a totally monotone Choquet capacity by definition. In particular, for any $A \in \mathcal{M}$ and any $\{A_i\}_{i=1}^n \in \mathcal{M}^n$ the successive differences are non-positive $\Delta_{A_n} \cdots \Delta_{A_1} \varphi(A) \leq 0$ for all $n \geq 1$. Set the value of $\varphi(\emptyset) = 0$ without loss of generality. This is the candidate capacity to generate the stochastic choice dataset. Using

Axiom (1) I build the revealed upper contour set or the transitive closure of $\succ$ (call it $\succ^*$) that is guaranteed to exists, pick any of these as the preference order, for all $a \in X$: $\mathbf{B}_A^{\succ^*}(a) = \{b \in A | b \succ^* a\}$. Now, I have to check if $p(a, A) = \varphi(\mathbf{B}_A^{\succ^*}(a) \cup \{a\}) - \varphi(\mathbf{B}_A^{\succ^*}(a))$. By Axiom (1), this is indeed the case. Notice, that we can set $p(a, A) = p(o, \mathbf{B}_A^{\succ^*}(a)) - p(o, \mathbf{B}_A^{\succ^*}(a) \cup \{a\})$ which provides the desired result. To see this is true, observe that by definition of $\mathbf{B}_A^{\succ^*}(a)$, we have $p(\mathbf{B}_A^{\succ^*}(a), \mathbf{B}_A^{\succ^*}(a) \cup \{a\}) = p(\mathbf{B}_A^{\succ^*}(a), \mathbf{B}_A^{\succ^*}(a))$. Assume not, then by Axiom (1) it must be the case that $p(\mathbf{B}_A^{\succ^*}(a), \mathbf{B}_A^{\succ^*}(a)) \neq p(\mathbf{B}_A^{\succ^*}(a), [\mathbf{B}_A^{\succ^*}(a) \cup \{a\}])$ but in that case this means that there is an element $b \in \mathbf{B}_A^{\succ^*}(a)$ such that $a \succ^* b$ which is a contradiction of Axiom (1). Then, $p(a, [\mathbf{B}_A^{\succ^*}(a) \cup \{a\}]) = p([\mathbf{B}_A^{\succ^*}(a) \cup \{a\}], [\mathbf{B}_A^{\succ^*}(a) \cup \{a\}]) - p(\mathbf{B}_A^{\succ^*}(a), \mathbf{B}_A^{\succ^*}(a))$ and $p(a, A) = p(a, [\mathbf{B}_A^{\succ^*}(a) \cup \{a\}])$, and by definition $p(\mathbf{B}_A^{\succ^*}(a), \mathbf{B}_A^{\succ^*}(a)) = 1 - p(o, \mathbf{B}_A^{\succ^*}(a))$. Thus we have build a pair $(\varphi, \succ)$ that generates the data. The final step is to construct the RC rule from the pair $(\varphi, \succ^*)$. First we construct the utility function $u : X \mapsto \mathbb{R}$, the transitive closure of $\succ$, that we called $\succ^*$ is a linear order as such it can be represented by an injective utility. We enumerate the elements of X according to $\succ^*$ such that $X = \{a_i\}_{i=1}^n$ with $a_i \succ^* a_{i+1}$ for all $i = 1, \cdots, n-1$, then let $u(a_i) = i$. Now we build the propensity of consideration map, to do so we use the dual Mobius inverse of φ, in fact we propose $m(D) = \sum_{A \subseteq D : D \in \mathcal{M}} (-1)^{|D \setminus A|}(1 - \varphi(X \setminus A))$ as the candidate, to verify that this is indeed a probability we use Axiom (2) such that φ is totally monotone, thus by Chateauneuf & Jaffray (1989) it is the case that $m(D) \geq 0$ and $\sum_{D \in \mathcal{M}} m(D) = 1$. Observe that m is defined uniquely in terms of the dataset and φ, because it is the dual Mobius inverse of φ. Finally we set the categories collection as the support of m, $\mathcal{D} = \{A \in \mathcal{M} : m(A) > 0\}$. We end the proof by noticing that $p(a, A) = \sum_{\{a\} \cap D \neq \emptyset ; \mathbf{B}_A(a) \cap D = \emptyset : D \in \mathcal{D}} m(D)$ for $\mathbf{B}_A(a) = \{b \in A : u(b) > u(a)\}$ because $\sum_{\{a\} \cap D \neq \emptyset ; \mathbf{B}_A(a) \cap D = \emptyset : D \in \mathcal{D}} m(D) = \varphi(\mathbf{B}_A^{\succ^*}(a) \cup \{a\}) - \varphi(\mathbf{B}_A^{\succ^*}(a))$. □

8.2 Proof of Corollary (1)

Proof. We prove that a complete stochastic choice dataset generated by a RC rule satisfies total monotonicity. By lemma (2) it follows that for a fixed $a \in X$, and any $A \in \mathcal{M}$ such that $a \in A$, $\Delta_{A_{n-1}} \cdots \Delta_{A_1} \Delta_{\{a\}} p_{RC}(a, A) = \Delta_{A_{n-1}} \cdots \Delta_{A_1} \Delta_{\{a\}} T(\mathbf{B}_A(a)) \geq 0$ for all $\{A_i\}_{i=1}^n \in \mathcal{M}^n$ for all $n \geq 1$, also $\Delta_{A_{n-1}} \cdots \Delta_{A_1} p_{RC}(o, A) = \Delta_{A_{n-1}} \cdots \Delta_{A_1} T(A) \geq 0$ for all $\{A_i\}_{i=1}^n \in \mathcal{M}^n$ for all $n \geq 1$. This is exactly Total Monotonicity. Finally because Total Monotonicity holds by Falmagne (1978) we conclude that the complete stochastic choice dataset admits a RUM representation. □

8.3 Proof of Theorem (2)

Before stating the proof we need a preliminary definitions and results.

Definition 9. (Attention capacity) An attention capacity is a map $\varphi : 2^X \mapsto [0, 1]$ such that: (i) (proper) $0 \leq \varphi(A) \leq 1$ for all $A \in 2^X$, (ii) (monotone) $\varphi(A) \leq \varphi(B)$ for all $A \subseteq B \in \mathcal{M}$.

The attention capacity is also known as a monotone Choquet set capacity. With this in hand we define the following probabilistic choice representation.

Definition 10. (Fuzzy Attention Model -FAM) It is a probabilistic choice rule $p_{\varphi, \succ^*}$, induced by a pair $(\varphi, \succ^*)$ where φ is an attention capacity and $\succ^*$ is a strict rational preference such that for any $a \in A$ $p_{\varphi, \succ^*}(a, A) = \varphi(\mathbf{B}_A^{\succ^*}(a) \cup \{a\}) - \varphi(\mathbf{B}_A^{\succ^*}(a))$ and $p_{\varphi, \succ^*}(o, A) = 1 - \varphi(A)$ (with $\mathbf{B}_A^{\succ^*}(a) = \{b \in A : b \succ^* a\}$).

Lemma 3. *A stochastic choice dataset is generated by a FAM if and only if Axiom (1) holds.*

Proof. If the DM is using a FAM then it satisfies the Axiom (1) trivially. The main part of the rest of the proof is to set a candidate attention capacity as $\varphi(A) = 1 - p(o, A)$ that defines a mapping over the whole power set $\varphi : 2^X \mapsto [0, 1]$, this is a Choquet capacity by definition. Set the value of $\varphi(\emptyset) = 0$ without loss of generality. This is the candidate capacity to generate the stochastic choice dataset. Using Axiom (1) I build the revealed upper contour set or the transitive closure of $\succ$ (call it $\succ^*$) that is guaranteed to exists, pick any of these as the preference order, for all $a \in X$: $\mathbf{B}^{\succ^*}(a) = \{b \in X | b \succ^* a\}$. Now, I have to check if $p(a, A) = \varphi(\mathbf{B}_A^{\succ^*}(a) \cup \{a\}) - \varphi(\mathbf{B}_A^{\succ^*}(a))$. Now, using Axiom (1) we show this is the case. Notice, that we can set $p(a, A) = p(o, \mathbf{B}_A^{\succ^*}(a)) - p(o, \mathbf{B}_A^{\succ^*}(a) \cup \{a\})$ which provides the desired result. To see this is true, observe that by definition of $\mathbf{B}^{\succ^*}(a)$, we have $p(\mathbf{B}_A^{\succ^*}(a), \mathbf{B}_A^{\succ^*}(a) \cup \{a\}) = p(\mathbf{B}_A^{\succ^*}(a), \mathbf{B}_A^{\succ^*}(a))$. Assume not, then by Axiom (1) it must be the case that $p(\mathbf{B}^{\succ^*}(a) \cap A, \mathbf{B}^{\succ^*}(a) \cap A) \neq p(\mathbf{B}^{\succ^*}(a) \cap A, [\mathbf{B}^{\succ^*}(a) \cup \{a\}] \cap A)$ but in that case this means that there is an element $b \in \mathbf{B}^{\succ^*}(a)$ such that $a \succ^* b$ which is a contradiction of Axiom (1). Then, $p(a, [\mathbf{B}^{\succ^*}(a) \cup \{a\}] \cap A) = p([\mathbf{B}^{\succ^*}(a) \cup \{a\}] \cap A, [\mathbf{B}^{\succ^*}(a) \cup \{a\}] \cap A) - p(\mathbf{B}^{\succ^*}(a) \cap A, \mathbf{B}^{\succ^*}(a) \cap A)$ and $p(a, A) = p(a, [\mathbf{B}^{\succ^*}(a) \cup \{a\}] \cap A)$, and by definition $p(\mathbf{B}^{\succ^*}(a) \cap A, \mathbf{B}^{\succ^*}(a) \cap A) = 1 - p(o, \mathbf{B}^{\succ^*}(a) \cap A)$. Finally, we notice that due to the probability axioms that imply that $p(a, A) \geq 0$ for all a, A we conclude that the candidate Choquet capacity is monotone thus being an attention capacity. We have build a FAM or a pair $(\varphi, \succ^*)$ that generates the data. □

Proof. We are going to prove that the following are equivalent.

(i) The complete stochastic choice dataset satisfies Axiom (1).

(ii) The complete stochastic choice dataset admits a FAM representation.

(iii) The complete stochastic choice dataset admits a GRC representation.

First, we prove that (i) iff (ii). This is established in lemma (3)

Second we prove that if (ii) holds then (iii) follows.

If a complete stochastic choice dataset admits a FAM $(\varphi, \succ^*)$ representation then we can build a GRC $(u, \mathcal{D}, m)$ representation by building a injective utility function $u : X \mapsto \mathbb{R}$ such that $u(a) > u(b) \iff a \succ^* b$. Then we build $\mu : 2^X \mapsto \mathbb{R}$ using the Mobius inversion such that $m(D) = \sum_{A \subseteq D : D \in \mathcal{M}} (-1)^{|D \setminus A|}(1 - \varphi(X \setminus A))$, we let $\mathcal{D} = \{D \in 2^X : \mu(D) \neq 0\}$. It is easy to check that $\varphi(A \cup \{b\}) - \varphi(A) = \sum_{b \cap D \neq \emptyset, A \cap D = \emptyset} \mu(D)$ for any A and b, thus establishing that $p_{\succ^*, \varphi}(a, A) = \sum_{a \cap D \neq \emptyset; \mathbf{B}_A^u(a) \cap D = \emptyset} \mu(D)$ for $\mathbf{B}_A^u(a) = \{b \in A | u(b) > u(a)\}$ can be generated by the constructed GRC rule. Finally, we prove that (iii) implies (ii). This follows immediately by noticing that we can build an attention capacity from a signed-measure over categories as follows: $\varphi(A) = \sum_{D \cap A \neq \emptyset} \mu(D)$. We can build $\succ^*$ on top of the utility u.

□

8.4 Proof of Claim (1)

Proof. I use the same setup that in example (2).

(i) Observe $p_{RC}(b, \{a, b\}) = m(\{b, d\})$, $p(b, \{b\}) = m(\{a, c, b\}) + m(\{b, d\})$ and $p_{RC}(\{b, o\}, \{a, b\}) = m(\{b, d\})$. (ii) $p_{RC}(a, \{a, b\}) = m(\{a\}) + m(\{a, c, b\})$, $p_{RC}(a, \{a\}) = m(\{a, c, b\}) + m(\{a\})$ and $p_{RC}(\{a, o\}, \{a, b\}) = m(\{a, c, b\}) + m(\{a\})$. (i) and (ii) imply that in general the RC rule does not satisfy the asymmetric sequential axiom ASI axiom. □