Random-Subset Voting

Guilherme Barros Correa de Amorim¹, Ana Paula Cabral Seixas Costa¹, Danielle Costa Moraes¹, Leandro Chaves Rego²

¹Department of Production Engineering, Federal University of Pernambuco, Av. Prof. Moraes Rego, 1235 Recife, PE 50.670-901, Brazil

²Department of Statistics and Applied Mathematics, Universidade Federal do Ceará, Bloco 910, Campus do Pici, 60455-760, Fortaleza, CE, Brazil

Correspondence should be addressed to guilherme.amorim@gmail.com

Journal of Artificial Societies and Social Simulation 21(1) 10, 2018

Doi: 10.18564/jasss.3610 Url: <http://jasss.soc.surrey.ac.uk/21/1/10.html>

Received: 13-01-2016

Accepted: 25-11-2017

Published: 31-01-2018

Abstract: Most of the voting procedures in the literature assume that voters have complete and transitive preferences over the alternatives. A basic premise is that any voter is able to evaluate any pair of alternatives in a set and select his/her preferred one, or indicate indifference between them. Nevertheless, some researchers have highlighted that voters, as humans, have limited capacity to deal with and consequently compare big sets of alternatives. In this paper, we propose the Random-Subset Voting, a voting procedure that through a random approach allows the voters to evaluate less alternatives. Instead of analyzing the entire set of alternatives, each voter will evaluate a random subset of a pre-determined size. We have proposed a theorem indicating that, for large sets of voters, the outcomes of traditional Borda and Random-Subset Borda converge. We have also implemented a web experiment and a Monte Carlo simulation in order to validate the proposed procedure and analyze how it behaves in several scenarios.

Keywords: Decision Theory, Voting Procedure, Random Subsets, Borda, Bounded Rationality

Introduction

- 1.1 In the last decades, the literature of decision theory has been facing a gradual shift from pure rational to fuzzy or probabilistic models of individual behavior. Simon (1955) and Allais (1953) mark important milestones in this trend and state the foundations of what was later referred to as “bounded rationality”.
- 1.2 For Simon (1955) and Allais (1953), the models that set the “economic man” as the ground for the developments in decision theory needed a drastic review in order to support the new findings of the psychologists, which demonstrated that in real situations the observed behavior of people substantially differed from the behavior preconized by such models.
- 1.3 Take for instance two core attributes of the preferences of rational individuals: completeness and transitivity. According to Arrow (1950, 1963), these attributes constitute the basic axioms of the individual preferences and the heart of the General Possibility Theorem.
- 1.4 Another attribute that is commonly presented as an important assumption for rational individual preferences is the independence of irrelevant alternatives (IIA). It states that if a voter v prefers alternatives x to y while choosing or ranking them from a choice-set A , adding or removing alternatives from A will not change the preference of v over x and y (Radner & Marschak 1954; Sen 1971)¹.
- 1.5 Conversely, some scholars have been diving into the realm of the non-rational, exposing that people sometimes fail to present transitive or complete preferences. (Fishburn 1991; Anand 1993; Camerer 1998; Nau 2006; Danan et al. 2006; Nurmi 2014). In addition, some researchers questions the validity of the IIA property (McFadden et al. 1977; Fry & Harris 1996; Cheng & Long 2007).
- 1.6 In this context, from a more practical point of view, some argue that people are not able to evaluate a large amount of data in order to take many decisions (Iyengar & Lepper 2000; Schwartz et al. 2002; Schwartz 2004). It is also verified that individuals can face an overwhelming effect when dealing with decisions with too many options (Scheibehenne et al. 2010).

- 1.7 This reasoning leads us to question whether the models we adopt in order to make social decisions take into consideration these non-rational aspects of human behavior. More specifically, we question whether the traditional voting systems consider the overwhelming effect individuals can face when large sets of alternatives have to be evaluated.
- 1.8 Even though some recent works have been focusing on this issue (Nurmi 2014), much work has to be done. In this paper, we propose a probabilistic voting procedure, called Random-Subset Voting, that addresses the choice overload issue through a random mechanism that reduces the number of alternatives. We propose a theorem indicating that, for large sets of voters, the outcomes of traditional Borda and Random-Subset Borda converge. We then evaluate the proposed procedure using a Monte Carlo simulation technique and a web experiment.
- 1.9 We organize this paper as follows: in Section 2, we present a brief review of the literature of probabilistic voting procedures; in Section 3, we formally present the Random-Subset Voting procedure – RSV – and include the mathematical modelling, followed by the simulations and the web experiment. In the last two sections, we discuss the results and present the conclusions of the paper.

Probabilistic Voting

- 2.1 The path taken in the last decades into the direction of less demanding assumptions with regard to the rationality of individuals has encouraged the development of models in which random or probabilistic features are embedded into the traditional deterministic schemes.
- 2.2 In fact, from a more theoretical perspective, one can show that a fundamental duality is on the basis of our democratic evolution. On one side, social choices purely made by chance with no mechanism of preference gathering. On the other side, social choices based on the preferences of the citizens with no room for any kind of sortition or fuzzy reasoning.²
- 2.3 Out of these extremes, a full range of mechanisms embeds both ideas and tries to find the proper combination for each decision.
- 2.4 Intriligator (1973, 1982) proposed a model in which the voters are not required to order (or select) alternatives but are asked to assign probabilities to each one indicating how prone he/she is to vote for each choice. These individual probability vectors are combined leading to the social probability vector. This vector is the input for the random mechanism that will generate the final social choice.
- 2.5 Zeckhauser (1969) proposed a method in which the set of alternatives is expanded. It included not only the alternatives themselves, but also lotteries between them. In some decisions, it might be suitable for the voters to opt for a lottery between alternatives a and b than taking alternative c for sure.
- 2.6 Barberá & Pattanaik (1986) argue that stochastic choice models are directly related to developments in psychology and model the preferences as distributions over orderings of alternatives. A function f maps values between 0 and 1 to each ordering w in the set W of all possible orderings.
- 2.7 In addition to these contributions, many authors have expanded the literature and reinforced the arguments in favor of probabilistic, or stochastic, models of social choice (Fishburn 1975; Fishburn & Gehrlein 1977; Gibbard 1977; Nitzan 1975; Barberá 1979; Coughlin & Nitzan 1981; Ehlers et al. 2002; Nurmi 2002; McKelvey & Patty 2006).
- 2.8 Inspired by Intriligator's idea, we propose a new three stage voting procedure. While the randomization step in Intriligator's process comes up in the third phase, after properly eliciting preferences for the alternatives, we set out a procedure in which the randomization takes place at the beginning, leaving to the voters the task of indicating their preferences from smaller sets of alternatives.
- 2.9 In the next section, we formalize the Random-Subset Voting procedure.

Random-Subset Voting (RSV)

- 3.1 Assume that a set V of n voters $\{v_1, v_2, v_3, \dots, v_n\}$ are required to indicate their preferences over m alternatives on a set $A = \{a_1, a_2, a_3, \dots, a_m\}$.
- 3.2 Random-Subset Voting states that instead of leaving the voters to indicate their preferences over A , each of them will vote from A_i , a random subset of A of size r . This means that, for each voter v_i , a mechanism will

generate the subset A_i by randomly picking r alternatives from A . We call RSVr, the Random-Subset Voting with r random alternatives. The voters will express their preferences over each random subset of alternatives.

3.3 In summary, the method is divided into three steps:

1. A fair random mechanism is used to generate n random subsets of r alternatives ($r < m$) and assign each of them to a voter v_i ;
2. Each voter v_i indicates his/her preferences over A_i ;
3. The individual preferences are aggregated and the result of the election is reached.

3.4 By fair, we mean that the alternatives will have the same probability of being chosen as elements of the random subsets.

3.5 In the following sub-section, we present the mathematical modeling of Random-Subset Borda.

Mathematical modeling

3.6 In this section, we state and prove a theorem showing that, under specific conditions, the final rankings of two elections, one using the Random-Subset Borda and one using Borda itself will be the same, with probability 1, for a large enough n .

3.7 Before going into the details, we provide some insights regarding the choice of Borda. Although any traditional voting rule might have been used in order to assess RSV, we have chosen Borda for the reasons that follows. First, in order to simplify the analytical modeling, with no risk of oversimplification, we have focused on one-stage procedures. Second, among the main one-stage procedures in the literature (Condorcet, Borda, plurality and approval), Borda and Condorcet are the procedures in which more information is required from the voter. Ordering a set of alternatives is more difficult than choosing one (plurality) or selecting some (approval). Therefore, as the number of alternatives increase in a social choice, the amount of information will increase rapidly and the task of evaluating them will become harder³ in comparison to other methods. Finally, we have chosen to focus on Borda, and not on Condorcet, because it allows choosing more than one alternative.

3.8 Based on Gehrlein & Fishburn (1978), we assume that large electorates are common and, then, focus on $n \rightarrow \infty$.

3.9 As stated in the last section, let A be the set of m alternatives and V be the set of n voters. The voters are considered to have preferences that are subject to transitivity, completeness and IIA.

3.10 Define $\mathcal{A}$ as the set of all possible complete orderings of A . Define $\mathcal{A}(j, k) \subseteq \mathcal{A}$ as the subset of in which the alternative a_j is in the k -th position in the ordering.

3.11 Define $p(x)$ as the fraction of voters whose preferences are described by the ordering $x \in \mathcal{A}$. So, considering the Borda method, in which the alternative in the k -th position gets $(m - k)$ points, the average score of alternative a_j is:

$$\overline{X}_j = \sum_{k=1}^{m-1} (m - k) \sum_{x \in \mathcal{A}(j, k)} p(x). \quad (1)$$

3.12 Now, consider the method Random-Subset Borda with r random alternatives ($2 \leq r < m$). The alternative in the q -th position in the randomly selected subset gets $(r - q)$ points.

3.13 Let $X_{i,j}^r$ be the random variable that indicates the score of the alternative a_j for the voter v_i . Note that, if $a_j \notin A_i$, then $X_{i,j}^r = 0$.

3.14 Define $\overline{X}_{n,j}^r$ as the random variable that indicates the average score of alternative a_j in the RSB election.

$$\overline{X}_{n,j}^r = \frac{1}{n} \sum_{i=1}^n X_{i,j}^r. \quad (2)$$

3.15 Lemma 1 proves that, if all voters satisfy independence of irrelevant alternatives (IIA), the expected average score of any alternative in RSB is equal to the average score in traditional Borda, except for a positive scale factor.

3.16 **Lemma 1:** Suppose that every voter in V satisfies transitivity, completeness and IIA and that $2 \leq r < m$. Thus, we have:

$$E(\overline{X}_{n,j}^r) = \frac{r(r-1)}{m(m-1)} \overline{X}_j.$$

3.17 Proof of Lemma 1:

Considering that alternative j can be in anyone of the m positions in the set A and that it can be in any of the r positions in the selected subset, it follows that:

$$E(\overline{X}_{n,j}^r) = \sum_{k=1}^{m-1} \left(\sum_{q=1}^{r-1} (r-q) \frac{\binom{k-1}{q-1} \binom{m-k}{r-q}}{\binom{m}{r}} \sum_{x \in \mathcal{A}(j,k)} p(x) \right) = \frac{1}{\binom{m}{r}} \sum_{k=1}^{m-1} \left(\sum_{q=1}^{r-1} (r-q) \binom{k-1}{q-1} \binom{m-k}{r-q} \sum_{x \in \mathcal{A}(j,k)} p(x) \right).$$

Thus, using the binomial coefficient definition and inverting the two innermost summations, we get:

$$\begin{aligned} E(\overline{X}_{n,j}^r) &= \frac{1}{\binom{m}{r}} \sum_{k=1}^{m-1} \left(\left(\sum_{x \in \mathcal{A}(j,k)} p(x) \right) \sum_{q=1}^{r-1} (m-k) \binom{k-1}{q-1} \binom{m-k-1}{r-q-1} \right) = \\ &= \frac{1}{\binom{m}{r}} \sum_{k=1}^{m-1} \left((m-k) \left(\sum_{x \in \mathcal{A}(j,k)} p(x) \right) \sum_{q=1}^{r-1} \binom{k-1}{q-1} \binom{m-k-1}{r-q-1} \right). \end{aligned}$$

Finally, using the definition of the hypergeometric distribution (Walpole et al. 2012) and Equation 1, we get:

$$E(\overline{X}_{n,j}^r) = \frac{\binom{m-2}{r-2}}{\binom{m}{r}} \sum_{k=1}^{m-1} \left((m-k) \left(\sum_{x \in \mathcal{A}(j,k)} p(x) \right) \right) = \frac{r(r-1)}{m(m-1)} \overline{X}_j.$$

3.18 Since the random variables $X_{i,j}^r$, with $i = 1, 2, \dots, n$, are independent and limited, the strong law of large numbers (Kolmogorov 1956) implies that:

$$\overline{X}_{n,j}^r \xrightarrow{n \rightarrow \infty} \frac{r(r-1)}{m(m-1)} \overline{X}_j. \quad (3)$$

3.19 Finally, we prove the theorem that states that, for a big enough electorate, with probability 1, the result of elections using Borda and Random-Subset Borda, with $2 \leq r < m$, are the same.

3.20 Theorem 1: Suppose that every voter in V satisfies transitivity, completeness and IIA and that $2 \leq r < m$. Thus, we have that:

$$\overline{X}_{j_1} > \overline{X}_{j_2} \iff P\left(\lim_{n \rightarrow \infty} \overline{X}_{n,j_1}^r > \lim_{n \rightarrow \infty} \overline{X}_{n,j_2}^r\right) = 1, \forall j_1, j_2 \in V.$$

3.21 Proof of Theorem 1: From Equation 3, we can affirm that

$$P\left(\lim_{n \rightarrow \infty} \overline{X}_{n,j_1}^r = \frac{r(r-1)}{m(m-1)} \overline{X}_{j_1}\right) = 1$$

and

$$P\left(\lim_{n \rightarrow \infty} \overline{X}_{n,j_2}^r = \frac{r(r-1)}{m(m-1)} \overline{X}_{j_2}\right) = 1.$$

Thus,

$$P\left(\lim_{n \rightarrow \infty} (\overline{X}_{n,j_1}^r - \overline{X}_{n,j_2}^r) = \frac{r(r-1)}{m(m-1)} (\overline{X}_{j_1} - \overline{X}_{j_2})\right) = 1.$$

3.22 In the next sub-section, we present how we have applied Monte Carlo techniques to simulate RSV in the context of the Borda voting rule.

Simulation

4.1 In the social choice field, many simulation approaches are found. Ludwin (1976) simulates a three-candidate single-seat election and compares voting methods. Bordley (1983) develops a general method for evaluating election schemes and uses it to compare six well-known voting schemes. Merrill III (1984) simulates elections in a randomly generated society and analyzes the Condorcet efficiency criteria for seven different methods. Bissey et al. (2004) presents a program to simulate preferences of voters and compare the representativeness and the

governability of eleven voting systems. Other important studies in the field are Chamberlin & Cohen (1978); Chamberlin (1986); Jones et al. (1995); Adams (1997); Johnson (1999); Gehrlein & Lepelley (2000); Kollman et al. (1992); Schumacher & Vis (2012).

- 4.2** In our approach, we use simulations to compare Borda and Random-Subset Borda (RSB). Even though we propose RSV as an alternative for dealing some already identified limits of human cognition in dealing with decisions, the simulation proposed assumes that voters have preferences that are complete, transitive and subject to IIA. The reason for this approach is primarily to assure that, when comparing voting systems, we have the same underlying premises supporting them. In other terms, the main point we want to evaluate is how RSB performs in comparison to Borda when the same premises regarding rationality are taken into account.
- 4.3** The simulation software was developed in Python and is publicly available at the CoMSES Net Computational Model Library⁴.
- 4.4** In order to facilitate the presentation, we propose the following definitions:
- *Configuration*: is represented by a tuple $(n; m)$ that indicates respectively the number of voters and the number of alternatives of a given simulation;
 - *Scenario*: is a randomly generated list of n rankings representing the preferences of the voters over m alternatives. For each configuration we can generate many different scenarios;
 - *RSBr*: Random-Subset Borda with r random alternatives;
 - *Unit Simulation*: for a given scenario, a unit simulation is the execution of Borda and RSBr. For each unit simulation, we set four parameters: number of voters (n), number of alternatives (m), number of random alternatives (r) and the distribution of preferences;
 - *Match index (MI)*: when comparing two rankings, the match index counts the number of matches between them. A full match ($MI = m$) happens when the number of matches is equal to the number of alternatives. On the other hand, when the number of matches count zero, we have minimal MI . This concept is as an adaptation of the Kemeny's distance (Kemeny 1959).
- 4.5** We divide this section into four subsections. In the first, we present how the random preferences for the voters were generated. In the second, we show the results of a single run of the simulation. In the third subsection, we present the results of the simulation for several runs and scenarios. Finally, in the last subsection, we identify thresholds of convergence between Borda and RSB.

Random preferences generation

- 4.6** In the context of modeling and simulating voting procedures, the random generation of preferences is by far the most important issue. Gehrlein & Fishburn (1981) describe three methods of generating profiles: impartial culture (IC), impartial anonymous culture (IAC) and maximal culture (MC). Apart from the details of each one, all three procedures are neutral regarding the alternatives and characterize "close" elections. Gehrlein (2004, 2006) evaluates how specific variations on IAC affects the probability of observing Condorcet's paradox. Chamberlin & Featherston (1986) present what Tideman & Plassmann (2012) denominated unique unequal probabilities (UUP). In this approach, it is assumed that each possible ranking has a constant probability of being used by a voter. An overview of the most common methods for profile generation is presented in Tideman & Plassmann (2012).
- 4.7** Our method is based on the idea of Chamberlin & Featherston (1986) but takes a different path. Instead of assuming that each possible ranking has a constant probability, we associate in our model each alternative to a Gaussian distribution with predetermined mean and variation. This distribution is the basis of the random numbers generator that will be used to rank the alternatives.
- 4.8** The algorithm of random preference generation works as follows:
1. For each alternative a_i , define a unique Gaussian distribution $N_i(\mu_i, \sigma_i)$;
 2. For each voter v_j :
 - (a) For each alternative a_i :
 - i. Generate x_i , random number generated from $N_i(\mu_i, \sigma_i)$;
 - (b) Order the alternatives according to x_i .

- 4.9 At the end of this process, the algorithm will produce a scenario, i.e. n random rankings of m alternatives.
- 4.10 For N_i , we propose a fixed value for $\sigma(0.5)$ and means ranging linearly from 1 to 0.
- 4.11 As an illustration of a configuration with 4 alternatives, the distributions for a_1, a_2, a_3 and a_4 are, respectively, $N_1(1.0, 0.5), N_2(0.67, 0.5), N_3(0.33, 0.5), N_4(0.0, 0.5)$. Figure 1 depicts the normal curves for each alternative.

Figure 1: Gaussian distributions.

- 4.12 Each distribution in Figure 1 guides the random number generation process of each alternative.
- 4.13 Note that, in spite of the arbitrariness of the values of the standard deviation and the range of means, its values were intended to guarantee that all alternatives have a chance (a very small chance for the least preferred ones) of being chosen as the first candidate.
- 4.14 This procedure will generate a set of 24 rankings of 4 alternatives. Table 1 and Figure 2 present the probabilities of each ranking for the distributions in Figure 1 when using the algorithm of random preference generation proposed.

#	ranking	Prob. (%)
1	a_1, a_2, a_3, a_4	22.42
2	a_1, a_2, a_4, a_3	12.25
3	a_1, a_3, a_2, a_4	13.24
4	a_1, a_3, a_4, a_2	4.44
5	a_1, a_4, a_2, a_3	4.64
6	a_1, a_4, a_3, a_2	2.92
7	a_2, a_1, a_3, a_4	12.25
8	a_2, a_1, a_4, a_3	6.59
9	a_2, a_3, a_1, a_4	4.64
10	a_2, a_3, a_4, a_1	1.00
11	a_2, a_4, a_1, a_3	1.67
12	a_2, a_4, a_3, a_1	0.70
13	a_3, a_1, a_2, a_4	4.44
14	a_3, a_1, a_4, a_2	1.41
15	a_3, a_2, a_1, a_4	2.91
16	a_3, a_2, a_4, a_1	0.60
17	a_3, a_4, a_1, a_2	0.46
18	a_3, a_4, a_2, a_1	0.31
19	a_4, a_1, a_2, a_3	1.00
20	a_4, a_1, a_3, a_2	0.60
21	a_4, a_2, a_1, a_3	0.70
22	a_4, a_2, a_3, a_1	0.28
23	a_4, a_3, a_1, a_2	0.32
24	a_4, a_3, a_2, a_1	0.21

Table 1: Rankings and probabilities.

Figure 2: Rankings and probabilities.

Unit simulation

4.15 Each unit simulation is the simulation of both Borda and RSB_r for a single scenario. It works as follows:

1. Create scenario $(n; m)$;
2. For each voter v_i , with i ranging from 1 to n :
 - (a) Generate A_i (random subset of alternatives of size r);
3. Compute the Borda count for the traditional Borda;
4. Compute the Borda count for RSB_r;
5. Calculate the Match Index (MI).

4.16 The generation of each random subset A_i have to guarantee that the alternatives will have the same chance of being assigned to any random subset.

4.17 Figure 3 shows the results of a unit simulation for a (250; 4) configuration with 2 random alternatives. The distributions of Figure 1 were used for generating the random preferences. For convenience, we have computed the evolution of the Borda count as the voters cast their votes. In practice, it is the equivalent of publishing the partial results of the election after each vote cast.

Figure 3: Gaussian distributions.

4.18 The lines in the graph represents the evolution of the Borda count for all four alternatives in both methods. Note that, as the number of voters increase, the rankings of Borda and RSB₂ tend to converge. In this simulation, the Match Index reached its peak (full match) after 221 votes.

4.19 In the next subsection, we show the multi- r simulations.

Multi-r simulation

- 4.20** For a given scenario, we define a multi-r simulation as the execution of all possible unit simulations for a single scenario. In other words, for each value of r (ranging from 2 to m), a unit simulation is executed and the match index is calculated.
- 4.21** In Figure 4, we present the results of the multi-r simulation of a (1,000; 10) scenario. The graph indicates how MI evolves as the number of random alternatives increases.

Figure 4: Multi-r simulation for (1,000; 10) .

- 4.22** Note that, as the number of random alternatives increase, the match index tends to reach m . It is a rather intuitive result, since the upper bound of the number of random alternatives is in this case 10. In this simulation, the value of MI for $r = 2$ was 3. It means that the comparison of the final rankings of Borda and RSB2 presented only 3 matches. For values of r ranging from 3 to 5, MI was, respectively, 5, 8 and 8. For the remaining values of r (from 6 to 10), MI was equal to 10 (full match).
- 4.23** In the next subsection, we show how we have executed the multi-r simulations for 15 pre-determined configurations.

Multi-configuration simulations

- 4.24** In this subsection, we present the results of the simulations for the configurations in Table 2.

m	Number of voters				
10	500	1,000	2,000	5,000	10,000
20	2,000	5,000	10,000	20,000	40,000
50	5,000	10,000	40,000	70,000	100,000

Table 2: Configurations.

- 4.25** For each configuration, we have generated 20 random scenarios and executed a multi-r simulation for each one. We have then calculated the means of MI for each configuration.
- 4.26** Figure 5 presents the results of the simulations for $m = 10$. Each line in the graph represents, for each configuration, the variation of the match index as the number of random alternatives increase. Note that, as expected, $r = 10$ implied $MI = 10$ for all configurations. This is the traditional Borda method. For small values of r , 3 for example, the mean of MI for the configurations (500; 10), (1,000; 10), (2,000; 10), (5,000; 10) and (10,000; 10) was, respectively, 4.4, 5.05, 7.65, 9.4 and 9.8. For $r = 8$, these values reached 9.6, 9.9, 10, 10 and 10. For n equals to 10,000, 5,000, 2,000, 1,000 and 500, the methods converged, i.e. presented the same rankings, respectively, at r equals to 5, 5, 8, 9 and 10.

Figure 5: Multi-configuration simulation for $m = 10$.

4.27 In Figure 6, the lines indicate the evolution of the match index for the configurations in which $m = 20$. One can easily observe that for $n = 40,000$, the match index reaches its peak at $r = 6$ staying unchanged for the remaining values of r . For n equals to 20,000, 10,000, 5,000 and 2,000 the methods converged, i.e. presented the same rankings, respectively, at r equals to 11, 12, 17 and 20.

Figure 6: Multi-configuration simulation for $m = 20$.

4.28 A similar analysis of Figure 7 shows that, for $m = 50$, the methods converge at r equals to 50, 50, 38, 30 and 25 for m equals to 5,000, 10,000, 40,000, 70,000 and 100,000, respectively.

Figure 7: Multi-configuration simulation for $m = 50$.

Convergence simulation

4.29 The results of the last subsection led us to the following question: For a given number of alternatives, how many voters and random alternatives are necessary to have a full match of Borda and RSBr?

4.30 In order to have a first glance of the answer, we have implemented a simulation in which, for a given number of alternatives, we gradually increase the number of voters and check whether the full match is reached. When it is reached, we decrement r and repeat the same process. This cycle repeats until $r = 2$. The algorithm that follows presents the simulation steps:

1. Define $r = m$
2. Define $n = 100$
3. Define $list = []$
4. While $r \geq 2$:
 - (a) Create scenario $(n; m)$
 - (b) Single-run for the scenario with r random alternatives
 - (c) Compute MI
 - (d) If $MI = m$:
 - i. $r = r - 1$
 - ii. Save tuple $(r; n)$ in list
 - (e) Increment n .

4.31 In summary, this simulation identifies the population in which the first full match is verified for each value of r .

4.32 For each value of m (10, 20 and 50) we have run 20 times the simulation. At the end of the process, we have a list of tuples that indicates the number of voters m in which a full match was verified with r random alternatives.

4.33 Figure 8 presents the results of simulations for $m = 10$. For each value of r , the graph depicts the mean of the population in which a full match of Borda and RSBr was first verified.

Figure 8: Convergence for $m = 10$.

4.34 The behavior of the curve indicates that, for values of r ranging from 10 to 3, the corresponding population rose steadily from 200 to 2,000. For $r = 2$, the number of voters increased exponentially and reached 8,000.

4.35 A similar exponential growth is verified in Figure 9 ($m = 20$). For r ranging from 20 to 5, one can identify a gradual increase in the population. Afterwards, the slope of the curve changes and the subsequent values of n grow rapidly.

4.36 For the simulations with $m = 50$ (Figure 10), the exponential behavior of the curve is even more tangible. After $r = 10$, when the number of voters was around 140,000, the curve begins a clear change in its behavior, leading to approximately 4,000,000 voters when $r = 2$.

Figure 9: Convergence for $m = 20$.

Figure 10: Convergence for $m = 50$.

4.37 In the next sub-section, we present the web experiment implemented to validate RSB.

Web experiment

4.38 In order to have a first glance of a real application of the Random-Subset Borda, we have implemented a very straightforward web experiment in which respondents were asked to rank alternatives, juices in this case, according to their preferences. Our goal was to evaluate how the scenario proposed in Figure 3 ($m = 4$, $r = 2$) performed when real people were required to order alternatives in RS Borda and in Borda.

4.39 In about 4 months, 365 students and federal employees from the Federal Institute of Education and Technology in Pernambuco and from the Federal University of Pernambuco, both in Brazil, participated in the experiment. The voting system was developed in Python/Django and was made available through the internet.

4.40 The entire voting process was divided into four steps:

1. present to the voters the goal of the experiment, which was, in summary, to rank juices according to their preferences;
2. fill in a form with basic personal information (age and sex);
3. rank 2 alternatives;
4. rank 4 alternatives.

4.41 We have chosen juices of four well known fruits in the region: passion fruit, strawberry, acerola and mangaba (*hancornia speciosa*).

4.42 In the third step, as proposed by RS Borda, we have randomly chosen two alternatives from the set of four juices.

4.43 The ranking process was accomplished by a drag and drop tool that allowed the voters to order the alternatives according to their preferences. To minimize any bias in this process, the alternatives were randomly presented to the voters.

4.44 From the set of 365 respondents, we had 160 men and 205 women. Figure 11 presents the distribution of their ages.

Figure 11: Distribution of ages.

4.45 The distribution of alternatives in the random subsets are presented in Figure 12.

Figure 12: Distribution of alternatives.

4.46 From Figure 12, one can notice that a uniform distribution was used to randomly select the alternatives for the random subsets, i.e., a fair mechanism selected the alternatives and assigned them to the respondents. As expected, in spite of the fairness of the mechanism, small variations in the count of alternatives in the subsets were verified. Passion fruit appeared in 184 subsets, while acerola appeared in 181. Strawberry and *mangaba* appeared 182 and 183 times, respectively.⁵

4.47 Figure 13 presents the results of the experiment.

Figure 13: Distribution of alternatives.

4.48 Table 3 presents the final counts for the alternatives in both methods.

Juice	Borda count	RSB2 count
Passion fruit	1049	306
Acerola	1027	287
Strawberry	842	259
Mangaba	732	243

Table 3: Evolution of the Borda and the RSB2 counts for the juices.

4.49 Figure 13 showed the evolution of the Borda count for Borda and for RSB2 for the 365 respondents. After 252 voters the methods converged.

Discussion

5.1 In this section, we present some discussions about the approaches used to validate RSB.

5.2 First, using analytical techniques, we have demonstrated that, for big electorates, RSB and Borda converge to the same outcome. This means that, as the number of voters increase, the results of both methods are the same with probability 1.

5.3 In order to have more practical insights regarding the proposed method, we have run several simulations. The results of the simulations indicate some relationships between the number of alternatives (m), the number of voters (n), the number of random alternatives (r) and the match index (MI):

1. On keeping m and n fixed, increasing r led to an increase in MI . In spite of the variations identified in the results this trend is verified in all simulations;
2. On keeping m and r fixed, increasing n led to an increase in MI ;
3. For a given m , in order to reach a full match ($MI = m$), one can either increase n or r .

5.4 For fixed values of m and r , the simulations also suggest the existence of thresholds for n . In other words, one can expect that after a certain number of voters the probability of having $MI = m$ tends to 1.

5.5 The findings here indicate that, considering the assumptions of transitivity, completeness and IIA, Borda and RSB methods lead to the same results in large electorates. This point was verified in the mathematical model, in all simulations and in the experiment.

5.6 It is important to notice that when comparing the simulation presented in Figure 3 ($m = 4$ and $r = 2$) and the results of the web experiment (Figure 13), a relatively small difference in the convergence threshold was verified. While in the simulation the threshold was 221 voters, in the experiment, it was 252. More experiments should be undertaken in future researches in order to verify how the implemented simulation behave in real scenarios.

Conclusion

6.1 In this paper, we have proposed the Random-Subset Voting procedure, a new way of combining probability and voting. With this method, a voter will not rank the entire set of alternatives, but a subset of it. This subset is selected by a fair random mechanism that guarantees that all candidates will have the same chance of being assigned to each voter.

6.2 We have applied the RSV idea in the context of Borda through the Random-Subset Borda and used three different approaches in the validation of the proposed method: a mathematical analysis of the convergence, a simulation and an experiment. The results of all approaches showed that when the number of voters is large enough RSB and Borda converge to the same outcome.

6.3 The results also suggest the existence of thresholds of convergence. We have used the simulations to estimate the number of voters in which a full match between Borda and RSB might occur for different scenarios.

- 6.4** The main goal of this proposal concerns the minimization of the overwhelming effect some voters could face when dealing with many alternatives. In a hypothetical large-scale scenario, we could implement the following experiment. Citizens in a city or a state have to evaluate (rank) 50 projects. After eliciting their preferences and aggregating the results, the top 5 projects will be implemented. Now, assume that instead of analyzing 50 projects, each citizen will be asked to evaluate only 2, chosen randomly. From a theoretical point of view, when the premises already mentioned hold, the results of RSB will match the results of Borda. It means that the voters do not need to evaluate all the projects in order to reach the same collective decision.
- 6.5** This example guides us to another important aspect of RSV: the possibility of bringing into large-scale decisions voting methods like Borda, Condorcet, ranked ballots, etc. These procedures, due to the considerable amount of information required to the voter, could not be used in several contexts. When deciding over 20 or more alternatives for example, some methods are, in practice, unsuitable. With RSV, they could represent feasible alternatives.
- 6.6** Another facet of RSV concerns the premise that humans are not always as rational as some models preconize and that completeness, transitivity and IIA are not verified. In the mathematical modeling and in the simulations, we have shown that RSB and Borda converge when these premises are considered. In real situations, however, one can anticipate that the rational assumptions do not necessarily hold and that the individual rankings of RSB will not be consistent with the rankings of Borda. It is expected that some voters change the relative position of some alternatives when evaluating fewer options. A voter could have a different opinion about some alternatives when evaluating small subsets. One could argue that these changes caused by this “non-rational” behavior would lead to better collective results, since dealing with fewer options will minimize the overwhelming effect that plays a role in such scenarios. However, further investigation of this point needs to be addressed in the future.
- 6.7** An additional concern regarding RSV is the applicability of the procedure in real scenarios. Some might argue that this work can only be taken from a theoretical point of view, but do not have practical implications. Even though we defend that much work has to be accomplished before bringing RSV into our political life, we do believe it could be a real-world tool for decision making, especially if, or when, we move from our traditional indirect democracy to a more direct and online-based one (Sousa Santos 1998; Sintomer 2010; Pateman 2012). The applied web experiment was a simple but relevant argument in favor of this idea.
- 6.8** It is important to emphasize that RSV is a concept that could be better applied to situations in which many alternatives are in the choice set. It is not yet clear, however, how many alternatives characterize “many”. Is it 10? 20? In fact, we believe that, depending on the nature of the alternatives and on the number of voters, reducing a set of 6 or 7 alternatives to a random subset of 2 might be appropriate. In other contexts, reducing sets of 20 or 30 to random subsets of 4 or 5 could lead to good social results. In future researches, this issue will be further investigated.
- 6.9** An additional point that must be brought into light is that one can defend the argument that some voters would not agree with the subset assigned to them and state they have been harmed by receiving “poor alternatives”.
- 6.10** Even though we could not yet identify to what extent this argument is valid, we argue that this argument can be considered when representatives are the alternatives. When people choose people, poor candidates might be assigned to many voters and lead to an increase in the abstention levels. However, there are evidences that the increase in the abstention levels can be identified in situations where the number of alternatives is big enough (Cunow 2013). This reasoning needs to be addressed in future researches.
- 6.11** In spite of this consideration, the “very poor alternatives” issue do not play a big role when taking projects or actions. In this situation, the goal of the voters will be to evaluate the alternatives assigned to them and declare their preferences. All the alternatives might be good enough, and the goal would be to define the subset of actions that will be implemented when a specific budget is made available, for example.
- 6.12** One last aspect that needs to be taken into account is the operationalization of the proposal. In fact, if this method was proposed a few decades ago, the applicability of it would have been challenged. Nowadays, however, our technology and ability to deliver private, secure and fast information suggest that dealing with the operational overhead of such a solution would not be a barrier.
- 6.13** Other future researches might include modeling, simulating and implementing experiments of other voting schemes and evaluating how they behave in the context of RSV.

Notes

¹There is some confusion in the literature concerning the different definitions of IIA; see Arrow (1950); Ray (1973); Bordes & Tideman (1991); McLean (1995). In general, these definitions split into two categories: social and individual. The social one, defined by Arrow (1950), concerns changes in relative positions of alternatives in the aggregated ordering when adding alternatives to or removing alternatives from the choice set. The individual one, proposed by Radner & Marschak (1954), and also presented by Sen (1971), under the alpha property definition, considers whether adding to or removing an alternative from the choice set will have impact on the preferences of a voter over the alternatives already in the choice set. It is also studied under the framework of psychology and consumer research as decoy or context effects (Trueblood et al. 2013). In this paper, we adopt the IIA definition from the individual perspective.

²Even though it might seem odd for many, lots have been used many times in our democratic history. In Athens, Venice and Florence, for example, many representatives were chosen by lot. Some scholars even argue that sortition was a core principle of the ancient Athenian democracy and played a crucial role at the time (Headlam [1891] 2014; Tridimas 2012; Bouricius 2013; Lockard 2004).

³Nurmi (1983) presents a detailed analysis regarding the difficulty and easiness of voting schemes.

⁴www.openabm.org/model/4758/

⁵A more strict version of the fair mechanism could consider forcing the same count of alternatives in the random subsets. In this case, some voters might get random subsets with $r + 1$ or $r - 1$ random alternatives.

References

- Adams, J. (1997). Condorcet efficiency and the behavioral model of the vote. *The Journal of Politics*, 59(4), 1252–1263
- Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'école Américaine. *Econometrica*, 21(4), 503–546
- Anand, P. (1993). The philosophy of intransitive preference. *The Economic Journal*, 103(417), 337–346
- Arrow, K. J. (1950). A difficulty in the concept of social welfare. *Journal of Political Economy*, 58(4), 328–346
- Arrow, K. J. (1963). *Social Choice and Individual Values*. New York, NY: Wiley
- Barberá, S. (1979). Majority and positional voting in a probabilistic framework. *The Review of Economic Studies*, 46(2), 379–389
- Barberá, S. & Pattanaik, P. K. (1986). Falmagne and the rationalizability of stochastic choices in terms of random orderings. *Econometrica*, 54(3), 707–715
- Bissey, M.-E., Carini, M. & Ortona, G. (2004). ALEX3: A simulation program to compare electoral systems. *Journal of Artificial Societies and Social Simulation*, 7(3), 3
- Bordes, G. & Tideman, N. (1991). Independence of irrelevant alternatives in the theory of voting. *Theory and Decision*, 30(2), 163–186
- Bordley, R. F. (1983). A pragmatic method for evaluating election schemes through simulation. *American Political Science Review*, 77(1), 123–141
- Bouricius, T. G. (2013). Democracy through multi-body sortition: Athenian lessons for the modern day. *Journal of Public Deliberation*, 9(1), 11
- Camerer, C. (1998). Bounded rationality in individual decision making. *Experimental Economics*, 1(2), 163–183
- Chamberlin, J. R. (1986). Discovering manipulated social choices: The coincidence of cycles and manipulated outcomes. *Public Choice*, 51(3), 295–313
- Chamberlin, J. R. & Cohen, M. D. (1978). Toward applicable social choice theory: A comparison of social choice functions under spatial model assumptions. *American Political Science Review*, 72(4), 1341–1356

- Chamberlin, J. R. & Featherston, F. (1986). Selecting a voting system. *The Journal of Politics*, 48(2), 347–369
- Cheng, S. & Long, J. S. (2007). Testing for IIA in the multinomial logit model. *Sociological Methods & Research*, 35(4), 583–600
- Coughlin, P. & Nitzan, S. (1981). Electoral outcomes with probabilistic voting and Nash social welfare maxima. *Journal of Public Economics*, 15(1), 113–121
- Cunow, S. (2013). Too much choice? Abstention rates, representation and the number of candidates. Working paper, University of California
- Danan, E., Ziegelmeyer, A. et al. (2006). Are preferences complete? An experimental measurement of indecisiveness under risk. Working paper, Max Planck Inst. of Economics, Strategic Interaction Group
- Ehlers, L., Peters, H. & Storcken, T. (2002). Strategy-proof probabilistic decision schemes for one-dimensional single-peaked preferences. *Journal of Economic Theory*, 105(2), 408–434
- Fishburn, P. C. (1975). A probabilistic model of social choice: Comment. *Review of Economic Studies*, 42(2), 297–301
- Fishburn, P. C. (1991). Nontransitive preferences in decision theory. *Journal of Risk and Uncertainty*, 4(2), 113–134
- Fishburn, P. C. & Gehrlein, W. V. (1977). Towards a theory of elections with probabilistic preferences. *Econometrica*, 45(8), 1907–1924
- Fry, T. R. L. & Harris, M. N. (1996). A Monte Carlo study of tests for the independence of irrelevant alternatives property. *Transportation Research Part B: Methodological*, 30(1), 19–30
- Gehrlein, W. V. (2004). Consistency in measures of social homogeneity: A connection with proximity to single peaked preferences. *Quality & Quantity*, 38(2), 147–171
- Gehrlein, W. V. (2006). *Condorcet's Paradox*, vol. 40. Berlin/Heidelberg: Springer
- Gehrlein, W. V. & Fishburn, P. C. (1978). Probabilities of election outcomes for large electorates. *Journal of Economic Theory*, 19(1), 38–49
- Gehrlein, W. V. & Fishburn, P. C. (1981). Constant scoring rules for choosing one among many alternatives. *Quality & Quantity*, 15(2), 203–210
- Gehrlein, W. V. & Lepelley, D. (2000). The probability that all weighted scoring rules elect the same winner. *Economics Letters*, 66(2), 191–197
- Gibbard, A. (1977). Manipulation of schemes that mix voting with chance. *Econometrica*, 45(3), 665–681
- Headlam, J. W. ([1891] 2014). *Election by Lot at Athens*. Cambridge: Cambridge University Press
- Intriligator, M. D. (1973). A probabilistic model of social choice. *The Review of Economic Studies*, 40(4), 553–560
- Intriligator, M. D. (1982). Probabilistic models of choice. *Mathematical Social Sciences*, 2(2), 157–166
- Iyengar, S. S. & Lepper, M. R. (2000). When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology*, 79(6), 995–1006
- Johnson, P. E. (1999). Simulation modeling in political science. *American Behavioral Scientist*, 42(10), 1509–1530
- Jones, B., Radcliff, B., Taber, C. & Timpone, R. (1995). Condorcet winners and the paradox of voting: Probability calculations for weak preference orders. *American Political Science Review*, 89(1), 137–144
- Kemeny, J. G. (1959). Mathematics without numbers. *Daedalus*, 88(4), 577–591
- Kollman, K., Miller, J. H. & Page, S. E. (1992). Adaptive parties in spatial elections. *American Political Science Review*, 86(4), 929–937
- Kolmogorov, A. N. (1956). *Foundations of the Theory of Probability*. New York, NY: Chelsea
- Lockard, A. A. (2004). Sortition. In C. K. Rowley & F. Schneider (Eds.), *The Encyclopedia of Public Choice*, (pp. 530–533). New York, NY: Kluwer

- Ludwin, W. G. (1976). Voting methods: A simulation. *Public Choice*, 25(1), 19–30
- McFadden, D., Tye, W. B. & Train, K. (1977). An application of diagnostic tests for the independence from irrelevant alternatives property of the multinomial logit model. *Transportation Research Record*, 637, 39–46
- McKelvey, R. D. & Patty, J. W. (2006). A theory of voting in large elections. *Games and Economic Behavior*, 57(1), 155–180
- McLean, I. (1995). Independence of irrelevant alternatives before Arrow. *Mathematical Social Sciences*, 30(2), 107–126
- Merrill III, S. (1984). A comparison of efficiency of multicandidate electoral systems. *American Journal of Political Science*, 28(1), 23–48
- Nau, R. (2006). The shape of incomplete preferences. *The Annals of Statistics*, 34(5), 2430–2448
- Nitzan, S. (1975). Social preference ordering in a probabilistic voting model. *Public Choice*, 24(1), 93–100
- Nurmi, H. (1983). Voting procedures: A summary analysis. *British Journal of Political Science*, 13(2), 181–208
- Nurmi, H. (2002). *Voting Procedures under Uncertainty*. Berlin: Springer
- Nurmi, H. (2014). Are we done with preference rankings? If we are, then what? *Operations Research and Decisions*, 24(4), 63–74
- Pateman, C. (2012). Participatory democracy revisited. *Perspectives on Politics*, 10(1), 7–19
- Radner, R. & Marschak, J. (1954). Note on some proposed decision criteria. In R. M. Thrall, C. H. Coombs & R. L. Davis (Eds.), *Decision Processes*, (pp. 61–68). Oxford: Wiley
- Ray, P. (1973). Independence of irrelevant alternatives. *Econometrica*, 41(5), 987–991
- Scheibehenne, B., Greifeneder, R. & Todd, P. M. (2010). Can there ever be too many options? A meta-analytic review of choice overload. *Journal of Consumer Research*, 37(3), 409–425
- Schumacher, G. & Vis, B. (2012). Why do social democrats retrench the welfare state? A simulation. *Journal of Artificial Societies and Social Simulation*, 15(3), 4
- Schwartz, B. (2004). *The Paradox of Choice*. New York, NY: Harper Collins
- Schwartz, B., Ward, A., Monterosso, J., Lyubomirsky, S., White, K. & Lehman, D. R. (2002). Maximizing versus satisficing: Happiness is a matter of choice. *Journal of Personality and Social Psychology*, 83(5), 1178–1197
- Sen, A. K. (1971). Choice functions and revealed preference. *The Review of Economic Studies*, 38(3), 307–317
- Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, 69(1), 99–118
- Sintomer, Y. (2010). Random selection, republican self-government, and deliberative democracy. *Constellations*, 17(3), 472–487
- Sousa Santos, B. d. (1998). Participatory budgeting in Porto Alegre: Toward a redistributive democracy. *Politics & Society*, 26(4), 461–510
- Tideman, T. N. & Plassmann, F. (2012). Modeling the outcomes of vote-casting in actual elections. In D. Felsenthal & M. Machover (Eds.), *Electoral Systems: Paradoxes, Assumptions, and Procedures*, (pp. 217–251). New York, NY: Springer
- Tridimas, G. (2012). Constitutional choice in ancient Athens: The rationality of selection to office by lot. *Constitutional Political Economy*, 23(1), 1–21
- 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
- Walpole, R. E., Myers, R. H., Myers, S. L. & Ye, K. (2012). *Probability and Statistics for Engineers and Scientists*. New York, NY: Macmillan
- Zeckhauser, R. (1969). Majority rule with lotteries on alternatives. *The Quarterly Journal of Economics*, 83(4), 696–703