The Robust Beauty of APA Presidential Elections: An Empty-Handed Hunt for the Social Choice Conundrum.*

Sergey V. Popov

National Research University Higher School of Economics

svp@hse.ru

Anna Popova

Department of Psychology,

University of Illinois

apopova2@illinois.edu

Michel Regenwetter†

Department of Psychology,

University of Illinois

regenwet@illinois.edu

March 30, 2012

*This work was supported by National Science Foundation grant SES # 08-20009 (PI: M. Regenwetter, University of Illinois at Urbana-Champaign), entitled *A Quantitative Behavioral Framework for Individual and Social Choice,* awarded by the Decision, Risk and Management Science Program. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of their colleagues, the National Science Foundation or the University of Illinois.

†Corresponding author.

Abstract

Social choice theory in Economics and Political Science has highlighted that competing notions of rational social choice are irreconcilable. This established wisdom is based on hypothetical thought experiments, mathematical impossibility theorems, and computer simulations. We provide new empirical evidence that challenges the practicality of these discouraging predictions. We analyze the ballots from thirteen presidential elections of the American Psychological Association. We report on an empirical comparison of the Condorcet, the Borda, the Plurality, the Anti-Plurality, the Single Transferable Vote, the Coombs, and the Plurality Runoff rules. We find that these rules frequently agree both on the winner and on the social order. Bootstrapping reveals that the coherence among competing rules is a property of the empirical distribution of voters' choices, and it is not specific to a particular sample. Our findings are highly robust to changes in the modeling assumptions that enter our analysis. These findings suggest many interesting open research questions for the emerging paradigm of behavioral social choice: Why do competing social choice procedures agree in real-world electorates? How broadly does the accumulated evidence against the social choice conundrum generalize to other electorates and other candidate choice sets?

Keywords: behavioral social choice, Condorcet cycles, Condorcet efficiency, voting rules.

1 Introduction

Choosing a president of a country or of a professional society, selecting a CEO of an organization, determining a good time for a group lunch, or ranking job applicants are just a few of the multitude of collective decisions that groups, organizations, and society face daily. Voting rules, i.e., mathematical preference aggregation mechanisms, have been studied intensely over a wide range of fields, especially in political science, mathematics, and economics, and also in philosophy and psychology.

To illustrate the fundamental conundrum of social choice, consider a hypothetical example of what can happen when a group of 13 people needs to rank 3 candidates. Writing $A \succ B$ to denote that A is strictly preferred to B, consider the preference rankings of 13 voters in Table 1. We briefly compare the outcomes of

Table 1: Hypothetical Preference Profiles of 13 Voters for Three Choice Options, A, B, and C.

Ranking	Number of Voters
$A \succ B \succ C$	5
$C \succ B \succ A$	3
$C \succ A \succ B$	1
$B \succ C \succ A$	3
$A \succ C \succ B$	1

four of the most heavily studied social choice procedures: the Borda, the Plurality, the Single Transferable Vote, and the Condorcet rules.

According to the centuries old *Borda* procedure (Borda, 1770), the first ranked candidate of each voter scores two points, and the second ranked scores one point. According to the Borda rule the social order is therefore $B \succ A \succ C$, with a $14 : 13 : 12$ point tally, suggesting that the social choice, here the *Borda winner*, ought to be B.

But is B really the best option for this electorate? Application of the most commonly used contemporary voting method, the *plurality* rule provides the social order $A \succ C \succ B$, with a $6 : 4 : 3$ tally. Here, each voter gives one vote to one option, namely the one whom he ranks first. From this perspective, the recommended social choice, here the *Plurality winner*, ought to be A.

Next, consider the most heavily promoted procedure for electoral reform across all types and levels of government in the United States. This is the popular multistage procedure labeled *instant runoff* in the popular media. This rule is also known as the *Hare* system, the *alternative vote*, and is often discussed as the *single transferable vote* (STV) among election scholars (Hare, 1857). STV chooses the Plurality winner when that option was ranked first by more than half of the voters. Otherwise, the option with the smallest number of plurality votes is eliminated, the remaining options are re-ranked, and a new plurality score is computed among the remaining options. Here, using the very same voter profile, B is eliminated, and STV elects option C, hence disagreeing with both the Borda and the Plurality procedures.

We have come full circle: each of the three alternatives B, A, and C has, in turn, been identified as best aggregate choice using a major voting procedure. In fact, the situation is even more dire!

One of the founding scholars of social choice theory, the Marquis de Condorcet (Condorcet, 1785), considered yet another, but very simple and attractive rule of social preference formation: according to the Condorcet rule, a candidate is the winner if she beats all competitors in pairwise elections. The group prefers candidate A over candidate B when a majority of voters prefer A over B. The critical problem with this rule is known as the famous *Condorcet paradox of majority cycles* because the rule can completely fail to generate a winner due to intransitive social preferences. The possibility of majority cycles threatens that a society could become paralyzed when facing a collective decision.

Table 2: The Borda, the Plurality, the Alternative Vote, and the Condorcet outcomes for the hypothetical profile of Table 1.

Voting Rule	Borda	Plurality	Alternative Vote	Condorcet
Winner	B	A	C	None
Loser	C	B	B	None
Social Order	$B \succ A \succ C$	$A \succ C \succ B$	$C \succ A \succ B$	Cycle

The first three methods yield three different 'legitimate' election winners, whereas the Condorcet rule yields an intransitive cycle without a winner.

Returning to our hypothetical voting profile, we find a Condorcet paradox, indeed: option A is preferred to option B by a majority (7 voters prefer A to B whereas only 6 voters prefer B to A), B is majority preferred to C, and C is majority preferred to A, hence yielding an intransitive cycle.

Table 2 summarizes the social choice conundrum we have just reviewed. Examples like this abound in the theoretical social choice literature (Arrow, 1951; Mueller, 2003; Saari, 1994, 2000; Sen, 1970). These examples illustrate how easily one can create a situation in which competing voting rules disagree massively on the legitimate winners and losers of an election and hence, can provide four completely different social orders. Extensive and sophisticated mathematical work, such as Kenneth Arrow's (Arrow, 1951) famous "impossibility theorem" explain why there is no way around the conundrum: different procedures are based on different, and mutually irreconcilable, mathematical properties (axioms). After intense study in the second part of the 20^{th} century, social choice theory has fallen out of fashion, with most scholars fully disillusioned about the prospect of finding a mathematical solution to the problem of rational social choice.

Because social choice theory rests on irrevocable mathematical truths, and because access to large and sufficiently rich empirical data sets is difficult, very few scholars have asked the question whether, where, and when the famous social choice conundrum bears out in empirical data. The theoretical predictions from classical theory hinge on the possibility, not necessity, of disagreement. Whether different social choice methods clash with each other depends on the profile being aggregated. For example, whenever the voters are unanimous in their preference rankings, say everyone has preference ranking $A \succ B \succ C$, then all major aggregation methods yield that same ranking $A \succ B \succ C$ as the social order. Much prior work on the likelihood of incompatible social choice outcomes across competing methods has hinged on distributional assumptions about the voter profiles, assumptions that many scholars consider empirically unrealistic.

We analyze ballots from presidential elections of the American Psychological Association (APA) for the years from 1998 to 2009. These elections provide a rich source of full and partial ranking ballots for five alternatives. The number of ballots in the APA elections typically exceeds 15,000. This is two to three orders of magnitude more than the sample size of most laboratory studies on consensus methods (see Hastie and Kameda, 2005). Another major advantage of these data sets over laboratory data is that, as Chamberlin et al. (1984) explain, the APA elections can be used as proxies for major political elections with real stakes.

Consistent with previous empirical studies (among others Felsenthal et al., 1993; Niemi, 1970; Regenwetter et al., 2006) we find no trace of a Condorcet paradox. Building on prior work, we also find that all six voting rules under consideration are coherent with Condorcet extremely frequently, especially when we concentrate our attention on the winners and the losers, i.e., the first and last ranked candidates in the social order.

Before we report on our empirical analysis, we first provide in Section 2 a summary of relevant background information about classical social choice results. Section 3 introduces the APA data and explains a few key methodological considerations. Section 4 describes and organizes our empirical findings. Section 5 lays out an open landscape of interesting questions for future behavioral social choice research.

2 The Social Choice Conundrum

The normative literature on social choice has highlighted in systematic ways how profoundly various voting rules differ mathematically from each other (see the summaries in Chamberlin and Cohen, 1978; Tideman, 2006), how different rules can systematically disagree with each other on the winners (see Gehrlein and Fishburn, 1978; Gehrlein and Lepelley, 2000), and how often most voting methods disagree with the Condorcet rule as a benchmark of rational social choice (see Adams, 1997; Gehrlein, 1985, 1992, 1999; Merrill, 1984, 1985). The conundrum becomes seemingly insurmountable once we consider the impossibility results. Impossibility theorems such as Arrow's theorem (see Arrow, 1951) and the Gibbard-Satterthwaite theorem (see Gibbard, 1973; Satterthwaite, 1975) state that every voting rule is flawed, because every conceivable rule must violate at least one property out of some set of reasonable properties of rational preference aggregation. For example, since the Condorcet rule can produce majority cycles, it violates the property of ”unrestricted domain” or ”universality”. In particular, the Condorcet rule does not necessarily provide a unique winner or a transitive social order. The Condorcet paradox of cyclical majorities has been explored extensively in the social choice literature (see among others Condorcet, 1785; Gehrlein and Fishburn, 1976b; Gehrlein, 1983; Saari, 1994).

The combination of issues like the ones we just reviewed has led social choice theorists to become extremely pessimistic about the possibility of meaningful consensus formation (Riker, 1982). The theoretical literature demonstrates a plethora of possible levels of disagreement among different voting rules, and it predicts that the Condorcet paradox will be ubiquitous on unrestricted domains. The intensity of pessimistic predictions was further enhanced through analytical and simultation results about the likelihood of Condorcet paradoxes based on the famous Impartial Culture assumption (see Gehrlein and Fishburn, 1976b). List and Goodin (2001) and Regenwetter et al. (2006) showed that the likelihood of a Condorcet paradox changes dramatically under even minute deviations from the Impartical Culture assumption. Whether a Condorcet paradox occurs in a random sample from a hypothetical culture depends entirely on the theoretical assumptions about the nature of that culture. "Cultures of indifference," in which all candidates are tied by majority rule at the level of the theoretical distribution, are misleading because the majority outcome in samples of odd size converges to the majority tie in the underlying distribution with probability zero as the sample size increases. Instead, that majority tie is replaced by artificial majority cycles. The literature extrapolated from the Impartial Culture assumption that, in large electorates and with more than three candidates, a majority winner is unlikely even to exist (Riker, 1982). Based on this misconception, the literature fosters a questionable policy recommendation that one should avoid the Condorcet rule in real elections (e.g. Shepsle and Bonchek, 1997).

Another branch of the social choice literature investigated domain restrictions that help avoid the Condorcet paradox (Black, 1948; Sen, 1970). One of the most prominent is Black's *single-peakedness* theorem. Black showed that if there is a linear ordering of the candidates relative to which each voter's preferences have only one peak, a Condorcet cycle is impossible. It appears not only that those results are hard to generalize to a multidimensional space, but also in spite of all the virtues of single peaked preferences and their importance in political science, real electorates generally do not satisfy those restrictions (Faliszewski et al., 2009).

A standard theoretical benchmark in the evaluation of the goodness of a voting rule is whether the rule's winner matches the Condorcet winner if one exists. A candidate who can beat any other candidate in pairwise comparison remains for many the most attractive social option, and therefore the Condorcet winner remains a central concept in social choice. The *Condorcet efficiency* of a rule is the probability that this rule's winner matches the Condorcet winner, given that one exists in a random sample of ballots. Similarly, one can consider *Borda efficiency*. The classical social choice literature estimates the Condorcet efficiency under the Impartial Culture assumption (see Gehrlein and Fishburn, 1978; Gehrlein and Lepelley, 2000; Gehrlein and Fishburn, 1978; Gehrlein, 1985, 1992). As in the situation with a Condorcet paradox, this leads to the questionable conclusion that with a large number of candidates the Condorcet efficiency decreases drastically. For example, for the class of weighted scoring rules, the Condorcet efficiency is 0.535 for three candidates, under the Impartial Culture assumption (Gehrlein, 1999).

We suggest to go beyond those two popular benchmarks and look at all pairwise agreements for both winners and losers among a larger set of rules. Thus, we extend the notion of *efficiency* to a much larger set of rules including the Plurality rule, the Anti Plurality rule, the Single Transferable Vote rule, and the Coombs rule, along with the Condorcet rule and the Borda rule. We also compare the social orders for all of these different rules.

The classical social choice literature is dominated by theoretical work, mathematical theorems and computer simulations based on distributional assumptions which are heavily debated. Alternatives to specific assumptions, such as the Impartial Culture assumption, the Unanimity assumption, and others, have not been developed. The scarcity of empirical evidence on voter preferences is one major reason for this lack of alternatives in classical theory. Another reason is that most available data sets contain, at best, partial rankings of options by voters.

Applying a large set of rules to data from real elections is challenging because most datasets from real elections are dominated by *partial ballots*, i.e. voters can and do rank as many candidates as they want. This implies that, while some voters provide full linear orders, others only rank a subset of candidates. However, most rules considered in the theoretical literature are only defined for preferences represented by full linear orders.

We use the methodology suggested by Regenwetter et al. (2006) and Regenwetter and Rykhlevskaia (2007), which allows us to work with a wide class of voting rules and make data analysis possible for partial rankings. We propose three models of partial rankings to transform election data into binary relations: the Weak Order model (WO), the Partial Order model (PO), and the Size Independent Linear Order model (SIM). All of these models work under completely different assumptions which complement partially reported individual preferences. These models allow us to test how sensitive the outcomes are to different underlying statistical models for partial ballots.

Another important issue we want to highlight is the robustness of election outcomes and the usage of inferential statistics in social choice. While the theoretical literature bases many of its results on sampling distributions derived from an Impartial Culture assumption or, in other words, a uniform distribution of voter preferences, empirical data analyzes rarely use statistical inference. Regenwetter and Rykhlevskaia (2007) and Regenwetter et al. (2006) encourage the use of inferential statistical methods. Regenwetter et al. (2006) and Regenwetter et al. (2009) demonstrate using a sampling approach that the behavior of the Condorcet rule and a class of scoring rules is extremely dependent on the underlying theoretical assumptions. It appears that the widely used Impartial Culture assumption is a knife-edge distribution over preference relationships which leads to the absence of a Condorcet winner and to ties for scoring rules, whereas even a small deviation from this distribution can lead to dramatically different results. In this paper, we use statistical methods, more specifically, bootstrap techniques, to show that the social orders computed using different rules and agreement between them are not very sensitive to small data perturbations.

In practice, some rules are easier to implement than others. The Borda and Condorcet methods, as well as the Single Transferable Vote, are among the most demanding ones — they technically require each voter to order all choice alternatives on her ballot. Others, such as the Plurality rule, require each voter only to report the highest ranked alternative. Nevertheless, we obtain high levels of agreement among all of these rules within each model of partial rankings and across models. This finding of high levels of agreement is not sensitive to small data perturbations.

We conclude that our empirical evidence does not support the predictions of the theoretical literature either on the prevalence of the Condorcet paradox or on the frequency of disagreement between rules. Felsenthal et al. (1993); Niemi (1970); Regenwetter et al. (2002) find no empirical evidence of the Condorcet paradox in a large number of surveys and ballot data, thereby suggesting that the theoretical predictions regarding the frequency of the Condorcet paradox are exaggerated. Regenwetter et al. (2007) and Regenwetter et al. (2009) report a high level of agreement on both winners and losers among a larger set of voting rules. In this study, we elaborate on those results.

Our findings suggest that in a number of situations, focused on gathering information rather than making a socially-relevant decision, it is not necessary to implement the Condorcet procedure to obtain the Condorcet winner. Instead, one could implement the less demanding Plurality procedure, similar to most presidential elections when each voter only provides her best option. The Plurality procedure requires voters to report less information, and the result is easier to tally. Thus, a less information-intensive procedure frequently determines the Condorcet winner correctly.

3 Methodology and Data

In our analysis, we use data from the presidential elections of the American Psychological Association (APA). The APA presidential elections present a unique dataset, which represents the distribution of an electorate in one of the largest scientific and professional organizations in the United States. Moreover, it is also a good approximation of a political election, due to the politicized structure of the association itself. APA presidential elections run yearly and provide a rich source of full and partial rankings for 5 candidates

Table 3: Full and Partial Rankings in the Data.

Election	Number of Candidates Ranked					Number of
Year	5	4	3	2	1	Voters
1998	0.57	0.01	0.08	0.14	0.20	18,723
1999	0.56	0.01	0.09	0.15	0.18	18,398
2000	0.56	0.01	0.10	0.15	0.19	20,239
2001	0.57	0.02	0.08	0.13	0.20	17,911
2002	0.60	0.02	0.11	0.11	0.16	17,094
2003	0.21	0.36	0.08	0.14	0.21	17,956
2004	0.58	0.02	0.07	0.13	0.19	16,245
2005	0.58	0.01	0.08	0.17	0.16	14,506
2006	0.59	0.02	0.10	0.14	0.16	16,836
2007	0.62	0.02	0.09	0.12	0.16	13,318
2008	0.54	0.05	0.07	0.13	0.21	18,286
2009	0.58	0.01	0.08	0.12	0.21	15,313

with an approximate size of the electorate of 20,000 voters. Meanwhile, most surveys and open source election ballots include a much smaller number of candidates (usually three) or a much smaller voter pool (usually fewer than 2,000 voters). The particular advantage of the APA data set is a combination of full (all candidates are ranked) and partial (some candidates are ranked) rankings provided by the voters, which allows us to construct a diverse set of statistical models to estimate population profiles under a variety of assumptions and to analyze various distributional characteristics based on these assumptions.

3.1 Data

The APA data contain a set of ballots with full or partial rankings of 5 alternatives. Table 3 demonstrates that partial rankings represent a non-negligible fraction of the data since they make up from 38% to 79% of all ballots. It also shows that partial rankings of all lengths are present in all 12 datasets.

One very common assumption in the theoretical social choice literature is the *Impartial Culture* assumption according to which each possible linear order is equally probable. For 5 candidates we obtain $5! = 120$ possible linear orders. A regular Chi-square test rejects the Impartial Culture hypothesis at the 0.001 significance level for each of the 12 data sets.[1] For illustrative purposes we graph the number of ballots containing each of the 120 linear orders for the year 1998 in Figure 1. Figure 1 compares the actual frequency of linear orders in the data (dashed lines) with the number predicted by the Impartial Culture assumption (solid lines). The 95% confidence intervals are also presented in Figure 1 (dotted lines). The other data sets give similar results and are omitted for brevity.

In each APA data set we observe all possible 120 linear orders. Therefore, another extreme theoretical assumption of single peakedness is also rejected by the data in each election. In order to make inferences about the empirical distribution of voters' preferences and measure the robustness of this distribution to small perturbations, we employ bootstrap techniques. We create ten thousand pseudosamples by sampling ballots from the empirical distribution with replacement. Then, we apply the voting rules to the pseudosamples and check whether the pseudosamples yield different social orders.[2]

3.2 Models of Partial Rankings

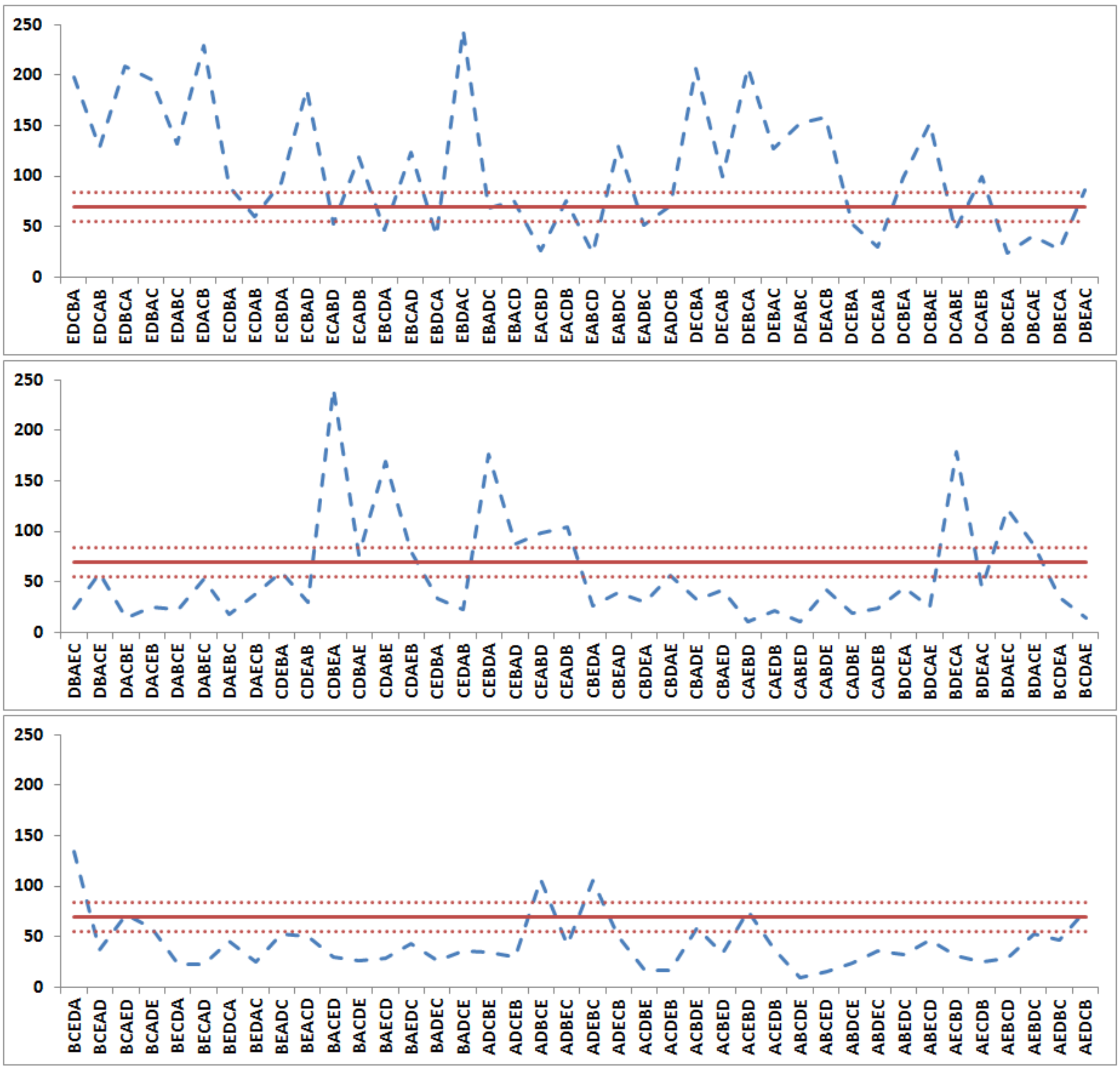
The APA data set is a rich source of full and partial rankings ballots. According to the structure of presidential elections in the APA, voters can rank as many candidates as they want. Therefore, while some voters provide linear orders by ranking all candidates, other voters only rank a subset of candidates and thereby provide partial rankings. Some voting rules are only defined for linear order preferences; this creates an obstacle for empirical analysis. A novel methodology suggested by Regenwetter et al. (2006) and Regenwetter and Rykhlevskaia (2007) allows us to work with a wide class of voting rules and make data analysis possible for partial rankings. We use three models of partial rankings mentioned earlier to transform the data into binary relations: the Weak Order model (WO), the Partial Order model (PO), and the Size Independent Linear Order model (SIM).

The *Weak Order* model assumes that all unranked candidates are tied at the bottom. Voters only express their preferences for their most preferred candidates and do not rank other candidates. This model assumes

[1] When only partial rankings are used, the hypothesis is rejected at an even lower significance level, because not all possible partial rankings are present in the data. A description of models of partial rankings used in the test is provided below.

[2] We use both nonparametric and parametric bootstrap methods assuming independence of the true distribution of preferences from the decisions of voters to report partial rankings.

Figure 1: Frequencies of Linear Orders in the 1998 APA Data.



that the voter prefers all ranked candidates to all unranked candidates, and does not discriminate between any two unranked candidates. For positional rules, i.e. rules which rank all candidates in the social ordering, we use the notion of generalized rank to specify a rank for each candidate. Suppose we have a voter's preferences for five candidates (A, B, C, D, E) where she ranks: A as the most preferable, B as the second, and C as the third preferred candidate. She does not state any preferences for candidates D and E. According to the Weak Order model she prefers A to everyone else, B to everyone else but A, C to D and E, and D is equivalent to E:

$$\begin{tikzcd} & A \arrow[d] & \\ & B \arrow[d] & \\ & C \arrow[dl] \arrow[dr] & \\ D & & E \end{tikzcd}$$

The *Partial Order* model works with different assumptions. It does not assume any binary relations for a pair of candidates when one or both of them are unranked. In the previous example, A is preferred to B, C, but not to D or E:

$$\begin{tikzcd} & A \arrow[d] & \\ D & B \arrow[d] & E \\ & C & \end{tikzcd}$$

The *Size Independent Linear Order* model (SIM) uses statistical inferences from linear orders in the data set. We parameterize the data generating process as follows. We assume each agent is endowed with a full linear ranking of alternatives, represented by a permutation of a sequence of alternatives (ABCDE would be an example of such a permutation). The ranking for each voter is independently picked from the set of all permutations according to a fundamental probability distribution over that set. After the ranking is chosen, a voter randomly picks the length of the reported ranking, independently from other voters and from the chosen alternative; the voter non-strategically reports the best k alternatives.

Denote the fundamental probabilities of orderings $ABCDE$ and $ABCED$, p_{ABCDE} and p_{ABCED} respectively. Let the probability of reporting just $k = 3$ alternatives be denoted by $p_{k=3}$. Then, the probability of observing a partial ranking ABC, p_{ABC}, in the previous example is computed as follows:

$$p_{ABC} = (p_{ABCDE} + p_{ABCED})\, p_{k=3}.$$

Similar formulas are used for other partially ranked ballots. This approach allows us to write down the likelihood of observing the data sample, estimate a probability distribution over full rankings, the probability of reporting a partial ranking of a specific length. This approach allows us to generate pseudosamples from the data-generating process using the estimated parameters. Since we impose structural assumptions and estimate structural parameters, we refer to this bootstrapping process as a ”parametric bootstrap” procedure.

3.3 List of Social Choice Rules

We selected seven well-known voting rules to analyze. These rules belong to different classes and have different algorithms, which vary in complexity.

- Plurality;
- Borda;
- Condorcet;
- Anti(Negative) Plurality;
- Single Transferable Vote (including Alternative Vote);
- Coombs;
- Plurality Runoff.

Detailed algorithms for each rule can be found in the Appendix.

4 Results

We analyze 10,000 bootstrapped pseudo-samples drawn with replacement from 12 data sets using 3 different models of partial rankings, which makes a total of 360,000 samples. Each pseudo-sample contains on the order of 20,000 voters — we always draw as many as in each original data set. Using each set of

10,000 bootstrapped pseudo-samples, we compute rates of agreement between pairs of voting rules for each of the 12 data sets for each of the 3 models of partial rankings. Thus, we obtain rates of agreement between all pairs of rules for 36 sets of data. To present our results graphically, we split the unit interval into five bins of equal size, which represent rates of agreement falling into the intervals [0,0.2], (0.2,0.4], (0.4,0.6], (0.6,0.8], (0.8,1]. For each pair of rules, we compute the number of sets of data out of 36, for which the rate of agreement falls into each of the five bins.

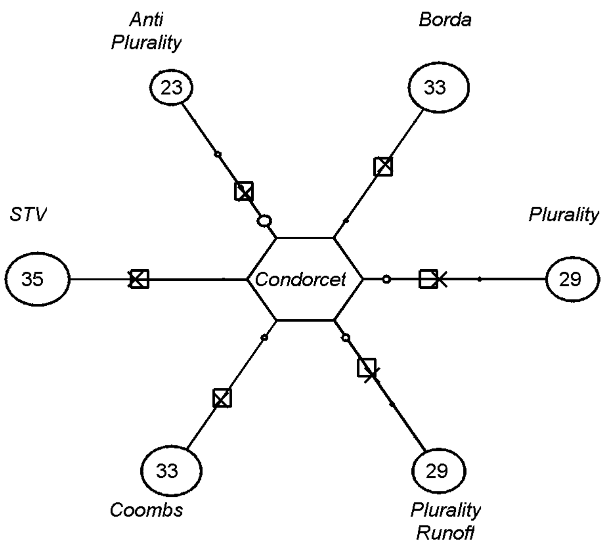
Figure 2 presents the rates of agreement on first and last ranked options, i.e. winners and losers, among seven different voting rules across twelve years and three partial rankings models. The upper left diagram of Figure 2 compares the rates of agreement between the Condorcet winner and the winners according to the other six voting rules. The voting rule in the center represents the reference rule (Condorcet). Each leg represents a unit interval with 0 at the reference rule and 1 at the end of the leg. Each leg can have up to five circles, which represent the five intervals mentioned before. The diameter of each circle visualizes the number of data sets and models for which the rates of agreement between the two rules fall into the corresponding interval. The number of data sets is also written inside each circle. The maximum number is 36.

Therefore, the position of a circle on each leg represents the level of agreement, while the size of the circle represents the frequency at which we observe this level of agreement. When the largest circle is at the end of a leg, the corresponding rule almost perfectly agrees with the reference rule on winners. Similarly, when the largest circle is at the start of the leg, the two rules almost never agree on the winners.

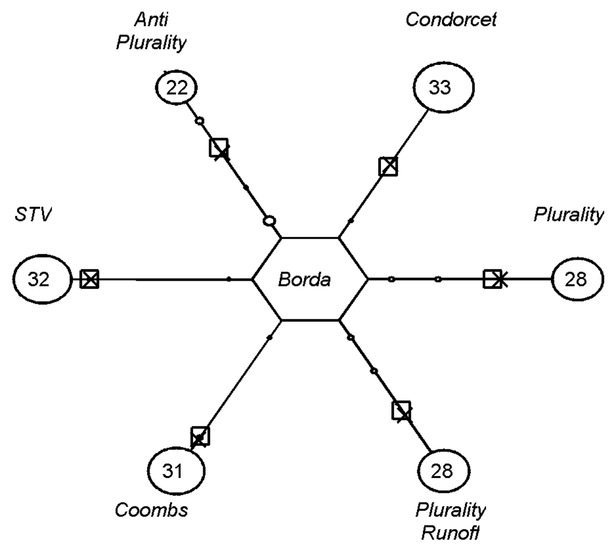
Similarly, the upper right diagram in Figure 2 presents the rates of agreement on winners between the Borda rule and the other six rules. The two lower diagrams in Figure 2 present the rates of agreement on losers between the Condorcet rule (on the left) and the Borda rule (on the right) with the other rules. We use two theoretical benchmarks to compare with our empirical results: agreement among rules under the Impartial Culture (IC) assumption and agreement among rules under the Impartial Anonymous Culture (IAC) assumption of Gehrlein and Fishburn (1976a). Both assumptions are widely used in the theoretical social choice literature. The rate of agreement under IC corresponds to the location of a square and of IAC to the location of an X on each leg of the four diagrams in Figure 2. We obtained these theoretical rates of agreement by Monte-Carlo simulation.

For instance, suppose we are interested in how often the Anti Plurality rule and the Condorcet rule agree on the winner in our 36 data sets. There are four circles on the leg which connects the Condorcet rule and

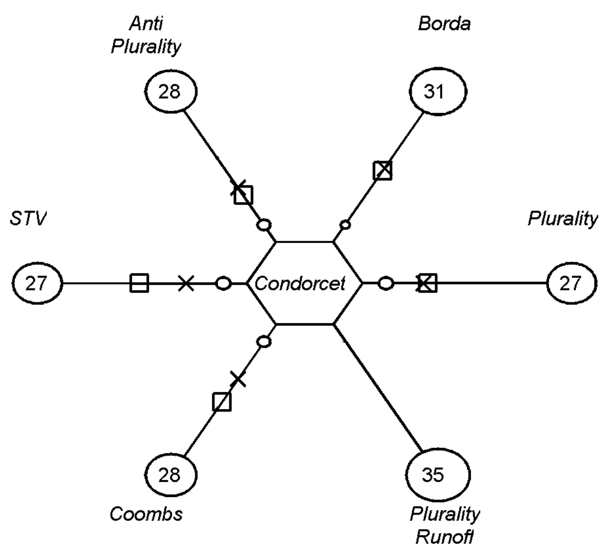
Figure 2: Agreement on Winners and Losers for Seven Rules



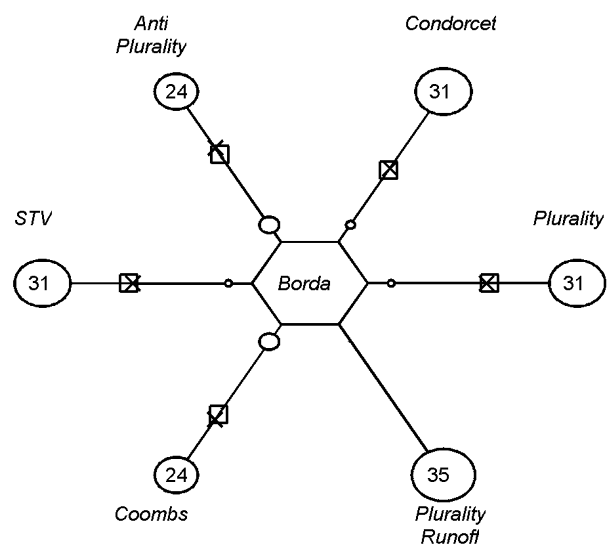
Condorcet rule, Winners



Borda rule, Winners



Condorcet rule, Losers



Borda rule, Losers

See the description of the diagrams in the text. When computing the rate of agreement with the Plurality Runoff rule, we assume that there is agreement if the loser of a rule belongs to the set of losers of the Plurality Runoff rule.

the Anti Plurality rule in the upper left panel of Figure 2. The largest circle is located at the end of the leg. This means that bootstrapped robustness is relatively high. In fact, this means that for 23 out of 36 data sets the rate of agreement on the winner between the Anti Plurality rule and the Condorcet rule in bootstrapped pseudo-samples is above 80%. The second largest circle corresponds to the first bin on the leg. This circle indicates, that for 8 out of 36 data sets the rate of agreement between the Anti Plurality rule and the Condorcet rule in bootstrapped pseudo-samples is below 20%.

Thus, the two upper panels of Figure 2 present in a visual form measures of Condorcet/Borda efficiency of the other six rules. If the position of a circle on a leg is far from (close to) the center, then the Condorcet/Borda efficiency of the corresponding rule is high (low). The upper left diagram of Figure 2 indicates that the Condorcet efficiency of all the rules, except Anti Plurality, is very high. Similarly, the upper right diagram indicates that the Borda efficiency of all rules is relatively high.

Figure 2 demonstrates that the rates of agreement between rules observed in our data sets for all voting rules are extremely high. Figure 2 also shows that this result has a very high level of bootstrapped confidence. These rates of agreement are much higher than those predicted by the theoretical literature, especially the rates obtained from simulations under the IC and IAC assumptions, as depicted in Figure 2. This disagreement between the empirical and theoretical results supports our intuition that neither IC nor IAC should be viewed as good approximations of distributions of electorate preferences. Figure 2 shows that our conclusions are broadly consistent across all models and all years, within each model and within each year. This suggests, that our results are robust to the choice of model of partial rankings and to the choice of data set.

Notice that, when applied to the APA data sets, the complex multistage rules (Single Transferable Vote, Coombs, Plurality Runoff) provide results very similar to the simple rules (Plurality, Anti-Plurality). Agreement among the Borda rule, the Plurality rule, the Condorcet rule, the Single Transferable Vote rule, and the Plurality Runoff rule on winners is virtually perfect. Agreement among the same rules on losers is also relatively high. This is consistent with a result of Saari (1999), that in absence of Condorcet cycles the Condorcet rule and the Borda rule have a good chance of providing the same social orders. Surprisingly, the Plurality rule agrees with many other rules on winners almost perfectly, even though it disregards a large part of the binary preferences and only takes into consideration first ranks.

For five candidates the theoretical probability of a Condorcet cycle under the Impartial Culture assumption for linear orders is 0.251 (Riker, 1982). In contrast to the theoretical result, what we observe in the data is the complete absence of the Condorcet paradox in more than 90 percent of all bootstrapped pseudo-samples. This result is robust to the choice of model of partial rankings and to the choice of data set.

Assumptions of single peaked and ”value restricted” preferences are typically used to explain the absence of the Condorcet paradox. However, these restrictions on the domain, aimed at eliminating the possibility of the Condorcet paradox, rule out some linear order preferences. Since we do observe all linear orders in all the datasets, we can conclude that the assumption of single—peaked preferences is also too restrictive as a description of the APA electorate. Thus, none of the standard assumptions on the distribution of preferences can be used to explain our results, suggesting a need for more behaviorally disciplined models of distributions of preferences.

4.1 Small Samples Results

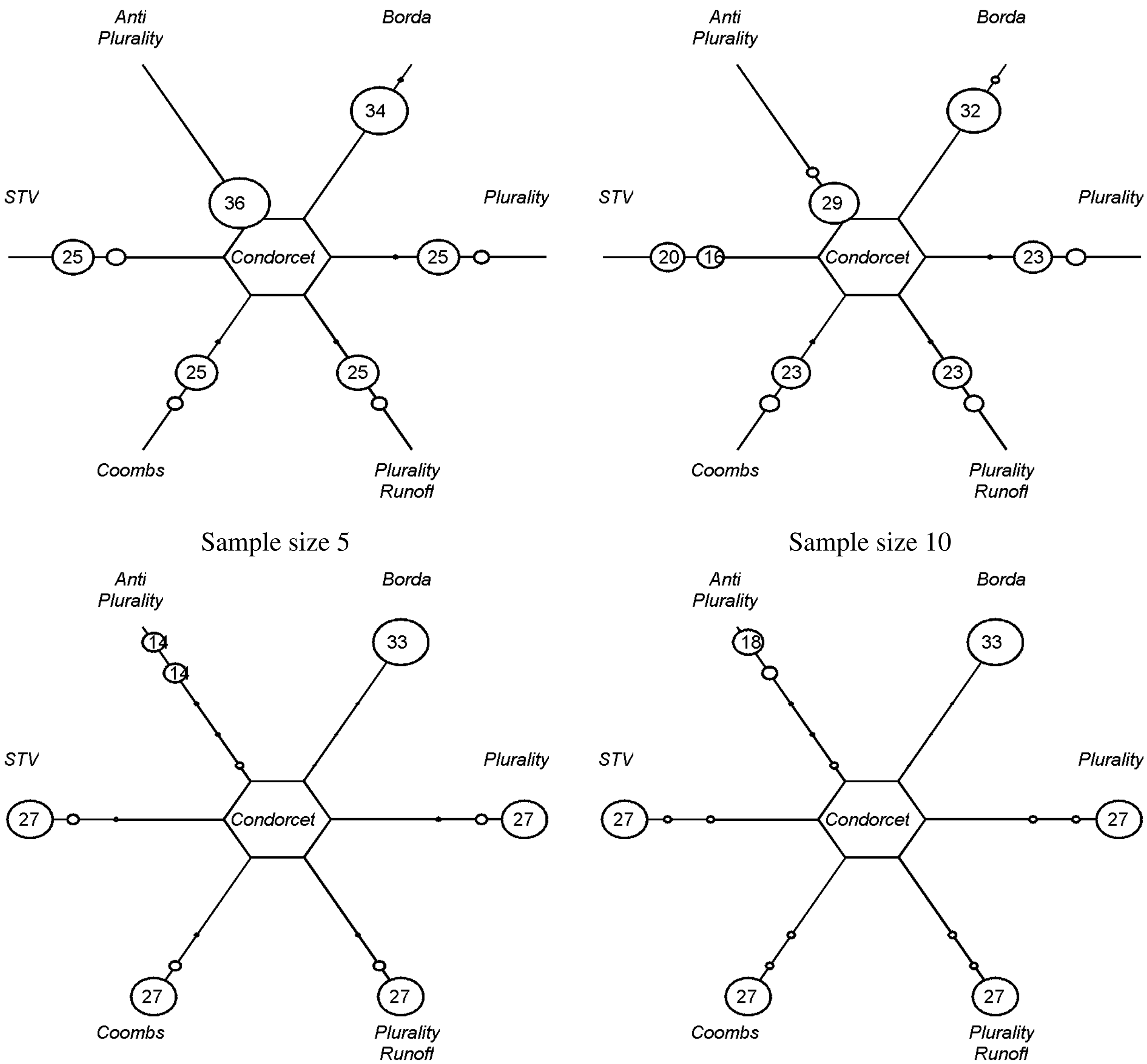
Another point of interest is the robustness of our results to the size of the sample. So far, we have analyzed original data sets and pseudo-samples drawn from them with the sample size equal to approximately 20,000 voters. We now consider pseudo-samples with much smaller sample sizes. Figure 3 presents the rates of agreement on winners between the Condorcet rule and the other six rules (akin to the upper left panel of Figure 2) for sample sizes 5, 10, 500 and 1,000.

The comparison of rates of agreement in Figures 3 and 2 demonstrates the high level of robustness of our results to sample size. Sample size of 1,000 voters is sufficient to obtain the same patterns as in the original data sets of size 20,000. Moreover, agreement rates converge very quickly. An increase from sample size 5 to sample size 500 already demonstrates substantial improvements in the rates of agreement on the winner.

The second robustness exercise we do is check the dependence of the results on the bootstrap size. Bootstrap size represents the number of bootstrap iterations. So far, we have analyzed pseudo-samples with the bootstrap size equal to 10,000. We now consider pseudo-samples with much smaller bootstrap sizes. We keep the size of pseudo-samples equal to the sample size. Figure 4 presents the rates of agreement on winners between the Condorcet rule and the other six rules (akin to the upper left panel of Figure 2) for bootstrap sizes 5, 10, 500 and 1,000.

The comparison of rates of agreement in Figures 4 and 2 demonstrates the high level of robustness of our results to bootstrap size. In fact, bootstrap size of 10 is sufficient to obtain essentially the same results as under bootstrap size 10,000. This demonstrates that, although results under bootstrap size 10,000 are more

Figure 3: Agreement on Winners for Different Sample Sizes



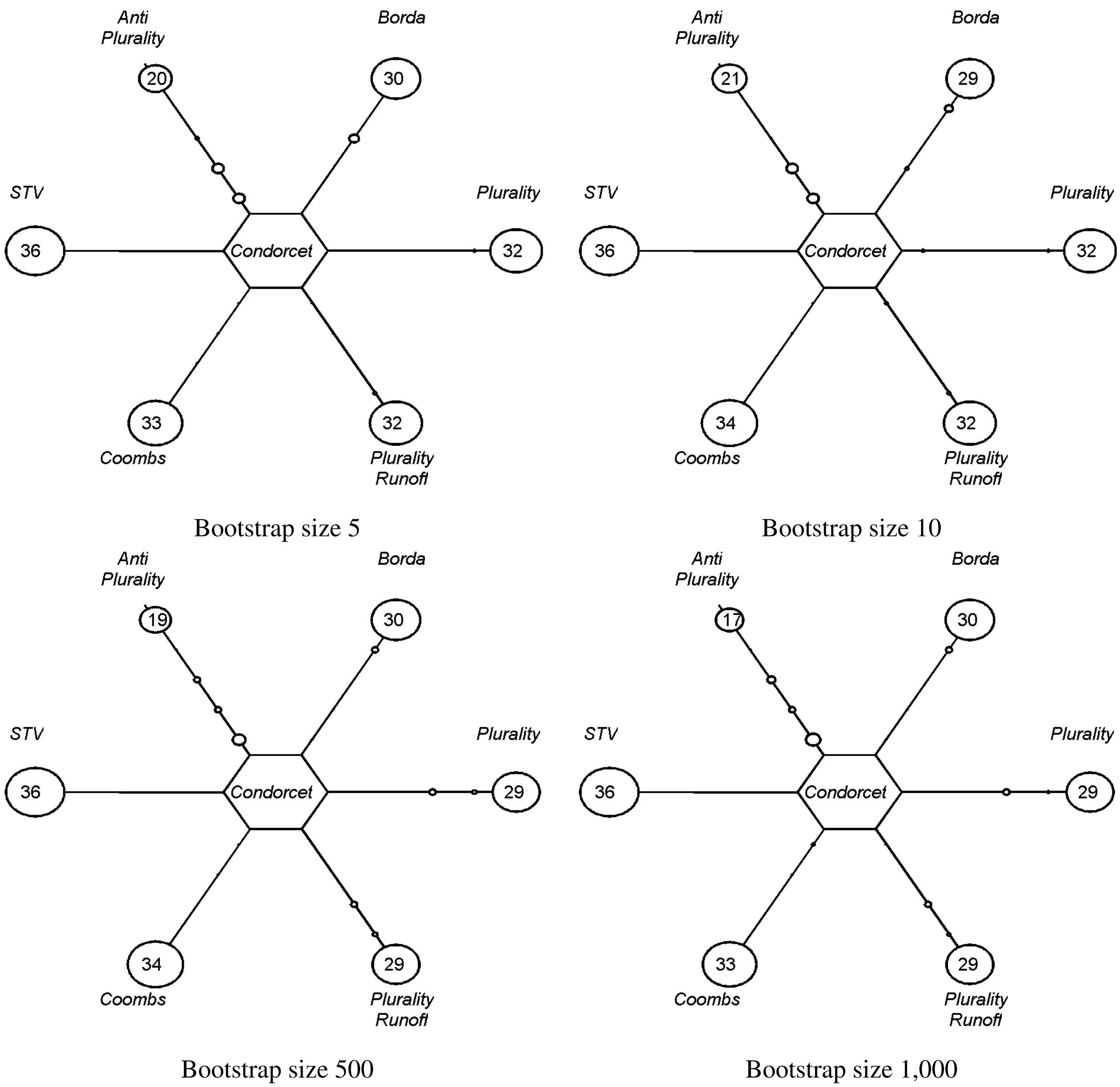
Sample size 5

Sample size 10

Sample size 500

Sample size 1,000

Figure 4: Agreement on Winners for Different Bootstrap Sizes



accurate, the increase from bootstrap size 10 to bootstrap size 10,000 is not necessary to get a good idea of the results.

5 Conclusions and Discussion

We study the properties of a set of well-known voting rules on real electorate large scale data from presidential elections of the American Psychological Association for the years from 1998 to 2009. The APA dataset presents a rare and rich source of information about full and partially ranked preferences of large electorates. These data allow us to apply a wide list of voting rules to the ballots. In addition, the complexity of the alternative vote procedure used in APA elections makes strategic voting difficult, thus, justifying our assumption of truthful revelation of preferences.

First, we applied a wide list of rules to the original data, and found strong agreement of all rules with respect to the choice of the first- and last-ranked candidate. To control for the influence of small random factors, we generated 10,000 samples of pseudo-data by drawing data from the original samples with replacement. We also verified whether our results were driven by the size of the sample. We documented the absence of Condorcet paradox in all 12 original data sets and in a set of pseudo-samples of size from 5 to 1,000. We found that in all 12 election data sets and pseudo-samples generated from them in virtually all Condorcet social orders were complete linear orders.

This finding stood out in contrast with the dominant view in the theoretical literature that Condorcet cycles are a problem when implementing social decision making. Nonetheless, this finding was consistent with previous studies on real datasets, which emphasized that the Condorcet rule performs well in real elections. We also ruled out single peaked preferences, the most common explanation for the absence of cycles. In our data, every complete linear order was used by at least one voter.

Second, we reported outstanding agreement among different rules on winners and on losers. This high level of agreement was obtained for rules of different intensity and complexity. For example, the one-stage Plurality rule agreed well on winners with the multistage Alternative Vote rule. Scoring rules (Plurality, Borda, Anti-Plurality) agreed with the more mathematically complex Condorcet rule, which exploits the pairwise comparison principle. Voting rules of very different mathematical nature and with very different properties (see rules comparison in Tideman, 2006) tended to agree almost perfectly on first- and last-rank candidates. In all 12 datasets we determined first- and last- rank candidates with high confidence for each rule, and they matched nearly perfectly throughout.

Third, we demonstrated the robustness of our results to different formats of data input. We did not discard partial ranking data since they represent up to 40% of ballots. Taking partial rankings data into account required additional simplifying assumptions to make all voting rules applicable. Our analysis demonstrated that, despite major differences in modeling assumptions, agreement among voting rules was consistently high.

Fourth, we studied the behavior of agreement rates as the size of the pseudo-samples decreased. Naturally, under an impartial culture, it was much easier to get Condorcet cycles randomly when the pseudo-sample was of size 5 than when it was of size 500. We found that not only the agreement rates were high in pseudo-samples of size 17,000 (like the data), but also that the agreement rates remained high for samples larger than 500. This implies that surveying 500 or so people would be sufficient to measure the preferences of the population and predict an election outcome consistent with all rules.

Our findings remove much of the gloom around the social choice literature. Although theoretical assumptions about electoral preference distributions lead to stark predictions about agreement rates between voting rules, we find remarkable agreement among these rules when applied to real election data. First, this implies that usual theorems about disagreement of rules or about the possibility of Condorcet cycles are relevant only for a narrow class of preference distributions. For instance, the Impartial Culture assumption, notorious for producing disagreements, does not seem to describe the observed patterns of social choice well.

Second, we find that simpler and less data-intensive rules frequently produce the same results as do complicated data-intensive rules. The agreement rate between the Plurality rule and the Condorcet rule is virtually 100% percent in our permuted-samples experiment, although the Plurality rule requires just asking about each voter's favorite candidate, whereas the implementation of the Condorcet rule involves collecting detailed preferences of all voters and then conducting all pairwise comparisons for every pairing of candidates.

Our findings support previous results that the assumptions used in the theoretical social choice literature are extreme. Our findings also highlight that there is demand for new and more behaviorally disciplined models of distributions of preferences that are consistent with real-life data. Future work will benefit from a more detailed analysis of large scale data which will help uncover features and characteristics of real electorate profiles leading to such outstanding agreement.

References

Adams, J. (1997). Condorcet efficiency and the behavioral model of the vote. *The Journal of Politics*, 59:1252–1263.

Arrow, K. J. (1951). *Social Choice and Individual Values*. Wiley, New York.

Black, D. (1948). On the rationale of group decision-making. *Journal of Political Economy*, pages 23–34.

Borda, J.-C. (1770). On elections by ballot. Orally presented to the French Academy.

Chamberlin, J., Cohen, J., and Coombs, C. (1984). Social choice observed: Five presidential elections of the american psychological association. *The Journal of Politics*, pages 479–502.

Chamberlin, J. and Cohen, M. (1978). Toward applicable social choice theory: A comparison of social choice functions under spatial model assumptions. *The American Political Science Review*, pages 1341–1356.

Condorcet, M. (1785). *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix (Essai on the application of the probabilistic analysis of majority vote decisions)*. Imprimerie Royale, Paris.

Faliszewski, P., Hemaspaandra, E., Hemaspaandra, L., and Rothe, J. (2009). The shield that never was: Societies with single-peaked preferences are more open to manipulation and control. In *Proceedings of the 12th Conference on Theoretical Aspects of Rationality and Knowledge*, pages 118–127. ACM.

Felsenthal, D. S., Maoz, Z., and Rapoport, A. (1993). An empirical evaluation of 6 voting procedures - do they really make any difference? *British Journal of Political Science*, pages 1–27.

Gehrlein, W. V. (1983). Concorcet's paradox. *Theory and Decision*, 15:161—197.

Gehrlein, W. V. (1985). Condorcet efficiency of constant scoring rules for large electorates. *Economics Letters*, 19:13—15.

Gehrlein, W. V. (1992). Condorcet efficiency of simple voting rules for large electorates. *Economics Letters*, 40:61—66.

Gehrlein, W. V. (1999). On the probability that all weighted scoring rules elect the condorcet winner. *Quality and Quantity*, 33:77—84.

Gehrlein, W. V. and Fishburn, P. C. (1976a). Concorcet's paradox and anonymous preference profiles. *Public Choice*, 26:1—18.

Gehrlein, W. V. and Fishburn, P. C. (1976b). The probability of the paradox of voting: A computable solution. *Journal of Economic Theory*, 13:14–25.

Gehrlein, W. V. and Fishburn, P. C. (1978). Coincidence probabilities for simple majority and positional voting rules. *Social Science Research*, 7(3):272 – 283.

Gehrlein, W. V. and Lepelley, D. (2000). The probability that all weighted scoring rules elect the same winner. *Economics Letters*, 66:191–197.

Gibbard, A. (1973). Manipulation of voting schemes: a general result. *Econometrica*, 41:587–601.

Hare, T. (1857). *The machinery of representation*. Oxford University.

Hastie, R. and Kameda, T. (2005). The robust beauty of majority rules in group decisions. *Psychological Review*, 112:494–508.

List, C. and Goodin, R. E. (2001). Epistemic Democracy: Generalizing the Condorcet Jury Theorem. *Journal of Political Philosophy*, pages 277–306.

Merrill, S. I. (1984). A comparison of efficiency of multicandidate electoral systems. *American Journal of Political Science*, 28:23—48.

Merrill, S. I. (1985). A statistical model for condorcet efficiency based on simulation under spatial model assumptions. *Public Choice*, 47:389—401.

Mueller, D. C. (2003). *Public Choice III*. Cambridge University Press, Cambridge, UK.

Niemi, R. (1970). The occurrence of the paradox of voting in university elections. *Public Choice*, pages 91—100.

Regenwetter, M., Grofman, B., Marley, A., and Tsetlin, I. (2006). *Behavioral Social Choice*. Cambridge University Press, Cambridge, UK.

Regenwetter, M., Grofman, B., Popova, A., Messner, W., Davis-Stober, C. P., and Cavagnaro, D. R. (2009). Behavioural social choice: A status report. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 364:833–843.

Regenwetter, M., Kim, A., Kantor, A., and Ho, M.-H. (2007). The unexpected empirical consensus among consensus methods. *Psychological Science*, pages 559–656.

Regenwetter, M., Marley, A. A. J., and Grofman, B. (2002). A general concept of majority rule. *Mathematical Social Sciences: special issue on random utility theory and probabilistic measurement theory*, 43:407–430.

Regenwetter, M. and Rykhlevskaia, E. (2007). A general concept of scoring rules: general definitions, statistical inference, and empirical illustrations. *Social Choice and Welfare*, 29(2):211–228.

Riker, W. H. (1982). *Liberalism v. Populism*. W. H. Freeman and Co., San Fransisco.

Saari, D. (1994). *Geometry of Voting*. Springer, New York.

Saari, D. G. (1999). Explaining all three-alternative voting outcomes. *Journal of Economic Theory*, pages 313–355.

Saari, D. G. (2000). Mathematical structure of voting paradoxes 1: Pairwise vote. *Economic Theory*, pages 1–53.

Satterthwaite, M. (1975). Strategy-proofness and Arrow's conditions:existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, 10:187–217.

Sen, A. K. (1970). *Collective Choice and Social Welfare*. Holden-Day, San Fransisco.

Shepsle, K. and Bonchek, M. (1997). *Analyzing Politics*. Norton, New York.

Tideman, N. (2006). *Collective decisions and voting: the potential for public choice*. Ashgate Publishing, Ltd.

A Rules list

Most of the rules are based on the concept of *generalized rank*, introduced in Regenwetter and Rykhlevskaia (2007). The representation of preferences in the form of generalized ranks is referred to as the *generalized rank matrix*. When the preferences are linear orders, the generalized rank is a simple ranking of the cadidates, with 1 corresponding to the best candidate and n corresponding to the worst candidate out of n. The generalized rank assigns noninteger ranks in case of indifference.

A.1 The Plurality Rule

The matrix of binary relations is transformed into a generalized rank matrix.

$$\text{Plurality score(Candidate)} = \begin{cases} 1, & \text{GR(Candidate)=1;} \\ 0, & \text{otherwise.} \end{cases}$$

The total score is the sum of all Plurality scores for a given candidate over all rows, representing voter preferences. The winner is the candidate with the maximum total score.

A.2 The Borda Rule

The matrix of binary relations is transformed into a generalized rank matrix. Each candidate has a generalized rank for any given binary preference row. Here and everywhere else n is the total number of candidates.

$$\text{Borda score(Candidate)} = n - \text{GR(Candidate)}$$

The total score is the sum of all Borda scores for a given candidate over all voters. The winner is the candidate with the maximum total score.

A.3 The Condorcet Rule

This concept is presented in detail in Regenwetter et al. (2006). Choice alternative X is majority preferred to choice alternative Y according to the Condorcet Rule if the majority of people prefer alternative X to alternative Y, i.e., if the number of relations containing $X \succ Y$ exceeds the number of relations containing $Y \succ X$. A candidate is the Condorcet winner if she beats each competitor by the Condorcet criterion. The

Condorcet paradox denotes the fact that the Condorcet winner may not exist. While the Condorcet criterion is generally viewed as the most natural implementation of democratic aggregation, a substantial theoretical literature has predicted an omnipresent threat of the Condorcet Paradox, and leading textbooks strongly caution against using this method (e.g., Shepsle and Bonchek (1997)). The Condorcet rule is rarely used in multi-candidate elections.

A.4 The Anti-Plurality Rule

The matrix of binary relations is transformed to a generalized rank matrix. Each candidate has a generalized rank for any given binary preference row.

$$\text{Anti Plurality score(Candidate)} = \begin{cases} 1, & \text{GR(Candidate)=n;} \\ 0, & \text{otherwise.} \end{cases}$$

The total score is the sum of all Anti-Plurality scores for a given candidate over all voters. The winner is the candidate with the minimum total score.

A.5 The Single Transferable Vote Rule

The STV rule provides the STV winner (for a one seat election) and the STV social order.

The matrix of binary relations is transformed to a generalized rank matrix. Each candidate has a generalized rank for a given binary preference row.

The STV Rule consists of 4 steps:

1. Calculate the Droop quota (DQ) for a particular number of seats;

$$DQ = \frac{\text{Number of Voters}}{\text{Number of Seats} + 1} + 1$$

2. Apply the Plurality rule and define the set of winners;

$$\text{Plurality score(Candidate)} \begin{cases} > DQ, & \text{Candidate is the winner;} \\ < DQ, & \text{Candidate is not a winner.} \end{cases}$$

3. If the number of winners is less than the number of seats, eliminate one candidate with the minimum Plurality score. Ties are broken at random.

4. Readjust voter preferences for (n-I) candidates and repeat steps 2 through 4 until the set of winners is defined. Here and everywhere else I is the number of eliminated candidates.

The STV order is defined by the sequence of additional candidates to get a seat as the number of seats increases. For example, for five candidates the one seat election winner is candidate A, for a two seat election the winners are candidates A and D; for a three seat election the winners are candidates E, D, and A; for four seats the winners are C, D, E, and A. Then, the social order is $A \succ D \succ E \succ C \succ B$ (we omit the $\succ$ sign hereafter).

A.6 The Coombs Rule

The rule consists of 4 steps:

1. Calculate the Droop quota (DQ) for a particular number of seats;

$$DQ = \frac{\text{Number of Voters}}{\text{Number of Seats} + 1} + 1$$

2. Apply the Plurality rule and define the set of winners;

$$\text{Plurality score(Candidate)} \begin{cases} > DQ, & \text{Candidate is the winner;} \\ < DQ, & \text{Candidate is not a winner.} \end{cases}$$

3. If no candidate (or fewer candidates than the number of seats) exceeds the Droop quota, eliminate the candidate with the maximum Anti Plurality score. Ties are broken at random.

4. Readjust voter preferences for (n-I) candidates and repeat steps 2 through 4 until the set of winners is defined. (For each next step the criterion for Anti Plurality changes.)

$$\text{Anti Plurality score(Candidate)} = \begin{cases} 1, & \text{GR(Candidate)=n+1-I;} \\ 0, & \text{Otherwise.} \end{cases}$$

The order is defined similarly to the STV rule.

A.7 The Plurality Runoff Rule

The rule consists of 4 steps:

1. Calculate the Droop quota (DQ);

$$DQ = \frac{\text{Number of Voters}}{2}$$

2. Apply the Plurality rule, calculate the plurality score and define the winner;

$$\text{Plurality score(Candidate)} \begin{cases} > DQ, & \text{Candidate is the winner;} \\ < DQ, & \text{Candidate is not a winner.} \end{cases}$$

3. If there is no winner, eliminate all but two candidates with the highest plurality score;

4. Readjust voter preferences for the 2 remaining candidates and apply the Plurality rule.

The Plurality Runoff rule selects a winner only. To define loser coincidence, we assume that all candidates but the Plurality Runoff winner form a set of Plurality Runoff losers. If the loser of the other rule is in this set, we count that as agreement on losers.

A.8 The Random Rule

This rule picks a voter on random and uses his preference as a social choice outcome. This is similar to the dictator rule, but since the dictator is assigned on random, every voter has an incentive to report her profile truthfully. The negative side of this rule is, obviously, that it does not provide a deterministic answer.

A.9 Mean Generalized Rank

The binary data is transformed into the generalized rank matrix. The generalized rank for each candidate is averaged across voters. The candidates are ordered according to those averages. The winner is the candidate with the smallest average.

A.10 Median Generalized Rank

The binary data is transformed into the generalized rank matrix. For each candidate, a median rank is calculated. The candidates are ordered according to those medians. The winner is the candidate with the smallest median.

A.11 Mode Generalized Rank

The binary data is transformed into the generalized rank matrix. For each candidate, the most probable rank is calculated (the one that is given by the largest proportion of voters). The candidates are ordered according to those ranks. The winner is the candidate with the smallest rank.

A.12 Mode

The binary data is transformed into the generalized rank matrix. The most frequent ordering is chosen as a social choice, and the social rank coincides with the ranking in this ordering. The winner is the candidate with the smallest rank.