DISAGREEMENT OF COUNTING METHODS IN SIMULATED RANKED PREFERENCE ELECTIONS

BRANDON CALIA AND JEROMY SIVEK

Abstract. This study addresses the fact that reasonable methods for determining the winner(s) of ranked choice elections can disagree with each other. We consider six well-known methods and check their agreement using a variety of simulation strategies. We explain some of the underlying causes of potential disagreement using examples. Most centrally, we attempt to determine the probability that the choice of counting method itself will determine the result of an election. In considering different models for voter behavior, we arrive at a range of numbers.

1. Introduction

In ranked choice voting, voters are given a set of candidates which they may rank in preferential order. Given a set of ballots, there exist various methods to determine the winner of the election. These methods need not pick the same winner. Methods that seem perfectly fair can produce results that appear unreasonable. This creates an issue - if different (seemingly fair) methods are picking different winners, then it can be difficult to settle on a single method for continued use.

Perhaps the most famous encapsulation of this issue is due to Kenneth Arrow [1]. Arrow's theorem has been widely discussed and we draw some clarity from Amartiya Sen [15].

Informed by the unavailability of a perfect decision-making method, this study aims to help clarify the differences between reasonable imperfect methods. We include Condorcet, Plurality, Instant-Runoff, Borda, Baldwin, and Range methods. Where the meanings of these methods are ambiguous (for example, when ballots are incomplete), we provide methodological details.

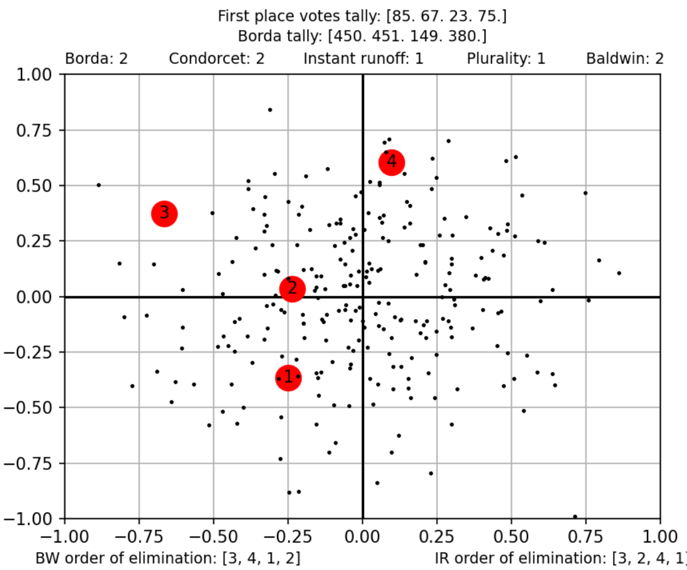
We focus on the frequency that these methods agree in simulated elections. Where possible, we attempt to explain the causes of the disagreements with an eye toward classifying such. We use four significantly different simulation strategies (and minor variations therein). Simulation strategies used here include simple and hybrid spatial models which are known to mirror real-world election dynamics [14], [12], [5]. As shown below, we found some expected and unexpected patterns in the agreement frequencies.

Before working through our simulation strategies in order of increasing complexity, we offer one helpful example here from the mixed spatial model described in section 3. Using a scatter plot of candidates and voters, we can see the location of each candidate in comparison to every voter's ideal point.

The example just below includes one of the most common method disagreements seen in the real elections we tested. The black dots represent voters, and the red circles represent candidates. The method used to generate their positions and the voters' ballots will be explained in section 3.

In this example, Instant Runoff and Plurality both pick a different winner than the rest of the methods. Both of these methods focus on first place votes, whereas other methods considered in this paper attempt to consider all preference information simultaneously. Candidate 2 (the Condorcet winner) gets fewer first place votes than 1 and 4, likely because it shares a quadrant with Candidate 3. In a pairwise match-up, however, 2 beats 1. This last fact is possibly because 2 is located between the other candidates, getting a large portion of the second place votes.

Date: October 17, 2023.



As mentioned above, we will proceed through simulation models in order of increasing complexity. We will pause at times to discuss what a given model can show us about why simple or advanced counting methods may disagree in a potential election.

We will see some patterns holding across models: that the more advanced methods agree more frequently with each other, that instant runoff agrees more frequently with plurality than other true ranked choice methods, and the frequency that even advanced methods disagree is non-trivial. We consider the data to support the assertion that using an advanced counting system is worth the consideration of political bodies. We hope that the collection of data presented here can help to inform the discussion of these choices.

Data from historical elections [4] informed changes to our simulation methods. First, voters may not rank every candidate. Voters could leave all but one candidate off the ballot, or rank the entire set of candidates. Simulation parameters were set to have this occurring as frequently in simulation as in historical data. And also the methods work differently with complete and incomplete ballots. Accounting for this required modifications to our implementation of the voting methods:

- Borda will now award 1 point to the candidate in last place on a ballot, as opposed to the traditional 0. Since no voter must put every candidate on the ballot, just being on the ballot should be awarded with a point.
- As previously discussed, Baldwin no longer needs to elect the Condorcet winner. This is possible even if a single ballot is left incomplete.
- Using the same logic as the first point, Condorcet should consider being on the ballot a pairwise win against all candidates not on it.
- Borda will now award 0 points to any candidate left off a ballot. It is worth noting that there are many possible solutions to dealing with candidates left off a ballot. Awarding 0 Borda points to those candidates ensures that no voter is forced to award points to any candidate they strongly dislike. However, it is possible that this provides a strategic incentive to submit an incomplete ballot.

2. Spatial Model

We now turn to simulations with more realistic ballot generation techniques. We consider the fact that some candidates (or policy options, etc.) are more similar to each other, so a high ranking for one makes a voter very likely to rank a similar candidate next. This phenomenon can be captured by a spatial model - widely considered the most accurate methods of election simulation [5].

Candidates are placed in multi-dimensional space, where each dimension is an issue, policy, or potential attribute of a candidate. Each voter has a point in this multi-dimensional space that best represents their opinions across multiple issues. A utility function can be used to model a voter's favor towards each candidate. A natural utility function is the euclidean distance, known as the proximity model: [11]

$$U(V, C) = -\sqrt{(v_1 - c_1)^2 + ... + (v_k - c_k)^2}$$

where k is the number of dimensions in the model. The voter then ranks candidates in order of decreasing utility.

We also chose to simulate the possibility of voters leaving incomplete ballots. In the algorithm generating the data present in this section, any candidate whose distance is more than twice as great as the candidate with the smallest distance to the current voter will be left off that voter's ballot.

Finally, we must decide on a number of dimensions to use for our model. Not knowing how this may affect results, we chose to test the effect of using a 2D vs. a 3D model on our results. The data is shown below. The code itself can generate simulations in n-dimensions where n is any number that is small enough for the computational hardware.

4 candidates, 1,000 simulations

	100 voters		**1,000 voters**		**10,000 voters**	
	2D	3D	2D	3D	2D	3D
Condorcet existence	.976	.984	.995	.999	.999	.999
Condorcet-Borda	.895	.935	.921	.951	.941	.956
Condorcet-Baldwin	.876	.907	.895	.916	.899	.930
Condorcet-IR	.914	.948	.942	.968	.937	.966
Condorcet-plurality	.806	.820	.801	.837	.802	.826
Borda-Baldwin	.942	.954	.950	.960	.960	.948
Borda-IR	.852	.921	.876	.920	.883	.923
Borda-Plurality	.753	.793	.745	.792	.756	.787
Plurality-IR	.850	.842	.837	.856	.844	.849
Plurality-Baldwin	.730	.769	.715	.763	.722	.764
Baldwin-IR	.830	.891	.845	.887	.839	.897

6 candidates, 1,000 simulations

	100 voters		**1,000 voters**		**10,000 voters**	
	2D	3D	2D	3D	2D	3D
Condorcet existence	.949	.963	.983	.995	.991	.999
Condorcet-Borda	.830	.855	.833	.903	.855	.895
Condorcet-Baldwin	.766	.813	.745	.858	.774	.855
Condorcet-IR	.831	.898	.843	.910	.838	.914
Condorcet-plurality	.652	.675	.667	.671	.632	.706
Borda-Baldwin	.907	.922	.903	.943	.907	.947
Borda-IR	.755	.817	.705	.829	.727	.818
Borda-Plurality	.586	.627	.562	.602	.545	.619
Plurality-IR	.721	.728	.748	.709	.717	.748
Plurality-Baldwin	.530	.592	.493	.567	.473	.585
Baldwin-IR	.687	.778	.621	.786	.646	.775

3. Mixed Spatial Model

Many have theorized that the Euclidean distance is not the only possible way for voters to determine their utility for a candidate. Beyond proximity, a spatial model with a directional component incorporates the observation [14] that voters can prefer the most extreme candidate on their side of an issue axis, calculating utility with a dot product of voter and candidate position.

Using a purely proximity model or a purely directional model could fail to account for all the possible aspects of voter behavior. Voters may prefer a candidate who is extremely close to them in ideological/ policy space, but might otherwise prefer the more extreme candidate on their side of an issue, but only up to a certain extent. For this reason, a mixed model is believed to be a sophisticated and accurate way to model a voter's utility for a candidate in space. The accuracy of such mixed models has been experimentally verified [10], [12].

We implemented such a "mixed" spatial model by adding a directional term as described above. Informed by the literature [11], we also added a small random term (R) which accounts for possible variations (misunderstanding, personal affinity, etc.) in voter behavior. The proximity and directional terms were calibrated to have roughly equal contributions to the average utility score. The R term was calibrated to contribute much less than other terms in every implementation.

$$U(V, C) = \alpha|V - C| + \beta V \cdot C + R(V, C)$$

Method(s)	Frequency 1*	Frequency 2*
Condorcet existence	.997	.876
Condorcet-Borda	.9230	.731
Condorcet-Baldwin	.933	.754
Condorcet-instant runoff	.8542	.763
Condorcet-plurality	.637	.582
Borda-Baldwin	.947	.854
Borda-instant runoff	.852	.759
Borda-plurality	.688	.735
Plurality-instant runoff	.701	.690
Plurality-Baldwin	.651	.640
Baldwin-instant runoff	.836	.751

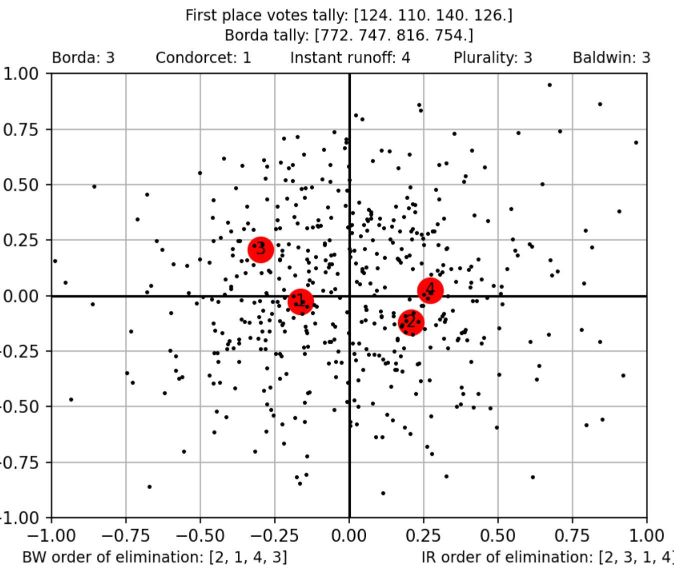
The above table states rates of agreement over two implementations of this model.

In model 1, the spatial axes run from -1 to 1. We found the directional and proximity components to have roughly equal sizes on average and so set $\alpha = \beta = 1$. We set $R(V, C)$ to sample from a uniform distribution on $[.15, .40]$. The data labeled Frequency 1* above comes with these parameters and 5000 runs with variable voter and candidate number. The overall pattern of agreement is similar to the purely proximity-based model (low agreement between plurality and advanced methods, interesting patterns otherwise).

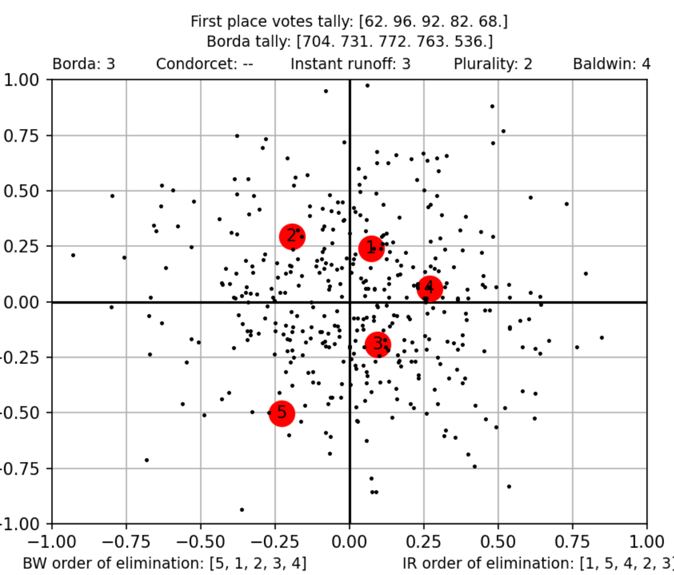
In model 2 ... 5 candidates, 500 voters, 1000 runs ... we set the probability that a voter would rank every candidate to 20%. After checking the historical databases we could access, we found this number to be closer to 8pct in reality [9]. The spatial dimensions were -100 to 100. We set $\alpha = 1$ and $\beta = .05$. $R(V, C)$ was sampled from a normal distribution with $\mu = 0$, $\sigma = 5$.

4. Visualization and Spatial Characteristics of Method Disagreements

Here are a few more examples that show simulations where methods disagree. These examples all used the mixed spatial model for the utility function.



In the above simulation, Condorcet picks a winner not selected by any other methods. Instant runoff also disagrees with plurality, even eliminating the plurality winner early. Borda and Baldwin agree, but Baldwin eliminates 1, the Condorcet winner, quickly. Candidate 1 is poorly placed once the lowest Borda score is eliminated, sharing "territory" with 3 while 4 has a large region to itself. But when all candidates are considered, 1 wins a pairwise comparison against every other candidate.



Finally, above we see the possible role of symmetry in creating the near-ties necessary for methods to disagree. There does not exist a Condorcet winner, instant runoff and plurality pick different winners, and Borda and Baldwin do not agree either. It is worth noting that the three top Borda scores here are very close.

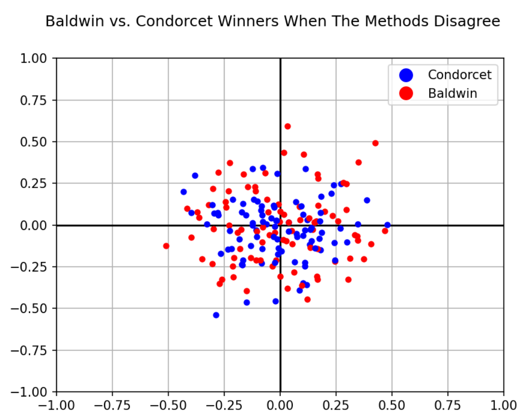
We see some apparent patterns in the winning candidates' positions. Being away from other candidates is most helpful in Plurality, where this leads to less competition for that candidate's neighborhood of first place votes.

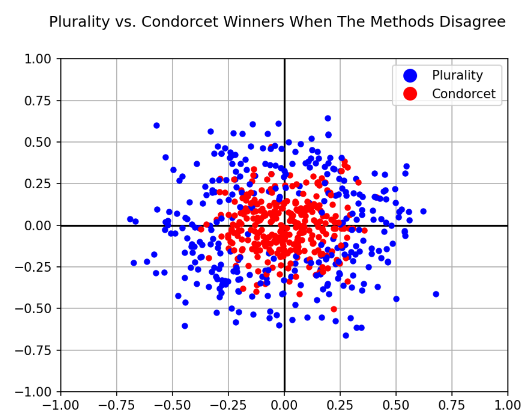
Having an overall better position, even if it is crowded, can be preferable in the elimination methods (Baldwin, IR). Being close to other candidates who are slightly less-well positioned means that those candidates will be eliminated first. And then the candidate who was positioned best with the voters to begin with will have that prime position to themselves.

We have attempted to capture these effects by measuring some numbers related to the spatial positions of method winners: their distance to the origin, their distance to the closest alternative candidate, and their average distance to all other candidates. The table below presents these numbers for 1100 simulations with 5 candidates and 500 voters.

We include information about the Range method (described in section 6). The Range and Plurality methods show the most unique spatial characteristics. We propose that Plurality winners have the largest average distance to the origin because this makes them most likely to have a region (even if less densely populated with voters) free of other candidates that would split that region's first place votes. We do not have as solid of an explanation for the Range method winners having unique spatial characteristics.

	Dist to Origin	Dist to Closest Alternative	Avg Dist to Alternatives
Range	.1928	.4885	.2679
Condorcet	.2177	.5437	.2967
Baldwin	.2243	.5475	.2977
Instant Runoff	.2345	.5543	.3055
Borda	.2290	.5540	.3065
Plurality	.2900	.6136	.3748

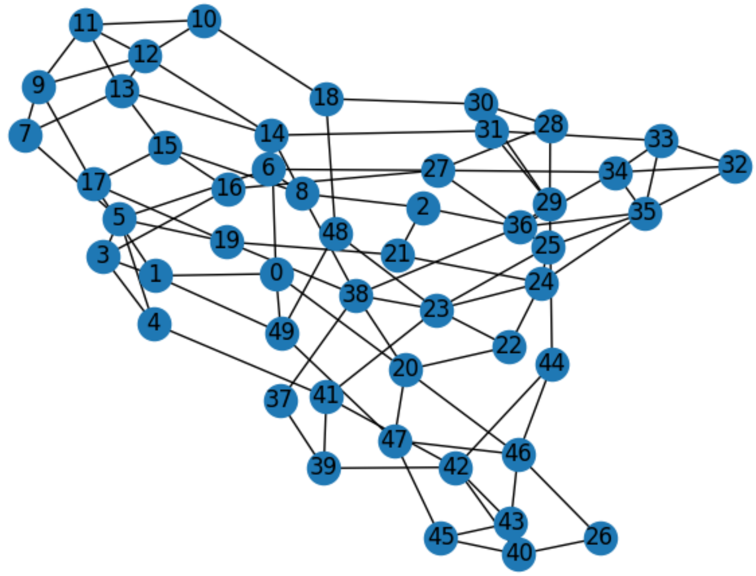




5. Graph Theoretic Voter Model for Ranked Preference Elections

This is just a placeholder paragraph. Every simulation method captures only one or two things. Above captured policy preference. The new model introduced in this section captures the influence of voters upon each other after each voter forms some initial opinions of the candidates. This has been modeled in various ways for 2-candidate elections [13]. We use an *update rule* adapted from pure graph-theoretic *Voter Models* (GTVM) [7], [8]. We believe that such a model for ranked preference elections is original here.

Using a Python script that implemented the NetworkX package, we generated such graphs with various documented algorithms for creating realistic social network models. Namely, we reference the Watts-Strogatz [18] and Barabási-Albert [2] algorithms. The former produces graphs which replicate the properties of small, real-world social networks. The latter generates scale-free networks, with examples including the internet or social media connections.



5.1. **Evolution Toward Consensus.** We believe that results about equilibria in 2-candidate GTVM [6] can be adapted here. As the multi-candidate GTVM is allowed to continue, the state of the election will converge toward and in fact reach a state where every voter has exactly the same ballot. This would lead to a consensus winner.

We conducted an experiment to see which counting methods (Baldwin, Borda, Condorcet, IR, Plurality) would be most likely to agree with the eventual consensus winner. In this initial round of data gathering, we started with a simulated ranked preference election with 5 candidates and 5000 voters using the mixed spatial model described above. We ran such simulations until encountering a case where at least two of the counting methods disagreed. We then ran loops of 10*[Voter Number] individual rolls of the update rule. We stopped the model if after any one of these loops the 5 counting methods all agreed and named their choice the *consensus winner*.

We discovered that this update rule favored Borda winners the most and disfavored Plurality winners the most. To understand this, recall that each Borda point is awarded to candidate X for being above candidate Y on some ballot. The update rule starts by randomly selecting consecutive candidates on a ballot and then polls a randomly chosen adjacent voter's preference between these candidates. The probability that X will be above Y on the randomly chosen adjacent ballot is correlated to their Borda scores. If Y tended to be ranked higher than X, it would be getting as many Borda points as X on those ballots and at least one additional point for being above X on each of those ballots.

The charts below focus on the cases where the eventual consensus winner was also the original winner of method X (where X is named in the upper-left corner of the chart) and displays the probability that method Y (where method Y is the column header) also agreed with that consensus winner.

First, we have results from 10,000 runs where at least two methods initially disagreed and the social network facilitating the updates was created with a Watts-Strogatz algorithm. In 96.40% of cases, the initial Borda winner became the consensus winner. And in 2.69% of cases, the initial Plurality winner became the consensus winner.

Agrees w/	Baldwin	Condorcet	IR	Plurality
Borda	.7854	.7848	.4718	.0215

Agrees w/	Baldwin	Borda	Condorcet	IR
Plurality	.5985	.7695	.5911	.2565

The last number has the potential to be somewhat misleading despite being accurately recorded. Because we are focused only on those cases where there is some initial disagreement and the original plurality winner eventually becomes the consensus winner, this number misses the frequency with which plurality and IR agreed generally. Of the 10,000 cases considered here where at least some methods disagreed, there were 5220 cases where IR and Plurality either both matched the eventual consensus winner or both failed to match the consensus. And as other data indicate, by focusing on disagreements, we are more likely to land on a disagreement involving either IR or Plurality than a more consensus-oriented method like Borda.

Second, we have results from 10,000 runs where the social network facilitating the updates was created with a Barabási-Albert algorithm. In 96.11% of these cases, the original Borda winner was the eventual consensus winner and in 2.86% of cases, the original Plurality winner became the consensus winner.

Agrees w/	Baldwin	Condorcet	IR	Plurality
Borda	.7913	.7904	.4672	.0234

Agrees w/	Baldwin	Borda	Condorcet	IR
Plurality	.6538	.7867	.6538	.2068

6. Range Voting Method

Range voting is similar in some ways to ranked preference voting. Instead of ranking candidates, however, voters provide a score (here between 0 and 99, inclusive) for each candidate. This score is meant to capture the extent to which the voter supports the candidate.

The winner is determined by averaging the scores for each candidate over all ballots. In range voting, a candidate may be left off of a ballot. The candidate's score is only taken over the ballots on which they appear. This is considered necessary for practical reasons (voters may simply not want to fill in every blank or be authentically indifferent to a candidate) and provides a convenient method for including partially damaged or illegible ballots (any score that cannot be read is simply excluded) [16].

The value of ranked choice voting to political communities is that it allows voters to express some of their opinions related to more than one candidate per election. Range voting allows for all of the expression of ranked voting and more, allowing voters to quantify the extent to which they prefer one candidate to another.

Range voting is natural and easily understood despite not being widely used. [3]

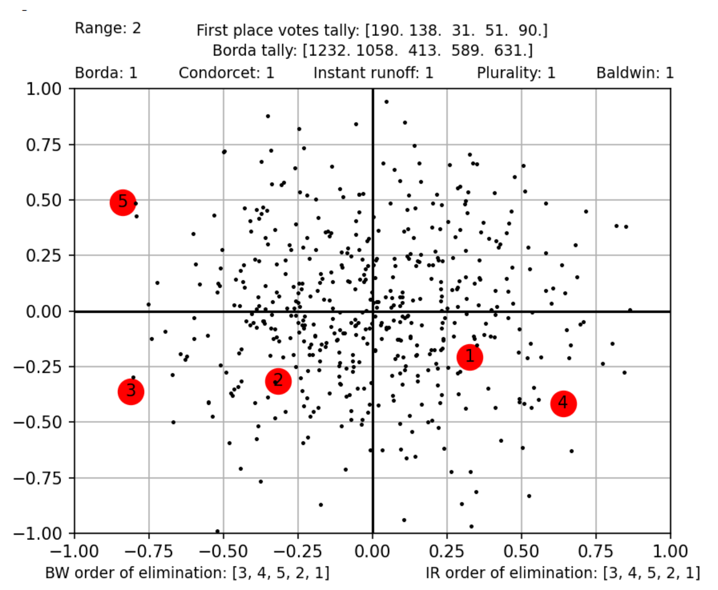
Given the natural link between a given voter's likely range score for a given candidate and the "utility" that voter expects to derive from that candidate's election, it has been shown that range voting exceeds the utility maximization of the previously discussed methods [17].

In our simulations, the Range method disagrees with other methods confusingly often. The table below comes from 500 simulations with 5 candidates and 500 voters. Other runs exhibited similar frequencies.

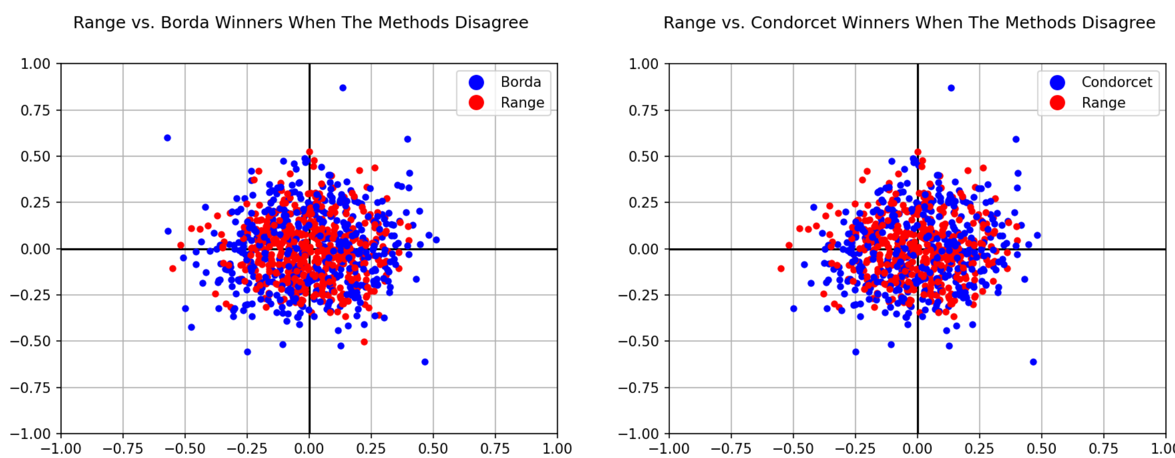
	Condorcet	Borda	Inst Runoff	Baldwin	Plurality
Frequency Agree w/ Range	.570	.530	.496	.566	.356

Above we saw that Range method winners have a unique spatial characteristic: closer on average to the median voter and closer on average to other candidates. We hypothesize that being close to other candidates (in our model and in real-world range voting) is much more advantageous in Range than in any other method. This can be understood by the fact that our model would give close candidates approximately identical Range scores (as a real voter might), whereas Borda and Baldwin force voters to give only one of them a valuable extra point (out of say 5). More dramatically, IR and Plurality would allow voters to support exactly one of the two close candidates (overall in plurality and at any given stage in IR).

Below we see an example where the Range method disagrees with all other methods. We hypothesize that the position of Candidate 5 is significant here.



We saw above that the Range method was much more likely to pick a centrally located candidate than other methods in this model. We see this in the scatterplots below as well. In the spatial characteristic data in section 4, we saw that plurality and Range are outliers in different ways. Range winners tended to be more centrally located as can also be seen below.



7. Uniformly Generated Ballots

This was above. Material on uniform ballots already submitted for publication. Weights array not submitted. Consider transitions.

The simplest model in this study generated ballots uniformly. Each alternative had the same probability of ending up at any given spot on any given ballot. We observed interesting variations in the frequency of agreement between methods as the numbers of candidates and voters changed. In particular, for many quantities of interest, the numbers appear to approach horizontal asymptotes as the number of voters increases.

Note that frequencies describing Condorcet's agreement with other methods are counted only in the cases where there exists a Condorcet winner. If one does not exist, it is not counted as a disagreement.

In the table below, the first row (9050 runs) comes from simulated elections with 100 to 1,000,000 voters and 5 candidates. There were more runs at smaller numbers of voters. We compress some of the data here and attempt to show the effect of changing voter number in graphs below. The second row (4430 runs) averages simulations with 5, 10, 20, and 50 voters.

Agrmnt Frcnqy	Cndrct-Bldwn	Borda-Bldwn	Plurality-IR	Plurality-Bldwn	Bldwn-IR
More Voters	.719	.762	.548	.551	.820
Fewer Voters	.552	.707	.593	.545	.806

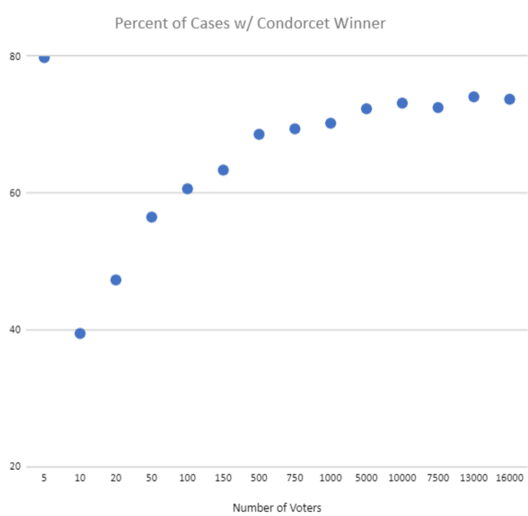
In the next table, we have agreement frequency data from more simulations focusing on simpler methods. The first row (90,000 runs) averages runs with 500 to 16,000 voters. The second row (60,000 runs) averages runs with 5 voters to 150 voters.

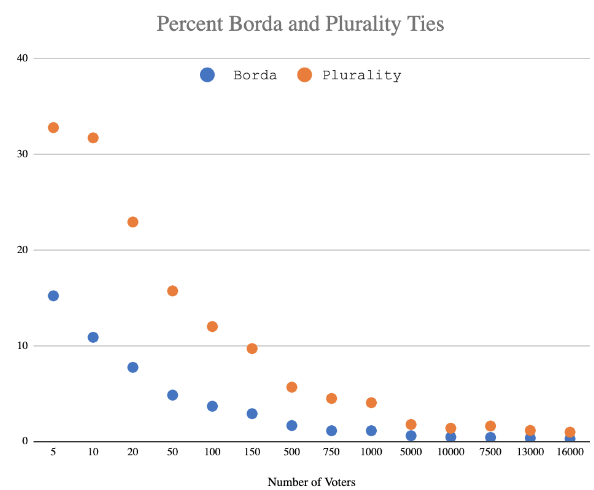
Agrmnt Frcnqy	Condorcet existence	Condorcet-Borda	Condorcet-Plurality	Borda-Plurality
More Voters	.621	.719	.409	.541
Fewer Voters	.578	.540	.388	.569

We note a couple trends in the data.

The frequency with which Plurality disagrees with the more advanced methods is striking. We consider this natural as plurality does not access the information about lower preferences at all. And we take it as evidence that, given the reality that lower preferences exist, the plurality method is not just failing to collect this information but also frequently making a different choice than the one that would be made by a reasonable method that saw this information.

We also note that agreement frequency seems to increase with voter number for most pairs of methods. For many pairs (data in appendix **??**), the frequency of agreement appears to stabilize as voter number increases. We find these potential horizontal asymptotes intriguing, observe them in many quantities in this simulation method, and cannot currently prove that they must exist.





In later sections, we describe simulation methods that attempt to realistically capture the behavior of an electorate. One imagines, for example, that voters may view certain candidates as "similar" and therefore be increasingly likely to rank them consecutively (or nearly consecutively). And the models below capture such behavior and more.

But this uniform model benefits from simplicity. With this study being concerned more with the differing choices made by the counting methods themselves, the uniform model may offer the best single number. If the apparent horizontal asymptotes are provably present for those methods, then the levels of those asymptotes may be the best simple "similarity scores" for pairs of methods.

One simple observed and provable fact is that this method, given k candidates running, generated elections closer to k-way ties as the number of voters increased. We note now that other models produced less-frequent disagreement as voter number increased and that it is easier to generate simple examples where the methods disagree if one starts with a near tie (as captured by Borda score or the number of first place votes, for example).

8. The Weights Array

One possible improvement to generating uniform ballots is to give some candidates more support than others. This may better simulate the circumstances of a real election, where some candidates are more popular, have more effective campaigns, etc. This intuition led to what we called the "weights array".

Each candidate receives a weight ($\{w_1, w_2, ..., w_{k-1}\}$, sampled from normal distribution with $\mu = 1/k$, $\sigma = 0.2/k$, and w_k chosen so that $\Sigma w_j = 1$ if possible, otherwise the run was discarded) which determines their probability of being put on the next available spot on the ballot. For example, if a candidate's weight is 0.4, their probability of ending up first on any given ballot is 0.4, after which the remaining candidates' probabilities are normalized and rolled again for second place.

Contrasting with the uniform method above, we observe that as voter number increases, the methods converge to total agreement, all choosing the candidate with the greatest weight. A proof that this observation must hold can be found after some data generated by this method.

8.1. Data From Weights Method of Simulation.

Below is a sample of data regarding frequency of method agreement using our weights method. (Here C=Condorcet, B=Borda, BW=Baldwin, IR=Instant Runoff, and P=Plurality.)

4 candidates, 1,000 simulations

	C existence	C-B	C-BW	C-IR	C-P	B-BW	B-IR	B-P	P-IR	P-BW	BW-IR
100 voters	.890	.843	.890	.874	.753	.896	.877	.802	.805	.800	.967
1,000 voters	.979	.947	.979	.972	.877	.953	.948	.896	.890	.884	.985
10,000 voters	.998	.989	.998	.998	.967	.990	.990	.969	.969	.969	1.00

5 candidates, 1,000 simulations

	C existence	C-B	C-BW	C-IR	C-P	B-BW	B-IR	B-P	P-IR	P-BW	BW-IR
100 voters	.836	.880	.863	.843	.659	.872	.848	.733	.723	.705	.938
1,000 voters	.981	.941	.981	.973	.864	.948	.945	.871	.871	.890	.989
10,000 voters	.999	.978	.999	.998	.958	.979	.978	.961	.959	.958	.999

6 candidates, 1,000 simulations

	C existence	C-B	C-BW	C-IR	C-P	B-BW	B-IR	B-P	P-IR	P-BW	BW-IR
100 voters	.830	.762	.830	.903	.571	.840	.830	.643	.652	.634	.913
1,000 voters	.983	.921	.983	.977	.835	.929	.925	.839	.841	.843	.989
10,000 voters	.998	.981	.998	.998	.946	.982	.981	.954	.946	.947	.999

8.2. Proof Regarding Weights Method.

This section refers to the claim that under the above weights method, choice functions will always pick the candidate with the greatest weight. In this proof we will show this holds true for Condorcet's method.

Claim: The winner of the weights roll (i.e. the candidate assigned the largest number in the weights array) will win every pairwise match-up.

Proof: Consider N candidates $\{c_1, c_2, ..., c_N\}$ in a simulated election using ranked preference ballots. Let the "weights" roll be $\{w_1, w_2, ..., w_N\}$ where these are the probabilities associated with each candidate for sampling (numpy.random.choice). We suppose that there is a "weights winner" c_m such that $w_m > w_j \ \forall j \neq m$.

In what follows, we fix $j \neq m$ and consider the pairwise matchup between c_j and c_m. We will write $c_j > c_m$ to communicate that c_j is a higher on a given ballot. We aim to show that it must be the case $c_m > c_j$ on more than half of the ballots for sufficiently large voter number. To this end, we define the random variable

$$X_i = \begin{cases} 1 & c_m > c_j \text{ on ballot } i \\ 0 & c_j > c_m \text{ on ballot } i \end{cases}$$

Let $\overline{X} = EV(X_i)$. We will prove below the main argument that $\overline{X} = \frac{w_m}{w_m+w_j}$. Taking this as given for now, we note that $\overline{X} = \dfrac{w_m}{w_m + w_j} > \dfrac{w_m}{w_m + w_m} = \dfrac{w_m}{2w_m} = \dfrac{1}{2}$.

If there are n voters, c_m beats c_j in the election in exactly the case where $\displaystyle\sum_{i=1}^{n} X_i > \frac{n}{2}$ which we note is the same as $\displaystyle\sum_{i=1}^{n} \frac{X_i}{n} > \frac{1}{2}$. The law of large numbers tells us that $\displaystyle\lim_{n\to\infty} \left(\sum_{i=1}^{n} \frac{X_i}{n} \right) = \overline{X} > \frac{1}{2}$.

Let $\epsilon = \dfrac{\overline{X} - 1/2}{2}$, which is greater than 0. By the definition of $\displaystyle\lim_{n\to\infty} \left(\sum_{i=1}^{n} \frac{X_i}{n} \right) = \overline{X}$, $\exists K \in \mathbb{N}$ such that $n > K$ gives

$$\left| \left(\sum_{i=1}^{n} \frac{X_i}{n} \right) - \overline{X} \right| < \epsilon \quad \Rightarrow \quad \overline{X} - \left(\sum_{i=1}^{n} \frac{X_i}{n} \right) < \frac{\overline{X} - 1/2}{2}$$

$$\Rightarrow \quad \overline{X} - \frac{\overline{X} - 1/2}{2} < \sum_{i=1}^{n} \frac{X_i}{n} \quad \Rightarrow \quad \frac{\overline{X} + 1/2}{2} < \sum_{i=1}^{n} \frac{X_i}{n}.$$

But $\overline{X} > 1/2$, giving that $\displaystyle\sum_{i=1}^{n} \frac{X_i}{n} > \frac{1}{2}$ meaning that c_m beats c_j whenever $n > K$ as desired.

To finish the proof, let's confirm that $\overline{X}$ is $\dfrac{w_m}{w_m + w_j}$.

We note first that $\overline{X}$ is also exactly the probability that c_m is ahead of c_j on a given ballot. The value of X_i in this event this is 1. The value of any other event is 0. So the probability of this event is the EV of X_i.

We note that the random.choice method uses the variable p, which we set equal to the weights array, to iteratively determine the candidate in the next ballot position. So at any stage in this iterative process, the matter can be resolved only if the choice of c_m or c_j is made. In any other case, the process continues to a next step.

With the basic conditional probability formula $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$ in hand, we consider the events $A =$ [c_m chosen at stage t] and $B =$ [the process stops at stage t]. We note that $A \cap B = A$ because choosing c_m stops the process. We also note that B is the event that either c_m or c_j is chosen.

In each case, the probability depends on the "leftover" weight in the weights array after the first $t-1$ choices. The probability of choosing c_m is $\frac{w_m}{leftover}$ and the probability of choosing either c_m or c_j is $\frac{w_m+w_j}{leftover}$. From this we see $P(A|B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{\frac{w_m}{leftover}}{\frac{w_m+w_j}{leftover}} = \dfrac{w_m}{w_m + w_j}$. □

This claim shows that the weights roll winner is the Condorcet winner immediately (because c_m wins pairwise matchups with c_j for all j). We expect that small modifications can prove the same for every other counting method as well.

9. Concluding Remarks

After much study, we remain interested in the natural questions about methods disagreeing in real elections and realistic simulations. What "causes" the disagreement? Can we classify candidates (using spatial position or some other characteristic) that are more likely to win one method while losing another?

We remain puzzled by the frequency of difference between Borda and Range winners. In our simulations, both scores were based closely on the underlying spatial score. Varying the parameters and having ballots complete or incomplete all resulted in Range making frequently different choices from other methods.

In advanced simulations and in historical elections, we found a surprisingly high frequency of Condorcet winner existence. This departs from the first data we had from the uniform method. In those simulations, being close to pairwise ties for every pair greatly increased the likelihood that no Condorcet winner would be present. The higher observed frequency of Condorcet winners in reality and realistic simulations suggests that this criterion could be used in some way. (We refer the interested reader to the easily researched method of Duncan Black. The second author has also proposed a method whereby a Condorcet winner that is also a Borda winner would be elected and otherwise the candidate with the lowest Borda score would be eliminated, as in Baldwin's method. In elections with complete ballots, this method always selects the Baldwin winner.)

As a matter of personal opinion, both authors would like to see one of the above methods used over plurality for elections affecting us because of the way they facilitate more meaningful and nuanced expression by voters. We also agree that methods such as Borda, Baldwin, Condorcet, and Range use the information submitted by voters in more reasonable ways than Instant-Runoff. One of the authors prefers the Range method for reasons including the practical value of its simplicity, that is the way that votes are counted being simpler will make it easier for the voting public to understand and therefore accept. Both authors like that any of these methods other than plurality allows for third parties and independents to participate without being "spoilers".

References

[1] Kenneth J Arrow. *Social choice and individual values*, volume 12. Yale university press, 2012.
[2] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.
[3] Antoinette Baujard, Frédéric Gavrel, Herrade Igersheim, Jean-François Laslier, and Isabelle Lebon. Compte-rendu de l'expérimentation des votes par approbation et par note lors des élections présidentielles françaises le 22 avril 2012 à saint-etienne, strasbourg et louvigny. *HAL Open Science*, 2012.
[4] Brandon Calia, Julia Conigliari, Gillian McGuire, and Jeromy Sivek. Comparative analysis of counting methods using documented ranked preference elections. *submitted*, 2023.
[5] James M Enelow and Melvin J Hinich. *The spatial theory of voting: An introduction*. Cambridge University Press, 1984.
[6] Thomas M Liggett. Coexistence in threshold voter models. *The Annals of Probability*, 22(2):764–802, 1994.
[7] Thomas M Liggett. Stochastic models of interacting systems. *The Annals of Probability*, 25(1):1–29, 1997.
[8] Thomas Milton Liggett and Thomas M Liggett. *Interacting particle systems*, volume 2. Springer, 1985.
[9] N. Mattei, T. Walsh, and R. Simon. Preflib: A library for preferences. `https://www.preflib.org/`, 2023.
[10] Samuel Merrill III. A comparison of efficiency of multicandidate electoral systems. *American Journal of Political Science*, pages 23–48, 1984.
[11] Samuel Merrill III. A probabilistic model for the spatial distribution of party support in multiparty electorates. *Journal of the American Statistical Association*, 89(428):1190–1197, 1994.
[12] Samuel Merrill III. Discriminating between the directional and proximity spatial models of electoral competition. *Electoral Studies*, 14(3):273–287, 1995.
[13] Hossein Noorazar. Recent advances in opinion propagation dynamics: a 2020 survey. *The European Physics Journal Plus*, 2020.
[14] George Rabinowitz and Stuart Elaine Macdonald. A directional theory of issue voting. *American political science review*, 83(1):93–121, 1989.
[15] Amartya Sen. *Collective choice and social welfare*. Harvard University Press, 2018.
[16] Warren Smith. Blanks in range voting. Center for Range Voting: `https://www.rangevoting.org/Blanks.html`.
[17] Warren Smith. Range voting. *Center for Range Voting*, `https://www.rangevoting.org/WarrenSmithPages/homepage/rangevote.pdf`, 2000.
[18] Duncan J Watts and Steven H Strogatz. Collective dynamics of 'small-world'networks. *nature*, 393(6684):440–442, 1998.